



Informatica® PowerExchange for Google
Cloud Spanner
10.5

User Guide

© Copyright Informatica LLC 2018, 2021

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

Informatica, the Informatica logo, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2021-04-01

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Google Cloud Spanner.....	 7
PowerExchange for Google Cloud Spanner Overview.	7
Introduction to Google Cloud Spanner.	7
 Chapter 2: PowerExchange for Google Cloud Spanner Configuration.....	 8
PowerExchange for Google Cloud Spanner Configuration Overview.	8
Prerequisites.	8
Administration of PowerExchange for Google Cloud Spanner.	9
 Chapter 3: Google Cloud Spanner Connections.....	 11
Google Cloud Spanner Connections Overview.	11
Google Cloud Spanner Connection Properties.	11
Creating an Google Cloud Spanner Connection.	12
 Chapter 4: PowerExchange for Google Cloud Spanner Data Objects.....	 13
Google Cloud Spanner Data Object Overview.	13
Google Cloud Spanner Data Object Properties.	13
Google Cloud Spanner Data Object Read Operation.	14
Output Properties of the Data Object Read Operation.	14
Google Cloud Spanner Data Object Write Operation.	16
Input Properties of the Data Object Write Operation.	16
Creating an Google Cloud Spanner Data Object.	18
Creating a Google Cloud Spanner Data Object Operation.	18
Rules and Guidelines for Google Cloud Spanner Mappings.	19
 Chapter 5: PowerExchange for Google Cloud Spanner Mappings.....	 20
PowerExchange for Google Cloud Spanner Mappings Overview.	20
Mapping Validation and Run-time Environments.	20

Chapter 6: Google Cloud Spanner Run-Time Processing..... 22

Google Cloud Spanner Run-Time Processing Overview. 22

Partitioning. 22

Parameterization for Google Cloud Spanner Sources and Targets. 23

Parameterization for Google Cloud Spanner Targets. 23

Appendix A: Google Cloud Spanner Data Type Reference..... 24

Data Type Reference Overview. 24

Google Cloud Spanner and Transformation Data Types. 24

Index..... 26

Preface

Use the *Informatica PowerExchange for Google Cloud Spanner User Guide* to learn how to read from and write to Google Cloud Spanner by using the Developer tool. Learn to create a Google Cloud Spanner connection, develop and run mappings in the native environment and in the Hadoop environments.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <https://network.informatica.com>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <https://search.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <https://docs.informatica.com>.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <http://velocity.informatica.com>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

To find online support resources on the Informatica Network, visit <https://network.informatica.com> and select the eSupport option.

CHAPTER 1

Introduction to PowerExchange for Google Cloud Spanner

This chapter includes the following topics:

- [PowerExchange for Google Cloud Spanner Overview, 7](#)
- [Introduction to Google Cloud Spanner, 7](#)

PowerExchange for Google Cloud Spanner Overview

You can use PowerExchange for Google Cloud Spanner to read data from and write data to Google Cloud Spanner.

You can use Google Cloud Spanner objects as sources and targets in mappings. When you use Google Cloud Spanner objects in mappings, you must configure properties specific to Google Cloud Spanner.

You can validate and run Google Cloud Spanner mappings in the native environment or on the Spark engine in the Hadoop environment. You can also run profiles against Google Cloud Spanner objects in the native environment.

Example

An enterprise application uses a relational database to store the customer details such as customerID, customerName, accountNum, phoneNum, address, and age. You can move the data from the relational database to Google Cloud Spanner to achieve scalability and high availability of the customer details. Use PowerExchange for Google Cloud Spanner to move the data from the relational database to a Google Cloud Spanner target.

Introduction to Google Cloud Spanner

Google Cloud Spanner is a fully managed relational database service that the Google Cloud Platform provides. Google Cloud Spanner is ideal for relational, structured, and semi-structured data that requires high availability, strong consistency, and transactional read and write operations. You can use Google Cloud Spanner for general purpose transactions using SQL (ANSI 2011 with extensions).

CHAPTER 2

PowerExchange for Google Cloud Spanner Configuration

This chapter includes the following topics:

- [PowerExchange for Google Cloud Spanner Configuration Overview, 8](#)
- [Prerequisites, 8](#)
- [Administration of PowerExchange for Google Cloud Spanner, 9](#)

PowerExchange for Google Cloud Spanner Configuration Overview

The PowerExchange for Google Cloud Spanner installs with Informatica Services. You can enable PowerExchange for Google Cloud Spanner with a license key.

To configure PowerExchange for Google Cloud Spanner, complete the prerequisites.

Prerequisites

To use PowerExchange for Google Cloud Spanner, complete the following tasks:

1. Install and configure Informatica Services.
2. Install and configure the Developer tool. You can install the Developer tool when you install Informatica clients.
3. Create and configure a Model Repository Service and a Data Integration Service in the Informatica domain.
4. Verify that you have write permissions on all the directories within the `<Informatica installation directory> directory`.
5. Ensure that the PowerExchange for Google Cloud Spanner license is activated.

Administration of PowerExchange for Google Cloud Spanner

Before you use PowerExchange for Google Cloud Spanner, you must complete the following tasks:

- Create a Google account to access Google Cloud Spanner.
- Ensure that you have the `client_email`, `project_id`, and `private_key` values for the service account. You will need to enter these details when you create a Google Cloud Spanner connection in the Developer tool.
- Enable billing for the project that you created.
- Ensure that you have enabled the Cloud Spanner API for your service account. PowerExchange for Google Cloud Spanner uses the Google Cloud Spanner API to integrate with Google Cloud Spanner.
- Ensure that you have selected the regional or multi-region configuration and specified the number of nodes that you want to use for the Google Cloud Spanner instance.

Note: When you increase the number of nodes, the cost charged for a billing account increases. Before you specify the number of nodes, see the Google Cloud Spanner documentation to estimate the cost of using Google Cloud Spanner. To evaluate Google's pricing details for Google Cloud Spanner, click the following URL:

<https://cloud.google.com/spanner/pricing>

- Ensure that you have the database name and table name in the Google Cloud Spanner instance. You will need to enter these details when you run a session in PowerCenter. For more information about creating a Google Cloud Spanner instance, database, and table, click the following URL:

<https://cloud.google.com/spanner/docs/quickstart-console>

- When you read data from or write data to a Google Cloud Spanner table, you must have the following permissions:

- `spanner.databases.beginOrRollbackReadWriteTransaction`
- `spanner.databases.beginPartitionedDmlTransaction`
- `spanner.databases.beginReadOnlyTransaction`
- `spanner.databases.getDdl`
- `spanner.databases.read`
- `spanner.databases.select`
- `spanner.databases.updateDdl`
- `spanner.databases.write`
- `spanner.sessions.create`
- `spanner.sessions.delete`
- `spanner.sessions.get`
- `resourcemanager.projects.get`
- `spanner.databases.list`
- `spanner.instances.get`
- `spanner.instances.list`

- When you only read data from a Google Cloud Spanner table, you must have the following permissions:
 - `spanner.databases.beginReadOnlyTransaction`

- `spanner.databases.getDdl`
- `spanner.databases.read`
- `spanner.databases.select`
- `spanner.sessions.create`
- `spanner.sessions.delete`
- `spanner.sessions.get`
- `resourceManager.projects.get`
- `spanner.databases.list`
- `spanner.instances.get`
- `spanner.instances.list`

For more information about creating a Google Cloud Spanner instance, database, and table, see the following site:

<https://cloud.google.com/spanner/docs/quickstart-console>

CHAPTER 3

Google Cloud Spanner Connections

This chapter includes the following topics:

- [Google Cloud Spanner Connections Overview, 11](#)
- [Google Cloud Spanner Connection Properties, 11](#)
- [Creating an Google Cloud Spanner Connection, 12](#)

Google Cloud Spanner Connections Overview

Use a Google Cloud Spanner connection to access a Google Cloud Spanner database.

Use the Google Cloud Spanner connection to import Google Cloud Spanner metadata, create data objects, preview data, and run mappings. When you create a Google Cloud Spanner connection, you define the connection attributes that the Developer tool uses to connect to the Google Cloud Spanner database.

Use the Developer tool, Administrator tool, or infacmd to create a Google Cloud Spanner connection.

Google Cloud Spanner Connection Properties

When you set up a Google Cloud Spanner connection, you must configure the connection properties.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes the Google Cloud Spanner connection properties:

Property	Description
Name	The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:~`!\$%^&*()-+={} \\"';<,>.?/
ID	String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. The ID must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name.
Description	Optional. The description of the connection. The description cannot exceed 4,000 characters.
Location	The domain where you want to create the connection.
Type	The connection type. Select Google Cloud Spanner.
Project ID	Specifies the project_id value present in the JSON file that you download after you create a service account. If you have created multiple projects with the same service account, enter the ID of the project that contains the bucket that you want to connect to.
Service Account ID	Specifies the client_email value present in the JSON file that you download after you create a service account.
Service Account Key	Specifies the private_key value present in the JSON file that you download after you create a service account.
Instance ID	Name of the instance that you created in Google Cloud Spanner.

Creating an Google Cloud Spanner Connection

Create a Google Cloud Spanner connection before you create a Google Cloud Spanner data object.

1. In the Developer tool, click **Window > Preferences**.
2. Select **Informatica > Connections**.
3. Expand the domain in the **Available Connections**.
4. Select the connection type **Enterprise Application > Google Cloud Spanner**, and click **Add**.
5. Enter a connection name and an optional description.
6. Select Google Cloud SpannerConnection as the connection type.
7. Click **Next**.
8. Configure the connection properties.
9. Click **Test Connection** to verify the connection to Google Cloud Spanner.
10. Click **Finish**.

CHAPTER 4

PowerExchange for Google Cloud Spanner Data Objects

This chapter includes the following topics:

- [Google Cloud Spanner Data Object Overview, 13](#)
- [Google Cloud Spanner Data Object Properties, 13](#)
- [Google Cloud Spanner Data Object Read Operation, 14](#)
- [Google Cloud Spanner Data Object Write Operation, 16](#)
- [Creating an Google Cloud Spanner Data Object, 18](#)
- [Creating a Google Cloud Spanner Data Object Operation, 18](#)
- [Rules and Guidelines for Google Cloud Spanner Mappings, 19](#)

Google Cloud Spanner Data Object Overview

A Google Cloud Spanner data object is a physical data object that uses Google Cloud Spanner as a source or target. A Google Cloud Spanner data object represents data based on a Google Cloud Spanner resource.

You can configure the data object read and write operation properties that determine how you can read data from Google Cloud Spanner sources and load data to Google Cloud Spanner targets.

Create a Google Cloud Spanner data object from the Developer tool. PowerExchange for Google Cloud Spanner creates the data object read operation and data object write operation for the Google Cloud Spanner data object. You can edit the advanced properties of the data object read or write operation and run a mapping.

Google Cloud Spanner Data Object Properties

Specify the data object properties when you create the data object.

The following table describes the properties that you configure for the Google Cloud Spanner data objects:

Property	Description
Name	Name of the Google Cloud Spanner data object.
Location	The project or folder in the Model Repository Service where you want to store the Google Cloud Spanner data object.
Connection	Name of the Google Cloud Spanner connection.

Google Cloud Spanner Data Object Read Operation

Create a mapping with a Google Cloud Spanner data object read operation to read data from Google Cloud Spanner.

Output Properties of the Data Object Read Operation

The output properties represent data that the Data Integration Service passes into the mapping pipeline. Select the output properties to configure advanced properties of the data object read operation.

The output properties of the data object read operation include general properties that apply to the data object operation. The output properties also include source, query, run-time, and advanced properties that apply to the Google Cloud Spanner data object.

You can view and change the output properties of the data object read operation from the **General**, **Sources**, **Query**, **Run-time**, and **Advanced** tabs.

General Properties

The general properties display the name and description of the data object read operation.

Sources Properties

The sources properties list the Google Cloud Spanner objects used in the data object read operation. You cannot join data from multiple sources of the Google Cloud Spanner data object in a read operation.

Ports Properties

The output ports properties display the data types, precision, and scale of the data object read operation.

The following table describes the output ports properties that you configure in the data object read operation:

Property	Description
Name	Name of the port.
Type	Data type of the port.

Property	Description
Precision	Maximum number of significant digits for numeric data types, or maximum number of characters for string data types. For numeric data types, precision includes scale.
Scale	Maximum number of digits after the decimal point for numeric values.
Description	Description of the port.

Query Properties

Use the **Query** tab to select specific records from a Google Cloud Spanner table.

The following table describes the query properties that you configure for a data object read operation:

Property	Description
Query	Filter value in a read operation. The filter specifies the where clause of select statement. Use a filter to reduce the number of rows that the Data Integration Service reads from the source. When you enter a source filter, the Developer tool adds a WHERE clause to the default query. You can use the Native or Platform expression to select specific records.

Run-time Properties

The run-time properties displays the name of the connection that the Data Integration Service uses to read data from the Google Cloud Spanner table.

You can define the partition type as key range partitioning to read data from a Google Cloud Spanner data object. When you configure key range partitioning, the Data Integration Service distributes rows of data based on a port or set of ports that you define as the partition key. You can define a range of values for each port. The Data Integration Service uses the key and ranges to send rows to the appropriate partition.

Advanced Properties

The Data Integration Service reads data from Google Cloud Spanner based on the data object read operation.

The Developer tool displays advanced properties for the Google Cloud Spanner data object operation in the Advanced view.

The following table describes the advanced properties for a Google Cloud Spanner data object read operation:

Property	Description
Index Directive	Specify the index directive in an SQL statement to indicate the index name to query a table. To specify a index directive, use the following syntax: FROM TableName@{FORCE_INDEX=TableIndex} For more information about index directive, click the following URL: https://cloud.google.com/spanner/docs/secondary-indexes
Batch Size	Minimum number of rows that PowerExchange for Google Cloud Spanner reads in a batch. Enter a number greater than 0. Default is 100.

Property	Description
Source Database	Optional. Overrides the Google Cloud Spanner database name specified in the Google Cloud Spanner data object read operation.
Source Table	Overrides the Google Cloud Spanner table name specified in the Google Cloud Spanner data object read operation.
EnableBulkRead	Reads data from large Google Cloud Spanner tables.

Google Cloud Spanner Data Object Write Operation

Create a mapping to write data to Google Cloud Spanner. Use the Google Cloud Spanner connection, and define the write operation properties to write data to Google Cloud Spanner.

You can perform insert, update, delete, and upsert operations on a Google Cloud Spanner target.

To perform a update, delete, or upsert operation, you must add an Update Strategy transformation to a mapping and flag the records in an update strategy expression.

Note: When you run a mapping on the Spark engine, you cannot add an Update Strategy transformation. Hence, you cannot perform update, delete, or upsert operations on a Google Cloud Spanner target when you run a mapping on the Spark engine.

For more information about configuring an Update Strategy transformation, see "Update Strategy Transformation" in the *Informatica Developer Transformation Guide*.

Input Properties of the Data Object Write Operation

Input properties represent data that the Data Integration Service writes to a Google Cloud Spanner table. Select the input properties to configure run-time properties of the data object write operation. You can also specify advanced data object write operation properties to write data to Google Cloud Spanner table.

The input properties of the data object write operation include general properties that apply to the data object write operation. Input properties also include source, run-time, and advanced properties that apply to the data object write operation.

You can view and change the input properties of the data object write operation from the **General**, **Sources**, **Run-time**, and **Advanced** tabs.

General Properties

The general properties list the name and description of the data object write operation.

Target Properties

The target properties list the Google Cloud Spanner table in the data object write operation.

Run-time Properties

The run-time properties displays the name of the connection that the Data Integration Service uses to write data to the Google Cloud Spanner table.

You can also use the **Run-time** tab to configure partitioning. You can configure dynamic partitioning for Google Cloud Spanner data object write operation.

Advanced Properties

Google Cloud Spanner data object write operation properties include advanced properties that apply to the Google Cloud Spanner data object.

The Developer tool displays advanced properties for the Google Cloud Spanner data object write operation in the **Advanced** tab.

You can configure the following advanced properties in the data object write operation:

Property	Description
UpdateMode	Determines the mode that the Data Integration Service uses to update rows in the Google Cloud Spanner target. You can select one of the following modes: <ul style="list-style-type: none">- Update As Update. The Data Integration Service updates all rows flagged for update in the Update Strategy transformation if the entries exist.- Update Else Insert. The Data Integration Service first updates all rows flagged for update in the Update Strategy transformation if the entries exist in the target. If the entries do not exist, the Data Integration Service inserts the entries. Default is Update As Update.
Target DataBase	Optional. Overrides the Google Cloud Spanner database name specified in the Google Cloud Spanner data object write operation.
Target Table	Optional. Overrides the Google Cloud Spanner target table name specified in the Google Cloud Spanner data object write operation.
Batch Size	Minimum number of rows in a batch. Enter a number greater than 0. Default is 100.
Create Table If Required	If the table is not present in the target database, PowerExchange for Google Cloud Spanner creates the target table with the name that you specify in the Target Table field. PowerExchange for Google Cloud Spanner creates the table with the same schema as the target table that you specified in the Google Cloud Spanner data object write operation.
Create DataBase If Required	If the database specified in the Target DataBase field is not present in the Google Cloud Spanner instance, PowerExchange for Google Cloud Spanner creates the target database with the name that you specify in the Target DataBase field.
Write Truncate	When you enable this property, PowerExchange for Google Cloud Spanner overwrites the existing data in the target table. Note: The Write Truncate option is applicable only when you perform an insert operation on a Google Cloud Spanner target.
Target Schema Strategy	Not applicable for PowerExchange for Google Cloud Spanner.

Creating an Google Cloud Spanner Data Object

Create a Google Cloud Spanner data object to add to a mapping.

1. Select a project or folder in the **Object Explorer** view.
2. Click **File > New > Data Object**.
3. Select **Google Cloud Spanner Data Object** and click **Next**.
The **Google Cloud Spanner Data Object** dialog box appears.
4. Enter a name for the data object.
5. Click **Browse** next to the **Location** option and select the target project or folder.
6. Click **Browse** next to the **Connection** option and select the Google Cloud Spanner connection from which you want to import the Google Cloud Spanner object.
7. To add a resource, click **Add** next to the **Selected Resources** option.
The **Add Resource** dialog box appears.
8. Select the checkbox next to the Google Cloud Spanner object you want to add and click **OK**.
9. Click **Finish**.
The data object appears under Data Objects in the project or folder in the **Object Explorer** view.

Creating a Google Cloud Spanner Data Object Operation

You can create the data object read, write, or lookup operation for Google Cloud Spanner data objects. You can add the Google Cloud Spanner data object operation to a mapping.

1. Select the data object in the **Object Explorer** view.
2. Right-click and select **New > Data Object Operation**.
The **Data Object Operation** dialog box appears.
3. Enter a name for the data object operation.
4. Select the type of data object operation. You can choose to create a read or write operation.
5. Click **Add**.
The **Select Resources** dialog box appears.
6. Select the Google Cloud Spanner data object for which you want to create the data object operation and click **OK**.
7. Click **Finish**.
The Developer tool creates the data object operation for the selected data object.

Rules and Guidelines for Google Cloud Spanner Mappings

Use the following rules and guidelines when you create a mapping:

- Ensure that the source and target table names contain only uppercase letters.
- You cannot use the OR operator in a filter condition.
- You must define a primary key in the target table. If you do not define a primary key in the target table, the mapping fails to delete the record or update the record in the target table.

CHAPTER 5

PowerExchange for Google Cloud Spanner Mappings

This chapter includes the following topics:

- [PowerExchange for Google Cloud Spanner Mappings Overview, 20](#)
- [Mapping Validation and Run-time Environments, 20](#)

PowerExchange for Google Cloud Spanner Mappings Overview

After you create a Google Cloud Spanner data object read or write operation, you can create a mapping to extract data from a Google Cloud Spanner source or load data to a Google Cloud Spanner target.

You can define properties in an operation to determine how the Data Integration Service must extract data from a Google Cloud Spanner source or load data to a Google Cloud Spanner target. You can extract data from one or more Google Cloud Spanner sources, and load data to one or more Google Cloud Spanner targets. When the Data Integration Service extracts data from the source or loads data to the target, it converts the data based on the data types associated with the source or the target.

Mapping Validation and Run-time Environments

You can validate and run mappings in the native environment or Spark engine.

The Data Integration Service validates whether the mapping can run in the selected environment. You must validate the mapping for an environment before you run the mapping in that environment.

Native environment

You can configure the mappings to run in the native or Hadoop environment. When you run mappings in the native environment, the Data Integration Service processes the mapping and runs the mapping from the Developer tool.

Spark Engine

When you run mappings on the Spark engine, the Data Integration Service pushes the mapping to a Hadoop cluster and processes the mapping on a Spark engine. The Data Integration Service generates an execution plan to run mappings on the Spark engine.

You can view the plan in the Developer tool before you run the mapping and in the Administrator tool after you run the mapping.

For more information about the Hadoop environment and Spark engines, see the *Informatica Data Engineering Integration User Guide*

CHAPTER 6

Google Cloud Spanner Run-Time Processing

This chapter includes the following topics:

- [Google Cloud Spanner Run-Time Processing Overview, 22](#)
- [Partitioning, 22](#)
- [Parameterization for Google Cloud Spanner Sources and Targets, 23](#)
- [Parameterization for Google Cloud Spanner Targets, 23](#)

Google Cloud Spanner Run-Time Processing Overview

When you create a Google Cloud Spanner data object read or write operation, you define properties that determine how the Data Integration Service reads data from or writes data to a Google Cloud Spanner database.

You can configure lookup caching, partitioning, and parameterization in the run-time properties.

Partitioning

When you read data from or write data to Google Cloud Spanner, you can configure partitioning to optimize the mapping performance at run time. You can configure partitioning for Google Cloud Spanner mappings that you run in the native or Spark engine.

Define the partition type as key range partitioning. To configure key range partitioning, open the Google Cloud Spanner data object read or write operation, and select the **Key Range** partition type option on the **Run-time** tab.

When you configure key range partitioning, the Data Integration Service distributes rows of data based on a port or set of ports that you define as the partition key. You can define a range of values for each port. The Data Integration Service uses the key and ranges to send rows to the appropriate partition.

Parameterization for Google Cloud Spanner Sources and Targets

You can parameterize the Google Cloud Spanner connection, data object read operation properties, and data object write operation properties to override the mapping properties at run time.

You can parameterize the following read operation properties for a Google Cloud Spanner source:

- Index Directive
- Batch Size
- Source Database
- Source Table

You can parameterize the following write operation properties for a Google Cloud Spanner target:

- Update Mode
- Target DataBase
- Target Table
- Batch Size

Parameterization for Google Cloud Spanner Targets

You can parameterize the Google Cloud Spanner connection and data object write operation properties to override the mapping properties at run time.

You can parameterize the following write operation properties for a Google Cloud Spanner target:

- Update Mode
- Target DataBase
- Target Table
- Batch Size

APPENDIX A

Google Cloud Spanner Data Type Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 24](#)
- [Google Cloud Spanner and Transformation Data Types, 24](#)

Data Type Reference Overview

Developer Tool uses the following data types in Google Cloud Spanner mappings:

- Google Cloud Spanner native data types. Google Cloud Spanner data types appear in Google Cloud Spanner definitions in a mapping.
- Transformation data types. Set of data types that appear in the transformations. They are internal data types based on ANSI SQL-92 generic data types, which the Data Integration Service uses to move data across platforms. They appear in all transformations in a mapping.

When the Data Integration Service reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When the Data Integration Service writes to a target, it converts the transformation data types to the comparable native data types.

Google Cloud Spanner and Transformation Data Types

The following table lists the Google Cloud Spanner data types that Developer Tool supports and the corresponding transformation data types:

Google Cloud Spanner Data Type	Transformation Data Type	Range and Description for the Transformation Data Type
BOOLEAN	String	True or False
BYTE	Binary	1 to 104,857,600 bytes

Google Cloud Spanner Data Type	Transformation Data Type	Range and Description for the Transformation Data Type
DATE	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
FLOAT64	Double	Precision 15
INT64	BigInt	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
STRING	String	1 to 104,857,600 characters
TIMESTAMP	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)

Note: PowerExchange for Google Cloud Spanner does not support the Array data type.

INDEX

A

advanced properties
input [17](#)

C

create
 data object operation
 create [18](#)
 Google Cloud Spanner data object [18](#)
creating
 Google Cloud Spanner connection [12](#)

D

data types [24](#)

G

general properties
 input [16](#)
Google Cloud Spanner
 data object properties [13](#)
 data object read operation [14](#)
 data object write operation [16](#)
 introduction [7](#)
Google Cloud Spanner connection
 properties [11](#)
Google Cloud Spanner connections
 creating [12](#)
 overview [11](#)
Google Cloud Spanner data object
 create [18](#)
 overview [13](#)
Google Cloud Spanner data types [24](#)
Google Cloud Spanner parameterization
 for sources [23](#)

Google Cloud Spanner parameterization (*continued*)
 for targets [23](#)
Google Cloud Spanner run-time processing
 parameterization [23](#)
 partitioning [22](#)

I

input properties [16](#)

N

native environment
 mappings [20](#)

P

PowerExchange for Google Cloud Spanner
 administration [9](#)
 configuration [8](#)
 example [7](#)
 installation [8](#)
 overview [7](#)
PowerExchange for Google Cloud Spanner mappings
 overview [20](#)

S

Spark engine
 mappings [20](#)

T

transformation data types [24](#)