

IBM InfoSphere Data Click  
Version 11 Release 3

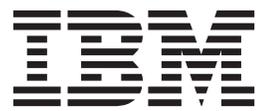
*User's Guide*





IBM InfoSphere Data Click  
Version 11 Release 3

*User's Guide*



**Note**

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 63.

---

# Contents

<b>Chapter 1. InfoSphere Data Click information roadmap</b>	<b>1</b>
---	----------

<b>Chapter 2. Overview of InfoSphere Data Click</b>	<b>3</b>
---	----------

<b>Chapter 3. Scenarios for integrating InfoSphere Data Click and InfoSphere BigInsights</b>	<b>9</b>
--	----------

<b>Chapter 4. Running InfoSphere Data Click activities</b>	<b>11</b>
--	-----------

Making assets available for use in activities	11
Creating InfoSphere Data Click authors and users	15
Creating activities	19
Running activities	22
Monitoring InfoSphere Data Click activities	24
Reference	25
The syncbi command (Linux)	26
Configuring access to Hive tables (AIX only)	27
Configuring access to HDFS (AIX only)	29
Checking activities in the Operations Console	32
Hive tables	34
Creating .osh schema files	37
Assigning user roles for InfoSphere Data Click installations that use LDAP configured user registries	39

Data connections in InfoSphere Data Click	42
Creating InfoSphere DataStage projects for InfoSphere Data Click activities	46
Configuring the workload management queue	49

<b>Chapter 5. Troubleshooting InfoSphere Data Click</b>	<b>51</b>
---	-----------

Troubleshooting InfoSphere Data Click activities that fail to run or finish with warnings	51
Troubleshooting general issues in InfoSphere Data Click	52

<b>Appendix A. Product accessibility</b>	<b>57</b>
--	-----------

<b>Appendix B. Contacting IBM</b>	<b>59</b>
-----------------------------------	-----------

<b>Appendix C. Accessing the product documentation</b>	<b>61</b>
--	-----------

<b>Notices and trademarks</b>	<b>63</b>
-------------------------------	-----------

<b>Index</b>	<b>69</b>
--------------	-----------



---

## Chapter 1. InfoSphere Data Click information roadmap

This document provides links to the information resources that are available for IBM® InfoSphere® Data Click.

For links to information resources that are available for all IBM InfoSphere Information Server products and components, see the [http://www.ibm.com/support/knowledgecenter/SSZJPZ\\_11.3.0/com.ibm.svg.im.iis.productization.iisinfsv.roadmap.doc/topics/iisinfsv\\_roadmap.html](http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/com.ibm.svg.im.iis.productization.iisinfsv.roadmap.doc/topics/iisinfsv_roadmap.html).

In addition to the InfoSphere Data Click information in IBM Knowledge Center, review these information resources:

### **IBM Information Server Integration and Governance for Emerging Data Warehouse Demands**

Published in July 2013, this IBM Redbooks® publication is intended for business leaders and IT architects who are responsible for building and extending their data warehouse and Business Intelligence infrastructure. It provides an overview of powerful new capabilities of InfoSphere Information Server in the areas of big data, statistical models, data governance, and data quality. The book also provides key technical details that IT professionals can use in solution planning, design, and implementation.



---

## Chapter 2. Overview of InfoSphere Data Click

IBM InfoSphere Data Click provides self-service data integration so that any business or technical user can integrate data between various systems both on and off-premise.

InfoSphere Data Click simplifies data integration for users across organizations. Analysts, data scientists, and line-of-business users can retrieve data and populate new systems on demand. For example, an analyst can optimize a business intelligence environment by integrating warehouse data, or line-of-business users can deliver their data for analysis.

Whether the data source includes one table in a database or thousands of objects in a Hadoop cluster, you can get the data that you need with just a few clicks by using InfoSphere Data Click.

### Define data integration from source to target

You can create activities to integrate data by using the InfoSphere Data Click browser-based interface. You can create multiple activities that each specify different source-to-target configurations.

When you create an activity, you define the source for the data. You can choose data that you require from a wide variety of data sources, including IBM PureData™ (DB2®) and Oracle. You can limit the source to the data that you require, whether it is a single table, multiple databases, or multiple objects that are stored in an Amazon S3 bucket.

When you run an activity that moves data into a Hadoop Distributed File System (HDFS) in IBM InfoSphere BigInsights™, InfoSphere Data Click automatically creates Hive tables and stores them in the target directory. A Hive table is created for each table that you select, and you can specify the location of the target directory where the data is stored. The data types of the columns in the Hive table are assigned based on metadata information about the data types of the columns in the source. You can then use IBM Big SQL in InfoSphere BigInsights to read and analyze the data in the tables.

You also set the policies for the activity, including the amount of data that you can integrate when the activity runs. The policy choices that you make are applied automatically to any runs of that activity.

The following figure shows an example of the New Activity page on which you can specify the source for the data. In this example, the user selected one table in a database as the source. The user expanded the list of assets to view the columns in the table. To view this page, the user first entered some basic information on the Overview page, including a name for the activity. Next, the user will select the target location.

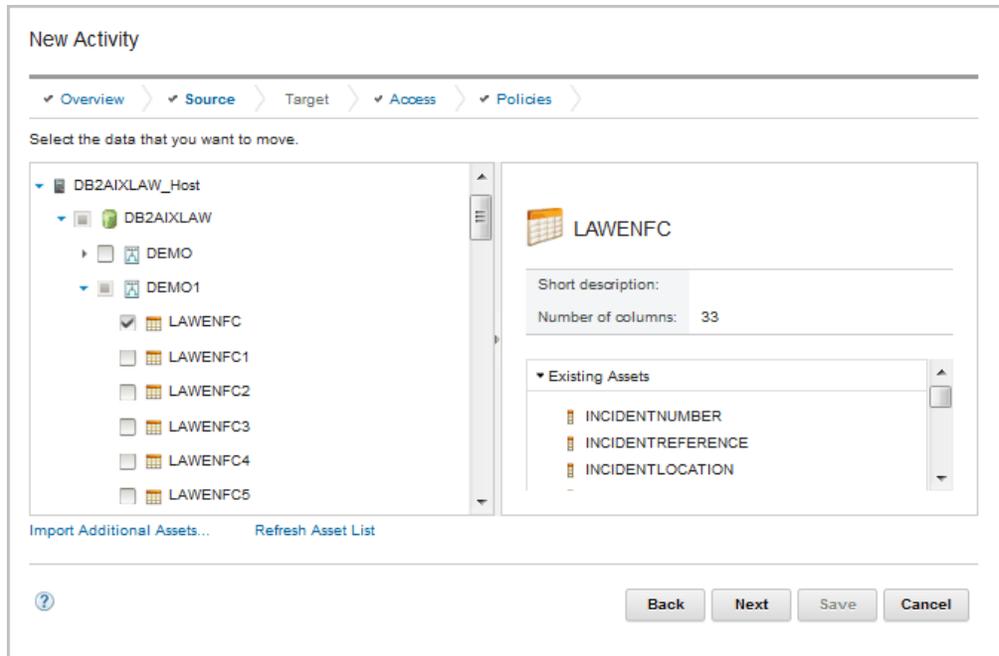


Figure 1. New Activity page in the InfoSphere Data Click interface

## Integrate data on demand

You can view the activities that you create on the InfoSphere Data Click home page with all the other activities that you can run. You can select a previously created activity and run it immediately, or you can review the activity and further customize it. For example, in an activity that integrates data from multiple tables, you can select a subset of the tables or a single table. You can also edit the default names of the target schema and tables.

The following figure shows an example of the page that is displayed when you select an activity and click **Run**. In this example, the user expanded the table in the available source database to review the table columns. The user can click **Finish** to run the activity without changes, or the user can remove some columns or click **Next** to review the target and the options.

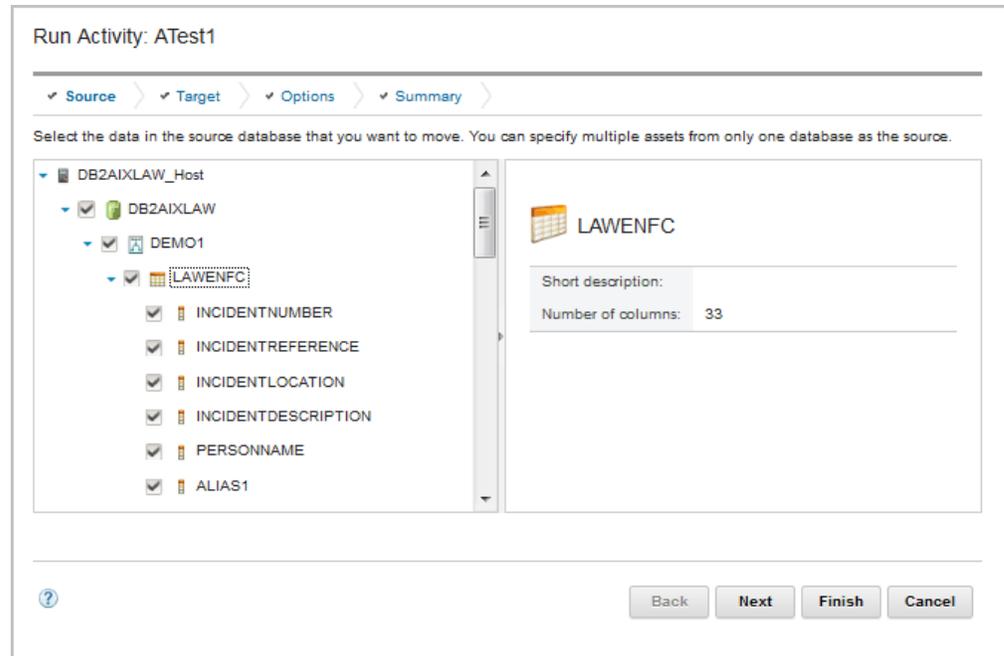


Figure 2. Run Activity page in the InfoSphere Data Click interface

## Monitor data integration activities

You can monitor data integration activities by using InfoSphere Data Click. You can review who ran an activity, the status of the run, the number of rows that were moved by an activity, the date that the activity was submitted, and the date that the activity completed. By reviewing this information, you can determine whether an activity ran successfully and whether the expected number of rows were processed.

The following figure, shows an example of the InfoSphere Data Click home page, including the Activities section and the Monitoring section. The headings for these two sections are highlighted.

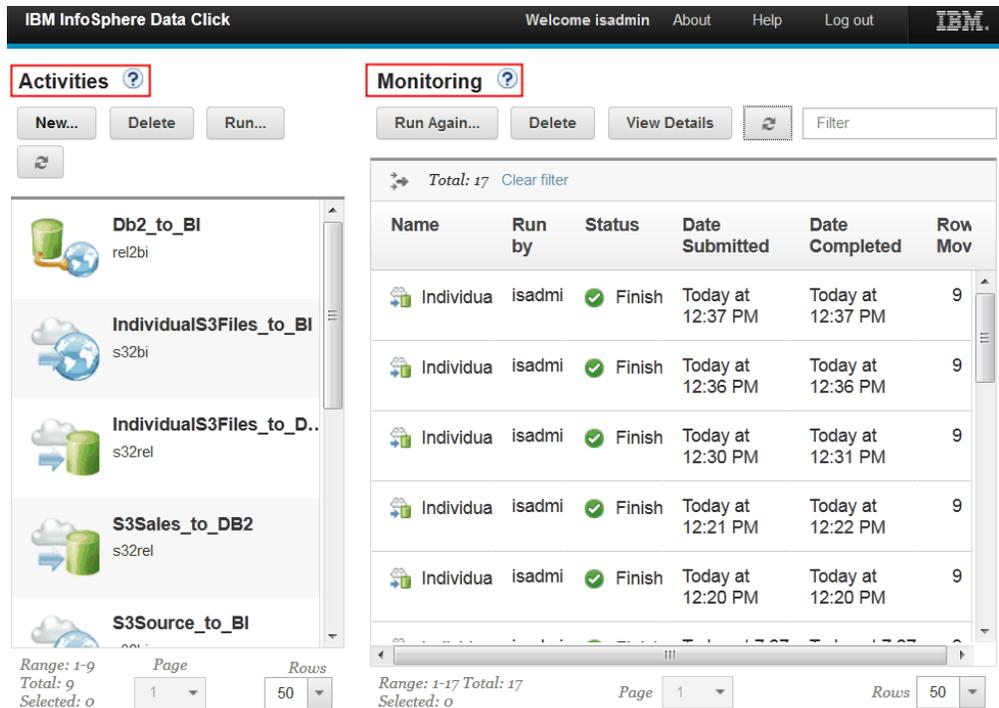


Figure 3. Activities and Monitoring areas on the home page of InfoSphere Data Click

You can also check the status of activities on the IBM InfoSphere DataStage® and QualityStage® Operations Console by using the **View Details** option from the InfoSphere Data Click home page. The Operations Console provides a more detailed view of the status of jobs that are generated when you run activities.

## Search, browse, and move assets from the Information Governance Catalog

You can also use Information Governance Catalog to identify the data that you want to make available in a target system. Information Governance Catalog is an interactive tool that you can use to search, browse, and query catalog assets from all your data sources. You can assign labels, stewards, and custom attributes to these assets. Assets can be grouped in a collection. You can then integrate the data directly into a target database, Amazon S3 bucket, or InfoSphere BigInsights by opening InfoSphere Data Click from Information Governance Catalog within the context of the assets that you identified. You can then use InfoSphere Data Click to move the assets.

## Control security

You can define and control which users integrate data by authorizing some users to create and run activities and authorizing other users only to run prebuilt activities. For example, you might authorize technical users to create activities for business users. So an information architect can define the source and targets for warehouse data in a business intelligence environment and authorize an analyst to integrate the data. An enterprise architect can identify the source and target of the data from line-of-business users and authorize those users to deliver their data.

The InfoSphere Data Click users who create activities define the data source and target, and set the policies. By setting policies, the users who create activities can

define the data flows and manage the stress on the system.

### **Powered by your ETL engine**

InfoSphere Data Click uses the highly scalable data integration engine of IBM InfoSphere DataStage to process the data integration activities and to manage the metadata that is associated with the activities.



---

## Chapter 3. Scenarios for integrating InfoSphere Data Click and InfoSphere BigInsights

You can use InfoSphere Data Click to integrate large volumes of data in a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights. InfoSphere BigInsights can then discover and analyze business insights that are hidden in the large volumes of data.

Because InfoSphere Data Click simplifies data movement, you do not have to be an experienced ETL developer to integrate data. Whether you are an analyst, data scientist, or line-of-business user, you can use InfoSphere Data Click to integrate data sets of any size, even into the terabyte or petabyte range.

After the data is copied into the HDFS, you can use InfoSphere BigInsights to derive business value from complex, unstructured information. For example, you can use IBM Big SQL, the IBM SQL interface to InfoSphere BigInsights, to summarize, query, and analyze data. InfoSphere BigInsights supports various scenarios that can help organizations grow by finding value that is hidden in data and data relationships.

### Scenario: Augmenting the data warehouse

An international retailer has a traditional data warehouse and wants to augment their business intelligence environment with new unstructured data sources that are only in an HDFS. To run new analytical models, the unstructured information must be combined with data from the traditional data warehouse.

To augment the data warehouse infrastructure, the retailer uses InfoSphere Data Click to copy customer and product information from the warehouse into an HDFS. The retailer then uses the BigSheets tool, which is part of InfoSphere BigInsights, to combine the warehouse data with structured, unstructured, and streaming data from other sources. Now, the retailer can filter and analyze the combined set of data.

By using the self-service integration features of InfoSphere Data Click with InfoSphere BigInsights, the retailer can move and analyze data in hours rather than days or weeks.

### Scenario: Using an HDFS as a landing zone

A financial services department receives data from external and internal sources; these sources send the data in different ways and in various formats. To make their processes more efficient, the department wants to use an HDFS as a common storage area, or *landing zone*, for the data that comes from their internal partners. Unfortunately, their internal partners lack the data integration skills to send data to the HDFS, which creates a roadblock to achieving the new corporate objective.

The company decides to use InfoSphere Data Click to help the internal partners get their data into the landing zone. Line-of-business users from the marketing, finance, and human resources departments can run InfoSphere Data Click activities to copy data from their data sources to the HDFS. In the HDFS, InfoSphere BigInsights processes can complete detailed analysis on complex sets of data in hours.

The line-of-business users can run InfoSphere Data Click activities on demand, providing current and comprehensive insights into the challenges that are faced by each part of the organization. The combination of InfoSphere Data Click and InfoSphere BigInsights gives the line-of-business users a simple infrastructure and process for both moving data and running analytics on that data.

### **Scenario: Identify data types to analyze in InfoSphere BigInsights**

Data scientists for a large insurance company need to run exploratory statistical models across huge amounts of data. The data scientists want to use InfoSphere BigInsights to analyze large volumes of varied data, but the data is scattered between many different databases and data sources. In addition, the data scientists do not know what kinds of data they have.

The data scientists use the Information Governance Catalog to search the data sources for specific types of data. For example, suppose that some of the data sources are indexed based on meaningful columns such as department, geographical region, and product name. Because the data scientists want to run statistical models on data from the South America geographical region, they search for South America. The Information Governance Catalog shows a list of database tables and other data sources, with an explanation of why the sources are related to the search term.

By browsing the search results, the data scientists identify the data sources that have the data that they want to analyze and the kinds of data in those data sources. The data scientists select the data to analyze and open InfoSphere Data Click directly from the Information Governance Catalog. When InfoSphere Data Click opens, the sources that were selected in the catalog are already selected in InfoSphere Data Click. The data scientists can use InfoSphere Data Click to copy the data to an HDFS and use InfoSphere BigInsights to analyze the data.

---

## Chapter 4. Running InfoSphere Data Click activities

To get started using InfoSphere Data Click, open the InfoSphere Information Server Launchpad.

You can access the Launchpad by constructing the following URL:  
`https://<server>:<port>/ibm/iis/launchpad.`

The *server* variable is the name of the computer where you have installed your application server. The default port number is 9443. For example,  
`https://my_computer:9443/ibm/iis/launchpad.`

From the Launchpad, you can access InfoSphere Data Click, InfoSphere Metadata Asset Manager, the IBM InfoSphere Information Server Administration Console, the IBM InfoSphere DataStage and QualityStage Operations Console, and the Information Governance Catalog.

You can construct the following URLs to access individual products without using the Launchpad:

- InfoSphere Data Click: `https://<server>:<port>/ibm/iis/dc/offload/ui/login.jsp`
- InfoSphere Metadata Asset Manager: `https://<server>:<port>/ibm/imam/console`
- IBM InfoSphere Information Server Administration Console:  
`https://<server>:<port>/ibm/iis/console`
- IBM InfoSphere DataStage and QualityStage Operations Console:  
`https://<server>:<port>/ibm/iis/ds/console`
- Information Governance Catalog: `https://<server>:<port>/ibm/iis/igc`

You can bookmark the URLs after you connect to them.

---

### Making assets available for use in activities

You must import metadata for all sources and targets that you want to make available for InfoSphere Data Click activities. One reason that you import the metadata is to ensure that connection information is available during the processing of activities.

#### Before you begin

Before you can successfully import metadata into the metadata repository by using InfoSphere Metadata Asset Manager, you must set up connections to all sources and targets that you want to use in InfoSphere Data Click.

Click on the link for each connector in the list below for information about configuring them. You must configure them before using InfoSphere Metadata Asset Manager or InfoSphere Data Click. The following connectors are supported in InfoSphere Data Click:

- ODBC connector
- JDBC connector
- DB2 connector

- Oracle connector
- Amazon S3 connector

**Note:** The Amazon S3 connector comes pre-configured for use with the InfoSphere Information Server suite. You do not need to configure the Amazon S3 connector.

Before you can move data from Amazon S3 by using InfoSphere Data Click, you must create an .osh schema file or specify the format of the file in the first row of each file that you want to move before you import the Amazon S3 metadata into the metadata repository by using InfoSphere Metadata Asset Manager. It is recommended that you create an .osh schema file because it provides more comprehensive details about the structure of the data in your files. If you used the Amazon S3 file as a target in a previous InfoSphere Data Click activity, then an .osh schema file was automatically created for the file by InfoSphere Data Click, so you do not need to create one.

To move data from Amazon S3, you must also verify that the character set of the Amazon S3 files match the encoding properties that are set for the InfoSphere DataStage project that you are using in your activity. For example, if the character set of the files in Amazon S3 is set to UTF-8, then the encoding for InfoSphere DataStage should also be set to UTF-8. You must update the APT\_IMPEMP\_CHARSET property in the InfoSphere DataStage project used for running the activity. For more information about updating the file encoding properties, see *Creating a project*.

If you are planning on using InfoSphere Data Click to move data into a relational database that connects to InfoSphere Data Click by using the JDBC connector, then you must verify that the relational database has the same NLS encoding as the local operating system.

If you are planning on using InfoSphere Data Click to move data into a Hadoop Distributed File System in InfoSphere BigInsights on a Linux computer, refer to the `snycbi` command for information on configuring your computer. See “Configuring access to Hive tables (AIX only)” on page 27 and “Configuring access to HDFS (AIX only)” on page 29 for information about setting up connections to InfoSphere BigInsights on Windows and AIX® computers.

To import metadata into the metadata repository through the InfoSphere Metadata Asset Manager, you must have the Common Metadata Importer role.

## About this task

The following task describes how to import the metadata by using InfoSphere Metadata Asset Manager.

You must import metadata for all sources and targets that you want to use in InfoSphere Data Click activities, as well as every asset that you want to use in the Information Governance Catalog. For every distributed file system that you import by using the HDFS bridge, you must import at least one associated Hive table. The distributed file system and the Hive table become associated when you create a separate data connection to each one, and you specify the same host name in the parameters. For example, in the data connection that connects to the distributed file system by using the HDFS connector, you might specify 9.182.77.14 in the Host field. When you import the associated Hive table, you must specify 9.182.77.14 as part of the URL field in the data connection parameters. For example, `jdbc:bigsql://9.182.77.14:7052/default`.

## Procedure

1. Log in to InfoSphere Metadata Asset Manager by clicking the icon in the InfoSphere Information Server Launchpad or by selecting **Import Additional Assets** in the InfoSphere Data Click activity creation wizard.
2. On the **Import** tab, click **New Import Area**.
  - a. Specify a unique name and a description for the import area.
  - b. Select the metadata interchange server that you want to run the import from.
  - c. Select the bridge or connector to import with. You choose the bridge or connector depending on the source metadata that it imports. Help for the selected bridge or connector is displayed in the **Import Help** pane.
  - d. Click **Next**.
3. Select or create a data connection. You can edit the properties of a selected data connection. If you create a new data connection, select **Save password** if you want InfoSphere Data Click users to be able to access the data without entering the user name and password for the database when they are creating activities. If you do not save the password, they will be prompted for credentials when they use data in the database in activities. If you are using an existing data connection, you might want to click **Edit** and update the **Save password** parameter.
4. Specify import parameters for the selected bridge or connector. Help for each parameter is displayed when you hover over the value field.
  - a. Optional: After you enter connection information for an import from a server, click **Test Connection**.
  - b. For imports from Amazon S3 buckets, you must select **Import file structure** if you want to use InfoSphere Data Click to move metadata from Amazon S3 files.

**Note:** You do not need to select the **Import file structure** option if you only want to move data to Amazon S3. If you want to use Amazon S3 as a target in InfoSphere Data Click, you do not need to select this option.

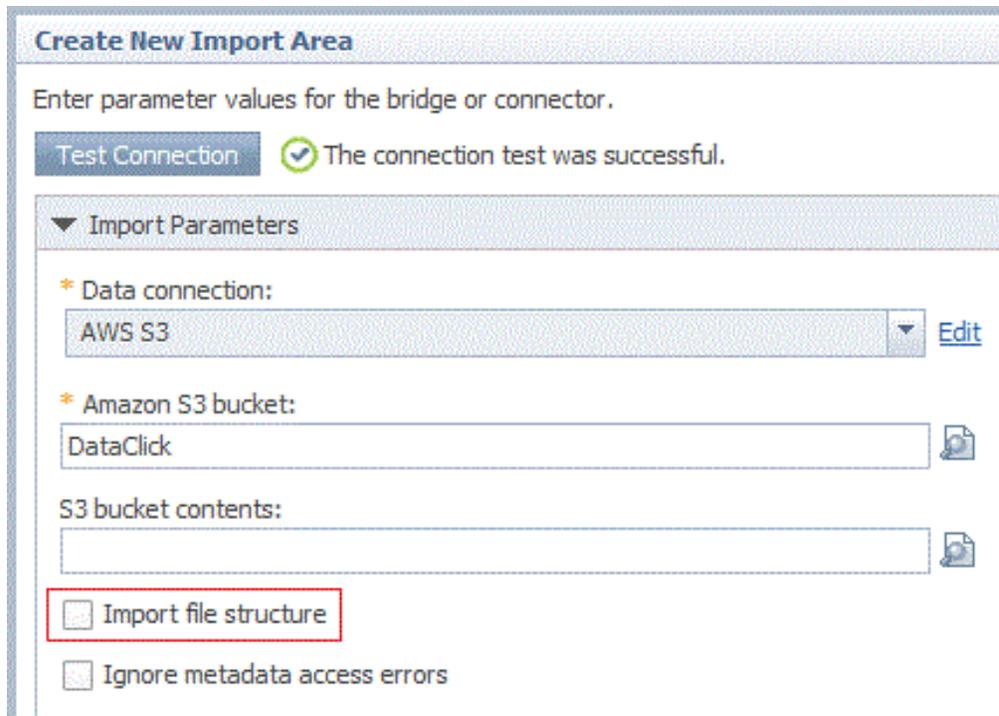


Figure 4. A view of the Import file structure option on the Import Parameters panel in InfoSphere Metadata Asset Manager

This imports the .osh schema files that define the structure of the data in the Amazon S3 files or it imports the metadata about the first row of each file, which contains information about the structure of each file.

- c. For imports from databases, repositories, and Amazon S3 buckets, browse to select the specific assets that you want to import.
- d. Click **Next**.
5. On the Identity Parameters screen, enter the required information. Click **Next**.
6. Type a description for the import event and select **Express Import**.
7. Click **Import**. The import area is created. The import runs and status messages are displayed.

Leave the import window open to avoid the possibility that long imports time out.

8. Repeat the preceding steps until you have imported all metadata assets that you want to make available to InfoSphere Data Click users.

## Results

After running an express import, take one of the actions that are listed in the following table.

Table 1. Choices after an express import

In this case	Take this action
If the analysis shows problems that you must fix	The <b>Staged Imports</b> tab is displayed. Review the analysis results. If necessary, reimport the staged event.

Table 1. Choices after an express import (continued)

In this case	Take this action
If your administration settings require a preview	The <b>View Share Preview</b> screen is displayed. Preview the result of sharing the import.
If your administration settings do not require a preview	The assets are shared to the metadata repository. The <b>Shared Imports</b> tab is displayed. You can browse the assets on the <b>Repository Management</b> tab and work with them in other suite tools.

## Creating InfoSphere Data Click authors and users

You must create InfoSphere Data Click authors and users before using InfoSphere Data Click, and assign them InfoSphere Information Server and InfoSphere DataStage roles by using the IBM InfoSphere Information Server Administration console and the InfoSphere Data Click administrator command-line interface.

### Before you begin

You must have InfoSphere Information Server administrator level access to create InfoSphere Data Click authors and users. You must have the DataStage and QualityStage Administrator or the Suite Administrator role to assign InfoSphere DataStage user roles and add users to projects by using the InfoSphere Data Click administrator command-line interface.

### About this task

InfoSphere Data Click authors and users need to be created and assigned roles before you can create and run activities in InfoSphere Data Click.

#### Data Click Authors

An author can:

- Prepare the InfoSphere Data Click environment
- Create and share activities with users
- Run activities

#### Data Click Users

A user can:

- Run activities
- Monitor activities

To begin the process of creating InfoSphere Data Click authors and users, you must create user names and passwords for each user in the IBM InfoSphere Information Server Administration console. You then create a group of users and a group of authors. After your groups are created, you grant user roles to each group of users by using the InfoSphere Data Click administrator command-line interface. You also have the option of creating individual users in the InfoSphere Information Server Administration console and granting them user roles by using the InfoSphere Data Click administrator command-line interface.

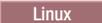
If you have configured InfoSphere Data Click to use a lightweight directory access protocol (LDAP) user registry, then all users are already created. You just need to assign the users the appropriate user roles by using the IBM InfoSphere

Information Server Administration console to assign InfoSphere Information Server roles and the DirectoryCommand tool to assign InfoSphere DataStage user roles and add users to projects.

If InfoSphere Data Click is not configured to use an LDAP user registry, use the InfoSphere Data Click administrator command line to assign InfoSphere Information Server and InfoSphere DataStage roles to InfoSphere Data Click users and assign them to InfoSphere DataStage projects before they can move data by using InfoSphere Data Click.

## Procedure

1. Create InfoSphere Data Click users.
  - a. Open the IBM InfoSphere Information Server Administration console.
  - b. Click the **Administration** tab, and then click **Users and Groups > Users**.
  - c. In the Users pane, click **New User**.
  - d. Specify the user name and password for the new InfoSphere Data Click author or user.
  - e. Click **OK**.
2. Create an InfoSphere Data Click Authors group and add users that you want to have the InfoSphere Data Click Author role to the group:
  - a. Click the **Administration** tab, and then click **Users and Groups > Groups**.
  - b. Click **New Group**.
  - c. Enter an ID in the **Principle ID** field and a name for the group in the **Name** field. For example, enter DataClick\_Authors as the principle ID and Data Click Authors as the name.
  - d. In the **Users** pane on the right, click **Browse** and select all users that you want to add to your group.
  - e. Click **OK**.
3. Create an InfoSphere Data Click Users group and add users that you want to have the InfoSphere Data Click User role to the group:
  - a. Click the **Administration** tab, and then click **Users and Groups > Groups**.
  - b. Click **New Group**.
  - c. Enter an ID in the **Principle ID** field and a name for the group in the **Name** field. For example, enter DataClick\_Users as the principle ID and Data Click Users as the name.
  - d. In the **Users** pane on the right, click **Browse** and select all users that you want to add to your group.
  - e. Click **OK**, and then **Save and Close**. Exit the IBM InfoSphere Information Server Administration console.
4. Use the InfoSphere Data Click administrator command-line interface to grant InfoSphere Information Server and InfoSphere DataStage roles.
  - a. Navigate to the InfoSphere Data Click administrator command line in the `IS_install_path/ASBNode/bin` directory.
  - b. Issue the following command to assign all users in the InfoSphere Data Click authors group that you created in 2 all the roles that they will need to use InfoSphere Data Click and add them to the default InfoSphere Data Click project:

-  

```
DataClickAdmin.sh -url URL_value -user admin_user -password password
                    -role role_type -grantgroup name_of_group
                    "Information_Server_engine_name/Data_Click_project_name"
```

- **Windows**

```
DataClickAdmin.bat -url URL_value -user admin_user -password password
                    -role role_type -grantgroup name_of_group
                    "Information_Server_engine_name/Data_Click_project_name"
```

Issue the command again, and specify the group of InfoSphere Data Click users that you created in 3 on page 16 as the `-grantgroup` parameter. Detailed information about the parameters is included after 5.

For example, if you are the InfoSphere Data Click Author `isadmin`, and you want to grant all users in the group `Data Click Users` the InfoSphere Data Click User role for the default InfoSphere Data Click project, `DataClick`, and the name of your engine is `"JSMITH-Q4J6VT"`, then you issue the following command:

```
DataClickAdmin.sh -url iisserver:9443 -user isadmin -password
temp4now -role "user" -grantgroup "DataClickUsers" -project
"JSMITH-Q4J6VT/DataClick".
```

5. Optional: Grant individual InfoSphere Data Click users the Data Click User or Data Click Author role by issuing the following command:

- **Linux** **UNIX** `DataClickAdmin.sh -url URL_value -primary -user admin_user -password password -authfile authfile_path -role role_type -grantuser name_of_user -project "Information_Server_engine_name/Data_Click_project_name"`

- **Windows** `DataClickAdmin.bat -url URL_value -primary -user admin_user -password password -authfile authfile_path -role role_type -grantuser name_of_user -project "Information_Server_engine_name/Data_Click_project_name"`

**Parameters:**

**-url**

Required

The base url for the web server where the InfoSphere Information Server Administration console is running. The url has the following format: `https://<host>:<https port>`. This parameter is not required if you specify the **-primary** parameter.

**-primary**

OptionalLog in to the primary services host if one is available. If this option is used, the **-url** parameter is ignored.

**-user**

Required

The name of the Suite Administrator. This parameter is optional if you use the **-authfile** parameter instead of specifying the **-user** and **-password** parameters. You can enter a user name without entering a password, which causes you to get prompted for a password.

**-password**

Required

The password of the Suite Administrator. This parameter is optional if you use the **-authfile** parameter instead of specifying the **-user** and **-password** parameters.

**-authfile (-af)**

Optional

The path to a file that contains the encrypted or non-encrypted credentials for logging on to InfoSphere Information Server. If you use the **-authfile** parameter, you do not need to specify the **-user** or **-password** parameter on the command line. If you specify both the **-authfile** option and the explicit user name and password options, the explicit options take precedence over what is specified in the file. For more information see the topic *Encrypt command* in the *IBM InfoSphere Information Server Administration Guide*.

**-role**

Required

The type of role that you want to assign to an individual or group. You can specify "author" or "user."

Table 2. InfoSphere Information Server roles

InfoSphere Data Click role	Roles that are assigned to each author or user
For Data Click Authors	<ul style="list-style-type: none"> <li>• Data Click Author</li> <li>• Suite User</li> <li>• DataStage and QualityStage User</li> <li>• Common Metadata Importer</li> <li>• DataStage Developer role for the default DataClick project or any InfoSphere DataStage project that you specify for the <b>Data Click project name</b>.</li> <li>• Information Governance Catalog Administrator</li> </ul>
For Data Click Users	<ul style="list-style-type: none"> <li>• Data Click User</li> <li>• Suite User</li> <li>• DataStage and QualityStage User</li> <li>• Common Metadata User</li> <li>• DataStage Operator for the default DataClick project or any InfoSphere DataStage project that you specify for the <b>Data Click project name</b>.</li> </ul>

**-grantgroup**

Optional

The name of the group that you want to assign author or user roles to. You create groups in the InfoSphere Information Server Administration console. All individuals in the group will be assigned the role that you specify.

**-grantuser**

Optional

The name of the user that you want to assign the author or user role to. You create users in the InfoSphere Information Server Administration console.

### **-project**

Required The name of the InfoSphere Data Click project. You must specify the project name in the following format: "*Information\_Server\_engine\_name/Data\_Click\_project\_name*." The name of the default project that is automatically installed with InfoSphere Data Click is DataClick. If you do not know the name of the InfoSphere Information Server engine, issue the following command by using the DirectoryCommand tool in the IS\_install\_path/ASBNode/bin directory to determine the name of the engine:

```
DirectoryCommand.sh -url URL_value -user admin_user -password password
-list DSPROJECTS
```

The following information is returned when you run that command:

```
----Listing DataStage projects.---
DataStage project: "Information_Server_engine_name/Data_Click_project_name"
-----.
```

For example,

```
----Listing DataStage projects.---
DataStage project: "DATANODE1EXAMPLE.US.IBM.COM/DataClick"
DataStage project: "DATANODE1EXAMPLE.US.IBM.COM/dstage1"
-----.
```

The name of the InfoSphere Information Server engine is, DATANODE1EXAMPLE.US.IBM.COM.

---

## **Creating activities**

You create activities to define repeatable data movement configurations. You use the activity creation wizard to specify the source that you want to move data from, the target that you want to move data to, policies that govern how the data is moved, the InfoSphere Information Server engines that you want to use to move the data, and the InfoSphere DataStage projects that you want to use in the activities. After you create an activity, users that have access to the activity can run the activity by using the activity run wizard.

### **Before you begin**

You need to have defined source and target assets. You define the assets that you want to use in InfoSphere Data Click by importing them into the metadata repository by using InfoSphere Metadata Asset Manager.

If you want to use an InfoSphere DataStage project, other than the DataClick default project in your InfoSphere Data Click activities, then you must import an InfoSphere Data Click job template into the InfoSphere DataStage project. The job template for the default InfoSphere Data Click project is automatically imported during the installation process.

If you want to use a specific InfoSphere Information Server engine in an activity, you must create an InfoSphere DataStage project and specify the particular engine that you want to use. Then, import an InfoSphere Data Click job template into the InfoSphere DataStage project.

You must have the Data Click Author role to create activities and an administrator must grant you access to the InfoSphere DataStage project that you are going to use in the InfoSphere Data Click activity.

## About this task

When you select the sources that you want to move data from and the targets that you want to move data to, you have the following options:

*Table 3. Assets that you can move by using InfoSphere Data Click*

Source	Assets you can move	Supported targets
Relational database	<ul style="list-style-type: none"> <li>• Database schemas</li> <li>• Tables in a schema,</li> <li>• Individual tables and columns</li> </ul>	<ul style="list-style-type: none"> <li>• Relational databases</li> <li>• Hadoop Distributed File Systems (HDFS) in InfoSphere BigInsights</li> <li>• Amazon S3</li> </ul>
Amazon S3	<ul style="list-style-type: none"> <li>• Buckets</li> <li>• Folders a bucket</li> <li>• Individual objects in the folders</li> </ul>	<ul style="list-style-type: none"> <li>• Relational databases</li> <li>• Hadoop Distributed File Systems (HDFS) in InfoSphere BigInsights</li> </ul>

If the metadata that you specify in InfoSphere Data Click activities changes, you need to reimport the metadata by using InfoSphere Metadata Asset Manager. After you reimport the metadata, existing activities automatically use the updated metadata.

Users are assigned to activities at the project level. If you want to add additional users to an activity, use the DirectoryCommand tool to add users to the InfoSphere DataStage project that you specified on the **Access** panel in the activity creation wizard.

When you delete an activity, all activity runs that are associated with the activity are also deleted.

## Procedure

1. Open InfoSphere Data Click.
2. Select **New** and then select **Relational Database to BigInsights**, **Relational Database to Relational Database**, **Relational Database to Amazon S3**, **Amazon S3 to Relational Database**, or **Amazon S3 to InfoSphere BigInsights**.

3. In the Overview pane, enter a name and description for the activity.

4. In the Sources pane, select the assets that you want to move data from.

The assets that you select are the sources that InfoSphere Data Click users are limited to extracting data from when they run the activity. Then, click **Next**.

**Note:** If you do not see the sources that you want to use in this activity, then click **Import Additional Assets**, which will launch InfoSphere Metadata Asset Manager.

5. Select the data connection that you want to use to connect to Amazon S3 or the source database and enter the credentials to access the source. Click **Next**. You are prompted to enter credentials only if the password for the data

connection was not saved when you imported metadata for the database by using InfoSphere Metadata Asset Manager or if the password expired or was changed.

6. In the Target pane, select the database, Amazon S3 bucket, or folder in the distributed file system, such as a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights, to move data to. Then, click **Next**.

The target database or folder that you select is the asset that InfoSphere Data Click users are limited to writing data to when they run the activity.

**Note:** If you do not see the target that you want to use in this activity, then click **Import Additional Assets**, which will launch InfoSphere Metadata Asset Manager.

7. Select the data connections that you want to use to connect to the target database, Amazon S3, or HDFS and enter the credentials to access them. If the target is an HDFS, enter the credentials to access the Hive table. Click **Next**. You are prompted to enter credentials only if the password for the data connection was not saved when you imported metadata for the database, HDFS, or the Hive table by using InfoSphere Metadata Asset Manager, or if the password expired or was changed.
8. In the Access pane, select the InfoSphere DataStage project that you want to use in the activity. If your instance of InfoSphere Data Click is configured to use more than one InfoSphere Information Server engine, select the engine that you want to use. The engine and the project that you select are used to process the InfoSphere DataStage job requests that are generated when you run the activity. The only projects that are displayed are:
  - Projects where you as the Data Click Author has the DataStage Developer role.
  - Projects that you imported as InfoSphere Data Click job templates.

The only engines that are displayed are engines that are specified in projects that you have access to.

Then click **Next**.

9. In the Policies pane, specify the maximum number of records that you want to allow to be moved for each table and the maximum number of tables to extract.
  - If the target you selected is an HDFS, specify the delimiter that you want to use between column values in the Hive table.

When the activity is run, a Hive table is automatically generated for each table that you select in the Source pane. The field delimiter that you specify separates columns in the Hive table. You can specify any single character as a delimiter, except for non-printable characters. In addition, you can specify the following control characters as a delimiter: \t, \b, and \f.
  - If the target that you selected is an Amazon S3 bucket, specify the storage class that defines how you want your data stored when it is moved to Amazon S3.

10. Click **Save** to save the activity and close the wizard.

## What to do next

To run the activity, select the activity in the **Monitoring > Activities workspace** and click **Run**. InfoSphere Data Click users who have access to the activity can run it. To add users to an activity, you must grant them the DataStage Operator role and add them to the InfoSphere DataStage project that you specified in Step 8.

---

## Running activities

InfoSphere Data Click authors and users can run activities that were created by InfoSphere Data Click authors. You run activities to move data from a source asset, such as a database or a bucket that is stored on Amazon S3, to a target database, bucket, or distributed file system, such as a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights.

### Before you begin

You must have the operations database schema for the IBM InfoSphere DataStage and QualityStage Operations Console installed and running before you run activities. You must also start the AppWatcher process before running your first activity. For information about starting the Operations Console, see Getting started with the Operations Console. For information about starting the AppWatcher process, see Starting and stopping the AppWatcher process.

You must have access to an activity that was created by an InfoSphere Data Click Author.

### About this task

You can run activities that appear in the list of activities on the **Home** tab. You can review activity runs in the **Monitoring** section on the **Home** tab. The configuration of activities that are available to you is set by the InfoSphere Data Click author.

The activity run wizard prompts you for information that is needed before you can run the activity. Each pane has a tab. A check mark on a tab indicates that the information in the pane is adequate and you do not need to provide additional information. If no check mark exists, you must provide information in the pane. If all tabs are marked with check marks, you can proceed to the **Summary** pane and run the activity with as few as two clicks.

### Procedure

1. Open InfoSphere Data Click and select an InfoSphere Data Click activity in the pane on the left side of the **Home** tab.
2. Click **Run**.

The InfoSphere Data Click wizard is displayed showing the source databases and schemas that you can move.

3. In the Source pane, select the data that you want to move. Then click **Next**.
4. In the Target pane, select the target that you want to move data to. Then click **Next**.
5. In the Options pane, review the name of the table that is created in the target database, the files that are created in the target Amazon S3 bucket, or the Hive table name that is automatically created when you run the activity. Expand the **Advanced** section to change the name that is automatically assigned.

Option	Description
<p><b>If you are moving data to a target HDFS</b></p>	<p>The default Hive table names are assigned by using the following format: &lt;schema_name&gt;.&lt;source_table_name&gt;. For example, Jane.NorthEastRegion. The default Hive schema name is the user name of the user that is running the activity. You can update the following components:</p> <ul style="list-style-type: none"> <li>• Hive table schema <p>You can use the schema name of the source table that you specified on the <b>Source</b> pane, or you can specify a new name. For example, I can specify JaneSmith_Customers. The Hive schema name would be JaneSmith_Customers, and the Hive table name would be JaneSmith_Customers.NorthEastRegion.</p> </li> <li>• Hive table <p>You can add a prefix or suffix to the Hive table name.</p> <ul style="list-style-type: none"> <li>– For example, if you specify 2014 as the table prefix, the Hive table name is Jane_Customers_2014NorthEastRegion_JSmith.</li> <li>– For example, if you specify critical as the table suffix, the Hive table name is Jane_Customers.NorthEastRegion_critical.</li> </ul> </li> </ul>
<p><b>If you are moving data to a target database</b></p>	<p>The default table name is &lt;prefix&gt;_&lt;source_table_name&gt;_&lt;suffix&gt;.</p> <p>If tables with the same name already exist in your target database, you can append the new tables to the existing tables by selecting <b>Append to existing tables</b>.</p>

Option	Description
<p>If you are moving data to a target Amazon S3 bucket</p>	<p>The default table name is &lt;prefix&gt;_&lt;source_table_name&gt;. Select <b>Include source schema in file name</b> if you want the name of the schema that you specified as your source in the name of the table that is created in the target Amazon S3 bucket.</p> <p>You can add a custom prefix to the file name. If you add a prefix to the file name, it affects where the file is stored.</p> <p>You can specify the file structure of the new file, how you want your data delimited, and so on.</p> <p>When you select <b>Create a file that documents its structure in the first row</b>, a file is created that documents the format of how your data is stored in the source database. For example, if you have a source database table that stores a customer name and a social security number with the column names <i>Customer_Name</i> and <i>SSN</i>, then the file will document the database's format in the first row as <code>Customer_Name:VarChar(20), SSN:VarChar(11)</code>. The comma is the column delimiter.</p>

Then click **Next**.

6. Review the information in the Summary pane. You see the name that was generated for the activity run and the policies that apply to the activity run.
7. Click **Run**. The activity is submitted for processing. When the job is processed:
  - Data is copied from the source that you selected and moved to the target that you selected.
  - The data that you are moving is registered in the metadata repository, and is accessible in Information Governance Catalog and other products in the InfoSphere Information Server suite.

A Hive table is also created for the source table for activities that move data from a database to a Hadoop Distributed File System (HDFS).

### What to do next

After you run an activity, you can view the status of the activity run in the Monitoring section of InfoSphere Data Click. You can select an activity and click **Run Again** to generate and run a new instance of the activity.

---

## Monitoring InfoSphere Data Click activities

After you run activities, you can monitor their status in the **Monitoring** section of InfoSphere Data Click.

The **Monitoring** section provides the following information:

- The name of the activity

- The name of the user who ran the activity
- The status of the activity
- The date that the job was submitted
- The date that the job was completed
- The total number of rows that were moved by the activity

Unless you have the role of InfoSphere Data Click Author, only the activities that you have been granted access to are displayed.

After activities are run, they are listed in the order that they are run. To find a particular activity, enter the name of the activity in the filter box. You can apply an advanced filter to display activities based on specific character strings in their name, the dates that jobs were submitted or completed, the person that ran them, and so on.

To refresh the status of activities that were run, click the refresh button in the top right of the screen. If the same activity has been run more than once, then there will be multiple entries for the activity.

By default, 50 activities are displayed on each page. You can navigate to other pages by clicking the **Page** option on the bottom of the screen.

If any errors occur while running an activity, you can view additional details about the job run by selecting the activity and clicking **View Details**. If you hover over the status icon, you can view:

- The overall status of the activity.
- The status of the job that moves data to the target.
- The status of the job that registers the data. After the data is registered, it is accessible in Information Governance Catalog and other products in the InfoSphere Information Server suite.

Select the individual jobs that you want to view in the Operations Console and select **View Details**. This will open a screen that displays the job details in the Operations Console.

---

## Reference

You specify where you want to move data to and where you want to move data from by creating an activity. After you create an activity, you run the activity to move the data.

To move data, you create an activity which specifies the source that you want to move data from and the target that you want to move data to. When you run the activity, data is copied from the source and moved to the target. When you create an activity, all databases and Amazon S3 buckets that you imported metadata for by using InfoSphere Metadata Asset Manager are available to use as sources. When you move the source data, you can either select the entire database or bucket, or you can move individual assets that the source contains. The data is then copied to an HDFS folder, an Amazon S3 bucket, or a relational database.

For example, if you are an InfoSphere Data Click Author and you want to enable business users to move data from the database `Customer_Demographics` to the target HDFS `Customer_Analytics`, then you would create the activity `Customer_Demographics_For_Team`, and select `Customer_Demographics` as the source and `Customer_Analytics` as the target. You would then save the activity and notify

your team that is available for them to run. Then, users with the InfoSphere Data Click User role can go into InfoSphere Data Click, select the activity `Customer_Demographics_For_Team`, and click **Run**. When they run the activity, they can accept the data that you preselected for them, or they can select a subset of the data available to them. After they run the activity, the data is moved to the `Customer_Analytics` HDFS, where it can be used in InfoSphere BigInsights.

You can also initiate data movement from the Information Governance Catalog. You can select data in the Information Governance Catalog that you want to move to target HDFS folders. The InfoSphere Data Click activity creation wizard is automatically launched from Information Governance Catalog when you select the option to move data to a target location.

## The `syncbi` command (Linux)

Run the `syncbi.sh` command to set up connections to Hadoop Distributed File Systems in IBM InfoSphere BigInsights.

The `syncbi.sh` command connects to the InfoSphere BigInsights computer and collects files and information that InfoSphere Data Click needs to access the HDFS.

You run the `syncbi.sh` command in the following circumstances:

- To initially connect to InfoSphere BigInsights.
- If a patch or fix pack has been installed to InfoSphere BigInsights.
- If the underlying Hadoop property files or configuration files on the InfoSphere BigInsights computer have been changed.
- If the host name or port number for InfoSphere BigInsights were changed, for example when moving from test into production.

### Procedure

1. Shut down any IBM InfoSphere Information Server Connectivity clients that might be running.
2. On the IBM InfoSphere Information Server engine tier computer, shut down the engine and node agents as either the root user or the DataStage administrator user. (The DataStage administrator user is `dsadm` if you accepted the default name during installation). For example, issue these commands:

```
cd <IS_install_path>/Server/DSEngine
. ./dsenv
./bin/uv -admin -stop
cd ../../ASBNode/bin
./NodeAgents.sh stop
```

3. If you intend to run the next step as the DataStage administrator user, first delete the `/tmp/ibm_is_logs` folder as the root user:

```
su - root
cd /tmp
rm -rf ibm_is_logs
```

4. Run the `syncbi.sh` command as either the root user or the DataStage administrator user. For example, issue the following command, which includes all of the required parameters:

```
cd <IS_install_path>/Server/biginsights/tools
./syncbi.sh -biurl protocol://bihost:biport -biuser username -bipassword
password -ishost hostname -isport portnumber
```

5. Restart the engine and node agents by issuing these commands as the DataStage administrator user:

```

su - dsadm
cd <IS_install_path>/Server/DSEngine
. ./dsenv
./bin/uv -admin -start
cd ../../ASBNode/bin
./NodeAgents.sh start

```

## Command syntax

```

<IS_install_path>/Server/biginsights/tools/syncbi.sh
-help |
-biurl value
-biuser value
-bipassword value
-ishost value
-isport value
[-{verbose | v}]

```

## Parameters

**-help** |

Displays command usage information.

**-biurl** *protocol://hostname:port*

Specifies the URL of the InfoSphere BigInsights computer.

*protocol*

Can be either https or http, depending on whether you want to connect to the InfoSphere BigInsights computer with or without SSL.

*hostname*

The host name of the InfoSphere BigInsights computer or master node. If the protocol is https, specify the host name exactly as it is specified in the InfoSphere BigInsights SSL certificate.

*port*

The InfoSphere BigInsights port number to use for the specified protocol.

**-biuser** *username*

Specifies the InfoSphere BigInsights administrator user name.

**-bipassword** *password*

Specifies the password for the InfoSphere BigInsights administrator user.

**-ishost** *hostname*

Specifies the host name of the IBM InfoSphere Information Server services tier computer.

**-isport** *portnumber*

Specifies the port number for the IBM InfoSphere Information Server services tier computer, for example, 9443 if the default installation port is used.

**[-{verbose | v}]**

**Optional:** Displays detailed information throughout the operation.

## Configuring access to Hive tables (AIX only)

You must configure access to Hive tables before you can use InfoSphere Data Click to move data into InfoSphere BigInsights. The Hive table is accessed by using the JDBC connector. You manually set up the driver configuration file for the JDBC connector and download the JDBC driver.

## About this task

The JDBC connector uses a driver configuration file to obtain information about the available JDBC drivers in your system.

If you are planning on using InfoSphere Data Click to move data into a Hadoop Distributed File System in InfoSphere BigInsights on a Linux computer, use the `syncbi` command to configure your computer. The `syncbi` command automatically creates and configures the `isjdbc.config` file.

## Procedure

1. Create or update the `isjdbc.config` file.
  - a. Open the existing driver configuration file or create a driver configuration file named `isjdbc.config` with read permission enabled for all the users. If you open the existing `isjdbc.config` file, then the file should be located in the `IS_HOME/Server/DSEngine` directory, where `IS_HOME` is the default installation directory. The driver configuration file name is case sensitive.
  - b. Open the file with a text editor and verify that the `CLASSPATH` entry points to the correct location of the Big SQL JDBC driver and that the `CLASS_NAMES` entry contains the Big SQL driver class name. If you create a new file, add the following two lines that specify the class path and driver Java™ classes:

```
CLASSPATH=$DSHOME/./biginsights/bigsql-jdbc-driver.jar
CLASS_NAMES=com.ibm.biginsights.bigsql.jdbc.BigSQLDriver
```
  - c. Save the `isjdbc.config` file in the `IS_HOME/Server/DSEngine` directory, where `IS_HOME` is the default installation directory. For example, the home directory might be `/opt/IBM/InformationServer`. If your installation consists of multiple hosts, this file must be available from the same location on all the hosts. You can make this file available from the same location on all the hosts, by configuring the `DSEngine` directory as a shared network directory.
2. Download and configure the JDBC driver.
  - a. While an instance of InfoSphere BigInsights is running, open a web browser and enter the following URL: `https://<host>:<port>/data/controller/FileDownload?target=bigsql`

**Note:** You can enter `http://<host>:<port>/data/controller/FileDownload?target=bigsql` if you are using an HTTP server that does not use an HTTPS connection.

### Host

Specifies the computer address that hosts the InfoSphere BigInsights web console.

### Port

Specifies the port that is used to communicate with the InfoSphere BigInsights web console. The default port number for the REST API is 8080. If you use the SSL encryption (https transport) the default port number is 8443.

- b. Click **Save** when you are prompted to save the `big-sql-client-kit.zip` file to your local file system.

- c. Move the downloaded `big-sql-client-kit.zip` file to the computer that is running the InfoSphere Information Server engine. For example, `/tmp/big-sql-client-kit.zip`.
- d. On the computer that is running the InfoSphere Information Server engine, log in as the InfoSphere DataStage administrator user. For example, log in as `dsadm` or the user that owns the `/opt/IBM/InformationServer/Server/` directory.
- e. Navigate to or create the following InfoSphere BigInsights directory on the InfoSphere Information Server computer: `/opt/IBM/InformationServer/Server/biginsights`.
- f. Unzip the `big-sql-client-kit.zip` file.
- g. Locate the `jdbc\bigsql-jdbc-driver.jar` file and move it to the following directory: `/opt/IBM/InformationServer/Server/biginsights`.

## What to do next

After making changes to the driver configuration file and downloading and moving the JDBC driver, you do not need to restart the InfoSphere Information Server engine, ISF agents or WebSphere® Application Server. The JDBC connector recognizes the changes that are made to this file the next time it is used to access JDBC data sources.

You must verify that node agents are up and running before you can use the connector. The following command starts the agents: `/IS_install_dir/ASBNode/bin/NodeAgents.sh start`. The following command stops the agents: `/IS_install_dir/ASBNode/bin/NodeAgents.sh stop`.

## Configuring access to HDFS (AIX only)

You must configure access to Hadoop Distributed File Systems (HDFS) in InfoSphere BigInsights. You create or update the `ishdfs.config` configuration file. Also, you must add the InfoSphere BigInsights client library and client configuration files to the computer that is running the InfoSphere Information Server engine.

### Before you begin

You must have the IBM InfoSphere DataStage administrator level access to configure the InfoSphere BigInsights client library and client configuration files.

### About this task

If you are planning on using InfoSphere Data Click to move data into a Hadoop Distributed File System in InfoSphere BigInsights on a Linux computer, use the `synbi` command to configure your computer. The `synbi` command automatically creates and configures the `ishdfs.config` file and sets up connections to HDFS.

You move and configure the InfoSphere BigInsights client library and client configuration files on the computer where the InfoSphere Information Server engine is installed.

When InfoSphere Information Server is configured to use the InfoSphere BigInsights REST API to communicate with InfoSphere BigInsights, it uses the same `ishdfs.config` configuration file as the HDFS bridge.

## Procedure

1. Create or update the `ishdfs.config` file.
  - a. Open the `ishdfs.config` configuration file or create a new one with read permissions enabled for all users of the computer where the InfoSphere Information Server engine is installed. The `ishdfs.config` file should be located in the `IS_HOME/Server/DSEngine` directory, where `IS_HOME` is the InfoSphere Information Server home directory. The configuration file name is case-sensitive.
  - b. Open a text editor and verify that the following line specifies the Java class path:

```
CLASSPATH= hdfs_bridge_classpath
```

If you are creating a new file, add the `CLASSPATH` line and specify the `hdfs_bridge_classpath` value. The value `hdfs_bridge_classpath` is the semicolon-separated list of file system artifacts that are used to access HDFS. Each element must be a fully qualified file system directory path or `.jar` file path. Below is the complete list of the required Java libraries and file system folders:

```
$DSHOME/../../ASBNode/eclipse/plugins  
/com.ibm.iis.client/httpclient  
-4.2.1.jar  
$DSHOME/../../ASBNode  
/eclipse/plugins/com.  
ibm.iis.client/  
httpcore-4.2.1.jar  
$DSHOME/../../PXEngine  
/java/biginsights-restfs-1  
.0.0.jar  
$DSHOME/../../PXEngine  
/java/cc-http-api.jar  
$DSHOME/../../PXEngine/  
java/cc-http-impl.jar:  
$DSHOME/../../biginsights  
/hadoop-conf  
$DSHOME/../../biginsights  
$DSHOME/../../biginsights/*
```

`$DSHOME` is the environment variable that points to the following directory: `IS_HOME/Server/DSEngine` directory. The default directory is `/opt/IBM/InformationServer/Server/DSEngine` directory. For example, create the `ishdfs.config` file and enter the following line:

```
CLASSPATH=$DSHOME/../../ASBNode/eclipse/  
plugins/com.ibm.iis.client/  
httpclient-4.2.1.jar:$DSHOME  
/../../ASBNode/eclipse  
/plugins/com.ibm.iis.client/  
httpcore-4.  
2.1.jar::$DSHOME/../../PXEngine/  
java/biginsights-restfs-1.0.0.  
jar:$DSHOME/../../PXEngine  
/java/cc-http-api.jar:$DSHOME  
/../../PXEngine/java/cc-http-impl.jar:  
$DSHOME/..  
/biginsights/hadoop  
-conf:$DSHOME/..  
biginsights:$DSHOME/../../biginsights/*
```

- c. Save the `ishdfs.config` file in the `IS_HOME/Server/DSEngine` directory, where `IS_HOME` is the InfoSphere Information Server home directory. The default directory is `/opt/IBM/InformationServer`. For example, `/opt/IBM/InformationServer/Server/DSEngine`. If the engine tier in your

InfoSphere Information Server installation consists of multiple hosts, this file must be available from the same location on all the hosts. You can make this file available from the same location on all the hosts, by configuring the DSEngine directory as a shared network directory.

2. Download and configure the InfoSphere BigInsights client library files.
  - a. While an instance of InfoSphere BigInsights is running, open a web browser and enter the following URL:

`https://<host>:<port>/data/controller/FileDownload?target=client`

**Note:** Enter `http://<host>:<port>/data/controller/FileDownload?target=client` if you are using an HTTP server that does not use an HTTPS connection.

**Host**

Specifies the computer address that hosts the InfoSphere BigInsights web console.

**Port**

Specifies the port that is used to communicate with the InfoSphere BigInsights web console. The default port number for the REST API is 8080. If you use the SSL encryption (https transport) the default port number is 8443.

- b. Click **Save** when you are prompted to save the `biginsights_client.tar.gz` file to your local file system.
  - c. Move the downloaded `biginsights_client.tar.gz` file to the computer that is running the InfoSphere Information Server engine. For example, `/tmp/biginsights_client.tar.gz`.
  - d. On the computer that is running the InfoSphere Information Server engine, log in as the InfoSphere DataStage administrator user. For example, log in as `dsadm` or the user that owns the `/opt/IBM/InformationServer/Server/` directory.
  - e. Create the following InfoSphere BigInsights directory on the InfoSphere Information Server computer: `/opt/IBM/InformationServer/Server/biginsights`.
  - f. Extract the `biginsights_client.tar.gz` file. For example:

```
dsadm> gunzip /tmp/biginsights_client
.tar.gz
dsadm> tar -C
/opt/IBM/InformationServer/Server/
biginsights -xvf /tmp/
biginsights_client.tar
```

The contents of the file are saved in the `/opt/IBM/InformationServer/Server/biginsights` directory.

3. Download and configure the InfoSphere BigInsights client configuration files.
  - a. While an instance of InfoSphere BigInsights is running, open a web browser and enter the following URL:

`https://<host>:<port>/data/controller/configuration`

**Note:** Enter `http://<host>:<port>/data/controller/configuration` if you are using an HTTP server that does not use an HTTPS connection.

**Host**

Specifies the computer address that hosts the InfoSphere BigInsights web console.

**Port**

Specifies the port that is used to communicate with the InfoSphere BigInsights web console. The default port number for the REST API is 8080. If you use the SSL encryption (https transport) the default port number is 8443.

- b. Click **Save** when you are prompted to save the configuration.zip file to your local file system.
- c. Move the downloaded configuration.zip file to the computer that is running the InfoSphere Information Server engine. For example, /tmp/configuration.zip.
- d. On the computer that is running the InfoSphere Information Server engine, log in as the InfoSphere DataStage administrator user. For example, log in as dsadm or the user that owns the /opt/IBM/InformationServer/Server/ directory.
- e. Navigate to or create the following InfoSphere BigInsights directory on the InfoSphere Information Server computer: /opt/IBM/InformationServer/Server/biginsights.
- f. Extract the configuration.zip file. For example:

```
dsadm> cp /tmp/configuration.zip
/opt/IBM/InformationServer/
Server/biginsights
dsadm> unzip
/opt/IBM/InformationServer/Server
/biginsights/configuration.zip
```

The contents of the file are saved in the /opt/IBM/InformationServer/Server/biginsights directory.

- g. Open the log4j.properties file in the /opt/IBM/InformationServer/Server/biginsights/ directory and add the following string:  
log4j.logger.org.apache.hadoop.util.NativeCodeLoader=ERROR

**Results**

After making changes to the configuration file and downloading the InfoSphere BigInsights client library and client configuration files, you do not need to restart the InfoSphere Information Server engine, ASB Agent, or WebSphere Application Server. The changes that you made are updated the next time the ishdfs.config file is used to access HDFS.

**Checking activities in the Operations Console**

You can check the status of activities in the IBM InfoSphere DataStage and QualityStage Operations Console. The Operations Console provides a more detailed view of the status of jobs that are generated when you run activities.

**About this task**

The Operations Console provides the most detailed view of the status of activities, however you can also check the status of activities in the **Monitoring** section of InfoSphere Data Click.

When you open the view of the InfoSphere Data Click activity, you might need to provide the user name and password that enables you to access the Operations Console.

## Procedure

1. In InfoSphere Data Click, select the activity that you want to view in the Operations Console and click **View Details**. Individual jobs that are associated with the activity are displayed.
2. Select the individual jobs that you want to view in the Operations Console and select **View Details**. In the Operations Console, you can see the details of the job of the activity.
3. To view other jobs associated with the activity, close the window and, while still in the Operations Console, search for the names of jobs that match the activity in the **Navigation** panel on the **Projects** tab.
  - a. Expand the **Jobs** folder, the folder with the name of the InfoSphere DataStage project that you are using in your activities, and the **OffloadJobs** sub-folder. The default InfoSphere DataStage project name is DataClick.

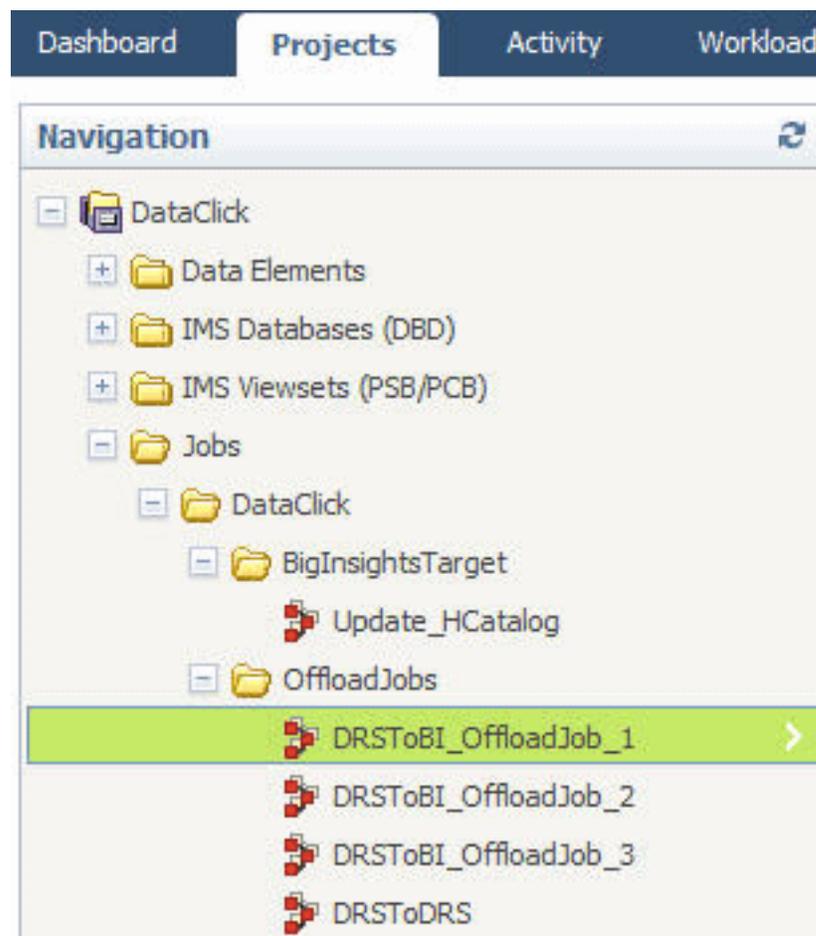


Figure 5. The navigation in the Operations Console

- b. Select the job for your data movement pattern.
  - DRSToBI\_OffloadJob moves data from a relational database to InfoSphere BigInsights.
  - DRSToDRS moves data from a relational database to a relational database.
  - DRSToS3 moves data from a relational database to Amazon S3.
  - S3ToBI\_OffloadJob moves data from Amazon S3 to InfoSphere BigInsights.

- S3ToDRS moves data from Amazon S3 to a relational database.
  - S3SchemaFileCreator is run when Amazon S3 is selected as a source.
- c. In the **Job Runs** section, select the particular data movement job that you want more details about and click **View Details**.

**Related information:**

 [Overview of the Operations Console](#)

## Hive tables

Hive tables are automatically created every time you run an activity that moves data from a relational database into a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights. One Hive table is created for each table in the source that you specify in the activity. After Hive tables are created, you can use IBM Big SQL in InfoSphere BigInsights to read the data in the tables.

Before you create or run activities in InfoSphere Data Click, you must import one existing Hive table into the metadata repository by using InfoSphere Metadata Asset Manager and specifying the JDBC connector.

After you import the Hive table, along with the sources and targets that you want to use, you can create and run activities. When you initially run an activity, a single folder is created and associated with the activity. For example, `Marketing_Reports`. Every time that you run the activity, a sub-folder is created inside of `Marketing_Reports`, and the sub-folder is given the name of the activity instance. The Hive tables that are generated during the activity run are then stored in the folder that was created for each run.

**Note:** If the names of the tables and columns that you specify as your source contain non-ASCII or extended ASCII characters, then a Hive table is not generated. Also, when you specify the name of a Hive table, the name you specify cannot contain non-ASCII or extended ASCII characters. Non-ASCII and extended ASCII characters are not supported in the names of Hive tables, however they are allowed in the actual content of the columns and tables.

When the Hive table is generated, all columns in the table are separated by the delimiter that was specified on the Policies pane when you created the activity. All rows in the table are delimited with the `\n` control character.

After you run an activity, a Hive table is created and automatically imported into the metadata repository. You can view the table on the Repository Management tab in InfoSphere Metadata Asset Manager. In the Navigation pane, expand **Browse Assets**, and then click **Implemented Data Resources**. The Hive table will be located under the host that was specified when you created the data connection to the original Hive table that you imported by using the JDBC connector. For example, in the image below, the data connection that uses the JDBC connector to connect to the Hive table specifies the host `IPSVM00104`. The Hive table is `bitest`.

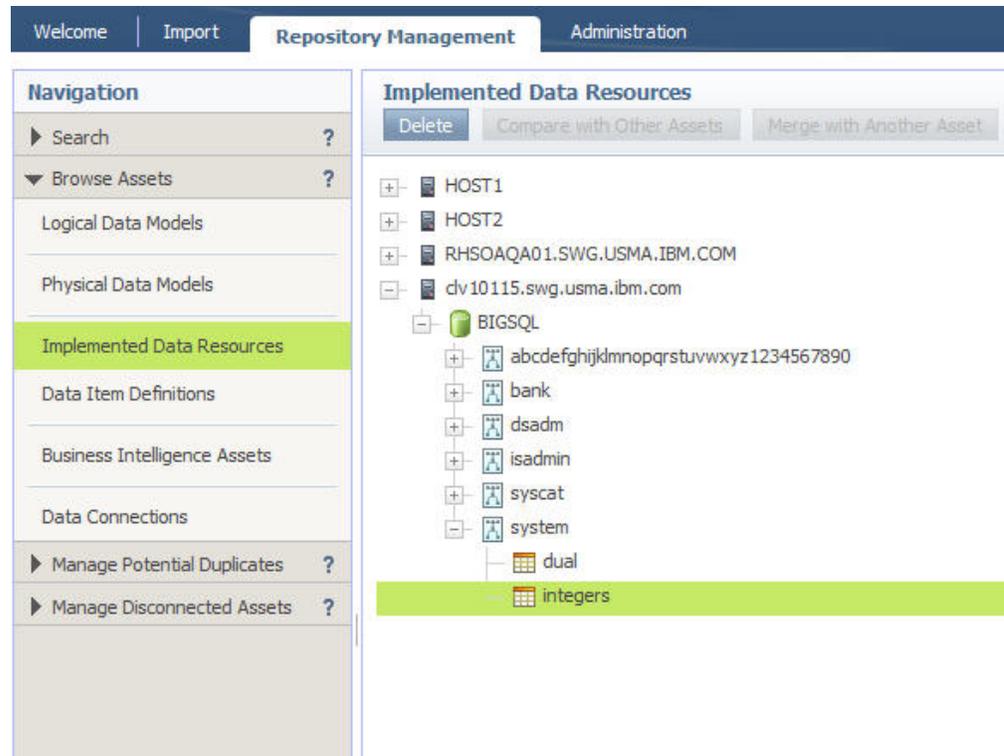


Figure 6. A view of the Hive table in InfoSphere Metadata Asset Manager

You can also view the Hive table in the Information Governance Catalog. In the image below, the original source database table that you selected to move data from in the InfoSphere Data Click activity is the PROJECT asset. The Hive table that was generated when the activity was run is the project database table.



Figure 7. A view of the Hive table in Information Governance Catalog

In Information Governance Catalog, you can add data from the Hive table to collections and view details about it.

For example, the user Jane Doe creates the activity, HRActivity to move the table, Customer (source) into the InfoSphere BigInsights folder MarketingReports (target). When she runs the activity, a Hive table is automatically created and is mapped to the /MarketingReports/HRActivity\_JaneDoe\_1381231526864/JaneDoe\_Customer folder. The Hive table contains relational information about the actual data that is in the Customer table. You can access the Hive table JaneDoe.Customer and then read the actual data in the table by using Big SQL.

### Mapping source table data to Hive tables

A Hive table is automatically created every time you run an activity that moves data from a relational database into a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights. The data type of each column in the Hive table is automatically assigned based on metadata information that is detected about the data type of each column in the source database that you specify when you run the activity.

InfoSphere Data Click uses Big SQL to create the Hive table. As shown in the following table, the data type of columns in the source table are mapped to Big SQL data types for the columns in the Hive table.

Table 4. Mapping the data type of columns in the source table to Big SQL data types for the columns in the Hive table

ODBC type in the metadata repository	Big SQL Data type used for creating the column in Hive
BIGINT	BIGINT
BINARY	BINARY ( <i>length</i> ). The <i>length</i> parameter is the length of the source column. If a length is not provided in the source, then the data type is BINARY.
BIT	BOOLEAN
CHAR	CHAR( <i>length</i> ). The <i>length</i> parameter is the length of the source column. If a length is not provided in the source, then the data type is CHAR.
DATE	VARCHAR(10)
DECIMAL	VARCHAR( <i>precision</i> +2). The precision value specifies the precision for the source column. If the <i>precision</i> parameter is not provided in the source, then the data type is STRING.
DOUBLE	DOUBLE
FLOAT	FLOAT
INTEGER	INT
LONGVARBINARY	VARBINARY(32768) <b>Note:</b> LONGVARBINARY is mapped to VARBINARY(32768). Big SQL places a restriction on the <i>length</i> of the VARBINARY column. Because of this restriction, data truncation might occur when you move data from a source table containing a LONGVARBINARY column. This can lead to the activity finishing with warnings.
LONGVARCHAR	STRING
NUMERIC	NUMERIC( <i>precision</i> , <i>scale</i> ). The precision and scale values specify the precision and scale for the source column. If the <i>precision</i> or <i>scale</i> parameters are not provided in the source, then the data type is NUMERIC.
REAL	FLOAT
SMALLINT	SMALLINT
TINYINT	INT
TIME	VARCHAR(8)
TIMESTAMP	VARCHAR(26)
VARBINARY	BINARY( <i>length</i> ). The <i>length</i> parameter is the length of the source column. If a length is not provided in the source, then the data type is BINARY.

Table 4. Mapping the data type of columns in the source table to Big SQL data types for the columns in the Hive table (continued)

ODBC type in the metadata repository	Big SQL Data type used for creating the column in Hive
VARCHAR	VARCHAR( <i>length</i> ). The <i>length</i> parameter is the <i>length</i> of the source column. If the <i>length</i> is not provided in the source, then the data type is VARCHAR.
WCHAR	CHAR( <i>length</i> ). The <i>length</i> parameter is the length of the source column. If a length is not provided in the source, then the data type is CHAR.
WLONGVARCHAR	STRING
WVARCHAR	VARCHAR( <i>length</i> ). The <i>length</i> parameter is the <i>length</i> of the source column. If the <i>length</i> is not provided in the source, then the data type is VARCHAR.

## Creating .osh schema files

Before you can move data from Amazon S3 to various targets, you might need to create an .osh schema file for each Amazon folder or for individual files that you are moving.

### About this task

OSH schema files define how data is formatted in Amazon S3 files and folders. If you are moving data from Amazon S3, you must create an .osh schema file for each folder that you move. If all files in an Amazon S3 folder have the same format, then you need to create only one .osh schema file per folder and name the .osh file the same name as the folder. If the files in a folder are a different format, then you need to create one .osh schema file per Amazon file and name the .osh file the same name as the Amazon file. For example, if all of the files in the folder Customers have the same file structure, except for one file named customer\_addresses.txt, then you need to create two .osh schema files. You need to create the .osh schema file Customers.osh that details the file structure of most of the files in the folder. You also need to create the .osh schema file customer\_addresses.txt.osh that details the file structure of the one file that has a different structure.

If you used InfoSphere Data Click to move data from a relational database or from InfoSphere BigInsights to Amazon S3, then .osh schema files were automatically created for the data. You do not need to create additional .osh schema files when you move that data from Amazon S3 to a new target. When you import the Amazon S3 metadata by using InfoSphere Metadata Asset Manager, select **Import file structure** to import the .osh schema files.

If you run an activity and it fails with an error that says a file on Amazon S3 cannot be moved because the parameters in the .osh schema files does not match the format of the Amazon S3 data, then you should go back to the file you are trying to move on Amazon S3, and update the file so the structure of the data matches the structure defined in the .osh schema file. If you decide to update the .osh schema file instead of updating your data, then you need to reimport the Amazon S3 data and osh file by using InfoSphere Metadata Asset Manager.

## Procedure

1. Open a blank new .osh schema file in a text editor.
2. Add a line to your file that specifies the formatting properties of your data. For example, // FileStructure: file\_format='csv', header='true', escape=''. You must specify each of the following properties:

Table 5. Amazon S3 formatting properties. Formatting properties you can use for the Amazon S3 connector.

Property	Supported values for the property
file_format	<ul style="list-style-type: none"> <li>• csv</li> <li>• delimited</li> <li>• redshift</li> </ul>
header	<ul style="list-style-type: none"> <li>• true</li> <li>• false</li> </ul>
escape	<i>escape_character</i>

3. For example, specify the record parameter if you are creating an .osh schema file for a spreadsheet that contains the columns SALES\_DATE, SALES\_PERSON, REGION, and SALES.

```
record { record_delim='\n', delim=',', final_delim=end, null_field='' } (
  SALES_DATE: nullable date;
  SALES_PERSON: nullable string[max=15];
  REGION: nullable string[max=15];
  SALES: nullable int32;
)
```

Click the property for specific details about how you specify the value in the .osh schema file.

Table 6. Properties from the OSH schema that you can use for the Amazon S3 connector.

Property	Supported values for the property
record_delim	<ul style="list-style-type: none"> <li>• 'one-character_delimiter'</li> <li>• null</li> </ul>
record_delim_string	<i>"delimiter_string"</i>
final_delim	<ul style="list-style-type: none"> <li>• end</li> <li>• 'one-character_delimiter'</li> <li>• null</li> </ul>
delim	<i>"delimiter_string"</i>
final_delim_string	<ul style="list-style-type: none"> <li>• 'one-character_delimiter'</li> <li>• null</li> </ul>
null_field	<ul style="list-style-type: none"> <li>• 'one-character_delimiter'</li> <li>• <i>"delimiter_string"</i></li> </ul>
quote	<ul style="list-style-type: none"> <li>• single</li> <li>• double</li> <li>• 'one-character_delimiter'</li> </ul>
charset	<i>character_set</i>
date_format	<i>format</i>
time_format	<i>format</i>

Table 6. Properties from the OSH schema that you can use for the Amazon S3 connector (continued).

Property	Supported values for the property
timestamp_format	<i>format</i>

- Save the file with the same file name as the Amazon S3 file or folder that you are creating the .osh schema file for. For example, if the Amazon S3 folder has the name DB\_Sales\_Data, then you need to name the .osh schema file DB\_Sales\_Data.osh. If the Amazon S3 file has the name Sales\_Data.txt, then you need to name the .osh schema file Sales\_Data.txt.osh.

### Example

Below is an example of an .osh file for an Amazon S3 folder that contains spreadsheet data with the following columns:

- Patient\_ID
- Date\_of\_visit
- Address
- Gender

Commas are used as delimiters in the spreadsheet.

```
// FileStructure: file_format='csv', header='true', escape=''
record { record_delim='\n', delim=',', final_delim=end, null_field='' } (
  PATIENT_ID: nullable string[max=27];
  DATE_OF_VISIT: nullable date;
  ADDRESS: nullable string[max=205];
  GENDER: nullable string[max=1];
)
```

## Assigning user roles for InfoSphere Data Click installations that use LDAP configured user registries

If you have configured InfoSphere Data Click to use a lightweight directory access protocol (LDAP) user registry, then all users are already created. You just need to assign the users the appropriate user roles and assign them to InfoSphere DataStage projects by using the IBM InfoSphere Information Server Administration console and the DirectoryCommand tool. The InfoSphere Data Click administrator command-line interface is not supported if you use LDAP configured user registries.

### Before you begin

You must have the DataStage and QualityStage Administrator or the Suite Administrator role to assign InfoSphere DataStage user roles and add users to projects.

### About this task

You need to assign InfoSphere Data Click authors and users roles before you can create and run activities in InfoSphere Data Click.

You must also assign InfoSphere DataStage roles to InfoSphere Data Click users and assign them InfoSphere DataStage projects before they can move data by using InfoSphere Data Click.

InfoSphere Data Click Authors must have the DataStage Developer role before they can create and run activities. InfoSphere Data Click Users must have the DataStage Operator role before they can run activities.

**Note:** The DataStage Developer and DataStage Operator roles are assigned at the project level. An InfoSphere Data Click Author might need the DataStage Developer role for one project, in order to create activities. However, the same InfoSphere Data Click Author might be just a user on a different project that another InfoSphere Data Click Author manages. In this situation, the author needs only the DataStage Operator role to run an activity in that project.

You use the IBM InfoSphere Information Server Administration console to assign InfoSphere Information Server roles to InfoSphere Data Click.

You use the DirectoryCommand tool to assign InfoSphere DataStage user roles and add users to projects.

## Procedure

1. To assign InfoSphere Information Server roles:
  - a. Open the IBM InfoSphere Information Server Administration console.
  - b. Click the Administration tab, and then click **Users and Groups > Users**.
  - c. In the Users pane, search for the user that you want to grant user roles to and edit the user profile.
  - d. Assign the following InfoSphere Information Server roles:

Table 7. InfoSphere Information Server roles

InfoSphere Data Click role	InfoSphere Information Server roles
For Data Click Authors	<ul style="list-style-type: none"> <li>• Data Click Author</li> <li>• Suite User</li> <li>• DataStage and QualityStage User</li> <li>• Common Metadata Administrator</li> <li>• Information Governance Catalog Administrator</li> </ul>
For Data Click Users	<ul style="list-style-type: none"> <li>• Data Click User</li> <li>• Suite User</li> <li>• DataStage and QualityStage User</li> <li>• Common Metadata User</li> </ul>

- e. Click **Save** to save the user information in the metadata repository.
  - f. Repeat steps 1c, 1d, and 1e to create additional users.
  - g. Click **Save and Close**, and then exit the InfoSphere Information Server Administration console.
2. To assign InfoSphere DataStage roles to InfoSphere Data Click users and assign them to InfoSphere DataStage projects:
  - a. Navigate to the IBM InfoSphere DataStage and QualityStage DirectoryCommand tool in the `IS_install_path/ASBNode/bin` directory.
  - b. Issue the following command to assign a InfoSphere Data Click Author the DataStage Developer role and add them to the InfoSphere Data Click project:
 

```
DirectoryCommand.sh -url URL_value -user admin_user -password password -authfile authfile_path -assign_project_user_roles
```

```
"DataStage_engine_name/  
Data_Click_project_name$user_name$DataStageDeveloper"
```

Parameters:

**-url**

Required

The base url for the web server where the InfoSphere Information Server web console is running. The url has the following format: `https://<host>:<https port>`.

**-user**

Required

The name of the Suite Administrator. This parameter is optional if you use the **-authfile** parameter instead of specifying the **-user** and **-password** parameters. You can enter a user name without entering a password, which causes you to get prompted for a password.

**-password**

Required

The password of the Suite Administrator. This parameter is optional if you use the **-authfile** parameter instead of specifying the **-user** and **-password** parameters.

**-authfile (-af)**

Optional

The path to a file that contains the encrypted or non-encrypted credentials for logging on to InfoSphere Information Server. If you use the **-authfile** parameter, you do not need to specify the **-user** or **-password** parameter on the command line. If you specify both the **-authfile** option and the explicit user name and password options, the explicit options take precedence over what is specified in the file. For more information see the topic *Encrypt command* in the *IBM InfoSphere Information Server Administration Guide*.

**DataStage engine name**

Required

The name of the InfoSphere Information Server engine. You can issue the following command to determine the name of the engine:

```
DirectoryCommand.sh -url URL_value -user admin_user -password password  
-list DSPROJECTS
```

The following information is returned when you run that command:

```
----Listing DataStage projects.---  
DataStage project: "DataStage_engine_name/Data_Click_project_name"  
-----
```

For example,

```
----Listing DataStage projects.---  
DataStage project: "DATANODE1EXAMPLE.US.IBM.COM/DataClick"  
DataStage project: "DATANODE1EXAMPLE.US.IBM.COM/dstage1"  
-----
```

The name of the InfoSphere Information Server engine is, `DATANODE1EXAMPLE.US.IBM.COM`.

### Data Click project name

Required

The name of the InfoSphere Data Click project. The name of the default project that is automatically installed with InfoSphere Data Click is DataClick.

### User name

Required

The name of the InfoSphere Data Click user that you want to grant the DataStage Developer role.

For example, if you want to add the InfoSphere Data Click Author isadmin to the InfoSphere Data Click project, DataClick, and the name of your engine is DATANODE1EXAMPLE.US.IBM.COM, then you issue the following command: /opt/IBM/InformationServer/ASBNode/bin/DirectoryCommand.sh -url https://DATANODE1EXAMPLE.US.IBM.COM:9443 -user isadmin -password temp4now -assign\_project\_user\_roles "DATANODE1EXAMPLE.US.IBM.COM/DataClick\$isadmin\$DataStageDeveloper".

- c. Issue the following command to assign an InfoSphere Data Click User the DataStage Operator role and add them to the InfoSphere Data Click project:

```
IS_install_path/ASBNode/bin/DirectoryCommand.sh -url <URL_value>  
-user <name_of_admin_user> -password <password>  
-assign_project_user_roles "<DataStage_engine_name>/  
<Data_Click_project_name>$<name_of_user>$DataStageOperator"
```

For example, if you are the InfoSphere Data Click Author isadmin, and you want to add the user JohnDoe1 to the InfoSphere Data Click project, DataClick, and the name of your engine is DATANODE1EXAMPLE.US.IBM.COM, then you issue the following command:

```
/opt/IBM/InformationServer/ASBNode/bin/DirectoryCommand.sh -url  
https://9.146.90.235:9443 -user isadmin -password temp4now  
-assign_project_user_roles "DATANODE1EXAMPLE.US.IBM.COM/  
DataClick$JohnDoe1$DataStageOperator".
```

To verify that InfoSphere Data Click users were successfully added to a project, you can issue the following command:

```
DirectoryCommand.sh -url URL_value -user admin_user -password password -list  
USERS -details
```

This provides a list of all of the projects that a user is assigned to.

## Data connections in InfoSphere Data Click

A data connection is a reusable connection between InfoSphere Data Click and Amazon S3, a database, a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights, or a Hive table. You create data connections in InfoSphere Metadata Asset Manager, and use them to share metadata to the metadata repository. After connections are created, they are stored and available for other users to use when they are moving data by using InfoSphere Data Click.

When you create a data connection by using InfoSphere Metadata Asset Manager, you should name it something that uniquely identifies the source that it is connecting to, and something that will be intuitive for other users to identify. You can save the password or access key of the data connection so it can be easily used by InfoSphere Data Click authors and users.

If you save the password or access key, then InfoSphere Data Click authors can create activities and select sources to move data from and targets to move data to without having to enter a password or key.

If you do not save the password or access key, authors must enter the password or key when they select a source or target that uses the data connection when they create an activity.

If you are creating a data connection to Amazon S3, you have the option of creating a credentials file or specifying an access key and a secret key. If you use a credentials file, InfoSphere Data Click authors and users will not be prompted to enter a password when they use the data connection. The data connection will authenticate by using the credentials stored in the credentials file. If you use an access key and a secret key, without saving the secret key, then users have to enter the access key and password when they try to use the data connection in InfoSphere Data Click. If you save the secret key, then users do not need to enter data connection credentials when they use the data connection in InfoSphere Data Click.

You can create two data connections to the same data source and configure one data connection to prompt users for a password, and save the password in the second data connection. This helps to manage who can access each data source.

To successfully create activities that move data into a Hadoop Distributed File System (HDFS) in InfoSphere BigInsights, you need to import at least one Hive table into the metadata repository by using InfoSphere Metadata Asset Manager and specifying the JDBC connector. One of the parameters in the data connection that you create is a URL.

**New Data Connection** [X]

\* Name:

Description:

\* URL:

Enter credentials if they are required for the connection:

User name:

Password:

Save password

Attributes:

?

Figure 8. The URL parameter in a data connection that uses the JDBC connector in InfoSphere Metadata Asset Manager

You must enter a URL that consists of the following components:  
 jdbc:bigsql://<host\_name>:<BigSQL\_Port>/default. For example,  
 jdbc:bigsql://9.182.77.14:7052/default. The <host\_name> parameter can be an IP  
 address, short host name, or fully qualified host name.

For every folder in the HDFS that you want to use as a target in InfoSphere Data  
 Click, you must have a data connection that uses the HDFS bridge to connect to  
 the HDFS. When you import the folder in the HDFS by using InfoSphere Metadata  
 Asset Manager, and create a data connection, you must specify a host.

The screenshot shows a 'New Data Connection' dialog box with the following fields and values:

- Name:** Marketing\_DB2\_database
- Description:** (empty)
- Host:** 9.182.77.14 (highlighted with a red box)
- User:** isadmin
- Password:** (masked with dots)
- Save password:**
- Port:** (empty)
- SSL:**  SSL (You must provide truststore file and password below.)
- Truststore file:** (empty)
- Truststore password:** (empty)

Buttons at the bottom: ? (help), OK, Cancel.

Figure 9. The Host parameter in a data connection that uses the HDFS bridge in InfoSphere Metadata Asset Manager

The host name must be the fully qualified host name of the computer that hosts the InfoSphere BigInsights web console. The host that you specify must match the host name that you specified as part of the URL field for at least one data connection that connects to a Hive table by using the JDBC connector. For example, if you specified `jdbc:bigsql://9.182.77.14:7052/default` as the URL in a data connection that uses the JDBC connector, then you must specify `9.182.77.14` as the host value for the data connection that uses the HDFS bridge. If you have a JDBC data connection that has a URL that includes a host name that is identical to an HDFS data connection that specifies an identical host in the data connection, you will be able to successfully create and run data click activities.

To view a full list of existing data connections, open InfoSphere Metadata Asset Manager and click the **Repository Management** tab. In the Navigation pane,

expand **Browse Assets**, and then click **Implemented Data Resources > Data Connections**. You can view details about each data connection in the panel on the right side of the screen.

## Creating InfoSphere DataStage projects for InfoSphere Data Click activities

InfoSphere Data Click is installed with the default InfoSphere DataStage project, DataClick. However, you can create additional InfoSphere DataStage projects to use in your InfoSphere Data Click activities.

### Before you begin

You must be an InfoSphere DataStage administrator to create projects.

### About this task

When you create and run InfoSphere Data Click activities, you must specify InfoSphere DataStage projects. Users are assigned access to activities at the project level, so having multiple projects helps you manage who has access to various activities that you create.

### Procedure

1. Use the **dsadmin** command or the IBM InfoSphere DataStage and QualityStage Administrator to create a new project.

Option	Description
To use the <b>dsadmin</b> command to create a new project	<p>Open the command by locating the directory where it is installed. If you used the default installation directory, the command is installed in the following directory:</p> <ul style="list-style-type: none"> <li>• <b>Linux</b>    <b>UNIX</b> /opt/IBM/InformationServer/Server/DSEngine/bin</li> <li>• <b>Windows</b>    C:\IBM\InformationServer\Server\DSEngine\bin</li> </ul> <p>For information about using this command, see Creating a project.</p>
To use the IBM InfoSphere DataStage and QualityStage Administrator to create a new project	<p>For information about creating a new project by using the IBM InfoSphere DataStage and QualityStage Administrator, see Adding projects.</p>

2. **Recommended:** Enable automatic purging of InfoSphere DataStage logs. If too many log files are generated and saved on your machine, it can cause performance issues, as well as create errors when you run activities. For information on disabling automatic purging of log files, see Enabling/disabling automatic purging of log files.

### What to do next

After you create your InfoSphere DataStage projects, you need add users to the projects and import the InfoSphere Data Click job templates into the projects.

## Importing job templates

Every time that you create a new InfoSphere DataStage project to use in InfoSphere Data Click activities, you must import InfoSphere Data Click job templates into the InfoSphere DataStage projects.

### Before you begin

You must have the InfoSphere DataStage Administrator role for the project that you are importing job templates into.

### About this task

You must import InfoSphere Data Click job templates into all InfoSphere DataStage projects that you want to use in InfoSphere Data Click, other than the default InfoSphere DataStage project DataClick. The job templates are automatically imported for the default project DataClick.

When users run activities to move data, the data movement is fulfilled through a job instance. All job instances are generated from jobs that are imported by using the InfoSphere DataStage and QualityStage Designer or the InfoSphere DataStage command line interface.

All jobs are pre-compiled, so you do not have to compile any jobs after you import the job templates.

### Procedure

Use the InfoSphere DataStage and QualityStage Designer or the InfoSphere DataStage command line interface to import job templates.

Option	Description
To use the InfoSphere DataStage and QualityStage Designer	<ol style="list-style-type: none"><li>1. Open the InfoSphere DataStage and QualityStage Designer client.</li><li>2. Enter your InfoSphere DataStage credentials and select the InfoSphere Data Click project. Click <b>Login</b>.</li><li>3. Select <b>Import &gt; DataStageComponents</b>.</li><li>4. In the <b>Import from file</b> field, go to the InfoSphere Data Click directory on the engine tier and select the .dsx file. For example, IBM/InformationServer/Server/DataClick/bin/DataClick.dsx.</li><li>5. Ensure <b>Import All</b> is selected and then click <b>OK</b> to import the template jobs and job parameters into repository folders.</li></ol>

Option	Description
<p>To use the InfoSphere DataStage command line interface</p>	<ol style="list-style-type: none"> <li>Open the following directory: <ul style="list-style-type: none"> <li><b>Linux</b> <b>UNIX</b> /opt/IBM/InformationServer/ASBNode/bin/</li> <li><b>Windows</b> C:\IBM\InformationServer\ASASBNode\bin\</li> </ul> <p>If you did not use the default installation directory, replace the /opt/IBM/InformationServer or C:\IBM\InformationServer part of the file path with the path to the directory that you specified during installation.</p> <p>You run the <b>DSXImportService</b> command from the InfoSphere DataStage command line interface and import the DataClick.dsx file. For information about running the <b>DSXImportService</b> command, see Importing objects from a .dsx file.</p> </li> <li>Issue the following command to import the job templates: <ul style="list-style-type: none"> <li><b>Linux</b> <b>UNIX</b> <code>IS_install_dir/ASBNode/bin/DSXImportService.sh -ISHost &lt;IS_HOST_NAME&gt;:&lt;IS_PORT&gt; -ISUser &lt;IS_ADMIN_USER_ID&gt; -ISPassword &lt;IS_ADMIN_PASSWORD&gt; -DSHost &lt;DS_HOST_NAME&gt;:&lt;DS_PORT&gt; -DSProject &lt;DataStage_Project_Name&gt; -DSXFile IS_install_dir/Server/DataClick/bin/DataClick.dsx -Overwrite</code></li> <li><b>Windows</b> <code>IS_install_dir\ASBNode\bin\DSXImportService.bat -ISHost &lt;IS_HOST_NAME&gt;:&lt;IS_PORT&gt; -ISUser &lt;IS_ADMIN_USER_ID&gt; -ISPassword &lt;IS_ADMIN_PASSWORD&gt; -DSHost &lt;DS_HOST_NAME&gt;:&lt;DS_PORT&gt; -DSProject &lt;DataStage_Project_Name&gt; -DSXFile IS_install_dir\Server\DataClick\bin\DataClick.dsx -Overwrite</code></li> </ul> </li> </ol>

Parameters for using the InfoSphere DataStage command line interface:

**-ISHost**

Required

Specifies the name of the computer that hosts the domain.

**-ISuser**

Required

The name of the Suite Administrator.

**-ISpassword**

Required

The password of the Suite Administrator.

**-DSHost**

Required

Specify the computer that hosts the engine tier.

**-DSProject**

Required

The name of the InfoSphere DataStage project that you want to import.

**-DSXFile**

Required

The file path to the .dsx file.

For example, if you want to import the InfoSphere Data Click template jobs into the InfoSphere DataStage project, you would issue the following command:  
`/opt/IBM/InformationServer/ASBNode/bin/DSXImportService.sh -ISHost  
isserver:9443 -ISuser isadmin -ISpassword temp4now -DSProject MyProject  
-DSXFile /opt/IBM/InformationServer/Server/DataClick/bin/DataClick.dsx  
-Overwrite.`

## Configuring the workload management queue

You can manage data movement jobs that are generated when you run InfoSphere Data Click activities by configuring the workload management settings on the **Workload Management** tab in the IBM InfoSphere DataStage and QualityStage Operations Console.

### About this task

When you run InfoSphere Data Click activities, jobs are automatically created and submitted to move the data that is defined in the activities. There is a predefined default InfoSphere Data Click queue on the **Workload Management** tab that you can use to set system policies, define the order in which jobs are run, set the thresholds for job counts, and manage CPU and memory use.

The CPU and memory use setting is particularly important to InfoSphere Data Click. The default CPU and memory use setting prevents jobs that are generated from activities from running if your machine has less than twenty percent of its memory available and less than twenty percent of its CPU.

You must have the DataStage and QualityStage Administrator role to complete this task.

### Procedure

1. Open the Operations Console and select the **Workload Management** tab.
2. Modify the workload management system policies and set thresholds for job counts, CPU usage, and memory usage. For more information on configuring these settings, see **Workload Management** tab.

## **What to do next**

After you run InfoSphere Data Click activities, you can move particular jobs to the top of the queue or remove jobs from the queue. For information on configuring the queue, see Queued Jobs panel.

---

## Chapter 5. Troubleshooting InfoSphere Data Click

Certain problems can occur while you are using InfoSphere Data Click. The issues documented in this section provide descriptions of the problems and steps to correct them.

---

### Troubleshooting InfoSphere Data Click activities that fail to run or finish with warnings

If the Monitoring section of InfoSphere Data Click indicates that an activity failed to run or that finished with warnings, you can use the Operations Console or check the job logs to troubleshoot the problem.

#### About this task

When you run activities, jobs are submitted to InfoSphere DataStage for processing. Below is a list of InfoSphere Data Click jobs that are run by InfoSphere DataStage:

- DRSToBI\_OffloadJob moves data from a relational database to InfoSphere BigInsights.
- DRSToDRS moves data from a relational database to a relational database.
- DRSToS3 moves data from a relational database to Amazon S3.
- S3ToBI\_OffloadJob moves data from Amazon S3 to InfoSphere BigInsights.
- S3ToDRS moves data from Amazon S3 to a relational database.
- S3SchemaFileCreator is run when Amazon S3 is selected as a target.

#### Procedure

1. View the details about a job failure in the Operations Console.
  - a. In the Monitoring section of InfoSphere Data Click, select the name of the activity that has errors or that failed to run and then click **View Details**. Then, select the individual jobs that you want to view in the Operations Console and select **View Details**.
  - b. In the Operations Console, review the details of the jobs that failed or had errors.
2. If you are unable to troubleshoot the issue by using the Operations Console, review job logs that are generated when an activity is run to gather more details about why the job failed or finished with warnings.
  - a. Navigate to the directory where job logs are stored. The default directory is *IS\_install\_path/Server/Projects/Name\_of\_DS\_project/DCLogs/name\_of\_activity*. The name of the default project that is automatically installed with InfoSphere Data Click is *DataClick*. For example, */opt/IBM/InformationServer/Server/Projects/DataClick/DCLogs/Activity1*.
  - b. Open one of the following job logs in the folder:

#### **DRSToBI\_OffloadJob**

If you see a job log for the DRSToBI\_OffloadJob job, then a problem occurred when moving data from a relational database to InfoSphere BigInsights.

### **DRSToDRS**

If you see the job log for the DRSToDRS job, then a problem occurred while moving data from a relational database to a relational database.

### **DRSToS3**

If you see the job log for the DRSToS3 job, then a problem occurred while moving data from a relational database to Amazon S3.

### **S3ToBI\_OffloadJob**

If you see the job log for the S3ToBI\_OffloadJob job, then a problem occurred while moving data from Amazon S3 to InfoSphere BigInsights.

### **S3ToDRS**

If you see the job log for the S3ToDRS job, then a problem occurred while moving data from Amazon S3 to a relational database.

### **S3SchemaFileCreator**

If you see the job log for the S3SchemaFileCreator job, then a problem occurred while moving data to Amazon S3. A problem occurred when a schema file was being created that details how data is formatted in the Amazon S3 files.

**Note:** By default, the logs are maintained for the most recent hundred times that a particular job is run across all activities. After each job is run a hundred times, the logs are deleted from the InfoSphere Information Server engine tier repository. However, all of the logs for jobs related to an activity are saved in the *IS\_install\_path/Server/Projects/Name\_of\_DS\_project/DCLogs/name\_of\_activity* directory. For example, if an activity is moving five hundred tables, and the movement of the tenth table fails, you have to go to the *IS\_install\_path/Server/Projects/Name\_of\_DS\_project/DCLogs/name\_of\_activity* directory to view the job log for the tenth table. The Operations Console displays the job logs for only the last hundred instances that a job was run.

3. If you are unable to troubleshoot the issue by manually looking at the job logs or by using the Operations Console, then run ISALite to collect log information and other system configuration information. For more information about using ISALite, see [http://pic.dhe.ibm.com/infocenter/bigins/v2r1m1/topic/com.ibm.svg.im.iis.productization.iisinfsv.install.doc/topics/wsinst\\_install\\_verif\\_troublshng.html](http://pic.dhe.ibm.com/infocenter/bigins/v2r1m1/topic/com.ibm.svg.im.iis.productization.iisinfsv.install.doc/topics/wsinst_install_verif_troublshng.html).

## **What to do next**

You must periodically delete the schema files in the *IS\_install\_path/Server/Projects/Name\_of\_DS\_project* folder that start with the prefix DC\_. The schema files are used to run the InfoSphere DataStage jobs and are stored in the *IS\_install\_path/Server/Projects/Name\_of\_DS\_project* folder. This helps to maintain the performance of your system and prevents activities from failing. We recommend that you delete the schema files every month.

Before you delete the files, verify that no activities are running. If you delete schema files while activities are running, the activities might fail to run.

---

## **Troubleshooting general issues in InfoSphere Data Click**

These problems and workarounds are related to using InfoSphere Data Click.

Based on the message or symptom, complete the action specified.

Table 8. Troubleshooting InfoSphere Data Click

Message or symptom	Action
<p>After migrating from InfoSphere Data Click, Version 9.1.2, you are unable to use your blueprints that contain InfoSphere Data Click activities.</p>	<p>Blueprints that contain InfoSphere Data Click activities that were created in Version 9.1.2 are no longer supported in the current version of InfoSphere Data Click. Create new activities to perform the data movement that was defined in the blueprints that were created in Version 9.1.2.</p>
<p>When you move XML data that is stored in a DB2 database by running an InfoSphere Data Click activity, the data is moved as binary data in the Hive table.</p>	<p>To resolve this issue, change the XML type setting in the <code>odbc.ini</code> file by adding the following line to the file:</p> <pre>XMLDescribeType=-10</pre> <p>For example:</p> <pre>[DB2DSN] Driver=/opt/IBM/Information Server/Server/branded_odbc/lib/ VMdb225.so Description=DataDirect DB2 Wire Protocol Driver AddStringToCreateTable= AlternateID= Database=DB2 UDB (Remove for OS/390 and AS/400) DynamicSections=100 GrantAuthid=PUBLIC GrantExecute=1 IpAddress=DB2 server host IsolationLevel=CURSOR_STABILITY LogonID= Password= Package=DB2 package name PackageOwner= TcpPort=DB2 server port WithHold=1 XMLDescribeType=-10</pre> <p>The default directory where the <code>odbc.ini</code> file is stored is <code>/opt/IBM/InformationServer/Server/DSEngine/</code>.</p>
<p>You receive an error message that says you must create a new JDBC data connection that connects to a Hive table by using InfoSphere Metadata Asset Manager.</p>	<p>When you run an activity that is moving data from a relational database to InfoSphere BigInsights, a Hive table is automatically created using the connection information that was specified when you imported metadata about the Hive table by using the JDBC connector in InfoSphere Metadata Asset Manager. If the host name in the URL that was specified when you imported the Hive table does not match the host name that was specified when you used the HDFS bridge to import the target distributed file system, you must create a new JDBC data connection to a Hive table. The host name that you specify as part of the URL for the JDBC data connection must be identical to the host name that you specified when you created the data connection that connects to the target distributed file system.</p>

Table 8. Troubleshooting InfoSphere Data Click (continued)

Message or symptom	Action
<p>You create a filter, but it is not applied.</p>	<p>You must select the filter icon in order to apply the filter changes. If you select <b>Close</b>, then the filter is not applied. Verify that you selected the filter button in order to apply your changes.</p>
<p>You are running InfoSphere Data Click activities that are moving a large amount of data from a database with an ODBC data connection to an Amazon S3 file, and you receive the following error message:</p> <pre>Source_DRS_Connector,0:   ODBC function "SQLGetData"   reported: SQLSTATE =   07009: Native Error Code = 0:   Msg = [IBM(DataDirect OEM) ][ODBC SQL Server Legacy Driver ]Invalid Descriptor   Index (CC_OdbcInputS tream::get TotalSize, file CC_OdbcInput Stream.cpp,   line 354)</pre>	<p>You need to use a different ODBC driver that supports large object (LOB) columns. It is recommended that you use the DataDirect SQL Server Native Wire Protocol driver.</p>
<p>You are trying to move data from Amazon S3 by running an activity and you receive the following error in your job log:</p> <pre>Amazon_S3_Read,1: The number of rows that   is specified by the Record   limit property (100,000)   was reached.</pre>	<p>When the activity was created, the person that created the activity specified a maximum number of records that can be moved by the activity on the <b>Policies</b> pane. The value that the person set has been reached. Create a new activity and specify a higher number of records to read per file on the <b>Policies</b> pane in order to avoid this error message.</p>
<p>You are trying to move data to InfoSphere BigInsights and you receive a connection refused error when you try to connect while all pool resources are in use. This situation might occur because of too many concurrent users.</p>	<p>Increase the value of the <b>maxPoolSize</b> setting. The recommended setting is two times the maximum number of expected concurrent users. For additional information about this setting, see Liberty profile: Configuration elements in the server.xml file. For general information about tuning the Liberty profile, see Tuning the Liberty profile.</p>

Table 8. Troubleshooting InfoSphere Data Click (continued)

Message or symptom	Action
<p>You are trying to move data to InfoSphere BigInsights and you experience data flow that is slow, intermittent, or both. This situation might occur because the InfoSphere Information Server computer and the data source are far away from each other geographically.</p> <p>Example message:            Message ID: IIS-DSEE-USBP-00002            BigInsights_Write,1: java.io.            FileNotFoundException:            Pipe error writing to            http://9.113.141.44:8080/data            /controller/dfs/user            /biadmin/Teradata            offload_isadmin_            1385548146984/isadmin_            _charges/isadmin_charges.            part00001, BigInsights            Console may be unreachable.</p>	<p>Increase the value of the <b>Socket readTimeout</b> setting. The recommended value for this settings is readTimeout="10m". For additional information about this setting, see Liberty profile: Configuration elements in the server.xml file.</p>



---

## 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.

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](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. Because the information center uses XHTML, you can set display preferences in your browser. This 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 9. IBM resources*

Resource	Description and location
IBM Support Portal	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>
Software services	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>
My IBM	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>
Training and certification	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/training">http://www.ibm.com/training</a>
IBM representatives	You can contact an IBM representative to learn about solutions at <a href="http://www.ibm.com/connect/ibm/us/en/">www.ibm.com/connect/ibm/us/en/</a>



---

## Appendix C. Accessing the product documentation

Documentation is provided in a variety of formats: in the online IBM Knowledge Center, in an optional locally installed information center, and as PDF books. You can access the online or locally installed help directly from the product client interfaces.

IBM Knowledge Center is the best place to find the most up-to-date information for InfoSphere Information Server. IBM Knowledge 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 IBM Knowledge Center from the installed product or from a web browser.

### Accessing IBM Knowledge Center

There are various ways to access the online documentation:

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

- Type the address in a web browser, for example, when you are not logged in to the product.

Enter the following address to access all versions of InfoSphere Information Server documentation:

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ/
```

If you want to access a particular topic, specify the version number with the product identifier, the documentation plug-in name, and the topic path in the URL. For example, the URL for the 11.3 version of this topic is as follows. (The ⇒ symbol indicates a line continuation):

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/⇒  
com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html
```

**Tip:**

The knowledge center has a short URL as well:

```
http://ibm.biz/knowctr
```

To specify a short URL to a specific product page, version, or topic, use a hash character (#) between the short URL and the product identifier. For example, the short URL to all the InfoSphere Information Server documentation is the following URL:

```
http://ibm.biz/knowctr#SSZJPZ/
```

And, the short URL to the topic above to create a slightly shorter URL is the following URL (The ⇒ symbol indicates a line continuation):

```
http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/⇒  
common/accessingiidoc.html
```

## Changing help links to refer to locally installed documentation

IBM Knowledge Center contains the most up-to-date version of the documentation. However, you can install a local version of the documentation as an information center and configure your help links to point to it. A local information center is useful if your enterprise does not provide access to the internet.

Use the installation instructions that come with the information center installation package to install it on the computer of your choice. After you install and start the information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product F1 and help links refer to. (The `⇒` symbol indicates a line continuation):

### Windows

```
IS_install_path\ASBServer\bin\iisAdmin.bat -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

### AIX Linux

```
IS_install_path/ASBServer/bin/iisAdmin.sh -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

Where `<host>` is the name of the computer where the information center is installed and `<port>` is the port number for the information center. The default port number is 8888. For example, on a computer named `server1.example.com` that uses the default port, the URL value would be `http://server1.example.com:8888/help/topic/`.

## Obtaining PDF and hardcopy documentation

- The 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>.

---

## Notices and trademarks

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

### 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.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, 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.

## Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session or persistent cookies. If a product or component is not listed, that product or component does not use cookies.

Table 10. Use of cookies by InfoSphere Information Server products and components

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
Any (part of InfoSphere Information Server installation)	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
Any (part of InfoSphere Information Server installation)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Enhanced user usability</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled

Table 10. Use of cookies by InfoSphere Information Server products and components (continued)

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
InfoSphere DataStage	Big Data File stage	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• User name</li> <li>• Digital signature</li> <li>• Session ID</li> </ul>	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere DataStage	XML stage	Session	Internal identifiers	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage Operations Console	Session	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Data Click	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Data Quality Console		Session	No personally identifiable information	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server web console	<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	User name	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> </ul>	Cannot be disabled
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> <li>• Session</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• User name</li> <li>• Internal identifiers</li> <li>• State of the tree</li> </ul>	<ul style="list-style-type: none"> <li>• Session management</li> <li>• Authentication</li> <li>• Single sign-on configuration</li> </ul>	Cannot be disabled
InfoSphere Information Analyzer	Data Rules stage in the InfoSphere DataStage and QualityStage Designer client	Session	Session ID	Session management	Cannot be disabled

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at <http://www.ibm.com/privacy> and IBM's Online Privacy Statement at <http://www.ibm.com/privacy/details> the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at <http://www.ibm.com/software/info/product-privacy>.

## Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com)<sup>®</sup> 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<sup>Link</sup>, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS<sup>Link</sup> licensee of the United States Postal Service.

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



---

# Index

## Special characters

.dsx files 47

## A

activities 25  
activity  
  creating 19  
  moving data 19  
  status of request 24, 32  
assigning user roles 39

## C

client 22  
components  
  viewer 22  
connectivity  
  express import 11  
  InfoSphere Data Click sources and target 11  
  InfoSphere Metadata Asset Manager importing assets 11  
credentials 15  
customer support  
  contacting 59

## D

Data Click job templates 47  
Data Click jobs  
  monitoring 24  
data connection 42

## E

error messages 51, 52

## H

HDFS bridge 42  
  configuration file 29  
  connections 29  
Hive table  
  configuring 36  
  connections 28  
  creating 36  
  type 36

## I

InfoSphere BigInsights  
  connection to InfoSphere Data Click 29  
InfoSphere Data Click  
  connection to InfoSphere BigInsights 29  
  servicing 51

InfoSphere Data Click (*continued*)  
  troubleshooting activities 51  
InfoSphere DataStage projects  
  creating 46  
integration  
  between InfoSphere Data Click and InfoSphere BigInsights 9

## J

JDBC connector 28, 42  
JDBC driver  
  driver configuration file 28  
job instances 24, 32  
job templates 47

## L

Launchpad 11  
LDAP configuration  
  assigning user roles 39  
legal notices 63  
limitations 51, 52

## M

moving data 25

## O

Operations Console 32  
  Data Click jobs 32  
  workload management settings 49  
osh schema files  
  creating 37

## P

product accessibility  
  accessibility 57  
product documentation  
  accessing 61

## R

run  
  activities 22

## S

security 15  
software services  
  contacting 59  
support  
  customer 59

## T

trademarks  
  list of 63  
troubleshooting  
  activities 51

## U

users  
  creating 15

## V

view  
  activities 22  
viewers 22







Printed in USA

SC19-4331-00

