IBM InfoSphere FastTrack
Version 9 Release 1

Tutorial
Before using this information and the product that it supports, read the information in "Notices and trademarks" on page 39.
## Contents

**Tutorial: IBM InfoSphere FastTrack**  
*Facilitated mapping creation*  
**1**  
Module 1: Set up the tutorial environment  
Lesson 1.1: Install the tutorial files  
Lesson 1.2: Load the tutorial data  
**3**  
Module 2: Extract customer information from the  
BANK1 database  
Lesson 2.1: Create a project  
Lesson 2.2: Add the target tables to the mapping  
specification  
Lesson 2.3: Discover matching source columns  
Lesson 2.4: Set the unmapped columns to null  
Lesson 2.5: Filter out nonvalid records  
Lesson 2.6: Generate a job  
Lesson 2.7: Review your work in the IBM  
InfoSphere DataStage and QualityStage Designer  
Lesson 2.8: Compile and run the job  
**7**  
Module 2 summary  
**8**  
Module 3: Extract customer information from the  
tables in the BANK2 schema  
Lesson 3.1: Import the mapping specification  
Lesson 3.2: Review and customize the mapping  
specification  
Lesson 3.3: Add a join to consolidate data from  
two tables  
Lesson 3.4: Add a lookup operator to associate  
demographic data with customer information  
Lesson 3.5: Add a transformation rule to  
aggregate account balances  
Lesson 3.6: Create a job that combines two  
mapping specifications  
Lesson 3.7: Compile and run the job  
Module 3: Summary  
**17**  
Module 4: Standardize information, use business  
terms, and create data extractions for customer  
marketing and service  
Lesson 4.1: Import and create a mapping  
specification for standardizing and cleansing data  
Lesson 4.2: Describe the mapping specification  
and review it  
Lesson 4.3: Import and create a mapping  
specification for identifying high-value customers  
Lesson 4.4: Identify the high-value customers in  
the mapping specification  
Lesson 4.5: Associate business terms with the  
BANK.MASTER table  
Lesson 4.6: Create a business term and publish  
all terms  
Lesson 4.7: Create extractions for marketing and  
customer service  
Lesson 4.8: Define conditions that populate the  
subtables  
Lesson 4.9: Generate your job  
Lesson 4.10: Modify, compile, and run your job  
Lesson 4.11: Create a report to review details  
about the mapping specification  
Module 4: Summary  
**31**  
Module 5: Integrate data from Bank 3  
Lesson 5.1: Import the mapping specification  
Lesson 5.2: Complete the mapping specification  
Module 5: Summary  
**47**  
Appendix A. Product accessibility  
**51**  
Appendix B. Reading command-line  
syntax  
**53**  
Appendix C. Contacting IBM  
**55**  
Appendix D. Accessing and providing  
feedback on the product  
documentation  
**57**  
Notices and trademarks  
**59**  
Index  
**63**

© Copyright IBM Corp. 2007, 2012
Tutorial: IBM InfoSphere FastTrack Facilitated mapping creation

In this series of modules, you will use IBM® InfoSphere® FastTrack to create an application that identifies customers with high value to your business. You will experience how InfoSphere FastTrack increases the ease and efficiency by which you create mapping specifications. From the specifications, you can generate InfoSphere DataStage® and QualityStage® jobs and reports.

To access the sample data associated with this tutorial, go to
<InformationServer_installed_location>/clients/fasttrack/samples.

For this tutorial, you will follow a fictional scenario about the First Midwest company. First Midwest is a financial institution that grew by acquisition. As First Midwest focused on acquisition activities, their competition pulled away several of their high-value customers. These customers were neglected or not treated preferably at First Midwest. Now the priority of First Midwest is to regain its customer base by improving its view of customer data. First Midwest defines two levels of high-value customers. Members who hold more than $100,000 in assets are gold level. Members who hold more than $300,000 in assets are platinum level. First Midwest wants to ensure that gold customers are offered new investment opportunities and that platinum customers are given premium customer service when they call in with issues.

First Midwest has these subsidiaries:

**BANK 1**

Holds only checking accounts. The data is in one table: BANK1.CHECKING. The table has a number of nonvalid accounts that are identified by the account balance = -999999.

**BANK 2**

Holds checking and savings accounts. The data is in these tables: BANK2.CUSTOMERS and BANK2.ACCOUNTS. Customers might have checking and savings accounts. Therefore, account balances for both checking and savings accounts must be aggregated to compute the total account balance for a customer. Bank 2 also keeps track of demographic data about customers in a separate table, BANK2.DEMOGRAPHICS.

**BANK 3**

Holds only savings accounts. The data is in one table: BANK3.SAVINGS.

The following steps illustrate the sequence of actions:

1. First Midwest created a standard customer database that its subsidiaries use to represent customer data. You will access this database in Module 1, in Lesson 1.2.

2. Through the modules, you create specifications and build an application to identify gold and platinum customers for marketing and customer service.

3. The application moves customer information from the bank subsidiaries into a standardized customer model.

4. The standardized information is used to build information about platinum customers for the customer service department and information about gold customers for marketing.
5. You then integrate the customer data from the BANK 3 subsidiary into the database containing the standardized customer information. These actions are reflected in Figure 1.

---

**Figure 1. First Midwest subsidiaries and the plan for how the data flows.**

These tasks are required to build the application that identifies high-value customers:

1. Extract customer information from the BANK 1 subsidiary and map it to the BANK.CUSTOMERS table in the bankdemo database.
2. Extract customer information from the BANK 2 subsidiary and map it to the BANK.CUSTOMERS table in the bankdemo database.
3. Identify gold customers as level B and platinum customers as level A. Standardize customer name and address information and add a business term that defines a level of service.
4. Move gold customer data appropriate for marketing (such as name, address, and gender) from the bankdemo.BANK.CUSTOMERS table to the bankdemo.BANK.MARKETING table and move platinum customer data appropriate for customer service (such as name, address, and tax ID) from the bankdemo.BANK.CUSTOMERS table to bankdemo.BANK.CUSTSERVICE table.

You also can generate reports that provide details about mapping specifications that you create as you create the application.
Learning objectives

By completing the modules, you will learn about these functions:

- Create mapping specifications that map data from the source to the target tables.
- Generate jobs that are used to build applications
- Generate reports to view mapping specifications statistics and characteristics

Time required

In the first module, you set up your environment, and the time required depends on your current environment. The remaining modules each take about 40-50 minutes to complete.

System requirements

The following components and applications must be installed on your system. In Module 1, you prepare for the tutorial.

- InfoSphere FastTrack
- InfoSphere Information Server with the following clients:
  - InfoSphere DataStage and QualityStage Designer
  - InfoSphere DataStage and QualityStage Director
  - InfoSphere DataStage and QualityStage Administrator
- InfoSphere Business Glossary
- IBM DB2® Version 9.7 for Linux, UNIX and Windows

Module 1: Set up the tutorial environment

You must prepare your system to run the tutorial. The complete IBM InfoSphere FastTrack consists of a server component and a client component.

Learning objectives

After completing this module, you will be able to run the tutorial.

Time required

The time needed to complete the setup depends on the overall performance of your system and the software that you already installed.

Prerequisites

- To install both the InfoSphere FastTrack server component and the InfoSphere FastTrack client component, you need approximately 4 GB of RAM on the computer where you install IBM InfoSphere Information Server. A workstation with 2 GB of RAM is adequate for the InfoSphere FastTrack client component.
- These instructions and the tutorial assume a Windows operating system:
  1. Complete the installation instructions for InfoSphere Information Server, which include the installation of InfoSphere FastTrack.
     - When you install InfoSphere Information Server, you also install IBM DB2 V9.7 and IBM IBM WebSphere® Application Server. Ensure that WebSphere Application Server is installed at X:\IBM\WebSphere\AppServer, where X is the specific drive.
  2. Perform the configuration procedures in WebSphere Application Server.
3. Complete the IBM InfoSphere DataStage user credentials steps.

4. Ensure that your IBM InfoSphere Business Glossary credentials are also enabled correctly:
   a. From the InfoSphere Information Server Web console, select the Administration tab.
   b. From the Navigation pane, select Users and Groups > Users.
   c. If your InfoSphere Information Server account name is not displayed in Select Users to Work With, specify your account name in the search field and click Search. Select your account name and click Open.
   d. Expand Suite Component under Roles. Ensure that Business Glossary Administrator, Business Glossary Author, and Business Glossary User are selected in the list under Suite Component. Also ensure that Common Metadata Administrator and Common Metadata User is selected. Click Save and Close.

5. Compilers must be installed:
   b. Microsoft Platform Software Development Kit (SDK), to complete the Visual C++ download.

6. To enable the compiler to find the SDK library and include files, update some system environment variables after you install the Microsoft Platform SDK.
   a. From the Control Panel, click System. Click the Advanced tab and click Environment Variables > System Variables.
   b. Set the LIB variable to the location of the lib directory for the SDK. A typical location is C:\Program Files\Microsoft Platform SDK for Windows Server 2003 R2\Lib.
   c. Set the INCLUDE variable to the location of the include directory for the SDK. A typical location is C:\Program Files\Microsoft Platform SDK for Windows Server 2003 R2\Include.
   d. Save your settings and restart your computer.

Lesson 1.1: Install the tutorial files

Install the files to use during the tutorial.

To install the software:

Extract the tutorial.zip file to a directory such as C:\. You can access the setup files at <the installed location of IBM InfoSphere Information Server>\clients\fasttrack\samples (for example, C:\IBM\InfoSphere\Clients\FastTrack\samples). You then have a set of sample working files in C:\tutorial\setup, where C:\ is the directory in which you extracted the tutorial.

Lesson 1.2: Load the tutorial data

Load data into DB2 tables, terms into IBM InfoSphere Business Glossary, and prepare the IBM InfoSphere DataStage environment.

To load the tutorial data:
1. Create the DB2 database named bankdemo that is accessed by the tutorial.
   a. From your Windows Start menu, click Run.
   b. Type db2cmd.
c. Change the directory to C:\tutorial\setup.
d. Type the following commands:
   
   - `db2 create database bankdemo`
   - `DB2SET DB2CODEPAGE=1252`
   - `db2 -tvf bankdemodata.db2`

   The sample tables are loaded with the following number of rows:

<table>
<thead>
<tr>
<th>DB2 Schema and Table name</th>
<th>Number of rows at setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANK1.CHECKING</td>
<td>201</td>
</tr>
<tr>
<td>BANK2.CUSTOMERS</td>
<td>204</td>
</tr>
<tr>
<td>BANK2.ACCOUNTS</td>
<td>206</td>
</tr>
<tr>
<td>BANK2.DEMOGRAPHICS</td>
<td>204</td>
</tr>
<tr>
<td>BANK3.SAVINGS</td>
<td>201</td>
</tr>
<tr>
<td>BANK.CUSTOMERS</td>
<td>0</td>
</tr>
<tr>
<td>BANK.CUSTSERVICE</td>
<td>0</td>
</tr>
<tr>
<td>BANK.MARKETING</td>
<td>0</td>
</tr>
<tr>
<td>BANK.MASTER</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Create a data source name (DSN) on the same computer that contains the IBM InfoSphere Information Server so that you can associate the DB2 database with IBM InfoSphere FastTrack.
   a. From the Windows Start menu, click **Start > Control Panel > Administrative Tools > Data Sources (ODBC)**.
   b. Click the **System DSN tab** and click **Add**.
   c. Select the type of DB2 ODBC driver that is installed, for example, IBM DB2 ODBC DRIVER. Click **Finish**.
   d. In the Data source name field, type bankdemo.
   e. Select bankdemo from the list to register for ODBC. Click **OK**.
   f. Select the bankdemo row that you just created. Click **Configure**.
   g. Type the DB2 user name and password for your computer.
   h. Click **Connect**. The following message is displayed: **Connection tested successfully**.
   i. Click **OK** on all of the current Control Panel windows.

3. Load InfoSphere Business Glossary with categories and terms.
   a. Open a web browser and enter a URL in the following format:
      
      http://<hostName>:<portNum>/bg
   b. Type the user name and password for your Information Server suite. Click **Enter**.
   c. Select the **Administration** tab.
   d. Expand **Import and Export** and select **Import XML**.
   e. Click **Browse** and navigate to BankBusinessTerms.xml.
   f. Click **Import**.

4. Create an InfoSphere DataStage and QualityStage project.
   a. From the Windows Start menu, click **Start > All Programs > IBM Information Server > IBM InfoSphere DataStage and QualityStage Administrator** and log in.
   b. From the pop-up dialog, ensure that the **Projects** tab is selected. Click **Add**.
5. Import the containers for the tutorial.
   a. From the Windows Start menu, click Start > All Programs > IBM Information Server > IBM InfoSphere DataStage and QualityStage Designer. Type in your credentials, then select DataStage project from the Project drop-down list.
   b. Click Import > DataStage Components (XML). Click the Browse icon and navigate to the C:\tutorial\setup\Bank2AccountBalanceAggregator.xml file. Click OK.
   c. Click Open to load the file into the Designer client. Click OK. If a message asks if you want to overwrite existing definitions, click Yes to All.
   d. Repeat Steps B and C above for the C:\tutorial\setup\StandardizeAndCleanse.xml and C:\tutorial\setup\StandardizeAndCleanDummy.xml files.

6. Import physical database metadata into the metadata repository.
   a. From the Windows Start menu, click Start > All Programs > IBM Information Server > IBM FastTrack. Enter the User Name and Password for InfoSphere Information Server. In the Server field, type the address of the computer where InfoSphere Information Server is installed and the port number, 9080.
      your.server.com:9080
   b. Click Metadata > Source Configuration on the Workspace Navigator, as shown.
      ![Figure 2. Connecting to the metadata.](image)
      c. From the Task list next to the table, click New Connection.
      d. In the Name field, type bankconnect. Select ODBCConnector from the Connector list.
      e. Type your DB2 administrator user name and password.
      f. Select bankdemo as the data source name from the Data Source drop-down list.
g. Click **Test Connection** to verify that the connection to the DB2 database is valid. Click **OK**. Click **Save and Close**.

h. In the Source Configuration tab, expand the **bankconnect > bankdemo** objects. Select the BANK, BANK1, BANK2, and BANK3 schemas by holding the shift key and clicking each one. From the Task list, click **Import Metadata**.

![Figure 3. Import metadata from several schemas as source.](image)

i. Select **New Host**. Type the name of the new host: bankconnect. Click **Next**. Select **New Database**. Type the name of the new database: bankdemo. Click **Finish**.

The following message is displayed: Importing metadata from... You do not need to do anything. When the import is complete, the window closes.

---

### Module 2: Extract customer information from the BANK1 database

In this module, you begin to consolidate relevant customer data into a table that follows the standard model of the company.

#### Learning objectives

In this module, you create a mapping specification and a job to extract customer information from the Bank 1 subsidiary and map that information to the BANK.CUSTOMERS table in the bankdemo database. At the completion of this module, you will know how to perform the following actions, as shown in [Figure 4](image):

1. Add target columns to a specification.
2. Find matching source columns.
3. Add transformation functions and a filter.
4. Generate, compile, and run a job.

![Figure 4. Extracting customer information from the BANK1.CHECKING table and mapping data to the BANK.CUSTOMERS table.](image)
Time required

This module takes approximately 40 minutes to complete.

Prerequisites

Ensure that you completed the setup steps in Module 1 and have tutorial files and data loaded on your computer system.

Lesson 2.1: Create a project

To start moving data into a consolidated location, you create an IBM InfoSphere FastTrack project first.

You must have administrative authorization to create an InfoSphere FastTrack project.

To create a project:
1. Log into the InfoSphere FastTrack client component.
2. From the menu, click Projects > New Project.
3. Type BankProject for the name of your new project.
4. Click Finish, then open the Mapping Specification workspace.

You now see the BankProject name on the Mapping Specifications list. Refer to Figure 5.

![Figure 5. The new BankProject in the Mapping Specifications list.](image)

Lesson 2.2: Add the target tables to the mapping specification

In this lesson, you put the target table BANK.CUSTOMERS, into a new mapping specification. Then you will match the target columns to the appropriate source columns from the BANK1.CHECKING table.

To create the mapping specification and add a target table to the specification:
1. Provide a name and description for your mapping specification. See Figure 6 on page 9.
a. If you closed the project from the previous lesson, from the IBM InfoSphere FastTrack menu bar, click Projects > Open Project. Click BankProject from the previous lesson.

b. Select the Mapping Specifications folder, and click New in the Task list. The Mapping Editor opens to the Overview.

Tip: If you do not see the Mapping tab, click View > Mapping Assets. Then, when the Mapping tab is displayed, select the BankProject project.

c. Type Bank 1 Extract in Name.

d. Type Extracts customer information from BANK1 in Description. Leave Status as In Progress. See Figure 6.

e. Optional: Select an owner from the Owner list. This is reflected in the jobs that you generate.

f. Click Save.

2. From the Mapping Editor list on the left, click Mappings.

3. Click the Browser tab on the right side of the view. Expand the Database metadata folder, and you will see a list of data sources, databases, schemas, and tables.

Tip: If you do not see the Database Metadata tab, click View > Browser.

4. Expand the bankconnect > bankdemo > BANK objects. Under the BANK schema, you see the tables that were imported in Module 1.

5. Drag and drop the BANK.CUSTOMERS table in the Target Fields column of the mapping specification table.

Figure 6. Provide a name and description for the specification.
When you drop the CUSTOMERS table into the mapping specifications table, all 13 columns within the CUSTOMERS table are listed.

6. In the Database Metadata tab, go back to the bankdemo database and expand BANK1 to see the CHECKING table.

7. Match the following columns under the BANK1.CHECKING table to the appropriate target by selecting each column and dragging it to the appropriate cell under Source Columns. You can see the resulting mapping specification in Figure 7.

Table 1. Columns in BANK1.CHECKING and BANK.CUSTOMERS

<table>
<thead>
<tr>
<th>Source Field columns from BANK1.CHECKING</th>
<th>Target Field columns from BANK.CUSTOMERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHECKING.ACCOUNT_BALANCE</td>
<td>CUSTOMERS.ACCOUNT_BALANCE</td>
</tr>
<tr>
<td>CHECKING.ADDR1</td>
<td>CUSTOMERS.ADDR1</td>
</tr>
<tr>
<td>CHECKING.ADDR2</td>
<td>CUSTOMERS.ADDR2</td>
</tr>
<tr>
<td>CHECKING.CITY</td>
<td>CUSTOMERS.CITY</td>
</tr>
<tr>
<td>CHECKING.CUSTOMER_ID</td>
<td>CUSTOMERS.CUSTOMER_ID</td>
</tr>
<tr>
<td>CHECKING.NAME</td>
<td>CUSTOMERS.NAME</td>
</tr>
<tr>
<td>CHECKING.STATE</td>
<td>CUSTOMERS.STATE</td>
</tr>
<tr>
<td>CHECKING.ZIP</td>
<td>CUSTOMERS.ZIP</td>
</tr>
</tbody>
</table>

Figure 7. Source and Target Field columns in the mapping specification.

You will work with the following columns in a later lesson:

CUSTOMERS.ONLINE_ACCESS
CUSTOMERS.YEARS_CLIENT
CUSTOMERS.GENDER
CUSTOMERS.LEVEL
CUSTOMERS.TAX_ID
Lesson 2.3: Discover matching source columns

You can discover and match source columns. For example, you need to find a match for the column CUSTOMERS.TAX_ID since it does not have an exact match in the tables of the BANK1 schema.

To find matching source columns for the mapping specification:

1. Scroll to the row that contains CUSTOMERS.TAX_ID in the Target Field column in the Mapping Editor.
2. Click the Columns cell under the Source header in the same row to select the cell.
3. Below the table in the Mapping Editor, click the Discover tab, then click Discover More.

   Tip: If you do not see the tabs below the table, click the arrow that is between the table and buttons at the bottom of the Mapping Editor. See Figure 8.

4. In the Discover More window, expand the bankconnect > bankdemo > BANK1 schema. Click the check box next to BANK1.
   By selecting certain objects, you are defining the scope of the search. When you select a single object, such as the BANK1 schema, you decide to narrow the scope to a single schema, which contains only certain tables. The larger the scope, the more work that the discover function can do.
5. Click OK.
   The Discover process takes all of the information that is available from the Target Field column and searches within the BANK1 schema for relevant source columns.
   The message Results available is displayed in the selected Source Field cell.
6. Examine the results of the discovery that appear in the Discover view. The column BANK1.CHECKING.SS_NUM is identified as a good match for TAX_ID, which is rated with 95% confidence.

Tip:
Select SS_NUM and click Match Details, to see why SS_NUM is a good match.
As you can see in Figure 9 on page 12 because you defined terms in the InfoSphere Business Glossary in “Lesson 1.2: Load the tutorial data” on page 4, the synonyms and abbreviations for TAX_ID were made available to the Discover process.

7. Select SS_NUM and click Add to Sources. Click Save.

Lesson 2.4: Set the unmapped columns to null

For this mapping specification, because there is no demographic data from the tables in the BANK1 schema, you indicate that some columns in the BANK1.CHECKING table must be set to null by means of the function setNull().

The remaining columns in the BANK.CUSTOMERS table do not have matches on the source table, BANK1.CHECKING. You will work with the BANK.CUSTOMERS.LEVEL column when you identify gold and platinum customers. The BANK.CUSTOMERS.YEARS_CLIENT, BANK.CUSTOMERS.GENDER, and BANK.CUSTOMERS.ONLINE_ACCESS columns contain demographic data not contained in the BANK1 schema. However, demographic data is collected by the set of tables in the BANK2 schema and therefore is on the BANK.CUSTOMERS table. You will want to retain the demographic data for marketing purposes.

To add a function that sets the unmapped columns to null:

1. In the table that is in the Mapping Editor, click the expander icon to the right of the Transformation column, as shown here:

![Transformation](image)

The Rule Expression column is displayed.

2. Type setNull() in the Rule Expression column next to the following columns:
Lesson 2.5: Filter out nonvalid records

The tables in the BANK1 schema have several nonvalid accounts that are identified by an account balance of -999999. You add a filter to the mapping specification to filter out nonvalid accounts.

The filter applies to all account records in the BANK1.CHECKING table. Therefore, you specify the filter at the table level instead of at the column level.

To add a filter to the mapping specification to filter out nonvalid record:

1. In the Mapping Editor, click Properties from the list on the left.
2. Under Source Tables, select the BANK1.CHECKING table.
3. Click the Filter tab.
   
   **Tip:** If you do not see the tabs below the table, click the arrow between the table and buttons at the bottom of the Mapping Editor.

4. In the Rule Expression field, type the expression: `ACCOUNT_BALANCE <> -999999`.
   You do not need to type a Filter name.

   **Tip:** If you enable validation, IBM InfoSphere FastTrack checks the validity of the column name in your expression. It can also check for simple expression syntax. For example, if you type `ACCOUNT_BALNCE <> -999999`, validation checking determines that the filter expression is incorrect. If you hover over the expression, you can see the message Variable "ACCOUNT_BALNCE" could not be resolved. When you correct the column name, the expression color changes from red to green.
Lesson 2.6: Generate a job

You generate jobs in IBM InfoSphere FastTrack. These jobs create your application.

To generate a job:
1. In the Mapping Editor, click Generate Job.
2. In the Name of New Job field, specify this name for the job:
   Bank_1_Extract_bankdemo.
3. From the DataStage and QualityStage Project or Folder list, select a project or folder within the InfoSphere DataStage and QualityStage project as the target for your job. The only projects that are visible in this window are those that you have authorization to access. You defined a project named DSProject in  
   [Lesson 1.2: Load the tutorial data on page 4]. Expand that project and select a folder to contain the job. The default folder is jobs. For the purposes of this tutorial, assume the folder name is jobs.
4. The Use Shared Table Definitions check box is selected by default. Clear the check box. For this tutorial, you are not going to associate your job with tables that are already defined in the metadata repository.
5. Click Finish.
6. Close the mapping specification.

Lesson 2.7: Review your work in the IBM InfoSphere DataStage and QualityStage Designer

To complete the consolidation of data that you started in IBM InfoSphere FastTrack, you use the InfoSphere DataStage and QualityStage Designer client to review your generated jobs.
To review your generated job:

1. Click **Start > All Programs > IBM InfoSphere > IBM InfoSphere DataStage and QualityStage Designer** and log in to the DSProject. Click **Cancel** on any windows that open.

2. The DSProject project is on the left navigation pane. Expand the **Jobs** folder within that project.

3. Double-click the Bank_1_Extract_bankdemo job to see these stages: CHECKING -> Filter_001 -> CUSTOMERS.

4. Double-click the Filter_001 stage. You can see how it filters out non-valid accounts through the link constraint.

5. On the right side of the Filter_001 stage, you can verify that the setNull() function is applied in the derivation column.

6. Click CUSTOMERS_TAX_ID to see that the column is mapped to the CHECKING_SS_NUM column.

7. Click **Cancel** to get back to the palette for the job, Bank_1_Extract_bankdemo.

**Figure 13. The generated job from the mapping specification, Bank 1 Extract.**

Lesson 2.8: Compile and run the job

After you review your generated job, you compile the job in the IBM InfoSphere DataStage client, and then run the job to consolidate the data in the BANK.CUSTOMERS table.

To compile and run a job:

1. On the palette for the Bank_1_Extract_bankdemo job, open each ODBC stage and type the password to connect to the data source.

2. On the toolbar, click the **Compile** icon.

When the compilation completes successfully, a window displays the message **Job compiled successfully with no errors.**

3. Click **Close**.

4. Click the **Run** icon.
5. Return to the palette. Look at the text near DSLink 3. One hundred and ninety-eight rows were added to the BANK.CUSTOMERS target table when the run completes.

**Tip:** To diagnose your job if there is a problem, use the InfoSphere DataStage and QualityStage Director client.

a. Click **Tools > Run Director**.
b. Select the Bank_1_Extract_bankdemo job from the list. A message indicates whether the job completed or failed.
c. Right-click and select **View Log** to see details of the job. If the job completed successfully, you see an Event: Finished Job Bank_1_Extract_bankdemo message. If not, you see an error message that you can use to diagnose the problem.

**Module 2 summary**

You consolidated customer data from the tables in the BANK1 schema to the BANK.CUSTOMERS table, that is based on the customer modeling scheme established by First Midwest.

**Lessons learned**

By completing this module, you learned how to do the following activities:

- Create a project.
- Create mappings.
- Discover related metadata.
- Add a function to a specification.
- Add a filter to a specification.
- Compile and run an IBM InfoSphere DataStage job.

**Module 3: Extract customer information from the tables in the BANK2 schema**

The customer information that First Midwest wants to integrate into its banking system is in multiple tables. In this module, you perform a join, add source columns based on lookup operators, and define business rules before you extract the customer information.

The goal for First Midwest is to identify its gold and platinum customers so that these high-value customers can be offered the appropriate services. Customers might have both checking accounts and savings accounts that need to be accessed so that the high-value customers can be tracked accurately.
Remember: The gold customers (accounts with balances greater than $100,000.00, but less than $300,000.00) get marketing updates. The platinum customers (accounts with balances of $300,000.00 or greater) get special customer service.

In Module 2, you extracted data from tables in the BANK1 schema and created a mapping specification that put pertinent data in a new table called BANK.CUSTOMERS.

In this module, you continue the process of consolidating relevant customer data into a table that follows the standard schema of First Midwest:

1. Create a mapping specification and a job to extract customer information from tables in the BANK2 schema and map that information to the BANK.CUSTOMERS table. The BANK2 schema has these tables:
   - BANK2.CUSTOMERS
   - BANK2.ACCOUNTS
   - BANK2.DEMOGRAPHICS
2. Join the BANK2.CUSTOMERS and BANK2.ACCOUNTS tables and then map those customers and accounts to the BANK.CUSTOMERS table.
   You use the demographics data in tables in the BANK2 schema to add the personal information of the customer to the final BANK.CUSTOMERS table.
3. Combine the mapping specification from Module 2 with the mapping specification in Module 3 to create standardized customer information. See Figure 14.

Learning objectives

After completing the lessons in this module, you will know how to do the following actions:

- Apply a lookup operator to append the customer information with demographic information such as social security number, the number of years as a client, customer gender, and access to online systems.
- Complete the mapping specification and then combine your new specification with the specification that was created in the second module, the Bank 1 Extract specification.
- Generate a job that combines data from tables in the BANK1 and BANK2 schemas.
- Add a business rule that describes an InfoSphere DataStage job aggregate operator to compute the total account balance for each customer from tables in the BANK2 schema.

**Time required**

This module takes approximately 40 minutes to complete.

**Prerequisites**

Ensure that you completed the lessons in Module 2 and have access to the mapping specification named Bank 1 Extract.

**Lesson 3.1: Import the mapping specification**

You start this module by using a mapping specification that was already populated with metadata.

To extract customer information from the tables in the BANK2 schema and map that information to the BANK.CUSTOMERS table, begin by importing a mapping specification. Rather than start by creating the mapping specification as you did in Module 2, you can save time by importing a partially completed mapping specification.

To import a mapping specification:

1. Optional: If you are not already in the project, from the IBM InfoSphere FastTrack menu, click **Projects > Open Project**. From the Open Project window, select BankProject and click **OK**.
2. Click **Mapping** from the menu to open the list of current mapping specifications that is displayed in the **Mapping** tab. Select the Mapping Specifications folder, then click **Import from CSV**.

   **Tip:** If you cannot see the **Task** list, look in the bottom left corner of the editor for the minimized Mapping Specification. Click on the minimized page to reopen.

3. In the **Import from CSV** tab, in the **Location** field, browse to the '<Location of the extracted tutorial files>Tutorial\Bank2Extract.csv' file and open it.
5. Optional: You can click the **UTF-8 Encoding** check box to ensure that the file is imported in the UTF-8 character set. By default, the file is imported in the encoding of the computer on which the InfoSphere FastTrack is installed.
6. Click **Next**. See [Figure 15 on page 20](#)
Twelve rows are accessed and ready to be imported. Some warning messages might say that the source columns cannot be resolved. You can ignore these messages because these are columns that you will resolve by defining a lookup operation.

7. Ensure that the Use First Row as Column Headers check box is selected and click Next.

8. For the purposes of this tutorial, there are no lookups in the CSV file that you are importing, therefore the Import Lookups view is disabled. If your CSV file had lookups, you could select Import Lookups and then define some lookup options. If you leave the check box blank, which is the default, you can define these lookups later. Click Next.

9. Check the Use Full Multipart Data Source Name check box. By default, the table and column names are associated with the full path, which includes the host name, the database name, and the schema name. Click Next to view the potential import. Your mapping specification will be populated with twelve rows.

Each column has an icon preceding it, indicating whether it is a resolved or candidate column. If the column selected for import is a candidate and cannot be resolved, you will receive one of the following warnings:

- The source column "table1.column1" could not be resolved. A candidate column was created.
- The target column "table1.column1" could not be resolved. A candidate column was created.
- The source term "table1.column1" could not be resolved. A candidate term was created.
- The target term "table1.column1" could not be resolved. A candidate term was created.

Ensure that the column is not a candidate. If it is, one of the following may have occurred:

- The metadata was not imported correctly.
- One or more of the options on the previous screen was not entered correctly (such as the metadata naming in the repository does not match the metadata naming in the file). Correct any discrepancies, then retry the import.

10. Click Finish to perform the import and to save the mapping specification. You are returned to the list of mapping specifications in BankProject.

You now have the Bank 2 Extract mapping specification in BankProject, with references to tables and columns from the BANK2 schema that includes BANK2.CUSTOMERS and BANK2.ACCOUNTS information. You have a defined target that contains a structure similar to the BANK2 structure. See Figure 16 on page 21.
Lesson 3.2: Review and customize the mapping specification

Now that you have a mapping specification that is partially built, you need to do a few things to customize the mapping specification for your customer needs.

To review and customize the mapping specification:
1. Click **Mapping** from the menu and click the imported specification, Bank 2 Extract.
2. From the **Task** list, click **Open**.
3. In the **Mapping Editor**, click **Overview** from the list on the left.
4. In the **Description** field, describe the specification, such as **Extracts customer information from Bank2** and click **Save**.
5. From the **Mapping Editor**, click **Mappings** from the list on the left.
   Many of the columns have a mapping. On the **Annotation** tab, some of the columns contain annotations that indicate additional work is required.
6. Scroll through the list of **Target Field** columns and click the **CUSTOMERS.LEVEL** cell.
   This column does not exist in the source table. On the **Annotation** tab, the description explains that this column will be computed later when you identify the high-value customers.
7. Scroll to the right of the **Target Field** column, **CUSTOMERS.LEVEL**. Click the expander arrow in the **Transformation** header to see the **Function** column. In the **Function** column for that row, specify the **setNull()** function, as you did in Module 2.
8. Click **Save**.

   You now have 12 rows with one **Target Field** column that contains a setNull() function.

Lesson 3.3: Add a join to consolidate data from two tables

To create a complete set of customer information from Bank 2 for your target table, you need to join the BANK2.CUSTOMERS table with the BANK2.ACCOUNTS table.

You use IBM InfoSphere FastTrack to help you find the right join key.
To add a join:

1. In the **Mapping Editor**, click **Properties** from the list on the left.
   
   The **Source Tables** field contains the tables bankconnect.bankdemo.BANK2.ACCOUNTS and bankconnect.bankdemo.BANK2.CUSTOMERS, and the **Target Tables** field contains the table bankconnect.bankdemo.BANK.CUSTOMERS.

2. Click the **Discover Join** tab and expand the **bankconnect > bankdemo** objects. Because you are interested only in the information that the BANK2 schema owns, you can narrow the scope to discover a valid join.

3. Select BANK2, and click **Discover Join**.
   
   In **Discovered Joins**, you see one discovered join between the CUSTOMERS.CUSTOMER_ID and ACCOUNTS.CUSTOMER_ID columns.

   **Tip**: You can preview the join results by clicking the **Sample Join** button. You must have access to the database. Provide your User name and Password in the authorization window. You then see a sampling of the returned rows that are based on the join. The number of rows that are returned depends on the number that you specified in Preferences. You can adjust that number at the time you sample the data.

4. Highlight the join and then click the **Select** button on the right to accept the join for your mapping specification.
   
   This operation adds the join operation to the mapping specification. See **Figure 17**.

5. Click **Save**.

![Figure 17. Discovery of the join key.](image)

**Lesson 3.4: Add a lookup operator to associate demographic data with customer information**

You define a lookup operator to associate BANK2 demographic data with BANK2 customer information. The result provides a more complete set of information of the customer accounts in the tables of the BANK2 schema.

The BANK2.CUSTOMERS and BANK2.DEMOGRAPHICS tables have the following information:
Table 2. Columns in BANK2.CUSTOMERS

<table>
<thead>
<tr>
<th>BANK2.CUSTOMERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADDR1</td>
</tr>
<tr>
<td>ADDR2</td>
</tr>
<tr>
<td>CITY</td>
</tr>
<tr>
<td>CUSTOMER_ID</td>
</tr>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>STATE</td>
</tr>
<tr>
<td>ZIP</td>
</tr>
</tbody>
</table>

Table 3. Columns in BANK2.DEMOGRAPHICS

<table>
<thead>
<tr>
<th>BANK2.DEMOGRAPHICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANKCARD</td>
</tr>
<tr>
<td>CUSTOMER_ID</td>
</tr>
<tr>
<td>GENDER</td>
</tr>
<tr>
<td>MARITAL_STATUS</td>
</tr>
<tr>
<td>ONLINE_ACCESS</td>
</tr>
<tr>
<td>PROFESSION</td>
</tr>
<tr>
<td>SAVINGS_ACCOUNT</td>
</tr>
<tr>
<td>SS_NUM</td>
</tr>
<tr>
<td>YEARS_CLIENT</td>
</tr>
</tbody>
</table>

Learn more about lookup operators: In InfoSphere DataStage and QualityStage, a lookup operator is used to retrieve more information about an entity from a lookup table, given a key.

To define a lookup operator:

1. From the Mapping Editor, click Mappings.
   a. In the table, click the row for the target column, CUSTOMERS.TAX_ID.
   b. On the Annotation tab, below the table, review the annotation for that column. It indicates that you need to define a lookup operator for the BANK2.DEMOGRAPHICS table to retrieve the social security information in that table.
   c. You will use this same lookup operator to review and return the DEMOGRAPHICS.YEARS_CLIENT, DEMOGRAPHICS.GENDER, and DEMOGRAPHICS.ONLINE_ACCESS columns.

2. In the Mapping Editor, click Lookup Definitions, and then click Add Lookup Definition.
3. In the Specify a Name for the Lookup Definition window, in the Name field type Demographic Data Lookup. Click OK.
   Demographic Data Lookup is entered in row 1.

4. Click the Tables tab.
5. Define the source table and then populate the columns with the lookup definitions.
   a. In the Browser, expand Database Metadata > bankconnect > bankdemo > BANK2.
b. Drag the DEMOGRAPHICS table to the **Lookup** field in the row that contains the lookup definition demographic data.

   c. Drag the CUSTOMERS table to the **Source Tables** field.

   d. Select the Demographic Data Lookup row.

6. Discover potential joins between the DEMOGRAPHICS table and CUSTOMERS table.
   a. Click the **Discover Join** tab.
   b. Expand the `bankconnect > bankdemo > BANK2` objects.
   c. Click the check box next to `BANK2` and click **Discover Join**.

      The result shows in the **Discover Joins** table. You can also see the message `Results available` is displayed in the Demographic Data Lookup row, in the `Join` column in the table.

   d. Select the result in the **Discovered Joins** table and click **Select** to add the join to the BANK2.CUSTOMERS table.

   e. Click **Save**.

7. Map the columns from the lookup operation.
   a. In the **Mapping Editor**, click **Mappings**.
   b. In the table, scroll to the row for the CUSTOMERS.TAX_ID **Target Field** column. Select the empty Source cell. Right-click the cell and select **Add Lookup Field**.
   c. From the window that is displayed, expand **Demographic Data Lookup > DEMOGRAPHICS > SS_NUM**.
   d. Click **OK**.
   e. Repeat the steps in the empty source cell next to each of the following target cells:

<table>
<thead>
<tr>
<th>Demographics lookup for Source field</th>
<th>Target field</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER</td>
<td>CUSTOMERS.GENDER</td>
</tr>
<tr>
<td>YEARS_CLIENT</td>
<td>CUSTOMERS.YEARS_Client</td>
</tr>
<tr>
<td>ONLINE_ACCESS</td>
<td>CUSTOMERS.ONLINE_ACCESS</td>
</tr>
</tbody>
</table>

   See [Figure 18 on page 25](#).

8. Click **Save**.
Lesson 3.5: Add a transformation rule to aggregate account balances

Bank 2 allows customers to have both savings and checking accounts. To compute the total account balance for customers from Bank 2, the InfoSphere DataStage and QualityStage developer computes the sum of the savings and checking account balances for that Bank 2 customer.

To add a transformation rule and aggregate account balances:

1. In the Mapping Editor, scroll to the row with CUSTOMERS.CUSTOMER_ID, CUSTOMERS.ACCOUNT_BALANCE in the Target Field column. Select the Rule cell under the Transformation header for the CUSTOMERS.ACCOUNT_BALANCE column.

   The columns in the Source Fields cell, ACCOUNTS.ACCOUNT_BALANCE and ACCOUNTS.CUSTOMERS_ID are the columns on which to aggregate the balances.

2. Type the following transformation rule in the Rule Description field for the CUSTOMERS.CUSTOMER_ID, CUSTOMERS.ACCOUNT_BALANCE column:

   Aggregate account balance over savings and checking accounts by customer ID to compute the total account balance for each customer.

3. Click Save, then click Close.

Lesson 3.6: Create a job that combines two mapping specifications

You are ready to save the mapping specification and then generate another job.

Recall that the goal of this lesson is to merge the customer information from Bank 1 and Bank 2. Now that you have two specifications, you can create one job that accomplishes that merge.

To generate a job that combines two mapping specifications, you must first define a mapping composition:

1. From the Mapping workspace, select the Bank_1_Extract and Bank_2_Extract specifications. in the Tasks list, select Create as Mapping Composition.
2. In the Mapping Composition editor (in the Overview page), type Bank1_Bank2_Extract_bankdemo in the Name field.


4. Click Save, then click Generate Job.

5. In the Name of New Job field, type a job name such as the following: Bank1_Bank2_Extract_bankdemo.

6. In the DataStage and QualityStage Project or Folder section, expand DSProject project, and select Jobs as the target folder.

7. The Use Shared Table Definitions check box is selected by default. Clear the check box. For this tutorial, you are not going to associate your job with tables that are already defined in the metadata repository.

8. Click Finish.

When the job is generated, the list of mapping specifications to manage is displayed.

---

**Figure 19. Composing two specifications into one job.**
Lesson 3.7: Compile and run the job

To complete the task that you started in IBM InfoSphere FastTrack, you compile and run the generated jobs in the InfoSphere Designer client.

To compile and run the job:

1. If the InfoSphere DataStage and QualityStage Designer client is not open on your computer, from your start menu, click Start > All Programs > IBM InfoSphere > IBM InfoSphere DataStage and QualityStage Designer and log in to the DSProject. Click Cancel on any windows that open.

2. From the Designer client menu, click Repository > Refresh.

3. On the navigation tree on the left of the window, expand the Jobs category. Double-click the Bank1_Bank2_Extract_bankdemo job to open the job that you just generated from InfoSphere FastTrack.

   You see a job that looks like: Figure 20.

   Note: The icon labels on your job might have different numeric indicators. The numbers are automatically generated.

   The upper branch extracts information from the files in the BANK1 schema, and the lower branch extracts information from the files in the BANK2 schema.

   Remember: You created the upper branch in Module 2.

4. Examine the lower branch of the diagram, because that diagram is the mapping specification from this module. The individual steps that make up a job are called stages.

   In the lower branch, a join operator was added as a result of the join predicate that you specified between the CUSTOMERS and ACCOUNTS tables. Also a lookup container appears after the CUSTOMERS source stage.

5. Double-click the lookup container to open it.

   The Lookup container contains a lookup stage for the DEMOGRAPHICS lookup definition to retrieve the SSN, GENDER, YEARS_CLIENT, and ONLINE_ACCESS data. See Figure 21 on page 28.
6. Select the tab at the bottom of the window to get back to the palette for the Bank1_Bank2_Extract_bankdemo job.

7. In the mapping container in the ACCOUNTS -> Transformer -> MappingContainer flow, double-click the mapping container to open it. The transformation rule for computing the aggregate represents instructions to the InfoSphere DataStage developer to complete the job. Your InfoSphere FastTrack developers created the job ahead of time for you and saved the job in a shared container. Shared containers represent work that InfoSphere DataStage developers normally do to write and share various operations or functions.

8. Replace the mapping container with the shared container that contains the aggregate function.
   a. Select the tab at the bottom of the window to get back to the palette for the Bank1_Bank2_Extract_bankdemo job.
   b. From the tree structure on the left, expand the Shared Containers folder, and drag the Bank2AccountBalanceAggregator to the palette under the mapping container. Click OK.
   c. Drag the arrow of the DSLink from the mapping container to the aggregate container. See Figure 22 on page 29.
d. Drag the output DSLINK away from the mapping container to the aggregate container so that the output comes from the aggregate container. See Figure 23.

![Figure 22. Drag the arrow of the DSLink from the mapping container to the aggregate container.](image1)

![Figure 23. Drag the output DSLINK away from the mapping container to the aggregate container.](image2)

e. Right-click the unattached mapping container and click **Delete** to remove the old container.

9. Right-click the aggregate container and select **Properties** to redefine the input and output links to the new container.

a. On the **Inputs** tab, select the link that will map from the input link, the DSLink_8 going from the aggregate container to the join. In the **General** tab, in the Link mapping area, select DSLink12 from the **Map to container link** list.
b. Click **Validate**. The message **Link map validated OK** is displayed.

c. On the **Outputs** tab, select the link that will map to the output link, **DSLink_9**. In the **General** tab, in the Link mapping area, select **DSLink82** from the **Map to container link** list.

d. Click **Validate**. The message **Link map validated OK** is displayed.

**Tip:** The input **DSLink12** and output **DSLink82** links are defined in the shared container.

10. Click **OK** to close the Properties window.

Recall that you composed two mapping specifications to generate the job: Bank 1 Extract and Bank 2 Extract. As a result, the funnel stage was added to merge the customer data that comes from the tables in the BANK1 schema with the customer data that comes from the tables in the BANK2 schema.

11. To erase the data from the BANK.CUSTOMERS target table that was created in Module 2, select the target CUSTOMERS stage. Open the stage and do the following modifications:

   a. On the **Properties** page, expand **Usage**, and select the **Before/After SQL** row.

   b. Click the text **No** to display a list. Select **Yes**.

   c. Go to the **Before SQL** row. In the empty column next to the **Before SQL**, insert the following text: **DELETE FROM BANK.CUSTOMERS**.

   d. Click **OK**.

12. On the palette for the Bank1_Bank2_Extract_bankdemo job, open each ODBC stage and type the password to connect to the data source. Set the password in the ODBC stage within the lookup container in this job.

13. To compile the job from the Designer client task bar, click **File > Save** to save the job. From the menu, click the **Compile** icon.
14. When you have a successful compilation, run the job by clicking on the Run icon.

Your result is 402 rows that are added to the BANK.CUSTOMERS table.
If you get less than 402 rows in your BANK.CUSTOMERS table (for example your result was 295 rows), and you are running on a Windows server with multiple processors, you might need to adjust the parallel options in the join stage. Right-click on the join stage, and go to Properties. Click the Input tab at the top. Select Entire as the partition type instead of Auto. See Figure 26.

Module 3: Summary
When you complete the module, you have a job that used two mapping specifications with look up definitions and joins.
Lessons learned

By completing this module, you learned how to do the following activities:

- Import a specification.
- Add a source column created by a lookup definition.
- Discover potential joins and add the join to the specification.
- Combine multiple specifications when you generate a job.

Module 4: Standardize information, use business terms, and create data extractions for customer marketing and service

In this module, you associate customer data with business terms and create data extractions for marketing and customer service.

In modules 2 and 3, you extracted data, created a mapping specification and put pertinent data in the BANK.CUSTOMERS table.

In this module, you standardize the BANK.CUSTOMERS table to create a master customer list. Subsequently, high-value customers are identified from the master customer list. You create the mapping specifications that enable an application to put appropriate marketing data into a BANK.MARKETING table and appropriate customer service data into a BANK.CUSTSERVICE table. Refer to Figure 27 on page 33.

Remember:
Gold customers have an account balance greater than $100,000, and platinum customers have an account balance greater than $300,000. First Midwest plans include the ability to offer new marketing programs to gold customers and premier customer service to its platinum customers.
You create two mapping specifications.

1. The first mapping specification provides transformation rules to the InfoSphere DataStage and QualityStage developer. The transformation rules help the developer to create operators in InfoSphere DataStage that standardize customer name and address information.

2. The second specification identifies the high-value customers. Gold customers are identified as level B for marketing and platinum customers as level A for customer service.

**Learning objectives**

After completing the lessons in this module, you will know how to do the following actions:

- Create and associate business terms with the table columns and save the relationships between terms and columns in the metadata repository.
- Apply a switch statement to populate two extract tables, one for marketing and another for customer service.
- Generate a sequential job that standardizes, cleanses, and then creates data extractions, by using the Designer client.
- Create reports that show details about your mapping specifications. You use IBM InfoSphere Information Server Web console to create the reports.

**Time required**

This module takes approximately 50 minutes to complete.
Prerequisites

Ensure that you know how to import specifications into IBM InfoSphere FastTrack and generate jobs.

Lesson 4.1: Import and create a mapping specification for standardizing and cleansing data

Similar to Module 3, you import a mapping specification to get started. The mapping specification that you import is a spreadsheet that describes how to standardize name and address information.

Note: You not only standardize, you also match and then remove duplicates so that the master customer list only contains unique customer entries.

These examples describe standardization and matching activities:

Standardizing
 Resolve Ron to be identified as Ronald.
 Split a name into a first name and a last name.
 Split an address into a street name and a house number.

Matching
 Identify duplicate entries according to a set match criteria that is defined in the match stage. This stage uses the standardized information as parameters.

To import a mapping specification for standardizing and cleansing:
1. From the IBM InfoSphere FastTrack menu, click Projects > Open Project.
2. From the Open Project window, select BankProject and click OK.
   The list of mapping specifications is displayed in the Mapping tab in the Mapping Specification folder, including Bank Extract 1 and Bank Extract 2.
3. Select the mapping specification folder in BankProject. Click Import from CSV in the Tasks list. In the Location field, browse to the C:\Tutorial\CleanseNameAddress.csv file and open it, where C:\ represents the location that you installed the data.
4. In Mapping Specification Name, type Cleanse Name Address and click Next.
5. Ensure that the Use First Row as Column Headers check box is selected. Click Restore Defaults. Click Next.
6. Click Next to accept the defaults about importing lookups. As in the previous lessons, there are no lookups in the CSV file that you are importing.
7. Ensure that the Use Full Multipart Data Source Name check box is selected. The Case Sensitive Matching for Data Source Names check box should be cleared by default.
   If you clear the Use Full Multipart Data Source Name check box, you can include the full path with the table and column names or omit some parts of the path. The full path includes host name, database name, and schema name.
8. Click Next to see the potential rows to import.
9. Click Finish to import the mapping specification.
   When the import completes, eleven rows are imported.

You now see the Cleanse Name Address mapping specification added to your list of mapping specifications.
Lesson 4.2: Describe the mapping specification and review it

Provide a description for the mapping specification and then review the functions.

To review the mapping specification:
1. From the Mapping tab, click the imported specification, Cleanse Name Address.
2. From the Task list, click Open. The Overview pane is displayed.
3. In the Description field, describe the mapping specification, such as Standardizes name and address information and click Save.
4. From the Mapping Editor, click Mappings from the list on the left.
   In the Transformation column, several transformation rules describe how to standardize name and address information. You will generate a job so that the InfoSphere DataStage developer can use the information as a guide to build a job with the correct operators.
5. Look in the Target Field column, at the MASTER.LEVEL row.
   In the Transformation Function column, you see this function: If ACCOUNT_BALANCE > 300000 then 'A' else if ACCOUNT_BALANCE > 100000 then 'B' else 'C'. This function sets the level for gold and platinum customers.
6. Click Close to return to the list of mapping specifications.
   You will not generate a job yet because you compose a job from this mapping specification and the mapping specification that you create next.

Lesson 4.3: Import and create a mapping specification for identifying high-value customers

As in Module 3, you import a partially completed mapping specification and create a mapping specification to load the marketing and customer service tables.

To import and create mapping specifications to identify the high-value customers:
1. In the Mapping Specifications view, highlight the BankProject project. Click Import from CSV from the Tasks list.
2. In the Import from CSV view, in the Location field, browse to the C:\Tutorial\IdentifyHighValueCustomers.csv file.
3. In Mapping Specification Name, type Identify High Value Customers and click Next.
4. Ensure that the Use First Row as Column Headers check box is selected. Click Next.
5. For the purposes of this tutorial, there are no lookups in the CSV file that you are importing, therefore the Import Lookups view is disabled. Click Next.
6. Ensure that the Use Full Multipart Data Source Name check box is selected.
   By default, the table and column names will be associated with the full path, which includes the host name, the database name, and the schema name. Click Next to view the potential imported information.
7. Click Finish to import the information. You are then returned to the list of mapping specifications in the BankProject.

   The new Identify High Value Customers mapping specification is now in the list. Eleven rows are imported.
Lesson 4.4: Identify the high-value customers in the mapping specification

You update the mapping specification to identify the high-value customers.

To identify high-value customers:
1. From the Mapping tab, click the imported specification, Identify High Value Customers.
2. From the Tasks list, click Open. The Overview pane is displayed.
3. In the Description field, describe the mapping specification, such as Identifies high value customers for Marketing and Customer Service.
4. From the Mapping Editor, click Mappings from the list on the left. You imported only source columns.
5. Look in the Source Field at the row that contains the MASTER.HOUSE_NUMBER column and other address columns.
   The columns are MASTER.HOUSE_NUMBER, MASTER.STREET_NAME, and MASTER.STREET SUFFIX in one source cell; MASTER FIRST_NAME, MASTER.MIDDLE_NAME, and MASTER.LAST_NAME in another source cell.
   You can see a function in Transformation Function column. Also you can see the function that is associated with the MASTER.FIRST_NAME column and the other customer name columns. These functions concatenate the name and the address to a single field.
6. Edit the functions because you notice that the imported function has a small error. The functions to concatenate the addresses and the names omitted spaces in the derived address and the derived name.
   a. Double-click the cell in the Function column in the row that contains MASTER.HOUSE_NUMBER, MASTER.STREET SUFFIX, MASTER.STREET_NAME.
   b. Move your cursor to the right of the first colon (the concatenation symbol).
   c. Type the following text to add a space between the words:
      \\:
   d. Continue to edit the text between each concatenated set for that row.
   e. Repeat the correction in the MASTER.FIRST_NAME,
      MASTER.MIDDLE_NAME, MASTER.LAST_NAME row.
   The result of the corrections looks similar to the text as shown in the following figure:
Lesson 4.5: Associate business terms with the BANK.MASTER table

Now that you have a standard customer table, BANK.MASTER, you can associate the columns of the table with business terms that are already in the metadata repository.

In these steps, you associate business terms with the columns that contain customer information.

Learn more about business terms: Terms are used to classify objects in the metadata repository. You use terms to standardize definitions of business concepts. IBM InfoSphere Business Glossary provides a Web-based tool for building a business-oriented classification system and the collaborative authoring of business metadata.

To associate business terms with columns in a table:
1. Ensure that you have the Identify High Value Customers mapping specification open. Click the Business Terms tab.
   You will see these categories: Tutorial: Customer and Tutorial: General Terms.

   **Tip:** If you do not see the Business Terms tab, click View > Business Terms.

2. Expand the Tutorial: Customer category to associate these terms with the source columns.
3. In the mapping specification table, click the expander icon to the right of the Source column heading to see the Business Term column.
4. On the Business Terms tab, associate the Street Address business term with these concatenated columns: MASTER.HOUSE_NUMBER,
MASTER.STREET_NAME, and MASTER.STREET_SUFFIX. To perform the association, select the Street Address term and drag it to the Source Business Term cell.

5. Continue associating terms to columns:
   - City
   - Customer ID
   - Gender
   - Has Online Access
   - Name
   - Number of Years as Client
   - State
   - Tax ID
   - Zip

6. Click Save.

Your “Identify High Value Customers” mapping specification now has these business terms that are associated with the source columns:

---

**Lesson 4.6: Create a business term and publish all terms**

The source column, MASTER.LEVEL, does not have a business term. You can create a term in IBM InfoSphere FastTrack.

To create a business term and publish terms:

1. In the Mapping Editor, type Level in the Source Business Terms cell in the row next to the source column, MASTER.LEVEL.
   All of the terms are annotated with [WBG] except Level. WBG indicates terms that are defined in IBM InfoSphere Business Glossary

2. On the Sources and Targets tab, Level is listed as the Source Business Term. Click Create.
   The term Level has not yet been created. You will create the term now. The Create Term pane indicates that Level is a candidate term. You can specify the category that the term is added to by selecting the Tutorial: Customer category. Specify the title as Level.
3. Enter the following description: A-platinum, B-gold, C-regular.
4. Click **Save & Close**.
   Now the term Level is defined as an InfoSphere Business Glossary term in the metadata repository.

5. Click **Save** to save your mapping specification. Ignore the warnings in the **Target Field** column.

6. Click **Close** to close the mapping specification.

7. From the list of mapping specifications, select the Identify High Value Customers specification, and click **Publish Relationships** on the **Tasks** list.
   An object list displays a list of terms and the columns to which they are associated. The list reflects only those relationships that are not already published.

8. Save all of these relationships by clicking **Select All**.
   You save the links that you just established between the columns in the BANK.MASTER table and the business terms that you associated with them.

9. Click **Publish**.
   These relationships are now saved to the metadata repository and are visible through InfoSphere Business Glossary.
Lesson 4.7: Create extractions for marketing and customer service

You return to the mapping specification and move columns from the BANK.MASTER table to two tables: BANK.CUSTSERVICE and BANK.MARKETING. Each department requires slightly different information about the customer.

To create data extractions for marketing and customer service tables:

1. On the Mapping tab, click the imported specification, Identify High Value Customers.
2. From the Tasks list, click Open.
3. In the Mapping Editor, click Mappings from the list on the left.
4. Select the subset of columns to be moved to the BANK.CUSTSERVICE table. Select the Database Metadata tab, and expand the bankconnect > bankdemo > BANK objects.

Tip: If you do not see the Database Metadata tab, click View > Database Metadata.

5. Open the BANK.CUSTSERVICE table.
   A set of columns contain data about high-value customers. The customer service department wants to track this data.

6. Drag each of the columns to the appropriate Target Field cell in the Mapping Editor. Map the following columns:

<table>
<thead>
<tr>
<th>Master customer list</th>
<th>Customer service</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASTER.CUSTOMER_ID</td>
<td>CUSTSERVICE.CUSTOMER_ID</td>
</tr>
<tr>
<td>MASTER.ZIP</td>
<td>CUSTSERVICE.ZIP</td>
</tr>
<tr>
<td>MASTER.TAX_ID</td>
<td>CUSTSERVICE.TAX_ID</td>
</tr>
<tr>
<td>MASTER.HOUSE_NUMBER</td>
<td>CUSTSERVICE.ADDRESS</td>
</tr>
<tr>
<td>concatenated columns</td>
<td></td>
</tr>
<tr>
<td>MASTER.FIRST_NAME</td>
<td>CUSTSERVICE.NAME</td>
</tr>
<tr>
<td>concatenated columns</td>
<td></td>
</tr>
<tr>
<td>MASTER.CITY</td>
<td>CUSTSERVICE.CITY</td>
</tr>
<tr>
<td>MASTER.STATE</td>
<td>CUSTSERVICE.STATE</td>
</tr>
<tr>
<td>MASTER.LEVEL</td>
<td>CUSTSERVICE.LEVEL</td>
</tr>
</tbody>
</table>

7. Expand the bankconnect > bankdemo > BANK objects and open the MARKETING table in the Database Metadata tab to extract data from the BANK.MARKETING table.

8. Use the columns NAME, ADDRESS, CITY, STATE, ZIP and LEVEL in the MARKETING table, as you used them in the CUSTSERVICE table. Drag the columns into the Columns cells under the Target header so that you have both BANK.MARKETING columns and BANK.CUSTSERVICES columns in the Columns cells under the Target header.

9. Map these additional columns to the BANK.MARKETING table:

<table>
<thead>
<tr>
<th>Master customer list</th>
<th>Marketing</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASTER.GENDER</td>
<td>MARKETING.GENDER</td>
</tr>
<tr>
<td>MASTER.ONLINE_ACCESS</td>
<td>MARKETING.ONLINE_ACCESS</td>
</tr>
</tbody>
</table>
Lesson 4.8: Define conditions that populate the subtables

Finally you define the conditions that populate the BANK.MARKETING and BANK.CUSTSERVICE tables.

To populate the subtables:
1. In the Mapping Editor, select Table Properties and click the Switch tab.
2. Add case statements that control which customers are moved to the BANK.MARKETING and BANK.CUSTSERVICE tables. Enter the following case statements in the Case cell for each of the target tables:
   - LEVEL = 'A' (Next to BANK.CUSTSERVICE)
   - LEVEL = 'B' (Next to BANK.MARKETING)

3. Click Save to save the mapping specification, then click Close.

Lesson 4.9: Generate your job

Now that you defined the data extraction, you can generate a job that combines the Cleanse Name Address and Identify High-Value Customers mapping specifications.

To generate your job:
1. From the BankProject project in the Mapping Specifications list, select the Identify High Value Customers specification and the Cleanse Name Address specification together by clicking on the first specification, holding the CTRL key, and clicking the second specification.
2. Click Generate Job.
3. Click Sequential Composition of Mapping Specifications.
4. Verify that the order of your mapping specifications is as follows:
   Cleanse Name Address
   Identify High Value Customers

5. If the order is not correct, select Cleanse Name Address and click Up to place
   the specification before the Identify High Value Customers mapping
   specification.

If your mapping specification contains validation errors, then you cannot
proceed to define the job. However, you can proceed if you have validation
warnings. Check the Validation Errors column in the Mapping Specification
list to ensure that you do not have errors that must be addressed.

6. Click Next.

7. In the Name of New Job field, type the following name for the job:
   Identify_High_Value_Customers_bankdemo.

8. In the DataStage and QualityStageProject or Folder field, expand the
   DSProject folder and select Jobs as the target folder.

9. The Use Shared Table Definitions check box is selected by default. Clear the
   check box. For this tutorial, you are not going to associate your job with tables
   that are already defined in the metadata repository.

10. Click Finish.

Lesson 4.10: Modify, compile, and run your job

You can now run your generated job. You can also look at the job and make some
modifications.

To modify, compile and run the job:

1. Open the InfoSphere DataStage and QualityStage Designer client on your
   computer. The Designer client might still be open from your previous work.

2. From the Designer client menu, click Repository > Refresh.

3. In the Repository of your project, select Jobs >
   Identify_High_Value_Customers_bankdemo and double-click the
   Identify_High_Value_Customers_bankdemo job.

   You can see the Identify_High_Value_Customers_bankdemo job with these
   stages: CUSTOMERS -> TRANSFORMER -> MAPPINGCONTAINER ->
   TRANSFORMER -> SWITCH -> MARKETING and CUSTSERVICE:

4. Open the mapping container to see the transformation rules for the name and
   address fields. The InfoSphere DataStage developer can use this information
to create the appropriate stages to cleanse and standardize customer name and address information. A shared container contains this required job. For this lesson, use one of the following shared containers:

- The **StandardizeAndClean** shared container. This job takes up to 10 minutes to run.
- The **StandardizeAndCleanDummy** shared container. This job bypasses InfoSphere QualityStage, reducing the run time of the job.

5. Select the **Identify_High_Value_Customers_ba**nk**d**emo tab to get back to the palette that shows the full job view. From the browser on the left, open **Shared Containers**, and drag the **StandardizeAndClean** container that you selected in the previous step to the palette under **MappingContainer**. Click **OK** on the window.

6. Drag the arrow of the input **DSLink** from the **MappingContainer** to the **StandardizeAndClean** container.

7. Drag the beginning of the output **DSLink** away from **MappingContainer** to the **StandardizeAndClean** container. Remove the mapping container from the palette, as shown in the following figure:

![Diagram showing the mapping of containers](image)

8. Right-click the **StandardizeAndClean** container and click **Properties**.

9. Select the **Inputs** and **Outputs** pages. For both the **Inputs** and the **Outputs** pages, in the **General** page, in the Link mapping area, select DSLink from the **Map to container link** list. Click **Validate**.

   The message: **Link map validated OK** is displayed.

10. Click **OK** to return to the palette.

11. Right-click the **Switch** stage, and select **Properties**.

   The switch logic sends customer data for level A customers, who are platinum, to the CUSTSERVICE table, and customers data for level B customers, who are gold, to the MARKETING table.

   **Tip:** Click the **Link Ordering** tab. Notice the association between Link Number and Link Name.

12. To account for customers who are neither gold nor platinum, in **Properties**, change the value of **If Not Found** = from Fail to Drop. Click **OK**.

13. Optional: If you have data already in the MARKETING and CUSTSERVICE tables due to running this job multiple times, erase the data. To erase the data from the MARKETING and CUSTSERVICE tables that was created earlier, open each container: MARKETING and CUSTSERVICE.

   a. On the **Properties** page, expand **Usage**, and select the **Before/After SQL** row.

   b. Click the text **No** to display a list. Select **Yes**.

   c. Go to the **Before SQL** row. In the empty column next to the **Before SQL**, insert the following text:

      - If you are in the MARKETING container properties, insert: DELETE FROM BANK.MARKETING.
• If you are in the CUSTSERVICE container properties, insert: DELETE FROM BANK.CUSTSERVICE.

d. Click OK.

14. On the palette for the Identify_High_Value_Customers_bankdemo job, open each ODBC stage and type the password to connect to the data source.

15. Modify the job properties.
   a. From the Designer client menu bar, select Edit > Job Properties.
   b. Select the Parameters tab.
   c. Click Add Environment Variable... and scroll down to the Execution Mode environment variable.
   d. Click Execution Mode and in the Job Properties pane, double-click on the Default Value to edit.
   e. In the Set Default drop-down, select One process.
   f. Click OK to until you are back on the job palette.

16. Click File > Save. Click Compile, and then click Close.

17. From the window that is displayed, click Run.

18. Return to the palette.
   The resulting view depends on the actual container that you selected in Step 4:

   ![Job Palette with Execution Mode Configuration]

Lesson 4.11: Create a report to review details about the mapping specification

In this lesson, you create a report that shows details about your mapping specifications.
In Modules 2, 3, and 4, you created mapping specifications and jobs that accomplish the following tasks:

- Extract customer information from Bank 1 and map it to the BANK.CUSTOMERS table.
- Extract customer information from Bank 2 and map it to the BANK.CUSTOMERS table.
- Identify gold customers as level B and platinum customers as level A and standardize customer name and address information.
- Move gold customer data appropriate for marketing (such as name, address, gender) from the BANK.CUSTOMERS table to the BANK.MARKETING table and move platinum customer data appropriate for customer service (such as name, address, customer ID number) from the BANK.CUSTOMERS table to the BANK.CUSTOMER.CUSTSERVICE table.

To create a report:

1. Open the IBM InfoSphere Information Server Web console by using one of the following methods:
   - Click on the IBM InfoSphere FastTrack icon in the left top corner of the window. Select **Open Web Console**.
   - Type the Web address of the application server, such as http://www.example.com:9080/, in the address field of your Web browser.
   - From the Windows start menu type **Start > Programs > IBM InfoSphere > IBM InfoSphere Web Console**.

2. Specify the application server user name and password.

   - You will create a detailed report for each mapping specification. Begin with Bank 1 Extract.


5. Complete the following entries:
   a. Type Bank 1 Extract Report in the **Name** field.
   b. Type Extracts customer information from BANK1 in the **Description** field.
   c. Select BankProject from the **Project Name** list.
d. Select Bank 1 Extract from the **Mapping Specification Name** list.

e. Select HTML as the **Output Format**.

f. Click **Save and Close**.

6. To run the report, click **Reports > View Reports** from the panel on the left of the window.

The Bank 1 Extract Report is listed in the middle panel.

7. Select Bank 1 Extract Report, and click **Run Now**.

8. While the Bank 1 Extract Report runs, set up reports for Bank 2 Extract and Identify High Value Customers. To do this, return to Step 2 and use the appropriate information.

9. To view a report, click **Reports > View Reports**.

10. Select a report and click **View Report Result** from the right pane.

The report is displayed in a separate browser, and you can page through the results.

**Module 4: Summary**

When you complete the module, you can see a job that was created with two mapping specifications that you placed in sequential order.

**Lessons learned**

By completing this module, you learned how to do the following activities:

- Create a mapping specification to standardize and cleanse customer name and address information.
- Create and customize a mapping specification that identifies high-value customers.
- Associate business terms to columns and create a new term.
- Publish terms.
- Define switch statements.
- Generate, compile, and run jobs.
- Create multiple reports.

**Module 5: Integrate data from Bank 3**

While you resolved issues with data in the Bank 1 and Bank 2 subsidiaries, the executive board at First Midwest acquired a new bank, Bank 3. Now you must integrate the customer data from Bank 3.

When you integrate the Bank 3 customer data with the master database, First Midwest can offer the same high-value customer service and new investment opportunities to Bank 3 customers as quickly as possible.

**Learning objectives**

Your task is to create a mapping specification to extract the customer information from the tables in the BANK3 schema and then to incorporate that customer information into the BANK.CUSTOMERS table.

The table in the BANK3 schema has only savings accounts information that is in the BANK3.SAVINGS table. The SAVINGS table is the only table in the BANK3 schema. Like the BANK1 schema, the BANK3 schema does not keep demographic
data, so there is no information on customer gender, years as a client, or access to online systems. Refer to Figure 28.

Figure 28. Moving customer data from the BANK1, BANK2, and BANK3 databases into the BANK.CUSTOMERS table.

Time required

This module takes approximately 40 minutes to complete.

Prerequisites

Ensure that you completed modules 1, 2, 3, and 4 and have access to the mapping specifications named Bank 1 Extract and Bank 2 Extract.

Lesson 5.1: Import the mapping specification

In this lesson you import another mapping specification to help you get started.

To import a mapping specification:
1. From the menu, click Projects > Open Project. Select the BankProject project from the Mapping Specifications list, and click OK.
   You will see a list of the mapping specifications that have been created in this project.
2. From the Tasks list, click Import from CSV.
3. Import the Bank3Extract.csv file. Make sure you do not have case sensitivity selected. The default is that it is not selected.
4. Modify the mapping specification by naming it Bank 3 Extract.
Lesson 5.2: Complete the mapping specification

Complete the mapping specification named Bank 3 Extract.

The metadata that you need to work with for the tables in the BANK3 schema was imported for you in Module 1. Now you are ready to use that data.

**Hint:** You have the information that you need to complete this lesson in the previous lessons. Feel free to refer back to those lessons.

To complete the mapping specification:

1. Match the columns between the BANK3.SAVINGS and the BANK.CUSTOMERS tables.

   **Hint:** Discover the columns, as you learned in “Lesson 2.3: Discover matching source columns” on page 11, to help you match the columns between the tables in the BANK3 database and the tables in the BANK database. You can discover columns in multiple rows at one time by selecting multiple columns before you click Discover Source.

2. Similar to “Lesson 3.6: Create a job that combines two mapping specifications” on page 25, combine the Bank 3 Extract mapping specification with the Bank 1 Extract mapping specification and the Bank 2 Extract mapping specification.

3. Generate a job that moves the customer data from the BANK1, BANK2, and BANK3 databases into the BANK.CUSTOMERS table.


The **Figure 29** shows your generated job before you modify it.

**Tip:** Be aware that the output of the final job might appear differently on your computer depending on the order that you combined the mapping specifications.

![Generated Bank_1_Bank_2_Bank_3_Extract job.](image)

**Figure 30 on page 50** shows you the compiled Bank_1_Bank_2_Bank_3_Extract job.

**Tip:** Remember that you need to modify the job before running it. Make sure that the inputs and outputs of the aggregator match the inputs and outputs of the data.
job flow. Look particularly at the prefixes. You want the input of the shared container to match the output to the transformer and the output of the shared container to match the input to the join.

Module 5: Summary

When you complete the module, you can see a job that integrates Bank 3 data into the master database.

Lessons learned

By completing this module, you practiced performing the following activities:

- Create a mapping specification that maps source columns to target columns.
- Combine mapping specifications to create jobs.
- Generate, compile, and run jobs.
Appendix A. Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible. The installation program installs the following product modules and components:
- IBM InfoSphere Business Glossary
- IBM InfoSphere Business Glossary Anywhere
- IBM InfoSphere DataStage
- IBM InfoSphere FastTrack
- IBM InfoSphere Information Analyzer
- IBM InfoSphere Information Services Director
- IBM InfoSphere Metadata Workbench
- IBM InfoSphere QualityStage

For information about the accessibility status of IBM products, see the IBM product accessibility information at http://www.ibm.com/able/product_accessibility/index.html.

Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in an information center. The information center presents the documentation in XHTML 1.0 format, which is viewable in most Web browsers. XHTML allows you to set display preferences in your browser. It also allows you to use screen readers and other assistive technologies to access the documentation.

The documentation that is in the information center is also provided in PDF files, which are not fully accessible.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.
Appendix B. Reading command-line syntax

This documentation uses special characters to define the command-line syntax.

The following special characters define the command-line syntax:

- **[ ]**: Identifies an optional argument. Arguments that are not enclosed in brackets are required.
- **...**: Indicates that you can specify multiple values for the previous argument.
- **|**: Indicates mutually exclusive information. You can use the argument to the left of the separator or the argument to the right of the separator. You cannot use both arguments in a single use of the command.
- **{}**: Delimits a set of mutually exclusive arguments when one of the arguments is required. If the arguments are optional, they are enclosed in brackets ([ ]).

**Note:**
- The maximum number of characters in an argument is 256.
- Enclose argument values that have embedded spaces with either single or double quotation marks.

For example:

**wsetsrc**

- *wsetsrc* [-S server] [-l label] [-n name] source

The source argument is the only required argument for the **wsetsrc** command. The brackets around the other arguments indicate that these arguments are optional.

**wlsac**

- *wlsac* [-l | -f format] [key...] profile

In this example, the -l and -f format arguments are mutually exclusive and optional. The *profile* argument is required. The *key* argument is optional. The ellipsis (...) that follows the *key* argument indicates that you can specify multiple key names.

**wrb**

- *wrb* -import {rule_pack | rule_set}...

In this example, the rule_pack and rule_set arguments are mutually exclusive, but one of the arguments must be specified. Also, the ellipsis marks (...) indicate that you can specify multiple rule packs or rule sets.
Appendix C. Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

*Table 5. IBM resources*

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description and location</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Support Portal</td>
<td>You can customize support information by choosing the products and the topics that interest you at <a href="http://www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server">www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server</a></td>
</tr>
<tr>
<td>Software services</td>
<td>You can find information about software, IT, and business consulting services, on the solutions site at <a href="http://www.ibm.com/businesssolutions/">www.ibm.com/businesssolutions/</a></td>
</tr>
<tr>
<td>My IBM</td>
<td>You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at <a href="http://www.ibm.com/account/">www.ibm.com/account/</a></td>
</tr>
<tr>
<td>Training and certification</td>
<td>You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at <a href="http://www.ibm.com/software/sw-training/">http://www.ibm.com/software/sw-training/</a></td>
</tr>
</tbody>
</table>
Appendix D. Accessing and providing feedback on the product documentation

Documentation is provided in a variety of locations and formats, including in help that is opened directly from the product client interfaces, in a suite-wide information center, and in PDF file books.

The information center is installed as a common service with IBM InfoSphere Information Server. The information center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open the information center from the installed product or from a Web browser.

Accessing the information center

You can use the following methods to open the installed information center.

- Click the Help link in the upper right of the client interface.

  **Note:** From IBM InfoSphere FastTrack and IBM InfoSphere Information Server Manager, the main Help item opens a local help system. Choose Help > Open Info Center to open the full suite information center.

- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

  **Note:** The F1 key does not work in Web clients.

- Use a Web browser to access the installed information center even when you are not logged in to the product. Enter the following address in a Web browser: http://host_name:port_number/infocenter/topic/com.ibm.swg.im.iis.productization.iisinfsv.home.doc/ic-homepage.html. The host_name is the name of the services tier computer where the information center is installed, and port_number is the port number for InfoSphere Information Server. The default port number is 9080. For example, on a Microsoft® Windows® Server computer named iisdocs2, the Web address is in the following format: http://iisdocs2:9080/infocenter/topic/com.ibm.swg.im.iis.productization.iisinfsv.nav.doc/dochome/iisinfsrv_home.html.

A subset of the information center is also available on the IBM Web site and periodically refreshed at [http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r7/index.jsp](http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r7/index.jsp)

Obtaining PDF and hardcopy documentation

- A subset of the PDF file books are available through the InfoSphere Information Server software installer and the distribution media. The other PDF file books are available online and can be accessed from this support document: [https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1](https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1)

- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at [http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss](http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss)
Providing comments on the documentation

Your feedback helps IBM to provide quality information. You can use any of the following methods to provide comments:

- To comment on the information center, click the Feedback link on the top right side of any topic in the information center.
- Send your comments by using the online readers' comment form at [www.ibm.com/software/awdtools/rcf/](http://www.ibm.com/software/awdtools/rcf/).
- Send your comments by e-mail to comments@us.ibm.com. Include the name of the product, the version number of the product, and the name and part number of the information (if applicable). If you are commenting on specific text, include the location of the text (for example, a title, a table number, or a page number).
Notices and trademarks

This information was developed for products and services offered in the U.S.A.

Notices

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user’s responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web
sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM’s future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to
IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

**Trademarks**

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS\textsuperscript{link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS\textsuperscript{link} licensee of the United States Postal Service.

Other company, product or service names may be trademarks or service marks of others.
Index

A
adding business rules 25
adding conditions 42
adding functions 21
adding joins 21
adding lookup definitions 22

B
business glossary terms 4
business rules 25
business terms 37, 38
business terms associated with metadata 32

C
cleansing data 34
comma separated values (CSV) 4
command-line syntax conventions 53
commands syntax 53
compiling DataStage jobs 16
composing a mapping specification from multiple specifications 25
creating a FastTrack project 8
creating a specification 7
creating business terms 38
customer support contacting 55

data repository use 8
DataStage jobs 15, 16, 25, 27, 42, 43, 49
DataStage shared containers 4
DataStage stages 16
defining InfoSphere DataStage credentials 3
defining user roles 3
derivations 16
discover function 11
discover option 21
discover options 47
discovering join candidates 21
discovering matching columns 11
drag into the mapping editor 8

defilters 14
functions 13

generating a DataStage job 25
generating DataStage jobs 15, 42

I
importing a specification 34
importing metadata 4
importing specifications 47
InfoSphere components 16
InfoSphere DataStage jobs
IBM InfoSphere FastTrack reports 46
reporting 46

J
join keys, adding to tables 22

L
legal notices 59
lookup operators 22
lookup tables 17

M
manipulating data 13
mapping business terms to columns 37
mapping source and target columns 41
mapping specification composition 17
mapping specification refinement 32
mapping specifications
importing 19, 34, 35, 48, 49
mapping specifications
updating 49
navigating 35
overview 21
updating 36
mapping specifications tab, find using view menu 8
mapping table information 7
moving columns in a specification 41

P
parallel composition of mapping specifications 25
preparing a DataStage job 27
product accessibility
accessibility 51
product documentation
accessing 57
publishing business terms 38

R
running DataStage jobs 16

S
sample data 4
sample files 4
sequential mapping composition 42
shared containers 43
shared containers in DataStage 27
software services
contacting 55
source and target mapping 8
special characters in command-line syntax 53
standardizing information 32
support
customer 55
switch conditions 42
syntax
command-line 53
system requirements 3

T
table functions 14
table mappings 14
trademarks list of 59
transformation functions 7
transformation rules 35
transformations 13

U
updating jobs generated from FastTrack 43
uploading mapping specifications 47
uploading terms into the business glossary 4

© Copyright IBM Corp. 2007, 2012

63