Note

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page [page].
## Contents

Chapter 1. Overview of InfoSphere Metadata Workbench ................................ 1

Chapter 2. Accessing the metadata workbench ..................................................... 3

Chapter 3. InfoSphere Metadata Workbench roles .............................................. 5

Chapter 4. Tutorial: Preparing data for lineage reports ....................................... 7
  Setting up the tutorial environment ................................................................. 7
  Copying the installed tutorial files .................................................................. 8
  Importing assets into the metadata repository .................................................. 9
  Importing databases, database tables, data files, and BI models ......................... 9
  Importing an IBM InfoSphere DataStage job .................................................... 16
  Creating application and file extended data source assets ................................ 17
  Importing extension mapping documents ....................................................... 19
  Module summary .............................................................................................. 22
  Completing tasks before running lineage reports ............................................. 22
    Managing lineage ........................................................................................... 22
    Identifying schemas as identical .................................................................... 23
    Module summary ............................................................................................ 26
  Running lineage reports .................................................................................. 26
    Running data lineage reports ....................................................................... 26
    Running business lineage reports .................................................................. 27
    Module summary ............................................................................................ 28

Chapter 5. Exploring metadata assets ................................................................. 29
  Tips for navigating the metadata workbench .................................................... 29
  Extended data lineage ....................................................................................... 30

  Information assets in the metadata workbench ................................................. 31
    Information asset actions .............................................................................. 32
    Asset types that are displayed in the metadata workbench ............................. 35
    Finding assets in the metadata workbench .................................................... 58
    Querying the metadata repository .................................................................. 59
    Queries ........................................................................................................... 59
    Creating queries ............................................................................................ 62
    Managing queries ......................................................................................... 64
    Running queries ............................................................................................ 65

Chapter 6. Creating data lineage, business lineage, and impact analysis reports in the metadata workbench ................................................................. 67
  Data lineage, business lineage, and impact analysis reports ............................ 67
    Asset types that are included in lineage and analysis reports .......................... 69
    Running data lineage reports ....................................................................... 70
    Running business lineage reports .................................................................. 72

Appendix A. Product accessibility ....................................................................... 75

Appendix B. Contacting IBM ............................................................................... 77

Appendix C. Accessing and providing feedback on the product documentation .... 79

Notices and trademarks ...................................................................................... 81

Index ................................................................................................................... 85
Chapter 1. Overview of InfoSphere Metadata Workbench

IT professionals such as business analysts, data analysts, and ETL developers use IBM® InfoSphere® Metadata Workbench to explore and analyze relationships between information assets in the metadata repository.

InfoSphere Metadata Workbench provides IT professionals with a design-time tool for managing and understanding the assets that are generated and used by the IBM InfoSphere Information Server suite, and for extending that analysis to assets and processes that are external to the suite.

By providing lineage reports, InfoSphere Metadata Workbench supports IT professionals who are responsible for compliance and governance initiatives that require lineage information (for example, Sarbanes Oxley or Basel II requirements). By providing an impact analysis that shows the effect of changes to information management environments, InfoSphere Metadata Workbench helps IT professionals to work most efficiently.

InfoSphere Metadata Workbench supports these tasks:

• Explore information assets in the metadata repository by using these features:
  – Details about the relationships of jobs, business intelligence (BI) reports, databases, data files, tables, columns, terms, stewards, servers, extended data sources, and other assets
  – Simple and advanced search and robust querying
  – Integrated cross-suite view of information assets
  – Details of extension mappings that describe external data flow and assets
  – Graphical view of asset relationships
• Analyze dependencies and relationships of key assets and business BI reports by doing these tasks:
  – Trace lineage through jobs and databases to BI reports
  – Perform lineage analysis to understand where data comes from or goes to by using shared table information, job design information, operational metadata from job runs, and extension mappings
  – Perform impact analysis to understand dependencies and the effects of changes to a column or job in IBM InfoSphere Information Server and beyond
  – View operational metadata from job runs
• Manage metadata to obtain in-depth analysis reports by doing these tasks:
  – Create and edit descriptions of information assets
  – Import or create assets that do not originate in InfoSphere Information Server:
    - Applications, stored procedures, and files that are defined as extended data sources
    - ETL processes that are defined as extension mappings
  – Assign terms, stewards, labels, and notes to information assets
Chapter 2. Accessing the metadata workbench

You access the metadata workbench by using a web browser.

Before you begin

- The supported web browsers are Microsoft Internet Explorer versions 6 and 7, and Mozilla Firefox version 2.
- To view graphical reports, Adobe® Flash Player version 10.0.22 or later must be installed. You can download the player from [http://www.adobe.com/](http://www.adobe.com/).
- To access the metadata workbench, you must have the role of Metadata Workbench Administrator or Metadata Workbench User. To view business lineage reports, you must have the role of Business Glossary User. The suite administrator of IBM InfoSphere Information Server assigns suite users to roles.
- You must have the URL to a server where IBM InfoSphere Metadata Workbench is installed.
- Enable JavaScript in the browser.
- Enable HTTP 1.1 in the Microsoft Internet Explorer browser.
- Enable cookies for the site.
- Your screen resolution must be set to 1024 by 768 or greater. Maximize the metadata workbench window.
- For detailed system requirements, see [http://www.ibm.com/software/data/infosphere/info-server/overview/requirements.html](http://www.ibm.com/software/data/infosphere/info-server/overview/requirements.html).

Procedure

1. Open a web browser and enter a URL in the following format:
   
   `protocol://host_server:port/workbench`

   where `protocol` is the communication protocol: either Hypertext Transfer Protocol (HTTP) or Hypertext Transfer Protocol Secure (HTTPS).

   The value of `host_server:port` depends on whether WebSphere® Application Server clustering is set up within your services tier configuration.

   Table 1. Host and port values for different configurations

<table>
<thead>
<tr>
<th>WebSphere Application Server cluster configuration</th>
<th>Value of <code>host_server</code></th>
<th>Value of <code>port</code> if HTTP is used</th>
<th>Value of <code>port</code> if HTTPS is used</th>
</tr>
</thead>
<tbody>
<tr>
<td>If clustering is set up</td>
<td>The host name or IP address and the port of the front-end dispatcher (either the HTTP server or the load balancer). Do not use the host name of a particular cluster member.</td>
<td>HTTP port of the front-end dispatcher (for example, 80).</td>
<td>HTTPS secure port (for example, 443).</td>
</tr>
</tbody>
</table>
Table 1. Host and port values for different configurations (continued)

<table>
<thead>
<tr>
<th>WebSphere Application Server cluster configuration</th>
<th>Value of host_server</th>
<th>Value of port if HTTP is used</th>
<th>Value of port if HTTPS is used</th>
</tr>
</thead>
<tbody>
<tr>
<td>If clustering is not set up</td>
<td>The host name or IP address of the computer where WebSphere Application Server is installed.</td>
<td>HTTP transport port (configured as WC_defaulthost in WebSphere Application Server. Default: 9080)</td>
<td>HTTPS transport secure port (configured as WC_defaulthost_secure in WebSphere Application Server. Default: 9443)</td>
</tr>
</tbody>
</table>

2. If HTTPS is enabled, then the first time that you access the metadata workbench, a message about a security certificate is displayed if the certificate from the server is not trusted. If you receive such a message, follow the browser prompts to accept the certificate, type your suite user name and password, and click Login.

**What to do next**

The default HTTP session timeout is 30 minutes. As a result, any web session that remains inactive for more than 30 minutes is terminated. You can change the default timeout value in IBM InfoSphere Information Server Web console (Administration tab > Session Management tab > Global Session Properties).
Chapter 3. InfoSphere Metadata Workbench roles

The suite administrator assigns roles that define the tasks that users of IBM InfoSphere Metadata Workbench can perform.

IBM InfoSphere Metadata Workbench has the following roles:

**Metadata Workbench Administrator**

- Runs the automated and manual analysis services, publishes queries, and explores metadata models. Performs all tasks that IBM InfoSphere Metadata Workbench users can perform.

The Metadata Workbench administrator must be familiar with the enterprise database metadata and data file metadata that is imported into the repository. The administrator must also be familiar with the metadata that is used in jobs.

**Metadata Workbench User**

- Finds and explores information assets, runs analysis reports, and creates, saves, and runs queries.
Chapter 4. Tutorial: Preparing data for lineage reports

In this series of modules, you use IBM InfoSphere Metadata Workbench to create a flow of data that can be displayed in a lineage report. You learn how to prepare the data for lineage reports by completing administrative tasks. Lastly, you learn to run lineage reports.

These tasks are required to create the flow of data that is displayed in lineage reports:

1. Import all assets that are needed for the tutorial into the metadata repository:
   - Relational databases with database tables and columns
   - Data files with structures and fields
   - Business intelligence (BI) reports with reports and models
   - Extended data source assets of type file and of type application
   - Extension mapping documents with source-to-target mappings
   - IBM InfoSphere DataStage® and QualityStage® jobs

   After you import the assets, you verify that they are in the metadata repository by using InfoSphere Metadata Workbench.

2. Perform administrative tasks that prepare the data for correct lineage:
   - Define a database alias so that the Manage Lineage utility can set relationships between stages of a job and database tables.
   - Run the Manage Lineage utility. This step links the target stage in one job to the source stage in the next job, and links views to database tables.
   - Define two schemas as identical. All database tables and database columns that are contained by identical schemas are also marked as identical when their names match.

3. Run lineage reports:
   - Run data lineage to display all assets in the complete flow of data.
   - Run business lineage to display selected assets in the flow of data.

Setting up the tutorial environment

You must prepare your system to run the tutorial. You access IBM InfoSphere Metadata Workbench by using your Web browser. Other product modules must be installed.

Learning objectives

After you complete the lessons in this module, the tutorial is ready for use.

Time required

The time needed to complete the setup depends on the overall performance of your system and on other IBM InfoSphere Information Server product modules that you installed.
Prerequisites

You must know the Web address of IBM InfoSphere Metadata Workbench and of InfoSphere Information Server.

You must know the user name and password of an account that has the Metadata Workbench Administrator role, the Suite Administrator role, and the DataStage and QualityStage Administrator role. These roles might be in different accounts or all roles might be in the same account.

The following software must be installed on the appropriate tier for InfoSphere Information Server:

- IBM InfoSphere Metadata Workbench
- IBM InfoSphere DataStage and QualityStage
- Istool (This software is typically installed in
  \$/InfoSphere_installation_directory/Clients/istools/cli
  where
  InfoSphere_installation_directory is the top-level installation directory of InfoSphere Information Server.)

The following software must be installed on the same computer from which you run the tutorial:

- InfoSphere DataStage and QualityStage Designer client
- IBM InfoSphere DataStage and QualityStage Administrator
- Any version of the Microsoft Windows operating system
- Microsoft Internet Explorer versions 6 or 7, or Mozilla Firefox version 2

Copying the installed tutorial files

You must copy the installed tutorial files to a temporary directory on your computer.

When IBM InfoSphere Metadata Workbench was installed on the services tier of IBM InfoSphere Information Server, the tutorial files were also installed. The tutorial files are in a compressed format.

To copy the installed tutorial files to your computer:

1. Go to the directory \IS_Installer\is-suite\TutorialData\WorkBench on the installation media.
2. Copy the file, tutorial.zip, to a temporary directory on your computer, such as C:\temp\tutorial.
3. Extract all files in tutorial.zip to the temporary directory.

The scripts are needed to import or to create the assets in the metadata repository. These scripts are extracted from tutorial.zip and are in the temporary directory:

- EWS.dsx
- EWS.isx
- EWS1.isx
- Report.isx
- ExtensionApplication_Source.csv
- Extended_File_Source.csv
- EWSMapping1.csv
- EWSMapping2.csv
Importing assets into the metadata repository

In this module, you import different types of metadata assets into the metadata repository.

Learning objectives

The lessons in this module explain how to do the following actions:

- Import databases, database tables, and data files from files that were created by vendor software.
- Import business intelligence (BI) reports and models that were created by IBM Cognos®.
- Create extended data source assets of type application and of type file by importing a file that is in a comma-separated value (CSV) format.
- Import extension mapping documents with two rows of source-to-target mappings. The extension mapping document is in a CSV format.
- Import jobs that were created by IBM InfoSphere DataStage and QualityStage Designer.

In this tutorial, you import assets into the metadata repository by using import files. Typically however, assets are created or imported into the metadata repository when an IBM InfoSphere Information Server product module connects to the data source.

Time required

This module takes approximately 45 minutes to complete.

Importing databases, database tables, data files, and BI models

You import databases, database tables, data files, and business intelligence (BI) models into the metadata repository. The import files are in an ISX format.

To import databases, database tables, data files, and BI models:

1. At the command prompt, go to the directory
   \InfoSphere_installation_directory\istools\cli, where
   \InfoSphere_installation_directory\ is the top-level installation directory of IBM InfoSphere Information Server. For example, the directory path might be C:\IBM\InformationServer\Clients\istools\cli.
2. At the command prompt, type the command istool.
3. Run this command on each tutorial file whose file extension is “isx”:

   ```
   import -d dom SERVERNAME -u USERNAME -p PASSWOrd -ar FULL_PATH_TO_FILE -cm
   ```

   where
   - `SERVERNAME` is the name or IP address of InfoSphere Information Server.
     - If clustering is set up, use the name or IP address and the port of the front-end dispatcher (either the web server or the load balancer). Do not use the host name and port of a particular cluster member.
     - If clustering is not set up, use the host name or IP address of the computer where WebSphere Application Server is installed and the port number that is assigned to the IBM InfoSphere Information Server Web console, by default 9080.
- **USRNAME** and **PASSWORD** is the user name and password of an account on InfoSphere Information Server with the Suite Administrator role.
- **FULL_PATH_TO_FILE** is the file name of an extracted tutorial file with the directory path. An example might be C:\temp\tutorial\EWS.ISX.
- Add two single quotation marks (" ) after the -cm parameter.

**Important:**
- You must add two single quotation marks (" ) after the -cm parameter.
- The command to import Report.isx is:
  ```
  import -dom SERVERNAME -u USRNAME -p PASSWORD -ar FULL_PATH_TO_FILE
  -cm " -replace
  ```

You can check that the databases, database tables, data files, and BI models are in the metadata repository by doing these steps:
1. Open the Web browser and connect to the Web address of IBM InfoSphere Metadata Workbench.
2. Type your user name and password, and click **Login**. The Welcome page of the metadata workbench is displayed.
3. In the left pane of the metadata workbench, click the **Discover** tab.
4. In the Additional Types list, select **Database** and click **Display**. The newly created databases are displayed in the right pane of the metadata workbench.

![Database Results](image)

*Figure 1. List of new databases in the metadata repository*

5. Right-click **DW_MART** and select **Open Details in New Window** to display its details in the asset information page.
6. In the Additional Types list in the left pane, select Data Files and click Display. The newly created data files are displayed.
7. Right-click C:\EWS\Prod\GlobalSales and select Open Details in New Window to display its details in the asset information page.

Figure 3. List of three new data files in the metadata repository
8. In the Additional Types list in the left pane, select BI Model and click Display. The newly created BI model EWS is listed.

9. Right-click EWS and select Open Details in New Window. Note the BI collections and BI report that are listed in the asset information page of the BI model EWS:

![Asset information page for a data file](image)

---

**Figure 4. Asset information page for a data file**

8. In the Additional Types list in the left pane, select BI Model and click Display. The newly created BI model EWS is listed.

9. Right-click EWS and select Open Details in New Window. Note the BI collections and BI report that are listed in the asset information page of the BI model EWS:
Lesson checkpoint
In this lesson, you imported databases, database tables, data files, and BI models. You can display these imported assets in the Browse tab of the left pane of the metadata workbench.

You imported these databases, schemas, database tables, and data files into the metadata repository:
You created the BI model EWS that contains BI collections, PROD_MRT and SALES_MRT. You created the BI report ProductionRunReport.

Figure 6. List of imported assets that are displayed in the Browse tab of left pane
Importing an IBM InfoSphere DataStage job

You must create a project and then import a job into the project. The project and its job are created and then imported into the metadata repository by using IBM InfoSphere DataStage and QualityStage Designer.

1. Open IBM InfoSphere DataStage and QualityStage Administrator in your desktop. Log in with the user name and password of an account with the DataStage and QualityStage Administrator role.
2. Click the Projects tab to list the Projects. Click Add.
3. In the Name field of the Add Project window, type EWS. Click OK. The EWS project is created and added to the list of projects.
4. Click Close to save the new project and to exit the InfoSphere DataStage and QualityStage.
5. Open IBM InfoSphere DataStage and QualityStage Designer in your desktop. Log in with the user name and password of an account with the DataStage and QualityStage Administrator role. Select the project EWS from the Project list. If a New window opens, click Cancel.
6. In the toolbar, click Import > DataStage Components.
7. In the DataStage Repository Import window, browse to the directory where you extracted the tutorial files. Select EWS.dsx as the import file.
8. Click OK to import the project.
9. Click File > Exit to close IBM InfoSphere DataStage and QualityStage Designer.

You can check that the jobs are in the metadata repository by doing these steps in IBM InfoSphere Metadata Workbench:

1. In the left pane of the metadata workbench, click the Discover tab.
2. In the Asset Types list, select Job and click Display. A list of all jobs is displayed.
3. Narrow your search to display only those jobs whose name begins with "EWS" by typing this string in the Narrow Your Results field.

Lesson checkpoint

In this lesson, you imported InfoSphere DataStage jobs into the metadata repository.

Creating application and file extended data source assets

In this lesson, you import files in a comma-separated value (CSV) format. The files list extended data source assets of type application or of type file. The import creates these assets in the metadata repository.

To create application and file extended data source assets:
1. In the left pane of the metadata workbench, click the Advanced tab and select Import Extended Data Sources.

2. Click Add in the Import Extended Data Sources window. Browse to the directory where you extracted the tutorial files and select these files:
   - ExtensionApplication_Source.csv
   - Extended_File_Source.csv

3. Click OK. The Status window displays the import status as the files are read. File and application assets are created in the metadata repository.

4. Click OK to close the Status window.

You can check that the extended data source assets are in the metadata repository by doing these steps:
1. In the left pane of the metadata workbench, click the Discover tab.
2. In the Asset Type list, select Application and click Find. Right-click the newly created application asset CRM and select Open Details in New Window to display its details.
3. In the Asset Type list, select **File** and click **Find**. Right-click the newly created file asset **Customer Data Upload** and select **Open Details in New Window** to display its details.

**Lesson checkpoint**

In this lesson, you imported two files in a CSV format by using InfoSphere Metadata Workbench. The import created new extended data source assets of type file and of type application in the metadata repository.

You learned the following tasks:

- How to create extended data source assets in the metadata repository by importing a CSV file.
- How to view the new extended data source assets in the metadata repository after the import.

**Importing extension mapping documents**

In this lesson, you import extension mapping documents into the metadata repository by using IBM InfoSphere Metadata Workbench. Each mapping row in the document maps a source asset to a target asset. The source assets and the target assets were created or imported into the metadata repository in the previous lessons.

To import extension mapping documents:

1. In the left pane of the metadata workbench, click the **Advanced** tab and select **Import Extension Mapping Documents**.
2. In the Import Extension Mapping Documents window, click **Add** in the top pane. Browse to the directory where you copied the tutorial files and select **EWS Mapping1.csv** and **EWS Mapping2.csv**. Leave the Source, Target, and Configuration fields blank.

3. Click **OK** and then click **Save** to save and import the extension mapping document into the metadata repository.

4. Click **OK** to close the Status window.

5. Right-click the extension mapping document **EWS Mapping 2.csv** and select **Open Details in New Window**.

   In the Extension Mappings pane of the asset information page, the two mapping rows, both called “SP Read”, are listed. Inventory Data is a source asset of type file and is mapped to the target assets AmericaProd and Plant. AmericaProd is an asset of type file structure. Plant is an asset of type file field.

![Figure 12. Asset information page of an extension mapping document with two mapping rows](image)

To see the asset information page of the source or target assets, right-click the asset name and select **Open Details in New Window**. In this example, the asset information page of InventoryData displays the same information about the extension mapping rows as the asset information page of the extension mapping document.
In this lesson, you created an extension mapping document with two mapping rows. You noted the source-to-target mapping assignment by looking at the asset information page of the extension mapping document, or of the source or target assets.

You learned the following tasks:
• How to import an extension mapping document with source and target assets.

Lesson checkpoint
In this lesson, you created an extension mapping document with two mapping rows. You noted the source-to-target mapping assignment by looking at the asset information page of the extension mapping document, or of the source or target assets.

You learned the following tasks:
• How to import an extension mapping document with source and target assets.

Figure 13. Asset information page of the source asset InventoryData
• How to see the source-to-target mappings when you view the asset information page of the extension mapping document.

Module summary
In this module, you imported assets into the metadata repository that are needed to create a lineage report.

You imported the following assets:
• Database
• Database file
• IBM Cognos business intelligence (BI) report
• Extended data source assets of type application and of type file
• Extension mapping documents with source-to-target mappings
• Compiled IBM InfoSphere DataStage and QualityStage jobs

The metadata repository now has the assets that are needed for the lineage report. The next step is to perform administrative tasks to prepare the data.

Lessons learned
In this module, you learned how to do the following tasks:
• Import databases, database files, BI reports, and compiled IBM InfoSphere DataStage and QualityStage jobs by using istool command-line interface.
• Create extended data source assets of type application and of type file by using IBM InfoSphere Metadata Workbench.
• Import extension mapping documents with source-to-target mappings by using InfoSphere Metadata Workbench.

Completing tasks before running lineage reports
In this module, you perform administrative tasks that are needed to prepare the data in the metadata repository for lineage reports.

Learning objectives
The lessons in this module explain how to do the following actions:
• Set relationships between stages and tables or data file structures, between stages, and between database tables and views.
• Map a database alias to ensure that stages and database tables are correctly linked in lineage reports.
• Check the identity of data sources to identify duplicate database and schemas.

Time required
This module takes approximately 30 minutes to complete.

Managing lineage
In this lesson, you learn how to manage lineage by setting relationships between stages and tables or file structures, between stages, and between base tables and views.
When the target stage in one job is matched to the source stage in the next job, the meta workbench reports display the cross-job analysis. When views are matched to their source base tables, the meta workbench displays the relationships in the lineage report.

The Manage Lineage utility in the meta workbench works with stages that connect to bases and files. The most commonly used stages are supported.

To run the Manage Lineage utility:
1. Click the Advanced tab in the left pane of the meta workbench and then click Manage Lineage.
2. Select the EWS project, which has new jobs, and then click the Detect Associations icon. This step detects associations between stages and sources.
3. Click the Map base Alias icon and do the following steps:
   a. In the Mapped base Aliases table, locate the row of the base alias, DW_Mart. Click Select in that row.
   b. In the Select a base for DW_Mart window, click Find and then select EWS in the results list. Click Select.
   c. Click Save to define EWS as the alias for the base DW_Mart.
   d. Repeat step 2 to update the base alias.
   These steps maps a base name to an alias name.

The Last Run field of the row in the Transformation Project results table displays the date and time that Manage Lineage was last run.

Lesson checkpoint
In this lesson, you assigned a base alias to a base to ensure that stages and base tables are correctly linked in lineage reports. You learned how to run the Manage Lineage utility.

Identifying schemas as identical
In this lesson, you learn how to specify that two schemas are identical. Database tables that the schemas contain are also marked as identical if the table names match.

To identify schemas as identical:
1. Click the Advanced tab in the left pane of the metadata workbench and then click Data Source Identity.
2. In Select a Database window, click Find.
3. Select DW_MART and click Select. In the Matched Schemas table, the database, DW_MART, has a single schema, SCHEMA1.
4. In the row for SCHEMA1, click Add.
5. In the Select a Schema for SCHEMA1 window, click Find and select SCHEMA1 of database EWS and of host EWS. Click Select.
6. Click Save to save your changes.
In the Browse tab of the left pane of the metadata workbench, you can see that the database tables of the matched schemas SCHEMA1 have the same table names. In this case, the database tables are also identified as identical.

Figure 14. Two schemas from different databases are identified as identical

In the Browse tab of the left pane of the metadata workbench, you can see that the database tables of the matched schemas SCHEMA1 have the same table names. In this case, the database tables are also identified as identical.
Lesson checkpoint
In this lesson, you identified SCHEMA1 of database DW_MART and SCHEMA1 of database EWS as identical schemas. Database tables in these schemas have the same name. As a result, these database tables are also identified as identical.
Module summary
In this module, you performed administrative tasks on the assets that you created or imported into the metadata repository. The administrative tasks prepared the data for correct lineage.

Lessons learned
In this module, you learned how to do the following tasks:
- Define a database alias so that the Manage Lineage utility can set relationships between stages of a job and database tables.
- Run the Manage Lineage utility. This step links the target stage in one job to the source stage in the next job, and links views to database tables.
- Define two schemas as identical. All database tables and database columns that are contained by identical schemas are also marked as identical when their names match.

Running lineage reports
In this module, you create reports that analyze the flow of data from data sources, through jobs and stages, and into databases, data files, and business intelligence (BI) reports.

Learning objectives
The lessons in this module explain how to do the following actions:
- Run data lineage reports
- Run business lineage reports

Time required
This module takes approximately 20 minutes to complete.

Running data lineage reports
Data lineage reports show the movement of data within a job or through multiple jobs and show the order of activities within a run of a job. In this tutorial, the data lineage report includes all assets that flow from an application asset to a business intelligence (BI) report.

1. Click the Discover tab in the left pane of the metadata workbench.
2. In the Asset Type list, select BI Report and click Find.
3. In the Find Results pane, right-click ProductionRunReport and select Data Lineage.
4. Select the Application tab and then select CRM. The results of data lineage report vary according to the assets that you select as your starting point for the data flow.

The data lineage report is displayed. A list of assets according to asset type is displayed in the right pane.
You can display information about the assets in a data lineage report in any of the following ways:

- Open each twisty of an asset type (in the right pane) to display the assets.
- Move the mouse pointer over the graphic of an asset in the left pane.
- Click the graphic of an asset to expand the asset information in the right pane.

You can get additional information about an asset by clicking the icons at the bottom of each graphic.

The results of data lineage vary according to the asset that you select as your start and end points for the data flow. If, instead of selecting CRM as the starting point for data lineage, you select EWS_SalesStaging job, then the data lineage report would be different.

**Figure 16. Data lineage from an application asset to a BI report**

**Lesson checkpoint**

In this lesson, you learned how to run a data lineage report.

**Running business lineage reports**

In this lesson, you create a business lineage report that displays only the flow of data, without the details of a full data lineage report.
Business lineage reports show a scaled-down view of lineage without the detailed information that is not needed by a business user. Business lineage reports show data flows only through those assets that have been configured to be included in business lineage reports. In addition, business lineage reports do not include extension mapping documents or jobs from IBM InfoSphere DataStage and QualityStage.

To run business lineage reports:
1. Click the Discover tab in the left pane of the metadata workbench.
2. In the Asset Type list, select BI Report. Click Find.
3. In the Find Results pane, right-click ProductionRun report and select Business Lineage.

The business lineage for a BI report is displayed.

![Business lineage report for a BI report](image)

**Figure 18. Business lineage report for a BI report**

**Lesson checkpoint**
In this lesson you learned how to create a business lineage report.

**Module summary**
In this module you ran data lineage and business lineage reports.

**Lessons learned**
In this module, you learned how to do the following tasks:
- How to configure and run data lineage reports.
- How to run business lineage reports.
Chapter 5. Exploring metadata assets

Users of the metadata workbench can search and query information assets in the metadata repository in order to view their properties and related assets and to create reports on the flow of data through these assets.

Tips for navigating the metadata workbench

You navigate to view assets in the metadata repository and to create reports about assets and their relationships.

You move through the metadata workbench by clicking hyperlinks and buttons in the application.

Use the following tips:

• Do not use the back and forward features of your Web browser to return to pages that you visit.
• You can start any task from one of the navigation tabs: Browse, Discover, or Advanced.
• To view context information for an asset, move your mouse pointer over the link to the asset in a results list or on an asset information page. The hover help displays assets that contain the object. For example, if your mouse pointer hovers over the link to a job, the names of the project and engine that contain the job are displayed, and you can click either name to open the asset information page for that asset.
• To choose from the list of tasks that you can do with an asset, right-click the asset name in a results list, or click the asset name to display its asset information page.
• To return to the Welcome page, click the Workbench tab in the upper left corner of the window.
• To return to any asset information page that you visit, click Add to Favorites to add the page to the list of bookmarks or favorites in your Web browser.
• To open an asset information page without navigating away from the page that you were on, right-click the asset name and click Open in New Window.
Extended data lineage

You can track the flow of data in your enterprise, even when you use external processes that do not write to disk, or ETL tools, scripts and other programs that do not save their metadata to the metadata repository.

The metadata repository stores information from the tools in the InfoSphere Information Server suite. The metadata workbench uses this information to display the flow of information from your data sources, through IBM InfoSphere DataStage and QualityStage jobs, and into target data structures.

However, there might be times you need to view data lineage that includes data flows that are not stored in the metadata repository. This scenario can happen in the following cases:

- Your enterprise uses third-party ETL (extract, transform, and load) tools whose information is not automatically stored in the metadata repository.
- Your InfoSphere DataStage job depends on information from stored procedures.
- You invoke Web services that are located elsewhere.
- You need to track lineage from mainframe applications and programs such as COBOL extracts.
- You receive flat files or other types of feeds from third parties.
- You often run scripts at the operating system level to copy or restructure files before processing them in jobs.

In these cases and others, the flow of data through tables and columns can extend beyond the metadata repository. By creating mappings in extension mapping documents and extended data sources in the metadata workbench you can track that flow and create data lineage reports from any asset in the flow.

Mappings in extension mapping documents

Mappings in extension mapping documents are source-to-target mappings that represent an external flow of data from one or more sources to one or more targets. The source or target must exist within the metadata repository, either as database or data file metadata or as a component of an extended data source.

By using mappings in the metadata workbench, you can create data lineage reports for the following types of flows:

- Data flows that happen completely outside of InfoSphere Information Server
• Data flows that happen both outside and inside InfoSphere Information Server

Mappings allow great flexibility in the types of assets that you can map to each other. You make your mapping decisions based on what information you want to see in your data lineage reports. For example, you might map an external stored procedure definition to a job, or you might map the output value of a column used in an ETL tool from an independent software vendor to an input parameter of a different ETL process.

**Extended data sources**

Before you create mappings, you can import the metadata from external databases and data files into the metadata repository. But some external processes, including Web services and stored procedures, do not write their data to disk. You might need to report on information about the data structure in these processes to get a clear picture of each step of the transformation process. You can capture this information by creating extended data sources and importing them into the metadata workbench.

You can create three distinct types of extended data sources: applications, stored procedure definitions, and files. The application type allows maximum flexibility for you to create an extended data source at any level of granularity. You can define object types, methods, and input and output parameters for applications to match the structure of your external data sources. While the parameters of an application might usually represent columns in a data flow, you can make them represent whatever structure is most appropriate to your external metadata.

**Related concepts:**

“Data lineage, business lineage, and impact analysis reports” on page 67

Data lineage and impact analysis reports show the movement of data within a job or through multiple jobs. These reports also show the order of activities within a run of a job. Business lineage reports show a scaled-down view of lineage without the detailed information that is not needed by a business user. Impact analysis reports show the dependencies between assets.

**Related reference:**

“Asset types that are included in lineage and analysis reports” on page 69

Certain asset types can be used to run data lineage, business lineage, or impact analysis reports.

**Information assets in the metadata workbench**

The objects stored in the metadata repository of InfoSphere Information Server are referred to as information assets in the metadata workbench. Each asset is an instance of an asset type.

Users and administrators of the metadata workbench find and display particular information assets, investigate the properties and relationships, and run reports on the assets.

Each asset has its own asset information page that displays the following information, depending on the type of asset:

• Properties of the asset
• Relationships that associate the asset to other assets such as synonyms, “Has A” relationships, and “Is A” relationships
• Related objects, including terms, stewards, mappings, rules, and objects that contain the asset
• Information that is generated by running automated services and performing manual linking actions
• Notes about the asset
• Blueprints that link to the asset
• Actions that you can perform on the asset

When your mouse pointer hovers over a link to an information asset in a results page or in an asset information page, the context of that asset is displayed. For example, if you hover over the link to a job, the names of the project and engine that contain the job are displayed. You can click either name to open the asset information page for that asset.

Related concepts:
“Asset types that are displayed in the metadata workbench” on page 35
Asset information pages show properties, relationships, and available actions that are appropriate to the selected type of asset.

Related tasks:
Chapter 6, “Creating data lineage, business lineage, and impact analysis reports in the metadata workbench,” on page 67
Users of the metadata workbench can create reports that analyze the flow of data from data sources, through jobs and stages, and into databases, data files, and business intelligence reports. You can report on the dependencies between assets of certain types. In addition, you can create a business lineage report that displays only the flow of data, without the details of a full data lineage report.

“Querying the metadata repository” on page 59
You can use queries to find and report on objects in the metadata repository.

“Finding assets in the metadata workbench” on page 58
You locate particular assets to investigate their properties and relationships and to run reports on them. You can browse, search, and query to locate assets.

**Information asset actions**

You can edit an asset, run reports on an asset, and take other actions that are specific to the asset. These actions are from the asset information page and from the right-click menu for an asset.

The available actions for an asset are displayed in these places:
• The right side of the asset information page
• The menu when you right-click an asset name in a list

Depending on the type of asset, some or all of the following actions are displayed:

**Add Note**
Create a note for the selected asset. Not all assets can have notes. An asset can have multiple notes.
You can remove the note assignment in IBM InfoSphere Business Glossary.

**Add to Favorites**
Add the URL of the asset information page for the selected asset to the Favorites list or Bookmarks list of your web browser.

**Assign to Label**
Assign a label to the asset.
Labels are simple, short descriptors for assets that have something in common that is meaningful to your enterprise. Labels can be used in search and filtering. You can think of labels as keywords or tags that you apply to assets.

Labels are created, edited, and deleted in InfoSphere Business Glossary.

**Assign to Rule ( Governed by )**

Assign a rule to govern the selected asset.

In the metadata workbench, you can assign only one rule to an asset. If a rule is already assigned to an asset, the rule that you select from the list replaces the original rule.

You can remove the rule assignment in InfoSphere Business Glossary.

**Assign to Rule (Implements)**

Assign a rule to the selected asset that implements the rule.

In the metadata workbench, you can assign only one rule to an asset. If a rule is already assigned to an asset, the rule that you select from the list replaces the original rule.

You can remove the rule assignment in InfoSphere Business Glossary.

**Assign to Steward**

Assign a steward to manage the selected asset.

Stewards that you assign are also displayed in InfoSphere Business Glossary, IBM InfoSphere FastTrack, and IBM InfoSphere Information Analyzer.

In the metadata workbench, you can assign only one steward to an asset. If a steward is already assigned to an asset, the steward that you select from the list replaces the original steward.

If a steward was assigned to the asset in IBM InfoSphere FastTrack or in InfoSphere Information Analyzer, more than one steward might be displayed. If an asset has more than one steward assigned to it and you assign a steward to the asset in the metadata workbench, then the previously assigned stewards are replaced by the single steward.

You can remove the steward assignment in InfoSphere Business Glossary.

**Assign to Term**

Assign a term to the selected asset. An asset can have multiple terms assigned to it. Terms are created in InfoSphere Business Glossary and IBM InfoSphere FastTrack. You can remove the term assignment in InfoSphere Business Glossary.

**Business Lineage**

Create a report that displays a business view of data lineage.

For this option to be displayed, the administrator of IBM InfoSphere Metadata Workbench must configure the asset to be included in business lineage reports. In addition, you must have at least the Business Glossary User role.

The following information is displayed, depending on the type of asset:

- The flow of data to or from the selected asset, through columns, and into business intelligence (BI) reports.
- The flow of data to or from the selected asset through database tables, database views, or data file structures, and into BI reports.
• The context, steward, and assigned terms of each asset in the report.

You cannot drill down into an asset in the business lineage report to get more information.

Copy Name
Copy the name of the asset to the clipboard.

Copy Shortcut
Copy the URL of the asset information page for the selected asset to the clipboard.

Data Lineage
Create a report that displays any of the following information, depending on the type of asset:
• The flow of data to or from the selected asset, through columns and stages, through one or more jobs, and into business intelligence (BI) reports.
• The flow of data to or from the selected asset through one or more jobs, through database tables, database views, or data file structures, and into BI reports.

Delete
Delete the asset from the metadata repository.

Download Blueprint
Download a blueprint that was published to the metadata repository to a directory that you select. Alternatively, you can open the blueprint in the Blueprint Viewer window.

Edit
Edit the description of the selected asset. You can add images to the asset information page for some types of assets. You can assign custom attributes to some types of assets.

Edit Aliases (Business Name)
Assign an alias name for the asset that gives a business meaning to the asset. For example, a BI report whose name is Source10 can be assigned the alias business name Risk_Level_3.

Edit Note
Edit every field of the note.

Exclude from (or Include in) Business Lineage
Remove the asset from business lineage reports if the asset is currently included. Include the asset in business lineage reports if it is currently excluded.

For this option to be displayed, you must have the Metadata Workbench Administrator role.

Graph View
Display a graphical model view of the asset, its related assets, and the relationships to them.

Model View
Display a model view of the asset that shows the relationship between logical data model, physical data model, and implemented data resource asset types.

Impact Analysis
Create a report that displays the assets that depend on the presence of the selected asset and the assets whose presence the selected asset depends on.
Open Details
Open the asset information page in the current window. Any changes that you made in the current window but did not save are lost.

Open Details in New Window
Open the asset information page in a new tab in the current window. The current window remains open.

Remove Note
Delete the note from the asset. The name of the deleted note is displayed with a strikethrough mark. When the page is refreshed, the deleted note is not displayed.

View Blueprint
Display the blueprint in a new Blueprint Viewer window. The current window remains open. You can open subdiagrams to view them.

You can also display context for a specific asset by moving your mouse pointer over the link to the asset. The hover help typically displays assets that contain the selected asset. For example, if you hover over the link to a job, the names of the project and engine that contain the job are displayed, and you can click either name to open the asset information page for that asset.

Asset types that are displayed in the metadata workbench
Asset information pages show properties, relationships, and available actions that are appropriate to the selected type of asset.

<table>
<thead>
<tr>
<th>Asset type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>A comment that is created by developers of IBM InfoSphere DataStage and QualityStage jobs.</td>
</tr>
<tr>
<td>Application</td>
<td>An extended data source asset that represents a program designed to perform a specific function directly for the user or for another application. Applications are the general collection of methods and parameters for reading or writing data.</td>
</tr>
<tr>
<td>BI collection</td>
<td>A structure that organizes data within a business intelligence (BI) model. BI collections are the data sources of BI reports.</td>
</tr>
<tr>
<td>BI collection member</td>
<td>The basic abstraction of a data value that is projected from a database column.</td>
</tr>
<tr>
<td>BI hierarchy</td>
<td>An organizational structure that defines an ordering or relationship of data within a BI collection.</td>
</tr>
<tr>
<td>BI level</td>
<td>A logical step within a BI hierarchy.</td>
</tr>
<tr>
<td>BI model collection</td>
<td>A grouping of BI collections that are relevant to a BI application.</td>
</tr>
<tr>
<td>BI report field</td>
<td>A field in a BI report that is typically sourced from a database column. Some BI report fields, including page numbers and section headers, are not data fields.</td>
</tr>
</tbody>
</table>
Table 2. Asset types that are displayed in the metadata workbench (continued)

<table>
<thead>
<tr>
<th>Asset type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>“BI report” on page 42</td>
<td>A report that is sourced from a database and is imported into the metadata repository of IBM InfoSphere Information Server.</td>
</tr>
<tr>
<td>Blueprint icon</td>
<td>A collection of diagrams that represents an information architecture for a project. A blueprint can be published to the metadata repository from IBM InfoSphere Blueprint Director.</td>
</tr>
<tr>
<td>Category</td>
<td>A type of directory or folder that contains and references terms and that organizes the InfoSphere Business Glossary in a hierarchy. A category can also contain other categories.</td>
</tr>
<tr>
<td>Column analysis summary</td>
<td>An IBM InfoSphere Information Analyzer process that describes the condition of data at the field level.</td>
</tr>
<tr>
<td>Column definition</td>
<td>A column-level data definition that stores data values within a InfoSphere DataStage and QualityStage table definition.</td>
</tr>
<tr>
<td>Column mapping</td>
<td>A row in an IBM InfoSphere FastTrack mapping specification that describes a transformation from one or more source columns and terms to one or more target columns and terms.</td>
</tr>
<tr>
<td>Connector</td>
<td>A software component that provides access from InfoSphere DataStage and QualityStage to an external source of data.</td>
</tr>
<tr>
<td>Custom attribute</td>
<td>A user-created attribute that stores additional information about glossary terms and categories in the metadata repository.</td>
</tr>
<tr>
<td>Data element</td>
<td>A user-defined data type.</td>
</tr>
<tr>
<td>“Data file” on page 48</td>
<td>A software file that stores data in the form of data file structures.</td>
</tr>
<tr>
<td>Data file field</td>
<td>A field within a data file structure. A data file field is equivalent to a database column and is the smallest data unit that is used to store the data values of an object.</td>
</tr>
<tr>
<td>Data file structure</td>
<td>A collection of data file fields in a data file. A data file structure is the file equivalent of a database table.</td>
</tr>
<tr>
<td>“Database” on page 48</td>
<td>A relational storage collection that is organized by schemas and procedures. A database stores data that is represented by tables.</td>
</tr>
<tr>
<td>Database column</td>
<td>A column in a database table.</td>
</tr>
<tr>
<td>Database connection</td>
<td>A connection for accessing a database or file, for example, an ODBC or Oracle connection.</td>
</tr>
<tr>
<td>Data rule definition</td>
<td>The data rule definitions of an InfoSphere Information Analyzer project that are assigned to a term.</td>
</tr>
</tbody>
</table>
Table 2. Asset types that are displayed in the metadata workbench (continued)

<table>
<thead>
<tr>
<th>Asset type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data rule set definition</td>
<td>A group of multiple data rule definitions of an InfoSphere Information Analyzer project that use the same source data.</td>
</tr>
<tr>
<td>Extension mapping</td>
<td>An extended data source asset that represents an external flow of data from one or more sources to one or more targets.</td>
</tr>
<tr>
<td>Extension mapping document</td>
<td>An extended data source asset that is a document with extension mappings.</td>
</tr>
<tr>
<td>File</td>
<td>An extended data source asset that represents a storage area for capturing, transferring, or reading data. A file is typically loaded and moved by using an FTP process. A file is often the source of ETL transactions.</td>
</tr>
<tr>
<td>Folder</td>
<td>A user-defined tree structure for storing the contents of InfoSphere DataStage and QualityStage projects.</td>
</tr>
<tr>
<td>Foreign key</td>
<td>A non-unique identifier that defines a relationship between 2 database tables. A foreign key in one table typically matches the primary key in the related table.</td>
</tr>
<tr>
<td>Foreign key definition</td>
<td>A foreign key relationship between pairs of table definitions.</td>
</tr>
<tr>
<td>Host</td>
<td>A computer that hosts databases or data files.</td>
</tr>
<tr>
<td>Host (Engine)</td>
<td>A computer that hosts the engine components of IBM InfoSphere Information Server products. The engine runs parallel, server, and sequencer jobs to extract, transform, load, and standardize data. The engine computer can also host databases and data files.</td>
</tr>
<tr>
<td>IMS™ database</td>
<td>The root object that defines an IMS database and its organization.</td>
</tr>
<tr>
<td>IMS field</td>
<td>A field of an IMS segment.</td>
</tr>
<tr>
<td>IMS segment</td>
<td>An IMS segment type, its position within the IMS hierarchy, and its relationships to other segments.</td>
</tr>
<tr>
<td>In parameter</td>
<td>An extended data source asset that delivers information from a client to a stored procedure definition.</td>
</tr>
<tr>
<td>IBM InfoSphere Information Server report</td>
<td>A report that is created and saved in the console or the Web console.</td>
</tr>
<tr>
<td>Information Governance Policy</td>
<td>A high-level, natural-language description of a governance subject area. Information governance policies are created in IBM InfoSphere Business Glossary.</td>
</tr>
<tr>
<td>Information Governance Rule</td>
<td>A language-based description of the criteria used to determine whether information assets are compliant with business objectives. Information governance rules are created in IBM InfoSphere Business Glossary.</td>
</tr>
<tr>
<td>Asset type</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Information service</td>
<td>A single operation or a collection of operations that exposes results from processing by information providers.</td>
</tr>
<tr>
<td>Information services application</td>
<td>A container for a set of services in IBM InfoSphere Information Services Director. All services within a single application are deployed or undeployed together.</td>
</tr>
<tr>
<td>Information services operation</td>
<td>A container for the business logic of an information service. The operation describes the actual task that is performed by the information provider. Examples of operations include jobs, IBM InfoSphere Federation Server queries, or the invocation of a stored procedure in an IBM DB2® database.</td>
</tr>
<tr>
<td>Information services project</td>
<td>A collaborative environment in IBM InfoSphere Information Services Director that contains applications, services, and operations.</td>
</tr>
<tr>
<td>InOut parameter</td>
<td>An extended data source asset that represents a parameter that combines the input parameter and the output parameter.</td>
</tr>
<tr>
<td>Input parameter</td>
<td>An extended data source asset that delivers information from a client.</td>
</tr>
<tr>
<td>&quot;Job&quot; on page 52</td>
<td>An InfoSphere DataStage and QualityStage job design specification. There are several types of jobs:</td>
</tr>
<tr>
<td></td>
<td>Mainframe job</td>
</tr>
<tr>
<td></td>
<td>Parallel job</td>
</tr>
<tr>
<td></td>
<td>Sequence job</td>
</tr>
<tr>
<td></td>
<td>Server job</td>
</tr>
<tr>
<td>Job run</td>
<td>A collection of the activities that is generated when a compiled job is run.</td>
</tr>
<tr>
<td>Job run activity</td>
<td>A single activity of a job run.</td>
</tr>
<tr>
<td>Job run event</td>
<td>The outcome of a job run or job run activity. Events indicate the set of resources that are affected by a job run, for example, the number of rows that are read from a particular table. Event types include read, write, and fail. Fail events have the following icon:</td>
</tr>
<tr>
<td>Link</td>
<td>The path that links and defines the flow of data between two stages in a job.</td>
</tr>
<tr>
<td>Local container</td>
<td>A grouping of job content and logic, such as stages and links, that can be reused within the same job.</td>
</tr>
</tbody>
</table>
Table 2. Asset types that are displayed in the metadata workbench (continued)

<table>
<thead>
<tr>
<th>Asset type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical data model</td>
<td>A set of related entities and their business associations that is defined in an entity-relationship model. A logical data model can be implemented by a physical data model or by a database schema.</td>
</tr>
<tr>
<td>Machine profile</td>
<td>The paths and parameters for accessing a mainframe computer. Machine profiles are created in IBM InfoSphere DataStage and QualityStage Designer.</td>
</tr>
<tr>
<td>Mapping project</td>
<td>A container that organizes mapping specifications and associated data resources in IBM InfoSphere FastTrack.</td>
</tr>
<tr>
<td>Mapping specification generation</td>
<td>A container for a set of mappings in IBM InfoSphere FastTrack. The mapping specification describes how data is extracted, transformed, or loaded from one data source to another.</td>
</tr>
<tr>
<td>Method</td>
<td>An extended data source asset that represents a function or a procedure. A method can contain input parameters and output values. A method either sends information by using an input parameter asset or receives information by using an output value asset.</td>
</tr>
<tr>
<td>Notes</td>
<td>Notes about an asset that are created by the user.</td>
</tr>
<tr>
<td>Object type</td>
<td>An extended data source asset that represents a grouping of methods or a defined data format that characterizes the input and output structures within a single application. For example, an object type could represent a common feature or business process within an application.</td>
</tr>
<tr>
<td>Out parameter</td>
<td>An extended data source asset that returns data to the stored procedure definition asset.</td>
</tr>
<tr>
<td>Output value</td>
<td>An extended data source asset that returns data to the client or to an application asset. An output value is the returned value for the database column or data field data.</td>
</tr>
<tr>
<td>Parameter</td>
<td>The runtime value or the default design value for a parameter that is used in a job, stored procedure, or stage type.</td>
</tr>
<tr>
<td>Parameter set</td>
<td>A group of job parameters that are used together and can be reused.</td>
</tr>
<tr>
<td>Physical data model</td>
<td>A design schema for information assets that defines the physical structures and relationships of data within a subject domain or application. Physical data models are independent of implementation or platform details. They can implement a logical data model and can be implemented by a database schema or by a data file.</td>
</tr>
<tr>
<td>Asset type</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Policy</td>
<td>Documentation of the constraints and maintenance procedures that apply to a particular object. A policy documents and captures additional information about business rules and processes.</td>
</tr>
<tr>
<td>Primary key</td>
<td>A unique identifier of a database table that can also be used to define relationships between tables.</td>
</tr>
<tr>
<td>Result column</td>
<td>An extended data source asset that represents the data that is returned from a database query.</td>
</tr>
<tr>
<td>Routine</td>
<td>A built-in or user-defined routine that is called in a derivation or constraint, or that is called before or after a job or stage.</td>
</tr>
<tr>
<td>“Database schema”</td>
<td>A named collection of related database tables and integrity constraints. A schema defines all or a subset of the data that is in a database.</td>
</tr>
<tr>
<td>Data server</td>
<td>A supertype of host that includes data servers and engines.</td>
</tr>
<tr>
<td>Shared container</td>
<td>A grouping of job content and logic, such as stages and links, that can be used by multiple jobs.</td>
</tr>
<tr>
<td>“Stage”</td>
<td>A component instance that performs a unit of work within an InfoSphere DataStage and QualityStage job or container. There are separate icons for each type of stage.</td>
</tr>
<tr>
<td>Stage column</td>
<td>A flow variable or column that is used to denote data flow items within a link or stage.</td>
</tr>
<tr>
<td>Stage type</td>
<td>A component type that provides the implementation and structure of a stage. Each stage is associated with a stage type.</td>
</tr>
<tr>
<td>Stage variable</td>
<td>A type of stage that is defined by IBM InfoSphere DataStage and that typically has an action, such as concatenate or a calculation, associated with it. The output of a stage variable is mapped to a stage column. Stage variables can be reused in the stage and can be used for lookup data.</td>
</tr>
<tr>
<td>Standardization component</td>
<td>A component file in a standardization rule set.</td>
</tr>
<tr>
<td>Standardization rule set</td>
<td>A series of customizable files that define how to process input data for the Standardize and Investigate stages in IBM InfoSphere QualityStage.</td>
</tr>
<tr>
<td>“Steward”</td>
<td>A user or group who is designated as responsible for one or more information assets in the metadata repository. stewards are created and managed in IBM InfoSphere Business Glossary.</td>
</tr>
<tr>
<td>Stored procedure</td>
<td>A procedure that is stored in the database to encode behavioral aspects of data manipulation.</td>
</tr>
<tr>
<td>Asset type</td>
<td>Definition</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Stored procedure definition</td>
<td>An extended data source asset that represents a procedure that is stored in the database to encode behavioral aspects of data manipulation, for example, assertions, constraints, and triggers. A stored procedure definition can also produce data in tabular form.</td>
</tr>
<tr>
<td>Table analysis summary</td>
<td>An InfoSphere Information Analyzer process that consists of primary key analysis and the assessment of multicolumn primary keys and potential duplicate values.</td>
</tr>
<tr>
<td>Table definition</td>
<td>A table-level data definition that structures data values within an InfoSphere DataStage and QualityStage project. Table definitions contain column definitions.</td>
</tr>
<tr>
<td>“Term” on page 44</td>
<td>A word or phrase that classifies one or more information assets that are in the metadata repository. Each term has a parent category. Terms and categories are created and managed in IBM InfoSphere Business Glossary.</td>
</tr>
<tr>
<td>Term History</td>
<td>The changes that were made to the description or other properties of a term since the term was first defined. Term history is displayed in IBM InfoSphere Business Glossary.</td>
</tr>
<tr>
<td>Transforms Function</td>
<td>A built-in or user-defined macro expression that is used in a derivation or constraint in InfoSphere DataStage and QualityStage.</td>
</tr>
<tr>
<td>“Transformation project” on page 55</td>
<td>The root of the InfoSphere DataStage and QualityStage folder tree. Projects hold collections of objects such as jobs, stages, and table definitions.</td>
</tr>
<tr>
<td>User group</td>
<td>A group of users of IBM InfoSphere Information Server. User groups are designated in the Administration tab of the Web console.</td>
</tr>
<tr>
<td>View</td>
<td>A dynamic or virtual database table whose data is computed or collated.</td>
</tr>
<tr>
<td>Warehouse mapping document</td>
<td>An extended data source asset that is a document with warehouse mappings from IBM InfoSphere Warehouse. A warehouse mapping represents an external flow of data from one or more source databases to one or more target databases.</td>
</tr>
</tbody>
</table>
The objects stored in the metadata repository of InfoSphere Information Server are referred to as information assets in the metadata workbench. Each asset is an instance of an asset type.

**Related tasks:**
- Chapter 6, “Creating data lineage, business lineage, and impact analysis reports in the metadata workbench,” on page 67
- “Querying the metadata repository” on page 59
- “Finding assets in the metadata workbench” on page 58
- “Creating queries” on page 62
- “Managing queries” on page 64

**BI report**
A business intelligence (BI) report is the metadata structure of a business intelligence report that is sourced from a database.

BI reports are instances of the ReportDef class of the Business Intelligence model.

Use a bridge to import BI reports into the metadata repository.

You can perform the following actions in the metadata workbench:
- Run impact analysis reports and lineage reports that trace the flow of information through jobs, stages, and databases into BI reports.
- Assign a term, steward, or label to a BI report.
- Add an image or edit the description in the asset information page of the BI report.

The asset information page for a BI report lists the properties of the BI report and displays related assets of the following types:

**BI Report**
Displays the general properties, the business name or the alias name of the BI report, whether the BI report is included in business lineage reports, the label of the report that describes its business meaning, the terms that are assigned to the report, the steward assigned to the report, and the database tables that are sources for the report. Displays rules that govern the BI report and rules that are implemented by the BI report.
Report fields are the database columns of the report. BI report collection is the group of database columns and user-defined columns of the database table that is used to build the report.

**Attributes**
Displays the custom attributes of the asset and their value.

**Blueprints**
Displays blueprints that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in a new tab of the Blueprint Viewer window.

**Notes**
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Extension Mappings**
Lists any extension mapping documents that include the asset in the source or target columns.

**Modification Details**
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

**Category**
A category is a directory or folder that contains a set of glossary terms. Typically, the terms contained in each category are related in some way that is meaningful to your organization. You can use IBM InfoSphere Business Glossary to define categories.

A category is an instance of the Category class in the Business Glossary Model.

To edit the category or to assign a steward to the category, use InfoSphere Business Glossary.

The asset information page for a category contains the following information:

**Category**
Displays the general properties of the category, including name, short and long descriptions, and steward.

**Attributes**
Displays the custom attributes of the asset and their value.

**Terms**
Displays a list of all the terms contained in this category and a list of terms that this category refers to. Click a term to display the details of that term.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is
highlighted. If the element is in a subdiagram of the blueprint, the
subdiagram is displayed in the Blueprint Viewer window.

Notes Displays notes about the asset that were created in the metadata
workbench or in other products in IBM InfoSphere Information Server.
Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

Modification Details
Displays the name of the user who created or last modified the asset. In
addition, displays the date and time of creation and last modification. This
section also displays what properties were changed.

Term
You use a term to classify, define, and group assets according to the needs of the
enterprise. A term is sometimes referred to as a glossary term or a business term.

A term is an instance of the Term class in the Business Glossary Model.

Terms are contained within categories, which make up the structure of the business
glossary. In IBM InfoSphere Metadata Workbench, IBM InfoSphere FastTrack, and
IBM InfoSphere Information Analyzer, you can assign terms to other assets. Terms
can be assigned to multiple assets, and assets can be assigned to multiple terms.

You can edit terms in IBM InfoSphere Business Glossary.

The asset information page for a term includes the following categories of
information:

Term Displays the properties of the term such as the context, or hierarchy, of the
parent category. Also displays the categories that refer to the term, its
steward, labels, and status.

Attributes Displays the custom attributes of the asset and their value.

Associated Terms Displays terms that are a synonym of, are related to, are replaced by, or are
assigned to this term.

In addition, the following relationships are displayed:

Is A Type Of
The concept of this term is an instance of the concept expressed by
another term that is typically broader in scope.
For example, the term Home Loan might have the Is A Type Of
relationship to the term Loan Info because a home loan is a type of
loan.
The Is A Type Of relationship is the inverse of the Has Type
relationship.

Has Type
The concept of this term has one or more subtypes that are
expressed by other terms.
For example, the term Loan Info might have the Has Type
relationship to the term Home Loan because one type of loan is a
home loan.
The Has Type relationship is the inverse of the Is A Type Of relationship.

Is Of  The concept of this term is part of a broader or larger concept expressed by another term.

For example, the term Home Loan might have the terms Hybrid_Rate, Fees, and Margin to indicate that hybrid rates, fees, and margins are parts of a home loan. These three terms have the Is Of relationship to the term Home Loan.

The Is Of relationship is the inverse of the Has A relationship.

Has A  The concept of this term is composed of one or more terms that are parts of this term.

For example, if term Home Loan is composed of terms Hybrid_Rate, Fees, and Margin, then the term Home Loan has the Has A relationship with terms Hybrid_Rate, Fees, and Margin.

The Has A relationship is the inverse of the Is Of relationship.

Assigned Assets  Displays assets that the term is assigned to. The assigned asset can be in the metadata repository, or external to the metadata repository. Displays mapping documents that reference the term.

Displays unpublished IBM InfoSphere FastTrack column mapping specifications that use the term.

Linked to Blueprint Elements  Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes  Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:

• Display the note in a new window
• Edit or delete the note, if you created it

Term History  Displays the name of the user who created or modified the term, and the properties that were changed.

Modification Details  Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Data file  A data file is a file-system storage medium for data. Data files contain one or more data file structures, which are the file equivalents of database tables.

A data file is an instance of the DataFile class in the Common Model.
You can edit the description and alias (business name). You can also assign a term, steward, or label. In addition, you can include or exclude from business lineage reports, and run lineage and impact analysis reports on the asset.

The asset information page for a data file includes the following categories of information:

**File**
Displays the properties of the data file, the server name and directory path where the data file is located, the business name or the alias name of the data file, whether the file is included in business lineage reports, the label of the file that describes its business meaning, and the data file structures that the data file contains.

Displays rules that govern the file and rules that are implemented by the file.

**Attributes**
Displays the custom attributes of the asset and their value.

**File Design Usage**
Displays jobs that read from or write to the data file, based on job design information that is interpreted by the Manage Lineage services.

**File Operational Usage**
Displays jobs that read from or write to the data file at run time, based on operational metadata that is interpreted by the Manage Lineage services.

**File User-Defined Usage**
Displays jobs that read from or write to the data file, based on the results of manual linking actions of the Manage Lineage services. Also displays any extension mapping documents with mappings between source and target assets.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

**Notes**
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Policy**
Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

**Modification Details**
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

**Data file structure**
A data file structure is created when you import a data file. It contains information about the structure of the imported file.
A data file structure is an instance of the DataCollection class in the Common Model.

You can edit the description and alias (business name). You can also assign a term, steward, or label. In addition, you can run lineage and impact analysis reports on the asset.

The asset information page for a data file structure contains the following information:

**File Structure**
Displays the general properties of the data file structure, including the name of the structure, the business name or the alias name of the data file structure, whether the data file structure is included in business lineage reports, the label of the file that describes its business meaning, short and long description, name of terms that are assigned to the structure, fields in the structure, and steward.

Displays rules that govern the file structure and rules that are implemented by the file structure.

**Attributes**
Displays the custom attributes of the asset and their value.

**File Structure Design Information**
Displays stages that write to and read from the data file structure, based on job design information that is interpreted by the Manage Lineage services.

**File Structure Operational Information**
Displays jobs that read from or write to the data file structure at run time, based on operational metadata that is interpreted by the Manage Lineage services.

**File Structure User-Defined Information**
Displays stages that read from or write to the data file structure, based on the results of manual linking actions that are performed by the Metadata Workbench Administrator. Also displays any extension mapping documents with mappings between source and target assets.

**Policy**
Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

**Notes**
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Modification Details**
Displays the name of the user who created or last modified the asset. In
addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Host
A host is the hardware that hosts a database, a data file, or a project whose metadata is imported into or referenced by IBM InfoSphere Information Server. A host can have both databases and engines.

Hosts are instances of the HostSystem class in the Common Model.

You can assign a steward or a label to a host.

The asset information page for a host includes the following information:

Host Displays the properties of the host, the network node, and the databases and data files that the host sources. In addition, display rules that govern the host and rules that are implemented by the host.

Attributes Displays the custom attributes of the asset and their value.

Linked to Blueprint Elements Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes Displays notes about the host that are created in IBM InfoSphere Business Glossary or in IBM InfoSphere Information Analyzer.

Right-click the name of the note to do the following actions to the note:
• Display the note in a new window
• Edit or delete the note, if you created it

Database
A database is a relational database or catalog that stores data that is defined by database tables and schemas.

A database is an instance of the Database class in the Common Model.

You can run impact analysis and lineage reports. You can assign a term, steward, and a label to a database.

The asset information page for a database includes the following information:

Database Displays the properties of the database, the server that hosts the database, the business name or the alias name of the database, and whether the database is included in business lineage reports. In addition, displays the label of the database that describes its business meaning, rules that govern the database, rules that are implemented by the database, and the schemas.

Attributes Displays the custom attributes of the asset and their value.
Database Design Information
Displays jobs that read from or write to the database, based on job design information that is interpreted by the Manage Lineage services.

Database Operational Information
Displays jobs that read from or write to the database at run time, based on operational metadata that is interpreted by the Manage Lineage services.

Database User-Defined Information
Displays jobs that read from or write to the database, based on the results of manual linking actions that are performed by the Metadata Workbench Administrator. Also displays any extension mapping documents with mappings between source and target assets.

BI Report Information
Displays business intelligence (BI) reports and BI report models that are sourced from the database tables. Right-click the name of the report or the report model to display a list of additional tasks that you can do.

Linked to Blueprint Elements
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

Policy
Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

Alias
Displays alternative names that InfoSphere DataStage jobs use to reference the database. The alias name is defined by the manual linking action, Database Alias.

Modification Details
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Database table
A database table is the structure that represents and stores columns within a schema. The metadata workbench displays information about schemas that were imported into IBM InfoSphere Information Server.

A database table is an instance of the DataCollection class in the Common Model.

You can edit the description and alias (business name) of the table. You can also assign a term, steward, or label to be associated with the table. In addition, you can run impact analysis, lineage, and model view reports.

The asset information page for a database table contains the following information:
Database Table
Displays the general properties of the table: name, the business name or the alias name of the database table, whether the database table is included in business lineage reports, the label of the database table that describes its business meaning, tool that created the database table, short and long description, term assigned to the table, and steward.

Displays the name of the database that contains the table, the name of the schema, the name of any identical tables, and the name of any view that refers to the database table.

Displays rules that govern the database table and rules that are implemented by the database table.

Attributes
Displays the custom attributes of the asset and their value.

Database Table Design Information
Displays the stages that write to and read from this table, based on the job design information that is interpreted when Manage Lineage services are run.

Database Table Operational Information
Displays the stages that write to and read from this table, based on the values of parameters at run time. This operational metadata is interpreted by Manage Lineage services.

Database Table User-Defined Information
Displays stages that write to and read from this table, based on the results of manual linking actions. Also displays any extension mapping documents with mappings between source and target assets.

BI Report Information
Displays the names of the business intelligence (BI) reports and the report collections that use information from this table.

Indexes and Analysis
Displays the primary key that is defined for the database table and the foreign keys that the database table refers to. Foreign keys are fields in a different table. These keys relate tables to each other. The name of the column and of the table where the foreign key is sourced are also displayed.

Displays the IBM InfoSphere Information Analyzer analysis summary report of the database table.

Linked to Blueprint Elements
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
• Display the note in a new window
• Edit or delete the note, if you created it
Policy  Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

Mapping Specifications  
Displays source and target mapping specifications from IBM InfoSphere FastTrack that refer to or use the database table.

Modification Details  
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Database schema  
A database schema is composed of database tables and can include all or part of the data in the database. The layout of a database schema outlines the way data is organized into tables.

A database schema is an instance of the DataSchema class in the Common Model.

You can edit the descriptions and alias (business name) of the database schema. You can also assign an image, term, steward, or label to the database schema. In addition, you can include the database schema in business lineage reports, as well as run impact analysis, lineage, and model view reports.

The asset information page for a database schema contains the following information:

Database Schema  
Displays the general properties of the database schema, including name, the business name or the alias name of the database schema, whether the database schema is included in business lineage reports, owner, short and long description, and steward. In addition, displays host server, database name, database tables, the label of the database schema that describes its business meaning, rules that govern the schema, rules that are implemented by the schema, and stored procedures.

Attributes  
Displays the custom attributes of the asset and their value.

Linked to Blueprint Elements  
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes  
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
• Display the note in a new window
• Edit or delete the note, if you created it

Policy  Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

Schema User-Defined Information  
Displays a list of extension mapping documents that have mappings between source and target assets.
Modification Details
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

View
A view is a virtual database table and can be imported into the metadata repository.

A view is an instance of the DataCollection class in the Common Model.

You can edit the description and alias (business name) of the view. You can also assign a term, label, or steward to the view.

The asset information page for a view contains the following information:

View Displays the general properties of the view: the view name, the business name or the alias name of the view, whether the view is included in business lineage reports, product that imported the view, the label of the view that describes its business meaning, the database that the view is derived from, a description of the view, the SQL expression that created the view, and the steward who manages the view.

Displays the database and schema that contain the view, and the columns that are defined in the view.

Displays rules that govern the view and rules that are implemented by the view.

Attributes Displays the custom attributes of the asset and their value.

View Design Information Displays the stages that write to and read from this database view, based on the design parameters of the job that are interpreted by the Manage Lineage services.

View Operational Information Displays the stages that write to and read from this database view at runtime, based on operational metadata that is interpreted by the Manage Lineage services.

View User-Defined Information Displays the stages that write to and read from this database view, based on the results of manual linking actions that are performed by the Metadata Workbench Administrator.

BI Report Information Displays the names of the business intelligence (BI) reports that contain data that is obtained from this view.

Indexes and Analysis Displays the primary key that is defined for the view and the foreign keys that the view refers to. Foreign keys are fields in a different view. These keys relate views to each other. The names of the column and of the view where the foreign key is sourced are also displayed.

Displays the IBM InfoSphere Information Analyzer analysis summary report of the view.
**Mapping Specifications**
Displays the IBM InfoSphere FastTrack source and target mapping specifications that refer to and use the view.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

**Notes**
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.
Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Policy**
Displays policies that are associated with the asset. Policies are rule sets that are created in IBM InfoSphere Information Analyzer.

**Modification Details**
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

**Job**
A job is a design specification that is created in IBM InfoSphere DataStage and QualityStage Designer to extract, transform, or load data. The job can be a DataStage job or a QualityStage job.

A job is an instance of the DSJob class in the Transformation model.

The job can be a parallel, server, or mainframe job.

You can run impact analysis and lineage reports. You can assign a term, steward, or label to a job.

The asset information page for a job includes the following information:

**Image**
Displays the job as it is displayed in the Designer client.

**Job**
Displays the properties of the job, the project and folder that it is in, and the stages and containers that it includes.

Displays rules that govern the job and rules that are implemented by the job.

**Job Design Information**
Displays data items that the job reads from or writes to. Displays the previous and next jobs based on job design information that is interpreted by the Manage Lineage services. Displays job design parameters and whether runtime column propagation is enabled.

**Job Operational Information**
Displays the previous and next jobs based on the values of parameters at run time, based on operational metadata that is interpreted by the Manage Lineage services.
Job User-Defined Information
Displays the data items that a job reads from or writes to, based on the results of manual linking actions that are performed by the Metadata Workbench Administrator. Also displays any extension mapping documents with mappings between source and target assets.

Sequence
Displays the job that sequenced the selected job.

Annotations
Displays annotations that are added to the job in the Designer client.

Linked to Blueprint Elements
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.
To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Notes
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.
Right-click the name of the note to do the following actions to the note:
• Display the note in a new window
• Edit or delete the note, if you created it

Mapping Specification
Displays the mapping specifications from IBM InfoSphere FastTrack that generates this DataStage job.

Information Services Usage
Displays related operations of IBM InfoSphere Information Services Director and whether a Web service is enabled.

Modification Details
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Stage
A stage is an individual step in an IBM InfoSphere DataStage job. Each stage defines a specific action or activity within the job.

A stage is an instance of the DSStage class in the Transformation model.

You can assign a term or label to the stage. The other properties of a stage are defined in InfoSphere DataStage.

The asset information page for a stage contains the following information:

Stage
Displays the general properties of the stage, including name, description, term, stage type, and job.
Displays the names of links, which contain the stage columns that are input and output for the stage. Click the twistie next to the name of the link, and then click the name of the stage column to display its details.

Stage Design Information
Displays the next and previous stages that are accessed, and the name of
the database tables or files that are written to and read from for this stage. The information is based on design parameters that are interpreted by the Managed Lineage services.

**Stage Operational Information**
Displays the next and previous stages that are accessed, and the name of the database tables written to and read from. The information is based on operational metadata that is interpreted by the Managed Lineage services.

**Stage User-Defined Information**
Displays the next and previous stages that are accessed, and the name of the database tables written to and read from. This information is based on the results of manual linking actions that are performed by the Metadata Workbench Administrator.

Also displays extension mapping documents that contain mappings between source and target assets.

**Parameters**
Displays a list of parameters and their values for the stage. Parameter values contain connection information to implemented data resources.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

**Notes**
Displays notes about the asset that were created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Modification Details**
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

**Transformation project**
A transformation project is a project that is created in IBM InfoSphere DataStage and QualityStage Designer. A transformation project contains jobs and table definitions.

A transformation project is an instance of the DSProject class in the Transformation model.

The asset information page for a transformation project includes the following information:

**Project**
Displays the name of the server on which the engine tier of IBM InfoSphere Information Server is installed and the folders or directories that contain the elements of the project.

**Includes Containers**
Displays the shared containers in this project.
Includes Jobs
Displays the jobs in this project. A job consists of stages that are linked together to describe the flow of data from a data source to a data target.

Includes IMS Databases
Displays the Information Management System (IMS) database that is used in the project.

Includes Machine Profiles
Displays the mainframe machine profiles that are used when IBM InfoSphere DataStage uploads generated code to a mainframe.

Includes Routines
Displays the routines that use a COBOL function from a library that is external to InfoSphere DataStage.

Includes Parameter Sets
Displays the set of parameters, or processing variables, for jobs.

Includes Stage Types
Displays the types of stages that are in the project.

Includes Table Definitions
Displays table definitions, a set of related columns definitions that are stored in the metadata repository and that can be loaded into stages.

Includes Transforms
Displays the list of built-in and custom transforms in the project. A transform changes data from one type to a different type.

Includes Standardization Rule Sets
Displays the standardization rule sets for specific countries. A rule set determines how fields in input records are parsed.

Linked to Blueprint Elements
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.
To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

Modification Details
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

Engine
An engine is the host on which the engine tier of IBM InfoSphere Information Server is installed. It can also be a server that hosts databases whose metadata is imported or referenced by IBM InfoSphere Information Server.

Engines are instances of the HostSystem class in the Common Model.

You can edit the short and long descriptions of an engine and attach an image. In addition, you can assign a term, steward, or label.

The asset information page for an engine includes the following information:

Engine
Displays the network node, data connection, and projects that the engine
hosts. Also lists the data connectors that implement stages that are used in a project. In addition, lists the databases and data files that the engine hosts if it is also a server.

**Attributes**
Displays the custom attributes of the asset and their value.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.

**Notes**
Displays notes about the engine that are created in the metadata workbench or in other products in IBM InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
- Display the note in a new window
- Edit or delete the note, if you created it

**Modification Details**
Displays the name of the user who created or last modified the asset. In addition, displays the date and time of creation and last modification. This section also displays what properties were changed.

**Steward**
A steward is a user or group that is responsible for assets in the metadata repository. The steward serves as the contact for information about those assets.

A steward is an instance of the Principal class in the Common Model.

Users and groups in IBM InfoSphere Information Server can be designated as stewards in IBM InfoSphere Business Glossary.

A steward can be assigned to categories and to terms by using InfoSphere Business Glossary. You can use the metadata workbench to assign a steward to other types of assets.

When you edit a steward in the metadata workbench, you can add an image to the asset information page, and you can edit the contact information for the steward.

The asset information page for a steward displays the following information:

**User**
Displays contact details of the steward.

**Manages Information Assets**
Displays assets that the steward is responsible for.

**Linked to Blueprint Elements**
Displays blueprint elements that link to the asset. Blueprints are created in IBM InfoSphere Blueprint Director.

To display the blueprint in the Blueprint Viewer window, click the blueprint link. The element of the blueprint that is linked to the asset is highlighted. If the element is in a subdiagram of the blueprint, the subdiagram is displayed in the Blueprint Viewer window.
Notes Displays notes about the steward that are created in the metadata workbench or in other products in InfoSphere Information Server.

Right-click the name of the note to do the following actions to the note:
• Display the note in a new window
• Edit or delete the note, if you created it

Modification Details
Displays the name of the user who created or last modified the steward, and the date and time corresponding to creation and last modification.

Finding assets in the metadata workbench
You locate particular assets to investigate their properties and relationships and to run reports on them. You can browse, search, and query to locate assets.

About this task
When you locate an asset, you can click it to display its asset information page, or right-click the asset name and choose a task.

You can also display assets, find assets, and run queries from the Welcome page.

Procedure
Choose a method from the following table.

<table>
<thead>
<tr>
<th>To do this action</th>
<th>Do this action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display all assets of a particular type</td>
<td>On the Discover tab, choose an asset type from the Asset Types list and then click Display.</td>
</tr>
<tr>
<td>Browse the projects and jobs that are contained by InfoSphere Information Server engines</td>
<td>On the Browse tab, click Engines.</td>
</tr>
<tr>
<td>Browse the data servers and databases whose metadata is imported into the metadata repository</td>
<td>On the Browse tab, click Hosts.</td>
</tr>
<tr>
<td>Find assets by using the name or short description</td>
<td>On the Discover tab, click Find, then select an asset type and optionally specify additional information.</td>
</tr>
<tr>
<td>Run an existing query to find assets</td>
<td>On the Discover tab, select an existing query from the Run Query list and click Run.</td>
</tr>
<tr>
<td>Create a query to find assets</td>
<td>On the Discover tab, click New Query.</td>
</tr>
</tbody>
</table>
You navigate to view assets in the metadata repository and to create reports about assets and their relationships.

The objects stored in the metadata repository of InfoSphere Information Server are referred to as information assets in the metadata workbench. Each asset is an instance of an asset type.

Asset information pages show properties, relationships, and available actions that are appropriate to the selected type of asset.

You can use queries to find and report on objects in the metadata repository.

Users of the metadata workbench can create reports that analyze the flow of data from data sources, through jobs and stages, and into databases, data files, and business intelligence reports. You can report on the dependencies between assets of certain types. In addition, you can create a business lineage report that displays only the flow of data, without the details of a full data lineage report.

Queries

Queries help users find and display information assets, their properties, and their relationships.

The metadata workbench has prebuilt queries that you can run to find information about assets. In addition, you can create and edit your own queries and use existing published queries as the basis for new queries.

Structure

Each query is based on a single asset type, but you can build queries so that they primarily display the features of related assets. For example, you might base a query on the database table asset type. But you might structure the display properties of the query to return detailed information about the business intelligence reports that are related to a single table.
You build a query by selecting an asset type. You then select the criteria and the display properties from the list of properties that are available for that asset type.

**Available Properties**

The Available Properties list displays the following information:

- Properties of the selected asset type, such as name and description.
- Possible relationships that the asset type can have to other asset types. For example, a term can have four types of relationships to other terms and two relationships to categories.
- Properties and relationships of all related asset types, and of their related asset types, and so on.

You can expand the list at any related asset type to select properties or objects to display in the query result. Alternatively, you can use criteria for returning results. For example, for a query based on database table you can specify that the results display the term that classifies the job that contains the stage that writes to the database table. Or you can specify that the query return only those tables that are written to by stages in jobs that are classified by a term that begins with the letter E.

You use the Available Properties list to populate the Criteria and Select tabs.

**Criteria**

On the Criteria tab, you specify the conditions under which results are returned. For example, for a query that is based on database tables, you might specify that the query returns only those tables that meet the following conditions:

- Read by a stage in a job
- Has no steward
- Classified by a term whose short description contains “customer”

You can add multiple conditions and subconditions, selecting properties and related objects from the Properties list to make the query results as precise as needed. By default, all conditions and subconditions must be met, but you can change the setting so that any condition can be met. You can set this value separately for each condition or each subcondition.

Depending on the type of the property, you can refine your query. For example, if the property is type text, you can narrow your query by using “Begins with,” “Is null,” “Is not,” “Contains.” If the property is type date, you can choose a date range by using “Is between” with begin and end dates. If the property is a relationship, you can choose “Is null” or “Is not null.”

**Note:** Pattern matching of text that you enter is case sensitive and affects the query results.

**Select**

On the Select tab, you specify the properties of the selected asset type. You also specify the properties of related asset types that you want to display in the query results. If you create a complex display, the query results are presented in tabs.

**Apply criteria to selected properties**

You can limit the query results to display only those assets that match the
criteria in the Criteria tab. If you select this check box, the criteria is applied to the asset on which you are making the query and on all selected relationships.

**Include the repository identifier of the asset in query results**
You can include the unique ID of the asset in the metadata repository.

**Sharing queries and results**

When a query is created, it is visible only to the user who created it. A user who has the Metadata Workbench Administrator role can share queries with other users of the metadata workbench.

The administrator can publish queries so that users of the same metadata workbench installation can see them. The administrator can also export queries in WBQ format. In this way, users of other installations can use them, and the administrator can import queries in WBQ format that users created in other installations of the metadata workbench. The WBQ format is a proprietary format, and you cannot edit the query files outside of the metadata workbench.

You can save query results in comma-separated value (CSV) and XLS format in order to use them programmatically or to distribute the results to others.

**Permissions of each role**

What you can do with queries depends on your metadata workbench role. The following table lists the permissions of each role.

*Table 3. Query tasks by role*

<table>
<thead>
<tr>
<th>Query tasks</th>
<th>Metadata Workbench Administrator role</th>
<th>Metadata Workbench User role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create queries</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Delete queries that you create</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Delete unpublished queries that others create</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Delete published queries</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>View and run prebuilt and published queries</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Publish queries so that other users can work with them</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Edit a prebuilt or published query and overwrite the original</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Edit a prebuilt or published query and save it as a new query</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Import queries in WBQ format.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Export queries in WBQ format.</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

**Example**

A prebuilt query, Job Run, returns the following information:
- All jobs that have runs
- All runs of each job
• The description, project, and steward of each job
• The status of each run

Any user can edit this query to display only those runs that occurred after a
certain data and time. A user with the Metadata Workbench Administrator role can
save the edited query and overwrite the original prebuilt query. A metadata
workbench user can save the query with a new name but cannot overwrite the
prebuilt query.

Related tasks:

Creating queries
Users of the metadata workbench can create simple and complex queries to find
assets in the metadata repository. Queries are based on the attributes and
relationships of a selected asset type.

Managing queries” on page 64
Metadata Workbench Administrators can edit, publish, import, and export queries.
Metadata Workbench Users can import queries, edit the queries that they create,
and modify published queries to save as new queries.

Running queries” on page 65
Users of the metadata workbench can run queries to find and report on
information assets.

Creating queries

Users of the metadata workbench can create simple and complex queries to find
assets in the metadata repository. Queries are based on the attributes and
relationships of a selected asset type.

About this task

You can create and save a query. You can share the query with other IBM
InfoSphere Metadata Workbench users by publishing the query.

Procedure

1. On the Discover tab, click Query.
2. From the Asset Type list, select the type of asset that you want to build the
   query on. The query returns the specified information about assets of this type
   and information about any related assets that you specify.
3. Optional: Specify criteria for returning results:

   a. On the Criteria tab, click the Add Condition button and click Add
      Condition.
   b. In the Available Properties list, select an attribute or a related asset. Related
      assets are indicated by a plus sign (+). You can expand a related asset to
      select its attributes or related assets. The selected property is displayed in
      the condition.
   c. On the Criteria tab, specify a value for the selected property.
   d. Optional: Click the numbered arrow on a condition to add subconditions or
      additional conditions. Add properties and specify values for each new
      condition or subcondition.

   Note: Pattern matching of text that you enter is case sensitive and affects
   the query results.

   e. Specify whether all or any of the criteria must be met for the query to
      return results. To change the specification, click All to change to Any, or
click Any to change to All. You must do this step for all conditions and for each individual condition or subcondition that has subconditions.

If you do not specify criteria, the query returns all assets of the type that you select in step 2.

4. Optional: Specify which results to display:
   a. In the Available Properties list, select an attribute or a related asset. You can expand a related asset to select its attributes or related assets.
   b. On the Select tab, click the Select button. The selected attribute or related asset is displayed in the Displayed Properties list.
   c. Optional: Select additional properties and add them to the list. You can reorder the displayed properties and rename the displayed properties. If you rename a displayed property, only the labels in the display results are affected by the name change.
   d. Optional: Select the Apply criteria to selected properties check box if you want to limit the query results to only those assets that match the criteria in the Criteria tab. If you select this check box, the criteria is applied to the asset on which you are making the query and on all selected relationships.
   For example, you want to create a query to display all databases that have a database table with the prefix WHS. If the Apply criteria to selected properties check box is clear, the query results include all database tables, even those database tables without the prefix WHS.

Figure 19. Query results when the Apply criteria to selected properties check box is clear

If the Apply criteria to selected properties check box is selected, the query results display only those database tables with the prefix WHS.

Figure 20. Query results when the Apply criteria to selected properties check box is selected

By default, Apply criteria to selected properties is selected in all new queries and in all queries that were created before InfoSphere Metadata Workbench, Version 8.5.

5. Optional: Save the query:
a. Click **Save**. The Save Query window opens.
b. Specify a name and description for the query.
c. Optional: Publish the query so that other users can see it and use it (administrators only). Published queries are displayed with a different icon.
d. Click **Save** in the Save Query window.

6. Optional: Click **Run**. The query runs and the results are displayed. For some complex displays, query results are presented with multiple tabs when multiple types of relationships are displayed.

### What to do next

You can save query results to a CSV or XSL file.

#### Related concepts:

“**Queries** on page 59

Queries help users find and display information assets, their properties, and their relationships.

“**Asset types that are displayed in the metadata workbench**” on page 35

Asset information pages show properties, relationships, and available actions that are appropriate to the selected type of asset.

#### Related tasks:

“**Managing queries**”

Metadata Workbench Administrators can edit, publish, import, and export queries. Metadata Workbench Users can import queries, edit the queries that they create, and modify published queries to save as new queries.

“**Running queries**” on page 65

Users of the metadata workbench can run queries to find and report on information assets.

### Managing queries

Metadata Workbench Administrators can edit, publish, import, and export queries. Metadata Workbench Users can import queries, edit the queries that they create, and modify published queries to save as new queries.

### About this task

Metadata Workbench Administrators can delete published queries. Metadata Workbench Users can delete only the queries that they create. If a query is not published, only the user who created the query can delete it.

Any user can edit a published query, but the Metadata Workbench User must save the published query with a new name as a user query, while the Metadata Workbench Administrator can change the published query.

Any user can import queries, but when a Metadata Workbench User imports a query it cannot be published.

Only the Metadata Workbench Administrator can export and publish queries.

### Procedure

1. On the Discover tab, click **Query**.
2. In the query builder, click **Manage**.
3. In the Manage Queries window, do any of the following tasks:
### To do this task | Do this action
--- | ---
**Edit a query** | 1. Select a query and click **Edit**.  
2. Modify the display properties and criteria as required.

**Change the name or description of a query** | 1. Select a query and click **Edit**.  
2. In the query builder, click **Save**.  
3. In the **Save Query** window, edit the description and click **Save**.

**Publish a query** | 1. Select a query and click **Edit**.  
2. In the query builder, click **Save**.  
3. In the **Save Query** window, select **Publish Query** and click **Save**.

**Import a query** | 1. Select a query and click **Import**.  
2. In the Import Queries window, browse to select a WBQ file and click **Import**.

**Export a query** | 1. Select a query and click **Export**.  
2. Save the query as a WBQ file.

**Delete a query** | Select a query and click **Delete**.

### Related concepts:
- **“Queries” on page 59**
  Queries help users find and display information assets, their properties, and their relationships.
- **“Asset types that are displayed in the metadata workbench” on page 35**
  Asset information pages show properties, relationships, and available actions that are appropriate to the selected type of asset.

### Related tasks:
- **“Creating queries” on page 62**
  Users of the metadata workbench can create simple and complex queries to find assets in the metadata repository. Queries are based on the attributes and relationships of a selected asset type.

- **“Running queries”**
  Users of the metadata workbench can run queries to find and report on information assets.

### Running queries
Users of the metadata workbench can run queries to find and report on information assets.

### About this task
You can also run a query when you create or edit the query.

### Procedure
1. Select the query:  
   - On the Welcome page, select the query in the **Available Queries** list.  
   - On the **Discover** tab, select the query in the **Run Query** list.
2. Click **Run**.
Related concepts:

“Queries” on page 59
Queries help users find and display information assets, their properties, and their relationships.

Related tasks:

“Creating queries” on page 62
Users of the metadata workbench can create simple and complex queries to find assets in the metadata repository. Queries are based on the attributes and relationships of a selected asset type.

“Managing queries” on page 64
Metadata Workbench Administrators can edit, publish, import, and export queries. Metadata Workbench Users can import queries, edit the queries that they create, and modify published queries to save as new queries.
Chapter 6. Creating data lineage, business lineage, and impact analysis reports in the metadata workbench

Users of the metadata workbench can create reports that analyze the flow of data from data sources, through jobs and stages, and into databases, data files, and business intelligence reports. You can report on the dependencies between assets of certain types. In addition, you can create a business lineage report that displays only the flow of data, without the details of a full data lineage report.

Related concepts:
- "Tips for navigating the metadata workbench" on page 29
  You navigate to view assets in the metadata repository and to create reports about assets and their relationships.
- "Information assets in the metadata workbench" on page 31
  The objects stored in the metadata repository of InfoSphere Information Server are referred to as information assets in the metadata workbench. Each asset is an instance of an asset type.

Related tasks:
- "Finding assets in the metadata workbench" on page 58
  You locate particular assets to investigate their properties and relationships and to run reports on them. You can browse, search, and query to locate assets.

Data lineage, business lineage, and impact analysis reports

Data lineage and impact analysis reports show the movement of data within a job or through multiple jobs. These reports also show the order of activities within a run of a job. Business lineage reports show a scaled-down view of lineage without the detailed information that is not needed by a business user. Impact analysis reports show the dependencies between assets.

You can start a report from the task list of an asset information page or from the right-click menu of an asset in a results list. You can track the flow from source to target or from target to source.

When you run reports, the metadata workbench displays information assets in the context of your enterprise goals. You see them not as isolated tables, columns, jobs, or stages, but as integrated parts of the process that extracts, loads, investigates, cleanses, transforms, and reports on your data.

Before you run reports, the Metadata Workbench Administrator must run the Manage Lineage utility and, where necessary, manual linking actions to set relationships between assets. To report on relationships that are created by operational metadata, you must first import operational metadata.

If a report does not return the expected results, take the following actions:
- Ensure that the Manage Lineage utility was run.
- Browse to the asset information page of information assets that you suspect are not properly linked. Expand the Design, Operational, and User-Defined information sections to validate that the correct relationships are set.
- Perform manual linking actions to set the necessary relationships.
Report types

You can run these types of reports:

Data Lineage and Impact Analysis

Both data lineage and impact analysis reports can show different types of information:

- The flow of data to or from a selected metadata asset, through stages and stage columns, through one or more jobs, into databases and business intelligence (BI) reports.

For example, a data lineage report might start with a database column that is read by a stage in a job. The report might show the following flow of data at the column level:
  - A stage in the first job reads the database column
  - Information flows through one or more stages in the first job until one of the stages writes to a table in a second database
  - A stage in a second job reads the database column in the second database
  - Information flows through one or more stages in the second job until one of the stages writes to a table in a third database
  - The data in the database column in the third database is captured in a BI report

- The flow of data to or from a selected metadata asset through one or more jobs, through database tables, views, or data file structures and into BI reports and information services operations.

For example, a data lineage report might show that the sources for a BI report come from three separate jobs that write to a single database table. It might then show that the table is bound to a BI report collection that is used by the BI report.

- The order of activities within a job run, including the tables that the jobs write to or read from and the number of rows that are written and read.

You can inspect the results of each part of a job run by drilling into the job run activities to see the links, stages, and database tables, or data file structures that the job reads from and writes to.

For example, for a simple job, a data lineage report might show that the first activity read six rows from a text file. It might then show that the second activity wrote six rows to a database table. For a more complex job, the data lineage report might show the order of activities that are responsible for every read, write, or lookup.

Business Lineage

Business lineage reports show data flows only through those information assets that have been configured to be included in business lineage reports. In addition, business lineage reports do not include extension mapping documents or jobs from IBM InfoSphere DataStage and QualityStage.

You are not required to specify the flow direction of data, the analysis type, or a target asset. The business lineage report displays the graphical and textual components for only those source, target, and intermediate assets that are configured to be included in business lineage.

The Metadata Workbench Administrator configures which information assets are displayed in business lineage reports. A report is generated from
the right-click menu of an asset that is configured for business lineage. The report is read-only and you cannot get further information about the data flow or about the assets themselves.

A user of IBM InfoSphere Business Glossary Anywhere, IBM InfoSphere Business Glossary, IBM InfoSphere Metadata Workbench, and external programs such as IBM Cognos, can create a business lineage report for an asset. The user must have at least the Business Glossary User role. The report is displayed in a new window in the web browser. For example, a business lineage report for a BI report might show the flow of data from one database table to another database table. From the second database table, the data flows into a BI report collection table and then to a BI report. The context of the database tables and the BI report collection table is displayed.

For each type of analysis, you can create a report that shows the flow of information from asset to asset that participates in the lineage or analysis flow.

**Related concepts:**

"Extended data lineage" on page 30
You can track the flow of data in your enterprise, even when you use external processes that do not write to disk, or ETL tools, scripts and other programs that do not save their metadata to the metadata repository.

**Related tasks:**

"Running data lineage reports" on page 70
You can run data lineage reports that combine information from job designs, operational metadata, and user-defined relationships between assets. You can see the flow of data through specific child assets, for example, the flow of data through selected database columns of a database table.

### Asset types that are included in lineage and analysis reports

Certain asset types can be used to run data lineage, business lineage, or impact analysis reports.

### Data lineage, business lineage, and impact analysis reports

You can run data lineage, business lineage, and impact analysis reports on the following types of assets and their children:

**Table 4. Lineage and impact analysis reports on these assets and their successive child assets**

<table>
<thead>
<tr>
<th>Asset</th>
<th>Successive child assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application (extended data source)</td>
<td>• Object type &gt; method &gt; input parameter&lt;br&gt;• Object type &gt; method &gt; output value</td>
</tr>
<tr>
<td>Business intelligence (BI)</td>
<td>• BI model &gt; BI collection &gt; BI collection member&lt;br&gt;• BI report &gt; BI report field</td>
</tr>
<tr>
<td>Database</td>
<td>• Schema &gt; database table &gt; database column&lt;br&gt;• Schema &gt; view &gt; database column</td>
</tr>
<tr>
<td>Data file</td>
<td>• Data file structure &gt; data file field</td>
</tr>
<tr>
<td>File (extended data source)</td>
<td>None</td>
</tr>
</tbody>
</table>
Table 4. Lineage and impact analysis reports on these assets and their successive child assets (continued)

<table>
<thead>
<tr>
<th>Asset</th>
<th>Successive child assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stored procedure definition (extended data source)</td>
<td>• In parameter</td>
</tr>
<tr>
<td></td>
<td>• Out parameter</td>
</tr>
<tr>
<td></td>
<td>• Inout parameter</td>
</tr>
<tr>
<td></td>
<td>• Result column</td>
</tr>
</tbody>
</table>

Data lineage and impact analysis reports

You can run data lineage and impact analysis reports on the following assets:

Table 5. Data lineage and impact analysis reports on these assets and their successive child assets

<table>
<thead>
<tr>
<th>Asset</th>
<th>Successive child assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job</td>
<td>• Stage &gt; stage column</td>
</tr>
<tr>
<td>Mapping in</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>• Extension mapping document from IBM InfoSphere Metadata Workbench</td>
</tr>
<tr>
<td></td>
<td>• Mapping specification from InfoSphere FastTrack</td>
</tr>
<tr>
<td></td>
<td>• Mapping document from IBM InfoSphere Warehouse</td>
</tr>
</tbody>
</table>

Related concepts:
“Extended data lineage” on page 30

You can track the flow of data in your enterprise, even when you use external processes that do not write to disk, or ETL tools, scripts and other programs that do not save their metadata to the metadata repository.

Related tasks:
“Running data lineage reports”

You can run data lineage reports that combine information from job designs, operational metadata, and user-defined relationships between assets. You can see the flow of data through specific child assets, for example, the flow of data through selected database columns of a database table.

Running data lineage reports

You can run data lineage reports that combine information from job designs, operational metadata, and user-defined relationships between assets. You can see the flow of data through specific child assets, for example, the flow of data through selected database columns of a database table.

Before you begin

The Metadata Workbench Administrator must have run the Manage Lineage and the Manage Data Source Identity utilities in the Lineage Administration section of the Administration tab. The Manage Lineage utility sets the relationships between stages, database tables, and database views so that the lineage report is accurate. The Manage Data Source Identity utility creates the relationships between the
schemas, and between the database tables and database fields of both schemas that have matching names.

**About this task**

Data lineage reports show the flow of information both to and from a selected asset. By default, the reports include design, operational, and user-defined relationships between assets. Duplicate assets are not displayed.

Some assets can contain child assets. The following examples of assets and their child assets can be displayed in the lineage report in a separate web browser window:

**Job asset whose child assets are stages.**
You can view the data as it flows through the stages that are contained in the job.

**Database table asset whose child assets are database columns.**
You can view the data as it flows through selected database columns of a database table. The relationship links between the selected database columns and other assets are redrawn. In addition, asset nodes that are not in the data flow are disabled. You cannot select child assets in the new browser window. For example, if you originally selected the database columns DM_SORTKEY, DM_TIMESTAMP, and DM_TXID from a database table, you cannot select a subset of these database columns from the database table in the new lineage report.

**Procedure**

1. Start the report by using either of the following methods:
   - In a results list, right-click the name of an asset, and choose **Data Lineage**.
   - In the task list on the asset information page for an asset, click **Data Lineage**.

2. Optional: Click ![arrow](image) and select which types of analysis relationships to display in the report.
   If you change the selection of analysis relationships, click **Refresh display** to redo the lineage graphic.

3. Optional: In the asset node, click the **Select [Column | Field | Member | Parameter]** link, if available, to trace data lineage through specific child assets. You can select no more than five child assets. You cannot select columns in a stage asset.

4. Optional: In an IBM InfoSphere DataStage job node, click the **Expand** link to see the flow of data between the stages that are contained in that job.

5. Optional: Click ![save](image) to save the text of the right pane, or to save the entire graphic, the part of the graphic that is visible in the window, or the part of the graphic that you selected, to a file.
   The text is saved to a file in a PDF format. The graphic is saved to a file in a JPEG image format.
   If you click **Save Graph**, in the Save Graph window, select **All** to save the entire graphic, **Current View** to save the part that is visible in the window, or **Selection** to save the part that you selected.
Results

The report displays available information about the flow of data through assets.

What to do next

In the image file of the lineage graphic, click the name of an asset to display its details page in a new browser window.

In the PDF file of the text, right-click the name of an asset to get list of actions that you can do on the asset.

Related concepts:

“Data lineage, business lineage, and impact analysis reports” on page 67
Data lineage and impact analysis reports show the movement of data within a job or through multiple jobs. These reports also show the order of activities within a run of a job. Business lineage reports show a scaled-down view of lineage without the detailed information that is not needed by a business user. Impact analysis reports show the dependencies between assets.

Related tasks:

“Running queries” on page 65
Users of the metadata workbench can run queries to find and report on information assets.

“Finding assets in the metadata workbench” on page 58
You locate particular assets to investigate their properties and relationships and to run reports on them. You can browse, search, and query to locate assets.

Related reference:

“Asset types that are included in lineage and analysis reports” on page 69
Certain asset types can be used to run data lineage, business lineage, or impact analysis reports.

Running business lineage reports

You can create business lineage reports that display the flow of information between assets that have been configured to be in business lineage reports.

Before you begin

• Only certain asset types, such as application, business intelligence (BI) report, and data file, can be included in business lineage reports. By default, all assets in these asset types are included in business lineage reports. The Metadata Workbench Administrator must configure any assets to be excluded from the business lineage report.

• You must have the Business Glossary User role to run a business lineage report.

About this task

A business lineage report is a read-only report that displays the flow of information between assets. Only those assets that have been configured to be included in business lineage reports are displayed in the lineage path.

Procedure

Run a business lineage report on a selected asset by doing either of these actions:
• In the task list on the asset information page for an asset, click **Business Lineage**.
• In a results list from a find or query action, or in the Manage Business Lineage window, right-click the name of an asset and choose **Business Lineage**.

**Results**

The report displays, in a new window, all of the assets that participate in the lineage path with the asset that you report on. The context, or parent, of each asset in the report is displayed. You cannot drill down into any asset in the report to obtain more information.

**What to do next**

You can zoom in to display selected areas of the report in more detail, save and print the report, and display the context, description, and steward of the asset. You can also send open your default email application to send feedback about the lineage report.
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. 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 6. 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 C. 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/dochrome/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.

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

- 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
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/
- 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).
- You can provide general product feedback through the Consumability Survey at www.ibm.com/software/data/info/consumability-survey
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
- asset term 44
- asset types
  - child assets 69
  - sources for impact analysis reports 69
  - sources for lineage reports 69
- assets
  - BI reports 42
  - business intelligence reports 42
  - category 43
  - data file 45
  - data file structure 47
  - database 48
  - database schema 51
  - database table 49
  - deleting 32
  - displaying details 32
  - editing details 32
  - engine 56
  - host 48
  - job 53
  - stage 54
  - steward 57
  - transformation project 55
  - view 52
- assets, performing tasks on
  - assets, deleting 32
  - assets, details 32
  - assets, editing details 32
  - blueprint, viewing 32
  - business lineage report, creating 32
  - business lineage report, excluding asset in 32
  - business lineage report, including asset in 32
  - business name, editing 32
  - custom attributes, assigning 32
  - data file report, creating 32
  - data file, graph view 32
  - data lineage report, creating 32
  - data lineage report, excluding asset in 32
  - data lineage report, including asset in 32
  - extended data lineage
    - extended data sources 30
    - function 30
    - mappings 30
  - extended data sources
    - definition 30
    - function 30
- finding assets
  - by creating query 58
  - by existing query 58
  - by short description 58
  - data servers and databases 58
  - particular type 58
  - projects and jobs 58
- graph view 32

### B
- BI reports 42
- blueprints
  - downloading 32
  - viewing 32
- business intelligence reports 42
- business lineage report
  - creating 32
  - excluding asset from 32
  - including asset in 32

### C
- category 43
- custom attributes
  - assigning 32
  - customer support contacting 77

### D
- data file 45
- data file structure 47
- data lineage report
  - creating 32
  - data lineage reports 67
  - database 48
  - database table 49

### E
- engine 56
- extended data lineage
  - extended data sources 30
  - function 30
- extended data sources
  - definition 30
  - function 30

### F
- finding assets
  - by creating query 58
  - by existing query 58
  - by short description 58
- data servers and databases 58
- particular type 58
- projects and jobs 58

### G
- graph view 32

### H
- Has A 44
- Has Type 44
- host 48

### I
- impact analysis 32
- impact analysis reports 67

### M
- mappings
  - definition 30
  - function 30
- metadata workbench
  - accessing 3
  - navigation tips 29
  - opening 3
  - overview 1
- metadata workbench reports
  - asset types as sources 69
  - business lineage 67, 72
  - data lineage 67, 70
  - impact analysis 67
  - model view 32

### N
- navigation tips for metadata workbench 29
- notes
  - adding 32
  - deleting 32
  - editing 32

### P
- product accessibility
  - accessibility 75
- product documentation
  - accessing 79

### Q
- queries 59
- creating 62
- editing 64

---

© Copyright IBM Corp. 2008, 2012
queries (continued)
  limiting results  62
  publishing  62
  running  65

R
  reports
    BI reports  42
    business intelligence reports  42
  roles
    InfoSphere Metadata Workbench
      Administrator  5
    Metadata Workbench User  5
  rule
    assigning  32
    governed by, definition  32
    implements, definition  32

S
  schema  51
  software services
    contacting  77
  stage  54
  steward  57
    assigning  32
  support
    customer  77

T
  term  44
  trademarks
    list of  81
  transformation project  55

V
  view  52