How-To Library

Informatica

Creating a Column Profile on a Logical Data Object in Informatica Developer

Abstract

To identify data quality issues in multiple data sources, you can create an enterprise discovery profile, or you can create a column profile on a logical data object. This article describes how to create and run a column profile on the logical data object in Informatica Developer to identify data quality issues, such as null values, inferred data types, and inferred data patterns.

Supported Versions

- Data Quality 10.0

Table of Contents

Overview

You can consolidate multiple physical data sources into a logical data object model. A logical data object model contains logical data objects and mappings. You can create and run a column profile on the logical data object to find the data quality of the data sources.

A logical view of data is a representation of data that exists in an enterprise. The data in an enterprise might reside in multiple disparate source systems, such as relational databases and flat files. You can create a logical data object model to represent the data from the perspective of the business regardless of the source systems.

The logical view of data includes the following components:

1. Logical data object model. A logical data object model describes the structure and use of data in an enterprise. You can create a logical data object model or use an existing model to create a logical data object model. You can use the logical data object model to study data, describe data attributes, and define the relationships among attributes.
2. Logical data object. An object in a logical data object model that describes a logical entity in an enterprise.
3. Logical data object mapping. A mapping that links a logical data object to one or more physical data objects. It can include transformation logic.

You can create and run a column profile on the logical data object to find the following type of information:

- Content of the data source that includes value frequencies and data types.
- Quality of the data source which includes unique values, null values, and data patterns.
- Structure of the data sources which includes keys and functional dependencies.
- Data domains in the data source which identify critical data characteristics within an enterprise.

The following sections explain how to create a logical data object model, logical data object, Joiner transformation, and to create and run a column profile on the logical data object.

Business Example

You are a data analyst in an International Bank. The bank acquires a National Bank. After the acquisition, the International Bank wants to find all the customers residing within a ZIP code location while identifying data quality issues associated with the selected records.

To accomplish the task, you must perform the following steps:

1. Create a logical data object model and name it Customer_LDOM.
2. Create a logical data object and name it Customer_LDO.
3. Create flat file data objects for the Customer tables in the International Bank database and National Bank database.
4. Create a mapping with a Joiner transformation, and map the flat file data objects to the logical data object.
5. Create a column profile on the logical data object, and view the profile results.

You can view the profile results in Informatica Developer and Informatica Analyst. In Informatica Analyst, you can view the null values, inferred data types, inferred patterns, values, and frequencies for all the columns in summary view.

The following tasks explain how to create a logical data object and create and run a column profile on the logical data object.

Creating a Logical Data Object Model

You can create a logical data object model to unify data from the Customer tables in the International Bank database and the National Bank database. You can create logical data objects and add mappings into the logical data object model.

1. In Informatica Developer, click **File** > **New** > **Project** to create a project named **ConsolidateCustomers**.

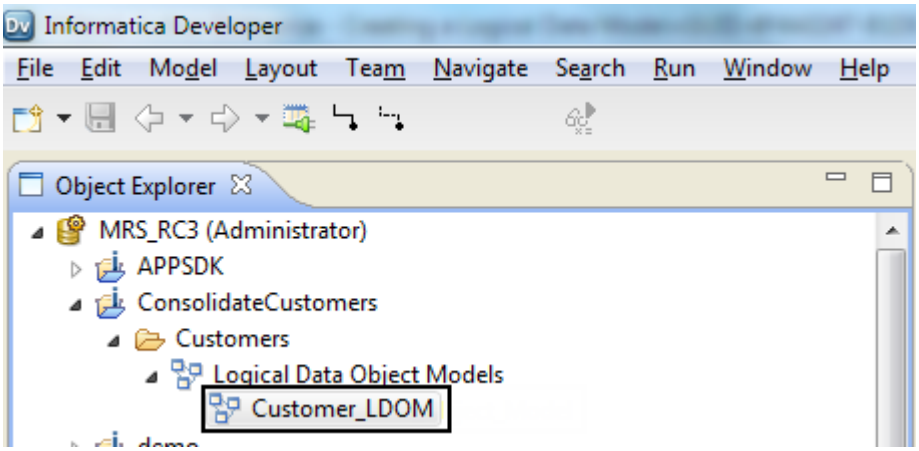
   The **ConsolidateCustomers** project appears in the **Object Explorer** view.
2. Click **File** > **New** > **Folder** to create a folder named **Customers**.
3. Select the **Customers** folder in the **Object Explorer** view, and click **File** > **New** > **Logical Data Object Model**.

   The **New** wizard appears.
4. Select the **Logical Data Object Model** wizard, and click **Next**.

   The **New Logical Data Object Model** screen appears.

5. In the **New Logical Data Object Model** screen, enter **Customer_LDOM** in the **Name** field. The Location field displays the location of the **Customers** folder. Click **Finish**.

   The **Customer_LDOM** appears in the **Object Explorer** view.
   The following image shows the **Customer_LDOM** logical data object model in the **Customers** folder and the logical data object model canvas:



Creating a Logical Data Object

Create a logical data object within a **Customer_LDOM** logical data object model. Add a logical data object read mapping and attributes to the logical data object. You can map the logical data object to one or multiple physical data objects.

1. Click **File** > **New** > **Data Object**.

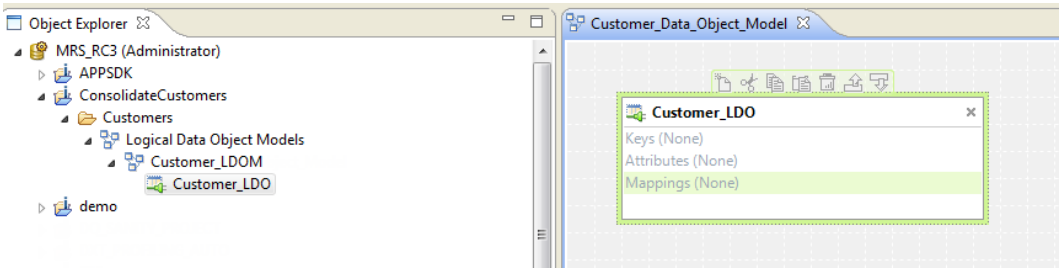
   The **New** wizard appears.

2. Select **Logical Data Object**, and click **Next**.

3. Enter the name of the data object as **Customer_LDO**. Click **Browse** to select the location of the **Customer_LDOM** logical data object model.

4. Click **Finish**.

   The **Customer_LDO** data object appears in the logical data object model canvas.
   The following image shows the **Customer_LDO** data object in the **Customer_LDOM** logical data object model canvas:



5. In the **Object Explorer** view, right-click the **Customer_LDO** data object. and click **Open**.

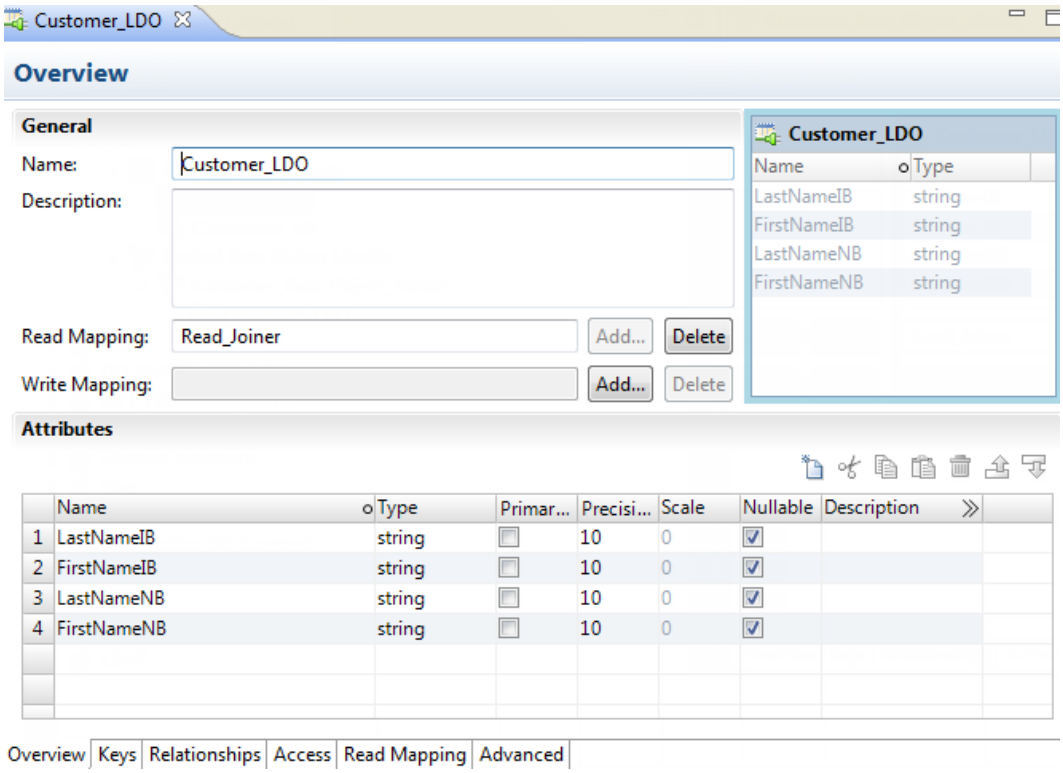
   The **Overview** tab appears.

6. In the **Overview** view, click **Add** in the **General** pane, to add a read mapping.

   The **Add Read Mapping** dialog box appears.

7. Enter a name for the mapping named **Read_Joiner**, and click **Finish**.
8. In the **Name** field in the **Attributes** pane, enter **LastNameIB**, **FirstNameIB**, **LastNameNB**, and **FirstNameNB**.
9. Click **File** > **Save** to save the attributes. Click **Finish**

   The following image shows the **Overview** tab for the logical data object:



Creating Flat File Data Objects

Create two flat file data objects and import the Customers data sources in the International Bank database and National Bank database.

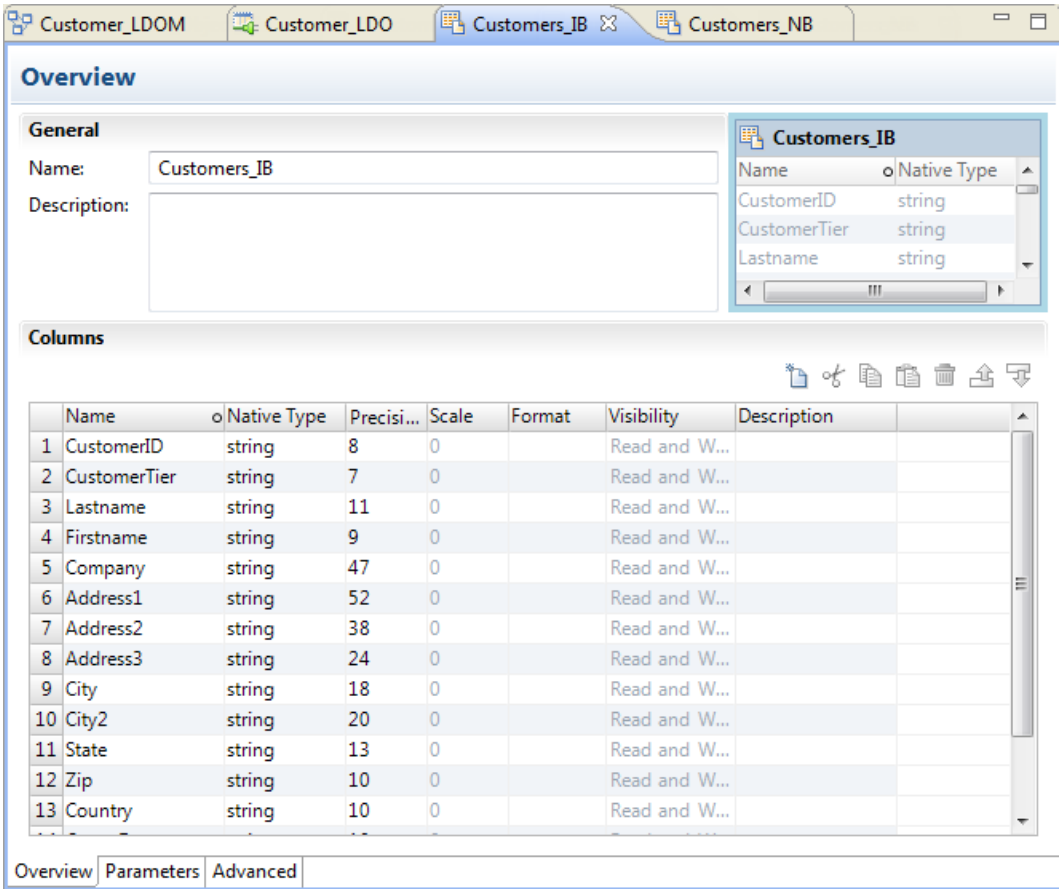
1. In Informatica Developer, click **File** > **New** > **Data Object**.

   The **New** wizard appears.
2. Select **Flat File Data Object**, and click **Next**.
3. Select **Create from an existing flat file** option, and click **Browse** to select the **Customers_IB** flat file from the folder where the file is located.
4. Click **Next**.
5. In the **Configure code page and format** dialog box, click **Finish**.

   The **Customers_IB** data object appears in the **Object Explorer** view.

6. Similarly, create a flat file data object for the **Customers_NB** flat file.

   The following image shows the **Customers_IB** and **Customers_NB** data objects in the **Object Explorer** view:



Creating a Joiner Transformation and Viewing the Transformation Results

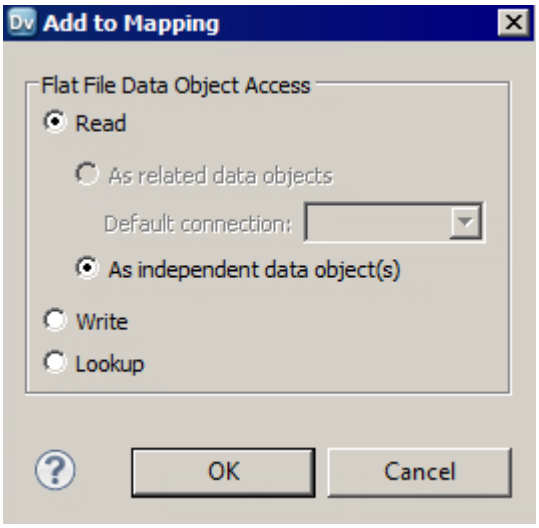
You can create a Joiner transformation to view all the customers residing in a particular ZIP code location. You must connect the ports from the physical data objects to the Joiner transformation, and connect the ports from the Joiner transformation to the logical data object. You can view the Joiner transformation results in the **Data Viewer** view.

1. In the **Object Explorer** view, select the **Customer_LDO** data object, and click **Read Mapping** tab.

   The **Read Mapping** view appears.

2. In the **Read Mapping** canvas, drag and drop the **Customer_IB** physical data object.

   The **Add to Mapping** dialog box appears.
   The following image shows the **Add to Mapping** dialog box:

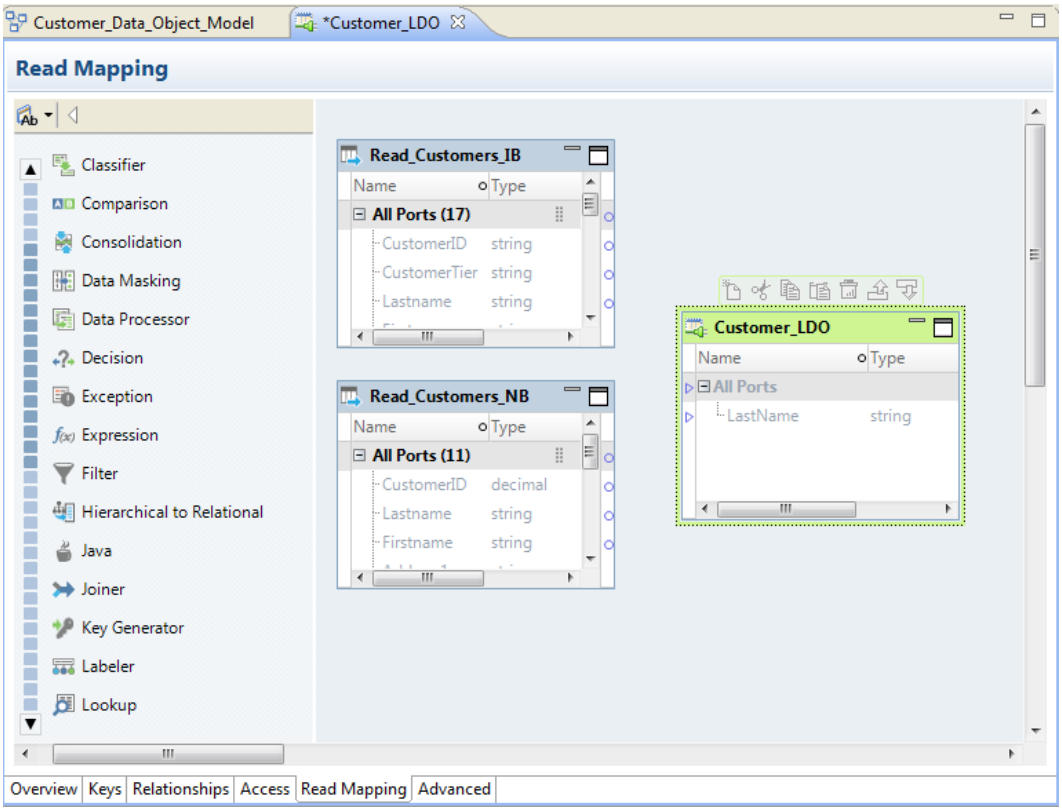


3. In the **Add to Mapping** dialog box, select **Read**, and click **OK**.

   The **Customer_IB** physical data object appears in the **Read Mapping** tab.

4. Similarly, drag and drop the **Customer_NB** physical data object into the **Read Mapping** tab.

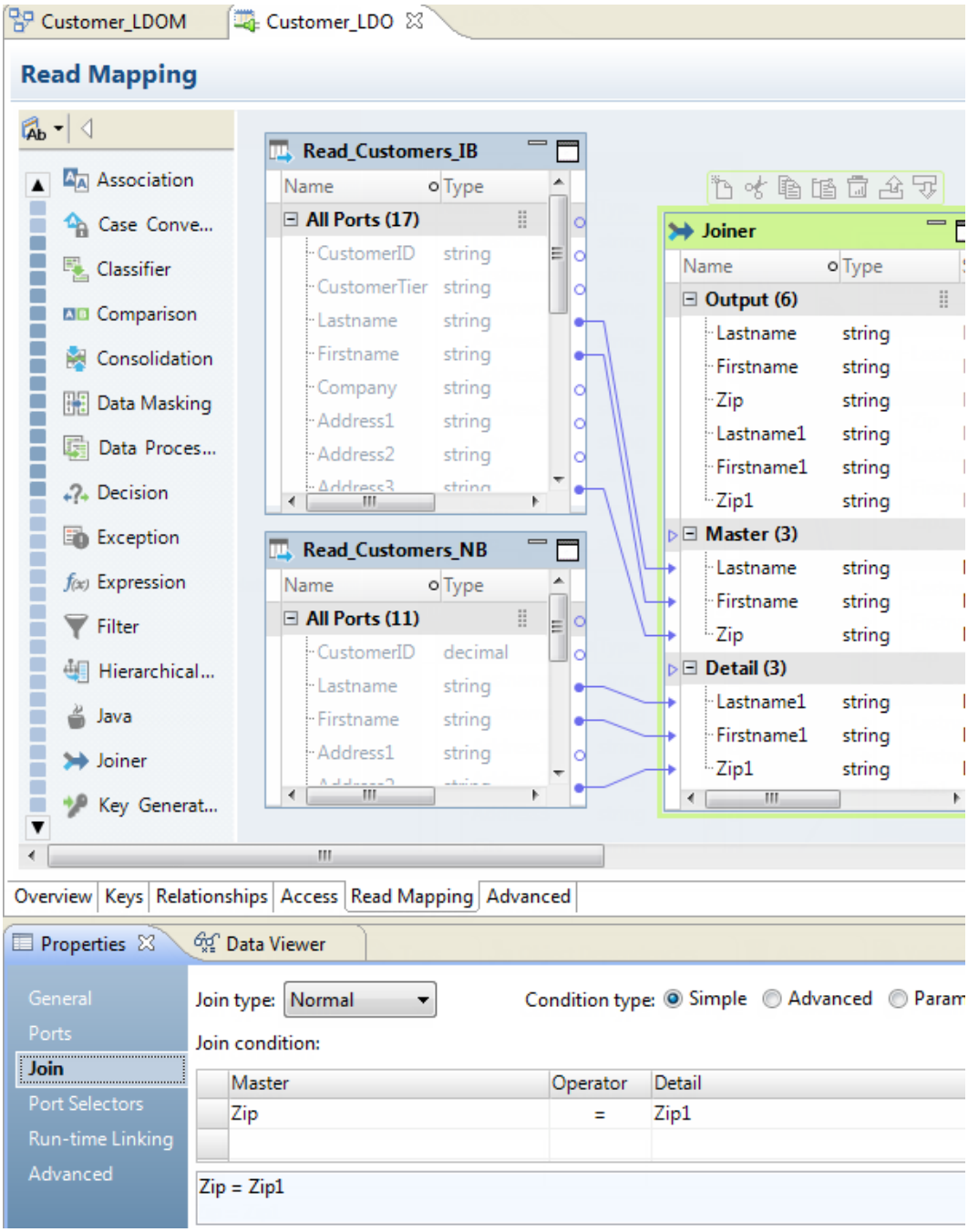
   The following image shows the **Customer_IB**, **Customer_NB**, **Customer_LDO** data objects in the **Read Mapping** tab.



5. In the **Read Mapping** tab, click **Joiner** transformation.
6. Drag the **Lastname**, **Firstname**, and **ZIP** in the **Read_Customers_IB** to **Master** in the **Joiner**. Similarly, drag the **Lastname**, **Firstname**, and **ZIP** in the **Read_Customers_NB** to **Detail** in the **Joiner** transformation.

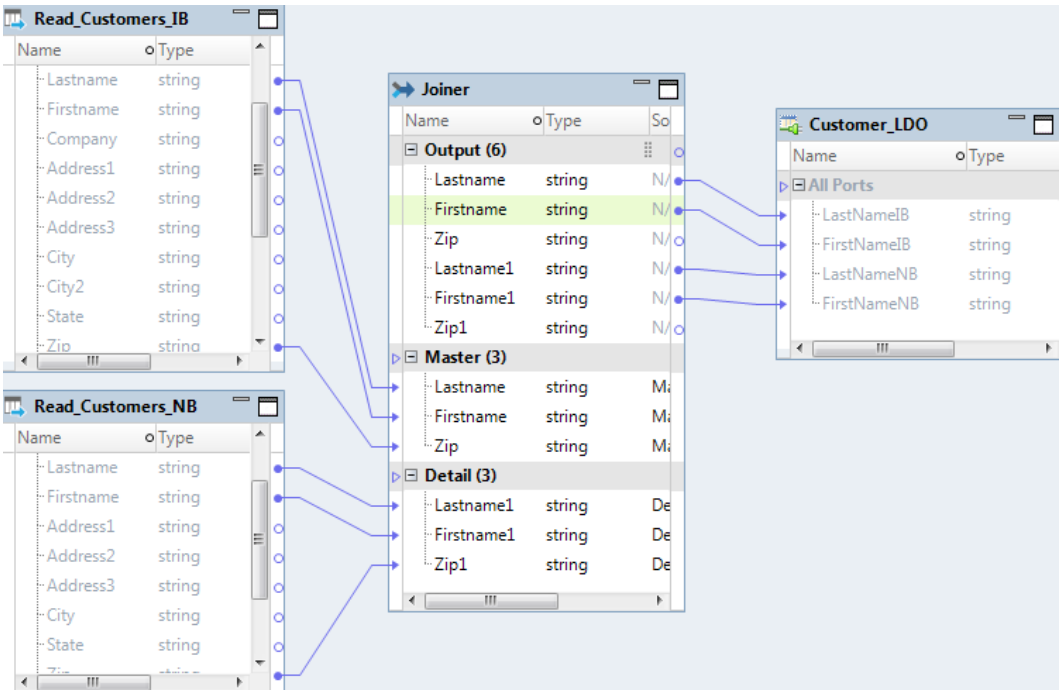
7. Click **Joiner** transformation, and click **Join** in **Properties** view.
8. Enter the join condition as `ZIP=ZIP1`.

   The following image shows the **Properties** view for the **Joiner** transformation with the join condition:



9. Drag the **Lastname**, **Firstname**, **Lastname1**, and **Firstname1** ports to the **Customer_LDO** data object.

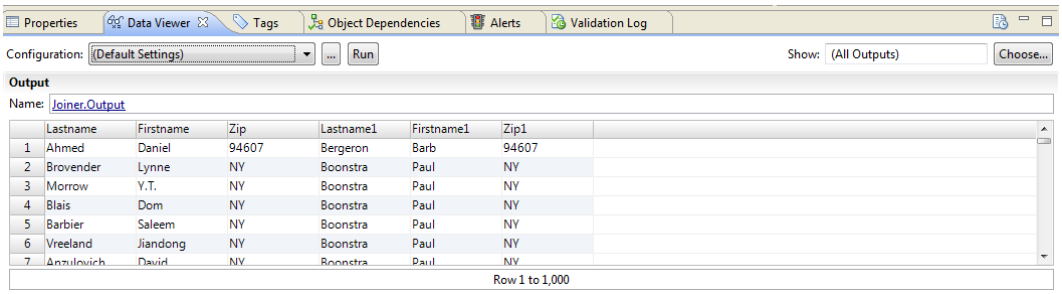
   The following image shows the Joiner transformation and the logical data object in the **Read Mapping** tab:



10. Right-click the Joiner transformation, and click **Run Data Viewer** to preview data.

    Informatica Developer displays the data in the **Output** section of the **Data Viewer** view. The **Output** section shows the combined data from the **Read_Customers_IB** source and **Read_Customers_NB** source.

    The following image shows the results for the join condition in the **Data Viewer** view:



Creating a Column Profile on the Logical Data Object

To view the data quality of the flat file data objects, you can create and run a column profile on the logical data object.

1. In the **Object Explorer** view, select the **Customer_LDO** data object, and click **File** > **New** > **Profile**.

   The **New** dialog appears.

2. Select **Profile** wizard, and click **Next**.

   The **Single Data Object Profile** screen appears.

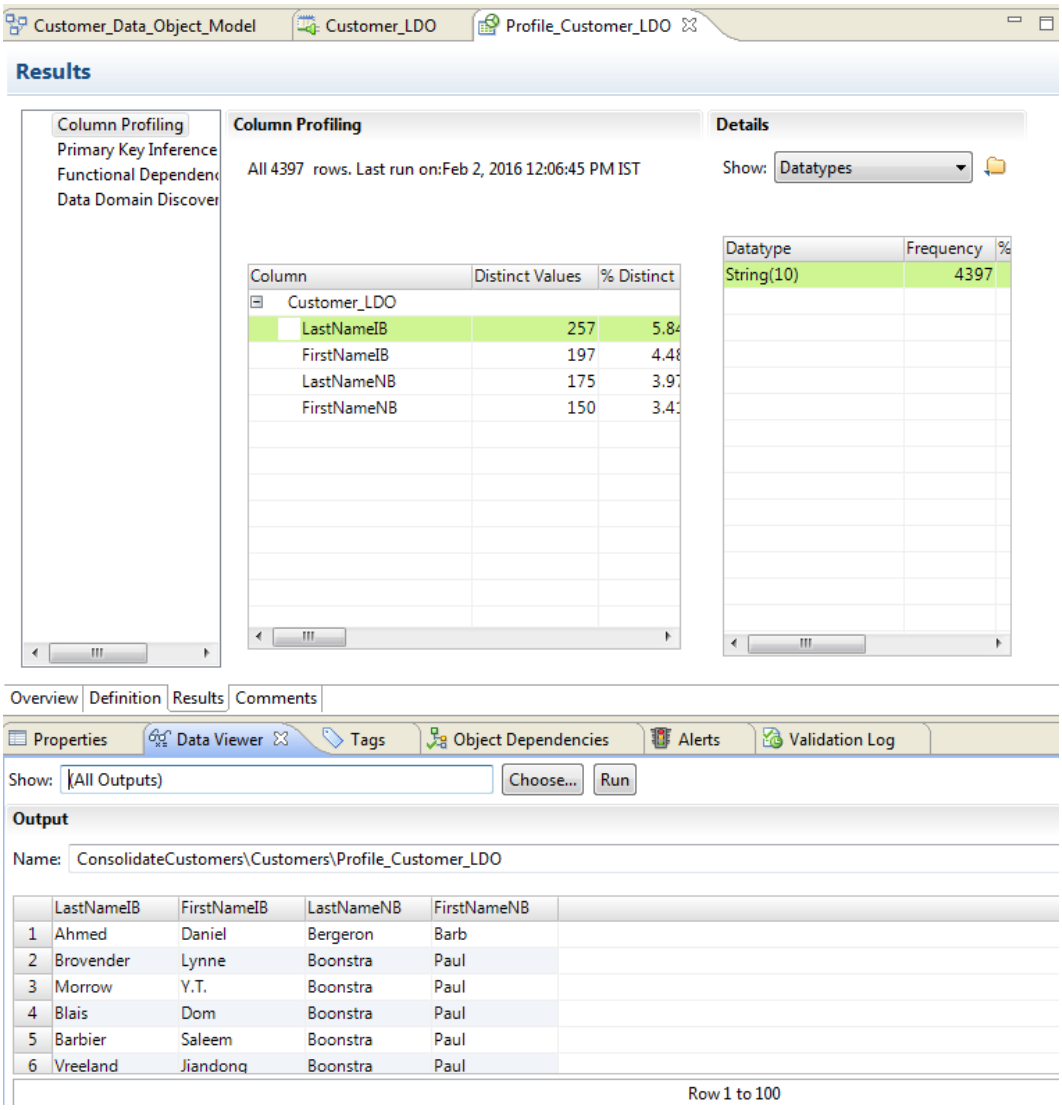
3. Enter **Profile_Customer_LDO** in the Name field, and click **Finish**.

   The **Overview** tab appears.

4. In the **Data Viewer** view, click **Run** to run the profile.

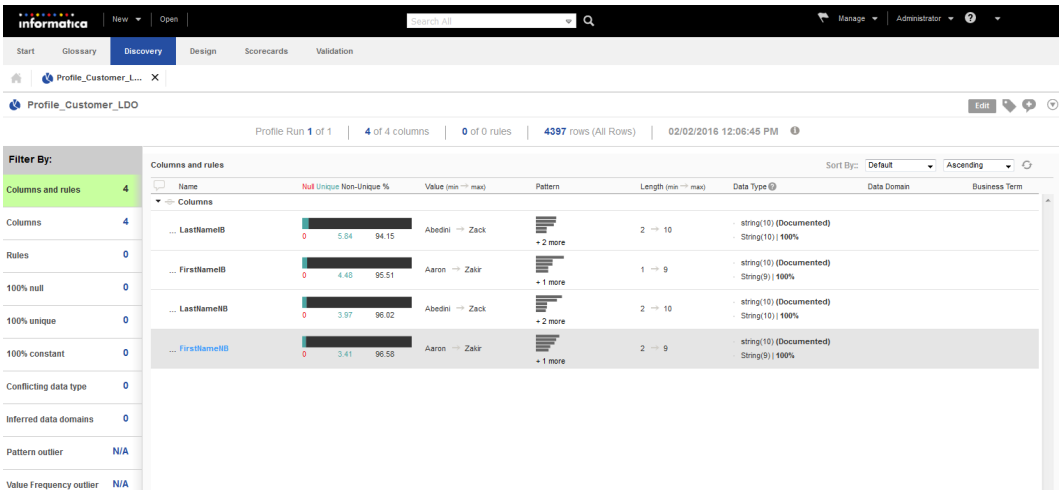
   The profile results appear in the **Results** tab.

The following image shows the profile results in the **Results** tab in Informatica Developer:



5. To view the profile results in Informatica Analyst, you can perform one of the following tasks:
   - In Informatica Developer, right-click the **Profile_Customer_LDO** profile, and click **Open with** > **Informatica Analyst**.
   - Log in to Informatica Analyst. In the **Library** workspace, click **Assets** > **Profiles** > **Profile_Customer_LDO**.

   The profile results appear in summary view.
   The following image shows the profile results in summary view in Informatica Analyst:



The profile results display all the customers residing within a ZIP code location and the data quality issues pertaining to the selected records.

Authors

Lavanya S