



PUBLIC

SAP Data Services

Document Version: 4.2 Support Package 14 (14.2.14.0) – 2022-09-17

What's New

Content

1	Introduction to SAP Data Services 4.2 Support Package 14.	6
2	SAP Data Services 4.2 SP 14 features.	7
2.1	Installation, administration, and monitoring.	8
2.2	Connectivity: Big data and CDC.	9
	Support Parquet and ORC for Google Cloud Storage load function.	9
	Support for CDP Private Cloud base.	10
	Support for X.509 authentication for SAP HANA datastore.	10
	User Authentication support for Google Cloud Storage (GCS) file location.	10
	OData Adapter V2 driver update.	12
	Add additional parameters to the end of OData URL requests.	13
	Kerberos authentication for Sybase IQ (SAP IQ).	13
2.3	Functions and transforms.	14
2.4	Integration.	15
2.5	Data Quality transforms.	15
	Global Address Cleanse transform.	16
	Email notifications with directory expiration warnings.	18
	NCOALink MoverID version updated to 6.6.	18
	USPS CASS Cycle O update.	19
3	SAP Data Services 4.2 SP 13 features.	20
3.1	Installation, administration, and monitoring.	20
3.2	Connectivity: Big data and CDC.	22
3.3	Data Quality transforms.	24
	Global Address Cleanse transform.	24
3.4	Functions and transforms.	25
4	SAP Data Services 4.2 SP 12 features.	26
4.1	Upcoming changes.	27
4.2	Installation, administration, and monitoring.	28
4.3	Connectivity: Big Data and CDC.	29
4.4	Data Quality transforms.	33
	Global Address Cleanse transform.	33
	USA Regulatory Address Cleanse.	37
4.5	SAP applications.	38
5	SAP Data Services 4.2 SP 11 features.	40
5.1	Installation, administration, and monitoring.	41

5.2	Usability.	42
5.3	Connectivity: Big data and CDC.	42
5.4	Data Quality transforms.	46
	USA Regulatory Address Cleanse.	46
	Global Address Cleanse transform.	48
6	SAP Data Services 4.2 SP 10 features.	51
6.1	Installation, administration, and monitoring.	52
6.2	Connectivity: Big data and CDC.	53
6.3	Data Quality transforms.	56
	Global Address Cleanse transform.	56
	USA Regulatory Address Cleanse transform.	58
6.4	Platform transforms.	59
	Data Mask transform.	59
6.5	Text Data Processing.	60
7	SAP Data Services 4.2 SP 9 features.	61
7.1	Installation, administration, and monitoring.	61
7.2	Connectivity: Big data and CDC.	62
7.3	Data Quality transforms.	65
	Global Address Cleanse transform.	66
7.4	Usability.	66
8	SAP Data Services 4.2 SP 8 features.	68
8.1	Connectivity: Big Data and CDC.	68
8.2	Data Quality transforms.	70
	Geocoder transform.	70
	Global Address Cleanse transform.	71
8.3	Platform transforms.	72
	DQM Microservices transform.	72
9	SAP Data Services 4.2 SP 7 features.	74
9.1	Connectivity: Big data and CDC.	75
9.2	Data Quality transforms.	80
	Geocoder transform.	80
	Global Address Cleanse transform.	81
	USA Regulatory Address Cleanse transform.	82
9.3	Platform transforms.	83
9.4	Text Data Processing.	83
10	SAP Data Services 4.2 SP 6 features.	87
10.1	Installation, administration, and monitoring.	88
10.2	Connectivity: Big data and CDC.	91

10.3	Functions and transforms.	95
10.4	Usability.	95
10.5	Data Quality transforms.	97
	Data Cleanse transform.	97
	Geocoder transform.	98
	Global Address Cleanse transform.	98
10.6	Platform transforms.	100
11	SAP Data Services 4.2 SP 5 features.	101
11.1	Installation, administration, and monitoring.	102
11.2	Connectivity: Big data and CDC.	103
11.3	Functions and transforms.	105
11.4	Usability.	107
11.5	Data Quality transforms.	107
	Geocoder transform.	107
	Global Address Cleanse transform.	109
11.6	Platform transforms.	111
	Data Mask transform.	111
11.7	Reserved words.	112
12	SAP Data Services 4.2 SP 4 features.	113
12.1	Installation, administration, and monitoring.	113
12.2	Connectivity: Big data and CDC.	114
12.3	Data Quality transforms.	115
	Data Cleanse transform.	115
	Global Address Cleanse transform.	116
12.4	Platform transforms.	117
	Data Mask transform.	117
13	SAP Data Services 4.2 SP 3 features.	119
13.1	Connectivity: Big data and CDC.	119
13.2	Functions and transforms.	120
13.3	Usability.	121
13.4	Data Quality transforms.	122
	Data Cleanse transform.	122
	Global Address Cleanse transform.	123
	USA Regulatory Address Cleanse transform.	129
13.5	Platform transforms.	130
13.6	Data Mask transform.	130
14	SAP Data Services 4.2 SP 2 features.	131
14.1	Connectivity: Big data and CDC.	131
14.2	Usability.	133

14.3	Text Data Processing.	134
14.4	Data Quality transforms.	135
	Data Cleanse transform.	135
	Geocoder transform.	136
	Global Address Cleanse transform.	137
15	SAP Data Services 4.2 SP 1 features.	139
15.1	Installation, administration, and monitoring.	140
15.2	Data Services Workbench.	140
15.3	Connectivity: Big data and CDC.	141
15.4	Text Data Processing.	143
15.5	Adapter SDK.	143
15.6	Data Quality transforms.	143
	Data Cleanse transform.	144
	Geocoder transform.	144
	Global Address Cleanse transform.	145
16	SAP Data Services 4.2 features.	148
16.1	Installation, administration, and monitoring.	149
16.2	Data Services Workbench.	152
16.3	Connectivity: Big data and CDC.	154
16.4	Usability.	155
16.5	Functions and transforms.	155
16.6	Text Data Processing.	156
16.7	Adapter SDK.	157
16.8	Data Quality transforms.	157
	Data Cleanse transform.	158
	Geocoder transform.	161
	Global Address Cleanse transform.	162

1 Introduction to SAP Data Services 4.2 Support Package 14

Welcome to SAP Data Services 4.2 Support Package 14 (version 14.2.14).

This *What's New* document highlights the new features available with this release, and the new features introduced in past releases and service packs.

For important information about this product release including installation notes, known issues, and fixed issues, see the *SAP Data Services Release Notes*.

Before using the latest version of SAP Data Services, see "Data Services behavior changes" in the *Upgrade Guide*.

Documentation

To obtain the latest version of the Data Services documentation, visit the SAP Help Portal at https://help.sap.com/viewer/p/SAP_DATA_SERVICES.

Related Information

[SAP Data Services 4.2 SP 14 features \[page 7\]](#)
[SAP Data Services 4.2 SP 13 features \[page 20\]](#)
[SAP Data Services 4.2 SP 12 features \[page 26\]](#)
[SAP Data Services 4.2 SP 11 features \[page 40\]](#)
[SAP Data Services 4.2 SP 10 features \[page 51\]](#)
[SAP Data Services 4.2 SP 9 features \[page 61\]](#)

2 SAP Data Services 4.2 SP 14 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP14.

The following links take you to descriptions of the features by category for version 4.2 SP14.



[Installation, administration, and monitoring \[page 8\]:](#)

- [BI/IPS 4.2 SP8 Patch 1 support \[page 8\]](#)
- [Certifications \[page 8\]](#)
- [Management Console \[page 9\]](#)



[Connectivity: Big data and CDC \[page 9\]:](#)

- [SAP Data Warehouse Cloud \[page 9\]](#)
- [Server-based connection to Amazon Redshift datastore \[page 9\]](#)
- [Support Parquet and ORC for Google Cloud Storage load function \[page 9\]](#)
- [Support for CDP Private Cloud base \[page 10\]](#)
- [Support for X.509 authentication for SAP HANA data-store \[page 10\] \(Patch 16\)](#)
- [User Authentication support for Google Cloud Storage \(GCS\) file location \[page 10\]](#)
- [OData Adapter V2 driver update \[page 12\]](#)
- [Add additional parameters to the end of OData URL requests \[page 13\]](#)



[Functions and transforms \[page 14\]:](#)

- [Get_BatchJob_By_TimeRange \[page 14\]](#)
- [Support ANSI concatenation semantics \[page 14\]](#)



[Integration \[page 15\]:](#)

- [REST Web services supports server and client authentication \[page 15\]](#)



[Data Quality transforms \[page 15\]:](#)

- [Global Address Cleanse transform \[page 16\]](#)
- [Email notifications with directory expiration warnings \[page 18\]](#)
- [NCOALink MoverID version updated to 6.6 \[page 18\]](#)
- [USPS CASS Cycle O update \[page 19\]](#)

2.1 Installation, administration, and monitoring

SAP Data Services 4.2.14 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

2.1.1 BI/IPS 4.2 SP8 Patch 1 support

Data Services 4.2.14 supports the following platforms:

- SAP BusinessObjects Business Intelligence platform (BI) 4.2 SP8 Patch 1
- SAP BusinessObjects Information platform services (IPS) 4.2 SP8 Patch 1

For more detailed information about the compatibility between Data Services and BI or IPS, see SAP Note [1740516](#).

2.1.2 Certifications

Data Services 4.2.14 adds support for the following certifications:

Category	Certification
Operating systems	<ul style="list-style-type: none">• Solaris 11.4• AIX 7.1 Power TL5 and AIX 7.2 Power TL3• SuSE 12 SP4• Windows 10 1909• Windows Server 2019• Red Hat 7.7
Databases	<ul style="list-style-type: none">• Oracle 19c• Teradata 16.20 on Windows, UNIX, and Linux Also see SAP Note 2936267.• MySQL 8.0 Server• Microsoft SQL Server 2019 (Cluster mode is not supported)
Applications	BW4HANA On-Premise
Cloud-based Databases	Amazon Web Services RDS SQL Server 2017 on EC2
HADOOP and its dependent databases	Hortonworks HDP 3.1

Certification information for SP14 patches can be found in SAP Note [2942023](#).

For more detailed information, see the Product Availability Matrix document located at <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

2.1.3 Management Console

Adobe Flash has been removed from SAP Data Services Management Console.


The end of Flash support is December 2020.

2.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP14 includes connectivity enhancements.

2.2.1 SAP Data Warehouse Cloud

Data Services 4.2.14 supports connectivity to the SAP Data Warehouse Cloud.

For more information, see SAP Note [2907747](#) .

2.2.2 Server-based connection to Amazon Redshift datastore

You can now create a server-based connection to an Amazon Redshift datastore.

For more information, see “Amazon Redshift datastores” in the *Data Services Supplement for Big Data*.

2.2.3 Support Parquet and ORC for Google Cloud Storage load function

With SAP Data Services version 4.2.14 Patch 4, use the Parquet or ORC file formats with the `load_from_gcs_to_gbq` built-in function.

Use the `load_from_gcs_to_gbq` function to transfer data from a Google Cloud Storage into Google BigQuery tables.

For more information about the function, see the *Reference Guide*.

2.2.4 Support for CDP Private Cloud base

CDP (Cloudera Data Platform) Private Cloud Base enables you to have the speed of cloud processing with consistent, on-premise security and governance.

Use SAP Data Services to access Hive and Impala data on your CDP Private Cloud Base. Support is for both Windows and Linux platforms.

Support for CDP Private Cloud Base begins with SAP Data Services version 4.2 SP14 Patch 6 and CDP version 7.1. The required version for the ODBC driver for Hive and Impala begins with version 2.6.9.

For more information about how Data Services supports CDP Private Cloud Base, see the *Supplement for Hadoop*.

2.2.5 Support for X.509 authentication for SAP HANA datastore

New in SAP Data Services 4.2.14 Patch 16, the SAP HANA database datastore now supports X.509 authentication. X.509 authentication is a more secure method of accessing SAP HANA than user name and password authentication.

Include X.509 authentication when you configure an SAP HANA datastore. When you use X.509 authentication, consider the following information:

- Use with or without SSL protocol.
- Applicable for both DSN and server named (DSN-less) connections.
- Uses the same Cryptographic libraries and global.ini settings as SSL protocol.
- Requires an X.509 key store file, which contains the following:
 - The X.509 client certificate.
 - The SSL server certificate, only when you use SSL encryption and you select to validate the server certificate in Data Services.




i Note

X.509 authentication is applicable for SAP HANA Server on premise 2.0 SP05 revision 56 and above and SAP HANA Cloud. For SAP HANA client, X.509 is applicable for SAP HANA ODBC client 2.7 and above.

2.2.6 User Authentication support for Google Cloud Storage (GCS) file location

Data Services now supports OAuth 2.0 User Authentication for GCS file location.

The following options were added to the GCS file location in Data Services 4.2 SP14 Patch 16:

<i>OAuth Mechanism</i>	<p>Specifies the authentication mechanism.</p> <ul style="list-style-type: none"> • <i>Service Authentication</i> (default): Data Services holds the credentials of a service account to complete authentication. • <i>User Authentication</i>: Data Services obtains credentials from the end user. You are required to sign into Google to complete the authentication if you don't have a refresh token. You can also use an existing refresh token. When <i>User Authentication</i> is specified, you must also populate <i>Client ID</i>, <i>Client Secret</i>, and <i>Refresh Token</i>.
<i>Client ID</i>	<p>Specifies the OAuth client ID for Data Services. To get the client ID, go to the Google API Console (https://console.developers.google.com/ .</p> <p>For more detailed instructions, see the Google documentation.</p>
<i>Client Secret</i>	<p>Specifies the OAuth client secret for Data Services. To get the client secret, go to the Google API Console (https://console.developers.google.com/ .</p> <p>For more detailed instructions, see the Google documentation.</p>
<i>Refresh Token</i>	<p>Specifies the refresh token that is required for the Google Cloud Storage file location connection when <i>OAuth mechanism</i> is set to <i>User Authentication</i>.</p> <p>To enter an existing refresh token, do the following:</p> <ol style="list-style-type: none"> 1. Click the ellipsis (...) next to the <i>Refresh Token</i> field. 2. In the <i>Get Refresh Token</i> window, select <i>Refresh Token</i>. 3. Click <i>OK</i>. The refresh token appears (masked for privacy and security reasons) in the <i>Refresh Token</i> field . <p>To generate a new refresh token, enter an authorization code.</p> <ol style="list-style-type: none"> 1. Click the ellipsis (...) next to the <i>Refresh Token</i> field. 2. In the <i>Get Refresh Token</i> window, select <i>Authorization Code</i>. 3. Enter an authorization code. To get an authorization code, follow the instructions in the Google documentation (https://developers.google.com/identity/protocols/oauth2/native-app#step-2:-send-a-request-to-googles-oauth-2.0-server . 4. Click <i>OK</i>. A new refresh token is generated and appears (masked for privacy and security reasons) in the <i>Refresh Token</i> field .

For more information, see “GCS file location option descriptions” in *Data Services Supplement for Big Data*.

2.2.7 OData Adapter V2 driver update

OData V2 now supports server-side pagination and batch processing, which can improve performance when extracting and loading data.

Server-side pagination

Starting in SAP Data Services SP14 Patch 19, OData V2 uses the Apache Olingo 2.0 library. The Apache Olingo 2.0 library supports only server-side pagination when extracting data. Prior to Data Services SP14 Patch 19, OData V2 used the odata4j library, which supported client-side pagination.

! Restriction

The loader actions CREATE LINK, UPDATE LINK, and DELETE LINK are not supported once you upgrade to the OData V2 driver that uses the Apache Olingo 2.0 library.

i Note

The *Number of concurrent threads* option was removed from the OData Source object configuration as it is no longer needed for V2.

To refresh the list of source and target options for an existing job, you need to edit an existing datastore or create a new datastore. If you edit an existing datastore, you don't need to re-import the table into the datastore to refresh the options, but you do need to re-import the table you are using as a source or target into the data flow.

For more information about choosing a driver version, see “OData adapter datastore configuration options” in the *Supplement for Adapters*.

For information about client-side pagination, see “OData pagination” in the *Supplement for Adapters* guide.

Batch processing

OData V2 now supports batch processing, allowing you to send multiple records in a single \$batch HTTP request. To use batch processing, set the *batchSize* parameter in the OData Target object configuration to a value greater than 1. To use non-batch mode, set *batchSize* to 1 (default) in new and existing jobs.

i Note

You must use a \$batch request and have a \$batch endpoint to load data using the create, update, merge, and delete loader actions in batch mode.

For V2 batch processing, a new [Loader action](#) option named [upsert \(IF-MATCH = *\)](#) was added in the OData target object configuration. The [upsert \(IF-MATCH = *\)](#) option supports the UPSERT function for endpoints that support the if-match = * header.

For more information, see “OData as a target” in the *Supplement for Adapters* guide.

2.2.8 Add additional parameters to the end of OData URL requests

Starting in Data Services 4.2 SP14 Patch 19, you can use the new [OData URL suffix adapter](#) option to add additional parameters to the end of all OData URLs that are sending requests.

For instance, you can add a URL parameter that helps route a connection to the correct SAP ERP system client. For example, you might add `sap-client=001`.

The parameter is then added to the end of all the request URLs either as `?sap-client=001` or `&sap-client=001`, based on existing parameters before it.


If you need to add more than one parameter, you must provide input (for example, `"param0¶m1¶m2"`).

This option is supported for OData V4 only and will be ignored if you use OData V2.

For more information, see the “OData adapter datastore configuration options” topic in the *Supplement for Adapters*.

2.2.9 Kerberos authentication for Sybase IQ (SAP IQ)

Starting in SP14 Patch 23, SAP Data Services supports Kerberos network authentication protocol for Sybase IQ (now called SAP IQ) on Windows and Linux.

To use Kerberos authentication for an SAP IQ connection, ensure that you have Kerberos installed and have the required Kerberos information to complete the configuration in the database datastore. Additionally, the Database Administrator must establish a Kerberos Key Distribution Center (KDC) server for authentication. Also, see SAP Note [3248440](#) .

Authentication methods include:

- [User](#) (default): Uses the traditional authentication mechanism (the SAP IQ database user name and password).
- [Kerberos](#): Uses the client principal name (for example, `dba@sapiq.com`) and password instead of the SAP IQ database user name and password.
- [Kerberos keytab](#): Uses the client principal name (for example, `dba@sapiq.com`) instead of the SAP IQ database user name (a password is not required) and the [Kerberos Keytab File](#) location.

For more information about Kerberos login features for SAP IQ, see the [Kerberos user authentication](#) topic in *SAP IQ Administration: User Management and Security*.

For more information about configuring Kerberos authentication for an SAP IQ connection in SAP Data Services, see [Kerberos authentication for Sybase IQ \(SAP IQ\)](#) in the *Designer Guide*.

2.3 Functions and transforms

SAP Data Services 4.2 SP14 includes enhancements to transforms and function support.

2.3.1 Get_BatchJob_By_TimeRange

Use `Get_BatchJob_By_TimeRange` to retrieve a list of all jobs that were running at any point during a specified time range for a repository. For more information, see [Get_BatchJob_By_TimeRange](#) in the *Integrator Guide*.

2.3.2 Support ANSI concatenation semantics

In Data Services 4.2.14 Patch 2, you can change how the SAP Data Services Job Server processes concatenation functions.

Change the Job Server from using the Data Services engine to using ANSI concatenation semantics for the following databases:

- Azure SQL, using `+`
- DB2, using `||`
- Google BigQuery, using `||` or `concat()`
- Microsoft SQL Server, using `+`
- MySQL, using `concat()` or `+`
- Netezza, using `concat()`
- PostgreSQL, using `||`
- Redshift, using `||`
- SAP HANA, using `||` or `concat()`
- SQL Anywhere, using `+`
- Teradata, using `||`

In addition, Data Services pushes down the processing of the concatenation statement for these databases to the database.

With the Data Services engine, a concatenated statement or function returns Null only when all variables in the statement are Null. However, when you configure the Job Server to use the ANSI concatenation semantics, a concatenated statement or function returns Null when any of the variables are Null.

❖ Example

WHERE clause:

The Data Services engine returns `FALSE` for the following Redshift WHERE clause:

```
redshift_string_types.c_nvarchar || null is null
```

However, the ANSI concatenation engine returns `TRUE`.

Function:

The Data Services engine returns 'NOT_ANSI' for the following Redshift function:

```
ifthenelse(redshift_string_types.c_nvarchar || NULL is NULL,  
'ANSI', 'NOT_ANSI')
```

However, the ANSI concatenation engine returns 'ANSI'.

For information about how to enable the ANSI concatenation semantics, see “Changing Job Server options” under the “Executing jobs” section in the *Designer Guide*.

2.4 Integration

SAP Data Services 4.2 SP14 includes the following enhancements to support SAP integrators:

2.4.1 REST Web services supports server and client authentication

Previously, you could configure only server certificate authentication when you created a REST Web services datastore. However, with this release, you can configure your REST Web services datastore for either client or server certificate authentication.

To support the two authentications, we've changed the *Advanced* options for datastore configuration in the following ways:

- Replaced the option *SSL PEM file* with the option *Server Certificate File*
- Added three new parameters for client certificate authentication:
 - *Client Certificate File*
 - *Key File*
 - *Passphrase*

For complete details about REST and SOAP Web services datastores and a description of options, see the *Integrator Guide*.

2.5 Data Quality transforms

Transforms work together to improve the quality of your data. 4.2.14 contains transform enhancements.

[Global Address Cleanse transform \[page 16\]](#)

In SAP Data Services 4.2 SP14, the Global Address Cleanse transform contains enhancements.

[Email notifications with directory expiration warnings \[page 18\]](#)

[NCOALink MoverID version updated to 6.6 \[page 18\]](#)

[USPS CASS Cycle O update \[page 19\]](#)

2.5.1 Global Address Cleanse transform

In SAP Data Services 4.2 SP14, the Global Address Cleanse transform contains enhancements.

2.5.1.1 Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2021 certification.

2.5.1.2 Support for complex primary ranges

The following enhancements apply to all countries, and may be particularly helpful for Estonia, Great Britain, and Ireland:

- Matching to complex primary ranges, including dashed and slashed ranges
- Matching to the first number of a dashed or slashed range
- Matching between a dashed range and a slashed range

For example, this is an exact match to the primary range `PRIMARY_NUMBER 68/1`:

Input		Output	
MULTILINE1		MULTILINE1	68/1 Park Avenue
MULTILINE2	68/1 Park Avenue	MULTILINE2	
MULTILINE3		MULTILINE3	
MULTILINE4		MULTILINE4	
LOCALITY1	Edinburgh	LOCALITY1	Edinburgh
REGION1	Scotland UK	REGION1	City of Edinburgh
POSTCODE		POSTCODE	EH15 1JPEH15 1JP
COUNTRY	UNITED KINGDOM	COUNTRY	GB

2.5.1.3 New Ireland directory

Ireland now has its own self-contained directory called SAP ADDR IRELAND [ga_ie_paf.dir] available for purchase. This directory enables street-level address assignment, which is an improvement over lastline-only assignment from the All World directory. SAP ADDR IRELAND works with Data Services 4.2 SP14 and higher.

i Note

The Ireland directory will be included starting in July in the quarterly International Directory Update letter, which is available on the [SAP Help Portal](#).

Since most delivery points in Ireland have a unique postcode, this release assigns the full address when the input includes only a postcode, a postcode and primary range, or a postcode and unit number. For example:

Input	Output
7	Hotel Saint George
D01 E176	7 Parnell Square East
IE	Dublin 1
	County Dublin
	D01 E176
	IE

For technical details about the directory SAP ADDR IRELAND, see [Directory listing and update schedule](#) in the *Installation Guide* and [Country coverage](#) in the *Reference Guide*.

2.5.1.4 China Geocoding directory

A new parcel level geocoding directory called SAP GEO DIR CHINA - HERE [geo_cn_nt.dir] is available.

For China geocoding, locality description is required. Use LOCALITY1_FULL and LOCALITY2_FULL from GAC output fields for mapping since they contain locality descriptions.

For details, see [Geocoder input fields](#) in the *Reference Guide* and [Directory listing and update schedule](#) in the *Installation Guide*.

2.5.1.5 Greece directory change

Data Services SP14 Patch 11 and higher provides an improved Greece Address Directory from a different vendor, specifically Universal Postal Union.

SAP recommends that you upgrade to 4.2.14.11 and higher to automatically receive the new directory.

If this cannot be done, you can obtain and extract the download package SAP ADDR GREECE, then copy the new `ga_parse.dct` to the `DataServices\DataQuality\gac` folder. However, if you update the directory and not the software, you could see unexpected results and lost assignments.

2.5.1.6 South Korea Geocoding directory

Effective July 2021, we support Geocoding for South Korea (often referred to as Korea (the Republic of)).

To ensure the most accurate assignment, install the SAP GEO DIR SOUTH KOREA directory package (`geo_kr_kl.dir`) at your earliest convenience.

For more detailed information, see the South Korea GEO Directory Update letter at https://help.sap.com/viewer/product/ADDRESSING_DIRECTORIES/2021.07/en-US.

2.5.1.7 Hong Kong International address directory

Starting in SP14 Patch 18, we support the Hong Kong address directory.


To ensure the most accurate assignment, install the SAP ADDR HONG KONG 4.X directory package (`ga_hk_paf.dir`) at your earliest convenience.

For more detailed information, see the International Directory Update letter at https://help.sap.com/viewer/product/ADDRESSING_DIRECTORIES/2022.04/en-US.

2.5.2 Email notifications with directory expiration warnings

Administrators can configure Data Services to send email notifications to users when address directories running in certified mode will expire in the near future.

For more information, see [Configuring e-mail notifications for directory expirations](#) in the Designer Guide.

In addition, we created a Directory Expiration Blueprint for you. For information about how to download and use the blueprint, see the *Data Quality Management Directory Expiration Blueprint User's Guide* on the [SAP Community Network](#) .

2.5.3 NCOALink MoverID version updated to 6.6

SAP's NCOALink solution, MoverID, has been updated.

- In 4.2 SP14 Patch 8, the version was updated from 6.4 to 6.5.
- In 4.2 SP14 Patch 16, the version was updated from 6.5 to 6.6.

i Note

The USPS is requiring every NCOALink user to re-certify using an SAP software version (4.2 SP14 Patch 16 or higher) that contains the fix for this issue.


All NCOALink customers must move to one of the latest supported packages with MoverID 6.6. You then need to re-certify with the USPS.

2.5.4 USPS CASS Cycle O update

The US Postal Service CASS Cycle O implementation is scheduled for August 1, 2023.

Since the USPS is ending support for CASS Cycle N in July of 2023, it is critical that customers who care about CASS changes or utilize the USPS CASS Certified Directories with the USA Regulatory Address Cleanse (URAC) transform do the following:

- Regularly review the SAP US Directory letters and the software documentation (Release Notes, What's New, etc.) for important guidance and release information.
- Be prepared to upgrade to the latest versions of SAP software as indicated in these resources.

The USPS continues to finalize the CASS Cycle O requirements and it has not set final NCOALink and DSF2 certification timelines. SAP will continue to work closely with the USPS to fully support these requirements and deadlines in a timely fashion. For more information about the CASS Cycle O schedule, see <https://postalpro.usps.com/certifications/cass> .

CASS O certification testing by software developers such as SAP will begin in Q2 of 2022. SAP plans to support CASS Cycle O features in a service pack of our upcoming Data Services 4.3 release. We are targeting to release the CASS Cycle O service pack in Q4 of 2022.

AIX and Solaris customers: Data Services 4.3 is planned to be released on Windows and Linux only, so please plan accordingly.

Disclaimer: Release dates are subject to change. SAP does not guarantee the availability of a delivery on the specified date.

3 SAP Data Services 4.2 SP 13 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 13.

The following links take you to descriptions of the features by category for version 4.2.13.



[Installation, administration, and monitoring \[page 20\]:](#)

- Certifications
- Python 3.x support
- Resolution of Connection Failure to a Microsoft SQL Server Contained Database
- RFC server interface logs



[Connectivity: Big data and CDC \[page 22\]:](#)

- Bulk loading for PostgreSQL
- Decrease local disk usage for upload to Google Cloud Storage
- Google BigQuery ODBC datastore
- Azure Data Lake Gen2 support
- SAP HANA proxy connection



[Data Quality transforms \[page 24\]:](#)

- Global Address Cleanse transform
 - Certified New Zealand options
 - Certified Canada options
 - Locality2 support for Estonia
 - Locality2_code support for Czechia



[Functions and transforms \[page 25\]:](#)

- sysutcdate
- New function: to_varchar

3.1 Installation, administration, and monitoring

SAP Data Services 4.2.13 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Certifications

Data Services 4.2.13 adds support for the following:

- BI/IPS 4.2 SP7
- Informix 12.1
- Citrix XenApp 7.8
- Solaris 11.3
- S/4 HANA 1909
- Oracle 18c

i Note

Data Services 14.2.13 no longer supports the `ojdbc6.jar` file. The `ojdbc8.jar` file is the only file Data Services 4.2.13 and higher supports, regardless of the Oracle server version.

For more information, see SAP Note [1741042](#).

For more detailed information, see the Product Availability Matrix document located at: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Python 3.x support

Data Services 4.2.13 now supports Python 3.x.

! Restriction

The new version of Python is not backwards compatible. Because of this, you must update any 2.x Python code being used in User-Defined or Match Best Record transforms before you can run data flows in Data Services 4.2.13 or higher.

In the Match Best Record transform, you must also edit each job that contains a standard strategy. To do this, navigate to the [Best Record Editor](#) and re-select an option from the [Best Record Strategy](#) drop-down. Once you save and re-run the job, the software will automatically update the 2.x Python code to the new 3.x Python code.

To update Python 2.x code to 3.x code, we recommend using the 2to3 conversion utility that ships with Data Services. The utility is located in `%LINK_DIR%\Data Services\DataQuality\python\Tools\scripts`. For detailed information about this utility, see the Python documentation.

High-level instructions for using this utility:

1. Identify User-Defined or Match Best Record transforms that use Python 2.x code. In this example, let's assume you need to update Python code in a User-Defined transform.
2. In the Data Services Designer, open the data flow that contains the transform.
3. Right-click on the transform and select [User-Defined Editor](#).
4. Click [Launch Python Editor](#) to access the [Python Expression Editor](#).
5. Copy and paste the Python 2.x code into a `.py` file. For example, `<example>.py`.
6. From the `%LINK_DIR%\Data Services\DataQuality\python` directory, run the 2to3 conversion utility. To run the utility, type the following:

```
python Tools\scripts\2to3 <example>.py -w
```

The `-w` in the command means that the tool will write the new code into `<example>.py` and save the original code to `<example>.py.bak`.

7. Copy the updated Python code from the `<example>.py` file and paste it back into the *Python Expression Editor*.
8. Run the data flow and confirm the new 3.x Python code works as needed.

Resolution of Connection Failure to a Microsoft SQL Server Contained Database

When you use a Microsoft SQL Server contained database, Data Services no longer generates an error message regarding a connection failure when attempting to log in.

RFC server interface logs

Two fields - *RFC Server Number of Logs* and *RFC Server Log Size in MB* - have been added to the RFC Server Configuration window in the Management Console so administrators can specify the number of logs to maintain for each RFC server interface configuration as well as the maximum file size of those logs. The default settings for these new fields are 3 logs with a maximum file size of 1 MB each, which administrators can then change as needed. This provides more flexibility for retaining important log information; previously, the system contained a fixed three logs that would be serially overwritten when the maximum file size of 500 Kb was reached. For more information about these fields, see the related Management Console Guide topic listed below.

Related Information

[Adding an RFC server interface](#)

3.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP13 includes connectivity enhancements.

Bulk loading for PostgreSQL

In this release, we support bulkloading for PostgreSQL. PostgreSQL bulk loading is applicable for DSN-less connections only. You must also use the PSQL tool from the official PostgreSQL website. Choose either to append or truncate the target when you configure PostgreSQL bulkloading.

For more information about bulkloading for PostgreSQL, see the *Data Services Supplement for Big Data*.

Decrease local disk usage for upload to Google Cloud Storage

This release has a new option for Google Cloud Storage named [Batch size \(rows\)](#). Use the option to decrease the amount of local disk space that Data Services uses when it uploads data to your Google Cloud Storage (GCS) account. Find the option in the target editor for flat files, XML files, and JSON files.

For complete information about the [Batch size \(rows\)](#) option, see the Google cloud storage section in the *Data Services Supplement for Big Data*.

Google BigQuery ODBC datastore

This release of Data Services introduces the Google BigQuery ODBC datastore for Windows and Linux platforms. The new datastore adds pushdown and bulk loading capabilities when you work with Google BigQuery data in Data Services.

The Google BigQuery ODBC datastore connects to your Google BigQuery account with the Magnitude Simba ODBC driver for BigQuery. The driver contains mechanisms that enable you to use the OAuth 2.0 authentication with either a Google user account or a Google service account. Configure the driver to connect using a data source name (DSN) connection or a server name (DSN-less) connection.

i Note

Beginning with Data Services 4.2.13, we recommend that you create a Google BigQuery ODBC datastore instead of the Google BigQuery datastore. The Google BigQuery ODBC datastore uses the Magnitude Simba ODBC driver for BigQuery, which supports standard SQL and more data types than the Google BigQuery datastore.

For complete information about configuring and using the Google BigQuery ODBC datastore, see the *Supplement for Google BigQuery*.

Azure Data Lake Gen2 support

This release supports Azure Data Lake Gen2 storage to enhance the upload and download of files. Both Service Principal and Shared Key authentications are supported. For more information, see the *Data Services Supplement for Big Data*.

You can also bulk load files using Gen2 of Azure Data Lake with shared key authentication. For more information, see the *Reference Guide*.

SAP HANA proxy connection

Enter proxy connection information to connect to SAP HANA data using an SAP HANA datastore. Also include your SAP Cloud connector account name to connect using the SAP Cloud connector instead of your virtual private network (VPN).

Use the proxy connection options for DSN-less connections on both Windows and Unix.

For proxy option descriptions, see the SAP HANA section in the *SAP Data Services Big Data Guide*.

3.3 Data Quality transforms

Transforms work together to improve the quality of your data. 4.2.13 contains transform enhancements.

3.3.1 Global Address Cleanse transform

In SAP Data Services 4.2 SP13, the Global Address Cleanse transform contains enhancements.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2020 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2020 certification.

Locality2 support for Estonia

Locality2 and the locality identifiers for Locality1-2 are now output for Estonia. The file size of the Estonia directory has increased from 19 MB to 607 MB.

Locality2_code support for Czechia

There is a new output field to include in your output mapping for Czechia addresses: [LOCALITY2_CODE](#). The new output field contains the LOCALITY2 code when an address has LOCALITY1 and LOCALITY2 information.

3.4 Functions and transforms

SAP Data Services 4.2.13 includes enhancements to functions and transforms.

sysutcdatetime

This release of Data Services supports the sysutcdatetime function. For more information, see “sysutcdatetime” in the *Reference Guide*.

New function: to_varchar

Use the new `to_varchar` function to convert a given date or numeric expression to varchar based on a specified format.

The return is a varchar string that describes the specified date expression or numeric expression based on the given format.

For complete information, see the topic “Descriptions of Built-in functions” in the *Reference Guide*.

4 SAP Data Services 4.2 SP 12 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 12.

The following links take you to descriptions of the features by category for version 4.2 SP 12.



[Upcoming changes \[page 27\]:](#)

- Python 3.x support



[Installation, administration, and monitoring \[page 28\]:](#)

- Certifications
- Tips for troubleshooting secure network communications (SNC) issues
- Debug package with runtime DSConfig.txt parameter information



[Connectivity: Big Data and CDC \[page 29\]:](#)

- Kerberos and SSL support for MapR 6.1
- Google Cloud Dataproc for Hadoop
- Support for PostgreSQL database tables and schemas
- Support for Snowflake database
- Access data from Kerberos-secured Hive cluster in Big Data Services
- Updating Hive target tables
- MySQL: SSL/TLS for MySQL DSN-less connections
- Hive: SSL/TLS for Hive remote server DSN-less connections
- DB2: SSL/TLS for DB2 DSN-less connections
- SAP HANA: SSL/TLS for SAP HANA DSN-less database datastores
- Oracle: SSL/TLS for Oracle TNS-less database datastores
- Microsoft SQL Server: Update to the SSL/TLS configuration



Data Quality transforms [page 33]:

- Global Address Cleanse transform
 - Japan data vendor change
 - Updated Russia output fields for Global Address Cleanse
 - Certified Australia options
 - Certified Canada options
 - Certified New Zealand options
 - New Bulgaria address data
 - Enhanced country identification features
 - Enhanced building address assignment
 - Medium change significance
- USA Regulatory Address Cleanse
 - CASS O features



SAP applications [page 38]:

- SAP BW/4HANA Advanced DataStore Object targets

4.1 Upcoming changes

The following changes are coming in a future release of Data Services. We are giving you advanced notice because action must be taken on your side.

Python 3.x support

Python 3.x support is planned for Data Services 4.2.13. The new version of Python is not backwards compatible. Because of this, you will have to update any 2.x Python code being used in User-Defined or Match Best Record transforms before you can run data flows in Data Services 4.2.13 or higher.

i Note

Release dates are subject to change without notice, and SAP does not guarantee release dates. For the availability of future versions, see the Maintenance release planning schedules at <https://support.sap.com/en/release-upgrade-maintenance/maintenance-information/schedules-for-maintenance-deliveries.html>.

To update Python 2.x code to 3.x code, we recommend using the 2to3 conversion utility that ships with Data Services. The utility is located in %LINK_DIR%\Data Services\DataQuality\python\Tools\scripts. For detailed information about this utility, see the Python documentation.

High-level instructions for using this utility:

1. Identify User-Defined or Match Best Record transforms that use Python 2.x code. In this example, let's assume you need to update Python code in a User-Defined transform.
2. In the Data Services Designer, open the data flow that contains the transform.
3. Right-click on the transform and select *User-Defined Editor*.
4. Click *Launch Python Editor* to access the *Python Expression Editor*.
5. Copy and paste the Python 2.x code into a .py file. For example, <example>.py.
6. From the %LINK_DIR%\Data Services\DataQuality\python directory, run the 2to3 conversion utility. To run the utility, type the following:

```
python Tools\scripts\2to3 <example>.py -w
```

The -w in the command means that the tool will write the new code into <example>.py and save the original code to <example>.py.bak.
7. Copy the updated Python code from the <example>.py file and paste it back into the *Python Expression Editor*.
8. Run the data flow and confirm the new 3.x Python code works as needed.

4.2 Installation, administration, and monitoring

SAP Data Services 4.2.12 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Certifications

Data Services 4.2.12 adds support for the following:

- MAPR 6.x
- Data Services certification with HANA HaaS Application Connector
- MongoDB 3.6

Data Services 4.2.12 Patch 3 adds support for the following:

- Red Hat Enterprise Linux 7.5 and 7.6
- HANA 2.0 SPS04
- BI 4.2 SP7

For more detailed information, see the Product Availability Matrix document located at: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Tips for Troubleshooting Secure Network Communications (SNC) Issues

Refer to SAP Note [2819167](#) for information about troubleshooting Secure Network Communications when integrated with Data Services.

Debug package with runtime DSConfig.txt parameter information

During a debug job execution, Data Services 4.2.12 can find DSConfig flags and their corresponding values that are being used in the job. This feature allows you, and SAP Support if necessary, to troubleshoot an environment when moving from a working to non-working environment.

You can enable this feature by setting the *Create Debug Package* trace option to *Yes* in the *Execution Properties* window of the Designer or by enabling *Debug Package* in the *Execution Batch Job* page of the Management Console Administrator.

When you run the job, a `dsconfig_flaglog.txt` file, which contains DSConfig flags and their corresponding values, is created and saved to the Job Server directory.

The debug package now contains the following files: `dsconfig_flaglog.txt` (new in SP12), `DSConfig.txt`, `tracelog.txt`, `errorlog.txt`, `job.atl`, and `sysinfo_tracelog.txt`.

For more information, see “Creating a debug package” in the *Designer Guide* and “Executing batch jobs” and “Downloading a debug package” in the *Management Console Guide*.

i Note

Jobs may take longer to run because Data Services is printing all traces.

Related Information

[Additional documentation on the SAP Help Portal](#)

4.3 Connectivity: Big Data and CDC

SAP Data Services 4.2 SP12 includes the following connectivity enhancements:

Kerberos and SSL support for MapR 6.1

Starting in version 4.2.12 Patch 3 (14.02.12.03), Kerberos and SSL are supported for MapR 6.1 on a Hadoop system.

For more information, see “Creating a DSN connection with SSL protocol in Windows”, “Configuring ODBC driver with SSL protocol for Linux”, and “Configuring Kerberos authentication for Hive connection” in the *Supplement for Hadoop*.

Google Cloud Dataproc for Hadoop

In version 4.2.12 Patch 1 (14.02.12.01), SAP Data Services provides tools for you to connect to an Apache Hadoop web interface running on Google Cloud Dataproc clusters.

Use a Hive database datastore to browse and view metadata from Hadoop and to import data for use in data flows. To upload processed data to Hadoop in a Google Cloud Dataproc cluster, use a WebHDFS file location and a Hive template table as a target in a data flow.

For complete information about working with Google Cloud Dataproc for Hadoop, see the topics in “Google Cloud Dataproc” in the *Supplement for Hadoop*.

Support for PostgreSQL database tables and schemas

Data Services 4.2.12 supports PostgreSQL version 10.X for Windows. Data Services 4.2.12 Patch 1 (14.02.12.01) supports PostgreSQL version 10.X for Linux.


Create a database datastore to import PostgreSQL tables and schemas. PostgreSQL datastores support the following features:

- DSN connection
- DSN-less (server-based) connection
- Data preview
- Push-down functions
- Template tables
- Use of one table with different schemas

Download and install an ODBC driver from the PostgreSQL official site. For more information about the PostgreSQL datastore, see the *Data Services Supplement for Big Data*.

Support for Snowflake database

Snowflake provides a data warehouse that is built for the cloud. After connecting to Snowflake, you can do the following:

- Import tables
- Read or load Snowflake tables in a data flow
- Create and load data into template tables
- Browse and import the tables located under different schemas (for example, Netezza)
- Preview data
- Push down base SQL functions and Snowflake-specific SQL functions. See SAP Note [2212730](#)  for more information.
- Bulk load data, which is possible through AWS S3 File Location or Azure Cloud Storage File Location

Download and install the Snowflake ODBC driver from the Snowflake website. Data Services supports Snowflake ODBC Driver version 2.16.0 and higher.

For more information about the Snowflake datastore, see the *Data Services Supplement for Big Data*.

Access data from Kerberos-secured Hive cluster in Big Data Services

This feature applies to users who have their Hive clusters in the SAP Cloud Platform Big Data Services, formerly known as Altiscale.

In previous releases, you could upload generated data from Data Services to your Hadoop account in Big Data Services. However, you could not import Hive tables and use them as sources and targets in SAP Data Services.

With this release, you can import tables from your Kerberos-secured Hive cluster, then use the tables as sources and targets in data flows.

! Restriction

Currently, you must use the Altiscale HIVE ODBC driver version 2.6.1 when you configure SAP Data Services for this feature.

For details about using this feature, see the section named “SAP Cloud Platform Big Data Services” in the *Supplement for Hadoop*.

Updating Hive target tables

Data Services 14.2.12 allows you to update Hive target tables using Update or Merge SQL. This enhancement is supported for Apache Hive Optimized Row Columnar (ORC) tables when using bulk loading and Hive version 1.2 and higher.

Data Services first loads all of the changes to a staging table, using the existing Hive bulk load options and HDFS file location, and then it applies the changes from the staging table to the target table by executing Update or Merge SQL.

To trigger bulk modification, the *Bulk Load* option must be enabled. In addition, one of the following must be true:

- Your data flow contains a Map_CDC_Operation transform.
- Your data flow contains a Map_Operation transform that outputs Delete or Update rows.
- The *Auto correct load* option is enabled.

If none of these requirements are met, Data Services assumes that the input data contains only Insert rows and it performs only bulk insert. Data Services does not use a staging table or execution of additional SQL.

MySQL: SSL/TLS for MySQL DSN-less connections

Now include SSL/TLS encryption for MySQL database datastores for server-named (DSN-less) connections. Previously you could use SSL/TLS encryption with DSN connections only.

i Note

Data Services supports SSL/TLS protocol for Windows, Linux, and Solaris. Data Services does not currently support SSL/TLS for AIX.

Data Services supports SSL/TLS for MySQL server 5.x, MySQL client 5.3, and MySQL client ODBC driver 5.3.x. See the Product Availability Matrix (PAM) for details about supported versions and platform.

For complete information about creating a MySQL database datastore, see the database datastore section of the *Designer Guide*.

Hive: SSL/TLS for Hive remote server DSN-less connections

You can now include SSL/TLS encryption for Hive database datastores for server-named (DSN-less) connections. For complete information, see the *Supplement for Hadoop* under “Connect to Hive”.

DB2: SSL/TLS for DB2 DSN-less connections

You can now include SSL/TLS encryption for IBM DB2 database datastores for server-named (DSN-less) connections. For complete information, see the *Designer Guide* under “IBM DB2 datastores.”

SAP HANA: SSL/TLS for SAP HANA DSN-less database datastores

You can now include SSL/TLS encryption for SAP HANA database datastores for server-named (DSN-less) connections. For complete information, see the *Data Services Supplement for Big Data*.

Oracle: SSL/TLS for Oracle TNS-less database datastores

You can now include SSL/TLS encryption for Oracle database datastores for server-named (TNS-less) connections. For details, see the topic “Oracle datastores” in the *Designer Guide*.

Microsoft SQL Server: Update to the SSL/TLS configuration

For Microsoft SQL Server, you can now configure SSL/TLS encryption parameters in the Microsoft SQL Server database datastore editor in Designer.

For details, see the topic “Microsoft SQL Server” datastores in the *Designer Guide*.

Related Information

[Additional documentation on the SAP Help Portal](#)

4.4 Data Quality transforms

Transforms work together to improve the quality of your data.

[Global Address Cleanse transform \[page 33\]](#)

In SAP Data Services 4.2 SP12, the Global Address Cleanse transform is enhanced with the following features:

[USA Regulatory Address Cleanse \[page 37\]](#)

In SAP 4.2 SP12 and subsequent patches, the USA Regulatory Address Cleanse transform has been enhanced with the following features.

4.4.1 Global Address Cleanse transform

In SAP Data Services 4.2 SP12, the Global Address Cleanse transform is enhanced with the following features:

Japan data vendor change

Starting with release 4.2 SP12 Patch 3, the Japan directory will be supplied by a new data provider. This new directory named `ga_jp_ipc.dir` contains updated and comprehensive address line data.

SAP customers must purchase this new directory in order to maintain functionality in SP12 Patch 3 and beyond. The new directory may be approximately 33 GB in size.

Note that the previous JPOST Japan directory `ga_jp_paf.dir` will remain available through the February of 2020 and will continue to work in support packages prior to SP12 SP3.

Updated Russia output fields for Global Address Cleanse

The Global Address Cleanse transform output fields named `Additional_Info1` through `Additional_Info8` have changed for Russia addresses.

! Restriction

The `Additional_Info3` and `Additional_Info4` output fields for the updated Russian FIAS data can contain information that is 74 to 256 characters long. In SP12 Patch 3, we increased the default component length for these fields to 256 characters to accommodate the increased field length.

To ensure the output for these fields is not truncated in jobs created in SP12 Patch 2 and prior, you will need to update existing queries and file formats that use the old maximum component length of 64 characters. To change the output component length, go to the [Schema Out](#) window in the Global Address Cleanse transform and double-click on any Additional_Info3 and Additional_Info4 output field to enter an output component length between 74 to 256 characters.

The following table contains the new Russia field descriptions:

Output field	Description
Additional_Info1	Russia A 4-digit tax code (IFNSUL) and a 4-digit territory code (TERRIFNSUL) for a legal entity. The two codes are separated with a pipe symbol () as in IFNSUL TERRIFNSUL.
Additional_Info2	Russia An 11-digit administrative territorial division code (OKATO) and an 8- or 11-digit municipality code (OKTMO). The two codes are separated with a pipe symbol () as in OKATO OKTMO.
Additional_Info3	Russia A 36-character identifier (AOID) and a 36-character global unique identifier (AOGUID) for the address object. The two codes are separated with a pipe symbol () as in AOID AOGUID.
Additional_Info4	Russia A 36-character identifier (HOUSEID) and a 36-character global unique identifier (HOUSEGUID) for the house number. The two codes are separated with a pipe symbol () as in HOUSEID HOUSEGUID.
Additional_Info5	Russia A 36-character global unique identifier (ROOMGUID) for the room number.
Additional_Info6	Reserved for future use.
Additional_Info7	Reserved for future use.
Additional_Info8	Reserved for future use.

Descriptions have also been updated in the “Global Address Cleanse output fields” topic in the *Reference Guide*.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2019 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2019 certification.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2019 certification.

New Bulgaria address data

New address data from HERE was added to the Bulgaria directory file, which improves address line (street/house number) assignment capability.

For more information, see the latest International Directory Update on the Help Portal.

Enhanced country identification features

This release contains enhanced country identification features in the form of new options and improvements to the way the engines search for and assign country information.

Identify country through locality search

A new option in the [Country ID](#) group enables you to select the locality resource the transform uses to determine the country for addresses with no country information. The [Identify Localities](#) option has the following values:

- [MAJOR](#): Uses the list of major localities in addition to the country dictionary, region dictionary, and parsing dictionary. [MAJOR](#) is the default setting.
- [EXTENDED](#): Uses the extended list of localities in addition to the country dictionary, region dictionary, and parsing dictionary.
- [ALL](#): Uses the Universal Postal Union (UPU) all-world directory in addition to the country dictionary, region dictionary, and parsing dictionary.

i Note

The major and extended localities lists are included with the installation at no extra cost. The UPU all-world directory is not included with the installation and is available for purchase.

If you select [ALL](#), but you do not own the UPU all-world directory, or the directory is not located in the directory path you specified in the [Reference files](#) group, the transform uses the [EXTENDED](#) option and issues a warning.

For more information about the [Identify Localities](#) option, see the *Reference Guide*

Enable suggestion lists for countries

A new option named [Enable Suggestion Lists](#) in the [Country ID](#) group enables suggestion lists for country names. To enable the suggestion list feature for countries, select [YES](#).

For more information about the [Enable Suggestion Lists](#) option, see the *Reference Guide*.

New suggestion list components

There are two new suggestion list components to include in suggestion lists:

- [ISO Country Code 2-Char](#): Returns the two-character ISO country code.
- [Sugg Confidence Score](#): Returns the confidence score.

Determine how the transform assigns country results.

A new option named [Assign Country ID Results](#) in the [Engines](#) group enables you to select how the transform determines the country to assign when there are multiple suggested candidate countries.

- [SINGLE](#): Assigns the address for the country with the highest confidence score from the list of candidate countries. When multiple countries have the same high confidence score, the transform takes one of the following actions:
 - Generates a country suggestion list when you have suggestion lists enabled.
 - Outputs the record as unassigned and generates an info code that indicates the country is unknown.
- [MULTIPLE](#): Assigns an address for each candidate country using the applicable directory for each country. If the transform finds the address assigned for only one country with a high confidence score, then it assigns the address using that country. If the transform finds the address assigned with more than one country with equal high confidence scores, the transform takes the following actions:
 - Generates a country suggestion list when you have suggestion lists enabled.
 - Outputs the record as unassigned and generates an info code that indicates the country is unknown.

For information about the new [Assign Country ID Results](#) option, see the *Reference Guide*.

Enhanced building address assignment

In this release, we included many enhancements to the way Data Services assigns addresses with building information. Specifically, we added the following new output fields:

- [BUILDING1](#): Contains the building name.
- [BUILDING2](#): Contains the building name for a second building when the address contains two buildings.
- [BUILDING_DESCRIPTION1](#): Contains the building descriptor.
- [BUILDING_DESCRIPTION2](#): Contains the building descriptor for a second building when the address contains two buildings.

Additionally, the output fields [BUILDING_NAME1](#) and [BUILDING_NAME2](#) contain the building descriptor and building name, when applicable.

For complete information about how Data Services assigns addresses when they include building information, see “Address assignment for buildings” in the *Designer Guide*.

Medium change significance

This release now identifies medium significant changes between original input addresses and output addresses.

The CLEANSE_CHANGE_INFO_ table currently identifies high and now identifies medium significance changes. An example of a medium significance is when a locality_addition is changed.

For more information about significance change, see “CLEANSE_CHANGE_INFO_” in the *Reference Guide*.

Parent topic: [Data Quality transforms \[page 33\]](#)

Related Information

[USA Regulatory Address Cleanse \[page 37\]](#)

[Additional documentation on the SAP Help Portal](#)

4.4.2 USA Regulatory Address Cleanse

In SAP 4.2 SP12 and subsequent patches, the USA Regulatory Address Cleanse transform has been enhanced with the following features.

CASS O features

Note

When upgrading to 4.2 SP11 and higher, you'll see new functionality that falls under CASS O requirements. We continue to work on USPS CASS O requirements and will have the necessary regulatory changes implemented by the CASS O cycle target date.

The USPS is allowing us to roll out these requirements as they're ready without needing to recertify. At this time, SAP software and CASS reports continue to be CASS N certified.

For more information about CASS O, see <https://postalpro.usps.com/certifications/cass> .

DPV Indicators

With the release of 4.2 SP12 Patch 3, we include new DPV indicators. The following table contains the DPV indicator, the name of the directory, and the related output field.

DPV indicators

Name	Directory	Output field
DPV Door Not Accessible	dpv_door_na.dir	DPV_Door_Not_Accessible
DPV No Secure Location	dpv_no_secure_loc.dir	DPV_No_Secure_Location
DPV PO Box Throwback	dpv_po_box_throwback.dir	DPV_Throwback_Indicator

The DPV indicator directories have been added to the list of DPV directories that you receive when you purchase the USPS CASS Certified address directories option. These files are installed alongside the other DPV

directory files. Data Services accesses the files when you enable DPV in your data flows. For complete information about the DPV indicators, see the section “Beyond the Basic Address Cleansing” in the *Designer Guide*.

For information about the USPS CASS Certified address directories, see the [latest directory letter](#).

Parent topic: [Data Quality transforms \[page 33\]](#)

Related Information

[Global Address Cleanse transform \[page 33\]](#)

4.5 SAP applications

SAP Data Services 4.2 SP12 includes the following enhancements to support SAP applications:

SAP BW/4HANA Advanced DataStore Object targets

Data Services supports loading data to SAP BW/4HANA Advanced DataStore objects (ADSO).

Access BW/4HANA ADSOs by creating a BW Target type datastore and importing ADSOs. Use the ADSOs as targets in your data flow. ADSOs load generated data to HANA.

i Note

Use ADSO target objects only after you have upgraded from SAP NetWeaver Business Warehouse (BW) to BW/4HANA.

For more information about using SAP BW/4HANA ADSOs as targets, see the *Supplement for SAP*.

For information about upgrading your old BW target objects to BW/4HANA, see the *Upgrade Guide*.

i Note

Currently, Data Services 4.2.12 supports BW/4HANA 2.0. However, when future releases occur, version support may change. Keep in mind that the Product Availability Matrix (PAM), located at <https://apps.support.sap.com/sap/support/pam>, maintains an updated list of supported versions.

Related Information

[Additional documentation on the SAP Help Portal](#)

5 SAP Data Services 4.2 SP 11 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 11.

The following links take you to descriptions of the features by category for version 4.2 SP 11.



[Installation, administration, and monitoring \[page 41\]:](#)

- Certifications



[Usability \[page 42\]:](#)

- Transaction control in continuous work flow
- Cast() function datetime data types



[Connectivity: Big data and CDC \[page 42\]:](#)

- Secure Sockets Layer (SSL) connectivity for SAP HANA databases
- Specify subfolders for cloud storage file locations
- New field for HTTP Adapter operation instance
- Support for Apache Knox gateway
- Azure blob service-level shared access
- Support for JSON key file for Google
- Support for SAP Vora
- S3 cloud storage classes



[Data Quality transforms \[page 46\]:](#)

- USA Regulatory Address Cleanse
 - 5-digit validation
 - Nondelivery days (NDD) for DPV and DSF2
 - PO Box only zone
 - DPV validation and single trailing alpha
 - Revisions to PS Form 3553, CASS Summary Report
- Global Address Cleanse transform
 - Suggestion list Sugg_Full_Addressline displays alias
 - China Post data vendor change
 - New India GEO directory

5.1 Installation, administration, and monitoring

SAP Data Services 4.2.11 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Certifications

Data Services 4.2.11 adds support for the following:

- BI 4.2 SP6
- DB2 UDB 11.1
- SAP HANA 2.0 SPS03 - Works with the Data Services 4.2.11 repository and can be used as a source or target in a dataflow.
- MSSQL Server 2017

! Restriction

The replication server change data capture (CDC) option of a datastore for SQL Server version 2016 and higher is disabled. Also, SQL Server 2017 does not support Azure DW and APS.

- MySQL 5.7
- Sybase IQ 16.1 SP2 and SP3 - Works with source and target tables.
- Chrome 67.0.3396.99
- Cloudera 5.13.2 - Works with HDFS format and location, Hive datastore through ODBC driver, Hive Server2 versions 1.1 and 1.2, Hive adapter, and the Impala datastore.
- MapR 6.0 - Works with Hive native ODBC as a source and target, Hive adapter, and HDFS format and location.
- Hortonworks HDP 2.6 - Works with Hive native ODBC as a source or target, Hive adapter, and HDFS format and location.

For more detailed information, see the Product Availability Matrix document located at: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

5.2 Usability

SAP Data Services 4.2 SP11 includes the following enhancements that improve the user experience.

Transaction control in continuous work flow

Data Services now supports transaction control in a continuous work flow for Oracle database tables only. Previously, you could not use transaction control in a continuous work flow.

For more information about transaction control, see “Transactional loading of tables” in the *Designer Guide*. For information about continuous work flows, see “What is a continuous work flow” in the *Designer Guide*.

Cast() function datetime data types

The format of datetime data types for the cast() function has been documented. For more information, see the “cast” topic in the *Reference Guide*.

5.3 Connectivity: Big data and CDC

SAP Data Services 4.2 SP11 includes the following connectivity enhancements.

Secure Sockets Layer (SSL) connectivity for SAP HANA databases

Starting in the Data Services SP11 Patch 2 release, you can use Secure Sockets Layer (SSL) connectivity for SAP HANA databases. This feature enables you to use the SSL (as well the newer Transport Layer Security, or TLS) standard security protocol.

To take advantage of this new feature, you must install Data Services SP11 Patch 2, and you must also update your Java security file, as explained in the “Edit java.security file” topic in the *Data Services Administrator Guide*. In addition, the KeyStore and TrustStore certificates must be available to the CMS (in Windows, the location usually means a local drive, not network drive; and in UNIX, the location doesn’t matter).

Finally, follow these steps to ensure that your SSL connection is set up properly.

In the *Add Data Services Repository* window, you must select the *Enable SSL* checkbox. The *Verify Server Certificate* option appears. Choose *Yes*. The following fields will appear.

1. In the *SSL KeyStore* field, enter the location of the Java keystore.
2. Select the *SSL KeyStore Type*, either *JKS* or *PKCS12*.

3. Enter the [SSL KeyStore Password](#) to access the private key from the Java keystore file.
4. In the [SSL TrustStore](#) field, enter the location of the Java truststore.
5. In [SSL TrustStoreType](#), select either [JKS](#) or [PKCS12](#).
6. Enter the [SSL TrustStore Password](#) to access the private key from the Java keystore file.

For more information, see “Registering a repository in the CMC” in the *Administrator Guide*.

Specify subfolders for cloud storage file locations

With the 4.2 SP11 Patch 1 release, you can specify more than one subfolder location for files in your cloud storage account. Use this feature based on the file storage type or the database type.

Cloud storage type

SAP has added a third parameter to the “copy” functions to specify the remote file directory for the following cloud storage types:

- Amazon S3
- Azure blob storage
- Azure Data Lake Store
- Hadoop file system
- Google Cloud storage

For specific cloud storage types, enter the remote path, including subfolders, for the source or target file location in your cloud storage account in the `copy_from_remote_system` and `copy_to_remote_system` functions.

Code Syntax

```
copy_from_remote_system("<file_location_object_name>", "<file_name>",
"<remote_sub_directory>")
```

Code Syntax

```
copy_to_remote_system("<file_location_object_name>", "<file_name>",
"<remote_sub_directory>")
```

For complete information, read the descriptions for the built-in functions in the *Reference Guide*.

Source or target editor definitions

Use the [File Name](#) or [File](#) options in the source or target editors to specify the full remote path, including subdirectories, for the source or target file location in your cloud storage account. Applicable file types include the following:

Data file type	Editor	Tab or Group	Option
Cobol copybooks	Source	Data File tab	File name
Excel workbooks	Source	Format tab	File name

Data file type	Editor	Tab or Group	Option
Flat files	Source	Data File(s) group	File Name(s)
	Target	Data File(s) group	File Name(s)
XML	Source	Source tab	File
	Target	Target tab	File
JSON	Source	Source tab	File
	Target	Target tab	File

For more information about including the subfolders in the source or target editors, see each data type, or read about source and target objects in the *Reference Guide*.

New field for HTTP Adapter operation instance

We have added a field for the HTTP Adapter in SAP Data Services 4.2 SP11 Patch 1. The new field is named [DOCTYPE header](#). Find the new field in the [Operation Details](#) tab of the Management Console.

Complete the new field when applicable for the Request/Reply operation and the Request/Acknowledge operation. Use the field to specify the DOCTYPE of the request header. Enter the value using the full string of `<!DOCTYPE...>`. Leave the field blank when the DOCTYPE of the request header is not required.

For complete information about using the options in the [Operation Details](#) tab, see the HTTP Adapter section of the *Supplement for Adapters*.

Support for Apache Knox gateway

Access the WebHDFS service in your Hadoop cluster using Apache Knox. Knox provides centralized authentication and access to the WebHDFS service. Find options to set up the Knox gateway in the HDFS file location object editor. For more information, see the *Reference Guide*, or the *Supplement for Hadoop*. For information about configuring Hadoop with Knox, consult your Apache Knox documentation.

Azure blob service-level shared access

Data Services now supports accessing a specific file (blob) or blobs in a container in an Azure Cloud Storage using a service-level storage access signature (SAS).

Set up the service-level SAS when you create a file location object for Azure Cloud Storage in Data Services. Complete the two options that support the service-level type of access: [Authorization type](#) and [Shared Access Signature URL](#).

[Authorization type](#) is an existing option for which we added two new values. Choose one of the following values for the service-level SAS:

- [File \(Blob\) Shared Access Signature](#)
- [Container Shared Access Signature](#)

The existing option, [Primary Shared Key](#) is available for accessing account-level SAS.

Also complete the new option named [Shared Access Signature URL](#), which contains the elements for accessing a specific file (blob) or blobs in a container. Use the following URL formats.

- To access blobs in a container, use the following format for the service-level SAS URL:
`https://<storage_account_name>/<container_name>/<signature_value>`
- To access a specific file (blob), use the following format for the service-level SAS URL:
`https://<storage_account_name>/<container_name>/<file_name>/<signature_value>`

For information about creating a file location object for Azure Cloud Storages, see the *Data Services Supplement for Big Data*.

Support for JSON key file for Google

With this release, you can use JSON file types for the service account private key file. Previously you could use only P12. You generate the key in your Google BigQuery account and save a local copy. Enter the location of the local copy when you create a Google BigQuery datastore or a Google Cloud Storage file location object.

For information about the service account private key file, see the *Supplement for Google BigQuery*.

Support for SAP Vora

This release includes new SAP Vora support for the following:

- Username and password authentication using an SAP HANA ODBC driver through an SAP HANA wire port.
- Additional target table options, including new bulk loading options, to support WebHDFS and Microsoft Azure Data Lake Storage.
- DSN configuration support for SAP Vora on both Windows and UNIX, including support for the Transport Layer Security (TLS) protocol on UNIX.

See the *Data Services Supplement for Big Data* for more information.

S3 cloud storage classes

When setting Amazon S3 protocol options, you can now choose a S3 cloud storage class to use to restore files. The values for the [Storage Class](#) option include:

- [STANDARD](#): Default storage class.
- [REDUCED_REDUNDANCY](#): For noncritical, reproducible data.
- [STANDARD_IA](#): Stores object data redundantly across multiple geographically separated availability zones.
- [ONEZONE_IA](#): Stores object data in only one availability zone.

i Note

The GLACIER storage class is not supported. Data Services is not able to specify this storage class during object creation.

For more information about file location options that are specific to the Amazon S3 protocol, see “Amazon S3 protocol options” in the *Data Services Supplement for Big Data* guide.

For more information about the storage classes, see the Amazon AWS documentation.

5.4 Data Quality transforms

In SAP Data Services 4.2 SP11, the set of transforms that work together to improve the quality of your data have been enhanced with the following features.

5.4.1 USA Regulatory Address Cleanse

In SAP Data Services 4.2 SP11 and subsequent patches, the USA Regulatory Address Cleanse transform has been enhanced with the following feature.

i Note

When upgrading to 4.2 SP11 and higher, you will see new functionality that falls under CASS O requirements. We continue to work on USPS CASS O requirements and will have the necessary regulatory changes implemented by the CASS O cycle target date.

The USPS is allowing us to roll out these requirements as they are ready without needing to re-certify. At this time, SAP software and CASS reports continue to be CASS N certified.

For more information about CASS O, see <https://postalpro.usps.com/certifications/cass> .

5-digit validation

The software validates the city, state, and ZIP code before it calculates the 5-digit total count that appears in the 5-digit Coded section of the CASS Summary Report (PS Form 3553).

Starting in 4.2 SP11 Patch 1, we will do the following in preparation for CASS Cycle O:

- If an address contains a valid 5-digit ZIP Code that corresponds with the city and state, we will include it in the 5-digit Coded count in the PS Form 3553.
- If an address contains an invalid 5-digit ZIP Code (for example, it doesn't correspond with the city and state), we will not include it in the 5-digit Coded count in the PS Form 3553.
- If an address does not have a 5-digit ZIP Code, we will not include it in the 5-digit Coded count in the PS Form 3553.


To see some examples, see <https://postalpro.usps.com/certifications/cass> .

Non-delivery days (NDD) for DPV and DSF2

In preparation for CASS Cycle O, the 4.2.11 Patch 1 release of the software allows businesses to request days on which they do not want mail delivered. For example, a business may be closed on Saturday so they don't want mail delivered on that day.

The software will use the following optional files, which are included in the SAP DELIVERY POINT VALIDATION (DPV) DATA directory, to determine non-delivery dates:

- `dpv_non_delivery_days_flag.dir`
- `dpv_non_delivery_days.dir`

You can download the DPV directory from the SAP ONE Support Launchpad on the SAP Support Portal at <https://launchpad.support.sap.com/#/softwarecenter> .

The following output fields have been added to the USA Regulatory Address Cleanse transform:

Output field	Description
DPV_NON_DELIVERY_DAYS	Indicates whether or not there are days when there is no mail delivery. This field can be set to <i>Y</i> , <i>N</i> , or blank.
DPV_NON_DELIVERY_DAY_SUN	The following are valid values: <ul style="list-style-type: none">• <i>Y</i>: Indicates a non-delivery day.• <i>N</i>: Indicates that mail is delivered on that particular day and not on all days.• <i>Blank</i>: Indicates that the DPV option is disabled or both of the new directory files are missing.
DPV_NON_DELIVERY_DAY_MON	
DPV_NON_DELIVERY_DAY_TUE	
DPV_NON_DELIVERY_DAY_WED	
DPV_NON_DELIVERY_DAY_THU	
DPV_NON_DELIVERY_DAY_FRI	
DPV_NON_DELIVERY_DAY_SAT	

PO Box only zone

In preparation for CASS Cycle O, the 4.2.11 Patch 1 release of the software includes an output field named `PO_Box_Only_Postcode`. The output field indicates whether the input ZIP Code is listed in the postal directory as a ZIP Code for post office box delivery only. Values include the following:

- *Y*: ZIP Code is for a PO Box only zone.
- *N*: ZIP Code is not for a PO Box only zone.
- *Blank*: No ZIP Code assigned.

Note

Data Services does not populate this output field if you are not using a postal directory that includes the new PO Box only information. See the latest directory letter for details.

DPV validation and single trailing alpha

Data Services processes primary address ranges that contain a single trailing alpha on input for DPV validation in accordance with CASS O requirements.

The USA Regulatory Address Cleanse transform drops the single trailing alpha and retries to DPV confirm the primary range without the single trailing alpha under the following circumstances:

- There is no secondary address on input.
- Address is flagged as a street type record in the US directory data.
- Address look-up with primary range and single trailing alpha on input fails DPV confirmation.

If the address DPV confirms without the single trailing alpha, the transform outputs the primary address with the single trailing alpha and assigns a ZIP+4 to the ZIP Code. The DPV_Status output field contains the code "S". The code "S" means that the software dropped the secondary address information or dropped the trailing alpha from the primary address to match the primary range to a confirmed delivery point.

Revisions to PS Form 3553, CASS Summary Report

We removed Direct Delivery Point Validation (DirectDPV) product information from the PS Form 3553, CASS Summary Report to satisfy USPS CASS O requirements. The Postal Service retired the DirectDPV product in November 2017.

Caution

Removing DirectDPV has changed the order of information on the report.

You can obtain a copy of the revised PS Form 3553, CASS Summary Report by going to <https://about.usps.com/forms/all-forms.htm> .

SAP Crystal Reports 2016 Support Pack 4 Version 14.2.4.2410 was used to make changes to PS Form 3553.

5.4.2 Global Address Cleanse transform

In SAP Data Services 4.2.11, the Global Address Cleanse transform has been enhanced with the following features.

Suggestion list Sugg_Full_Addressline displays alias

We changed the way the software presents suggestions for records that have an input street alias.

Sometimes when input data contains records that have addresses on the same street, suggestions appear to contain duplicates. The suggestions, however, are not duplicates, but suggestions for separate records that have the same street name.

The suggestion content for the suggestion list component [Sugg_Full_Addressline](#) now contains the input street alias information so the suggestions do not appear to be duplicates. The software places the alias street name in parentheses next to the suggested official street address.

❖ Example

Bliss Road is an alias street name for Main Street. Previously, when there were two input addresses, one with the official street name Main St, and the other with the alias Bliss Rd, the suggestions looked like they were for duplicate addresses. However, with this release, the suggestions are differentiated by the alias street name in parentheses. The following table shows how the suggestions looked in the previous version of Data Services, and how they look in Data Services 4.2.11.

Comparison of returned suggestions

Input	Primary_Name1	Primary_type1	Previous Data Services version: Sugg_Full_Addressline	Data Services version 4.2.11 and later: Sugg_Full_Addressline
100 Main Street	MAIN	ST	MAIN ST	MAIN ST
250 Bliss Rd	MAIN	ST	MAIN ST	MAIN ST (BLISS RD)

China Post data vendor change

Effective July 2018, SAP started using a new provider with more accurate and current China address directory data. Data Services 4.2.11 works only with the new data, so you must download the new directory named `ga_cn_nav.dir`. Data Services versions prior to 4.2.11 will work only with the old data.

⚠ Caution

SAP is discontinuing using the old vendor and has terminated our contract with that vendor. If you wish to continue to cleanse China data, you will need to purchase a new subscription for China Address Directory that will be available with Data Services 4.2.11. Older service pack versions of Data Services will continue to support the old China Post data directory (`ga_cn_paf.dir`). However, we will be updating and posting this directory file for the last time in July 2018. This allows you to process data with the old data directory file through the end of the year.

The new data has limited coverage on postcodes. If a real postcode cannot be obtained, Data Services assigns a generated locality level postcode.

The following status code values were added to indicate when Data Services assigns a default locality level postcode:

Sixth character value	Description
P	A software generated default locality level postcode was assigned. Global Address engine (China).

Sixth character value	Description
Q	An alias and a software generated default locality level postcode was assigned. Global Address engine (China).

For more information, see “Global Address Cleanse status codes” and “Output fields for the Global Address Cleanse transform” in the *Reference Guide*.

New India GEO directory

Effective August 2018, we support Geocoding Parcel for India by MMI data. Download the India GEO directory (`geo_in_mi.dir`) from the SAP ONE Support Launchpad at <https://launchpad.support.sap.com/#/softwarecenter>.

Supported geocoding mode:

- AG: Address geocoding
- RG: Reverse geocoding
- POITS: POI textual search

Supported centroid level:

- L1: Locality1
- L2: Locality2
- L3: Locality3

Note

To get the best results when connecting GAC output to GEO input fields, be sure to map the locality2 and locality3 input.

For information about point of interest (POI) types, see the "Geocoding Parcel for India by MMI" directory letter on the help portal (https://help.sap.com/viewer/p/ADDRESSING_DIRECTORIES).

6 SAP Data Services 4.2 SP 10 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 10.

The following links take you to descriptions of the features by category for version 4.2 SP 10.



[Installation, administration, and monitoring \[page 52\]:](#)

- Certifications
- Central Object Library adds Latest Version column
- Support TCPS protocol for Oracle repository



[Connectivity: Big data and CDC \[page 53\]:](#)

- Apache Impala support
- Apache Hive Server connectivity using an ODBC driver
- Azure Data Lake Store file location object
- Hadoop Distributed File System (HDFS) file location object
- SAP HANA access moves to multitenancy database containers
- Microsoft Graph API as source and target
- Multi-Schema for Netezza
- SAP Data Services supports integration with SAP Vora: Phase 2
- Support more data types for Google BigQuery
- Time delay option for ODP sources
- Amazon S3 data encryption
- OData adapter upsert for SuccessFactors endpoint



[Data Quality transforms \[page 56\]:](#)

- Global Address Cleanse transform
 - Certified Australia options
 - Certified Canada options
 - Certified New Zealand options
 - La Poste ratification
 - New NW_Firm input field for Global Address Cleanse
 - Enhancements to suggestion lists
 - Support for single-line address processing
- USA Regulatory Address Cleanse transform
 - Support for additional military address descriptors



Platform transforms [page 59]: Data Mask transform

- Date and Numeric Generalization added to Data Mask transform
- Date Parsing Options group relocated
- New Default Century Threshold date parsing option



Text Data Processing [page 60]:

- Updates to the *Language Reference Guide*

6.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP10 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Certifications

Data Services 4.2.10 adds support for the following:

- BI 4.2 SP5
- HANA 2.0 SPS 02
- Apache Ranger
- Oracle 12cR2 and above - Data Services supports long table/column names (up to 128 bytes).

For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <http://support.sap.com/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Central Object Library adds Latest Version column

When you view objects in the Central Object Library in Designer, you can quickly assess the version history of an object. A new column named *Latest Version* shows the following information about an object:

- Version number
- Date and time the software saved the version

Read about the Central Object Library in the *Designer Guide*.

Support TCPS protocol for Oracle repository

When you configure your Oracle repository in the Central Management Console (CMC), you can choose to use the TCPS (TCP with SSL) protocol.

Before you configure TCPS for your Oracle repository, install the required Oracle Advanced Security files and comply with Oracle's SSL requirements. The Advanced Security files contain the necessary files for configuring TCPS for an Oracle repository in Data Services. Consult your Oracle documentation for details.

For complete information about configuring the TCPS protocol for an Oracle repository, see the *Administrator Guide*.

6.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP10 includes the following connectivity enhancements.

Apache Impala support

Apache Impala is an open source database for Apache Hadoop. You connect to Impala by using the Cloudera ODBC driver.

After configuring an Impala database datastore, you can import tables and use data from Impala as a source or target in your data flow.

For more information, see "Apache Impala" and "Impala connectivity using the Cloudera ODBC driver" in the *Reference Guide*.

Apache Hive Server connectivity using an ODBC driver

There is a new way to connect to the Hive Server. Users can now use supported ODBC drivers (MapR, Hortonworks, Cloudera) to connect remotely to the Hive Server.

Using this option for connection means that you don't have to install Data Services in the cluster for connection.

For more information, see "Apache Hive" and "Using an ODBC driver to connect to a remote Hive Server" in the *Reference Guide*.

Azure Data Lake Store file location object

Use an Azure Data Lake Store file location object to access data from your Azure Data Lake Store. Use the data as a source or target in Data Services 4.2.10.

For more information, see the *Reference Guide* Go to Objects, Description objects, File Location object.

Hadoop Distributed File System (HDFS) file location object

Use an HDFS file location object to access data from your HDFS. Use the data as a source or target in Data Services 4.2.10. With an HDFS file location object, the Data Services installation can be outside of the Hadoop cluster.

For more information, see the *Reference Guide*. Go to Objects, Description of objects, File Location object.

SAP HANA access moves to multitenancy database containers

This feature applies to SAP Data Services 4.2 support packs 7, 8, 9, or 10. We refer to these versions as 4.2 SPx.

Data Services 4.2 SPx supports SAP HANA 2.0 SPS 01 multitenancy database containers (MDC). Any new database datastores that you create in Data Services 4.2 SPx with SAP HANA 2.0 SPS 01 MDC and later must be for tenant databases.

If you are not familiar with the tenant database concept in SAP HANA 2.0 SPS 01, refer to the SAP HANA Platform *Release Notes* for 2.0 SPS 01.

For upgrade considerations, see the *Upgrade Guide*. Find descriptions for all SAP HANA database datastore options in the *Reference Guide*.

Microsoft Graph API as source and target

You can now connect to Microsoft Graph API using the OData adapter, and choosing the OAuth 2.0 Authentication Type. This will enable you to access other Microsoft services such as Outlook, OneDrive, One Note, Planner, and Office Graph through a single endpoint and with a single access token.

The Microsoft Graph API provide access to data, including users, groups, contracts, devices, files, domains, sites, and drives. It uses Azure Active Directory (Azure AD) to provide secure authentication and authorization to the user's Microsoft Graph API data. Azure AD implements authorization flows per the OAuth 2.0 protocol.

For OAuth 2.0, SAP Data Services 4.2 SP10 can support only the grant types password and client_credentials. Refer to the *Supplement for Adapters Guide*.

Multi-Schema for Netezza

Netezza 7.x now supports the ability for users to define multiple schemas within a database.

You can use privileges to restrict access to schemas and the objects in a schema.

[Schema name](#) was also added to the list of objects for analysis options.

For more information, see “Analysis Options” in the *Management Console Guide*.

SAP Data Services supports integration with SAP Vora: Phase 2

Use SAP Vora tables as sources and targets in Data Services data flows using the new SAP Vora datastore. In Phase 1 of our integration with SAP Vora, we implemented connectivity with SAP Vora through the ODBC datastore. Users with SAP Vora version earlier than version 2 can still use that method. For SAP Vora 2.0 and later versions, users can use the new SAP Vora datastore.

For more information about the SAP Vora datastore, see the Database Datastores section in the *Reference Guide*. Go to Objects, Description of objects, Datastores.

Support more data types for Google BigQuery

Data Services now supports the following data types for Google BigQuery: Date, Time, and Datetime.

Time delay option for ODP sources

This release includes a new option that is displayed any time you create a new ODP source, or edit an existing ODP source: [Delay between calls by continuous workflow \(sec\)](#).

This option specifies the wait period between the reader close and reader open for each continuous workflow iteration. The software uses the delay between these calls to create a new time delta; otherwise the software could extract the same rows repeatedly if the delta is not large enough.

See the “ODP source” topic in the *Data Services Supplement for SAP* for more information.

Amazon S3 data encryption

With this release, Data Services now includes four server-side encryption methods for Amazon S3 data:

- Encryption Algorithm
- AWS KMS Key ID
- AWS KMS Encryption context
- Customer Key

You can choose the encryption method to use for Amazon S3 data any time you create a new Amazon S3 file location object, or edit an existing S3 file location object. The default is no encryption.

See the “Amazon S3 protocol options” topic in the *Data Services Reference Guide* for details on each of these encryption methods.

OData adapter upsert for SuccessFactors endpoint

With this release, use a new method for upsert when you use OData version 2 and SuccessFactors endpoints. Choose the *Upsert function* option for *Loader Action* in the target configuration in your data flow.

You can still use the *Upsert* option for *Loader Action* when you use any version of OData. The *Upsert* option uses a two-request method based on the OData version you use.

For more information, see the *Supplement for Adapters*.

6.3 Data Quality transforms

In SAP Data Services 4.2 SP10, the set of transforms that work together to improve the quality of your data have been enhanced with the following features.

6.3.1 Global Address Cleanse transform

In SAP Data Services 4.2 SP10, the Global Address Cleanse transform has been enhanced with the following features.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2018 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2018 certification.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2018 certification.

La Poste ratification

The France processing options meet all requirements for La Poste 2018 certification.

New NW_Firm input field for Global Address Cleanse

Global Address Cleanse has a new NW input field, NW_Firm, which holds an input firm name. When you include a firm name in some addresses, it helps the software better assign and verify addresses. Also, a firm name is required for some postal authorities for specific discounts. For example, a firm name is required for France PO Box addresses that use the CEDEX postal code numbering system. CEDEX is designed for recipients of large volumes of mail.

For more information about NW input fields, see the Data Quality fields section of the *Reference Guide*.

Enhancements to suggestion lists

The enhanced suggestion capabilities in the Global Address Cleanse transform enables the software to present more relevant suggestions by improving the efficiency in the engines. Additional settings enable users to view all of the suggestions, and includes range type information when the suggestions include ranges.

New items

- To enable user access to all suggestions when the suggestions result in additional pages, include the new output field More_Suggestions_Available.
- To have the software indicate the type of range when the suggestion list includes ranges, select **YES** for the component Sugg_Range_Type.

For details about enabling More_Suggestions_Available and Sugg_Range_Type, see the related topics in the Global Address Cleanse transform section of the *Reference Guide*.

Support for single line address processing

The Global Address engine in the Global Address Cleanse transform now better assigns input addresses that appear on one line.

This enhancement is applicable when the user sets up the input mapping as follows:

MULTILINE1-12
COUNTRY
FIRM (optional)

The software automatically applies the single address line enhancement when an input address is all on Multiline1 and any other multilines are blank.

❖ Example

The following input mapping shows examples of the different ways an address appears all on one line:

Input:

MULTILINE1 = 1811 S Alpine Rd Rockford IL 61108
COUNTRY = US

Input:

```
MULTILINE1 = 1811 S Alpine Rd Rockford IL 61108  
MULTILINE2  
MULTILINE3  
MULTILINE4  
COUNTRY = US
```

Input:

```
FIRM = Allied Advertising  
MULTILINE1 = 1811 S Alpine Rd Rockford IL 61108  
COUNTRY = US
```

For more information about preparing your input data, see the Address Cleanse section of the *Designer Guide*.

6.3.2 USA Regulatory Address Cleanse transform

In SAP Data Services 4.2 SP10, the USA Regulatory Address Cleanse transform has been enhanced with the following feature.

Support for additional military address descriptors

The USA Regulatory Address Cleanse transform now supports the following special street name designators for the military:

- UMR (Unit Mail Room)
- OMC (Official Mail Center)

For more information about USA Regulatory Address Cleanse transform, see the *Reference Guide*.

6.4 Platform transforms

In SAP Data Services 4.2.10, the set of transforms that you use for general data movement operations have been enhanced with the following features.

6.4.1 Data Mask transform

In SAP Data Services 4.2 SP10, the Data Mask transform has been enhanced with the following features.

Date and Numeric Generalization added to Data Mask transform

Increase your date and number masking capabilities by using the new [Date Generalization Group](#) and [Number Generalization Group](#) options in the [Data Mask](#) transform. The following table describes both types of generalization.

New feature	Description
Number generalization	Generalize all number-based input data by masking the output data with a common label or a generalized number range.
Date generalization	Generalize all date-based input data by masking the output data with a common label or a date range. You define the range, or you choose to have the software generate the range.

For more information about the new number and date generalization capabilities, see the Data Mask section in the *Reference Guide*.

Date Parsing Options group relocated

To support both the new [Date Generalization Group](#) and the [Date Variance Group](#), we moved the [Date Parsing Options](#) group out of the [Date Variance Group](#) options and placed them following [Processing Options](#). The [Date Parsing Options](#) settings now apply to both Date Variance and Date Generalization groups.

For information about upgrading to 4.2.10, and the effects on existing [Date Parsing Options](#) groups, see the *Upgrade Guide*.

New Default Century Threshold date parsing option

We have added a new option to the [Date Parsing Options](#) group in the [Data Mask](#) transform. The new option is [Default Century Threshold](#).

The value that you enter for this option indicates whether an input date with a two-digit year is considered part of the 20th century or the 21st century. Enter a value from 0 to 99. The default value is 25.

For more information about the Default Century Threshold option, see the option descriptions for the Date Parsing Options in the [Reference Guide](#).

6.5 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Updates to the Language Reference Guide

The "Voice of the Customer" sentiment analysis elements of Fact Extraction have been added for Japanese. Examples are shown for Sentiment extraction, Emoticon extraction, and Request extraction.

Over 100 minor corrections and additions have been made in the Guide.

7 SAP Data Services 4.2 SP 9 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 9.

The following links take you to descriptions of the features by category for version 4.2 SP 9.



[Installation, administration, and monitoring \[page 61\]:](#)

- Null monthly reports for NCOA Licensee ID
- New Web Services operations



[Connectivity: Big data and CDC \[page 62\]:](#)

- HANA 2.0 support
- S/4 HANA 1610 support
- RedHat 7.3 platform support
- Performance optimization for Google BigQuery large data extraction
- Additional functions for SQL pushdown in Teradata and HP Vertica
- Better performance loading HP Vertica data source
- Support for Microsoft APS
- Support for Microsoft Azure DW (Data Warehouse)
- Support for ORC and AVRO format in Hadoop Hive
- SAP Data Services supports integration with SAP Vora
- Data Services connects to your Hadoop cluster in the cloud



[Data Quality transforms \[page 65\]:](#) Global Address Cleanse transform

- Puerto Rico address assignment enhancement
- New directories available for Global Address Cleanse



[Usability \[page 66\]:](#) Support Arabic right to left in COBOL Copybooks

7.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP9 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Null monthly reports for NCOA Licensee ID

You can now export a null monthly report for a specific NCOA Licensee ID by using the [Export Null monthly reports for NCOA Licensee ID](#) option under ► [Management](#) ► [Certification Logs](#) ► [NCOALink](#) ►.

For more information, see “Exporting NCOALink report data” in the *Management Console Guide*.

New Web Services operations

The following new Web Services are available for use with this release of Data Services.

Web Service	Description
Get_BatchJob_Detail()	Retrieves a list of all system configurations, global variables, and substitution parameters.
Get_BatchJob_Run_ExeDetail	Retrieves a list of all job execution details for a stated runID for the selected repository and job.
Get_BatchJob_Options	Retrieves a list of job options for the named batch job. Results in a list of each option and value from the job Properties information for the named batch job.

The following Web Service is updated with additional functionality.

Web Service	Updated functionality
GetJobServerList()	Updated to return Job Server group names in addition to a list of job servers registered to a repository.

For details about these Web Services, see the *Integrator Guide*.

7.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP9 includes the following connectivity enhancements.

HANA 2.0 support

Data Services is certified with HANA 2.0. For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <https://apps.support.sap.com/sap/support/pam> 🗑️. To go directly to Data Services, enter “Data Services” in the search field.

S/4 HANA 1610 support

Data Services is certified with S/4 HANA 1610. For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

RedHat 7.3 platform support

Data Services supports the RedHat 7.3 platform. For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Performance optimization for Google BigQuery large data extraction

With SAP Data Services 4.2 SP9, we offer a new option in Google BigQuery. The performance for Google BigQuery large data extraction will be enhanced if you specify a GCS file location object in this option.

Before you decide to use the GCS file location object for optimization, consider the following factors:

- Fees: Compare the time saved using optimization against the potential fees from using your GCS account in this manner.
- File size: The optimization may not be beneficial for smaller data files of less than or equal to 10 MB.

For more information about Google BigQuery file optimization for data extraction, see the *Supplement for Google BigQuery*.

Additional functions for SQL pushdown in Teradata and HP Vertica

SQL pushdown transfers part of the transformation logic to the target database in the form of generated SQL statements. SQL pushdown optimizes data transfer by transferring some internal processing to the database to reduce time-consuming data transfers between database server memory and Data Services memory.

Teradata functions

We have added the following functions for Teradata SQL pushdown:

- decode
- ifthenelse
- lpad
- ltrim
- ltrim_blanks
- rpad
- rtrim

- rtrim_blanks
- trunc

For a complete list of Teradata functions for SQL pushdown, see SAP Note [2212730](#).

HP Vertica functions

For a complete list of supported HP Vertica functions, see SAP Note [2212730](#).

Better performance loading HP Vertica data source

Enabling the *Auto Correct Load* option in HP Vertica target table options enhances loading performance. Starting with Data Services 4.2 SP9, when you choose *Auto Correct Load*, the Optimizer automatically implements a MERGE statement to improve the performance of the Auto Correct Load functionality. For more information about Auto Correct Load for HP Vertica, see the *Reference Guide*.

Support for Microsoft APS

Data Services now supports Microsoft Analytics Platform System (APS). For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Support for Microsoft Azure DW (Data Warehouse)

Data Services now supports Microsoft Azure Data Warehouse (DW). For more information, see the Product Availability Matrix document located in the SAP Service Marketplace: <https://apps.support.sap.com/sap/support/pam>. To go directly to Data Services, enter “Data Services” in the search field.

Support for ORC and AVRO format in Hadoop Hive

Data Services now supports the ORC and AVRO format in Hadoop Hive. For more information, see the *Supplement for Adaptors*.

SAP Data Services supports integration with SAP Vora

With the SAP Data Services 4.2 SP 9 release, we've implemented phase 1 support for Data Services integration with SAP Vora. Data Services supports Vora as an ODBC datastore. Vora users can now execute Vora SQL in Data Services using the sql() function, import Vora tables in Data Services designer, and read from Vora tables.

To take advantage of this feature, download and install the “Hortonworks ODBC Driver for SparkSQL (1.5 & 1.6)” (subject to Hortonworks license agreements), which you’ll use to set up the connection between SAP Vora and SAP Data Services.

You can obtain this ODBC driver and documentation from the Hortonworks download site here: <https://hortonworks.com/downloads> . Refer to Hortonworks documentation for details on how to set up the driver.

For the Vora connection, select **SparkThriftServer** for the *Spark Server Type*, **SASL** for the *Thrift Transport* option, and enable **Use native Query** in *Advanced Options*.

This ODBC driver relies on one of the following ODBC driver managers on Linux:

- iODBC 3.52.7 or later
- unixODBC 2.2.14 or later

As a result, in the `$LINK_DIR/bin/ds_odbc.ini` file, specify the driver manager library for the Vora DSN. For example:

```
[VORA_DSN]
```

```
Driver=/usr/local/unixODBC/2.3.4/lib/libodbc.so
```

This ODBC driver is supported on Windows and SUSE Linux Enterprise Server 11.

Once you’ve made the connection between SAP Vora and SAP Data Services, you can start designing your jobs for SAP Vora in the SAP Data Services Designer. Refer to the *SAP Data Services Designer Guide* for details.

Additional SAP Vora functionality will be included in future releases of SAP Data Services.

Data Services connects to your Hadoop cluster in the cloud

SAP Cloud Platform Big Data Services

Big Data Services is your Hadoop distribution in the cloud. It performs all Hadoop upgrades and patches for you, and provides Hadoop support. enables you to access your Hadoop cluster in the cloud, without having to maintain and update a Hadoop installation.

Now you can upload your big data files from SAP Data Services to Big Data Services enables you to access using SSH tunneling directly from your computer. For example, access data from S3 (Amazon Simple Storage Service) and use the data as a source in a Data Services data flow. Execute the job and upload the target data to your HDFS that resides in Big Data Services.

Read all about this new feature in the *Supplement for SAP Cloud Platform Big Data Services*.

7.3 Data Quality transforms

In SAP Data Services 4.2 SP9, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

7.3.1 Global Address Cleanse transform

In SAP Data Services 4.2 SP9, the Global Address Cleanse transform has been enhanced with the following features.

Puerto Rico address assignment enhancement

The US engine of the Global Address Cleanse transform is now more consistent with all of the other supported country engines when parsing the primary street name and primary street type.

New directories available for Global Address Cleanse

The following directories provide a more consistent behavior for Global Address Cleanse (GAC) suggestion lists:

- SAP ADDR CANADA 4.X ENH
 - ga_ca_paf.dir
- SAP ADDR UNITED STATES 4.X ENH
 - ga_us_paf.dir

To use these directories to process US and CA addresses using the Global Engine, enable the Global Engine and disable the US and CA Engines. If you leave the US and CA Engines enabled, GAC will use the old US and CA Engine directories to process the addresses.

If you are happy with the current suggestion list behavior, you do not need to download these directories. You can always download the directories later to take advantage of functionality added in future service packs.

For more information, see “Downloading directory files” in the *Windows or UNIX Installation Guide* and “Global Address Cleanse” in the *Reference Guide*.

7.4 Usability

SAP Data Services 4.2 SP9 includes the following enhancements that improve the user experience.

Support Arabic right to left in COBOL Copybooks

Data Services supports EBCDIC Arabic characters from the CP420 code page for COBOL copybooks. The software outputs Arabic characters in the left to right order along with all other characters such as English words or numbers.

To change the output order of Arabic characters from the CP420 code page, add a flag named `REVERSE_RTL=TRUE` to the `DSConfig.txt` file.

Caution

The flag affects every COBOL reader in your system that uses the CP420 code page. Therefore, make sure that your other projects that use the CP420 code page are not adversely affected by setting this flag.

Note

After you set the flag, the software maintains the usual order, left to right, for all non-Arabic output characters.

Before you set the flag, perform a test to determine if your data system already outputs Arabic from right to left.

After you set the flag, perform a test to make sure that the software is outputting Arabic characters in the right to left order.

This information is documented in [241158](#) .

Follow these steps to set the flag:

1. Open `DSConfig.txt` located in `<DS_COMMON_DIR>/conf/`.
2. Find the `[AL_Engine]` section.
3. Add a new line and enter the following command: `REVERSE_RTL=TRUE`.
4. Save and close the `DSConfig.txt` file.

8 SAP Data Services 4.2 SP 8 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 8.

The following links take you to descriptions of the features by category for version 4.2 SP 8.



[Connectivity: Big Data and CDC \[page 68\]:](#)

- New support for Hadoop Hive
- Amazon Redshift support
- Google BigQuery template table
- New Google built-in function for performance optimization



[Data Quality transforms \[page 70\]:](#)

- Geocoder transform
 - Group output fields in Result_List XML output fields
- Global Address Cleanse transform
 - Certified Australia options
 - Certified Canada options
 - Certified New Zealand options
 - La Poste ratification
 - Unique error numbering
 - Suppress Suggestion List with No Range option



[Platform transforms \[page 72\]:](#) DQM Microservices transform

8.1 Connectivity: Big Data and CDC

SAP Data Services 4.2 SP8 includes the following connectivity enhancements.

New support for Hadoop Hive

With this release, SAP Data Services supports edge nodes (also known as “gateway” nodes) running on Hadoop Hive clusters. Edge node support is for Linux only; SAP Data Services does not support using edge nodes on Windows.

In addition edge node support, SAP Data Services now supports creating template tables in either Parquet or Table format.

For more information, see “Using Hive template tables” in the *Data Services Supplement for Adapters*

Amazon Redshift support

Amazon Redshift is a data warehouse application for relational database query and management systems.

You can connect to the Redshift cluster database on Windows and Linux platforms by using the Amazon Redshift ODBC driver. For information about downloading and installing the Amazon Redshift ODBC driver, see the Amazon Redshift documentation on the Amazon website.

After creating a Redshift database datastore, you can do the following:

- Use Redshift tables as sources or targets in your data flows.
- Preview data.
- Create and import template tables.
- Load S3 data files into a Redshift table using the `load_from_s3_to_redshift` function.

For more information, see the following topics in the *Reference Guide*:

- “Amazon Redshift”
- “Amazon Redshift source”
- “Amazon Redshift target table options”
- “Amazon Redshift data types”
- “load_from_s3_to_redshift”

For information about template tables and data preview, see the *Designer Guide*.

Google BigQuery template table

Use a Google BigQuery template table as a target in a data flow. When you execute the data flow, the software automatically creates the table in your Google account in the specified project and dataset. The software creates a schema for the template based on your Google BigQuery dataset, and accesses your Google account using information from the applicable Google BigQuery application datastore.

To use the template table as a source, include a transform as a reader to pull specified data from the target table. For example, include a SQL transform in your data flow that includes a SQL statement that pulls the first 1000 records from the specified target table.

See the *Supplement for Google BigQuery* for more information.

New Google built-in function for performance optimization

Use the Google built-in function named `gbq2file` to optimize your software performance when you extract large volumes of data from Google BigQuery results to your local machine.

This function exports results of a Google BigQuery to files in your Google Cloud Storage (GCS), and then transfers the data from GCS to a user-specified file on your local machine.

For details about the syntax and return value for the new function, see the *Reference Guide*.

8.2 Data Quality transforms

In SAP Data Services 4.2 SP8, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

8.2.1 Geocoder transform

Group output fields in Result_List XML output fields

Result records in the Result_List XML output field can now be grouped so that records that represent the same address are in the same group. There are two common reasons why the search results may include multiple occurrences of the same address:

- A street may have multiple names. For example, “North Ave” and “Highway 5” may be names for the same street. In this case, the search results may have two records for “1800 North Ave” and “1800 Highway 5”, both with the same latitude and longitude values.
- Multiple organizations may reside in the same building. For example, “Highland Insurance Group” and “Granite Communications” are on different floors in a building with the address “510 Lakeshore Dr”. In this case, the search results may have two records, both with “510 Lakeshore Dr” and both with the same latitude and longitude values, but the records have different organization names in the POI_Name field.

The following output fields have been added to the Geocoder transform Result_List XML output fields:

Output field name	Category	Description
Group_Number	Group	An identifier for multiple records that represent the same address. For example, one record might have an alias primary name of another. This output field has a value only when the same address is returned in multiple entries. When entries have the same value in Group_Number, it means that the entries have the same address.

Output field name	Category	Description
Group_Rank	Group	<p>Indicates whether a record in a group is a master or a subordinate. For each group of entries that have the same Group_Number value, one of them has Group_Rank of "M" (master) and all other entries in the group have "S" (subordinate). You may obtain only one of each address returned by filtering where the Group_Rank is either "M" or a blank string.</p> <p><i>M</i>: Master</p> <p><i>S</i>: Subordinate</p>

These output fields are used only for reverse geocoding.

i Note

To accommodate the new group output fields, the maximum length of the Result_List output field has been changed from 60,000 to 100,000. This may require that you update your existing jobs that put a length limitation of less than or equal to 60,000 on the field. Otherwise, the Result_List output field may be truncated.

8.2.2 Global Address Cleanse transform

In SAP Data Services 4.2 SP8, the Global Address Cleanse transform has been enhanced with the following features.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2016 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2016 certification.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2016 certification.

La Poste ratification

The France processing options meet all requirements for La Poste 2016 certification.

Unique error numbering

Some of the more common Global Address Cleanse and Geocoder engine error messages have been enhanced to provide unique message IDs rather than a generic error message identifier.

Suppress Suggestion List with No Range option

An option has been added to the Global Address engine that lets you specify whether a suggestion list is generated when "S/N" or no range is found on input and the only records found in the reference data are range records.

Option	Description
Suppress Suggestion List With No Range	<p>Specifies whether a suggestion list is generated when "S/N" or no range is found on input and the only records found in the reference data are range records.</p> <p>Yes: A suggestion list is not generated, information code 3030 is returned, and a primary name assignment level match is made.</p> <p>No: A suggestion list is generated with the range values in the reference data record. This is the default value.</p>

8.3 Platform transforms

In SAP Data Services 4.2 SP8, the set of transforms that you use for general data movement operations has been enhanced with the following features.

8.3.1 DQM Microservices transform

The DQM Microservices transform helps you configure and execute SAP Data Quality Management, microservices for location data services within Data Services.

SAP Data Quality Management, microservices for location data offers cloud-based microservices for address cleansing, geocoding, and reverse geocoding. You can embed address cleansing and enrichment services within any business process or application so that you can quickly reap the value of complete and accurate address data.

The DQM Microservices transform helps you configure and execute a subset of DQM microservices for location data services within Data Services. The DQM Microservices transform supports the following DQM microservices services:

- **addressCleanse service:** With address cleansing, you can quickly and easily correct, parse, standardize, validate, and enhance address data, and assign geocodes.
- **reverseGeo service:** Reverse geocoding translates latitude and longitude coordinates into address(es).

For more information about DQM microservices, see the documentation set at <https://help.sap.com/viewer/d95546360fea44988eb614718ff7e959/Cloud/en-US>.

i Note

The DQM Microservices transform functionality is available when DQM microservices is generally released. For information about the availability of SAP Data Quality Management, microservices for location data, see [2392827](#) - Availability of DQM microservices for location data in SAP Data Services 4.2 SP8. A trial landscape is also available.

For more information, see “DQM Microservices” in the Transforms chapter of the *Reference Guide* and “Connecting to SAP Data Quality Management, microservices for location data” in the *Supplement for SAP*.

9 SAP Data Services 4.2 SP 7 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 7.

The following links take you to descriptions of the features by category for version 4.2 SP 7.



[Connectivity: Big data and CDC \[page 75\]:](#)

- SSL protocol for HP Vertica
- Hive on Spark
- New Beeline command-line interface
- Support for parallel OData calls
- Template table indexes for SQL Server template tables
- Moving files to and from Azure blob containers
- Microsoft Azure PaaS and Azure VM support on SQL Server
- Google Cloud Storage support
- Template documents for MongoDB
- MongoDB document data preview
- HP Vertica database datastore
- SSL encryption for SAP HANA and Microsoft SQL Server database datastores
- DP Bridge Outlook adapter
- Support for Amazon SE
- Support for SAP HANA Identity column



[Data Quality transforms \[page 80\]:](#)

- Downloading directories via the SAP ONE Support Launchpad
- Geocoder transform: Enhancements to TomTom reference data
- Global Address Cleanse transform
 - German character conversion
 - Assignment using only lastline information
 - Enhanced suggestion lists
 - Redefine suspect counts
- USA Regulatory Address Cleanse transform: SHA-256 directory format support



[Platform transforms \[page 83\]:](#) User-Defined transform in the platform



Text Data Processing [page 83]:

- Updates to the *Language Reference Guide* and the *Extraction Customization Guide*
- Expanded Language Support
- Redesigned Voice of Customer (VOC) mechanics
- Dictionary format has been modified
- Improved XML processing

9.1 Connectivity: Big data and CDC

SAP Data Services 4.2 SP7 Corrections includes the following connectivity enhancements.

SSL protocol for HP Vertica

SSL encryption for HP Vertica database datastore is available in SAP Data Services version 4.2 SP7 Patch 1 or later.

To configure and enable SSL protocol for HP Vertica datastore, you install MIT Kerberos and configure it for HP Vertica using the specifications from HP Vertica documentation. Additionally, you create an SSL data source (DSN) for the datastore using the ODBC Data Source Administrator. See the *Administrator Guide* for more information.

Hive on Spark

The ability to load data using the Spark engine on Hive becomes a supported feature with the 4.2 SP7 Patch 1 release.

Before you can take advantage of Hive on Spark, you need to make sure that Spark is enabled, and that you choose either [Spark](#) or [Map Reduce](#) as the execution engine on the [Adapter Source](#) tab of the Hadoop Hive adapter source options.

For details, see the topic discussing Hadoop Hive adapter source options in the *SAP Data Services Supplement for Adapters* guide.

New Beeline command line interface

Starting with the 4.2 SP7 Patch 1 release, SAP is deprecating the Hive command line interface (CLI), and replacing it with the Beeline CLI. Data Services will continue to function in the same way as it did with the Hive CLI.

Before you can use the Beeline CLI, your Hadoop system administrator will need to create an HDFS working directory with permission 755 for all of the directories in the path, and this should be specified in the new field *HDFS Working directory* in the Hive datastore configuration options window. If you don't provide the path, SAP Data Services will use `/user/sapds_hivetmp` as the default.

With Beeline CLI, the user name value in the Hive datastore is mandatory. This user name should be the same as the user name in the Data Services job server that is installed and running.

Refer to the *Supplement for Adapters* to learn more about Hive adapter datastore configuration options.

Support for parallel OData calls

Starting with version 4.2 SP7 Patch 1, Data Services now supports parallel fetches using an OData adapter. The following enhancements are supported in this release:

- Supports push-down datetime type to OData \$filter.
- Uses base64 encoding for data of binary type.
- Configure the default size of varchar, instead of using 100, using the new OData adapter datastore configuration option *Default Varchar Length*.
- Control the number of concurrent threads in the reader by using the new option *Number of Concurrent Threads*.
- Memory size (-Xmx) depends on 'batch size', 'thread numbers', 'response data size', 'Concurrent reader' For example (8 CPU, 8G Memory):
batchsize=10,000; threadNum=10; response data size = 0.35k/row; Concurrent reader=1
Memory= 128M*1+(10,000*10*0.35k)/1024 * 10 =470M, set -Xmx=512

Template table indexes for SQL Server template tables

Starting with version 4.2 SP7 Patch 1, you can create indexes for SQL Server template tables and define which columns the software uses to create the index. The index can be clustered, non-clustered, or unique.

When you create a template table, the software creates an index for that table, which you can edit by going to **► Properties ► Indexes ►**. You can edit template table indexes created from supported datastores only (SQL Server).

For more information, see "Template table indexes" in the *Designer Guide*.

Moving files to and from Azure blob containers

Data Services lets you move files from local storage such as a local drive or folder to an Azure container. You can use an existing container or create one if it does not exist. You can also import files (called "blobs" when in a container) from an Azure container to a local drive or folder. The files can be any type and are not internally manipulated by Data Services. Currently, Data Services supports the block blob in the container storage type.

You use a file format to describe a blob file and use it within a data flow to perform extra operations on the file. The file format can also be used in a script to automate upload and local file deletion.

For more information, see “Moving files to and from Azure containers” in the *Designer Guide* and “Azure Cloud Storage protocol” in the *Reference Guide*.

For information about supported platforms, see the *Product Availability Matrix* (PAM) at <https://apps.support.sap.com/sap/support/pam>.

Microsoft Azure PaaS and Azure VM support on SQL Server

In this release, Data Services adds support for Azure platform as a service (PaaS) and Virtual Machines (VMs) on Microsoft SQL Server databases a new option for datastore subtype.

- [SQL Server as a service on Azure](#)

You can install Data Services locally to a network and use an Azure Cloud SQL Server database for a datastore. When you use a remote Azure Cloud database as a datastore, it is recommended that you set the target accordingly to [Azure PaaS](#) or [Azure VM](#). The selection depends on whether the database is an Azure Managed PaaS database or a SQL Server installed on an Azure Virtual Machine. Setting the proper database subtype ensures a secure connection by encrypting the connection using SSL.

- [SQL Server running on an Azure VM](#)

It is recommended to install the repository on the same computer as the Azure Virtual Network. It can be on the same virtual machine as where you install Data Services or another Azure Virtual Machine. This is strictly for performance reasons as a remote repository can cause a serious impact on job execution performance.

For more information, see “Microsoft SQL Server” in the Datastore section of the *Reference Guide*.

For information about supported platforms, see the *Product Availability Matrix* (PAM) at <https://apps.support.sap.com/sap/support/pam>.

Google Cloud Storage support

Data Services has added support for Google Cloud Storage so that you can:

- Upload and download files to Google Cloud Storage and on local storage.
- Load objects in Google Cloud Storage into Google BigQuery.

The `load_from_gcs_to_gbq` function has been added to help you transfer data from a Google Cloud Storage into Google BigQuery tables.

For more information, see “Google Cloud Storage protocol” and “load_from_gcs_to_gbq” in the *Reference Guide*.

Template documents for MongoDB

Use template documents as a target in one data flow or as a source in multiple data flows. Template documents are particularly useful in early application development when you are designing and testing a project.

Template documents can be found in the [Template Documents](#) node under the MongoDB datastore.

Importing a template document converts it to a regular document. You can use the regular document as a target or source in your data flow.

i Note

Template documents are available in Data Services 4.2.7 and later. If you are upgrading from a previous version, you need to edit the MongoDB datastore and then click [OK](#) to see the [Template Documents](#) node and any other template document related options.

For more information, see “Creating template documents” and “Converting a template document to a regular document” in *Supplement for Adapters*.

MongoDB document data preview

Data preview allows you to view a sampling of MongoDB data from documents.

To preview MongoDB document data, right-click on a MongoDB document name in the Local Object Library or on a document in the data flow and then select [View Data](#).

You can also click the magnifying glass icon on a MongoDB source and target object in the data flow.

i Note

By default, the maximum number of rows displayed for data preview is 100. To change this number, use the [Rows To Scan](#) adapter datastore configuration option. Enter -1 to display all rows.

For more information, see “Previewing MongoDB document data” and “MongoDB adapter datastore configuration options” in *Supplement for Adapters*.

HP Vertica database datastore

Now connect natively to data in HP Vertica using an HP Vertica database datastore.

After you create the HP Vertica datastore you can import HP Vertica tables. Use the tables as source or target in a data flow so that you can read from or load to your HP Vertica database. You can additionally create HP Vertica template tables.

Read about HP Vertica database datastores in the *Reference Guide* in “HP Vertica”.

SSL encryption for SAP HANA and Microsoft SQL Server database datastores

Add the security of SSL encryption for both SAP HANA and Microsoft SQL Server databases with this release of SAP Data Services. SSL encrypts data that is transmitted across a network between the database server and SAP Data Services. Create a database datastore with SSL enabled and use it as a source or target in a data flow.

For example, configure your Microsoft SQL Server database for SSL, and then create a Microsoft SQL Server database datastore with SSL enabled. When you use the datastore in a data flow, data transmits from the database server to SAP Data Services with the added security of SSL encryption.

Learn more about configuring and using SSL encryption for SAP HANA and Microsoft SQL Server databases in the following topics in the *Administrator Guide*:

- Configure SSL for Microsoft SQL Server
- Configure DSN SSL for SAP HANA

And see the following topics in the *Designer Guide*:

- Creating a Microsoft SQL Server database datastore with SSL encryption
- Creating an SAP HANA database datastore with SSL encryption

DP Bridge Outlook adapter

Now you can access and import message data from a local Outlook PST file to use as a source in an SAP Data Services data flow. Use the new Data Provisioning (DP) Bridge adapter and create an instance in SAP Data Services Management Console. Then create a DP Bridge adapter in Data Services specifically for Outlook data. Use the DP Bridge Outlook adapter datastore as a data flow source to import a message table and an attachment table. The message table contains information such as subject, sender name, creation time, and recipient names. The attachment table contains information such as file name, comments, and content.

i Note

Several fields from the message and attachment tables are in large object format (such as CLOB or LOB). Therefore, when you create the adapter datastore, you can choose whether to import large objects. If you choose to import large objects, set the field size so the software doesn't import the entire field. Importing large data objects will slow job processing.

For complete information about creating an adapter instance and creating an adapter datastore for the DP Bridge Outlook adapter, see “Creating an adapter datastore” and “DP Bridge Outlook adapter datastore options” in the *Supplement for Adapters*.

Support for Amazon S3

SAP Data Services now supports importing and exporting data to and from Amazon's cloud product, Amazon S3 (Simple Storage Service). To take advantage of this new enhancement, you simply create a file location object that tells SAP Data Services where the file is located.

For complete information on creating a file location object for Amazon S3, see the *Reference Guide*.

Support for SAP HANA Identity column

Prior to this release, if you imported a table containing an SAP HANA Identity column, you would receive an error. With this release, SAP Data Services now supports SAP HANA Identity columns. To take advantage of this enhancement, you'll need to map the Identity column, and make sure the *Identity Upsert* option in the *General options* tab in *General settings* is turned *On*.


If you have any existing SAP HANA jobs that you plan on updating to SAP Data Services 4.2 Service Pack 7, you'll need to re-import the target SAP HANA table. The job will behave the same as it did prior to this release.

For complete information setting up the Identity column, see the *Reference Guide*.

9.2 Data Quality transforms

In SAP Data Services 4.2 SP7, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

Downloading directories via the SAP ONE Support Launchpad

Now you can use the SAP ONE Support Launchpad at <https://launchpad.support.sap.com/#/softwarecenter>  to easily select all of the address and geocoding directories that you want to download and add them to your Download Basket. For more information, see the *Installation Guide* or the directory update letter for the appropriate directories.

The Directory Update Assistant utility has been deprecated.

9.2.1 Geocoder transform

In SAP Data Services 4.2 SP7, TomTom reference data for the Geocoder transform has been enhanced with the following features, depending on the country.

- *Map coverage*. More streets are available in some regions.

- *Post data coverage.* More addresses from post data (GDM directories) are covered by new directories
- *Postcode1 (P1) centroid assignment level.* With TomTom data, almost all countries provide P1 assignment level in centroid cascading. Previously, the directories provided only L1 assignment.
- *Better postal friendly locality.* Locality name from TomTom data gets closer to address directory and well aligns the flavor of post.
- *Higher quality in postcode data.*

9.2.2 Global Address Cleanse transform

In SAP Data Services 4.2 SP7, the Global Address Cleanse transform has been enhanced with the following features.

German character conversion

“ß” character conversion. Starting with Data Services 4.2 SP7 Patch 1, the German “ß” character is converted to “ss” during the normalization process in all cases. It outputs as "ss" when converting to US ASCII in the standardizer for best components, and all match components convert it to "ss". During the phrase matching process, "s" is no longer considered a match to “ß”, but "ss" is a match for “ß”. One exception is the street type of "strasse", where "strase", "strasse", and "straße" are considered equal.

Umlaut conversion. Starting with Data Services 4.2 SP7 Patch 1, the Global Address Cleanse transform converts umlauts correctly. The transform creates a normalized version of the input address by converting all non-English characters to an English alphabet equivalent. This includes converting all accented characters to a common base character; for example, all accented versions of the letter “e” are converted to an “e”. In the German language, an “ä” becomes “ae”, “ö” becomes “oe”, and “ü” becomes “ue”. This affects the match scores to determine if a component is corrected and final output when converting to ASCII output in the standardizer.

Assignment using only lastline information

The Global Address Cleanse engines now allow address processing with only lastline input mapped. Previously, at least one address line input field was required to process any address, and the Lastline input field was not required if one or more multiline input lines were defined. Now, an address with only lastline information is processed and given the best possible lastline assignment, and an address line-related information codes will not be issued. The possible input line formats for lastline are:

- Lastline
- Locality1–4 only
- Postcode1–2 only
- Locality1–4 and Postcode1–2

i Note

The Region1–2 input field is optional.

The Country input field must be mapped.

If you want to assign on lastline information, you must remove the address line input.

Enhanced suggestion lists

Suggestion lists have been enhanced in the the following ways:

- *Partial primary name and/or locality matching for suggestion lists* Previously, Data Services has supported partial primary name matching for suggestion lists, but locality names had to be entered completely to find matches. Now Data Services finds locality matches with as few as three characters on input for all countries. Both begins-with and ends-with matching are supported.
- *Follow-up lastline suggestion lists.* Previously, suggestion list processing only allowed a single lastline list before it moved on to an address line list or complete processing. This prevented you from choosing all sublocality data and, in some cases, the postcode. Now, Data Services provides follow-up lastline suggestion lists so that all lastline data can be selected before moving on to the address-line data.
- *Simplified suggestion list output by combining adjacent ranges*

Redefine suspect counts

In this release, Data Services now calculates the number of suspects in the Generate Cleanse Statistics table (CLEANSE_STATISTICS_TABLE) differently. The Generate Cleanse Statistics table tracks the number of records, the number of blanks, the number of suspects, and the number of high significant changes. Previously, the number of suspects was a count of records whose assignment information is set to "I". Now, the number of suspects is based on a subset of information codes. Any information code less than 4000 plus 5000, 5010, 5040, and 5050 is flagged as a suspect address.

For more information about the Generate Cleanse Statistics table, see "Data quality statistics settings" in the *Reference Guide*.

9.2.3 USA Regulatory Address Cleanse transform

In SAP Data Services 4.2 SP7, the USA Regulatory Address Cleanse transform has been enhanced with the following feature.

SHA-256 directory format support

SAP supports SHA-256 directory format starting in SAP Data Services 4.2 SP7 Patch 1. The SHA-256 directories are planned to be released in Q3 2016.

The USPS announced a change to its encrypted data products (including DPV, DSF2, LACSLink, NCOALink, SuiteLink, and RDI) that will require implementation of new compliant SHA-256 code in all SAP solutions that

leverage encrypted USPS data before August 2017. The USPS will ship the last old style SHA1 databases in July 2017.

For a limited time, Data Services and Data Quality Management SDK will support both the previous SHA1 and the new SHA-256 directory formats; however, you cannot mix the SHA1 and SHA-256 formats. You must use one format or the other. SAP will continue to provide and support the previous directory format into 2017, but will discontinue support before July 2017.

i Note

This change does not affect the U.S. National Directories. Programs and customers that do not require CASS or other USPS certifications should not be affected.

Starting in the July 2016 directories, the U.S. directories include a header file that the USPS ships with RDI called `cdhdr01.dat`. This is required to support the SHA-256 format, but does not affect current processing.

9.3 Platform transforms

In SAP Data Services 4.2 SP7, the set of transforms that you use for general data movement operations has been enhanced with the following features.

User-Defined transform in the platform

The User-Defined transform has been moved from the Data Quality category to the Platform category. This makes the User-Defined transform with its flexible Python code available to more Data Services users.

This change should have no impact on User-Defined transforms in existing Data Services jobs.

For more information, see “User-Defined” in the Transforms chapter and the Python chapter of the *Reference Guide*.

9.4 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Updates to the Language Reference Guide and the Extraction Customization Guide

Updates to the *Extraction Customization Guide* are minor.

The *Language Reference Guide* has been extensively reorganized to reduce redundancy. The linguistic information has been gathered into one section, organized by feature, then by language. The entity extraction material has been gathered into a separate section. Please explore the table of contents to see the new structure.

The guide is now published only in HTML, and uses some new features such as a dynamic table (datatable) to display the extensive list of predefined language-specific entity type examples. Readers can select a set of languages to display, and can search and sort and filter the content. See “Language-specific Entity Type Examples” in the *Text Data Processing Language Reference Guide*.

- General
 - The entity type NOUN_GROUP is now supported for all languages except Greek.
 - PERSON and PRODUCT extractions have been improved for all languages except Arabic, Farsi, and Russian.
 - Part-of-speech tags can now be used in custom extraction (CGUL) rules for all languages except Greek.
 - Automatic language identification has been improved for Farsi, Polish, Thai, and Turkish.

Expanded Language Support

- Chinese
 - (New) Simplified and Traditional Chinese processing has been consolidated to provide consistent features across Simplified and Traditional Chinese, lower storage requirements, and allow for mixed Simplified and Traditional Chinese text in a single document.
 - Improved word segmentation
 - (New) Voice of Customer (VOC) now also supports Traditional Chinese. See the Sentiment Extraction examples in the *Language Reference Guide*.

i Note

See the information in the Behavior Changes in Version 14.2.7

- (New) Emoticon and profanity extraction
 - Traditional Chinese supports additional predefined entity types
 - Improved social media extraction
 - Improved numerical extraction
- Dutch
 - (New) Emoticon and profanity extraction
 - Expanded support for nonstandard spellings
- English
 - Expanded support for nonstandard spellings
 - Voice of Customer (VOC) topics now omit determiners, which simplifies topic aggregation for downstream analysis

- Improved Enterprise fact extraction
- German
 - Expanded support for nonstandard spellings
 - Voice of Customer (VOC) topics now omit determiners, which simplifies topic aggregation for downstream analysis
- Indonesian
 - (New) Standard language support, including:
 - NOUN_GROUP extractions
 - Part-of-speech tagging (used in custom extraction rules)
- Italian
 - (New) Voice of Customer (VOC) support. See the Sentiment Extraction examples in the *Language Reference Guide*.
 - (New) Emoticon and profanity extraction
 - Expanded support for nonstandard spellings
- Japanese
 - Improved social media extraction
- Korean
 - Improved predefined entity extraction
- Portuguese
 - (New) Voice of Customer (VOC) support. See the Sentiment Extraction examples in the *Language Reference Guide*.
 - (New) Emoticon and profanity extraction
- Russian
 - (New) Voice of Customer (VOC) support. See the Sentiment Extraction examples in the *Language Reference Guide*.
 - (New) Emoticon and profanity extraction
 - Additional predefined core extraction
- Serbian
 - (New) Cyrillic script support

Voice of Customer (VOC) mechanics have been redesigned for much easier customization and improved classifications

Sentiment keywords have been moved into separate “thesaurus” dictionaries, which allows you to customize VOC by simply adding or changing entries in a dictionary.

i Note

See the information in the Behavior Changes in Version 14.2.7

Dictionary format has been modified

To reduce storage requirements, the dictionary format has been modified. Because of the format change, dictionaries compiled with new TDP dictionary compiler cannot be used with previous releases of TDP. However, dictionaries compiled using an older TDP dictionary compiler will continue to work with this version of TDP –that is,, they do not have to be recompiled.

Improved XML processing

The textual content of XML documents is now processed as a single field. Previously, text data processing was performed on each XML element separately. This results in better language detection and generally produces more desirable TDP output for most applications.

10 SAP Data Services 4.2 SP 6 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 6.

The following links take you to descriptions of the features by category for version 4.2 SP 6.



[Installation, administration, and monitoring \[page 88\]:](#)

- Configuring Data Services for CTS+
- Windows upgrade: Change location for DS_COM-MON_DIR
- Job Server and Access Server availability
- Download a debug package
- Improved RFC tracing
- Use SAP Server Manager to update SLD settings
- Ensure only one instance of a job runs at one time



[Connectivity: Big data and CDC \[page 91\]:](#)

- OData V4 support
- Spatial data support in data Services
- MongoDB as a target
- SAP HANA partition support
- New MongoDB authentication type
- Re-importing a MongoDB schema
- Use Google BigQuery table data as a source
- Enhanced append features for file location objects
- Template table support for Hive
- SASL-QOP support with Kerberos on Hive
- Hadoop support for Windows
- Varchar and Char data type support for Hive tables
- RAW and STRING data type support



[Functions and transforms \[page 95\]:](#) Push-down functions, transform, and operators



[Usability \[page 95\]:](#)

- Viewing objects before import
- Documentation set updates
- View related objects in the object library and the Variables and Parameters window
- Enhanced filters for substitution parameter search



[Data Quality transforms \[page 97\]:](#)

- Data Cleanse transform
 - Generate data quality statistics
 - Firm_Line input field
- Geocoder transform
 - Generate data quality statistics
 - Brazil support
- Global Address Cleanse transform
 - Certified Australia options
 - Certified Canada options
 - Certified New Zealand options
 - La Poste ratification
 - Enhanced France data directories
 - Generate data quality statistics
 - South Korea five-digit postcode
 - Brazil quick data entry
 - Macao address cleansing



[Platform transforms \[page 100\]:](#) Support for identity column

10.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP6 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Configuring Data Services for CTS+

Starting with SAP Data Services 4.2 SP6 Patch 1, Data Services Management Console contains tools to help you configure Data Services for the Enhanced Change and Transport System (CTS+).

CTS+ can transport changed objects between SAP systems in your landscape. For example, use CTS+ to export changed data flows from your Data Services development system to your test and production systems. You can transport the following objects between Data Services systems: projects, jobs, work flows, data flows, file formats, functions, and datastores.

CTS+ and the CTS system automatically keeps track of your system landscape changes, and protects against all uncoordinated, parallel changes from being made to objects. CTS makes sure that all of the appropriate objects are transferred in an automatic and controlled manner.

Data Services Management Console contains new CTS tabs for configuring Data Services for CTS+. Additionally there is a Promotion History node that provides information about completed exports and imports. The new tools are in the Administrator application under the Object Promotion node.

Object Promotion nodes

Node	Description
Export Configuration	Complete the options in the CTS tab to connect to the CTS Export Web Services.
Import Configuration	Specify the target repository for the SAP target systems.
Promotion History	Information about all completed CTS exports and imports including operation logs.

For more information, see the new *SAP Data Services Configuration Guide for CTS+*.

Windows upgrade: Change location for DS_COMMON_DIR

If you are updating your current Data Services installation on Windows, you can change the DS_COMMON_DIR location during upgrade.

The DS_COMMON_DIR directory contains configuration files and log files that Data Services components read from and write to during installation, upgrade, and runtime. The directory that you change to must be a fixed, local directory, have sufficient disk space for the configuration files and log files, and be available for all users.

i Note

There will be optional post installation steps to follow to move your previous log files to the new directory.

For more information, see “Change DS_COMMON_DIR location” in the *Installation Guide*.

Job Server and Access Server availability

You can now retrieve information about Job Server and Access Server availability. The following new web service operations have been added:

- Validate_SessionID
- Get_AS_Info
- Get_RTService_Info
- Get_RTService_Statistic
- Get_Jobserver_List
- Get_Jobserver_Status

For more information, see “Supported web service operations” in the *Integrator Guide*.

Download a debug package

You can download a job debug package from the Management Console from its location on the machine where the Job Server is installed. To download the debug package, click the Download Logs/Debug Package option on the Batch Job Status page and open or save the .zip or .tar file. The download option is available even if the package has not been created or is not available. If the package is not available on the Job Server, a package containing log files is downloaded and a message alerts you.

The debug package includes the ATL and trace, monitor, and log files. You can send the generated .zip or .tar file to SAP customer support to help you troubleshoot execution-related issues. You can generate the debug package as you execute the job either from the Execution Properties window of the Designer or in the Execution Options page of the Management Console Administrator.

For more information, see “Downloading a debug package” in the *Management Console Guide*.

Improved RFC tracing

In this release, we have improved tracing by providing configurable trace information for SAP RFC function calls. With the new Trace RFC function option in the Designer and Management Console, you can enable RFC tracing during job execution. Its default value is off. You can also choose whether to output brief, verbose, or full details using the RFC Trace Level option in the Designer datastore editor for SAP Applications or in the Management Console.

For more information, see the *Management Console Guide* and *Supplement for SAP*.

Use SAP Server Manager to update SLD settings

Now you can use SAP Server Manager and the new SLD (System Landscape Directory) tab to register Data Services with SLD. Previously, you may have used a utility named ACTool, which is no longer necessary with this release.

For more information, see the topic “Managing System Landscape Directory registration” in the *Administrator Guide*.

Ensure only one instance of a job runs at one time

You can now use the Single instance option to tell the system to abort a second instance of any batch job.

For more information, see the topic “Single instance job” in the *Reference Guide*.

10.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP6 includes the following connectivity enhancements.

OData V4 support

Starting with Data Services 4.2 SP6 Patch 3, Data Services supports OData V4. A new value (AUTO) has been added to the OData Version configuration option that automatically detects the OData version that you are using from the URL.

i Note

Job migration between OData V2 and V4 is not supported because each version uses different metadata. OData V3 is also not supported.

The OData adapter uses the odata4j (supports V1 and V2) and Apache Olingo (supports V2 and V4) libraries. For more information about OData libraries, see <http://www.odata.org/libraries/> .

In addition, Data Services now supports the following OData data types for V4:

- TimeOfDay
- Duration
- Date

For more information, see “OData adapter datastore configuration options” and “Metadata mapping for OData” in the *Supplement for Adapters*.

Spatial data support in Data Services

Now you can use Data Services spatial data capabilities for reading Microsoft SQL server databases. For complete information about spatial data support, see the *Supplement for SAP*.

MongoDB as a target

You can now use MongoDB as a target in your data flow.

After you create your MongoDB adapter instance and datastore, you can set the following options in the *Adapter Target* tab of the target document editor:

- Use auto correct
- Write concern level
- Use bulk
- Use ordered bulk

- Documents per commit
- Bulk maximum rejects
- Delete data before loading
- Use audit

For more information about these options, see “MongoDB as a target” in *Supplement for Adapters*.

SAP HANA partition support

SAP HANA supports partition support for column store tables. SAP Data Services supports the SAP HANA partition feature for parallel reading and loading using physical partitions and logical partitions. You can import an SAP HANA partition table metadata for hash partitioned tables used for parallel reading and loading.

The following enhancements have been made to SAP HANA partition support in this release:

- You can now import SAP HANA hash and range partition information into Data Services Designer and view the details. You cannot modify a hash partition.
- You can read from SAP HANA hash and range partitioned tables done in parallel using partition range values. On SAP HANA SP 10 and later, a SELECT statement with partition clauses is used when possible; otherwise, a WHERE clause is used for a range partition as in previous versions.
- Composite range partitions are imported as a user-modified single range partition and WHERE clauses are used for parallel processing.

For more information, see the *Performance Optimization Guide*.

New MongoDB authentication type

A new [Authentication Type](#) has been added to the MongoDB adapter. [SCRAM-SHA-1](#) authenticates a user's credentials by using the user's name, password, and the database on which the user was created.

i Note

SCRAM-SHA-1 is the preferred mechanism for MongoDB versions 3.0 and later. It is not supported in earlier versions.

For more information, see “MongoDB adapter datastore configuration options” in *Supplement for Adapters*.

Re-importing a MongoDB schema

You can re-import a MongoDB schema into the Local Object Library. The software honors the MongoDB adapter datastore settings when re-importing.

- To re-import a single document, right-click on the document and click [Reimport](#).
- To re-import all documents, right-click on a MongoDB datastore or on the Documents node and click [Reimport All](#).

For more information, see “Re-importing schemas” in *Supplement for Adapters*.

Use Google BigQuery table data as a source

The Google BigQuery application datastore contains access information and passwords to open your Google BigQuery account on your behalf. Now you can use the application datastore to both extract data from and load data to your Google tables. Previously, you could only load generated Data Services data to your Google BigQuery tables for running queries.

See the *Supplement for Google BigQuery* for details.

Enhanced append features for file location objects

File location objects contain details about a specific file transfer protocol so that the software safely transfers data between remote and local servers. Now there are append options for the data files in your remote server location.

For information about file location objects, and file transfer protocols, see the *Designer Guide*: “File location objects”.

Template table support for Hive

The Hive adapter datastore now supports the creation of template tables in a data flow. You can use template tables as a source in a data flow or you can import them as regular Hive tables.

The *Drop and re-create table before loading* option was added to the *Adapter Target* tab to support template table creation. This option allows you to drop an existing table and create a new one with the same name before loading it.

For more information, see “Hadoop Hive adapter target options” in *Supplement for Adapters*.

SASL-QOP support with Kerberos on Hive

To enable SASL-QOP support, set the Hive adapter *Authentication* option to *Kerberos* and enter one of the following values into the *Additional Properties* field (the value you enter should match the value on the Hive server):

- Use `;sas1.qop=auth;` for authentication only.
- Use `;sas1.qop=auth-int;` for authentication with integrity protection.
- Use `;sas1.qop=auth-conf;` for authentication with integrity and confidentiality protection.

For more information, see “Hive adapter datastore configuration options” in *Supplement for Adapters*.

Hadoop support for Windows

Data Services now supports Hadoop on the Windows platform (Hortonworks HDP 2.2.6 only).

The following Hadoop features have been implemented for the Windows platform:

- Using Hive tables as a source or target in your data flow.
- Using HDFS files as a source or target in your data flows (using PIG script or the HDFS library API).
- Ability to stage non-Hive data in a data flow (using the Data_Transfer transform) before joining it with a Hive source. You would then pushdown the `Join` operation to Hive.
- View data for HDFS files and Hive tables

Items of importance:

- The Data Services Job Server must be installed in one of the nodes of the Hadoop cluster.
- The Job Server runs as a service, so all system environment variables like `PATH`, and `CLASSPATH` must be set up correctly.
- HDFS/Hive users must adhere to all HDFS file system permission requirements.

For more information, see “Hadoop support for the Windows platform” and “Setting up HDFS and Hive on Windows” in the *Reference Guide*.

Varchar and Char data type support for Hive tables

Data services now supports Varchar and Char data types for Hive tables.

You can now do the following:

- Create and use Hive tables that contain Varchar and Char columns.
- Use Varchar and Char columns in Hive queries, `Join` pushdowns, SQL functions, and SQL transforms.
- Create and use Hive template tables that contain Varchar and Char columns.

Data Services converts Char into Varchar. For example, if you import a Hive table that has a column with `char(<n>)`, Data Services converts that column to `varchar(<n>)`.

Also note:

- Data Services truncates values if the string value exceeds the length specifier.
- If the length (number of characters) is less than the Char column size, Data Services replaces missing characters with a space. This applies to null and blank value mappings as well.
- For the Varchar data type, a blank mapping is seen as null.
- Integers are automatically converted to Varchar or Char.

The Varchar data type was introduced in Hive 0.12.0 and the Char data type was introduced in Hive 0.13.0.

For information about the supported data types in Hive, go to <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Types> .

RAW and STRING data type support

This version of Data Services includes support for RAW and STRING data types in SAP systems that use / SAPDS. SAP tables with these data types can be used as a source in both regular data flows and ABAP dataflows.

Data Services will import these data types as the following way: ()

SAP data type	Data Services data type
Raw	Varchar
String	Long
String	Varchar
(When DSConfig.txt flag Import_SAP_String_As_Char = True)	

For this feature to work properly, SAP Note [2166986](#) must be applied to the source SAP system. If the source SAP system is on a higher level than the validity listed in the SAP Note, the functionality is already included in the installed SP and no further action is necessary. If the source SAP system is on a lower level than the low boundary of the validity, this feature cannot be used with that system.

10.3 Functions and transforms

SAP Data Services 4.2 SP6 includes enhancements to transforms and function support.

Push-down functions, transform, and operators

Now you can find out which functions, transforms, and operators you can use as push-down functions.

To access the list, see SAP Note [2212730](#).

10.4 Usability

SAP Data Services 4.2 SP6 includes the following enhancements that improve the user's experience.

Viewing objects before import

When importing an ATL or XML file, you can select the objects you want to import before beginning the import process.

1. Click **Tools** > **Import from file**, or in the object library, right-click and choose **Repository** > **Import from File**.
2. Choose an ATL or XML file and click **Open**. A list of objects displays in the **Import Plan** window.
3. Highlight objects and use the buttons at the bottom of the window to include or exclude them during import. You can also right-click on an object to access these options.

For more information about the available options, see “Importing from a file” in the *Administrator Guide*.

Documentation set updates

- Hive, HTTP, and JMS adapter information was moved from the *Management Console Guide*, *Integrator Guide*, and *Reference Guide* into the *Supplement for Adapters* guide.
- A new topic called “SSL connection support” was added to the *Supplement for Adapters* guide.
- General information about installing and configuring adapters was moved from the *Management Console Guide* into the *Supplement for Adapters* guide.

View related objects in the object library and the Variables and Parameters window

Now there are even more places in Data Services where you can see parent and child relationships between objects in the object library by looking at the **Usage** column. The **Usage** column now appears in the following object library tabs:

- Jobs
- Work flows
- Data flows
- Transforms (Data Quality, Text Data Processing, and aBAP transforms in the Platform tab)
- Formats

Additionally, you can see where you have used variables and parameters in the enhanced Variables and Parameters window.

For more information about the Usage column and using View Where Used tools, see the following sections in the *Designer Guide*:

- Variables and Parameters, Accessing View Where Used from Variables and Parameters window
- Design and Debug, Using View Where Used

Enhanced filters for substitution parameter search

Open the [Substitution Parameter Editor](#) and use the new filter tool to shorten the list of parameters to search. Enter the name or partial name of the parameter in the [Filter](#) text box and the list of parameters automatically displays a subset of the substitution parameters that match your filter.

For more information about substitution parameters and using the filter tool for easier search, see the Substitution parameters topics in the *Designer Guide*.

10.5 Data Quality transforms

In SAP Data Services 4.2 SP6, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

10.5.1 Data Cleanse transform

In SAP Data Services 4.2 SP6, the Data Cleanse transform has been enhanced with the following features.

Generate data quality statistics

The data quality statistics tables in the repository contain summary and record-level statistics about the cleansing and assignment processes performed on your data. You use the supplemental content information file (TASK_LOCALIZATION) to translate values that appear in the statistics tables into meaningful information.

You can enable data quality statistics in the Data Cleanse, Geocoder, and Global Address Cleanse transforms. The software sends the statistics to the applicable repository table.

There are two main reasons for generating and using the data quality statistics information:

- To view aggregated results and summary information about the quality of your data.
- To drill into records to gain information about how the software cleansed specific data elements.

For information about data quality statistics, see the *Reference Guide*.

Firm_Line input field

The software added four additional lines to the firm line input field for a total of six lines: Firm_Line1-6.

10.5.2 Geocoder transform

In SAP Data Services 4.2 SP6, the Geocoder transform has been enhanced with the following features.

Generate data quality statistics

The data quality statistics tables in the repository contain summary and record-level statistics about the cleansing and assignment processes performed on your data. You use the supplemental content information file (TASK_LOCALIZATION) to translate values that appear in the statistics tables into meaningful information.

You can enable data quality statistics in the Data Cleanse, Geocode, and Global Address Cleanse transforms. The software sends the statistics to the applicable repository table.

There are two main reasons for generating and using the data quality statistics information:

- To view aggregated results and summary information about the quality of your data.
- To drill into records to gain information about how the software cleansed specific data elements.

For information about data quality statistics, see the *Reference Guide*.

Brazil support

Support for Brazil is added in this release. In addition, many Brazil addresses include Primary_Name2–4 and Primary_Type2–4.

10.5.3 Global Address Cleanse transform

In SAP Data Services 4.2 SP6, the Global Address Cleanse transform has been enhanced with the following features.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2015 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2015 certification.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2015 certification.

La Poste ratification

The France processing options meet all requirements for La Poste 2015 certification.

Enhanced France data directories

In the October International directory update, we added two data sets to the France (FR) directory in support of La Poste ratification:

- Building names
- More precise primary range data to provide better assignments and more accurate results. This may cause some records that used to get primary range assignments to drop to primary name assignments.

Generate data quality statistics

The data quality statistics tables in the repository contain summary and record-level statistics about the cleansing and assignment processes performed on your data. You use the supplemental content information file (TASK_LOCALIZATION) to translate values that appear in the statistics tables into meaningful information.

You can enable data quality statistics in the Data Cleanse, Geocode, and Global Address Cleanse transforms. The software sends the statistics to the applicable repository table.

There are two main reasons for generating and using the data quality statistics information:

- To view aggregated results and summary information about the quality of your data.
- To drill into records to gain information about how the software cleansed specific data elements.

For information about data quality statistics, see the *Reference Guide*.

South Korea five-digit postcode

For this release, the South Korea reference data is updated to the new five-digit postcode standard.

Brazil quick data entry

In this release, you can do a quick postcode-only lookup for Brazil.

Macau address cleansing

The Global Address Cleanse transform has been enhanced to cleanse and validate Macau addresses.

10.6 Platform transforms

In SAP Data Services 4.2 SP6, the set of transforms that you use for general data movement operations has been enhanced with the following features.

Support for identity column

A new option in the target table editor named *Identity Insert* enables users to control how the software populates identity column content for Microsoft SQL Server database target tables.

On: Identity column values are generated by the software. You must include the identity column in the output map.

Off: Identity column values are generated by SQL Server.

For more information see the topic named “Microsoft SQL Server target tables options” in the *Reference Guide: Objects, Description of objects, Target tables*.

11 SAP Data Services 4.2 SP 5 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 5.

The following links take you to descriptions of the features by category for version 4.2 SP 5.



[Installation, administration, and monitoring \[page 102\]:](#)

- Improved visibility for real-time environment status
- Ability to trace failed requests from real-time jobs
- Ability to create a debug package
- Continuous work flow log file management



[Connectivity: Big data and CDC \[page 103\]:](#)

- SLT SP8 integration
- File location objects
- MongoDB support
- Hive adapter datastore support for SQL function and transform
- Pushing the JOIN operation to Hive
- SSL connection for Hive



[Functions and transforms \[page 105\]:](#)

- New file management functions
- New file location object functions
- Additional supported transforms in continuous work flows



[Usability \[page 107\]:](#) Switch repositories in designer



[Data Quality transforms \[page 107\]:](#)

- Geocoder transform: New Status_Code output field
- Global Address Cleanse transform
 - New Processing Timeout option
 - Inter-Script Matching support fields
 - Updated AMAS and SERP report information



[Platform transforms \[page 111\]:](#)

- Data Mask transform: Pattern variance



11.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP5 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Improved visibility for real-time environment status

New options were added to the Status Interval page in the Administrator for real-time jobs. You can now specify the time period for which the Administrator displays job information on the Real-Time History page.

You can filter the information for real-time jobs in the following ways:

- By the last execution of each job
- By number of days

i Note

The default is 5 days.

- By range of dates

For more information, see “Setting the status interval” in the *Management Console Guide*.

Ability to trace failed requests from real-time jobs

You can now trace failed requests from real-time jobs in the Access Server trace log. This enhancement allows you to troubleshoot more efficiently.

Previously, you were only able to trace all requests, making it difficult and time consuming to pick out the failed requests.

You can enable this feature in one of the following ways:

- In Data Services Server Manager, enter 256 (the value for tracing failed requests) into the *Parameter* field in the *Access Server Properties* window.
- In Server Manager on UNIX, use the command prompt to set the trace parameter (-T) to 256 (enable failed requests). For example, you might enter `-A4000 -MWIN2K8R2 -R"C:\temp\as1" -NoSSL -T256`.

- In the Management Console, navigate to the [Access Server Log Configuration](#) tab and select the [Failed_log](#) checkbox.

For more information, see the *Administrator Guide* and *Management Console Guide*.

Ability to create a debug package

You can create a job debug package that includes the ATL and trace, monitor, and log files. Send the generated .zip or .tar file to SAP customer support to help you troubleshoot execution-related issues. You can generate the debug package as you execute the job either from the Execution Properties window of the Designer or in the Execution Options page of the Management Console Administrator. The software generates a zip or tar file in the following location: `<DS_COMMON_DIR>\log\<Job_Server_Name>\<repository name>` on the machine where the Job Server is installed.

For more information, see “Creating a debug package” in the *Designer Guide*.

Continuous work flow log file management

The trace log file for continuous work flows has been enhanced to limit the size of the file and switch to a second archived file when the first file reaches the size limit. After the first log file reaches the size limit, the archive file is truncated and the content of the first file content is moved to the archive file. The default size of the trace log file is 8MB.

The SAP Data Services Designer and Management Console display the active trace file content, not the archived trace file.

11.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP5 includes the following connectivity enhancements.

SLT SP8 integration

Data Services now supports SAP Landscape Transformation Replication Server (SLT) running on the NetWeaver Platform. SLT is the ideal solution for all SAP HANA customers who need real-time (and non-realtime) data replication sourcing from SAP ERP or non-SAP systems into SAP HANA.

File location objects

File locations are first-class reusable objects that define specific file transfer protocols (FTP, SFTP, for example) between a remote server and a local server for which you (as the user) have access permission. The software stores file location objects under [File Locations](#) in the [Format](#) tab of the Data Services Designer object library.

You must associate a file location object with a specific file format to use it in a data flow as a source or target.

Applicable file formats include flat files, nested schema files (DTD, JSON, XML files) COBOL copybooks and Excel workbooks. You can also use the new built-in functions in a script to provide the information in a file location object:

- `copy_to_remote_server`
- `copy_from_remote_server`

Find out more about file location objects in "Description of objects" in the *Reference Guide*. For more information about the built-in functions, see [Functions and transforms \[page 105\]](#). Also see "Descriptions of built-in functions" in the *Reference Guide*.

MongoDB support

The MongoDB adapter allows you to read data from MongoDB to other Data Services targets.

This release includes the following MongoDB enhancements:

- Lightweight Directory Access Protocol (LDAP) authentication
- Kerberos authentication
- Support for Sharded Cluster connection
- Ability to connect to MongoDB using SSL with or without a PEM file

For more information, see "MongoDB adapter datastore configuration options" in *Supplement for Adapters*.

Hive adapter datastore support for SQL function and transform

The Hive adapter datastore can now process data using the SQL function and the SQL transform.

After connecting to a Hive datastore, you can do the following in Data Services:

- Use the SQL Transform to read data through a Hive adapter datastore. Keep in mind that the SQL transform supports a single SELECT statement only.

i Note

Select table column plus constant expression is not supported.

- Use the `sql()` function to:
 - create, drop, or INSERT Hive tables
 - return a single string value from a Hive table

- select a Hive table that contains aggression functions (max, min, count, avg, and sum)
- perform inner and outer joins

For more information, see “Using Hive metadata” in the *Supplement for Adapters*.

Pushing the JOIN operation to Hive

You can now stage non-Hive data in a dataflow (using the `Data_Transfer` transform) before joining it with a Hive source. You would then pushdown the `Join` operation to Hive.

Staging data is more efficient because Data Services no longer has to read all the data from the Hive data source into memory before performing the join.

Before staging can occur, you must first enable the [Enable automatic data transfer](#) option for the Hive datastore. This option can be found in the [Create New Datastore](#) or [Edit Datastore](#) window.

For more information, see “Using Hive metadata” in the *Supplement for Adapters*.

SSL connection for Hive

You can now use a Secure Socket Layer (SSL) connection when connecting to a Hive server.

You can configure SSL connection settings when adding a Hive adapter datastore.

For more information, see “Hive adapter datastore configuration options” in the *Supplement for Adapters*.

11.3 Functions and transforms

SAP Data Services 4.2 SP5 includes enhancements to transforms and function support.

New file management functions

There are three new built-in functions for file management:

File management built-in functions

Function name	Category	Description
file_copy	Miscellaneous	Copies an existing file to a different location using the same file name or a different file name. Copies a group of files indicated by a wildcard (*) to a different existing directory. Overwrites any existing target file when overwrite flag is set to 1. Source file still exists in the original location after <code>file_copy</code> .
file_delete	Miscellaneous	Deletes an existing file, or deletes a group of files indicated by a wildcard.
file_move	Miscellaneous	Moves an existing file to a different location using the same file name or a different file name. Moves a group of files indicated by a wildcard (*) to a different existing directory. Overwrites any existing target file when overwrite flag is set to 1. Source file does not exist in the original location after <code>file_move</code> .

New file location object functions

There are two new built-in functions that may be used in scripts to specify a specific file location object, which specifies file transfer protocols for certain files.

File location object built-in functions

Function name	Category	Description
copy_from_remote_server	Miscellaneous	Copy a file from a remote server to a local server before reading and processing the file in a work flow. Also copies a group of files indicated by a wildcard in the file name (either * or ? character).
copy_to_remote_server	Miscellaneous	Copy a file from a local server to a remote server after the software has processed the file in a work flow. Also copies a group of files indicated by a wildcard in the file name (either * or ? character).

Read about built-in functions in the *Reference Guide* under “Functions and procedures”. For information about the file location objects, see “File location objects” in the “Description of objects” section of the *Reference Guide*.

Additional supported transforms in continuous work flows

The following transforms can now be used in a continuous work flow:

- Data Quality transforms:
 - Country ID
 - Data Cleanse
 - DSF2 Walk Sequence
 - Geocoder
 - Global Address Cleanse
 - Global Suggestion List
 - USA Regulatory
 - User-Defined
- Table_Comparison transform
- Text Data Processing transform: Entity Extraction

11.4 Usability

SAP Data Services 4.2 SP5 includes the following enhancements that improve the user's experience.

Switch repositories in Designer

In SAP Data Services Designer, you can now switch between repositories, with no need to log out of and back into Designer. To switch repositories, right-click in the local object library and select [Switch Object Library](#).

11.5 Data Quality transforms

In SAP Data Services 4.2 SP5, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

11.5.1 Geocoder transform

In SAP Data Services 4.2 SP5, the Geocoder transform has been enhanced with the following feature.

New Status_Code output field

The Status_Code output field has been added to the Geocoder transform that displays a four-character status code that always starts with an S. This code represents the fuzzy matches made to the address/point of interest (POI) component during address geocoding processing.

- The first character is always S (for Status).
- The second character is associated with any last-line fuzzy matches.
- The third character is associated with any address-line fuzzy matches.
- The fourth character is associated with any POI data fuzzy matches.

Note

The Status_Code output field is only used for the address geocoding process.

First character

The value of the first character is always S for Status.

Second character

The value of the second character depends on fuzzy matches to the postal code, region, or locality.

Value	Description
0	No significant difference between the input data and the reference data.
1	Fuzzy match on postal code.
2	Fuzzy match on region.
3	Fuzzy match on postal code and region.
4	Fuzzy match on locality.
5	Fuzzy match on postal code and locality.
6	Fuzzy match on region and locality.
7	Fuzzy match on postal code, region and locality.

Third character

The value of the third character depends on fuzzy matches to the primary name prefix, primary name postfix, primary type, primary name, and primary range.

Value	Description
0	No significant difference between the input data and the reference data.
1	Fuzzy match on primary name prefix/primary name postfix.
2	Fuzzy match on primary type.
3	Fuzzy match on primary name prefix/primary name postfix and primary type.
4	Fuzzy match on primary name.
5	Fuzzy match on primary name prefix/primary name postfix and primary name.
6	Fuzzy match on primary type and primary name.
7	Fuzzy match on primary name prefix/primary name postfix, primary type, and primary name.
8	Fuzzy match on primary range.
9	Fuzzy match on primary name prefix/primary name postfix and primary range.
A	Fuzzy match on primary type and primary range.
B	Fuzzy match on primary name prefix/primary name postfix, primary type, and primary range.
C	Fuzzy match on primary name and primary range.
D	Fuzzy match on primary name prefix/primary name postfix, primary name, and primary range.
E	Fuzzy match on primary type, primary name, and primary range.
F	Fuzzy match on primary name prefix/primary name postfix, primary type, primary name, and primary range.

Fourth character

The value of the fourth character depends on fuzzy matches to POI name.

Value	Description
0	No significant difference between the input data and the reference data.
1	Fuzzy match on POI name.

11.5.2 Global Address Cleanse transform

In SAP Data Services 4.2 SP5, the Global Address Cleanse transform has been enhanced with the following features.

New Processing Timeout option

The Global Address engine has a new option that allows you to specify a timeout for processing.

Option	Description
Processing Timeout	<p>Specifies the timeout duration in milliseconds. After the duration is exceeded, processing stops and information code 7000 is issued. You can use a substitution variable for this option.</p> <p><i>0 or less</i>: Disables the timeout. This is the default value.</p> <p><i>> 0</i>: Indicates the amount of time after which processing aborts.</p> <div> <p>Note</p> <p>It is recommended that the timeout is set at or above 100 milliseconds.</p> </div>

The following information code is issued when the processing timeout is exceeded:

Information code	Description	Engine(s)
7000	The user-defined processing timeout threshold (set in the Processing Timeout option) has been reached, preventing further corrective action on this address.	Global

Inter-Script Matching support fields

The Global Address cleanse transform includes output fields that can be used in the Match transform during the comparison process. For China, Russia, South Korea, and Taiwan, non-Latin scripts are now transliterated to Latin to support inter-script matching for the following Match_* output fields:

- Match_Block_Number
- Match_Building_Name
- Match_Country
- Match_Floor_Number
- Match_Locality
- Match_Locality2
- Match_Postcode1
- Match_Primary_Directional
- Match_Primary_Name
- Match_Primary_Name2
- Match_Primary_Number
- Match_Primary_Type
- Match_Region
- Match_Stairwell_Name
- Match_Unit_Number
- Match_Wing_Name

Updated AMAS and SERP report information

The values of fields that are displayed on the AMAS Report (Address Matching Processing Summary Report) for Australia have been updated as follows:

Field name	Previous value	Current value
Company Name	SAP Business Objects	SAP
Software Name and Version	Data Services - 4.2.4.0	Data Services - 4.2.5.0

The values of fields that are displayed on the SERP (Statement of Address Accuracy) report for Canada have been updated as follows:

Field name	Previous value	Current value
Software Vendor Name	SAP Business Objects	SAP
Product Name and Version	ACE CANADA 7.91c	SAP Data Services 4.2
Software Recognition Expiry Date	31-Dec-2014	31-Dec-2015

11.6 Platform transforms

In SAP Data Services 4.2 SP5, the set of transforms that you use for general data movement operations has been enhanced with the following features.

11.6.1 Data Mask transform

This release of SAP Data Services includes the following enhancement to the [Data Mask](#) transform.

Pattern variance

Use the Pattern Variance Group to mask an input field substring with a specific pattern variance type: preserve, character, string, or default.

Pattern variance types

Type	Description
Preserve	Outputs the defined substring the same as it is input.

Type	Description
Character	Masks the defined substring by randomly replacing each of the characters in the defined substring with values that you specify in the <i>Value</i> field. Retains any special input characters (such as *,-,:) and spaces in the output field.
String	Masks the defined substring by randomly replacing the entire substring with values that you specify in the <i>Value</i> field. Does not retain special input characters (such as *,-,:) in the output field.
Default	Masks each applicable character with like characters for alpha and numeric content. Retains any special input characters (such as *,-,:) in the output field.

For details about each pattern variance type, and for examples, see “Pattern Variance Group” in the *Reference Guide*.

11.7 Reserved words

New reserved words for this release.

When you name work flows, data flows, transforms, or other design elements that you create, there is a list of reserved words that you should not use. Further, you cannot use reserved words as user names when you create a Data Services repository.

Two new words that are related to the new file location objects feature (see [Connectivity: Big data and CDC \[page 103\]](#)) are included in the reserved word list for this release:

- Job Server
- Local

Reserved words are reserved with any combination of upper- and lower-case letters. If you use reserved words you must put double quotation marks around them.

For more information about reserved words, and to see a complete list, see the *Reference Guide*.

12 SAP Data Services 4.2 SP 4 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 4.

The following links take you to descriptions of the features by category for version 4.2 SP 4.



[Installation, administration, and monitoring \[page 113\]](#): Retrieving the time zone of a Management Console machine



[Connectivity: Big data and CDC \[page 114\]](#):

- Google BigQuery application datastore
- MongoDB support
- HDFS data preview
- Additional database support for Replication Server real-time CDC
- SAP HANA synonym support
- Merge SQL support for SAP HANA
- Pushdown decode() to Sybase IQ



[Data Quality transforms \[page 115\]](#):

- Data Cleanse transform: Data Cleanse support for Turkey
- Global Address Cleanse transform
 - Certified Australia options
 - Certified Canada options
 - Certified New Zealand options
 - Lastline drilldown
 - Transliteration support for South Korea and Taiwan



[Platform transforms \[page 117\]](#):

- Data Mask transform
 - Data Variance feature for Data Masking

12.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP4 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Retrieving the time zone of a Management Console machine

The new `Get_MC_Machine_Timezone` operation allows you to retrieve the time zone of the Management Console machine.

i Note

`Get_MC_Machine_Timezone` returns only time zone information for the Management Console machine. It does not return Central Management Server (CMS) or job server time zone information. Be aware that the Management Console, the CMS, and the job server you use may be deployed across different time zones.

For more information, see the *Integrator Guide*.

12.2 Connectivity: Big data and CDC

SAP Data Services 4.2 SP4 includes the following connectivity enhancements.

Google BigQuery application datastore

The Google BigQuery application datastore allows SAP Data Services to access your Google projects on your behalf. Use it to load data from Data Services into your Google project tables for Google BigQuery analysis.

Data Services has the capacity to access data from many sources such as Microsoft SQL Server, Oracle, and Peoplesoft. Before you load data to your Google account, use Data Services to transform, enrich, and join data so that you run effective and efficient queries in Google. This data can consist of relational tables that have a nested schema.

MongoDB support

The MongoDB adapter allows you to read data from MongoDB to other Data Services targets.

MongoDB is an open-source document database which uses JSON-like documents (MongoDB calls the format BSON) with dynamic schemas instead of a traditional table-based relational database structures.

MongoDB is schema-free, but Data Services needs metadata for task design and execution. Data Services generates schema data based on a certain number of records and allows you to provide a JSON file that the software can use to generate a schema for each collection.

After you create an adapter instance and a datastore, you can browse and import MongoDB entities, which are similar to database tables.

For more information, see *Supplement for Adapters*.

HDFS data preview

You can now preview HDFS file data in Data Services for delimited and fixed width file types. For more information about this enhancement, see “Previewing HDFS file data” in the *Reference Guide*.

Additional database support for Replication Server real-time CDC

In Data Services 4.2.3, we introduced a simplified way to capture changed data with Replication Server. This CDC method uses a continuous work flow, built-in functions, and Replication Server to capture changed data from Oracle databases.

This releases includes additional database support for SAP ASE, DB2, and Microsoft SQL Server.

SAP HANA synonym support

SAP Data Services now supports synonyms in SAP HANA.

Merge SQL support for SAP HANA

SAP Data Services now supports Merge SQL statements in SAP HANA.

Pushdown decode() to Sybase IQ

SAP Data Services now supports pushdown for the decode() function as SQL to Sybase IQ.

12.3 Data Quality transforms

In SAP Data Services 4.2 SP4, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

12.3.1 Data Cleanse transform

SAP Data Services 4.2 SP4 includes the following enhancements to the Data Cleanse transform.

Data Cleanse support for Turkey

The global person and firm cleansing package that you use in the Data Cleanse transform now supports parsing Turkey person, firm, and date data.

12.3.2 Global Address Cleanse transform

In SAP Data Services 4.2 SP4, the Global Address Cleanse transform has been enhanced with the following features.

Certified Australia options

The Australia processing options meet all requirements for AMAS 2014 certification.

Certified Canada options

The Canada processing options meet all requirements for SERP 2014 certification.

Certified New Zealand options

The New Zealand processing options meet all requirements for SendRight 2014 certification.

Lastline drilldown

The new Enable Lastline Drilldown option extends the suggestion list capabilities of the Global Engine engine of the Global Address transform. You can extract another level of information with the content depending on the selection of its parent component. In other words, given an input country, you are provided a list of available regions; then after specifying the region, you are provided a list of available locality1s; then after specifying the locality1, you are provided a list of available locality2s; and so on. For example, for China, you can select the region, then the cities within the selected region, and then the districts within the selected city.

A suggestion list using the Enable Lastline Drilldown option is the same as a typical lastline suggestion list, except that there are no postcode or address type components in the suggestion entry. After you are provided a complete lastline after drilldown, you can add additional address information such street and house number, and send the entire address to the Global Address Cleanse transform for validation and correction.

This option is only supported for real-time jobs for the countries that have the data available, and the country is required on input. For a list of supported countries, see the topic “Countries supported by lastline drilldown” in

the *SAP Data Services Reference Guide*. The Enable Lastline Drilldown option is not available for Israel, New Zealand, Serbia, or Singapore because their localities are not linked with region in the data directory. For some countries, such as India, some localities may not be returned in the suggestion list because localities are not always linked with region in the data directory. For countries and regions with a large number of localities, the suggestion list may perform slower. The option is not available for the Canada and USA engines.

The following values have also been added to the Status output field of the Global Address transform:

Output field name (Global Address Cleanse)	Description
Status	<p>Specifies the suggestion status generated as the result of looking up the current record and performing suggestion processing.</p> <p><i>L1</i>: Locality1 list available.</p> <p><i>L2</i>: Locality2 list available.</p> <p><i>L3</i>: Locality3 list available.</p> <p><i>L4</i>: Locality4 list available.</p> <p><i>R1</i>: Region1 list available.</p> <p><i>R2</i>: Region2 list available.</p>

Transliteration support for South Korea and Taiwan

With the transliteration support added in this release, the Global Address Cleanse transform now helps you cleanse and validate your South Korean and Taiwanese address data more effectively, either in the native or Latin script.

12.4 Platform transforms

In SAP Data Services 4.2 SP4, the set of transforms that you use for general data movement operations has been enhanced with the following features.

12.4.1 Data Mask transform

SAP Data Services 4.2 SP4 includes the following enhancement to the *Data Mask* transform.

Date Variance feature for Data Masking

New [Date Variance Group](#) enhances the [Data Mask](#) feature so that the transform now randomizes date values on output.

Upgrade to Data Services version 4.2 SP4 to use the new [Date Variance Group](#) in the Data Mask transform. The [Date Variance Group](#) enhances data masking so that the transform randomizes date values on output. Use the new [Date Variance](#) options to alter input fields (Date, DateTime, or Character) based on a date variance type (set number of days, months, years), or by date range. Retain referential integrity using a seed value to keep the altered date values the same when you run a job multiple times.

For more information about the Date Variance Group, see “Date Variance Group” in the *Reference Guide*.

13 SAP Data Services 4.2 SP 3 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 3.

The following links take you to descriptions of the features by category for version 4.2 SP 3.



[Connectivity: Big data and CDC \[page 119\]:](#)

- Shapefile support
- Previewing Hive table data
- Replication Server real-time CDC



[Functions and transforms \[page 120\]:](#) New built-in functions added



[Usability \[page 121\]:](#)

- New *Supplement for Adapters* guide
- Bypassing specific work flows and data flows in a job
- Designer search capabilities



[Data Quality transforms \[page 122\]:](#)

- Data Cleanse transform
 - International phone parsing enhancements
 - New information codes
 - New output field Match_Person
- Global Address Cleanse transform
 - Input and output fields for use with SAP software
 - Use Firm To Assign option
 - Locality 4 input field
 - Suggestion Lists Reply
- USA Regulatory Address Cleanse transform
 - NCOALink Utility



[Platform transforms \[page 117\]:](#) New Data Mask transform

13.1 Connectivity: Big data and CDC

SAP Data Services 4.2 SP3 includes the following connectivity enhancements.

Shapefile support

You can now create a Shapefile adapter instance in the Management Console of the Administrator.

The Shapefile adapter option allows Data Services to read geospatial vector data from Shapefiles and load the data into the HANA database for further analysis.

For more information, see *Supplement for Adapters*.

Previewing Hive table data

You can now preview Hive table data by right-clicking on a Hive table name in the Local Object Library and clicking [View Data](#). Alternatively, you can click the magnifying glass icon on Hive source and target objects in a data flow or [View Data](#) tab of the Hive table view.

i Note

Hive table data preview is only available with Apache Hive version 0.11 and later.

For more information, see "Hadoop" in the *Reference Guide*.

Replication Server real-time CDC

This release includes a simplified way to capture changed data with Replication Server. You no longer need to use PowerDesigner modeling. This CDC method uses a continuous work flow, built-in functions, and Replication Server to capture changed data from Oracle databases (additional databases will be supported in future releases).

13.2 Functions and transforms

SAP Data Services 4.2 SP3 includes enhancements to transforms and function support.

New Built-in Functions added

You can now use the following five built-in function capabilities:

- `local_to_utc` - Converts the input datetime of any time zone to Coordinated Universal Time (UTC)
- `regex_replace` - Matches the whole input string to the pattern that is specified with regular expressions and flags and replaces the matching part of the input string with the replacement string provided
- `string_to_number` - Returns the integer sum of all characters from the input string

- `translate` - Returns the translated input string with all occurrences of each character in the original string replaced by their corresponding character in the new string
- `utc_to_local` - Converts the input in Coordinated Universal Time (UTC) to the desired timezone value

13.3 Usability

SAP Data Services 4.2 SP3 includes the following enhancements that improve the user's experience.

New Supplement for Adapters guide

We have consolidated existing adapter information into a single guide named *Supplement for Adapters*. This new guide replaces the following:

- *Supplement for OData*
- *Supplement for Salesforce.com*
- *Supplement for SuccessFactors*

Bypassing specific work flows and data flows in a job

You can now bypass the execution of a work flow or data flow during design time. The Bypass attribute can help speed up the testing process when designing jobs by allowing you to run certain work flows or data flows in the job instead of having to run them all.

For more information, see "Bypassing specific work flows and data flows" and "Changing properties of a data flow" in the *Designer Guide*.

Designer search capabilities

Using SAP Data Services Designer, you can search for a text string in every part of the object, such as table name and variable name. A new status line at the bottom of the [Search](#) window shows where the search was conducted (Local or Central), the total number of items found, and the amount of time it took to complete the search.

For more information, see "Searching for objects" in the *Designer Guide*.

13.4 Data Quality transforms

In SAP Data Services 4.2 SP3, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

13.4.1 Data Cleanse transform

Information about what's new for Data Cleanse in SAP Data Services 4.2 SP3.

International phone parsing enhancements

We have added enhancements to the Data Cleanse phone parsing process for more efficient and complete international phone parsing. Changes include a new group of options in the Data Cleanse transform *Options* tab named *Phone Options*.

The *Phone Options* group consists of five parameters, four of which we moved from ► *Options* ► *Standardization Options* ► *Other* ► and a new parameter named *ISO2 Country Code Sequence*. The new *Phone Options* group contains the following parameters:

- *ISO2 Country Code Sequence* (New)
- *North American Phone Parens Area*
- *North American Phone Delimiter After Area* (Moved from *Other* group)
- *North American Phone Delimiter* (Moved from *Other* group)
- *Phone Extension Text* (Moved from *Other* group)

You use the *ISO2 Country Code Sequence* parameter to create a series of country codes (separated with "|" pipes) in the order in which you want Data Cleanse to parse phone information.

Click the ellipses button at the end of the parameter to open the *Order Options* dialog box. The listed country codes are those from the cleansing package. Select the applicable two-character codes from the left side list and add them to the right side list. You can list a code only once in your sequence. The order in which you place the codes in the sequence determines the order in which Data Cleanse searches for phone information.

For more information about international phone parsing behavior, see the topic "How Data Cleanse parses phone numbers" in the *Designer Guide*. For a complete description of the new *ISO2 Country Code Sequence* parameter, see the *Reference Guide*.

New information codes

There are three new information codes related to phone parsing as described in the table below.

New phone information codes

Information code format	Description
T#02	The transform parsed phone data using a different country than the country listed in the Option_Country field.
T#03	The transform parsed phone data by prepending a country code to the incoming phone data.
R428	The cleansing package does not recognize the country code provided in the Option_Country input field.

For details about information codes see the topic "Information codes (Data Cleanse)" in the *Reference Guide*.

New output field Match_Person

A new person output field, Match_Person, is available for standardized output for the Data Cleanse transform in Data Services 4.2.3.

Generated field name	Content type	Description
Match_Person	Person	A form of Person that you can use in the Match transform during the comparison process. Match outputs the data in uppercase, removes apostrophes and replaces other punctuation with a single space. It also removes data that is extraneous for matching purposes. Extraneous data includes pre-given name, pre-family name, and pre-name as well as honorary and maturity post names and name designators.

13.4.2 Global Address Cleanse transform

In SAP Data Services 4.2 SP3, the Global Address Cleanse transform has been enhanced with the following features.

Input and output fields for use with SAP software

Several enhancements have been made to the Global Address Cleanse transform to allow it to work seamlessly with other SAP software, such as SAP Customer Relationship Management.

Updated input field names

The following fields that were added in a previous release have been renamed for consistency.

Previous input field name	New input field name
NetWeaver_Building	NW_Building
NetWeaver_Floor	NW_Floor_Num
NetWeaver_House_Num1	NW_House_Num1
NetWeaver_House_Num2	NW_House_Num2
NetWeaver_Location	NW_Location
NetWeaver_PO_Box	NW_PO_Box
NetWeaver_RoomNumber	NW_Room_Num
NetWeaver_Str_Suppl1	NW_Str_Suppl1
NetWeaver_Str_Suppl2	NW_Str_Suppl2
NetWeaver_Str_Suppl3	NW_Str_Suppl3
NetWeaver_Street	NW_Street

New input fields

The following fields have been added in this release that map to the data model of SAP software.

Caution

Use the NW input fields properly to avoid unexpected results in your data. You cannot map multiline or Address_Line input fields when you use the NW input fields. Although the NW input fields appear discrete, they behave and are processed as multiline fields. They are mapped internally to Multiline1-12 before normal Global Address Cleanse processing is performed. If a NW input field is not mapped, the multiline that would have been mapped to it is mapped to the next available NW input field. For more information about using the NW input fields, see “Mapping NW input fields” in the *Reference Guide*.

NW input field name (Global Address Cleanse)	Description	Engine
NW_City1	Contains the locality. When you map NW input fields, this is a required field. The NW_City1 and NW_City2 input fields must be mapped in sequence.	All engines
NW_City2	Contains additional locality or district information.	All engines
NW_Country	Contains the country. When you map NW input fields, this is a required field.	All engines
NW_Home_City	Contains additional locality information.	All engines
NW_PO_Box_City	Contains the locality. If any of the NW_PO_Box input fields are mapped, then all of them must be mapped.	All engines
NW_PO_Box_Country	Contains the country. If any of the NW_PO_Box input fields are mapped, then all of them must be mapped.	All engines
NW_PO_Box_Postcode	Contains the postcode. If any of the NW_PO_Box input fields are mapped, then all of them must be mapped.	All engines
NW_PO_Box_Region	Contains the state, province, or region. If any of the NW_PO_Box input fields are mapped, then all of them must be mapped.	All engines

NW input field name (Global Address Cleanse)	Description	Engine
NW_Postcode	Contains the postcode. When you map NW input fields, this is a required field.	All engines
NW_Region	Contains the state, province, or region. When you map NW input fields, this is a required field.	All engines

Updated output field names

The following fields that were added in a previous release have been renamed for consistency.

Previous output field name	New output field name
NetWeaver_Formatted_Postcode	NW_Formatted_Postcode
Postcode_In_NetWeaver_Supported_Format	NW_Postcode_In_Supported_Format

New output field names

The following fields have been added in this release that map to the data model of SAP software.

Output field name (Global Address Cleanse)	Description	Engine
Block_Full	A compound output field consisting of the Block_Description and Block_Number output fields.	G
Building_Name1_2	A compound output field consisting of the Building_Name1 and Building_Name2 output fields.	G
Building_Primary_Addr_Delivery_Dual	A compound output field consisting of the Building_Name1, Building_Name2, Primary_Address (delivery) and Primary_Address (dual) output fields.	All engines
Building_Primary_Secondary_Addr_Delivery_Dual	A compound output field consisting of the Building_Name1, Building_Name2, Primary_Secondary_Address (delivery), and Primary_Secondary_Address (dual) output fields.	All engines
Delivery_Installation_Full	A compound output field consisting of the Delivery_Installation_Name, Delivery_Installation_Qualifier, and Delivery_Installation_Type output fields.	C, G
Floor_Full	A compound output field consisting of the Floor_Description, Floor_Number, and Floor_Qualifier output fields.	All engines
Locality1_2_Full	A compound output field consisting of the Locality1_Full and Locality2_Full output fields.	All engines
Locality1_2_Name	A compound output field consisting of the Locality1_Name and Locality2_Name output fields.	All engines
Locality1_4_Full	A compound output field consisting of the Locality1_Full, Locality2_Full, Locality3_Full, and Locality4_Full output fields.	All engines
Locality1_4_Name	A compound output field consisting of the Locality1_Name, Locality2_Name, Locality3_Name, and Locality4_Name output fields.	All engines

Output field name (Global Address Cleanse)	Description	Engine
Locality2_4_Full	A compound output field consisting of the Locality2_Full, Locality3_Full, and Locality4_Full output fields.	All engines
Locality2_4_Name	A compound output field consisting of the Locality2_Name, Locality3_Name, and Locality4_Name output fields.	All engines
Locality3_4_Full	A compound output field consisting of the Locality3_Full and Locality4_Full output fields.	All engines
Locality3_4_Name	A compound output field consisting of the Locality3_Name and Locality4_Name output fields.	All engines
Point_Of_Reference1_2	A compound output field consisting of the Point_of_Reference1 and Point_of_Reference2 output fields.	All engines
Primary_Address_Delivery_Dual	A compound output field consisting of the Primary_Address1-4 (delivery) and Primary_Address1-4 (dual) output fields.	All engines
Primary_Name_Full1_2	A compound output field consisting of the Primary_Name_Full1 and Primary_Name_Full2 output fields.	All engines
Primary_Name_Full1_4	A compound output field consisting of the Primary_Name_Full1, Primary_Name_Full2, Primary_Name_Full3, and Primary_Name_Full4 output fields.	All engines
Primary_Name_Full3_4	A compound output field consisting of the Primary_Name_Full3 and Primary_Name_Full4 output fields.	All engines
Primary_Secondary_Addr_Delivery_Dual	A compound output field consisting of the Primary_Secondary_Address (delivery) and Primary_Secondary_Address (dual) output fields.	All engines
Region1_2_Full	A compound output field consisting of the Region1_Full and Region2_Full output fields. USA : Does not include Region2_Full.	All engines
Region1_2_Name	A compound output field consisting of the Region1_Name and Region2_Name output fields. USA : Does not include Region2_Name.	All engines
Remainder_Extra_PMB_Full	A compound output field consisting of the Remainder_Full, Extra1, Extra2, and PMB_Full output fields.	All engines
Room_Full	A compound output field consisting of the Unit_Description (if it contains "room" or a variant) and Unit_Number output fields.	All engines
Room_Number	The unit number for units that are variations of "room" (for example, RM, RMS, ROOM, ROOMS, RM., RMS, 号室, 室, 号).	All engines

Output field name (Global Address Cleanse)	Description	Engine
Secondary_Address_No_Floor_No_Room	A compound output field consisting of all Secondary_Full output fields except Floor_Full and Room_Full.	All engines
Secondary_Address_No_Floor	A compound output field consisting of all Secondary_Full output fields except Floor_Full.	All engines
Secondary_Address_No_Room	A compound output field consisting of all Secondary_Full output fields except Room_Full.	All engines
Stairwell_Full	A compound output field consisting of the Stairwell_Description and Stairwell_Name output fields.	All engines
Unit_Full	A compound output field consisting of the Unit Description, Unit_Number, and Unit_Qualifier output fields.	All engines
Wing_Full	A compound output field consisting of the Wing_Description and Wing_Name output fields.	All engines

New NW_PO_Box output fields

The NW_PO_Box output fields are populated only when fields are mapped to NW input fields and are used only for the PO Box address portion of SAP business suite software.

The following is a list of the available NW_PO_Box output fields. The content of each NW_PO_Box field is identical to its corresponding output field without the prefix.

- NW_PO_Box_Assignment_Info
- NW_PO_Box_Assignment_Level
- NW_PO_Box_Assignment_Type
- NW_PO_Box_Delivery_Point
- NW_PO_Box_Info_Code
- NW_PO_Box_ISO_Country_Code_2Char
- NW_PO_Box_ISO_Script_Code
- NW_PO_Box_Locality1_Full
- NW_PO_Box_Match_Block_Number
- NW_PO_Box_Match_Building_Name
- NW_PO_Box_Match_Country
- NW_PO_Box_Match_Floor_Number
- NW_PO_Box_Match_Locality
- NW_PO_Box_Match_Locality2
- NW_PO_Box_Match_Postcode1
- NW_PO_Box_Match_Primary_Directional
- NW_PO_Box_Match_Primary_Name
- NW_PO_Box_Match_Primary_Name2
- NW_PO_Box_Match_Primary_Number
- NW_PO_Box_Match_Primary_Type
- NW_PO_Box_Match_Region
- NW_PO_Box_Match_Stairwell_Name

- NW_PO_Box_Match_Unit_Number
- NW_PO_Box_Match_Wing_Name
- NW_PO_Box_NW_Formatted_Postcode
- NW_PO_Box_NW_Postcode_In_Supported_Format
- NW_PO_Box_Postcode_Full
- NW_PO_Box_Postcode_In_Valid_Format
- NW_PO_Box_Primary_Address
- NW_PO_Box_Primary_Number
- NW_PO_Box_Primary_Secondary_Address
- NW_PO_Box_Region1_Full
- NW_PO_Box_Region2_Full
- NW_PO_Box_Region2_Name
- NW_PO_Box_Status_Code

For information about NW input and output fields and options, see the *SAP Data Services Reference Guide*.

Use Firm To Assign option

This option is available for the Canada, Global Address, and USA engines.

Option	Description
Use Firm To Assign	Specifies whether the firm is used to make an assignment and is displayed in a suggestion list. <i>Yes:</i> Uses and displays the firm. This is the default. <i>No:</i> Does not use or display the firm.

Locality 4 input field

A locality field was added to support additional locality information in China and Japan.

Input field name (Global Address Cleanse)	Description	Engine
Locality4	Any additional city, town, or suburb information. <i>China:</i> Village and neighborhood level localities, such as administrative villages (行政村), neighborhood committees (社区居民委员会), neighborhoods or communities (社区), or village committees (村民委员会). <i>Japan:</i> Any additional district, village, sub-district (aza, bu, chiwari, sen, and tori), or super block (joh).	G

Suggestion Lists Reply

We have enhanced the Suggestion_Reply input field to make the suggestion lists more flexible. You can continue to use all six reply fields or now you can contain all replies in the Suggestion_Reply1 field, delimited with the pipe (|). You can also allow end users to tell you what addresses they want to accept and when they are done with the street address.

Input field name (Global Address Cleanse)	Description	Engine
Suggestion_Reply1-6	<p>Used to input the index number that corresponds to a specific last line suggestion, an address line suggestion, or secondary list suggestion. These fields can also be used to input a street primary range or a street secondary range.</p> <p><i>Suggestion_Reply1</i>: If you do not want to use a suggestion list, make the value of this field 0 and the suggestion list will be ignored.</p> <p>If you want to use one field to hold all of the replies (rather than using all six reply fields), you can use the Suggestion_Reply1 field and separate the replies with a pipe ().</p> <p>When using the Suggestion_Reply1-6 fields for SAP software for street and PO Box addresses, you can insert the following symbols to indicate whether the user has accepted changes made to the street address and when they are done with the street address:</p> <ul style="list-style-type: none">• <i>asterisk plus</i> (*+): The user accepts the changes made to the street address up to the specified point and is done with the street address.• <i>asterisk minus</i> (*-): The user does not accept the changes made to the street address up to the specified point and is done with the street address.	All engines


13.4.3 USA Regulatory Address Cleanse transform

In SAP Data Services 4.2 SP3, the USA Regulatory Address Cleanse transform has been enhanced with the following feature.

NCOALink Utility

A new command-line utility is installed with Date Services to replace the NCOALink DVD Verification utility graphical user interface. The utility extracts and uncompresses the compressed NCOALink directory files from the USPS.

- If you use a version of the software that automatically installs the NCOALink Utility, see the *SAP Data Services Reference Guide*.

- If you use a version of the software that does not install the NCOALink Utility, you can download the application from the SAP Support Portal at <https://support.sap.com/software/address-directories.html> .
- If you still have the graphical user interface installed from a previous version of Data Services, you can continue to use it, or you can use the new command-line utility.

13.5 Platform transforms

In SAP Data Services 4.2 SP4, the set of transforms that you use for general data movement operations has been enhanced with the following features.

13.6 Data Mask transform

The Data Mask transform enables you to protect personally identifiable information in your data.

The Data Mask transform is a new Platform transform that uses mask out and number variance techniques to anonymize, obfuscate, and mask personal identifiable information in your databases. Personal identifiable information includes data such as U.S. Social Security Numbers, bank account numbers, salary and income information, and so on.

For complete information about the Data Mask transform, see “Data Mask” in the *Reference Guide*.

14 SAP Data Services 4.2 SP 2 features

Learn about the new and changed features introduced in SAP Data Services 4.2 SP 2.

The following links take you to descriptions of the features by category for version 4.2 SP 2.



[Connectivity: Big data and CDC \[page 131\]:](#)

- JSON file format support
- SAP HANA partition support
- Hadoop requirements
- Generic JDBC adapter instance
- New Web Services function, Get_Scheduled_Task
- Columnar support for DB2 10.5 and Teradata 14.10
- Allow merge or upsert supported for Teradata
- OData support



[Usability \[page 133\]:](#)

- Expanded Designer search capabilities
- Operational Dashboard redesign
- WSDL Labels
- Failed Request Log



[Text Data Processing \[page 134\]:](#)

- Dynamic input field added
- Substitution parameters added



[Data Quality transforms \[page 135\]:](#)

- Data Cleanse transform
 - Family Name Full and Given Name Full output fields
- Geocoder transform
 - New country directory data
- Global Address Cleanse transform
 - New Global Address Cleanse solutions for South Korea
 - Simplified Suggestion List output fields
 - Support major city translation
 - Enhanced support for SAP software

14.1 Connectivity: Big data and CDC

SAP Data Services 4.2 SP2 includes the following connectivity enhancements.

JSON file format support

Similar to XML files, SAP Data Services now supports JSON files and messages as batch or real-time sources and targets. As a source, a JSON file translates incoming JSON-formatted data into data that the software can process. As a target, a JSON file translates the data produced by a data flow, including nested data, into a JSON-formatted file. You can access the JSON format in the new Nested Schemas category of the Designer's Format tab. You can also use a Nested Schemas template to create a single-use JSON file or an XML file that matches a particular input schema.

For more information, see the *Reference Guide* and *Designer Guide*.

SAP HANA partition support

SAP HANA supports partition support for column store tables. SAP Data Services supports the SAP HANA partition feature for parallel reading and loading using physical partitions and logical partitions. You can import an SAP HANA partition table metadata for range partitioned tables used for parallel reading and loading. For a logical partition of SAP HANA tables, range mixed with list (similar to physical range partition syntax) is supported.

HADOOP requirements

Data Services 4.2.2 supports only Apache HiveServer2 and Hive version 0.11 and higher. Adapter connections will fail if you do not migrate to the latest versions.

For more information about HADOOP, see the *Reference Guide*.

Generic JDBC adapter instance

The new JDBC Adapter data store is designed for data sources that provide JDBC connections, it uses Data Services Adapter SDK, it's recommended for data sources which only support JDBC connections.

For more information on the JDBC adapter instance, see the *Designer Guide*.

New Web Services function, Get_Scheduled_Task

The Get_Scheduled_Task function retrieves the list of all scheduled tasks and detailed schedule information for the selected repository (for example, periodicity and next scheduled start time).

For more information on this function, see the *Integrator Guide*.

Columnar support for DB2 10.5 and Teradata 14.10

The new target table option, Table type is added to enable you to create tables organized row or column for DB2 10.5 and Teradata 14.10

For more information on this option, see the *Reference Guide*.

Allow merge or upsert supported for Teradata

The target table option, Allow merge or upsert is now supported for Teradata.

For more information on this option, see the *Reference Guide*.

OData support

You can now create an OData adapter instance in the Management Console of the Administrator.

OData (Open Data Protocol) is a standardized protocol for creating and consuming data APIs. For example, you can load and extract data from new OData based objects in the SuccessFactors API. In addition, when you customize objects or extensions, the data can be loaded through OData objects only.

i Note

Older objects, like SuccessFactors CompoundEmployee and BizX tables, cannot be exposed through OData at this time.

For more information, see the *Supplement for OData Guide*.

14.2 Usability

SAP Data Services 4.2 SP2 includes the following enhancements that improve the user's experience.

Expanded Designer search capabilities

Using SAP Data Services Designer, now you can search for a text string in every part of the object, such as table name and variable name.

For more information, see "Searching for objects" in the *Designer Guide*.

Operational Dashboard redesign

The Operational Dashboard has a new and improved user interface that provides you with graphical depictions of SAP Data Services job execution statistics.

Being able to see task statuses across repository organizations allows you to view the status and performance of your job executions for one or more repositories over a given time period.

For more information, see "Operational Dashboard" in the *Management Console Guide*.

WSDL Labels

Labels allow you to create WSDL files with only a subset of the API exposed. This makes it easier for you to manage your SOAP based applications. You can then open a filtered view of the API by adding a label parameter to the end of the URL for the WSDL file.

For more information, see "Using custom WSDL labels" in the *Integrator Guide*.

Failed Request Log

You can now have Data Services put failed request information into a log file. Data Services does not log failed requests by default because it can slow down performance. If you want failed request information written to a log file, you need to enable this feature in the Management Console Administrator.

For more information, see "Failed Request log" in the *Integrator Guide*.

14.3 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Dynamic input field added

Prior to this release, the TDP extraction transform required that every DTP option be defined at design time. The dynamic input field LANGUAGE has been added as an input field of type varchar(32) which is to be optionally mapped from some column of the input schema, thus enabling it to be specified at run time.

Substitution parameters added

Prior to this release, the TDP extraction transform did not support substitution parameters for its option values, and performed design-time verification of each option. Support for substitution parameters for the following option values allows using different configurations for development, test, and production environments:

- LANGUAGE
- DEFAULT_LANGUAGE
- DICTIONARY_ONLY
- ADVANCED_PARSING
- PROCESSING_TIMEOUT
- DOCUMENT_PROPERTIES
- DICTIONARY_FILE
- RULE_FILE

The value of a substitution parameter is validated at run time. If such a value is found invalid, an error is issued and the thread is stopped.

For more information, see the *Data Services Reference Guide*.

14.4 Data Quality transforms

In SAP Data Services 4.2 SP2, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

14.4.1 Data Cleanse transform

In SAP Data Services 4.2 SP2, the Data Cleanse solution has been enhanced to include the following functionality.

Family Name Full and Given Name Full output fields

The new composite field `Given_Name_Full` provides a single cleansed field for first names that represent the combination of `Given_Name1` and `Given_Name2` fields. The new composite field `Family_Name_Full` provides a single cleansed field for given names that represent the combination of `Family_Name1` and `Family_Name2`.

With the new fields, you can output the cleansed full dual first name and cleansed full dual last name without having to consolidate fields in a post processing step.

❁ Example

Input data	Parsed data		Composite
Ms. Mary Ann Smith Jones, CPA Account Mgr. Jones Inc.	Prenome	Ms.	
	Given Name1	Mary	Given_Name_Full
	Given Name2	Ann	Mary Ann
	Family Name1	Smith	Family_Name_Full
	Family Name2	Jones	Smith Jones
	Honoray Postname	CPA	
	Title	Account Mgr.	
	Firm	Jones, Inc.	

See the *Reference Guide* for descriptions of the two new fields.

14.4.2 Geocoder transform

In SAP Data Services 4.2 SP2, the Geocoder transform has been enhanced with the following features.

New country directory data

For this release, directory data is available for four additional countries:

- Estonia
- Lithuania
- Luxembourg
- Turkey

The Geocoder transform is flexible enough to accept new country directory data immediately after the directory data is released. There is no need to wait for the next Data Services release to begin using new country directory data.

Check with your sales representative for a list of the most current country directories available.

14.4.3 Global Address Cleanse transform

In SAP Data Services 4.2 SP2, the Global Address Cleanse transform has been enhanced with the following features.

Note

Important! To display the new Designer options and input and output fields introduced in this release, you must import the `gac.atl`, which by default is in the following location: `${LINK_DIR}/admin/repo`

- On UNIX: `$HOME/sap_businessobjects/data_services/admin/repo`
- On Windows: `C:\Program Files (x86)\SAP BusinessObjects\Data Services\admin\repo`

New Global Address Cleanse solutions for South Korea

The Global Address Cleanse transform has been enhanced to cleanse and validate South Korean addresses in native script and validate equivalent transliterated versions in Latin script down to the house-number level. South Korea roll out new standard from . The Global Address Cleansing transform can also convert old South Korea land-lot addresses into the new road name address standard by setting the Address Line Alias and Assign Locality options to Convert.

A new South Korea blueprint is available that contains a sample Global Address Cleanse transform configuration with best practice settings for cleansing address data in South Korea. You can access blueprints on the [SAP Data Services Blueprints](#) page of the SAP Community Network.

Simplified Suggestion List output fields

In addition to the powerful and flexible discrete output fields that you can output to suggestions lists, Data Services provides a simpler way to quickly output suggestions into a single address line or lastline. The following options have been added to enable the simplified suggestions lists:

Option	Description
Sugg Full Addressline	Select Yes to output the complete address line, including secondary address, and dual address (street and postal) as appropriate for the country. The default is No.
Sugg Full Lastline	Select Yes to output the locality, region, and postal code together in one component as appropriate for the country. The default is No.
Sugg Single Address	Select Yes to output the combined result of the full addressline and full lastline in the order appropriate for the country. The default is No.

Support major city translation

The Global Address Cleanse transform now allows you specify whether to translate major cities to English. This is specified with the Translate Major Locality option in the Global Address engine.

Option	Description
Translate Major Locality	<p>Specifies whether to translate the Locality1 output field for major localities.</p> <p><i>English</i>: Translates the output to English, if available.</p> <p><i>Preserve</i>: Outputs the locality data as it was input. This is the default value.</p> <div><p>i Note</p><p>The Translate Major Locality option takes precedence over the Assign Locality option. It also takes precedence over the Output Address Script option for the locality.</p></div>

Enhanced support for SAP software

Several enhancements have been made to the Global Address Cleanse transform to allow it to work seamlessly with other SAP software, such as SAP Customer Relationship Management.

- *New postal code check output field*. The Global Address Cleanse transform now performs a postal code format check, so that you can ensure that the cleansed results can be loaded to SAP software successfully.

Output field	Description	Engine
Postcode_In_Valid_Format	Indicates whether the postcode is in the correct format as defined by the postal authority for that country.	All engines

- *Additional Match_* output fields*.
The following output fields have been added that may be used in the Match transform during the comparison process.
 - Match_Country
 - Match_Locality2
 - Match_Postcode1
 - Match_Primary_Name2
 - Match_Region

15 SAP Data Services 4.2 SP 1 features

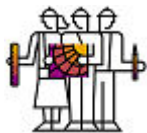
Learn about the new and changed features introduced in SAP Data Services 4.2 SP 1.

The following links take you to descriptions of the features by category for version 4.2 SP 1.



[Installation, administration, and monitoring \[page 140\]:](#)

- Separate Information Platform Services environment
- Updated Windows installation
- REST Web services



[Data Services Workbench \[page 140\]:](#) File as target



[Connectivity: Big data and CDC \[page 141\]:](#)

- Big data loading
- Support for long database identifiers
- SAP LT Replication Server integration
- Spatial data support in data Services



[Text Data Processing \[page 143\]:](#) Multithreaded TDP Entity Extraction transform



[Adapter SDK \[page 143\]:](#)

- Adapter SDK binary communication
- Adapter SDK simplification



[Data Quality transforms \[page 143\]:](#)

- Data Cleanse transform: Reports
- Geocoder transform: New country directory data
- Global Address Cleanse transform
 - Language and script conversion
 - Improved Turkish address assignment
 - Enhanced support for SAP software
 - Combining overlapping ranges

15.1 Installation, administration, and monitoring

SAP Data Services 4.2 SP1 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Separate Information platform services environment

This release recommends installing Data Services (and Information Steward) on a separate Information platform services system (instead of Business Intelligence platform) to provide flexibility to upgrade Data Services (and Information Steward) independently from Business Intelligence platform.

Updated Windows installation

The Data Services installation now includes two installation options.

Option	Description
<i>Install with default configuration</i>	This option installs and configures mandatory components using default values. The installation will configure a Data Services repository on the user-selected database via a server-based connection. This repository will be associated with the default job server that is configured during the installation.
<i>Install without configuration</i>	This option allows advanced users to install specific Data Services features and configure them post-installation. The user must configure the Data Services repository (repository creation and registration on the CMS) and Job Server.

REST web services

Representational State Transfer (REST or RESTful) web service is a design pattern for the World Wide Web. Data Services now allows you to call the REST server and then browse through and use the data the server returns.

For more information, see “REST web services” in the *Integrator Guide*.

15.2 Data Services Workbench

SAP Data Services 4.2 SP1 includes the following enhancement that improves the user's experience.

File as target

In this version you can choose to use file as target rather than datastore. This option enables you to replicate to multiple systems, replicate multiple times, and perform replication of source and target data that is located on different networks. The following changes are available to support the file as target functionality:

Window	New options
Replication Job editor	Replication Target Type – Choose Datastore or Files. Target File Properties – When selecting a target type of Files, enter the file properties. Generate File formats - Generates the file formats containing the schema for each file that is loaded. These file formats contain the exact schema (column names, data types and so on.) as defined in the target options
Data Flow Editor	Template File Loader - Includes the complete information needed to use the file loader which includes all the file format properties applicable for the loader and the loader specific properties as well (Such as Validate decimal Data and Delete File). The schema of the file is generated based on the output schema of the data flow element connected to the template file loader in the dataflow diagram.
Quick Replication Wizard	The Configure target settings window now includes the options, Target type (Files or Datastore) and Target file properties, and Target datastore properties.

15.3 Connectivity: Big data and CDC

SAP Data Services 4.2 SP1 includes the following connectivity enhancements.

Big data loading

Data Services 4.2 provides the big data file loading feature with the recovery option. Big data file loading is an extension to the Recovery Mechanism. When reading the big source flat file into the database tables, you can turn on the job recovery mode. The job recovery mode checks for the recoverability, divides the source file into smaller batches, and then processes each batch at a time. Therefore, when a job fails, it can be resumed from the last checked point and run forward instead of restarting.

For more information and limitations, see “Loading Big Data file with recovery option” in the *Designer Guide*.

Support for long database identifiers

The maximum length of most importable database metadata names and Data Services internal identifiers has increased from 64 to 256, depending on the Data Service repository type. For most repository types the maximum length is 256, for MySQL the length is 64, and for MS SQL server the length is 128.

SAP LT Replication Server integration

Data Services has been enhanced to integrate with SAP LT Replication Server (SLT) by leveraging the new version of ODP API. The existing extractor interface in Data Services has been enhanced and replaced with ODP in the Object Library and Metadata browsing. ODP allows uniform access to all contexts provided by ODP API.

In case the new interface is not available on the SAP system, the old extractor interface (ODP or native) will be used.

A new option has been added to the SAP datastore: "ODP Context ". The SAP datastore in Data Services has been enhanced to support the SLT objects. Working with these objects is similar to the way Data Services users work with SAP extractor today. The ODP context allows you to connect to both the extractors and the SLT.

For more information about ODP in SAP Data Services, see the *Supplement for SAP*.

SLT enhances the CDC (Change Data Capturing) scenario in Data Services, because with the trigger-based technology SLT adds delta-capabilities to every SAP or non-SAP source table which then allows for using CDC and transferring the delta data of the source table. For more information on the SLT for ODP scenario refer to the Installation Guide in ► www.service.sap.com/instguides ► [SAP Components](#) ► [SAP Landscape Transformation Replication Server](#) ► [Operational Data Provisioning with SAP LT Replication Server](#) ►.

Spatial data support in Data Services

In this release, Data Services supports spatial data (for example, point, line, polygon, collection, or a heterogeneous collection) for the following databases:

- Oracle—reading
- SAP HANA—reading and loading

When you import a table with spatial data columns into Data Services, the spatial type columns are imported as character-based large objects (clob). The column has an attribute, Native Type, which has the value of the actual data type in the database; for example, SDO_GEOMETRY for Oracle or ST_GEOMETRY for SAP HANA.

Limitations

- Because spatial columns are imported into Data Services as clob, creating template tables with spatial types is not supported.
- Spatial data cannot be manipulated inside a data flow because the spatial utility functions are not supported.

For more information about loading data from Oracle into SAP HANA, see the *SAP Data Services Supplement for SAP*. For more information about SAP HANA spatial data support, see the SAP HANA documentation.

15.4 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Multithreaded TDP Entity Extraction transform

The TDP Entity Extraction transform now permits multithreading. Previously, scaling was accomplished only by using multiple processes. Using multithreading, less memory is required when scaling, and better scaling can be achieved when processing custom extraction rules.

15.5 Adapter SDK

Adapter SDK binary communication

This version of the SDK Adapter allows you to use binary communication when creating a source table implementation through the Java Project wizard.

Adapter SDK simplification

This version of the SDK Adapter simplifies the process for creating an adapter in the following ways:

- The definition of externalized properties is simplified
- The need to generate and package the configuration XML files has been removed
- Many command line options now have defaults to simplify debugging setup
- The `getRowType` property is now externalized

For more information, see the *Data Services Adapter SDK Guide*.

15.6 Data Quality transforms

In SAP Data Services 4.2, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

15.6.1 Data Cleanse transform

In SAP Data Services 4.2 SP1, the Data Cleanse solution has been enhanced to include the following functionality.

Reports

To complement the many existing Data Quality reports, Data Services has added two reports. Use the Data Cleanse Information Code Summary report to view how the Data Cleanse transform modified the data. Use the Data Cleanse Status Code Summary report to view record counts generated for each status code.

For more information, see the *Data Services Reference Guide* and *Data Services Designer Guide*.

15.6.2 Geocoder transform

In SAP Data Services 4.2 SP1, the Geocoder transform has been enhanced with the following features.

New country directory data

For this release, directory data is available for nine additional countries:

- Czech Republic
- Finland
- Greece
- Liechtenstein
- Netherlands
- Norway
- Poland
- Portugal
- US TomTom (parcel)

The Geocoder transform is flexible enough to accept new country directory data immediately after the directory data is released. There is no need to wait for the next Data Services release to begin using new country directory data.

Check with your sales representative for a list of the most current country directories available.

15.6.3 Global Address Cleanse transform

In SAP Data Services 4.2 SP1, the Global Address Cleanse transform has been enhanced with the following features.

Language and script conversion

Output option for US ASCII

The Global Address Cleanse transform has an extended ability to cleanse address data in both local and international formats. If input address data contains local special characters, you can choose whether the output address data contains the special characters or their international equivalent. For example, if the street name “Østerbrogade” is input, you can preserve the local character “Ø” or convert it to the international data format “Osterbrogade” in the cleansed output.

For Latin script records, the new Convert Latin Output To US ASCII option converts any extended ASCII characters in the Best component to US ASCII characters, if a character conversion is available. Any extended ASCII characters for which there is no conversion (for example, the degree symbol or inverted exclamation and question marks), are left as is. By default, the option does not convert extended ASCII characters.

Romanization for Chinese and Kanji characters

Previously, when given a Chinese or Russian address, the Global Address Cleanse transform would identify the input script. If the input script was in native Chinese, the transform cleansed and validated it to the address-line level; however, if the input script was Latin, the address was validated only to the locality level. After processing, the address was output in the same script as it was input. The transform did not transliterate the address to the desired script.

With the transliteration support added in this release, the Global Address Cleanse transform now helps you cleanse and validate your Chinese and Russian address data more effectively, either in the native or Latin script. Data Services adds two features:

- Cleansing and validating Chinese and Russian addresses in Latin script to the address-line level.
- Transliterating native Chinese and Cyrillic addresses to Latin script.

A new engine option, Output Address Script, has been added for transliterating native Chinese and Russian addresses into Latin script in your output data.

Given a Chinese or Russian address, the transform first determines the input character script (CJKK, CYRL, or LATN). If the script is in CJKK or CYRL, the output data can be formatted and output as Latin (by setting the Output Address Script option to Latin) or the native script (by setting the option to Preserve). If the script is Latin, the address is output as Latin.

Improved Turkish address assignment

In this version, the Turkish address cleansing solution is enhanced in the following ways:

- A new address directory content that provides more comprehensive and accurate Turkish address data.

- The existing parsing dictionary and rule files have been enhanced to improve parsing accuracy.

Enhanced support for SAP software

Several enhancements have been made to the Global Address Cleanse transform to allow it to work seamlessly with other SAP software, such as SAP Customer Relationship Management.

- *ISO standard region code support.* You can now input and output ISO standard region codes (ISO 3166-2) from the Global Address Cleanse transform.
- *ISO standard region2 code support.* The Global Address Cleanse transform now supports region2 output so that both Region1 and Region2 components are available on output.
- *Discrete input field support for address components.* The Global Address Cleanse transform includes several new input fields, NetWeaver_<fieldname>, that map to the data model of SAP software.

⚠ Caution

Use the NetWeaver input fields properly to avoid unexpected results in your data. You cannot map multiline or Address_Line input fields when you use the NetWeaver input fields. Although the NetWeaver input fields appear discrete, they behave and are processed as multiline fields. They are mapped internally to Multiline1-12 before normal Global Address Cleanse processing is performed. If a NetWeaver input field is not mapped, the multiline that would have been mapped to it is mapped to the next available NetWeaver input field.

- *New output field to contain unused data.* A new output field, Remainder_Full, has been added that concatenates all extra and reminder data, including Address_Line_Remainder1-4 and Lastline_Remainder1-4, into a single output field.
- *New output field to combine primary name and secondary data.* A new output field, PName_Secondary_Addr, has been added that contains the full primary name (with no associated primary number) and the full secondary address.
- *New output field for private mailbox data.* A new output field, PMB_Full, contains private mailbox information for U.S. addresses.
- *Standard postal code format check.* The new NetWeaver_Formatted_Postcode output field lets you output a postal code in a format required by SAP software. The Postcode_In_NetWeaver output field indicates whether the NetWeaver_Formatted_Postcode output field is populated and is therefore a valid postal code.
- *Improved Country ID performance.*

Combining overlapping ranges

A new Combine Overlapping Ranges option has been added to the Global Address Suggestion List Options group that lets you specify whether individual suggestions with overlapping ranges are combined. You might set this option to Yes if you want to limit the number of total suggestions presented to your user. However, you might not see gaps of invalid ranges that would be apparent if this option were set to No.

For example, a suggestion list might show the following suggestions if this option is set to No:

1000-1099 Maple Ave

1100-1199 Maple Ave

But would only show this suggestion if set to Yes:

1000-1199 Maple Ave

16 SAP Data Services 4.2 features

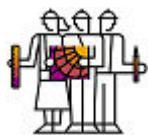
Learn about the new and changed features introduced in SAP Data Services 4.2.

The following links take you to descriptions of the features by category for version 4.2.



[Installation, administration, and monitoring \[page 149\]:](#)

- Operational statistics enhancements
- Object promotion management
- User application rights



[Data Services Workbench \[page 152\]:](#)

- Introduction to Data Services Workbench
- Create data flows
- Data Flow editor
- Validation in the Data Flow editor
- Query transform editor: Manage expression macros
- Create multiple file formats
- Changes to the Replication job editor
- Using data cleansing solutions from Data Cleansing Advisor



[Connectivity: Big data and CDC \[page 154\]:](#) Native Microsoft SQL Server support on UNIX



[Usability \[page 155\]:](#)

- Enhancements to variables and parameters
- Enhancements to global variables



[Functions and transforms \[page 155\]:](#)

- Mapping expression functionality added to the Map_Operation transform
- Batch Mode functionality added to the XML_Map transform



[Text Data Processing \[page 156\]:](#)

- Language identification
- Expansion of Dutch and Portuguese extraction
- New Simplified Chinese Voice of the Customer rules
- Expansion of Emoticon and Profanity extraction in French, German, and Spanish
- Expansion of Russian and Korean extraction



[Adapter SDK \[page 157\]](#): New interfaces



[Data Quality transforms \[page 157\]](#):

- Data Cleanse transform
 - Remove diacritical characters
 - Information and status codes
 - Phone parser
 - Parse discrete input
 - New Data Cleanse output fields
 - New Data Cleanse input fields
 - Memory cache
- Geocoder transform: New country directory data
- Global Address Cleanse transform
 - Address-line validation for Russia
 - Improved Austrian address assignment

16.1 Installation, administration, and monitoring

SAP Data Services 4.2 includes enhancements for administrators and IT personnel responsible for Data Services installation, administration, and monitoring.

Operational statistics enhancements

Additional web service operations were added to this release of Data Services. These operations give you the ability to generate dashboards to further analyze and review operational statistics.

The new operations include the following:

Operation	Description
Get_Repository_List	Retrieves a list of repositories at log in.
Get_BatchJob_ExeDetail	Retrieves a list of all job executions for the selected repository and job. You can also filter the list by time range.
Get_BatchJob_FlowDetails	Retrieves details about the tasks included in a job.
Get_DF_Auditdata	Shows audit information for a data flow.
Get_DF_Monitor_Log	Shows the runtime statistics for single data flow execution.
Get_DF_Monitor_Log_Pivot	Shows the runtime statistics as pivoted for a single data flow execution.

Object promotion management

This release of Data Services provides a new tool, Data Services Object Promotion Management, which is launched from the Administrator in the Data Services Management Console. The Object Promotion Management tool enables you to securely copy one or more Data Services objects from a development environment to a production environment. To ensure security, these environments typically do not have direct access to one another. Object promotion in Data Services is accomplished over secure FTP or over a shared directory that only an Administrator or a user with an account assigned to the Administrator group can configure for this purpose, maintaining security in both the source and target environments.

Users with [View](#) access to a repository can export its objects from a development environment to a shared directory created by authorized users. Users who have Full Control access to the repository can import objects after an Administrator or a user with an account assigned to the Data Services Administrator group has granted them the new [Manage Object Promotion Import](#) user application right.

i Note

The user who starts the Server Intelligence Agent (SIA) on the server that has access to the shared directories that are used for exporting and importing promoted objects must have full read and write access to those shared directories. If the server is started by a user who does not have full access to the shared directory, all configurations for export and import operations for that shared directory will fail validation and the configurations cannot be completed until the issue is resolved. If another user without full read write privileges to the shared directory starts SIA after configurations have been successfully validated, all export and import operations will fail until SIA is restarted by a user with the required credentials.

Objects with the same name and type are imported sequentially, based on the date when the export was initiated. When you view the list of exported objects in the [Import Configuration](#) page in the Administrator, the exported objects are grouped together with objects of the same type and object name, listed in order by the date they were exported. You can import the same object more than once as long as that version of the object has not yet been imported by itself or as part of an earlier import operation.

User application rights

This release of Data Services introduces new or modified user application rights to support best practices for security in message logs, create or modify object promotion export and import configurations, use an object import promotion configuration to import objects, or to edit substitution parameters.

User Right	Description
View internal information in log	<p>Messages displayed to a user from warning and error logs often contain internal system information which can provide details that might pose a potential security risk (for example, domain names or stack traces). A new user right included in this release, View internal information in log, provides the Administrator or a member of the Data Services Administrator group a way to limit who gets to see internal system information.</p> <p>If a user is not explicitly granted the View internal information in log user right through the Central Management Server (CMS), any internal system information displayed in warning or error messages is masked (replaced by a series of asterisks). A user whose account has been granted the View internal information in log user right will see the internal system information associated with warning or error messages in its unmasked form. The unmasked internal system information portion of each warning and error message is still retained in the server's log files.</p> <p>The View internal information in log user right is automatically granted to the following users when you install or upgrade to this release of Data Services:</p> <ul style="list-style-type: none"> • Data Services Administrator • Data Services Operator Users • Data Services Profiler Administrator Users <div> <p>Note</p> <p>SAP Information Steward, SAP Data Services On Demand, or SAP Data Services Workbench users who run the Metadata Browsing or View Data services and who do not have the View internal information in log right will see masked error and warning messages. If these users run any other services are used, they will see both the masked and unmasked versions of warning and error messages, even if their account has not been assigned the View internal information in log right.</p> </div>
Manage datastore and substitution param configurations	<p>An existing Data Services user right, Manage datastores, has been expanded to also grant permission for users to edit substitution parameters. When you upgrade to this release, all existing user accounts who were granted the Manage datastores right will automatically be updated to receive the Manage datastores and substitution parameters user right. The Administrator or a member of the Data Services Administrator group can grant this right to another user or group to allow them to modify substitution parameters. Members of the Data Services Administrator group are automatically granted this right.</p>
Manage Object Promotion Configurations	<p>By default, only a user whose account is a member of the Data Services Administrator group can create and manage object promotion import and export configurations. Members of the Administrator group or an Administrator can assign the Manage Object Promotion Configurations right to another user or group account to grant permission to edit configurations in the Administrator to use for exporting and importing objects.</p>
Manage Object Promotion Import	<p>By default, only a user whose account is a member of the Data Services Administrator group can run an object promotion import configuration. Members of the Administrator group or an Administrator can assign the Manage Object Promotion Import right to another user or group account to grant permission to use the object import configurations in Data Services Management Console to import objects from a shared directory into the production environment. The user must also have Full Access rights to the repository from which the objects were exported.</p>

16.2 Data Services Workbench

SAP Data Services 4.2 includes several enhancements that improve the user's experience.

Data Services Workbench

The Data Services Workbench is an application that simplifies the migration of data and schema information between different database systems.

In previous versions, migrating data and schema information required you to create many data flows in the Designer, with each data flow reading from a single source table and writing to a template target table. This process could take days. In addition, incompatibilities between the source and target database types could require manual schema and data corrections.

The Data Services Workbench automates this migration process. Instead of creating many data flows manually, you now provide connection information for the source and target databases and select the tables that you want to migrate. The Workbench automatically creates Data Services jobs, workflows, and data flows and imports them into a Data Services repository. You can execute and monitor these jobs from within the Workbench. In addition, the Workbench supports more advanced options such as bulk loading and delta loading.

Jobs created in the Workbench can be scheduled with the Data Services Management Console, and the generated objects can also be used as a starting point for further editing within the Data Services Designer. For example, you might want to add more advanced transformations that are not available directly in the Workbench.

For a complete list of supported sources and targets, as well as more information about using the Workbench to migrate data and schema information, see the *Workbench Guide*.

Create data flows

This version of Workbench enables you to create data flows and include them in the replication job. The transforms supported in this release are, Query transform, Case transform, Merge transform, Row Generation transform, Map Operation transform, Date Generation transform, and SQL transform. You can reference existing data flows from the Replication Job by dragging and dropping the data flows in from the Explorer. You can verify all data flows and specify whether the data flow is for initial load, delta load, or both.

You also have the ability to convert replication tables to data flows. Highlight multiple replication tables, right click the selection, and select convert data. You have the option to generate a single data flow or one data flow per replicated table.

Data Flow editor

The Workbench Data Flow editor includes the Data Flow diagram (top), list of input schemas for the selected transform (bottom left), and a tabbed area containing the properties of the object currently selected in the Data Flow diagram (bottom right).

Data Flow diagram	The Data Flow diagram is composed a graphical area (left side) and a palette of elements you can drop on the graphical area. The diagram toolbar includes options, such as Zoom slider, Snap to Grid, Highlight Related Element, and so on.
Input Schemas panel	This panel displays all the available input schemas / fields that the user can drag n drop to the properties area. Every field that is being used in the current property tab is marked with a little black triangle. Some fields are marked with a key icon indicating they are part of the primary key.

The options for Table Reader properties, File Reader properties, Table Loader properties, Merge properties, Case properties, SQL properties, Row Generation properties, Map Operation properties, Date Generation properties, and the Effective Date properties are the same as the Data Services Designer.

The Template Table Loader uses the same database specific loader options as Data Services Designer as well as some extra options:

- Use NVARCHAR for VARCHAR
- Drop and re-create table
- Double quote names
- Table Type (for HANA only)

Validation in the Data Flow editor

In this version, objects containing errors will be underlined in red and a tooltip will show the list of errors/warnings for an element.

Query transform editor: Manage expression macros

In this version you can define expression macros that you can re-use to apply the same function to a selection of fields with a single mouse click.

Create multiple file formats

In this version, you are able to create multiple file formats using the new wizard. You can select multiple delimited text files from either the job server or the local file system. The wizard then scans the files and automatically detects multiple file format parameters such as the row and column delimiters, the text qualifier and so on. The File Format editor is similar to the one in Data Services and currently supports only text-delimited files.

Changes to the Replication job editor

In this version, the Replication job editor includes the following changes:

- Files are supported as a source for replication.
- A replication job now allows using tables from different sources for replication.
- The source panel has been removed. Now you can use the Project Explorer as the source when selecting tables.

Using data cleansing solutions from Data Cleansing Advisor

Data Services Workbench now includes the ability to consume and model data cleansing solutions. This feature allows a technical ETL developer with no data quality expertise to create a data flow containing a data cleansing solution and validate the results in a production environment.

A data steward uses the Data Cleansing Advisor feature of SAP Information Steward to create and publish a data cleansing solution. The data cleansing solution is stored in the SAP BusinessObjects Business Intelligence platform's Central Management Server (CMS). In the Workbench, the ETL developer then models the data cleansing solution as a Data Cleansing Solution transform within the context of a data flow, maps the transform's input and output schema to production sources and targets, and validates the results in the product environment. Finally, the ETL developer generates and deploys the data flow as ATL to the Data Services repository for execution in the Data Services engine.

For more information about using data quality solutions in Workbench, see the *SAP Data Services Workbench Guide*. For more information about Data Cleansing Advisor, see the *SAP Information Steward User Guide*.

16.3 Connectivity: Big data and CDC

SAP Data Services 4.2 includes the following connectivity enhancements.

Native Microsoft SQL Server support on UNIX

Data Services 4.2 provides native Microsoft SQL Server support on UNIX as a source or a target (similar to what is available on Windows).

When using the UNIX job server or engine to access the MS SQL Server datastore, the following functionality is available:

- CDC support, which includes the Replication method for SQL Server 2008 and later and the CDC and Change Tracking methods for SQL Server 2008 and later.
- Bulk loading support.
For more information, see "To use the SQL Server ODBC bulk copy API" in the *Performance Optimization Guide*.

- [Allow merge or upsert](#) option support for SQL Server 2008 and later.
- Linked datastores support, which provide a one-way communication path from one database server to another.

The following limitations apply:

- You cannot use SQL Server as a repository on UNIX.
- Windows authentication for the MS SQL Server datastore is not supported on the UNIX platform.

16.4 Usability

SAP Data Services 4.2 includes the following enhancements that improve the user's experience.

Enhancements to variables and parameters

In the Data Services Designer, the [Variables and Parameters](#) window has been enhanced with the following items in the right-click menu:

- [Replicate](#): Lets you copy an existing variable or parameter and then modify it as necessary.
- [Insert Before](#) and [Insert After](#): Let you insert a variable or parameter before or after an existing variable or parameter.
- [Move Up](#) and [Move Down](#): Let you specify the order of variable or parameters.

Enhancements to global variables

In the Data Services Designer, the [Global Variables](#) tab of a job's properties has been enhanced in the following ways:

- Improved column resizing.
- **Filter**: Lets you filter the displayed global variables by name.
- **Multi-select**: Lets you select multiple global variables at one time.
- **Bulk value update**: Lets you modify the values of multiple global variables.

For more information about variables and parameters, see “Variables and Parameters” in the *Designer Guide*.

16.5 Functions and transforms

SAP Data Services 4.2 includes enhancements to transforms and function support.

Mapping expression functionality added to the Map_Operation transform

You can now use the Map_Operation transform to modify data based on current operation codes and mapping expressions. The operation codes can then be converted between data manipulation operations.

Writing map expressions per column and per row-type (INSERT/UPDATE/DELETE) allows you to:

- change the value of data for a column.
- execute different expressions on a column, based on its row type.
- use the new before_image function to access the before image value of an UPDATE row.

i Note

Mapping expressions are optional. If there are no mapping expressions, Data Services performs operation codes only.

For more information, see the “Map_Operation” topic in the *Reference Guide*.

Batch Mode functionality added to the XML_Map transform

You can now use the XML_Map transform in Batch mode to accumulate data as a block of rows instead of a single row. This block of rows is sent as a unit to the next transform.

i Note

Only one input is allowed in Batch mode.

For more information, see the “XML_Map” topic in the *Reference Guide*.

16.6 Text Data Processing

Text Data Processing analyzes content and automatically identifies and extracts entities and facts in multiple languages. It includes a set of transforms that work together to improve the quality of your data. It has been enhanced with the following features:

Language identification

The Entity Extraction transform automatically identifies the input language to enable both targeted extraction and selection of language-specific dictionaries and rule files for that language.

Expansion of Dutch and Portuguese extraction

Dutch and Portuguese entity extraction has been expanded to recognize 31 entity types.

New Simplified Chinese Voice of the Customer rules

Simplified Chinese extraction has been expanded to recognize sentiments and problems, emoticons, and general & contact requests.

Expansion of Emoticon and Profanity extraction in French, German, and Spanish

Emoticons and profanities are now extracted for not only English, but also French, German, and Spanish.

Expansion of Russian and Korean extraction

Provides new predefined types, such as REGION/MAJOR, CONTINENT, and DATE.

16.7 Adapter SDK

In SAP Data Services 4.2, we have added four new interfaces to the Adapter SDK. These interfaces allow you to:

- implement CDC (Change data capture) datastores for adapters
- implement finer controls for pushdown operations.

For more information, see the *Data Services Adapter SDK Guide*.

16.8 Data Quality transforms

In SAP Data Services 4.2 SP1, the set of transforms that work together to improve the quality of your data has been enhanced with the following features.

16.8.1 Data Cleanse transform

In SAP Data Services 4.2, the Data Cleanse solution has been significantly enhanced to include the following functionality.

Remove diacritical characters

Use the new [Remove Diacritical Characters](#) option to replace diacritical characters with the ASCII equivalent version. The addition of this option eliminates the need to use a Query transform to replace the characters.

The option is defaulted to [No](#) for all configurations except for the US North American configuration where the option is defaulted to [Yes](#) to replace the diacritical characters. All other configurations retain the diacritical characters. These settings make the option compatible with previous versions of Data Services.

To view the conversion chart, see the *Data Services Reference Guide*.

Information and status codes

The Data Cleanse transform now includes information and status codes to help you understand how the data is processed. Status codes provide information about corrections and standards applied to the data. Information codes provide information about data that is suspect and might require a manual review.

To view the Data Cleanse information and status codes, see the *Data Services Reference Guide*.

Phone parser

The new Phone parser parses both North American phone numbers and international phone numbers. You can still use the individual North American and International phone parsers. However, when using Multiline1-12, the new Phone parser is used. If both North_American_Phone and International_Phone parsers are selected, the first parser selected is changed to Phone, and the second parser is removed.

Prior to upgrade	Post upgrade
SSN DATE NORTH_AMERICAN_PHONE INTERNATIONAL_PHONE PERSON_OR_FIRM	SSN DATE PHONE PERSON_OR_FIRM
INTERNATIONAL_PHONE NORTH_AMERICAN_PHONE	PHONE
INTERNATIONAL_PHONE EMAIL NORTH_AMERICAN_PHONE	PHONE EMAIL

You can use the existing North_American_Phone and International_Phone output fields or you can use the new phone output fields.

Generated field name	Description
International_Phone	The entire international phone number, including extra items such as the country code.
International_Phone_Country_Code	The country code of an international phone number.
International_Phone_Country_Name	The name of the country of origin of an international phone number.
International_Phone_Line	The portion of the international phone number that is not the country code or the city code.
International_Phone_Locality_Code	The locality code of an international phone number
North_American_Phone	An entire North American Numbering Plan (NANP) phone number.
North_American_Phone_Area_Code	The area code parsed from the phone number.
North_American_Phone_Extension	An extension parsed from the phone number.
North_American_Phone_Line	The last four numbers (excluding an extension) parsed from a phone number. In (123) 456-7890, 7890 is returned.
North_American_Phone_Prefix	The middle three numbers parsed from a phone number. In (123) 456-7890, 456 is returned.
North_American_Phone_Type	The type of phone number that was parsed, if it is included with the input. For example, Home or Work.
Phone	Shows the phone number that was identified as either North American or International.

See the *Data Services Reference Guide* and *Data Services Designer Guide*.

Parse discrete input

When parsing discrete person input fields, you might get better discrete output if you set the [Parse Discrete Input](#) option to [Yes](#) in the Data Cleanse transform. With this option on, the transform concatenates the discrete person input fields into a single name-line field, and then parses and standardizes the data into discrete output fields.

In the following example, you can see how the input data is output differently based on how this option is set.

Input data

Column	Field
Person1_Given_Name1	Mr John T
Person1_Family_Name1	Smith Iii

Output data

Column	Option=No	Option=Yes
Person1.Prenome	<blank>	Mr
Person1.Given_Name1	Mr John T	John
Person1.Given_Name2	<blank>	T

Column	Option=No	Option=Yes
Person1.Family_Name1	Smith Iii	Smith
Person1.Maturity_Postname	<blank	III

New Data Cleanse output fields

Previously, you had to perform additional processing to prepare the data for other tasks that you want to accomplish. When you want to use the standardized data for downstream processes such as creating break keys in the Match transform, use the following output fields.

Generated field name	Content type	Description
Match_Family_Name	Family_Name1	The combined standardized form of FamilyName1 and FamilyName2 with a space between used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space. PreFamilyName data is removed.
Match_Firm	Firm	A form of Firm that may be used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space, and data that is extraneous for matching purposes is removed. This extraneous data includes business types such as Ltd. and GmbH, and noise words such as The, And, and Of.
<div>Note</div> <div>Some words are classified to be removed from all domains, while others are language-specific and are classified to be removed in specific cultural domains.</div>		
Match_Given_Name1	Given_Name1	The standardized form of GivenName1 used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space. PreGivenName data is removed.
Match_Given_Name2	Given_Name2	The standardized form of GivenName2 used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space. PreGivenName data is removed.
Match_Maturity_Postname	Postname	The standardized form of MaturityPostname used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space.
Match_Phone	Phone	The standardized form of Phone used in the Match transform during the comparison process. Data is output as a string of digits. Spaces, punctuation, alphabetical characters and leading zeros are removed.

Generated field name	Content type	Description
Match_Prename	Prename	The standardized form of Prename used in the Match transform during the comparison process. Data is output in uppercase, apostrophes are removed, and other punctuation is replaced with a single space.

To see the Data Cleanse output fields, see the *Data Services Reference Guide*.

New Data Cleanse input fields

Several input fields have been added to automatically assign the content domain sequence and output format.

Input fields	Description
Option_Country	The content domain sequence and output format usually can be automatically generated based on the Option_Country data. However, there are a few countries where Option_Language and Option_Region data is helpful to make the assignment, for example, Switzerland, Belgium, and Canada. Using Option_Language and Option_Region is optional, and is only used to determine the most appropriate content domain and output format.
Option_Language	
Option_Region	
	These input fields should be mapped from the following Global Address Cleanse output fields in this order:
	<ul style="list-style-type: none"> • ISO_Country_Code_2Char • Language • Region1

These input fields override the Data Cleanse [Content Domain Sequence](#) and [Output Format](#) transform options. However, the Option_Output_Format and Option_Content_Domain_Sequence dynamic input fields override the input fields list above as well as the Data Cleanse transform options.

To see the Data Cleanse input fields, see the *Data Services Reference Guide*. For more information about assigning the content domain sequence and output format options, see the *Data Services Designer Guide*.

Memory cache

To improve performance, you can set the cache in kilobytes (KB) to allocate a certain amount of memory for use in the Data Cleanse transform. For example, if you have a 39 MB cleansing package, you can set the cache size to 40000 KB and use the rest of available memory for processing other transforms. Likewise, if your cleansing package is larger, you can increase the value.

16.8.2 Geocoder transform

In SAP Data Services 4.2, the Geocoder transform has been enhanced with the following features.

New country directory data

For this release, directory data is available for six additional countries:

- Austria
- Belgium
- Denmark
- Italy
- Spain
- Sweden

The Geocoder transform is flexible enough to accept new country directory data immediately after the directory data is released. There is no need to wait for the next Data Services release to begin using new country directory data.

Check with your sales representative for a list of the most current country directories available.

16.8.3 Global Address Cleanse transform

In SAP Data Services 4.2, the Global Address Cleanse transform has been enhanced with the following features.

Address-line validation for Russia

The Global Address Cleanse transform can now cleanse Russian address-line data, so that you can parse, validate, and correct Russian address information. To support this feature, Data Services provides the following:

- New Russia address directory to enable validation to the address-line component level.
- Enhanced parsing dictionary and rule file to improve parsing accuracy.
- Transliteration support to support Latin or Cyrillic scripts for address assignment.

Improved Austrian address assignment

Austrian address assignment has been improved with the following enhancements:

- The reference data and parsing rules have been enhanced to parse building and block data.
- Assignment for addresses that have no house number but have a suffix (stair, group, block, and so on) and are valid addresses.

Building and block data

In the following examples both inputs are assigned to the same output, which illustrates how suffixes can be either written in full or separated with a forward slash (/).

Example 1:

Input 1			Input 2		
Artillerieplatz 1 Haus 6 Tür 1			Artillerieplatz 1/6/1		
1110 Wien			1110 Wien		
Postcode1	Locality1_Name	Pri- mary_Name_Full1	Primary_Number	Building_Name1	Secondary_Ad- dress
1110	WIEN	ARTILLERIEPLATZ	1	HAUS 6	1

Example 2:

Input 1			Input 2		
Klg Feuchterweg 10 Gruppe 5			Klg Feuchterweg 10/5		
1110 Wien			1110 Wien		
Postcode 1	Locality1_Name	Primary_Name_Full1	Primary_Number	Building_Name1	
1110	WIEN	KLK FEUCHTER	10	GRUPPE 5	

Addresses with a suffix and no house number

In the following examples, the suffixes have been assigned and validated for addresses that have no house number.

Example 1:

Input 1			Input 2		
Klg Gaswerk Gruppe 5 Parzelle 308			Klg Gaswerk Gruppe 5/308		
1110 Wien			1110 Wien		
Postcode1	Locality1_Name	Primary_Name_Full1	Primary_Number	Building_Name1	
1110	WIEN	KLK GASWERK		GRUPPE 5 PARZELLE 308	

Example 2:



Input 1					
Karlsplatz Objekt 26					
1040 Wien					
Postcode1	Locality1_Name	Primary_Name_Full1	Primary_Number	Building_Name1	
1110	WIEN	KARLSPLATZ		OBJECT 26	

Important Disclaimers and Legal Information

Hyperlinks

Some links are classified by an icon and/or a mouseover text. These links provide additional information.

About the icons:

- Links with the icon  : You are entering a Web site that is not hosted by SAP. By using such links, you agree (unless expressly stated otherwise in your agreements with SAP) to this:
 - The content of the linked-to site is not SAP documentation. You may not infer any product claims against SAP based on this information.
 - SAP does not agree or disagree with the content on the linked-to site, nor does SAP warrant the availability and correctness. SAP shall not be liable for any damages caused by the use of such content unless damages have been caused by SAP's gross negligence or willful misconduct.
- Links with the icon  : You are leaving the documentation for that particular SAP product or service and are entering a SAP-hosted Web site. By using such links, you agree that (unless expressly stated otherwise in your agreements with SAP) you may not infer any product claims against SAP based on this information.

Videos Hosted on External Platforms

Some videos may point to third-party video hosting platforms. SAP cannot guarantee the future availability of videos stored on these platforms. Furthermore, any advertisements or other content hosted on these platforms (for example, suggested videos or by navigating to other videos hosted on the same site), are not within the control or responsibility of SAP.

Beta and Other Experimental Features

Experimental features are not part of the officially delivered scope that SAP guarantees for future releases. This means that experimental features may be changed by SAP at any time for any reason without notice. Experimental features are not for productive use. You may not demonstrate, test, examine, evaluate or otherwise use the experimental features in a live operating environment or with data that has not been sufficiently backed up.

The purpose of experimental features is to get feedback early on, allowing customers and partners to influence the future product accordingly. By providing your feedback (e.g. in the SAP Community), you accept that intellectual property rights of the contributions or derivative works shall remain the exclusive property of SAP.

Example Code

Any software coding and/or code snippets are examples. They are not for productive use. The example code is only intended to better explain and visualize the syntax and phrasing rules. SAP does not warrant the correctness and completeness of the example code. SAP shall not be liable for errors or damages caused by the use of example code unless damages have been caused by SAP's gross negligence or willful misconduct.

Bias-Free Language

SAP supports a culture of diversity and inclusion. Whenever possible, we use unbiased language in our documentation to refer to people of all cultures, ethnicities, genders, and abilities.

© 2022 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company. The information contained herein may be changed without prior notice.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

Please see <https://www.sap.com/about/legal/trademark.html> for additional trademark information and notices.