



Informatica ETL

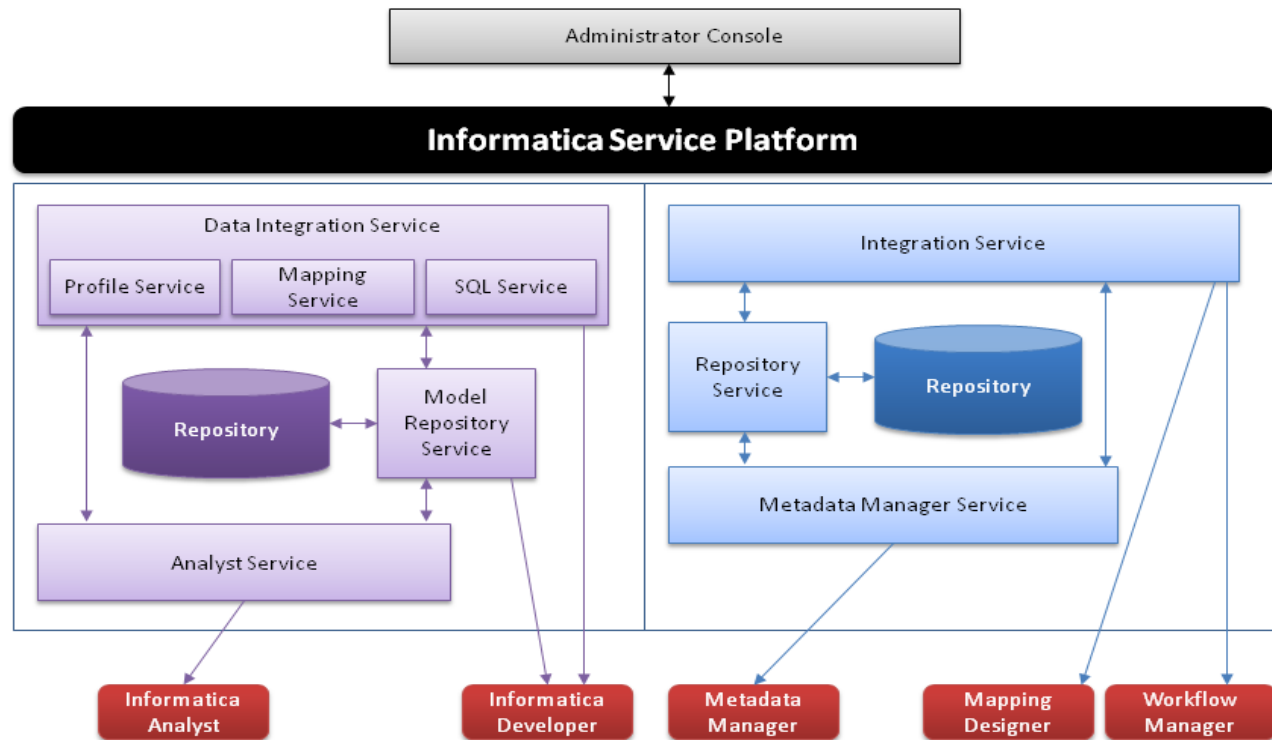
Presented by : Prashant Tikone

This presentation is the intellectual property of Cybage Software Pvt. Ltd. and is meant for the usage of the intended Cybage employee/s for training purpose only. This should not be used for any other purpose or reproduced in any other form without written permission and consent of the concerned authorities.

Agenda

Introduction to Informatica Platform
Power Center Overview
Power Center Express Architecture
Source & Targets
Mappings
Transformations
Profiling
Workflows
Monitoring
Demos

Platform Overview



Informatica Power Center

Description	PC Express Personal	PC Express Paid	PC Enterprise Editions
Data Integration Design and Development	X	X	X
Data Profiling	Column-level only	Column-level only	Full profiling
Data Quality			X
Number of users	1	Upto 5 max	Unlimited
Data Volume	250K rows / day	Unlimited	Unlimited
Simultaneous jobs	1	1	Unlimited
Connectivity	Basic (Relational, FF, XML, Unstructured, Social Media)	Basic (Relational, FF, XML, Unstructured, Social Media)	Advanced (Including SAP, CDC, Mainframes, SFDC, Hadoop, Applications, Messaging)
Central Administration and Management	X	X	X
Built-in Performance Optimizer	X	X	X
Parallel Processing (Partitioning, Grid, Load Balancer)			X
High Availability			X
Real-time processing			X
Data Virtualization			X
Metadata Manager			X
Versioning			X
Data Validation for Testing			X
Reporting			X
Mapping Templates			X
Big data processing			X
On-line Knowledge Base	X	X	X
Support	On-line, community-based	Personal web and email-based, in addition to on-line, community-based support	Enterprise-class 24x7, premium support

Power Center Express

Used to design and implement data integration solutions.

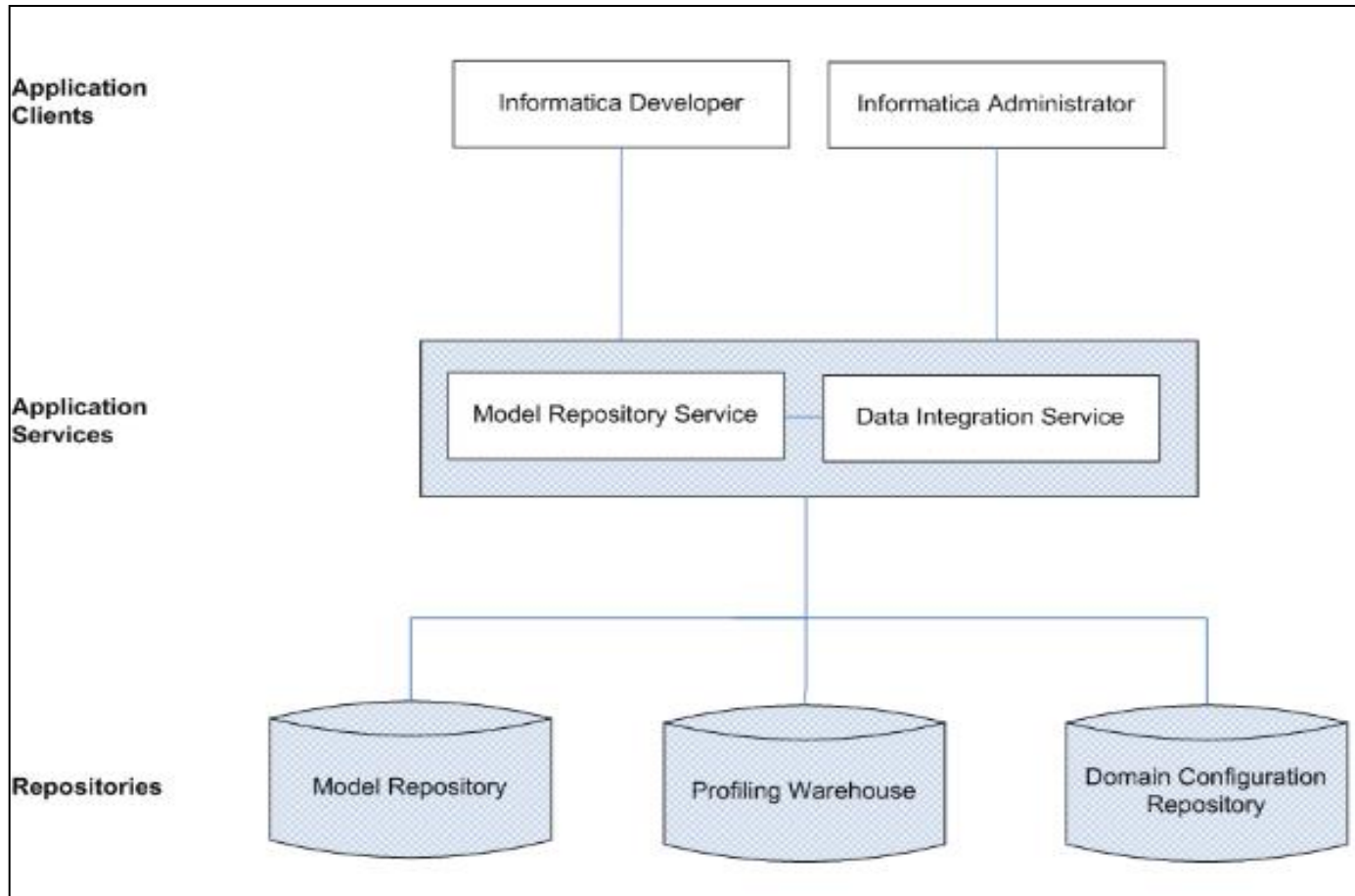
Extract data from multiple sources, transform the data according to business logic that you build in the client application, and load the transformed data to targets.

Profile to analyze the structure and content of your data, and to determine the quality of your data.

Access data in relational databases, flat files, web services, and social media web sites.

- Powercenter Express includes the
 - Informatica domain,
 - Application services
 - Repositories to store metadata,
 - Informatica Administrator
 - Informatica Developer

Architecture



Architecture

Application Clients

Informatica Developer – Its Developer tool that developers can use to design and implement data integration and data federation solutions.

Informatica Administrator –

- Administrator tool that consolidates the administrative tasks for domain objects such as services, connections, and licenses.
- Administrators manage the domain and the security of the domain through the Administrator tool.

Architecture

Application Services

Data Integration Service -

- Runs data integration jobs for the Developer tool and external clients.
- Previewing data and running profiles, mappings, and workflows.

Model Repository Service - Application service that manages the Model repository.

Architecture

Repositories

Repositories are a group of relational databases that store metadata about objects and processes required to handle user requests from application clients.

Model repository –

It is a relational database that stores the metadata for projects. The Model repository also stores run-time and configuration information for applications that are deployed to a Data Integration Service.

Domain configuration repository –

The domain configuration repository is a set of domain metadata tables stored in a relational database. Each time an administrator makes a change to the domain, the Service Manager writes the change to the domain configuration repository.

Profile warehouse –

The profile warehouse is a relational database that the Data Integration Services uses to store profile results.

Data Integration Process

Create connections

Create connections to access data from relational databases, third-party web services, or social media web sites.

Import metadata to create data objects

Import metadata to create data objects for sources and targets that you want to use in a mapping.

Run a profile

Run a profile to analyze the structure and content of your data, and to determine the quality of your data.

Develop mappings

A mapping is a set of inputs and outputs that represent the data flow between sources and targets. Link the sources and targets with transformation objects that define the rules for data transformation.

Data Integration Process

Create and run workflows

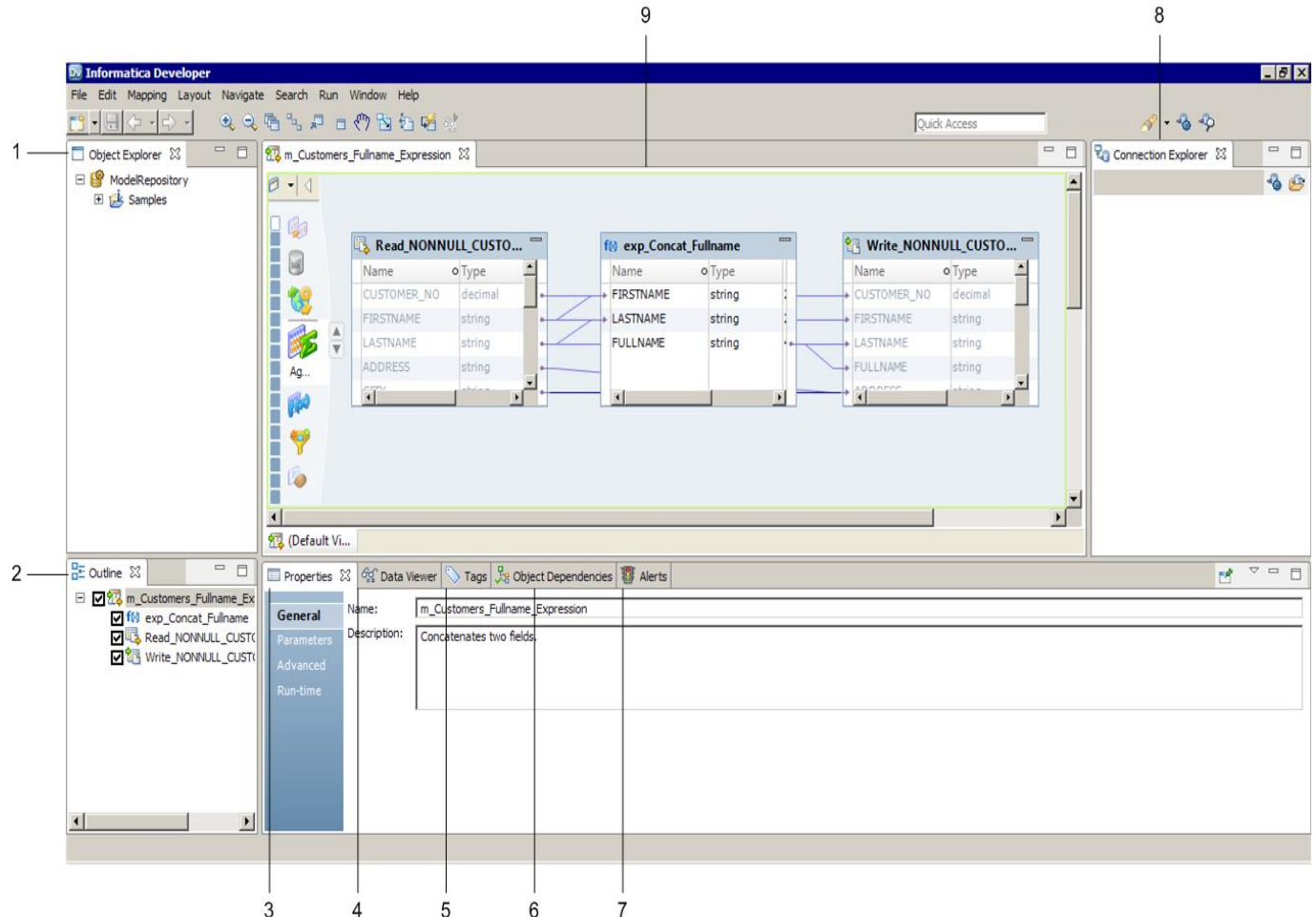
Create a workflow to define a sequence of events, tasks, and decisions based on a business process. Then, deploy the workflow to the Data Integration Service and run the workflow.

Monitor workflows

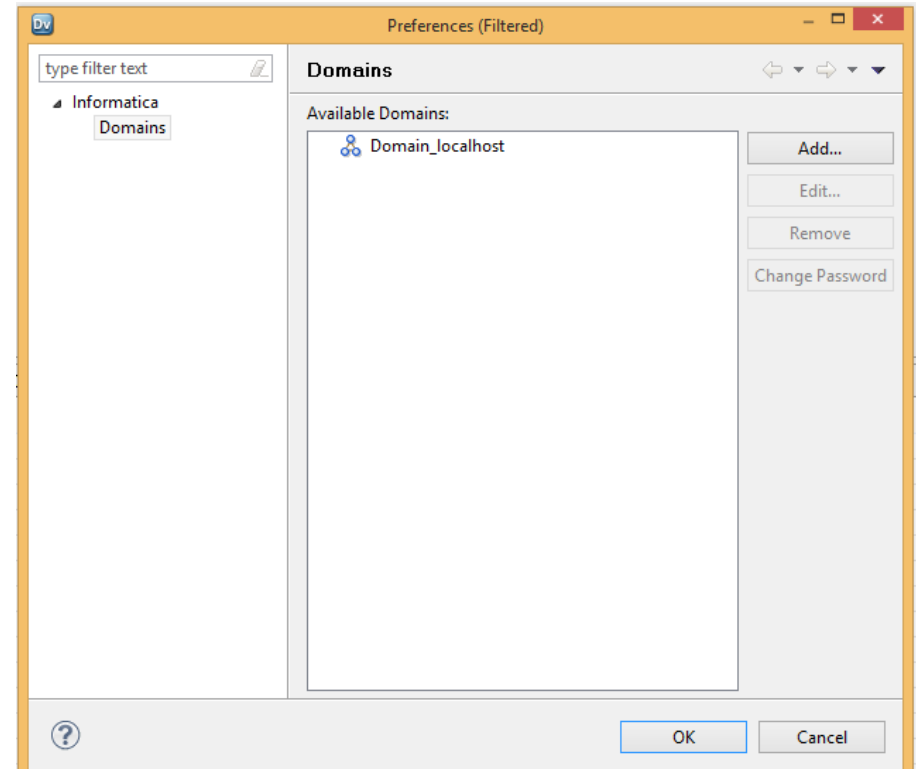
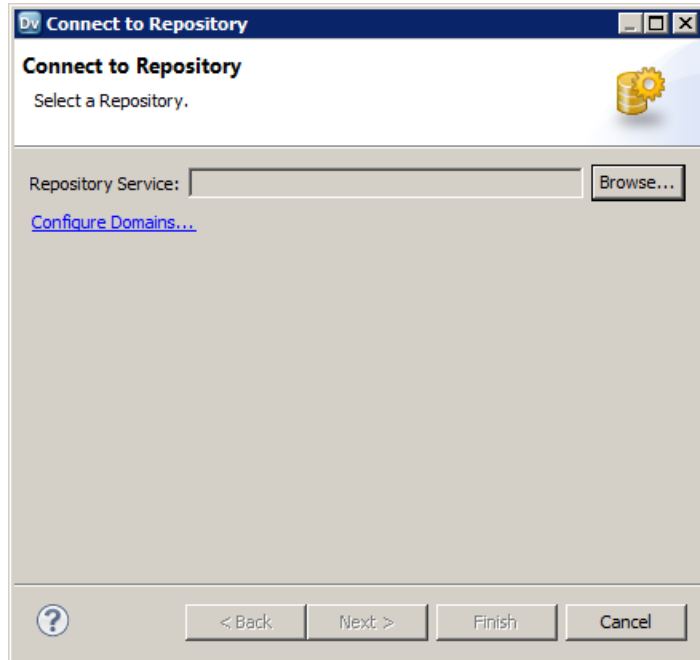
- Monitor the workflow instance on the Monitoring tab of the Administrator tool.
- Get status of running workflow and workflow object instances.
- **Abort or cancel** a running workflow instance in the Monitoring tool.
- **View logs for workflow** instances and to view workflow reports

User Interface

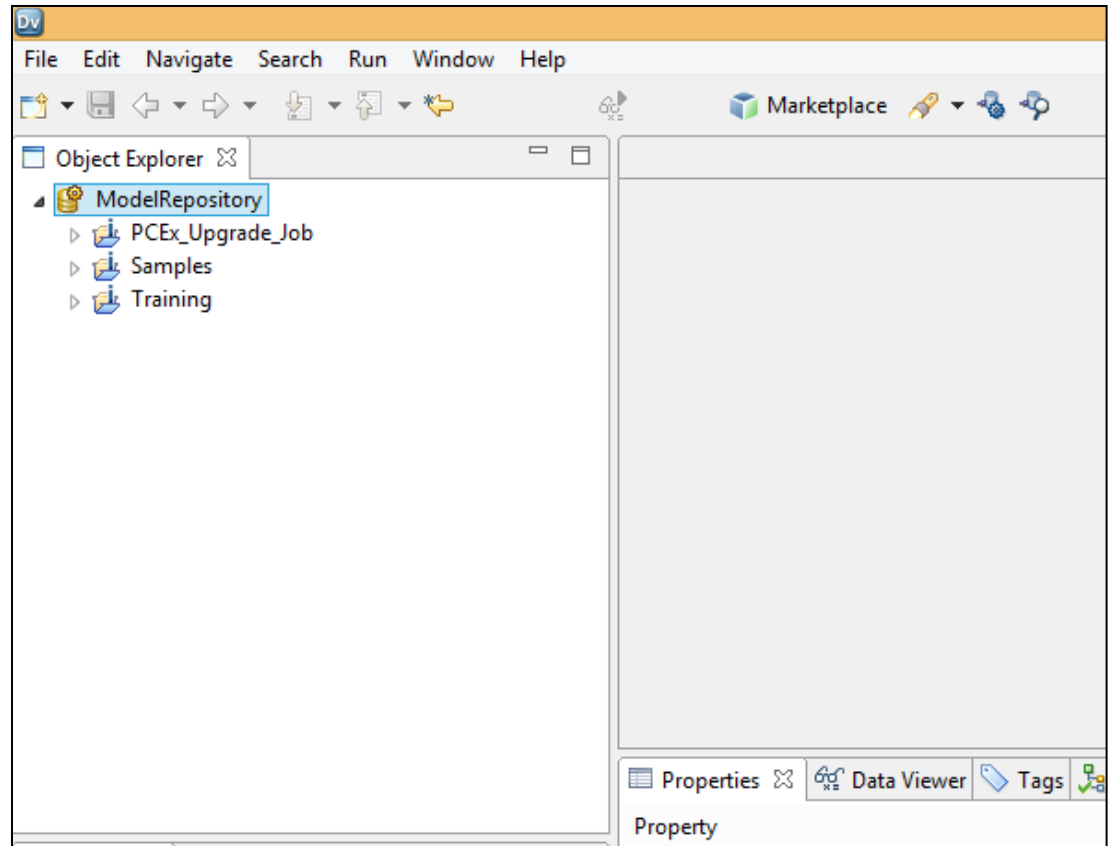
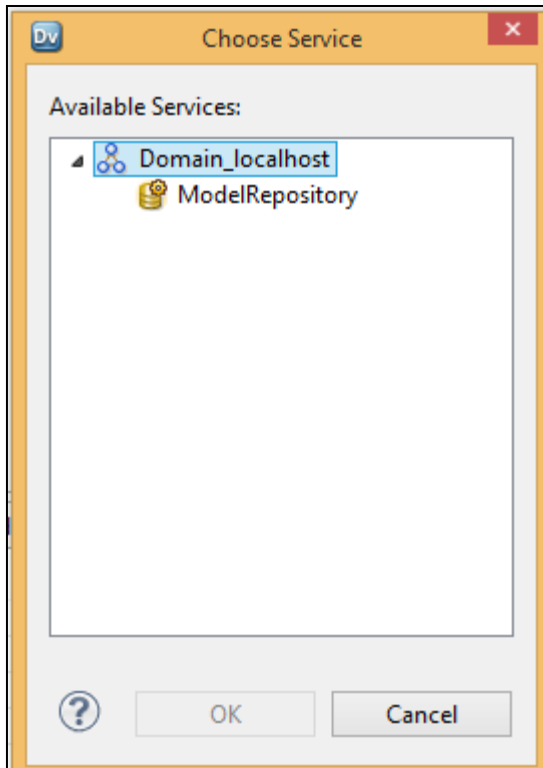
1. Object Explorer view
2. Outline view
3. Properties view
4. Data Viewer view
5. Tags view
6. Object dependencies view
7. Alerts view
8. Connection Explorer view
9. Editor



Adding Domain



Connecting Repository



Model Repository

- The Model repository is a relational database that stores the metadata for projects and folders.
- **Application** - Deployable object that can contain data objects, mappings, and workflows.
- **Folder** - 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.
- **Logical data object** - Object in a logical data object model that describes a logical entity in an enterprise. It has attributes and keys, and it describes relationships between attributes.
- **Logical data object mapping** - Mapping that links a logical data object to one or more physical data objects.

Model Repository

Logical data object model - A data model that contains logical data objects and defines relationships between them.

Mapping - Set of inputs and outputs linked by transformation objects that define the rules for data transformation.

Mapplet - Reusable object that contains a set of transformations that you can use in multiple mappings or validate as a rule.

Physical data object - Physical representation of data that is used to read from, look up, or write to resources.

Model Repository

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.

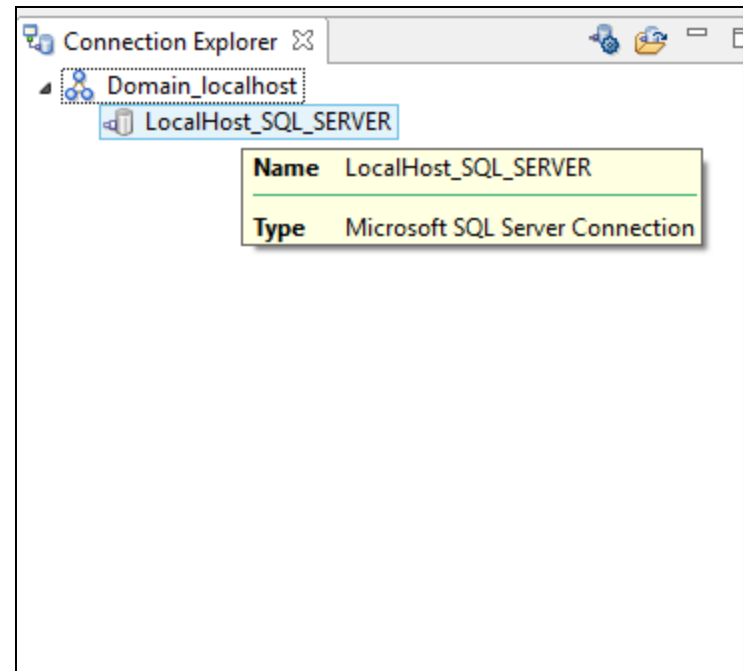
Rule - Business logic that defines conditions applied to source data when you run a profile. It is a midstream mapplet that you use in a profile.

Transformation - A repository object in a mapping that generates, modifies, or passes data. Each transformation performs a different function. A transformation can be reusable or non-reusable.

Workflow - A graphical representation of a set of events, tasks, and decisions that define a business process.

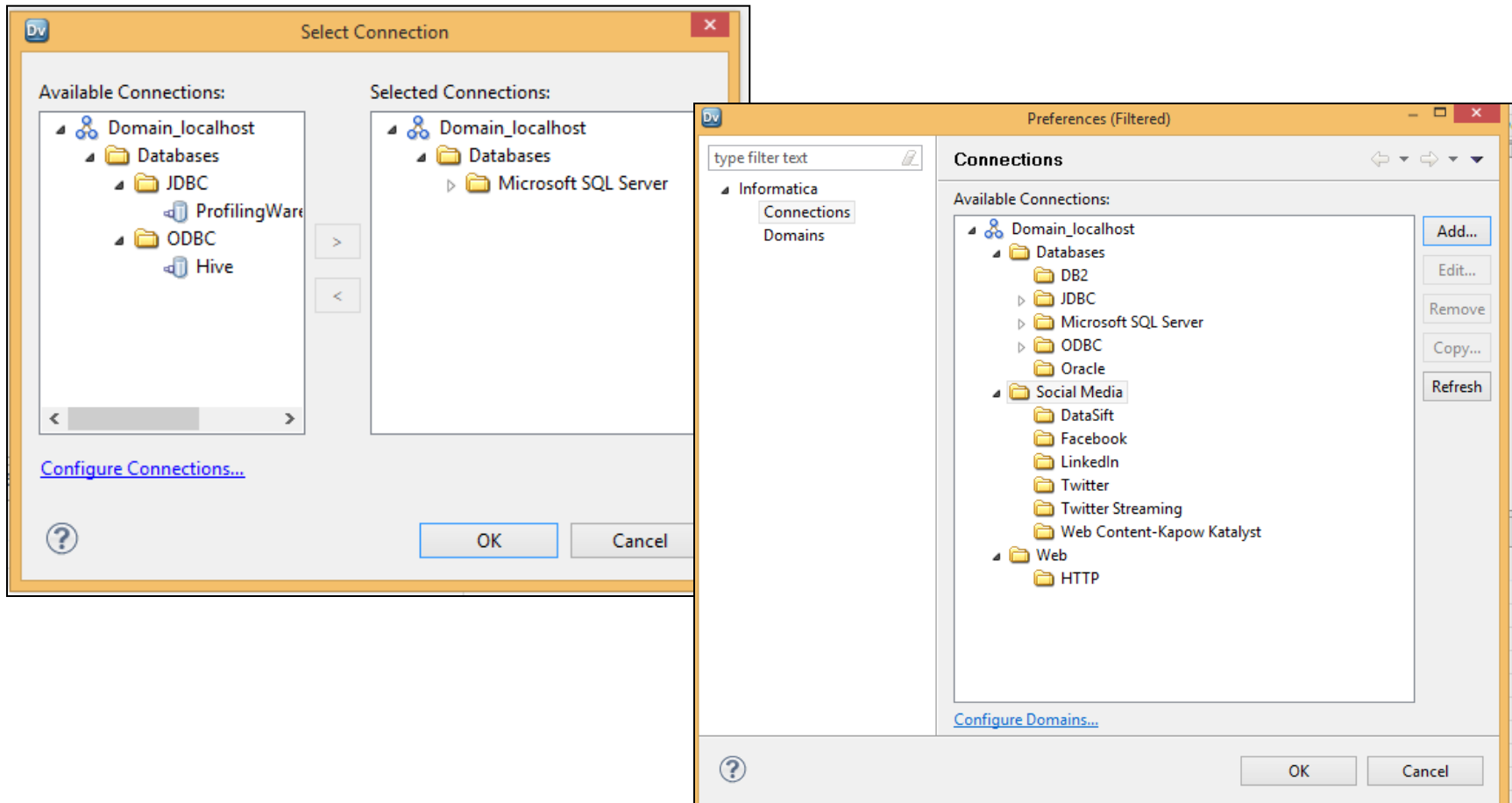
Connection Explorer

- Create relational or non-relational data objects.
- View relational or non-relational database connections



Connection Explorer

- Select Available Connections or Add new connection



Physical Data Objects Overview

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

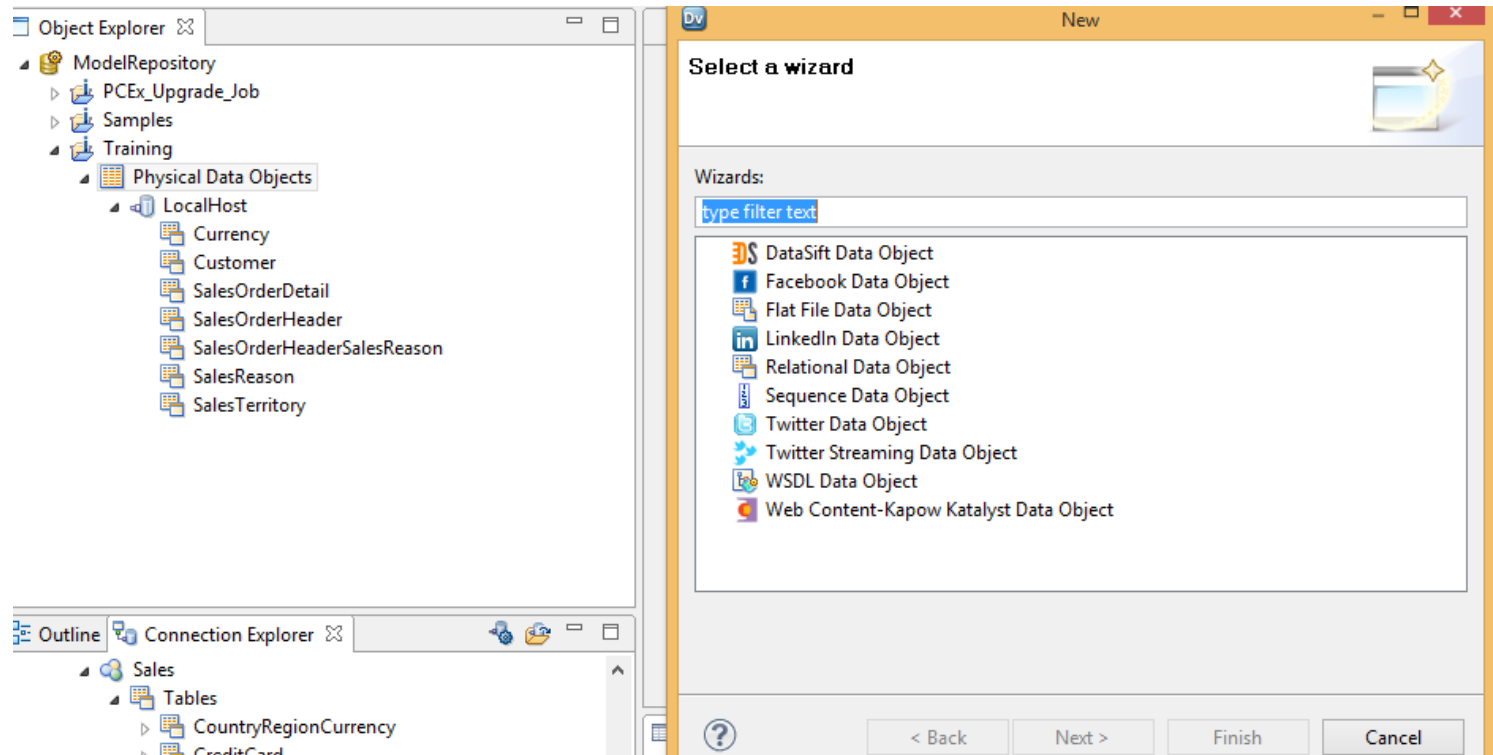
Relational data object A physical data object that uses a relational table, view, or synonym as a source. Use it as source, a target, or a Lookup transformation.

Customized data object A physical data object that uses one or multiple related relational resources or relational data objects as sources.

Flat file data object A physical data object that uses a flat file as a source.

WSDL data object A physical data object that uses a WSDL file as a source.

Physical Data Objects Overview



Create from Database

New Relational Data Object

Create a relational data object.

Connection: LocalHost_SQL_SERVER Browse...

☒ Create data object from existing resource

Resource: Currency Browse...

☐ Create customized data object

Name: Currency1

Location: ModelRepository\Training Browse...

? < Back Next > Finish Cancel

Select a Resource

Filter: *

Resource:

You can select only one resource.

☐ Show selected ☐ Show Default Schema Only

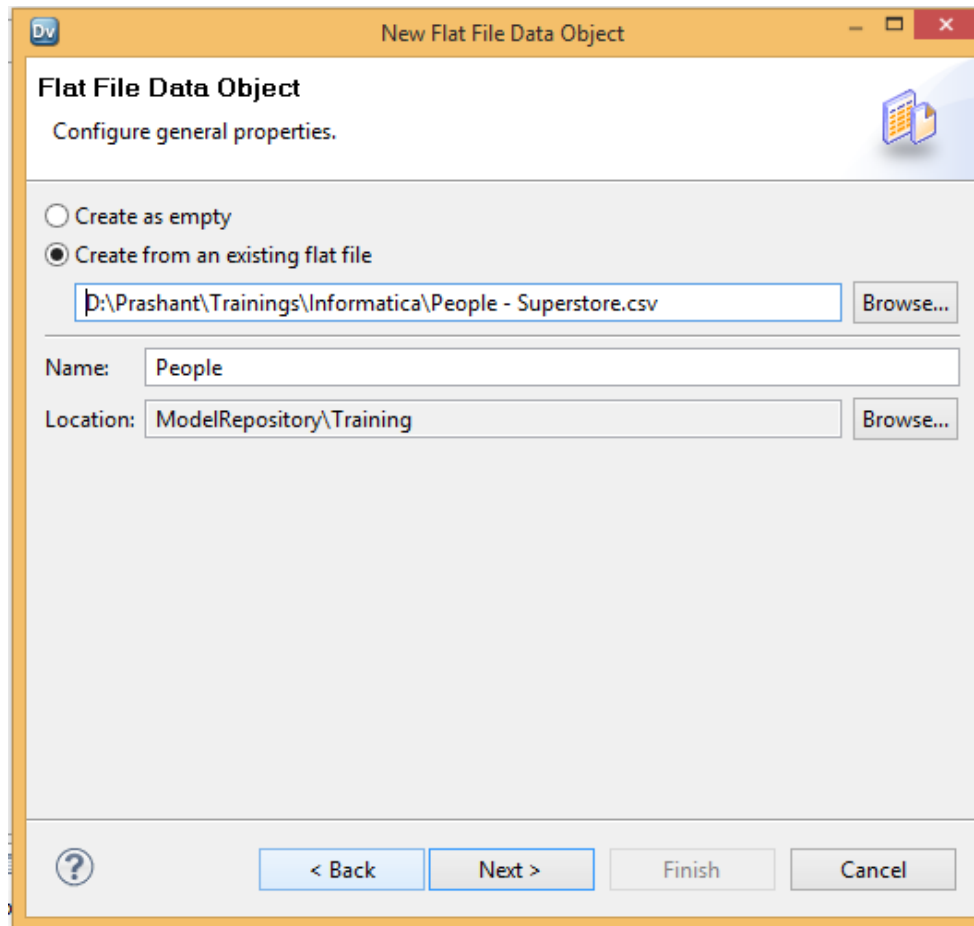
- Purchasing
- Sales
 - Tables
 - CountryRegionCurrency
 - CreditCard
 - ☒ Currency
 - Dim_SalesReason
 - Dim_SalesTerritory
 - Extract_Customer
 - Fact_Orders

Name Currency

Type Table

? OK Cancel

Create from File



The image shows a Windows-style dialog box titled "New Flat File Data Object". The dialog has a yellow title bar with standard minimize, maximize, and close buttons. Inside, the title "Flat File Data Object" is followed by the instruction "Configure general properties." and a small icon of a document with a grid. There are two radio buttons: "Create as empty" (unselected) and "Create from an existing flat file" (selected). Below the radio buttons is a text field containing the file path "D:\Prashant\Trainings\Informatica\People - Superstore.csv" and a "Browse..." button. Below this is a "Name:" label followed by a text field containing "People". Below that is a "Location:" label followed by a text field containing "ModelRepository\Training" and another "Browse..." button. At the bottom, there is a help icon (question mark in a circle) and four buttons: "< Back", "Next >", "Finish", and "Cancel".

New Flat File Data Object

Flat File Data Object
Configure general properties.

☐ Create as empty
☒ Create from an existing flat file

D:\Prashant\Trainings\Informatica\People - Superstore.csv **Browse...**

Name: People

Location: ModelRepository\Training **Browse...**

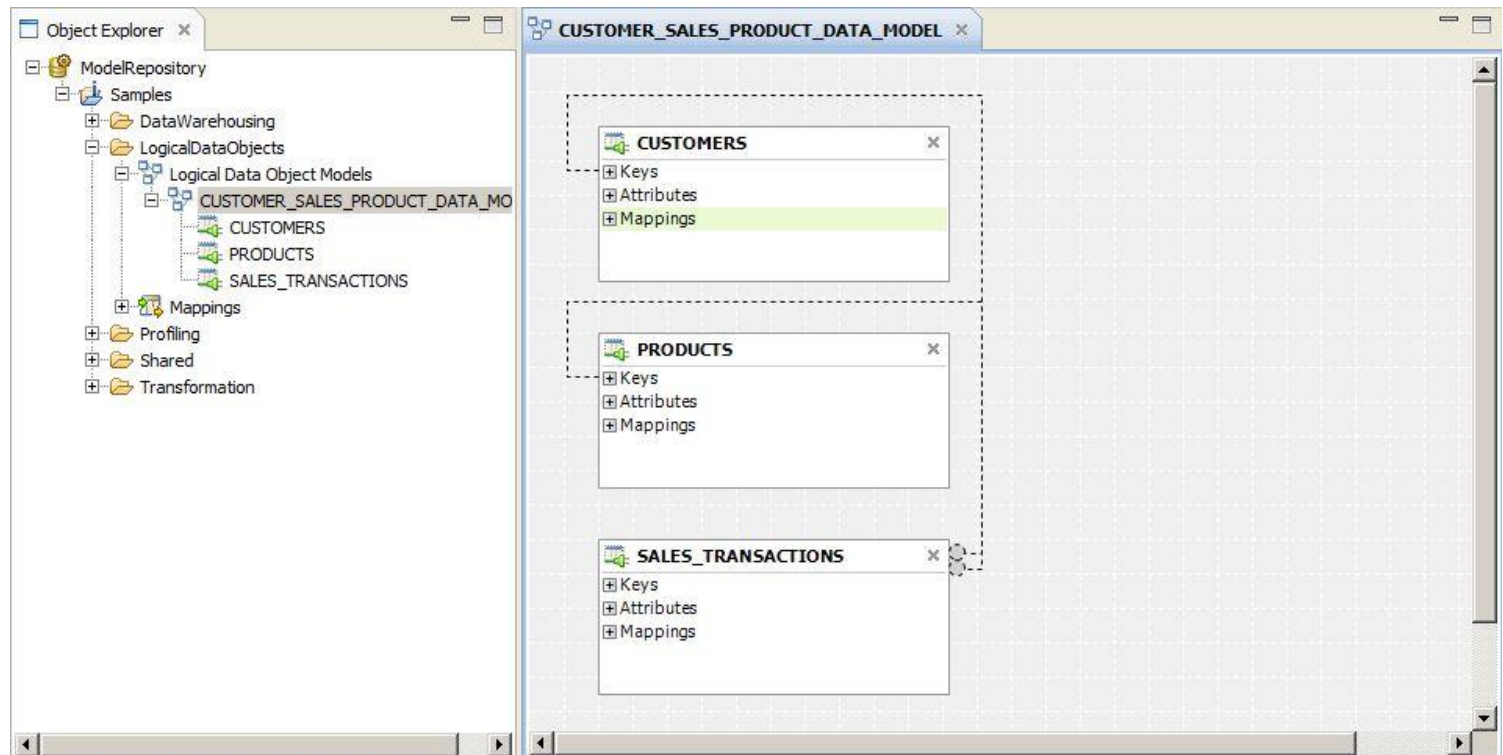
? < Back Next > Finish Cancel

Demo

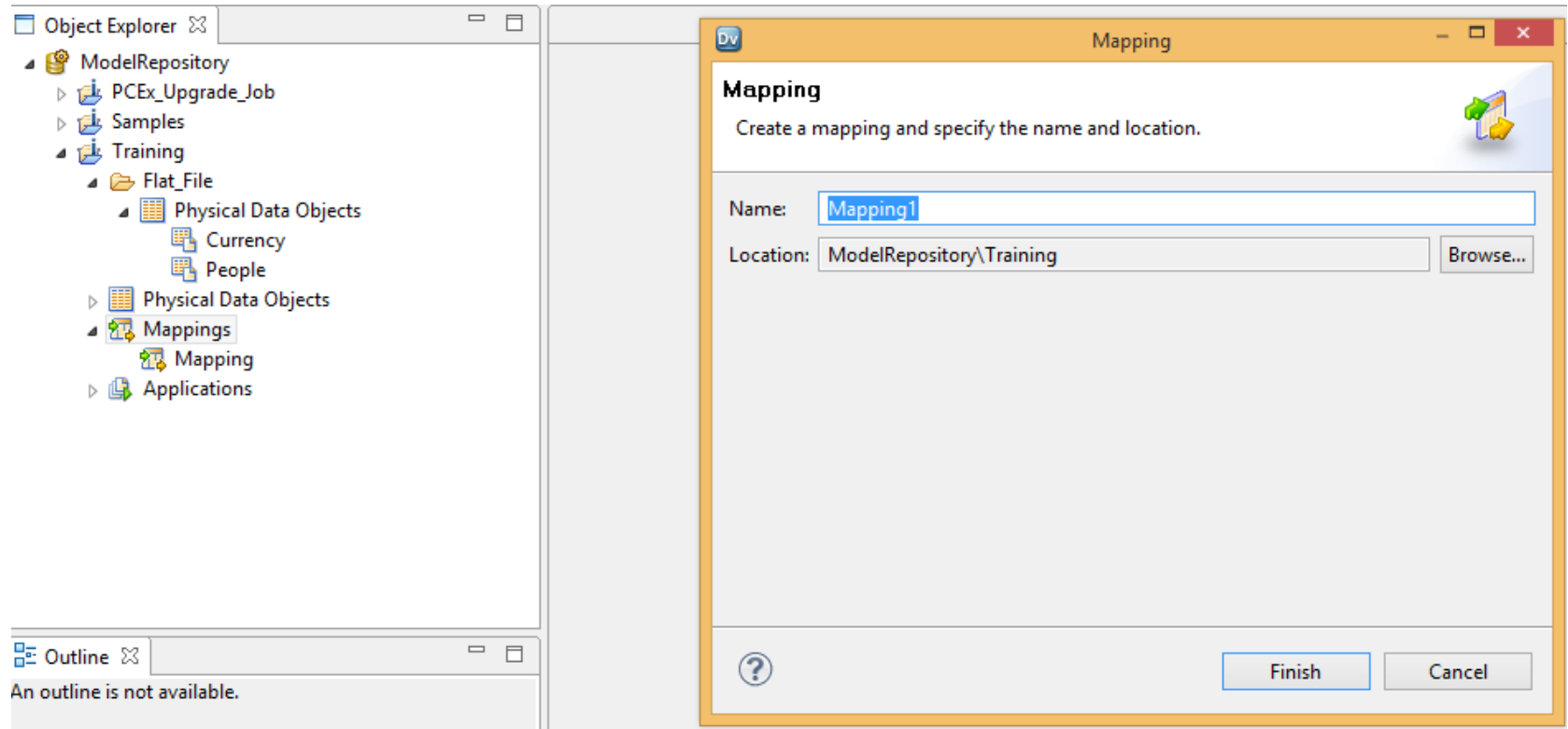
- Create Source and Target objects
- Export the data from Source

Logical View of Data

A logical view of data includes a logical data model, logical data objects, and logical data object mappings.

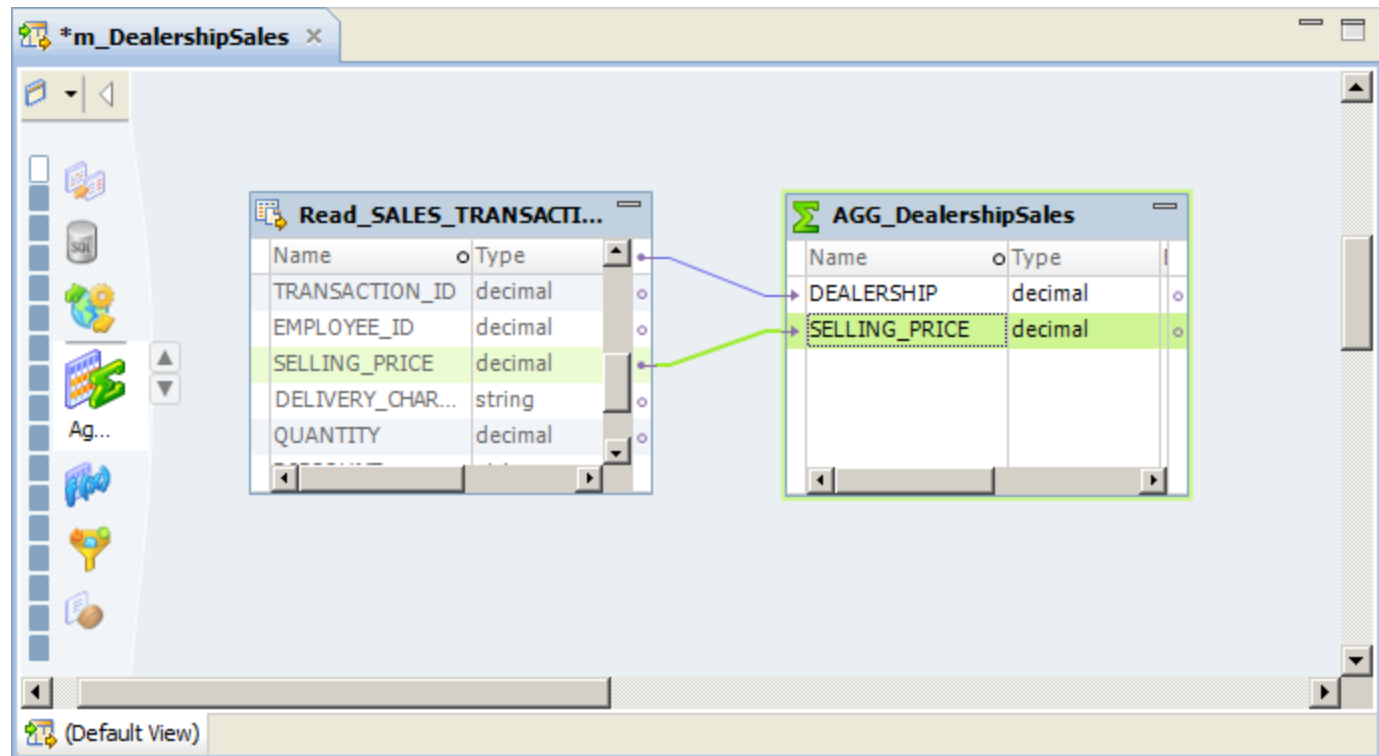


Creation of simple mapping



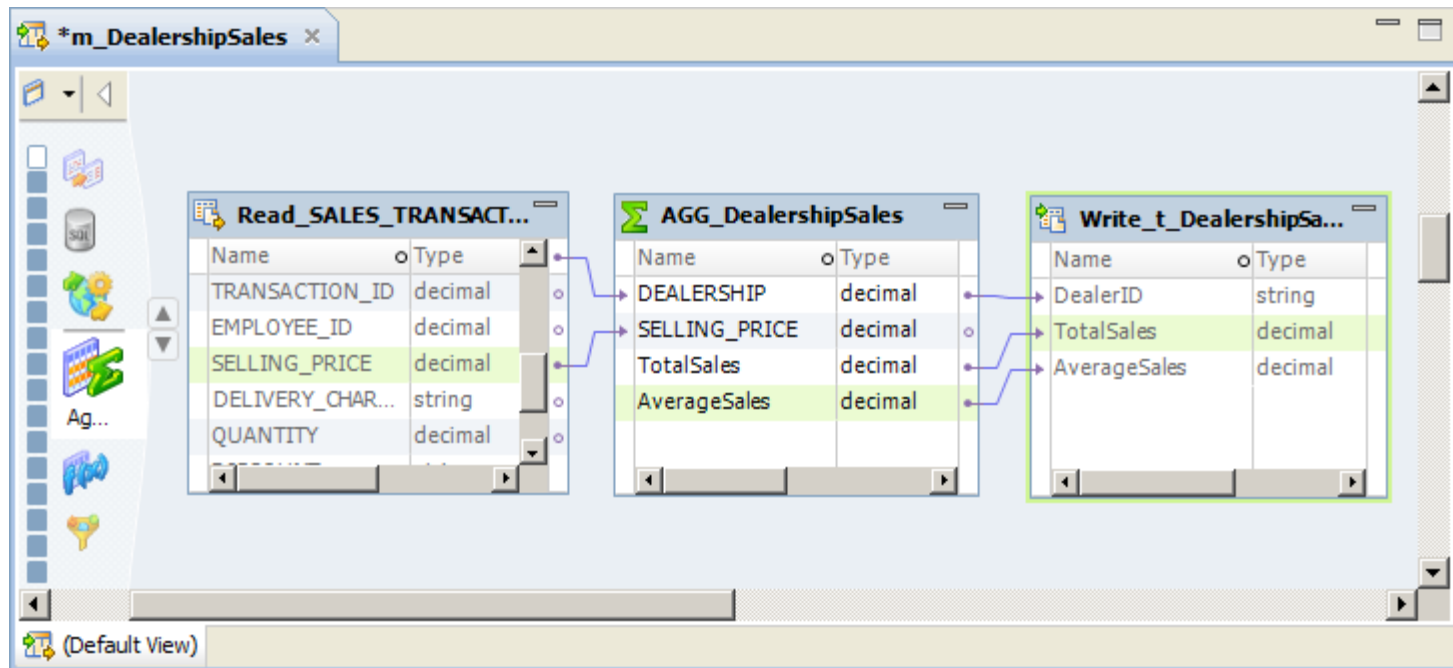
Creation of simple mapping

- Right-click in the editor, and select Add Transformation
- Or Add Transformation from scroll Menu



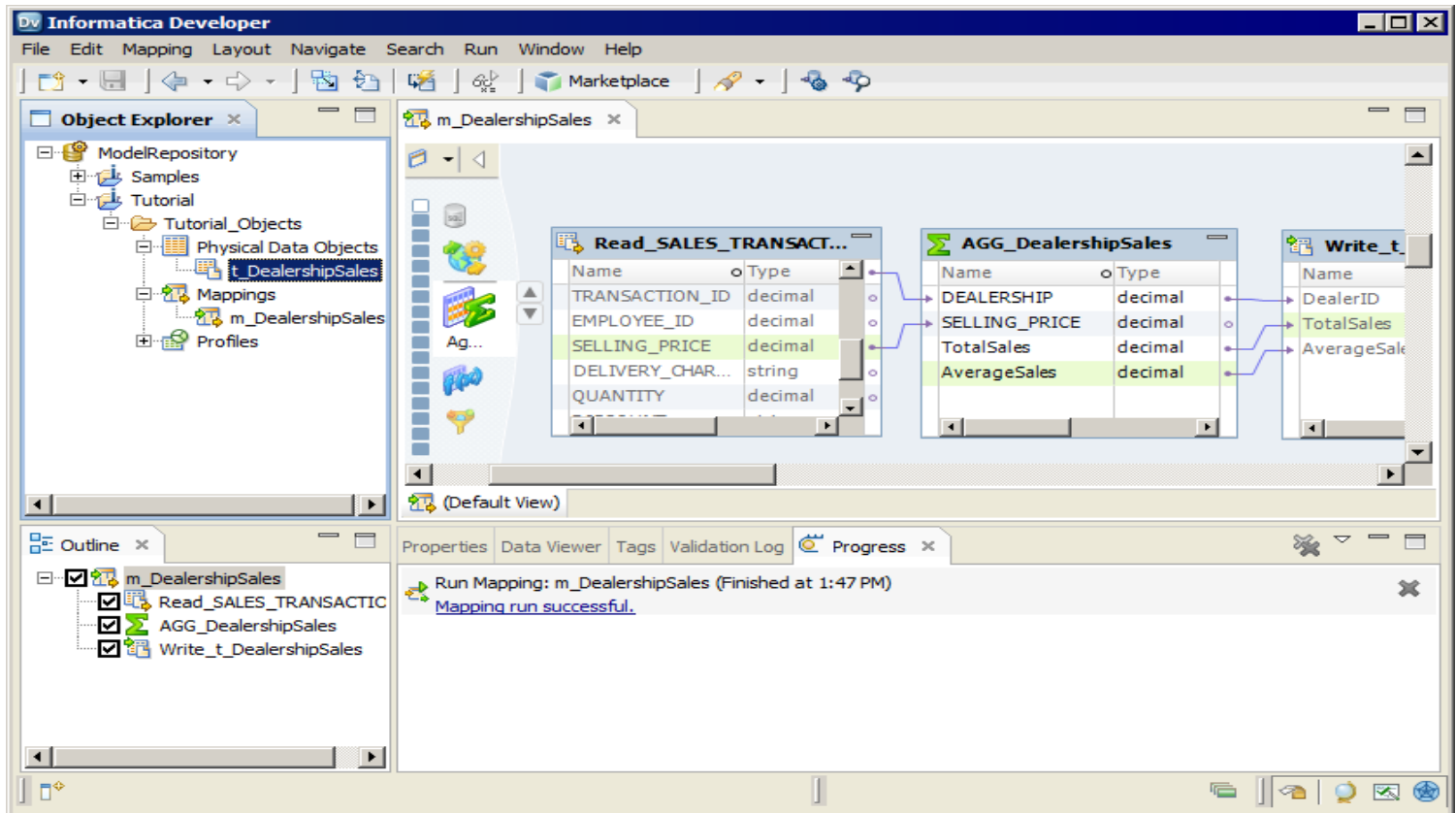
Creation of simple mapping

Add a Target to the Mapping



Creation of simple mapping

Run the Mapping and Review Results



The screenshot displays the Informatica Developer interface. The **Object Explorer** on the left shows the project structure: ModelRepository > Samples > Tutorial > Tutorial_Objects > Physical Data Objects > t_DealershipSales > Mappings > m_DealershipSales. The **Outline** pane at the bottom left lists the mapping components: m_DealershipSales, Read_SALES_TRANSACTION, AGG_DealershipSales, and Write_t_DealershipSales, all of which are checked. The main workspace shows the mapping diagram with three tables: **Read_SALES_TRANSACTION**, **AGG_DealershipSales**, and **Write_t_DealershipSales**. Arrows indicate the data flow from the source table to the aggregate table and then to the target table. The **Properties** pane at the bottom right shows the status: **Run Mapping: m_DealershipSales (Finished at 1:47 PM)** with the message **Mapping run successful.**

Name	Type
TRANSACTION_ID	decimal
EMPLOYEE_ID	decimal
SELLING_PRICE	decimal
DELIVERY_CHAR...	string
QUANTITY	decimal

Name	Type
DEALERSHIP	decimal
SELLING_PRICE	decimal
TotalSales	decimal
AverageSales	decimal

Name
DealerID
TotalSales
AverageSale

Transformations

The operations that the Data Integration Service performs on the data. Data passes through transformation ports that you link in a mapping or mapplet.

Active Transformations

An active transformation changes the number of rows that pass through a transformation. Or, it changes the row type.

Passive Transformations

A passive transformation does not change the number of rows that pass through the transformation, and it maintains the row type.

Unconnected Transformations

An unconnected transformation is not connected to other transformations in the mapping. An unconnected transformation is called within another transformation, and returns a value to that transformation.

Transformations

Active transformations

Aggregator

performs aggregate calculations

Filter

serves as a conditional filter

Router

serves as a conditional filter (more than one filters)

Joiner

allows for heterogeneous joins

Source qualifier

represents all data queried from the source

Passive transformations

Expression

performs simple calculations

Lookup

looks up values and passes to other objects

Sequence generator

generates unique ID values

Stored procedure

calls a stored procedure and captures return values

Update strategy

allows for logic to insert, update, delete, or reject data

Aggregator Transformation

Use the Aggregator transformation to perform aggregate calculations

The Data Integration Service performs aggregate calculations as it reads and stores data group.

Aggregator transformation is an active transformation.

Ports :

- **Input** - Receives data from upstream transformations.
- **Output** - Provides the return value of an expression. The expression can include non-aggregate expressions and conditional clauses.
- **Pass-Through** Passes data unchanged.
- **Variable** Used for local variables.
- **Group by** - Indicates how to create groups. Group port can be any input, input/output, output, or variable port. When grouping data, the Aggregator transformation outputs the last row of each group unless otherwise specified.

Aggregator Transformation

Properties Data Viewer Tags Object Dependencies Alerts Progress Validation Log

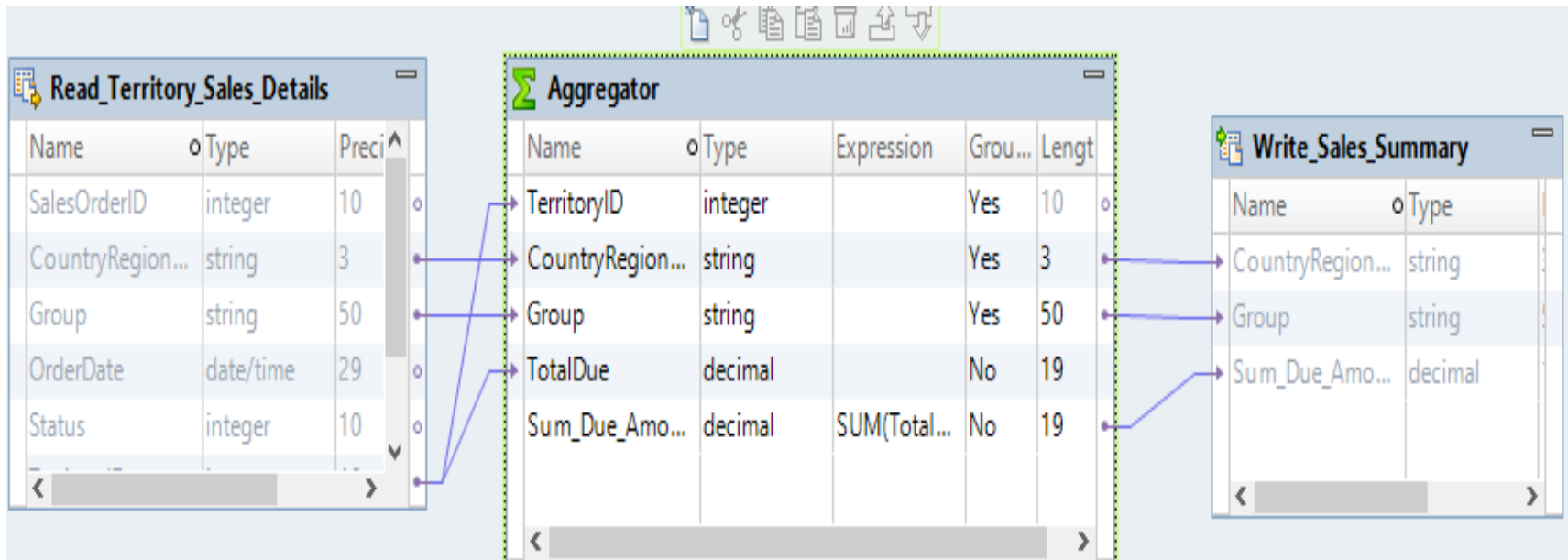
General Hide categories

Ports

Advanced

	Name	Type	Precisi...	Scale	Input	Output	Variable	Expression	Group...	Default Value
Input Only (1)										
1	TotalDue	decimal	19	4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	
Output Only (1)										
1	Sum_Due_Amo...	decimal	19	4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SUM(TotalDue)	<input type="checkbox"/>	ERROR('transf...
Passthrough (3)										
1	TerritoryID	integer	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	
2	CountryRegion...	string	3	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	
3	Group	string	50	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	
Variable (0)										
Group By (3)										
1	TerritoryID	integer	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	
2	CountryRegion...	string	3	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	
3	Group	string	50	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input checked="" type="checkbox"/>	

Aggregator Transformation



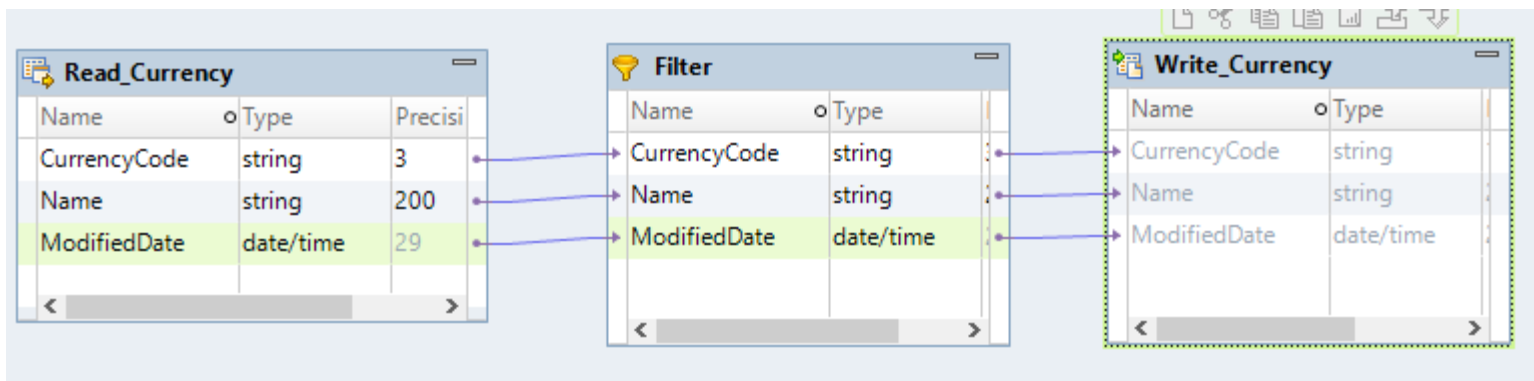
Aggregator Transformation

Aggregate Expressions:

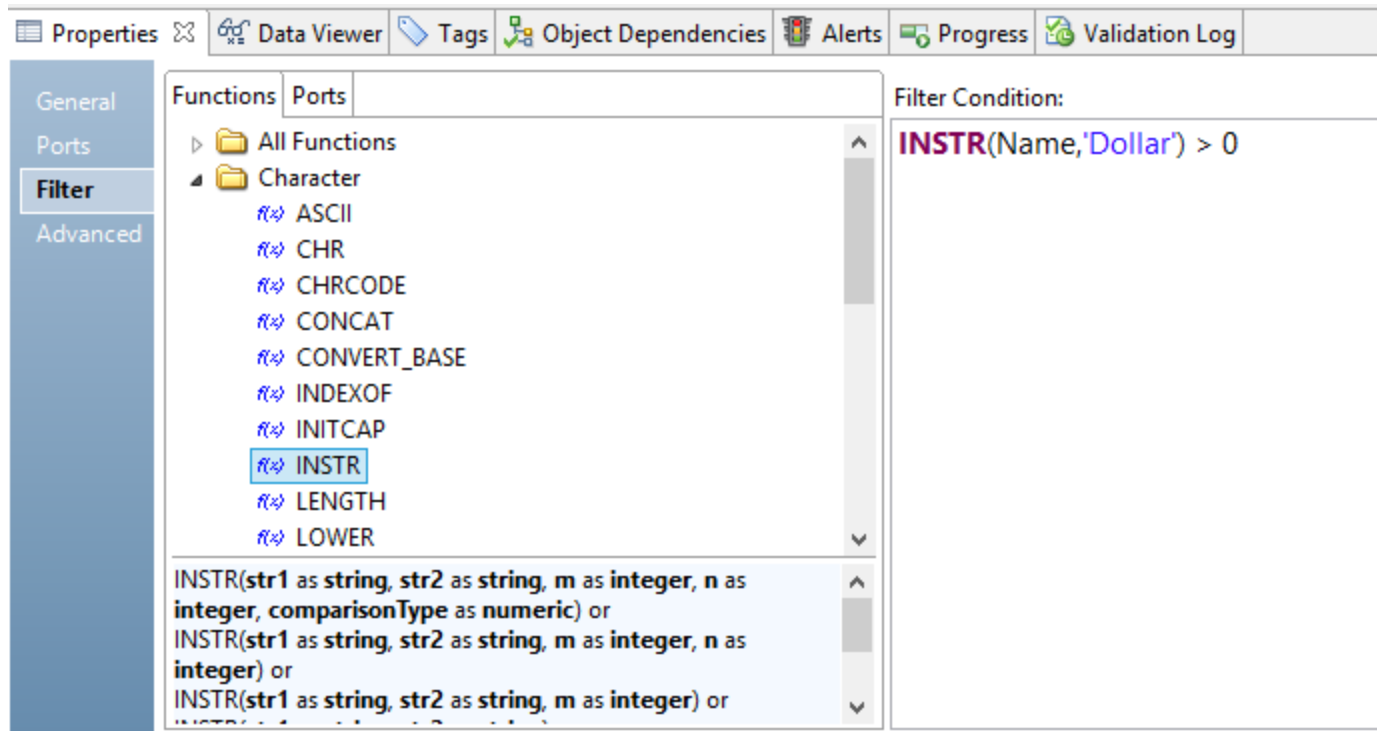
- AVG
- COUNT
- FIRST
- LAST
- MAX
- MEDIAN
- MIN
- PERCENTILE
- STDDEV
- SUM
- VARIANCE

Filter Transformation

- Filter out rows in a mapping. As an active transformation, the Filter transformation may change the number of rows passed through it.
-
- Allows rows that meet the specified filter condition to pass through.
- Drops rows that do not meet the condition.
- Filter data based on one or more conditions.



Filter Transformation




Router Transformation

- Active transformation that routes data into multiple output groups based on one or more conditions.
- It is similar to a Filter transformation because both transformations use a condition to test data.
- A Router transformation tests data for one or more conditions and can route rows of data that do not meet any of the conditions to a default output group.
- Use router transformation if we need to filter same input data based on multiple conditions, instead of creating multiple Filter transformations to perform the same task.

Router Transformation

1. Input group
2. Input ports
3. Default output group
4. User-defined output groups



Name	Type	Length/...
Input (4)		
COUNTRY	string	13
CUSTOMER_NO	decimal	10
FIRSTNAME	string	10
LASTNAME	string	10
Default (4)		
COUNTRY	string	13
CUSTOMER_NO	decimal	10
FIRSTNAME	string	10
LASTNAME	string	10
France (4)		
COUNTRY	string	13
CUSTOMER_NO	decimal	10
FIRSTNAME	string	10
LASTNAME	string	10
Japan (4)		
COUNTRY	string	13
CUSTOMER_NO	decimal	10
FIRSTNAME	string	10
LASTNAME	string	10
USA (4)		
COUNTRY	string	13
CUSTOMER_NO	decimal	10
FIRSTNAME	string	10
LASTNAME	string	10

Router Transformation

Properties

Data Viewer

Tags

General

Ports

Groups

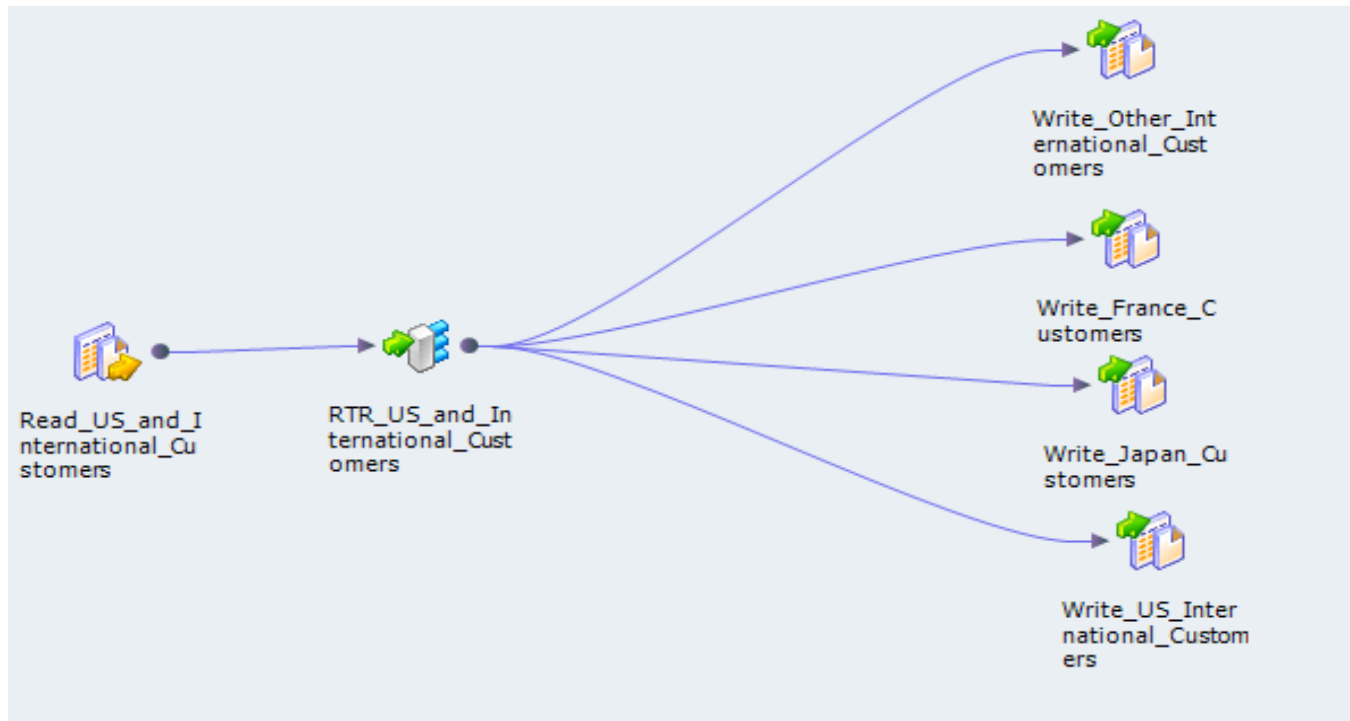
Advanced

Output Groups:

Group Name	Group Filter Condition
Default	
France	COUNTRY='France'=TRUE
Japan	COUNTRY='Japan'=TRUE
USA	COUNTRY='USA'=TRUE

Description:

Router Transformation



Joiner Transformation

- Joiner transformation joins source data from two related heterogeneous sources from different locations or from different file systems.
- Joiner transformation is a multiple-group active transformation.
- Uses a condition that matches one or more pairs of columns between the two sources.
- The two input pipelines include a master pipeline and a detail pipeline or a master and a detail branch.
- The master pipeline ends at the Joiner transformation, while the detail pipeline continues to the target.

Joiner Transformation

Properties Data Viewer Tags Object Dependencies Alerts Progress Validation Log

General

Ports

Join

Advanced

	Name	Type	Precisi...	Scale	Master	Default Value	Description
MASTER (3)							
1	CUSTOMER_NO	decimal	5	0	<input checked="" type="checkbox"/>		
2	FIRSTNAME	string	20	0	<input checked="" type="checkbox"/>		
3	LASTNAME	string	20	0	<input checked="" type="checkbox"/>		
DETAIL (14)							
1	CUST_NO	decimal	4	0	<input type="checkbox"/>		
2	PRODUCT	decimal	5	0	<input type="checkbox"/>		
3	DEALERSHIP	decimal	2	0	<input type="checkbox"/>		
4	PAYMENT_DESC	string	12	0	<input type="checkbox"/>		
5	PROMO_ID	decimal	3	0	<input type="checkbox"/>		
6	DATE_ID	decimal	7	0	<input type="checkbox"/>		
7	TRANSACTION...	string	21	0	<input type="checkbox"/>		
8	TRANSACTION...	decimal	7	0	<input type="checkbox"/>		
9	EMPLOYEE_ID	decimal	5	0	<input type="checkbox"/>		
10	SELLING_PRICE	string	8	0	<input type="checkbox"/>		
11	DELIVERY_CHA...	string	7	0	<input type="checkbox"/>		
12	QUANTITY	decimal	1	0	<input type="checkbox"/>		
13	DISCOUNT	string	0	0	<input type="checkbox"/>		

Properties Data Viewer Tags Object Dependencies Alerts Progress Validation Log

General

Ports

Join

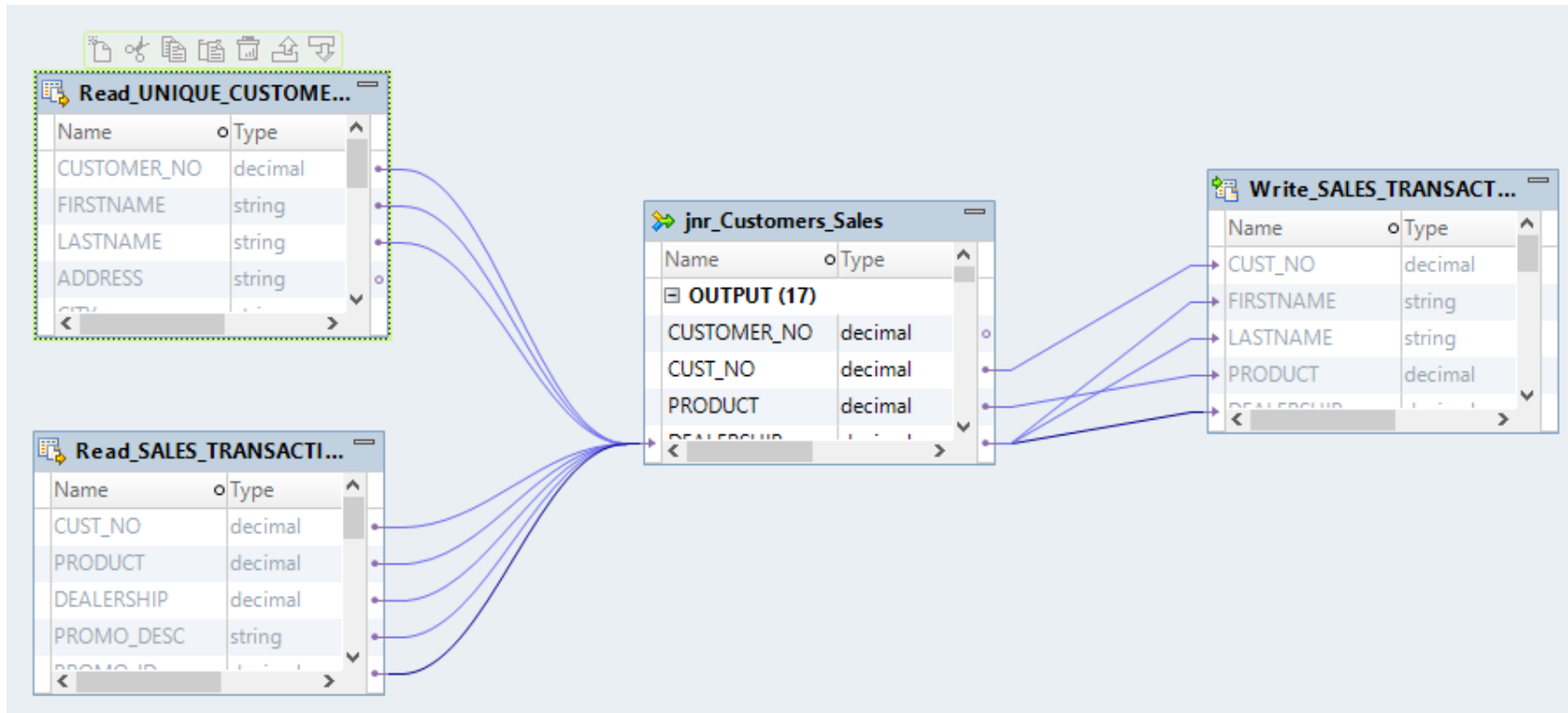
Advanced

Join type: Normal Condition type: ☒ Simple ☐ Advanced

Join condition:

Master	Operator	Detail
CUSTOMER_NO	=	CUST_NO

Joiner Transformation



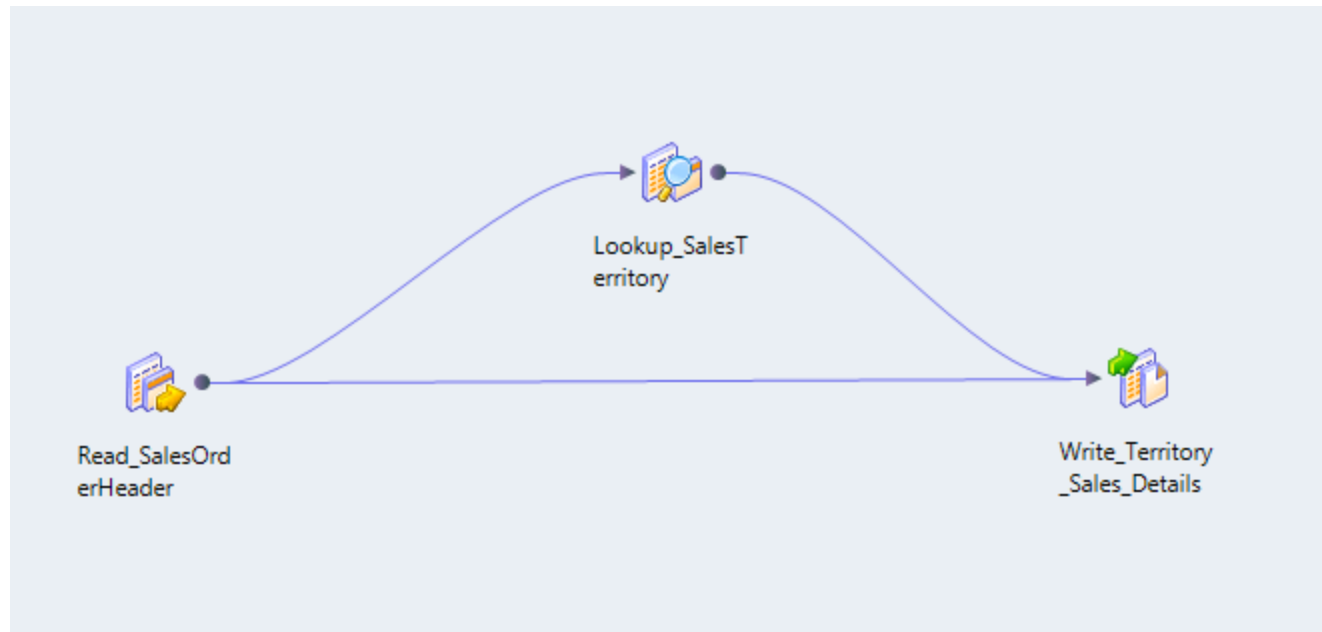
Lookup Transformation

The Lookup transformation is a passive or active transformation that looks up data in a flat file, logical data object, reference table, or relational table. The Lookup transformation can return one row or multiple rows from a lookup.

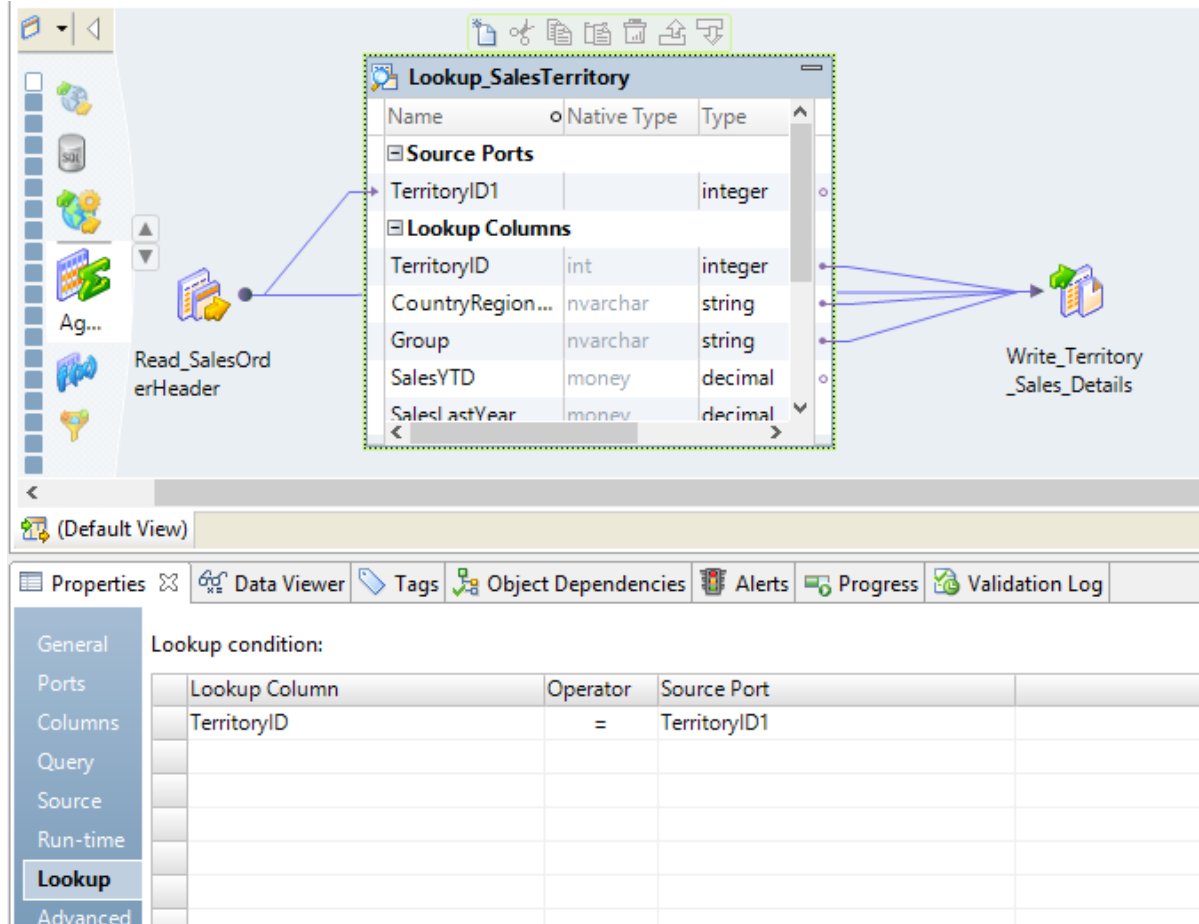
Lookup transformation can perform below tasks -

- Get a related value. Retrieve a value from the lookup source based on a value in the input data.
-
- Retrieve multiple rows from a lookup source.
- Perform a calculation. Retrieve a value from a lookup table and use the value in a calculation.
- Perform an unconnected lookup with a :LKP expression in a transformation that accepts expressions.

Lookup Transformation



Lookup Transformation



Lookup_SalesTerritory

Name	Native Type	Type
Source Ports		
TerritoryID1		integer
Lookup Columns		
TerritoryID	int	integer
CountryRegion...	nvarchar	string
Group	nvarchar	string
SalesYTD	money	decimal
SalesLastYear	money	decimal

Read_SalesOrderHeader → Lookup_SalesTerritory → Write_Territory_Sales_Details

(Default View)

Properties | Data Viewer | Tags | Object Dependencies | Alerts | Progress | Validation Log

Lookup

Lookup condition:

Lookup Column	Operator	Source Port
TerritoryID	=	TerritoryID1

Expression Transformation

The Expression transformation is a passive transformation used to perform non-aggregate calculations in a single row.

Expression transformation can test conditional statements before passing the results to a target or other transformations

Port types :

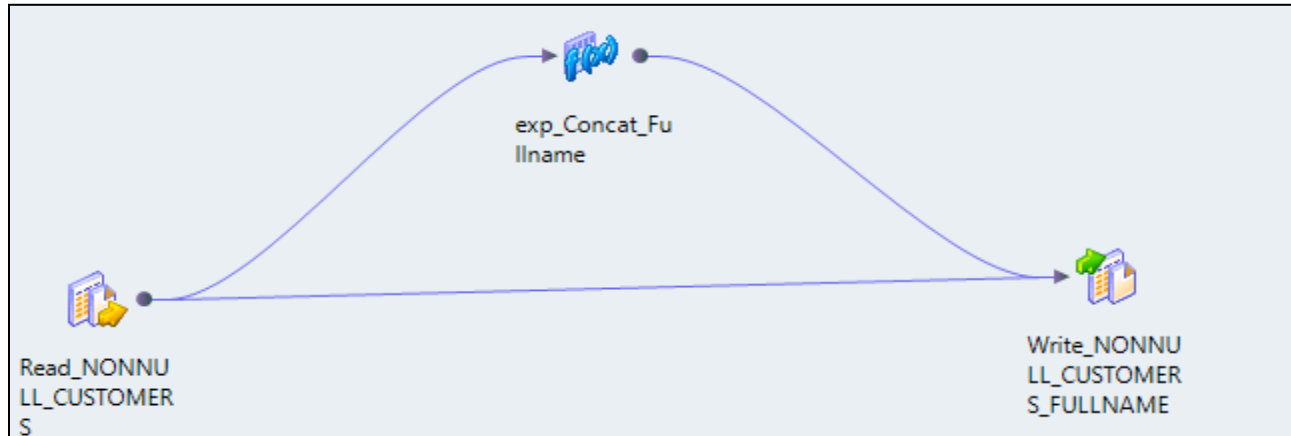
Input - Receives data from upstream transformations. Input ports provide values used in a calculation.

Output - Provides the return value of the expression. Enter the expression as a configuration option for the output port.

Pass-Through Passes data unchanged.

Variable - Used for local variables

Expression Transformation



Properties

Data Viewer

Tags

Object Dependencies

Alerts

Progress

Validation Log

General

Ports

Dependencies

Advanced

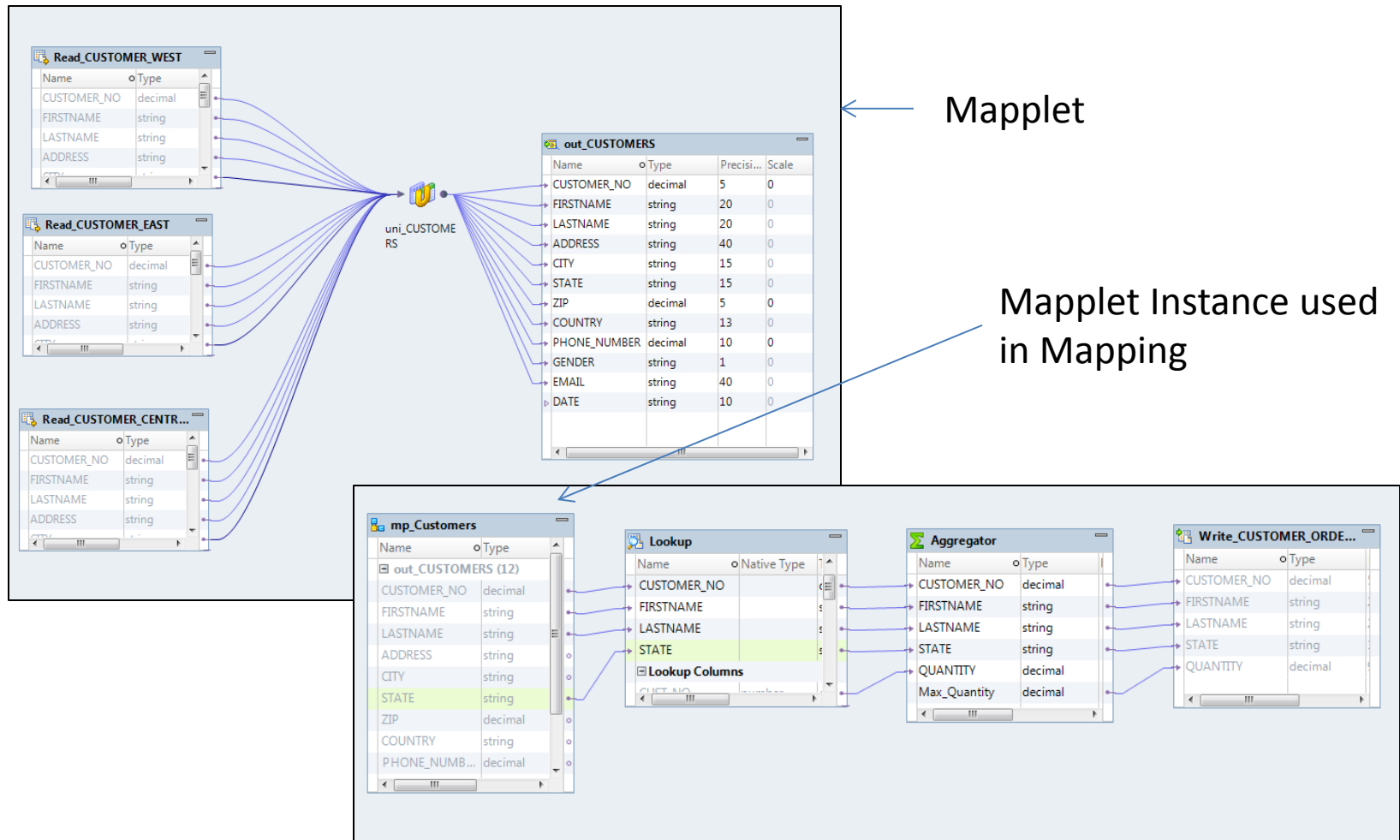
Hide categories

	Name	Type	Precisi...	Scale	Input	Output	Variable	Expression
Input Only (2)								
1	FIRSTNAME	string	20	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2	LASTNAME	string	20	0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Output Only (1)								
1	FULLNAME	string	40	0	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	CONCAT(CONCAT(FIRSTNAM...
Passthrough (0)								
Variable (0)								

Mapplet

- A mapplet is a reusable object containing a set of transformations that you can use in multiple mappings.
- When you use a mapplet in a mapping, you use an instance of the mapplet.
- Any Change made to the mapplet is inherited by all instances of the mapplet.
- Mapplets can contain other mapplets. You can also use a mapplet more than once in a mapping or mapplet.

Mapplet



Workflows

A workflow is a **graphical representation** of a set of events, tasks, and decisions that define a business process.

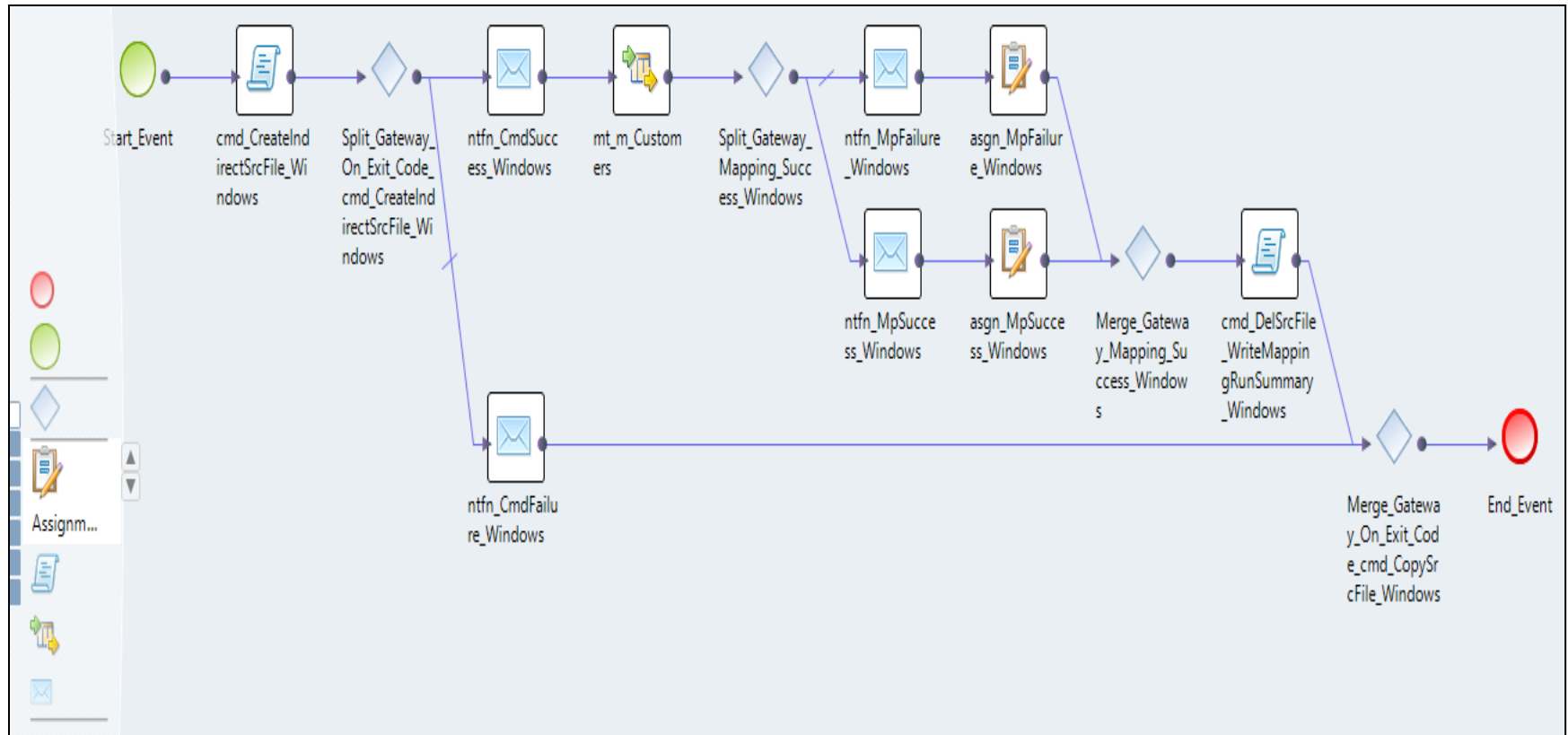
Add objects to a workflow and to connect the objects with sequence flows.

Data Integration Service uses the instructions configured in the workflow to run the objects.

Workflow object can be - **event, task, or gateway**.

- **Event** - starts or ends the workflow.
- **Task** - is an activity that runs a single unit of work in the workflow, such as running a mapping, sending an email, or running a shell command.
- **Gateway** - Makes a decision to split and merge paths in the workflow.

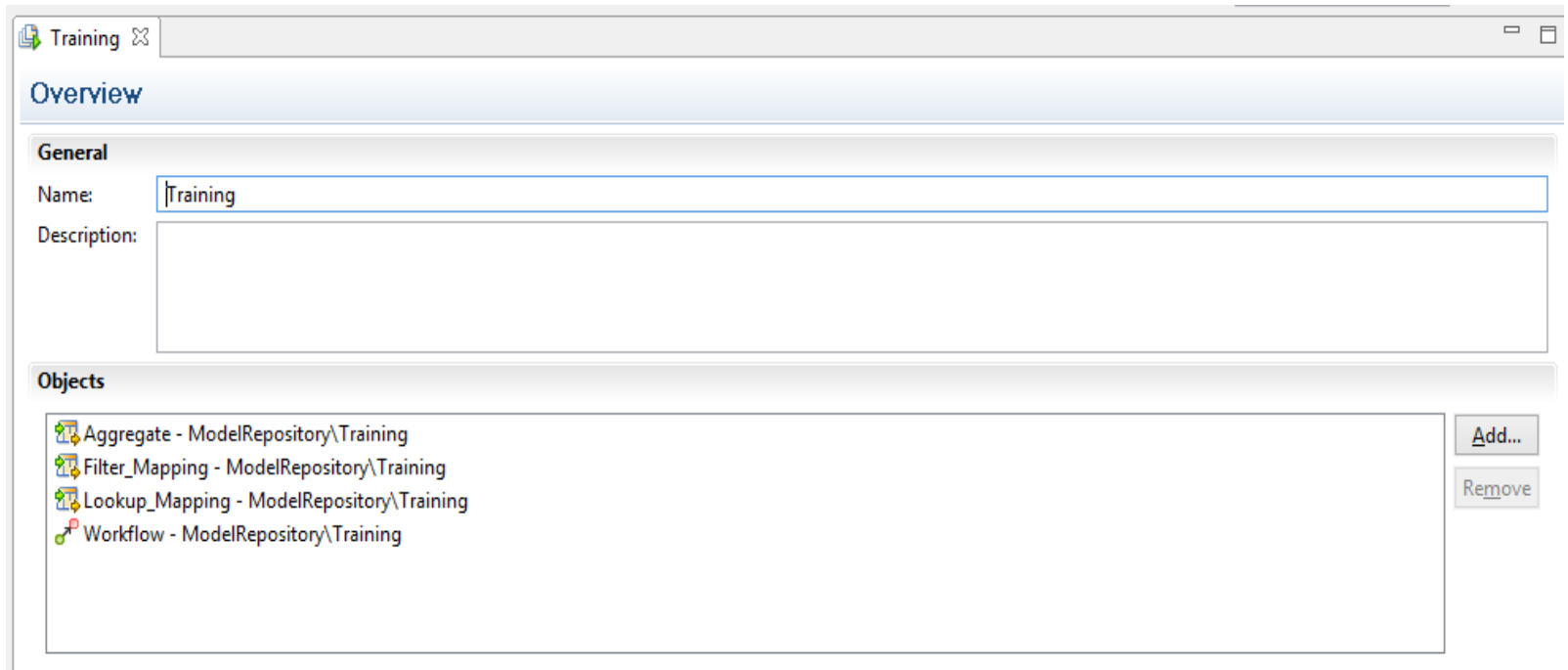
Workflow



Workflow Deployment

- Develop a workflow in the Developer tool via task, events and gateways object.
- Add the workflow definition to an application.
- Deploy the application to the Data Integration Service.

Workflow Deployment



The screenshot displays the 'Training' workflow configuration window. The window has a title bar with 'Training' and a close button. Below the title bar is a tab labeled 'Overview'. The main content area is divided into two sections: 'General' and 'Objects'.

General

Name:

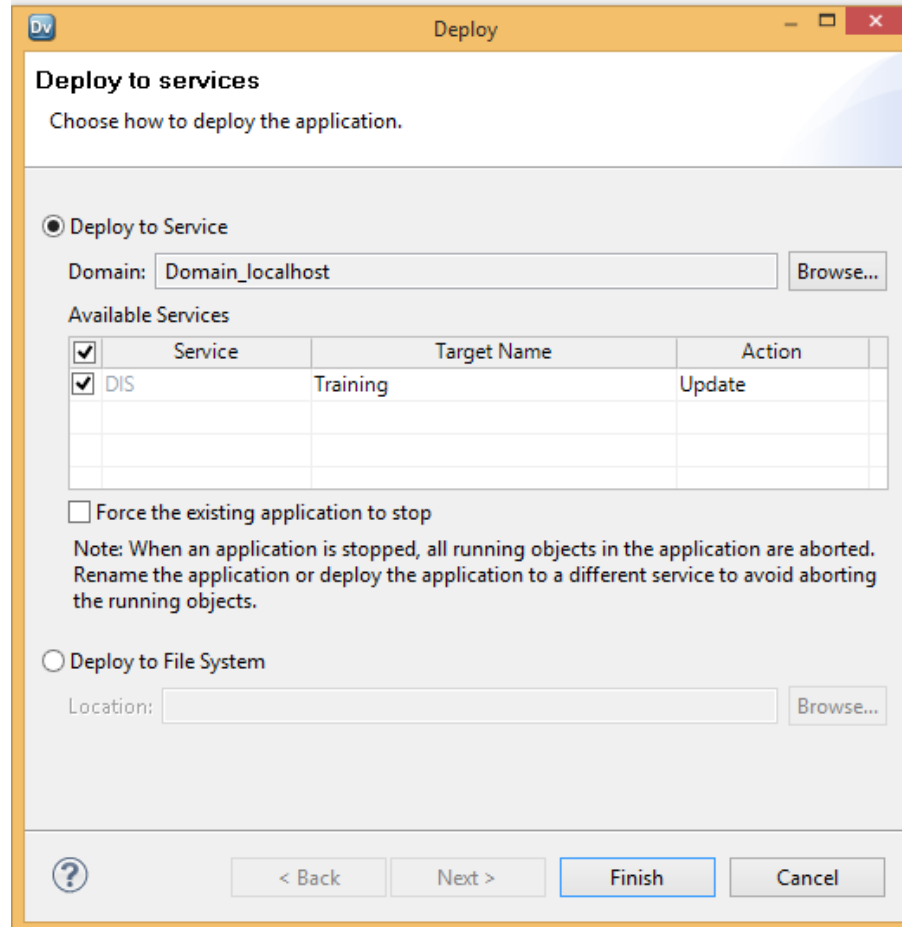
Description:

Objects

- Aggregate - ModelRepository\Training
- Filter_Mapping - ModelRepository\Training
- Lookup_Mapping - ModelRepository\Training
- Workflow - ModelRepository\Training

Buttons:

Workflow Deployment



Deploy

Deploy to services
Choose how to deploy the application.

☒ **Deploy to Service**

Domain:

Available Services

<input checked="" type="checkbox"/>	Service	Target Name	Action
<input checked="" type="checkbox"/>	DIS	Training	Update
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			

☐ Force the existing application to stop

Note: When an application is stopped, all running objects in the application are aborted. Rename the application or deploy the application to a different service to avoid aborting the running objects.

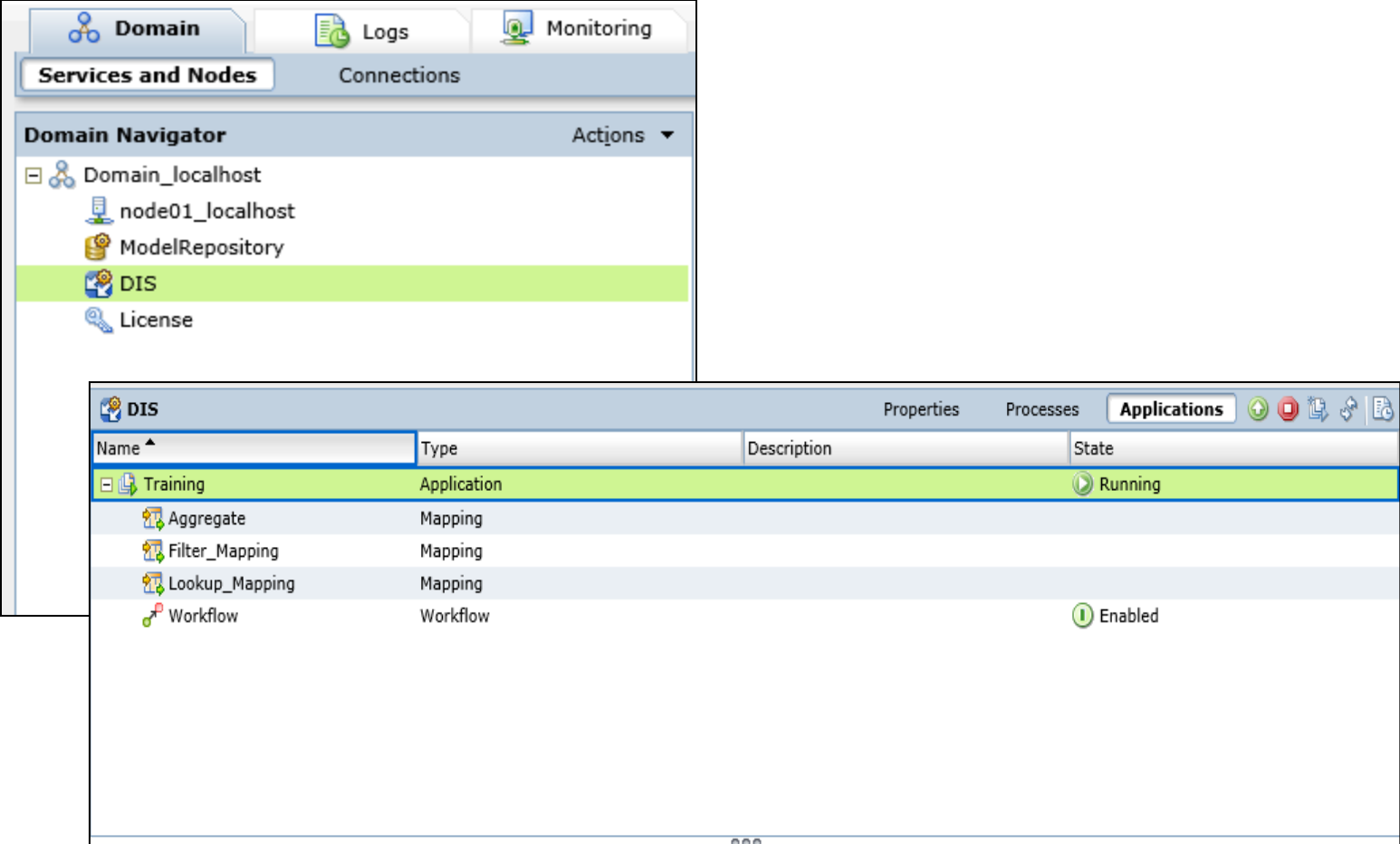
☐ **Deploy to File System**

Location:

Workflow Monitor

- Monitor a workflow instance run in the Monitoring tool.
- The Monitoring tool is a direct link to the Monitoring tab of the Administrator tool.
- The Monitoring tool shows the status of running workflow and workflow object instances.
- Use the Monitoring tool to view logs for workflow instances and to view workflow reports.

Workflow Monitor



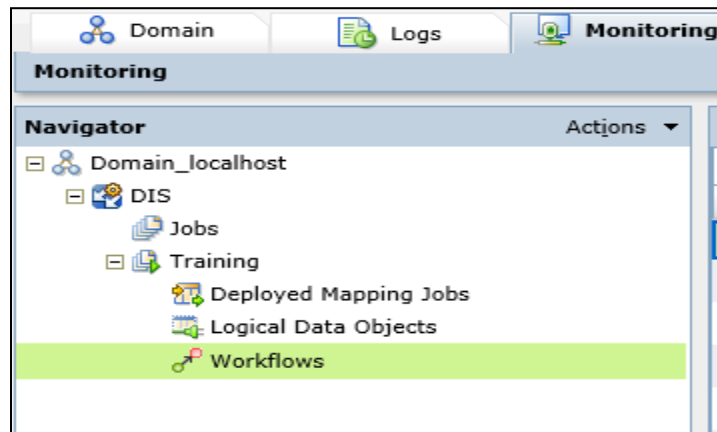
Domain Navigator

- Domain_localhost
 - node01_localhost
 - ModelRepository
 - DIS**
 - License

DIS Properties - Applications Tab

Name	Type	Description	State
Training	Application		Running
Aggregate	Mapping		
Filter_Mapping	Mapping		
Lookup_Mapping	Mapping		
Workflow	Workflow		Enabled

Workflow Monitor



Workflows							
Name	Type	State	ID	Started By	Start Time	Updated At	Elapsed Time
Workflow	Workflow	Completed	U:GfGqAHZTE...	Administrator	09/09/2016 11:33:17	09/09/2016 11:33:21	00:00:03
Lookup_Mapping	Mapping Task	Completed	G3_r1HZTEak...	Administrator	09/09/2016 11:33:19	09/09/2016 11:33:20	00:00:01
Filter_Mapping	Mapping Task	Completed	GttHEXZTEak...	Administrator	09/09/2016 11:33:18	09/09/2016 11:33:19	00:00:00
Aggregate	Mapping Task	Completed	GhpAr3ZTEak...	Administrator	09/09/2016 11:33:17	09/09/2016 11:33:18	00:00:01
Workflow	Workflow	Failed	U:1IKM1XZSE...	Administrator	09/09/2016 11:31:20	09/09/2016 11:31:25	00:00:04
Workflow	Workflow	Failed	U:L_VD0XZRE...	Administrator	09/09/2016 11:19:35	09/09/2016 11:19:48	00:00:13

Question?





Thank You