



Informatica®

10.4.0

Glossary

© Copyright Informatica LLC 2016, 2020

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 and the Informatica logo 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.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://curl.haxx.se/docs/copyright.html>. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies. The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.dom4j.org/license.html>.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://dojotoolkit.org/license>.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at <http://www.opensource.org/licenses/mit-license.php>.

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.

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

Publication Date: 2020-06-26

Table of Contents

Preface 4

Informatica Resources. 4

 Informatica Network. 4

 Informatica Knowledge Base. 4

 Informatica Documentation. 4

 Informatica Product Availability Matrices. 5

 Informatica Velocity. 5

 Informatica Marketplace. 5

 Informatica Global Customer Support. 5

Appendix A: Glossary..... 6

Preface

Refer to the *Informatica® Glossary* to learn about the terminology for Informatica products that you can use with Informatica Administrator, Informatica Analyst, and Informatica Developer.

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.

APPENDIX A

Glossary

application

An object that you can deploy to a Data Integration Service. It can contain data objects, mappings, SQL data services, web services, and workflows.

application patch

An object that you can deploy to a Data Integration Service. It contains the metadata that you can use to update a deployed incremental application.

application service

A service that runs on one or more nodes in the Informatica domain. You create and manage application services in Informatica Administrator or through the infacmd command program. Application services include services that can have multiple instances in the domain and system services that can have a single instance in the domain. Configure each application service based on your environment requirements.

big data

A set of data that is so large and complex that it cannot be processed through standard database management tools.

Blaze executor

A component of the DTM that can simplify and convert a mapping to a Blaze execution plan that runs on a Hadoop cluster.

candidate key

A column or a set of columns that uniquely identifies each source row in a database table.

Cloudera's Distribution Including Apache Hadoop (CDH)

Cloudera's version of the open-source Hadoop software framework.

column name rule

Reusable business logic that identifies a column by its name as belonging to a particular data domain.

column profile

A type of profile that determines the characteristics of columns in a data source, such as value frequency, percentages, patterns, and data types.

command task

The pre-processing or post-processing task for local data for a Blaze engine workflow.

complex data type

A data type that allows you to represent multiple data values in a single column position. The data values are called elements.

complex data type definition

A reusable representation of the schema of the data that you reference in a struct port or in a complex port that contains elements of type struct. One or more complex ports can use the complex data type definition.

complex function

A type of pre-defined function in which the input or the return type is of complex data type.

complex operator

A type of operator to refer to element names or to access elements in a complex data type.

complex port

A port type that is assigned a complex data type such as an array, struct, or map to pass hierarchical data.

CompressionCodec

Hadoop compression interface. A codec is the implementation of a compression-decompression algorithm. In Hadoop, a codec is represented by an implementation of the CompressionCodec interface.

conditional sequence flow

A sequence flow that includes an expression that the Data Integration Service evaluates to true or false. If the expression evaluates to true, the Data Integration Service runs the next object in the workflow. If the expression evaluates to false, the Data Integration Service does not run the next object in the workflow.

container

An allocation of memory and CPU resources on a node with the compute role. An application service uses the container to remotely perform computations on the node. For example, a Data Integration Service that runs on a grid can remotely run a mapping within a container on a node with the compute role.

cost-based optimization

Optimization method that reduces the run time for mappings that perform join operations. With cost-based optimization, the Data Integration Service creates different plans to run a mapping and calculates a cost for each plan. The Data Integration Service runs the plan with the smallest cost. The Data Integration Service calculates cost based on database statistics, I/O, CPU, network, and memory.

curation

The process of validating and managing discovered metadata of a data source so that the metadata is fit for use and reporting.

customized data object

A physical data object that uses one or more related relational resources or relational data objects as sources. You use a customized data object to perform tasks such as join data from related resources or filter rows. Customized data object uses a single connection and SQL statement for the source tables.

data domain

A predefined or user-defined Model repository object that represents the functional meaning of a column based on the column data or column name. Examples are Social Security number, credit card number, and email ID.

data domain discovery

The process that identifies all the data domains associated with a column based on column values or name.

data domain glossary

A container for all data domains and data domain groups in the Analyst tool or Developer tool.

data domain group

A collection of data domains under a specific data domain category.

Data Integration Service

An application service that performs data integration jobs for Informatica Analyst, Informatica Developer, and external clients. Data integration jobs include previewing data and running mappings, profiles, SQL data services, web services, and workflows.

DataNode

An HDFS node that stores data in the Hadoop File System. An HDFS cluster can have more than one DataNode, with data replicated across them.

data object profile

A repository object that defines the type of analysis you perform on a data source.

data object read operation

Repository object that contains properties required to perform certain run-time operations on sources. A data object read operation is associated with a source data object.

data object write operation

Repository object that contains properties required to perform certain run-time operations on targets. A data object write operation is associated with a target data object.

Data Processor event

An occurrence during the execution of a Data Processor transformation.

data rule

Reusable business logic that identifies a column by its values as belonging to a particular data domain.

data service

A collection of reusable operations that you can run to access and transform data. A data service provides a unified model of data you can access through a web service or run an SQL query against.

default sequence flow

The outgoing sequence flow from an Exclusive gateway that always evaluates to true. When all other conditional sequence flows evaluate to false, the Data Integration Service runs the object connected to the default outgoing sequence flow.

dependent column

In a functional dependency, the column containing values that are determined by a determinant column.

deploy

To make objects within an application accessible to end users. Depending on the types of objects in the application, end users can then run queries against the objects, access web services, run mappings, or run workflows.

design-time application

An application that you edit it in the Developer tool. The application contains design-time instances of application objects.

determinant column

In a functional dependency, a set of columns that determines the value of the dependent column. If the determinant has zero columns, the dependent is a constant.

direct match

In a global search, a direct match is an asset that matches the entire search query. In discovery search, a direct match is a match with some or all the metadata of the asset that matches the search query.

discovery search

A type of search in the Analyst tool that identifies assets based on direct matches to the search query as well as relationships to other objects that match the search query.

documented key

A declared primary key in the source database.

document processor

A component that operates on a document as a whole, typically performing preliminary conversions before parsing.

DTM instance

A specific, logical representation of the execution Data Transformation Manager (DTM) that the Data Integration Service creates to run a job. Based on how you configure the Data Integration Service, DTM instances can run in the Data Integration Service process, in a separate DTM process on the local node, or in a separate DTM process on a remote node.

DTM process

An operating system process that the Data Integration Service starts to run DTM instances. Based on how you configure the Data Integration Service, the service can run each DTM instance in a separate DTM process on the local or on a remote node.

dynamic email address

An email address defined in a workflow parameter or variable.

dynamic email content

Email content defined in a workflow parameter or variable.

dynamic mapping

A mapping in which you can change sources, targets, and transformation logic at run time based on parameters and rules that you define. You can configure dynamic mappings to allow metadata changes to sources and targets. You can determine which ports a transformation receives, which ports to use in the transformation logic, and which links to establish between transformation groups.

dynamic port

A port that can receive one or more columns from an upstream transformation and create a generated port for each column.

dynamic recipient

A notification recipient defined in a workflow parameter or variable.

dynamic source

A flat file or relational source for a mapping that can change at run time. Read and Lookup transformations can get definition or metadata changes directly from the source. If you use a parameter for the source, you can change the source at run time.

dynamic target

A flat file or relational target for a mapping that can change at run time. Write transformations can define target columns at run time based on the mapping flow or from an associated target. Write transformations can also drop and replace the target table at run time.

early projection optimization

Optimization method that reduces the amount of data that moves between transformations in the mapping. With early projection optimization, the Data Integration Service identifies unused ports and removes the links between the ports in a mapping.

early selection optimization

Optimization method that reduces the number of rows that pass through the mapping. With early selection optimization, the Data Integration Service moves filters closer to the mapping source in the pipeline.

enterprise discovery

The process that finds column profile statistics, data domains, primary keys, and foreign keys in a large number of data sources spread across multiple connections or schemas.

enterprise discovery profile

A profile type that you use to perform enterprise discovery.

event

A workflow object that starts or ends the workflow. An event represents something that happens when the workflow runs. The editor displays events as circles.

example source document

A sample of the documents that a Data Processor transformation processes.

Exclusive gateway

A gateway that represents a decision made in a workflow. When an Exclusive gateway splits the workflow, the Data Integration Service makes a decision to take one of the outgoing branches. When an Exclusive gateway merges the workflow, the Data Integration Service waits for one incoming branch to complete before triggering the outgoing branch.

execution Data Transformation Manager (DTM)

The compute component of the Data Integration Service that extracts, transforms, and loads data to complete a data transformation job.

folder

A container for objects in the Model repository. Use folders to organize objects in a project and create folders to group objects based on business needs.

foreign key discovery

The process that finds columns in one data source that matches the primary key columns in the parent data source.

full application

An application that you must redeploy to the Data Integration Service after editing application objects.

functional dependency

The relationship between a set of columns in a given table, in which the determinant column functionally determines the dependent column.

functional dependency discovery

The process that finds functional dependency relationships between columns in a data source.

gateway

A workflow object that splits and merges paths in the workflow based on how the Data Integration Service evaluates expressions in conditional sequence flows. The editor displays gateways as diamonds.

generated port

A port within a dynamic port that represents a single column. The Developer tool generates ports based on one or more input rules.

grid mapping

An Informatica mapping that the Blaze engine compiles and distributes across a cluster of nodes.

grid segment

Part of a grid mapping that is contained in a grid task.

grid task

A parallel processing job request. When the mapping runs in the Hadoop environment, the Blaze engine executor sends the request to the Grid Manager. When the mapping runs in the native environment and the Data Integration Service runs in remote mode, the Data Integration Service process sends the request to the Service Manager on the master compute node.

Hadoop cluster

A cluster of machines that is configured to run Hadoop applications and services. A typical Hadoop cluster includes a master node and several worker nodes. The master node runs the master daemons JobTracker and NameNode. A slave or worker node runs the DataNode and TaskTracker daemons. In small clusters, the master node may also run the slave daemons.

Hadoop Distributed File System (HDFS)

A distributed file storage system used by Hadoop applications.

Hadoop environment

An environment that you can configure to run a mapping or a profile on a Hadoop Cluster. You must configure Hadoop as the validation and run-time environment.

hierarchical data

A set of data that is hierarchically related. The hierarchical relationship is represented as schema. Informatica transformations use complex data types to represent hierarchical data.

Hive

A data warehouse infrastructure built on top of Hadoop. Hive supports an SQL-like language called HiveQL for data summarization, query, and analysis.

Hive execution plan

A series of Hive tasks that the Hive executor generates after it processes a mapping or a profile. A Hive execution plan can also be referred to as a Hive workflow.

Hive executor

A component of the DTM that can simplify and convert a mapping or a profile to a Hive execution plan that runs on a Hadoop cluster.

Hive scripts

Script in Hive query language that contain Hive queries and Hive commands to run the mapping.

Hive task

A task in the Hive execution plan. A Hive execution plan contains many Hive tasks. A Hive task contains a Hive script.

incremental application

An application that you can update by deploying an application patch to the Data Integration Service.

indirect match

A match in discovery search results that is linked to the asset that directly matches some or all of the search query.

inferred key

A column or a set of columns that the Analyst tool or Developer tool infers as a candidate key based on column data.

Informatica Administrator

Informatica Administrator (the Administrator tool) is an application that consolidates the administrative tasks for domain objects such as services, nodes, licenses, and grids. You manage the domain and the security of the domain through the Administrator tool.

Informatica Developer

Informatica Developer (the Developer tool) is an application that you use to design data integration solutions. The Model repository stores the objects that you create in the Developer tool.

Informatica Mass Ingestion

Informatica Mass Ingestion (the Mass Ingestion tool) is an application that you can use to configure, deploy, run, and monitor mass ingestion specifications.

Informatica Monitoring tool

Informatica Monitoring tool (the Monitoring tool) is an application that provides a direct link to the Monitor tab of the Administrator tool. The Monitor tab shows properties, run-time statistics, and run-time reports about the integration objects that run on a Data Integration Service.

input rule

A rule that determines which generated ports to create within a dynamic port.

intelligent structure

A model of the pattern and types of fields of data that Intelligent Structure Discovery identifies in an input file. The model includes both simple elements and complex elements, depending on the organization of data in the file.

Intelligent Structure Discovery

A service in Informatica Intelligent Cloud Services that uses machine learning algorithms to automatically identify the types of information in an input file. Intelligent Structure Discovery creates an intelligent structure, a model of the pattern and types of fields of data discovered in the file. The model includes both simple elements and complex elements, depending on the organization of data in the file.

JobTracker

A Hadoop service that coordinates map and reduce tasks and schedules them to run on TaskTrackers.

join profile

A type of profile that determines the degree of overlap between a set of one or more columns in one data source and a similar set in the same or a different data source.

logical data object

An object that describes a logical entity in an organization. It has attributes and keys, and it describes relationships between attributes.

logical data object mapping

A mapping that links a logical data object to one or more physical data objects. It can include transformation logic.

logical data object model

A data model that describes data in an organization and the relationship between the data. It contains logical data objects and defines relationships between them.

logical data object read mapping

A mapping that provides a view of data through a logical data object. It contains one or more physical data objects as sources and a logical data object as the mapping output.

logical data object write mapping

A mapping that writes data to targets using a logical data object as input. It contains one or more logical data objects as input and a physical data object as the target.

logical Data Transformation Manager (LDTM)

A service component of the Data Integration Service that optimizes and compiles jobs, and then sends the jobs to the execution Data Transformation Manager (DTM).

mapping

A set of inputs and outputs linked by transformation objects that define the rules for data transformation.

mapplet

A reusable object that contains a set of transformations that you can use in multiple mappings or validate as a rule.

MapReduce

A programming model for processing large volumes of data in parallel.

MapReduce job

A unit of work that consists of the input data, the MapReduce program, and configuration information. Hadoop runs the MapReduce job by dividing it into map tasks and reduce tasks.

mass ingestion

The movement of large amounts of data between a relational database and a data lake or a Hadoop cluster.

Mass Ingestion Service

An application service that manages the mass ingestion specifications that you create in the Mass Ingestion tool. The Mass Ingestion Service validates specifications, schedules specifications to run on a Data Integration Service, and monitors the ingestion statistics.

mass ingestion specification

A configuration that determines how source data from a relational database is ingested to a specific location in a data lake or a Hadoop cluster.

metastore

A database that Hive uses to store metadata of the Hive tables stored in HDFS. Metastores can be local, embedded, or remote.

metric

A column of a data source or output of a rule that is part of a scorecard.

metric group

A user-defined group of metrics.

metric group score

The computed weighted average of all the metric scores in the metric group.

metric score

The percentage of valid values in a metric.

metric weight

An integer greater than or equal to 0 assigned to a metric. A metric weight defines the contribution of the metric to the metric group score.

Model Repository Service

An application service in the Informatica domain that runs and manages the Model repository. The Model repository stores metadata created by Informatica products in a relational database to enable collaboration among the products.

NameNode

A node in the Hadoop cluster that manages the file system namespace, maintains the file system tree, and the metadata for all the files and directories in the tree.

native environment

The default environment in the Informatica domain that runs a mapping, a workflow, or a profile. The Data Integration Service performs data extraction, transformation, and loading. Alternatively, the Data Integration Service might push processing to a non-native environment, which is a distributed cluster outside of the Informatica domain, such as Hadoop or Databricks.

nested complex port

A complex port that contains a nested complex data type definition.

nested data type

A complex data type that contains at least one element of a complex data type. For example, a struct data type that contains an element of type array.

nested data type definition

A complex data type definition that references other complex data type definitions.

node

A representation of a level in the hierarchy of a web service message.

node role

The purpose of a node. A node with the service role can run application services. A node with the compute role can perform computations requested by remote application services. A node with both roles can run application services and locally perform computations for those services.

non-native environment

A distributed cluster outside of the Informatica domain, such as Hadoop or Databricks, where the Data Integration Service can push run-time processing. If the Data Integration Service does not push processing to a non-native environment, it processes the job in the native environment of the Informatica domain.

operating system profile

A type of security that the Data Integration Services on UNIX or Linux uses to isolate the run-time user environment. The operating system profile contains the operating system user name, service process variables, environment variables, and permissions. The Data Integration Service runs mappings, workflows, profiling jobs, and scorecards with the system permissions of the operating system user and the properties defined in the operating system profile.

operation mapping

A mapping that performs the web service operation for the web service client. An operation mapping can contain an Input transformation, an Output transformation, and multiple Fault transformations.

Outlier

An outlier is a pattern, value, or frequency for a column in the profile results that does not fall within an expected range of values.

output document

A document that is the result of a Data Processor transformation.

partitioning

The process of dividing the underlying data into subsets that can run in multiple processing threads. When administrators enable the Data Integration Service to maximize parallelism, the service increases the number of processing threads, which can optimize mapping and profiling performance.

partition point

A boundary between stages in a mapping pipeline. When partitioning is enabled, the Data Integration Service can redistribute rows of data at partition points.

physical data object

A physical representation of data that is used to read from, look up, or write to resources.

pipeline

A source and all the transformations and targets that receive data from that source. Each mapping contains one or more pipelines.

predicate expression

An expression that filters the data in a mapping. A predicate expression returns true or false.

predicate optimization

Optimization method that simplifies or rewrites the predicate expressions in a mapping. With predicate optimization, the Data Integration Service attempts to apply predicate expressions as early as possible to increase mapping performance.

preprocessor

A document processor used to perform an overall modification of a source document, before the main transformation.

primary key discovery

The process to identify a column or combination of columns that uniquely identify a row in a data source.

primitive data type

A data type that allows you to represent a single data value in a single column position.

profile

An object that contains rules to discover patterns in source data. Run a profile to evaluate the data structure and verify that data columns contain the type of information that you expect.

profiling warehouse

A relational database that stores profiling information, such as profile results and scorecard results.

project

The top-level container to store objects created in Informatica Analyst and Informatica Developer. Create projects based on business goals or requirements. Projects appear in both Informatica Analyst and Informatica Developer.

pushdown optimization

Optimization method that pushes transformation logic to a source or target database. With pushdown optimization, the Data Integration Service translates the transformation logic into SQL queries and sends the SQL queries to the database. The database runs the SQL queries to process the data.

recipient

A user or group in the Informatica domain that receives a notification during a workflow.

recursive data type definition

A nested data type definition where one of the complex data type definitions at any level is the same as that of a parent.

Resource Manager Service

A system service that manages computing resources in the domain and dispatches jobs to achieve optimal performance and scalability. The Resource Manager Service collects information about nodes with the compute role. The service matches job requirements with resource availability to identify the best compute node to run the job. The Resource Manager Service communicates with compute nodes in a Data Integration Service grid. Enable the Resource Manager Service when you configure a Data Integration Service grid to run jobs in separate remote processes.

result set caching

A cache that contains the results of each SQL data service query or web service request. With result set caching, the Data Integration Service returns cached results when users run identical queries. Result set caching decreases the run time for identical queries.

rule

Reusable business logic that defines conditions applied to source data when you run a profile. Use rules to further validate the data in a profile and to measure data quality progress. You can create a rule in Informatica Analyst or Informatica Developer.

run-time application

An application that is deployed on a Data Integration Service and contains run-time instances of application objects.

run-time environment

The environment you configure to run a mapping or a profile. The run-time environment can be native or Hive.

run-time link

A group-to-group link that uses a policy or a parameter or both to determine which ports to link between the groups at run time.

schema

A definition of the structure of data. Complex ports of struct data type use complex data type definitions to represent schema.

scorecard

A graphical representation of valid values for a source column or output of a rule in profile results. Use scorecards to measure data quality progress.

scorecard lineage

A diagram that shows the origin of data, describes the path, and shows how the data flows for a metric or metric group in a scorecard. In the scorecard lineage analysis, boxes or nodes represent objects. Arrows represent data flow relationships.

semi-join optimization

Optimization method that reduces the number of rows extracted from the source. With semi-join optimization, the Data Integration Service modifies the join operations in a mapping. The Data Integration Service applies the semi-join optimization method to a Joiner transformation when a larger input group has rows that do not match a smaller input group in the join condition. The Data Integration Service reads the rows from the smaller group, finds the matching rows in the larger group, and performs the join operation.

sequence flow

A connector between workflow objects that specifies the order that the Data Integration Service runs the objects. The editor displays sequence flows as arrows.

source document

A document that is the input of a Data Processor transformation.

Sparkline

A sparkline is a line chart that displays the variation in a null value, unique value, or non-unique value across the latest five consecutive profile runs.

SQL data service

A virtual database that you can query. It contains virtual objects and provides a uniform view of data from disparate, heterogeneous data sources.

SQL Service Module

The component service in the Data Integration Service that manages SQL queries sent to an SQL data service from a third-party client tool.

startup component

The runnable component that Data Transformation starts first when it runs a Data Processor Transformation.

stateful variable port

A variable port that refers to values from previous rows.

system service

An application service that can have a single instance in the domain. When you create the domain, the system services are created for you. You can enable, disable, and configure system services.

system workflow variable

A workflow variable that returns system run-time information such as the workflow instance ID, the user who started the workflow, or the workflow start time.

task

A workflow object that runs a single unit of work in the workflow, such as running a mapping, sending an email, or running a shell command. A task represents something that is performed during the workflow. The editor displays tasks as squares.

task input

Data that passes into a task from workflow parameters and variables. The task uses the input data to complete a unit of work.

tasklet

A partition of a grid segment that runs on a separate DTM.

task output

Data that passes from a task into workflow variables. When you configure a task, you specify the task output values that you want to assign to workflow variables. The Data Integration Service copies the task output values to workflow variables when the task completes. The Data Integration Service can access these values from the workflow variables when it evaluates expressions in conditional sequence flows and when it runs additional objects in the workflow.

TaskTracker

A node in the Hadoop cluster that runs tasks such as map or reduce tasks. TaskTrackers send progress reports to the JobTracker.

team-based development

The collaboration of team members on a development project. Collaboration includes functionality such as versioning through checking out and checking in repository objects.

transformation

A repository object in a mapping that generates, modifies, or passes data. Each transformation performs a different function.

type configuration

A set of complex port properties that specify the data type of the complex data type elements or the schema of the data.

type definition library

An object in the Model repository that stores complex data type definitions for a mapping or a mapplet.

user-defined workflow variable

A workflow variable that captures task output or captures criteria that you specify. After you create a user-defined workflow variable, you configure the workflow to assign a run-time value to the variable.

user role

A collection of privileges that you assign to a user or group. You assign roles to users and groups for the domain and for some application services in the domain.

validation environment

The environment you configure to validate a mapping or a profile. You validate a mapping or a profile to ensure that it can run in a run-time environment. The validation environment can be Hive, native, or both.

virtual data

The information get when you query virtual tables or run stored procedures in an SQL data service.

virtual database

An SQL data service that you can query. It contains virtual objects and provides a uniform view of data from disparate, heterogeneous data sources.

virtual schema

A schema in a virtual database that defines the database structure.

virtual stored procedure

A set of procedural or data flow instructions in an SQL data service.

virtual table

A table in a virtual database.

virtual table mapping

A mapping that contains a virtual table as a target.

virtual view of data

A virtual database defined by an SQL data service that you can query as if it were a physical database.

Web Service Module

A component in the Data Integration Service that manages web service operation requests sent to a web service from a web service client.

web service transformation

A transformation that processes web service requests or web service responses. Examples of web service transformations include an Input transformation, Output transformation, Fault transformation, and the Web Service Consumer transformation.

workflow

A graphical representation of a set of events, tasks, and decisions that define a business process. You use the Developer tool to add objects to a workflow and to connect the objects with sequence flows. The Data Integration Service uses the instructions configured in the workflow to run the objects.

workflow instance

The run-time representation of a workflow. When you run a workflow from a deployed application, you run an instance of the workflow. You can concurrently run multiple instances of the same workflow.

workflow instance ID

A number that uniquely identifies a workflow instance that has run.

workflow parameter

A constant value that you define before the workflow runs. Parameters retain the same value throughout the workflow run. You define the value of the parameter in a parameter file. All workflow parameters are user-defined.

Workflow Service Module

A component in the Data Integration Service that manages requests to run workflows.

workflow variable

A value that can change during a workflow run. Use workflow variables to reference values and record run-time information. You can use system or user-defined workflow variables.

XMap

A Data Processor transformation object that maps an XML input document to another XML document.

XML schema

A definition of the elements, attributes, and structure used in XML documents. The schema conforms to the World Wide Web Consortium XML Schema standard, and it is stored as an *.xsd file.

XPath

A query language used to select nodes in an XML document and perform computations.

XSD schema file

An *.xsd file containing an XML schema that defines the elements, attributes, and structure of XML documents.