

IBM InfoSphere Information Server
Version 8 Release 7

*Connectivity Guide for IBM InfoSphere
Change Data Capture*



IBM InfoSphere Information Server
Version 8 Release 7

*Connectivity Guide for IBM InfoSphere
Change Data Capture*



Note

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 53.

Contents

Product accessibility	1	Ending the subscription	31
Accessing product documentation	3	Example: Applying change data by using a CDC Transaction stage	33
Overview	5	Setting up source and target tables for the example	33
IBM InfoSphere Change Data Capture overview	5	Setting up replication	34
CDC Transaction stage overview.	5	Generating a template for the job	37
Flow of change data in a CDC Transaction stage job	6	Importing and configuring the job.	38
Limitations	7	Starting the subscription and running the job	41
Software prerequisites and supported installation topologies	8	Writing change data to the target database	42
Software prerequisites	8	Troubleshooting.	43
Supported installation topologies	9	Value of DM_TIMESTAMP column in the target table is incorrect.	43
Configuring the software for change data capture	13	Job fails to start automatically	43
Configuring InfoSphere CDC	13	Job fails while processing large records	44
Setting up the ODBC data source for bookmark information	14	Job fails with authentication failure message	45
Setting up replication in the InfoSphere CDC Management Console	15	Job fails with bookmark table missing message	45
Developing a job that processes replicated data from InfoSphere CDC	17	Job fails with connection error message	45
Generating a template for the job	17	Job fails with multiple bookmark links or no bookmark link error message	46
Setting up target tables for the job.	18	Job fails with table name for subscription not specified on link error message.	46
Record format for change data records	18	Job fails with no response from server error message	46
Bookmark record format	21	Job fails with ODBC access failure message.	47
Configuring the CDC Transaction stage to read change data from InfoSphere CDC.	22	Output data is truncated	47
Supported data type conversions	23	Subscription stops while still in "Starting" state	48
Adding and configuring additional stages	24	Frequently asked questions about the CDC Transaction stage.	49
Configuring the target database stage to write change data	25	Links to non-IBM Web sites.	51
Compiling the job	26	Notices and trademarks	53
Starting the subscription and running the job.	27	Contacting IBM	57
Monitoring the status of the job and the subscription.	29	Index	59

Product accessibility

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

The IBM InfoSphere® Information Server product modules and user interfaces are not fully accessible. The installation program installs the following product modules and components:

- IBM InfoSphere Business Glossary
- IBM InfoSphere Business Glossary Anywhere
- IBM InfoSphere DataStage®
- IBM InfoSphere FastTrack
- IBM InfoSphere Information Analyzer
- IBM InfoSphere Information Services Director
- IBM InfoSphere Metadata Workbench
- IBM InfoSphere QualityStage™

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

Accessible documentation

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

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.

Accessing product documentation

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

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

Accessing the information center

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

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

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

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

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

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

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

Obtaining PDF and hardcopy documentation

- A subset of the PDF file books are available through the InfoSphere Information Server software installer and the distribution media. The other PDF file books are available online and can be accessed from this support document: `https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1`.
- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at `http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss`.

Providing feedback about the documentation

You can send your comments about documentation in the following ways:

- Online reader comment form: `www.ibm.com/software/data/rcf/`
- E-mail: `comments@us.ibm.com`

Overview

The CDC Transaction stage integrates the replication capabilities that are provided by IBM InfoSphere Change Data Capture (InfoSphere CDC) with the ETL capabilities that are provided by IBM InfoSphere DataStage.

You can use these products together to perform real time, continuous replication with guaranteed delivery and transactional integrity in the event of failure.

Before using the CDC Transaction stage, you must first install InfoSphere CDC.

IBM InfoSphere Change Data Capture overview

IBM InfoSphere Change Data Capture (InfoSphere CDC) is a replication solution that captures database changes as they happen and delivers the changes to target databases, message queues, or an ETL solution, such as IBM InfoSphere DataStage.

The unit of replication within InfoSphere CDC is called a subscription. A subscription contains mapping details that specify how data in a source data store is applied to a target data store. A subscription can be scheduled to run at a specific time and stop when the updates to the target data store are complete, or a subscription can be set to run on a continuous basis to apply updates to the target data store as the changes occur. Subscriptions are created, configured, run, and monitored in the IBM InfoSphere Change Data Capture Management Console.

When a subscription runs, InfoSphere CDC captures changes on the source database. InfoSphere CDC delivers the change data to the target, and stores sync point information in a bookmark table in the target database. InfoSphere CDC uses the bookmark information to monitor progress of the InfoSphere DataStage job and determine restart points in the event of a failure. Synchronous updates to the bookmark information ensure that no updates are lost even if InfoSphere DataStage is not able to process all available updates. The bookmark information can also be used to determine the log retention policy for the source database.

InfoSphere CDC supports the following connection methods for writing change data to InfoSphere DataStage:

Direct connect

Uses TCP/IP as the transport protocol to stream data from InfoSphere CDC to InfoSphere DataStage. When you use this method, you can configure options that cause the job to start automatically when you run the subscription.

Flat File

Uses a file system to deliver source changes to InfoSphere DataStage.

This documentation describes how to use the direct connect method. For more information about using the flat file method, see the documentation for InfoSphere CDC. The documentation for InfoSphere CDC also contains additional details about using the direct connect method.

CDC Transaction stage overview

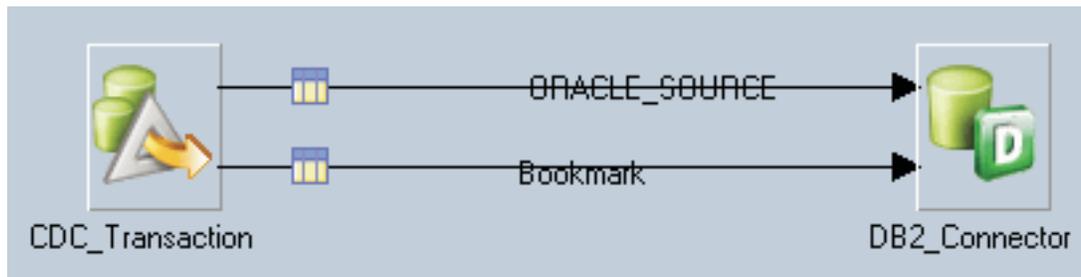
You can use the CDC Transaction stage in an IBM InfoSphere DataStage job to read data that is captured by IBM InfoSphere Change Data Capture (InfoSphere CDC) and apply the change data to a target database.

To use the CDC Transaction stage, you first set up replication in the InfoSphere CDC Management Console, and then use the InfoSphere CDC Management Console to generate a template for the InfoSphere DataStage job. You import the template into InfoSphere DataStage and modify the job, as required.

A basic CDC Transaction stage job includes the following stages and links:

- A CDC Transaction stage that specifies details about accessing your installation of InfoSphere CDC.
- The following output links from the CDC Transaction stage:
 - An output link for each table in the subscription. These output links transfer change data from InfoSphere CDC to another stage or to a target database. If a subscription contains more than one table, the changes for each table can follow a different path through the job.
 - A bookmark link that transfers bookmark information to a bookmark table in the target database.
- Additional stages, as required, that process the data.
- A database connector stage that specifies details about the target database.

For example, the following basic job contains a CDC Transaction stage, an output link, a bookmark link, and a DB2 connector stage that defines the details of the target database where the change data is applied:



A more complex job might include additional stages that process the data before it is delivered to the target database. For example, you can create jobs that perform the following types of actions:

- Transform change data by using prebuilt stages that are provided by InfoSphere DataStage
- Integrate data by performing lookups
- Apply custom business logic to the change data before delivering the data to a target database

Flow of change data in a CDC Transaction stage job

To understand how the CDC Transaction stage works, you should understand how data flows from the source database to the target database.

Figure 1 on page 7 shows how data flows when IBM InfoSphere Change Data Capture (InfoSphere CDC) captures changes at a source database and uses IBM InfoSphere DataStage to deliver the change data to a target database.

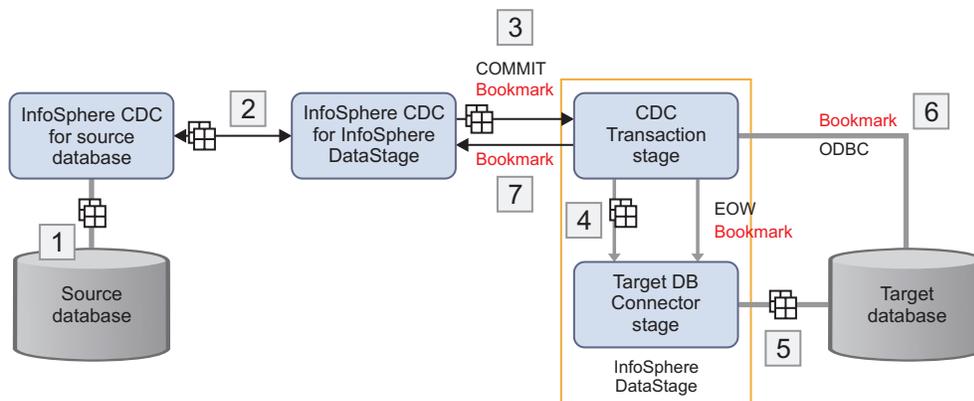


Figure 1. Data flow in a CDC Transaction stage job

1. On the computer where the source database is installed, the InfoSphere CDC service for the database monitors and captures the change.
2. InfoSphere CDC transfers the change data according to the replication definition.
3. The InfoSphere CDC for InfoSphere DataStage server sends data to the CDC Transaction stage through a TCP/IP session that is created when replication begins. Periodically, the InfoSphere CDC for InfoSphere DataStage server also sends a COMMIT message (along with bookmark information) to mark the transaction boundary in the captured log.
4. In the InfoSphere DataStage job, the data flows over links from the CDC Transaction stage to the target database connector stage. The bookmark information is sent over a bookmark link. For each COMMIT message sent by the InfoSphere CDC for InfoSphere DataStage server, the CDC Transaction stage creates end-of-wave (EOW) markers that are sent on all output links to the target database connector stage.
5. The target database connector stage connects to the target database and sends data over the session. When the target database connector stage receives an end-of-wave marker on all input links, it writes bookmark information to a bookmark table and then commits the transaction to the target database.
6. Periodically, the InfoSphere CDC for InfoSphere DataStage server requests bookmark information from a bookmark table on the target database. In response to the request, the CDC Transaction stage fetches the bookmark information through ODBC and returns it to the InfoSphere CDC for InfoSphere DataStage server.
7. The InfoSphere CDC for InfoSphere DataStage server receives the bookmark information, which is used for the following purposes:
 - To determine the starting point in the transaction log where changes are read when replication begins. (The starting point in the transaction log is the ending point from the previous replication, if the replication ended successfully.)
 - To determine if the existing transaction log can be cleaned up.

The bookmark is committed synchronously with the data, so even if the job fails, the bookmark information and the written data are consistent. If the job fails, replication begins at the point that is indicated by the bookmark, and there is no loss of data.

Limitations

Some limitations apply when you use the CDC Transactions stage to read change data from IBM InfoSphere Change Data Capture (InfoSphere CDC) and apply the change data to a target database.

LOB data type truncation

Large object (LOB) data is supported, but there is a maximum LOB size above which the content is truncated. You configure LOB truncation size in the InfoSphere DataStage Properties window of the IBM InfoSphere Change Data Capture Management Console when you configure the subscription.

Define a truncation size for LOB data that ensures that the total length of the output record does not exceed the size of the transport block that can pass between the stages of an IBM InfoSphere DataStage job. The size of the transport block is specified by the `APT_DEFAULT_TRANSPORT_BLOCK_SIZE` environment variable. By default, the value of this environment variable is 128 KB. You can increase the size of the transport block. However, increasing this size affects all links and might increase overall memory and disk usage.

For more information about specifying the LOB truncation size, see the InfoSphere CDC documentation.

Transactional integrity

If either of the following statements is true, the overall transactional integrity of the job is not guaranteed:

- Reject links are used on the target database connector stage of the job.
- The job does not process the data sequentially.

Encryption of sensitive data

Other than connection keys, data transferred between InfoSphere CDC and the CDC Transaction stage is unencrypted. SSL connections are not supported. To provide additional data security, use VPN or SSH tunneling.

Instance recreation

After creating an instance of InfoSphere CDC for InfoSphere DataStage, you cannot delete and recreate the instance without making subscriptions that use the instance unusable. By deleting the instance, you destroy the relationship between source and target configuration for the subscription. If you destroy this relationship, you must delete the subscription and start again.

DDL awareness

Some versions of InfoSphere CDC do not support DDL awareness. Therefore, if a source table is dropped during replication, no error occurs for some InfoSphere CDC sources. The subscription and corresponding CDC Transaction stage job might continue to run even though the source table no longer exists.

Software prerequisites and supported installation topologies

Before using the CDC Transaction stage in an IBM InfoSphere DataStage job, you must install and configure the prerequisite IBM InfoSphere Change Data Capture (InfoSphere CDC) software. When you plan your installation, make sure that you use a supported installation topology.

Software prerequisites

You must install and configure the required IBM InfoSphere Change Data Capture (InfoSphere CDC) components.

For a list of supported software versions, see www.ibm.com/support/docview.wss?uid=swg27020317.

IBM InfoSphere Change Data Capture Access Server

Controls all non-command line access to the replication environment.

IBM InfoSphere Change Data Capture Management Console

Provides a user interface for configuring and monitoring replication.

IBM InfoSphere Change Data Capture for the source database

Monitors source tables for changes and sends change data to the replication engine.

InfoSphere CDC provides components that are compatible with each supported database. You install the InfoSphere CDC component that is compatible with the source database that you have installed. For example, if you are using IBM DB2® Database for Linux, UNIX, and Windows as the source database, you install InfoSphere CDC for DB2 for Linux, UNIX, and Windows.

The log retention policy for the source database must be set to retain logs.

IBM InfoSphere Change Data Capture for IBM InfoSphere DataStage

Processes changes delivered from InfoSphere CDC that can be used by InfoSphere DataStage jobs.

If you plan to use the auto-start feature to start InfoSphere DataStage jobs automatically, install IBM InfoSphere Information Server before you install InfoSphere CDC. Installing the software components in this order ensures that the InfoSphere CDC for InfoSphere DataStage server has access to the **dsjob** command during installation.

For more information about installing InfoSphere CDC, see the InfoSphere CDC documentation.

Supported installation topologies

You can install InfoSphere CDC components on the same computer, or you can design a topology that includes several different computers. The term computers refers to separate physical servers, logical partition (LPARs), or virtual machines.

Figure 2 on page 10 illustrates a topology with components that are installed as separately as possible on different computers. In this topology, InfoSphere CDC for InfoSphere DataStage is not installed on the same computer as InfoSphere DataStage. Although this topology illustrates what is possible, it has limitations because the auto-start feature is unavailable, and InfoSphere DataStage jobs must be started manually.

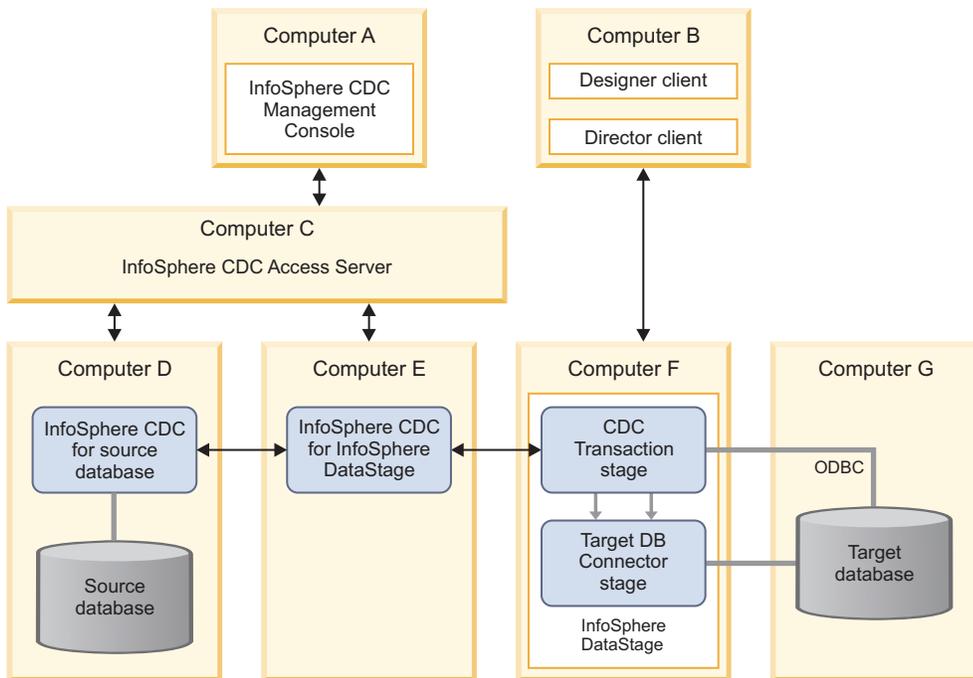


Figure 2. Topology with separate components installed on many different computers

In Figure 2, the components are installed on the following computers. Unless otherwise noted, solid black arrows in the diagram represent TCP/IP connections. Solid gray lines represent database connections. Solid gray arrows represent InfoSphere DataStage links.

Computer A

InfoSphere CDC Management Console

Computer B

InfoSphere DataStage and QualityStage Designer client
 InfoSphere DataStage and QualityStage Director client

Computer C

InfoSphere CDC Access Server

Computer D

InfoSphere CDC for the source database
 Source database

Computer E

InfoSphere CDC for InfoSphere DataStage

Computer F

InfoSphere DataStage

Computer G

Target database

Figure 3 on page 11 illustrates a topology with components that are installed on shared computers. The auto-start feature is available, and the coexistence of client components on the same computer simplifies subscription and job monitoring.

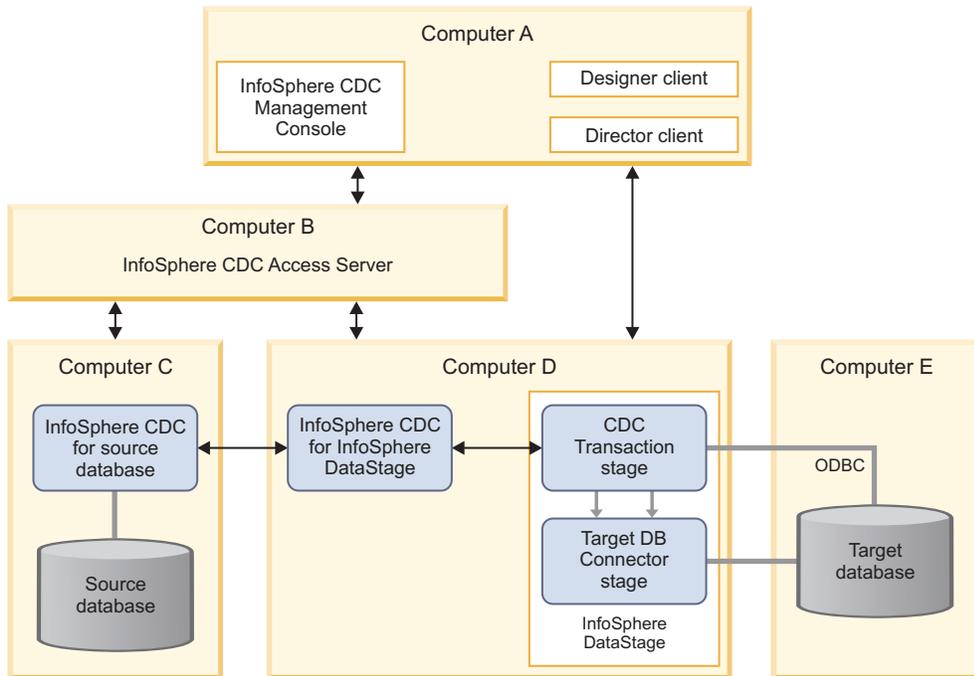


Figure 3. Topology with components installed on shared computers

In Figure 3, the components are installed on the following computers. Unless otherwise noted, solid black arrows in the diagram represent TCP/IP connections. Solid gray lines represent database connections. Solid gray arrows represent InfoSphere DataStage links.

Computer A

- InfoSphere CDC Management Console
- InfoSphere DataStage and QualityStage Designer client
- InfoSphere DataStage and QualityStage Director client

Computer B

- InfoSphere CDC Access Server

Computer C

- InfoSphere CDC for the source database
- Source database

Computer D

- InfoSphere CDC for InfoSphere DataStage
- InfoSphere DataStage

Computer E

- Target database

If required, you can install all the software on a single computer.

Configuring the software for change data capture

After installing InfoSphere CDC, you configure the software for change data capture before developing a job that uses the CDC Transaction stage.

Configuring InfoSphere CDC

After installing InfoSphere CDC, the installation program starts a configuration tool. You use the configuration tool to add and configure InfoSphere CDC instances for your environment. You must configure InfoSphere CDC for the source database and for the target (InfoSphere DataStage) before you can start replication.

Before you begin

Install InfoSphere CDC. For more information, see the InfoSphere CDC documentation.

About this task

This procedure describes the general steps that you must perform to configure InfoSphere CDC to work with the CDC Transaction stage. For more information about configuring InfoSphere CDC, see the InfoSphere CDC documentation that describes how to add and configure new instances.

Tip: If you plan to configure your job to start automatically when the subscription starts, ensure that the PATH environment variable contains the path to the dsjob executable file before you add the new CDC instances.

Procedure

1. Configure InfoSphere CDC for the source database. For example, if you are using IBM DB2 Database for Linux, UNIX, and Windows as a source database, you configure InfoSphere CDC for DB2 for Linux, UNIX, and Windows.
 - a. On the computer where InfoSphere CDC for the source database is installed, start the InfoSphere CDC configuration tool (if it is not already running).
 - b. Add a new instance of InfoSphere CDC.
 - c. Specify configuration details about the instance, including the following settings:
 - For the **Server Port**, specify the port number that other servers will use to communicate with this instance of InfoSphere CDC for the source database. Write down the port number that you specify. You provide this port number when you specify access parameters for this data store in the Access Manager perspective in InfoSphere CDC Management Console.
 - Specify connection details for the database that contains the tables that you want to replicate.
 - d. Start the instance.
2. Configure InfoSphere CDC for InfoSphere DataStage.
 - a. On the computer where InfoSphere CDC for InfoSphere DataStage is installed, start the InfoSphere CDC configuration tool (if it is not already running).
 - b. Add a new instance of InfoSphere CDC.

Important: After creating this instance, you cannot delete and recreate it without making subscriptions that use the instance unusable.

- c. Specify configuration details about the instance, including the following settings:
 - For the **Server Port**, specify the port number that other servers will use to communicate with this instance of InfoSphere CDC for InfoSphere DataStage. Write down the port number that

you specify. You provide this port number when you specify access parameters for InfoSphere DataStage in the Access Manager perspective in InfoSphere CDC Management Console.

- Specify a password for the user named `tsuser`. Write down the password that you specify. You provide the `tsuser` ID and password when you specify connection parameters for InfoSphere DataStage in the Access Manager perspective in the InfoSphere CDC Management Console.

What to do next

After configuring InfoSphere CDC, you set up the ODBC data source on the computer where InfoSphere DataStage is installed.

Setting up the ODBC data source for bookmark information

Before using the CDC Transaction stage, you must set up a new ODBC data source name (DSN) on the computer where InfoSphere DataStage is installed. The CDC Transaction stage uses ODBC to return bookmark information from the target database. You specify the ODBC DSN as a stage property when you configure the InfoSphere DataStage job.

Procedure

On the computer where InfoSphere DataStage is installed, add a new system data source name (DSN) for ODBC access to the target database.

- On a Windows computer, use the ODBC Data Source Administrator tool to add a new System DSN. To ensure that the InfoSphere DataStage process can find the DSN, you must add a System DSN, not a User DSN.
- On a Linux or UNIX computer, ensure that the `ODBCINI` environment variable setting specifies the correct location of the DSN configuration file (this file is typically named `.odbc.ini`).

What to do next

After setting up the ODBC data source, you set up replication in the InfoSphere CDC Management Console.

Setting up replication in the InfoSphere CDC Management Console

You set up replication in the InfoSphere CDC Management Console to define how change data is captured and sent to the target. To set up replication, you add data stores for the source and target, add a new subscription, and map one or more tables to the target system.

Before you begin

Configure the software for change data capture.

About this task

This procedure describes the general steps that you must perform to set up replication in the InfoSphere CDC Management Console. For more information about setting up replication, see the InfoSphere CDC documentation. (When searching for information about adding data stores, note that the InfoSphere CDC documentation spells data store as one word, `datastore`.)

Procedure

1. Log on to the InfoSphere CDC Management Console. Enter the connection information that you specified for InfoSphere CDC Access Server during the installation.
2. In the Access Manager perspective, add data stores for the source and the target.
 - a. In the **Datastore Management** tab, add a data store for the source database. Specify details about the data store, including the following settings:
 - For the server name, enter the host name of the computer where you installed InfoSphere CDC for the source database.
 - For the port number, enter the number that you specified when you configured InfoSphere CDC for the source database.
 - b. Add a data store for the target, InfoSphere DataStage. Specify details about the target, including the following settings:
 - For the server name, enter the host name of the computer where you installed InfoSphere CDC for InfoSphere DataStage.
 - For the port number, enter the number that you specified when you configured InfoSphere CDC for InfoSphere DataStage.
 - For the connection parameters, specify the credentials for `tsuser`, the user that you created when you configured InfoSphere CDC for InfoSphere DataStage.
 - c. Optional: In the **User Management** tab, add a new user. You can add a new user or use an existing one.
 - d. In the **Connection Management** tab, assign a user to each data store.
3. In the Configuration perspective of the InfoSphere CDC Management Console, add a new subscription. A subscription contains mapping details that specify how data in a source data store is applied to a target data store.
 - a. For the source, select the data store that contains the tables that you want to replicate.
 - b. For the target, select the data store that you created for InfoSphere DataStage.
4. In the Configuration perspective of the InfoSphere CDC Management Console, use the Map Tables wizard to map one or more tables to the target, InfoSphere DataStage. When you run the wizard, make sure that you select the following options.
 - a. In the Select Map Typing page, select **InfoSphere DataStage**.

- b. In the Select InfoSphere DataStage Connection Method page, select **Direct Connect**.
- c. In the Select Source Tables page, select the tables that you want to replicate.
- d. In the InfoSphere DataStage Direct Connect page, select the type of record format to use for the change data.

Single Record

Both the "before image" (the record before the update) and the "after image" (the record after the update) are sent in a single record.

Multiple Records

The "before image" and the "after image" are sent in separate records.

After you have mapped a table to a target, the table is no longer available for another mapping within the same subscription.

5. In the Configuration perspective of the InfoSphere CDC Management Console, specify InfoSphere DataStage properties.
 - a. In the **Subscriptions** tab, right-click the subscription, and select **InfoSphere DataStage > InfoSphere DataStage Properties**. The subscription must be open for editing when you define these properties, or your changes are not saved.
 - b. In the Direct Connect area, specify the project name, job name, and connection key for the InfoSphere DataStage job. The connection key ensures that only the job with the correct connection key information is used by InfoSphere DataStage when receiving change data from InfoSphere CDC. You specify this connection key as a stage property when you configure the InfoSphere DataStage job.
 - c. Optional: If you want to configure the job to start automatically when the subscription starts, select **Auto-start InfoSphere DataStage Job**.

Important: For the auto-start function to work correctly, InfoSphere CDC must be able to load and execute the **dsjob** command and the dependent shared library. The path to this library must be specified in the library path environment variable on the computer where InfoSphere CDC for InfoSphere DataStage is installed.

On UNIX or Linux, you can view and modify the setting for this environment variable in the `dsenv` file. After setting this environment variable, source the `dsenv` file and then start the InfoSphere CDC for InfoSphere DataStage server to set these environment variables for the server. For example, issue one of the following commands to source the `dsenv` file:

```
$ . dsenv
```

or

```
$ source dsenv
```

Then issue the following command to start the InfoSphere CDC for InfoSphere DataStage server:

```
$ dmsts64 -I <instance name>
```

The environment variables for the process, `dmsts64`, are set as specified in the `dsenv` file.

Developing a job that processes replicated data from InfoSphere CDC

After setting up replication in the InfoSphere CDC Management Console, you generate a template for the InfoSphere DataStage job, import the template into InfoSphere DataStage and QualityStage Designer, and configure the stages of the job.

Before you begin

- Make sure the prerequisite software is installed.
- Configure the software for change data capture.
- Set up replication in the InfoSphere CDC Management Console.

About this task

An InfoSphere DataStage job that processes change data from InfoSphere CDC includes the following stages:

- A CDC Transaction stage that specifies details about InfoSphere CDC.
- A database connector stage that specifies details about the target database.
- Zero or more stages that exist between the CDC Transaction stage and the database connector stage. These additional stages process the change data before delivering it to the final database connector stage.

To ensure transactional integrity, all stages in the job must run sequentially. Reject links are not supported.

Perform the following tasks to develop a job that uses a CDC Transaction stage.

Generating a template for the job

After creating a subscription in the InfoSphere CDC Management Console, you generate a template for the InfoSphere DataStage job. You can use the template as a starting point for the job and insert your own logic, or you can cut and paste parts of the job, such as the CDC Transaction stage, into another job.

Before you begin

Set up replication in InfoSphere CDC Management Console.

Procedure

1. Log on to the InfoSphere CDC Management Console.
2. Generate the template for the job:
 - a. In the **Subscriptions** tab of the Configuration perspective, right-click the subscription, and click **InfoSphere DataStage > Generate InfoSphere DataStage Job Definition**. This option is available only after the subscription is correctly set up.
 - b. In the dialog box that is displayed, browse to select the location for the .dsx file, and click **Save**.
3. Copy the .dsx file to a location that is accessible by the computer where InfoSphere DataStage and QualityStage Designer is installed.

What to do next

Set up target tables for the job.

Setting up target tables for the job

Before configuring the InfoSphere DataStage job, you set up tables in the target database to hold the change data and bookmark information. The target database must contain a unique bookmark table for each subscription.

Before you begin

You must be authorized to create tables in the target database.

Procedure

1. Connect to the target database.
2. Create the target tables for storing the change data records. Use the appropriate format for either single record format or multiple record format. For more information, see "Record format for change data records."

To determine the format of the target tables, also consider the design of the job and the transformations that need to be done between the CDC Transaction stage and the target database connector stage. The change data that is output by the CDC Transaction stage includes the before and after images of the data, along with control columns. The control columns, which have the prefix `DM_`, provide additional details about the change data, such as when the change occurred and the type of operation that was performed. You can use these control columns to transform the data in subsequent processing stages, or you can simply drop the columns if they are not needed.

3. Create the bookmark table. Use the appropriate format for bookmark records. For more information, see "Bookmark record format" on page 21.

What to do next

Configure the CDC Transaction stage.

Record format for change data records

InfoSphere CDC delivers change data to the CDC Transaction stage in either a single record format or a multiple record format. You select the type of record format to use when you set up the subscription.

Single record format

When single record format is specified, both the "before image" (the record before the update) and the "after image" (the record after the update) are passed in a single record.

The following example illustrates the format of the change data that is inserted in the target table when single record format is specified. In this example, the following SQL statements are issued on the source database:

```
INSERT INTO T01 VALUES(1, 'John')
UPDATE T01 SET NAME='Jack' WHERE ID=1
DELETE T01 WHERE ID=1
COMMIT
```

After these statements execute and the job runs, each record in the target table contains the "before image" and "after image" of the data.

Table 1. Example of change data that has a single record format

DM_OPERATION_TYPE	BEFORE_ID	BEFORE_NAME	ID	NAME
T			1	'John'
U	1	'John'	1	'Jack'
D	1	'Jack'		

In this example, the table contains a record for each operation that was performed. The DM_OPERATION_TYPE column indicates the type of operation that was performed on the source table. The BEFORE_ID and BEFORE_NAME columns contain the data before the change, and the ID and NAME columns contain the data after the change.

Multiple record format

When multiple record format is specified, the "before image" and "after image" are passed in separate records.

The following example illustrates the data that is inserted in the target table when multiple record format is specified. In this example, the following SQL statements are issued on the source database:

```
INSERT INTO T01 VALUES(1, 'John')
UPDATE T01 SET NAME='Jack' WHERE ID=1
DELETE T01 WHERE ID=1
COMMIT
```

After these statements execute and the job runs, the target table contains separate records for the "before image" and "after image" of the data.

Table 2. Example of change data that has a multiple record format

DM_OPERATION_TYPE	ID	NAME
'I'	1	'John'
'B'	1	'John'
'A'	1	'Jack'
'D'	1	'Jack'

In this example, the table contains one record for the INSERT operation that was performed, two records for the UPDATE operation, and one record for the DELETE operation. The DM_OPERATION_TYPE column indicates the type of operation that was performed on the source table. If the operation was an update, the DM_OPERATION_TYPE column contains either 'B' to indicate that the record contains the "before image" or 'A' to indicate that the record contains the "after image."

Column definition for change data records

Change data records include the following columns. The SQL types in this table are InfoSphere DataStage data types. The compatible native data types on the target database might differ depending on the target stage type or target database version. For example, some databases use the datetime data type to store data that InfoSphere DataStage stores in the Timestamp field. For more information, see the documentation for the target stage or the target database.

Table 3. Column definition for change data records

Column name	Key	SQL type	Length	Nullable
DM_SORTKEY	Yes	Numeric	20	Yes
DM_OPERATION_TYPE	Yes	Char	1	Yes
DM_TIMESTAMP	No	Timestamp	Not applicable	Yes
DM_TXID	No	Numeric	24	Yes
DM_USER	No	NVarChar	30	Yes
BEFORE_COL_NAME ^{1,2}		Derived from source database	Derived from source database	Yes
COL_NAME ²		Derived from source database	Derived from source database	Yes

Table 3. Column definition for change data records (continued)

Column name	Key	SQL type	Length	Nullable
Table notes:				
1. This column is used only with the single record format. Omit this column definition for multiple record format.				
2. <i>COL_NAME</i> is the name of the column in the source table. The change data record includes a column with this name for each column in the source table. For example, if the source table contains two columns, DEPT and DEPTNO, the change data record for single record format includes the following columns:				
<ul style="list-style-type: none"> • BEFORE_DEPT • BEFORE_DEPTNO • DEPT • DEPTNO 				

The record format includes the following columns:

DM_SORTKEY

Contains an unsigned 64 bit integer, starting at 1, that is incremented for each output row. When multiple record format is specified, the value of this column is the same for both the "before image" and "after image." The CDC Transaction stage has its own counter and counts from 1 at every job run. When the value of this column reaches the maximum value for an unsigned, 64 bit integer, a warning is raised in the job log, and the value is set to 0.

DM_OPERATION_TYPE

Contains a single character that indicates the type of operation:

- I** Insert
- D** Delete
- U** Update (in single record format)
- A** "After image" of an update (in multiple record format)
- B** "Before image" of an update (in multiple record format)

DM_TIMESTAMP

Contains the date and time of when the change was made on the source table. This column contains the value from the &TIMESTAMP (Record Modification Time) journal control field. By default, the value of DM_TIMESTAMP is changed to UTC timezone before the InfoSphere CDC for InfoSphere DataStage server passes the value to the target table. You can change this behavior by adding a system parameter to the InfoSphere CDC for InfoSphere DataStage data store. The system parameter must have the parameter name ds_output_timestamp_utc and the value false. For more information about adding a system parameter to a data store, see the InfoSphere CDC documentation.

DM_TXID

Contains the transaction ID of the transaction with the update. This column contains the value from the &CCID (Commit Cycle ID) journal control field.

DM_USER

Contains the user ID that changed the source table. This column contains the value from the &USER (Record Modification User) journal control field.

BEFORE_COL_NAME

For single record format, contains the "before image" of the data. If the value of DM_OPERATION_TYPE is I (for Insert), this column contains NULL.

COL_NAME

For single record format, contains the "after image" of the data. If the value of DM_OPERATION_TYPE is D (for Delete), this column contains NULL.

For multiple record format, contains the "before image" or "after image" of the data. The value of the DM_OPERATION_TYPE column indicates whether this column contains the "before image" or "after image."

For more information about journal control fields, see the InfoSphere CDC documentation.

Bookmark record format

The bookmark table in the target database stores sync point information. InfoSphere CDC uses the bookmark information to monitor progress of the InfoSphere DataStage job and determine restart points in the event of a failure. The target database must contain a unique bookmark table for each subscription.

The bookmark table consists of two columns: An ID that is used to uniquely identify the record, and the bookmark block.

Table 4. Column definition for bookmark record

Column name	Key	SQL type	Length	Nullable
DM_KEY	Yes	Smallint	Not applicable	No
DM_BOOKMARK	No	VarChar	1024	No

The record format includes the following columns:

DM_KEY

An ID that uniquely identifies the bookmark block.

DM_BOOKMARK

A variable-length string with a maximum length of 1024 bytes. When the bookmark data is longer than 1024 bytes, the CDC Transaction stage splits the data into 1024 byte blocks. The blocks are ordered by the DM_KEY, which begins at 1. An end of bookmark record is inserted at the end of all the blocks.

Example

The following example demonstrates the contents of a bookmark record.

```
DM_KEY DM_BOOKMARK
-----
1      MIRROR;JOURNAL;000301000000000F64669400000000F64669400000000F646694
2
```

The following example demonstrates the contents of a bookmark record when the bookmark is split into more than one block:

```
DM_KEY DM_BOOKMARK
-----
1      MIRROR;JOURNAL;000000....000 (1024 chars)
2      000000.....000 (1024 chars)
3      000000.....00FF ( 952 chars)
4                                     (blank to mark the last record)
```

Important: The actual format of the DM_BOOKMARK string varies depending on the data source and other InfoSphere CDC configuration details. The strings that are shown in these examples are used only to demonstrate the format of the bookmark record.

Configuring the CDC Transaction stage to read change data from InfoSphere CDC

The CDC Transaction stage specifies details for connecting to and reading change data from InfoSphere CDC. When you generate the template job, some of the stage properties are set to values that are accurate for the subscription. However, you must update the job to specify settings that were not known when the job was generated.

Before you begin

Generate a template for the job.

Set up target tables for the job.

Procedure

1. In the Designer client, click **Import > DataStage components**, and specify the path to the .dsx file that you generated. Click **OK**.
2. Open the job that you imported.
3. Configure the CDC Transaction stage.
 - a. On the Job canvas, double-click the CDC Transaction stage.
 - b. In the upper left corner of the stage editor, verify that the stage is selected in the diagram.
 - c. In the **Properties** tab, update the following ODBC properties. These properties are used to retrieve bookmark information.

Bookmark DSN

Enter the ODBC DSN that you created for bookmark information when you configured the software for change data capture. To select from a list of available DSNs, click the **Bookmark DSN** button.

ODBC user name

Enter the user name that you specified for the bookmark DSN.

ODBC password

Enter the password for the ODBC user that you specified for the bookmark DSN.

Tip: Click the **Test** link to test the ODBC connection. Clicking this link does not test the connection to the InfoSphere CDC for InfoSphere DataStage server.

The values for other CDC Transaction stage properties are already set based on properties that are specified in the subscription.

4. Configure the output links for the CDC Transaction stage. The CDC Transaction stage includes an output link for each table in the subscription. The order of output links is not important.
 - a. In the upper left corner of the stage editor, select the output link for a table.
 - b. In the **Properties** tab, verify that the **Table name** field contains the name of the table that you want to replicate. Typically, you accept the value that is set for this field. If you need to change this value, specify a fully qualified table name.
 - c. In the **Columns** tab, verify that the record format matches the format that you expect.
 - d. Repeat these steps for each output link that represents a table.
5. Configure the bookmark output link. A CDC Transaction stage job can have only one bookmark link.
 - a. In the upper left corner of the stage editor, select the bookmark link.
 - b. In the **Properties** tab, in the **Table name** field, replace the template bookmark name, BOOKMARKTABLE, with the real name of the bookmark table. The target database must contain a unique bookmark table for each subscription. The table name can be specified with or without the schema name. If the table name is not qualified with the schema name, the schema name is

implicitly determined by the ODBC driver or the target database. For more information about how the table name is implicitly qualified, see the documentation for ODBC or your database.

- c. Optional: In the **Columns** tab, view the record format of the bookmark table.
6. Click **OK**, and then save the job.

What to do next

If required, add and configure additional stages. Otherwise, configure the target database stage.

Supported data type conversions

The data types for output links on the CDC Transaction stage are defined when you generate the InfoSphere DataStage job. However, you can change the data types on the output links from the CDC Transaction stage.

Important: When the CDC Transaction stage and the InfoSphere CDC for InfoSphere DataStage server communicate, they use the original data type. Be careful when you change to a data type that has a smaller precision. If the InfoSphere CDC for InfoSphere DataStage server sends a value that is out of the valid range for the data type, the job fails.

In the following tables, each row shows the original data type. An X in a column indicates that the original data type can be changed to the target data type listed in the column header. If the original data type and the target data type do not appear in the same table, then the original data type cannot be changed to the target data type. If you need to perform a data type conversion that is not allowed by the CDC Transaction stage, then perform the conversion in a downstream stage that does support the conversion.

Table 5. Allowed data type conversions for Character data types

	NVarChar	VarChar	LongVarChar	LongNVarChar	NChar	Char
NVarChar	X	X ¹	X ¹	X	X	X ¹
NChar	X	X ¹	X ¹	X	X	X ¹

Table notes:

- Character conversion occurs from UTF-16 to the system default Character set.

Table 6. Allowed data type conversions for Decimal data types

	Numeric	Decimal
Numeric	X	X

Table 7. Allowed data type conversions for Binary data types

	LongVarBinary	VarBinary	Binary
LongVarBinary	X	X	X

Table 8. Allowed data type conversions for Integer data types

	TinyInt	SmallInt	Integer	BigInt
SmallInt	X	X	X	X
Integer	X	X	X	X
BigInt	X	X	X	X

Table 9. Allowed data type conversions for Floating point number data types

	Real	Double	Float
Real	X	X	X
Double	X	X	X

Table 10. Allowed data type conversions for Date and Time data types

	Time	Date	Timestamp
Time	X	X	X
Date	X	X	X
Timestamp	X	X	X

Adding and configuring additional stages

You can add stages to the job to process the change data before delivering it to the target database connector stage.

Before you begin

Configure the CDC Transaction stage.

About this task

When you add stages to the job, use the following basic guidelines in the overall design of the job:

- If a subscription contains more than one table, the changes for each table can follow a different path through the job.
- To ensure transactional integrity, set all stages in the job to run sequentially.
- You can add stages to the template job. However, the job can contain no more than one CDC Transaction stage and one target database connector stage. If more than one CDC Transaction stage exists, the job fails. If more than one target database stage exists, the stages operate in separate transactions, and no transactional integrity is expected.
- Each output link from the CDC Transaction stage forms a stream of end-of-wave (EOW) markers to notify the target database connector of the commit timing. The target database connector stage commits the transaction when it receives a set of EOW markers from all the input links.
- You can split the stream by using stages that support multiple output links, such as a Filter stage or a Copy stage. You can then mix those streams at a downstream stage that supports multiple input links, such as a Funnel, Join, or Merge stage. However, do not mix a stream that originated from the CDC Transaction stage with a stream that originated from a different source. For example, some stages, such as source database stages or the Wave Generator stage, also produce EOW markers. In a CDC Transaction stage job, you do not want the EOW markers from those stages to reach the target database connector. If the extra EOW markers reach the target database connector, the transaction fails to commit as expected.
- Avoid processing data on the bookmark link. The bookmark link is used to pass bookmark information in an internal format. The bookmark link directly connects the CDC Transaction stage to the target database connector. Set the write mode on the bookmark link to "Update then insert" to overwrite the existing bookmark information on the target database connector stage.
- You can use stages that write data to external objects, such as a file or database other than the target database. However, transactional integrity is not expected for the data that is written to those external objects. If the job fails and the subscription restarts, the CDC Transaction stage re-sends the data. Duplicate data will probably be written to the external objects, but not to the target database.

- As much as possible, design the job so that workloads are equally distributed among output streams. The target stage needs to wait for the incoming data or EOW markers on all the input links. Therefore, processing heavy workloads on one of the change data streams between the CDC Transaction stage and the target stage affects the overall throughput of the job.

Procedure

1. In the Designer client palette area, open the section of the palette that contains the stage that you want to use.
2. Select the stage icon, and drag the stage to your open job.
3. Select the end of the output link from the CDC Transaction stage and drag it to the new stage.
4. Right-click the new stage and draw a link from the new stage to the next stage on the canvas.
5. Double-click the new stage, and configure it. In the **Advanced** tab, make sure that you select **Sequential** for the **Execution mode**. For information about configuring the stage, click **Help**.
6. Optional: Repeat these steps to add more stages.

What to do next

Configure the target database stage.

Configuring the target database stage to write change data

When you generate the template job, it contains a DB2 connector as the target stage. If your target database is not a DB2 database, you can manually edit the job to remove the DB2 connector stage and add any supported connector type.

Before you begin

- Configure the CDC Transaction stage.
- If required, add and configure additional stages.

About this task

This procedure describes how to configure the target database stage. For specific configuration details, see the documentation for the connector that you are using.

Procedure

1. Configure the bookmark input link.
 - a. On the Job canvas, double-click the connector stage.
 - b. In the upper left corner of the stage editor, select the bookmark link in the diagram.
 - c. In the **Properties** tab, in the **Write mode** field, verify that **Update then insert** is selected. When this option is selected, the incoming bookmark overwrites the existing one.
 - d. In the **Table name** field, replace the template bookmark name, BOOKMARKTABLE, with the real name of the bookmark table that you created for the subscription. The target database must contain a unique bookmark table for each subscription. The table name can be specified with or without the schema name. If the table name is not qualified with the schema name, the schema name is implicitly determined by the ODBC driver or the target database. For more information about how the table name is implicitly qualified, see the documentation for ODBC or your database.
 - e. Optional: In the **Columns** tab, view the record format of the bookmark table.
2. Configure the input links for the connector stage. The connector stage includes an input link for each table in the subscription.
 - a. In the upper left corner of the stage editor, select the input link for a table.

- b. In the **Properties** tab, in the **Table name** field, specify the fully qualified name of the change data table in the target database.
 - c. Specify usage properties for the input link. For more information about specifying these properties, click **Help**.
 - d. Optional: In the **Columns** tab, verify that the record format matches the format that you expect.
 - e. Repeat these steps for each input link that represents a table.
 - f. After configuring all input links, click the **Link Ordering** tab, and specify the order in which you want the links to execute.
3. Configure the connector stage.
 - a. In the upper left corner of the stage editor, select the connector stage.
 - b. In the **Properties** tab, specify connection details for the target database. For example, specify values for the following types of properties: **Instance**, **Database**, **User name**, and **Password**. For more information about these properties, click the **Help** button.
 - c. Optional: After specifying connection details, click the **Test** link to test the connection.
 - d. In the **Advanced** tab, verify that the value for **Execution mode** is **Sequential**. All stages in the job must run sequentially, or transactional integrity is not guaranteed.

What to do next

After configuring all stages in the job, you compile the job.

Compiling the job

When you finish designing a job, compile it. Then use the log file to gather information that might help you adjust the job configuration or correct error conditions.

Procedure

1. In the Designer client, open the job that you want to compile.
2. On the toolbar, click the **Compile** toolbar button (or press F7).
3. If the Compilation Status area shows errors, edit the job to resolve the errors. After resolving the errors, click **Re-compile**.

What to do next

Start the subscription and run the job.

Starting the subscription and running the job

You start the subscription in the InfoSphere CDC Management Console. If auto-start is selected, starting the subscription automatically starts the InfoSphere DataStage job. If auto-start is not selected, you must start the job manually.

Before you begin

If the subscription has been running recently, verify that the previous InfoSphere DataStage job run has completed and that all data has been processed before starting the subscription. If replication begins before the previous job run has completed, the job might produce errors or fail.

Also verify that the job is compiled and that it is ready to run.

Procedure

1. Log on to the InfoSphere CDC Management Console. Specify a user ID that is authorized to run the subscription.
2. Open the Monitoring perspective.
3. In the **Subscriptions** tab, right-click the name of the subscription, and select one of the following replication methods:

Option	Description
Start Mirroring	Select this option to continuously monitor the source for updates.
Start Refresh	Select this option to replicate all existing data on the source to the target. This option provides a simple snapshot of the source. For more information about this option, see the InfoSphere CDC documentation.

4. Select one of the following options to control how the job runs:

Option	Description
Continuous	Select this option to replicate continuously until the subscription is stopped. All updates are captured.
Scheduled end	Select this option to replicate until a specified time or location in the log is reached.

5. If the job is not configured to start automatically, start the job. In the IBM InfoSphere DataStage and QualityStage Designer client, open the job, and click **Run**. When the job is running, the links turn blue, and you can see that the row count for the bookmark link increments.
6. Optional: After changes are committed to the source database, view the data that is written to the target database. For example, for DB2 connector:
 - a. In the Designer client, open the database connector stage.
 - b. Select an input link that is associated with a table. In the **Properties** tab, click **View Data**. The View Data window shows all change data records that are stored in the target database table. Repeat this step to see data for other tables.

You can also run, schedule, and monitor the job in the InfoSphere DataStage and QualityStage Director client.

What to do next

Monitor the status of the job and the subscription.

Monitoring the status of the job and the subscription

You use the InfoSphere CDC Management Console to monitor the status of a subscription, and the InfoSphere DataStage and QualityStage Director client to monitor the status of the job. By using these tools together, you can view statistics about latency and throughput, and identify error and warning conditions.

Before you begin

Start the subscription and run the job.

Procedure

1. Monitor the status of the subscription by using the InfoSphere CDC Management Console.
 - a. Log on to the InfoSphere CDC Management Console. Specify a user ID that has Monitor role privileges.
 - b. In the Monitoring perspective, select the **Subscriptions** tab.
 - c. In the **Subscriptions** tab, select the subscription that you want to monitor.
 - d. In the statistics pane, select **Collect Statistics**.
 - e. Click one of the following icons to display the type of information that you want to view:

Overview

View an overview of statistics about latency, mirroring activity, and current replication events.

Latency

View the current, high, low, and average values for latency.

Activity

View the current and high values for bytes per second or operations per second, and the cumulative count of bytes and operations.

Events

View a list of the events that have occurred over a specific time period. Use this view to see errors and warnings about the subscription.

For more information about monitoring subscriptions, see the InfoSphere CDC documentation.

2. Monitor the status of the job in the InfoSphere DataStage and QualityStage Director client.
 - a. Log on to the Director client. If you are already logged on to InfoSphere DataStage and QualityStage Designer, select **Tools > Run Director**.
 - b. To view the status of the job, select **View > Status**. To view summary information about the job, such as the number of rows updated and performance statistics, right-click the job, and select **Monitor**.
 - c. To view the event log, select **View > Log**. Use this view to see errors and warnings about the job. To view details about a specific event, double-click the event in the log.

For more information about monitoring jobs, see the *IBM InfoSphere DataStage and QualityStage Director Client Guide*.

What to do next

When you no longer want to replicate the source data, end the subscription.

Ending the subscription

After you start a CDC Transaction stage job, it runs continuously until the subscription ends and the InfoSphere CDC for InfoSphere DataStage server sends a termination message to the CDC Transaction stage. To end the subscription, you specify one of the termination modes in the InfoSphere CDC Management Console.

Before you begin

You must have a user ID that is authorized to end the subscription.

About this task

When you end replication by using the InfoSphere CDC Management Console, the job ends gracefully. However, if a failure occurs that prevents the InfoSphere CDC for InfoSphere DataStage server and the CDC Transaction stage from communicating, there might be a delay before the InfoSphere CDC for InfoSphere DataStage server or InfoSphere DataStage recognizes that a failure has occurred. If the job ends unexpectedly, the InfoSphere CDC for InfoSphere DataStage server uses bookmark information to ensure that no data is lost when the subscription resumes.

Procedure

1. Log on to the InfoSphere CDC Management Console.
2. In the Monitoring perspective, click the **Subscriptions** tab.
3. In the list of subscriptions, right-click the name of the subscription, and select **End Replication**.
4. In the End Replication window, select one of the following methods to end replication:

Option	Description
Normal	<p>InfoSphere CDC finishes in-progress work and then ends replication. If a refresh is in progress, the refresh finishes for the current table before replication ends.</p> <p>Normal is the most appropriate option for most business requirements and is the preferred method for ending replication in most situations.</p>
Immediate	<p>InfoSphere CDC stops all in-progress work and then ends replication. If a refresh is in progress, the refresh for the current table is interrupted and then replication ends.</p> <p>Use this option if business reasons require replication to end faster than when Normal is specified. Selecting this option results in a slower start time when you resume replication on the subscription.</p> <p>For the CDC Transaction stage, this method is identical to Abort. The CDC Transaction stage signals the other stage to stop processing. Data that is not committed on the target database is discarded. The job ends, and the log contains fatal errors.</p>

Option	Description
Abort	<p>InfoSphere CDC stops all in-progress work and then ends replication rapidly. If a refresh is in progress, the refresh is interrupted and the target stops processing any data that has not been committed before replication ends.</p> <p>Use this option if your business reasons require a rapid end to replication and you are willing to tolerate a much slower start time when you resume replication on the subscription.</p> <p>For the CDC Transaction stage, this method is identical to Immediate. The CDC Transaction stage signals the other stage to stop processing. Data that is not committed on the target database is discarded. The job ends, and the log contains fatal errors.</p>
Scheduled End	<p>InfoSphere CDC processes all committed changes to the indicated point in the database log and then ends replication normally.</p> <p>Use this option if your business reasons require a scheduled end to replication.</p>

For more information about these options, see the InfoSphere CDC documentation.

5. Click **OK**. In the list of subscriptions, the state changes to Inactive. The InfoSphere CDC for InfoSphere DataStage server sends a termination message to the CDC Transaction stage, and the job stops gracefully.

Example: Applying change data by using a CDC Transaction stage

This example demonstrates how to use IBM InfoSphere Change Data Capture (InfoSphere CDC) to read changes that occur on tables in an Oracle database and then use IBM InfoSphere DataStage to replicate the changes to a target DB2 database.

When you use InfoSphere CDC to capture changes, the change data includes the before and after images of the data, along with control columns. The control columns provide additional details about the change data, such as when the change occurred and the type of operation that was performed. Depending on how replication is configured, the change data might be sent in a single record that includes the before and after images, or in multiple records.

Simply inserting the change data record into the target database might not meet your business needs. To remove the extra control columns and replicate only the change data, you can create a stored procedure and call it from the target connector stage.

Setting up source and target tables for the example

For this example, you set up a source table in an Oracle database and a target table in a DB2 database. You also create a stored procedure that is called when the target DB2 connector stage runs.

Before you begin

The log retention policy for the source database must be set to retain logs.

Procedure

1. Connect to the source Oracle database and issue the following DDL statement to create the source table, S100:

```
create table S100 (C1 number(5), C2 VARCHAR2(10));
```

If you want to use a DB2 database as the source instead of an Oracle database, modify the syntax of the create table statements, as required, and create the tables in a DB2 database.

2. Connect to the target DB2 database and issue the following DDL statement to create the target table, T100:

```
create table T100 (C1 decimal(5), C2 varchar(10))
```

3. Issue the following DDL statement to create the bookmark table in the target DB2 database:

```
create table BOOKMARKTABLE (DM_KEY smallint not null primary key, DM_BOOKMARK  
    varchar(1024))
```

Important: The target database must contain a unique bookmark table for each subscription. If a table with the name BOOKMARKTABLE already exists in the target database, specify a different name for this table. Use this name wherever the steps in this example indicate that you need to specify BOOKMARKTABLE.

4. Create the following stored procedure in the target DB2 database. You call this stored procedure from the target DB2 connector stage. The stored procedure uses the value of the DM_OPERATION_TYPE column to apply only the change data to the target table, T100:

```
create procedure P1 (  
    IN DM_OPERATION_TYPE CHAR(1),  
    IN BEFORE_C1 decimal(5),  
    IN BEFORE_C2 varchar(10),  
    IN AFTER_C1 decimal(5),
```

```

    IN AFTER_C2 varchar(10)
)
language sql
BEGIN
  DECLARE dummy INT;
  DECLARE SQLSTATE CHAR(5) DEFAULT '00000';
  CASE DM_OPERATION_TYPE
    WHEN 'I' THEN
      INSERT INTO T100 (C1, C2) VALUES (AFTER_C1, AFTER_C2);
    WHEN 'D' THEN
      BEGIN ATOMIC
        DECLARE CURSOR_FOR_DELETE CURSOR FOR SELECT 1 FROM T100
        WHERE C1=BEFORE_C1 AND C2=BEFORE_C2 FOR UPDATE;
        OPEN CURSOR_FOR_DELETE;
        FETCH FROM CURSOR_FOR_DELETE INTO dummy;
        IF (SQLSTATE = '00000') THEN
          DELETE FROM T100 WHERE CURRENT OF CURSOR_FOR_DELETE;
        END IF;
        CLOSE CURSOR_FOR_DELETE;
      END;
    WHEN 'U' THEN
      BEGIN ATOMIC
        DECLARE CURSOR_FOR_UPDATE CURSOR FOR SELECT 1 FROM T100
        WHERE C1=BEFORE_C1 AND C2=BEFORE_C2 FOR UPDATE;
        OPEN CURSOR_FOR_UPDATE;
        FETCH FROM CURSOR_FOR_UPDATE INTO dummy;
        IF (SQLSTATE = '00000') THEN
          UPDATE T100 SET C1=AFTER_C1, C2=AFTER_C2
          WHERE CURRENT OF CURSOR_FOR_UPDATE;
        END IF;
        CLOSE CURSOR_FOR_UPDATE;
      END;
  END CASE;
END@

```

Setting up replication

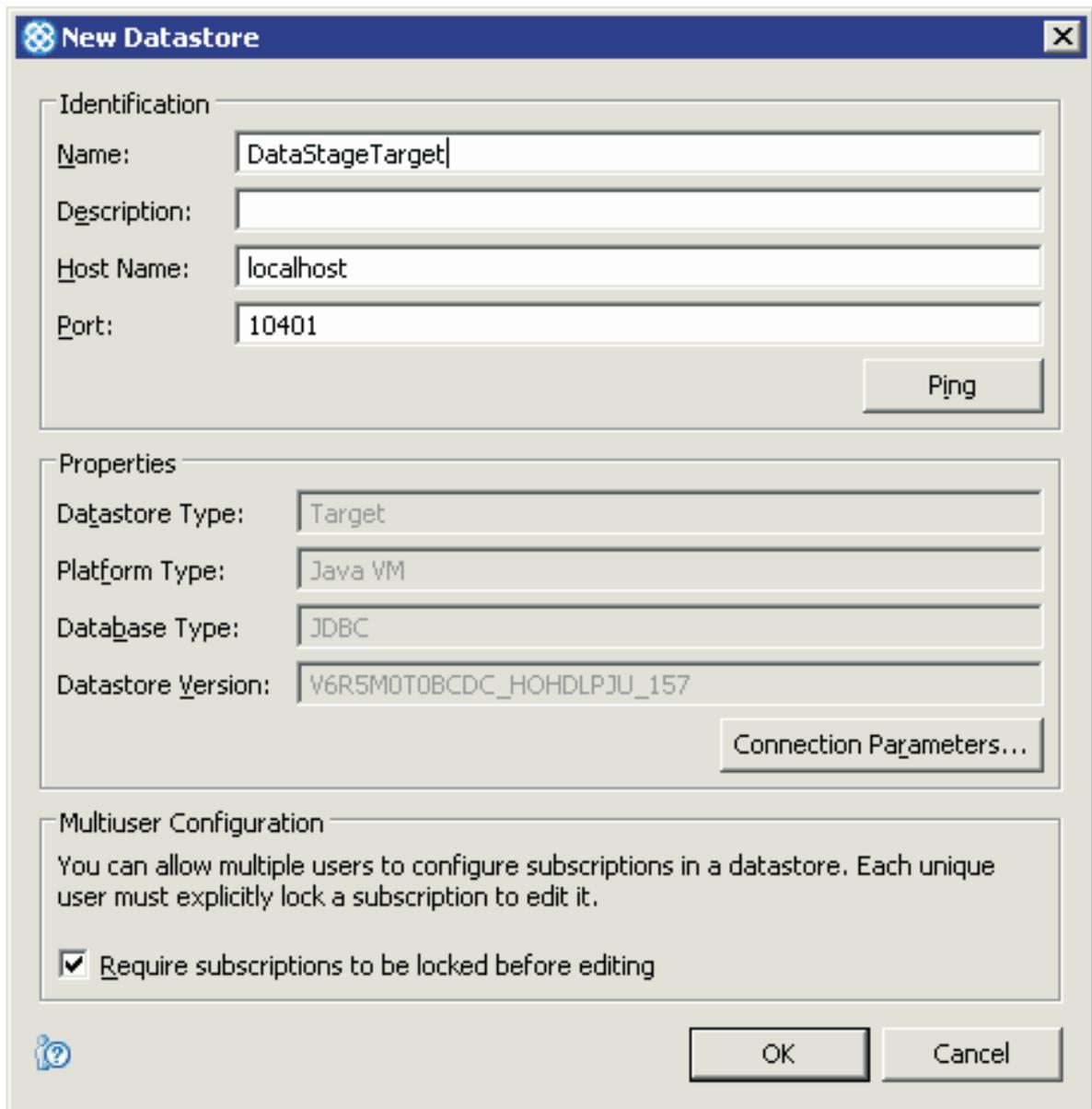
You use the IBM InfoSphere Change Data Capture Management Console to add and configure a new subscription. Before you can perform these steps, the InfoSphere CDC software must be installed and configured.

Before you begin

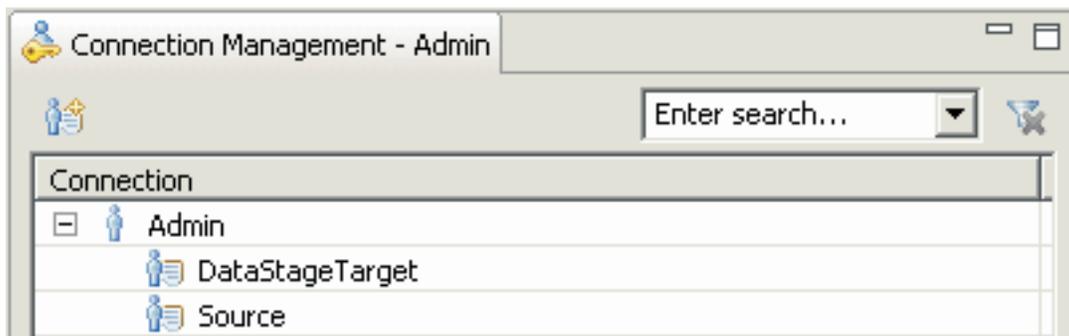
Install and configure the InfoSphere CDC software.

Procedure

1. Log on to the InfoSphere CDC Management Console.
2. In the Access Manager perspective, add new data stores for the source and the target.
 - a. Add a new data store called Source to access the Oracle database.
 - b. Add a new data store called DataStageTarget to access InfoSphere DataStage. For example, the following image shows values that are specified for the new data store.

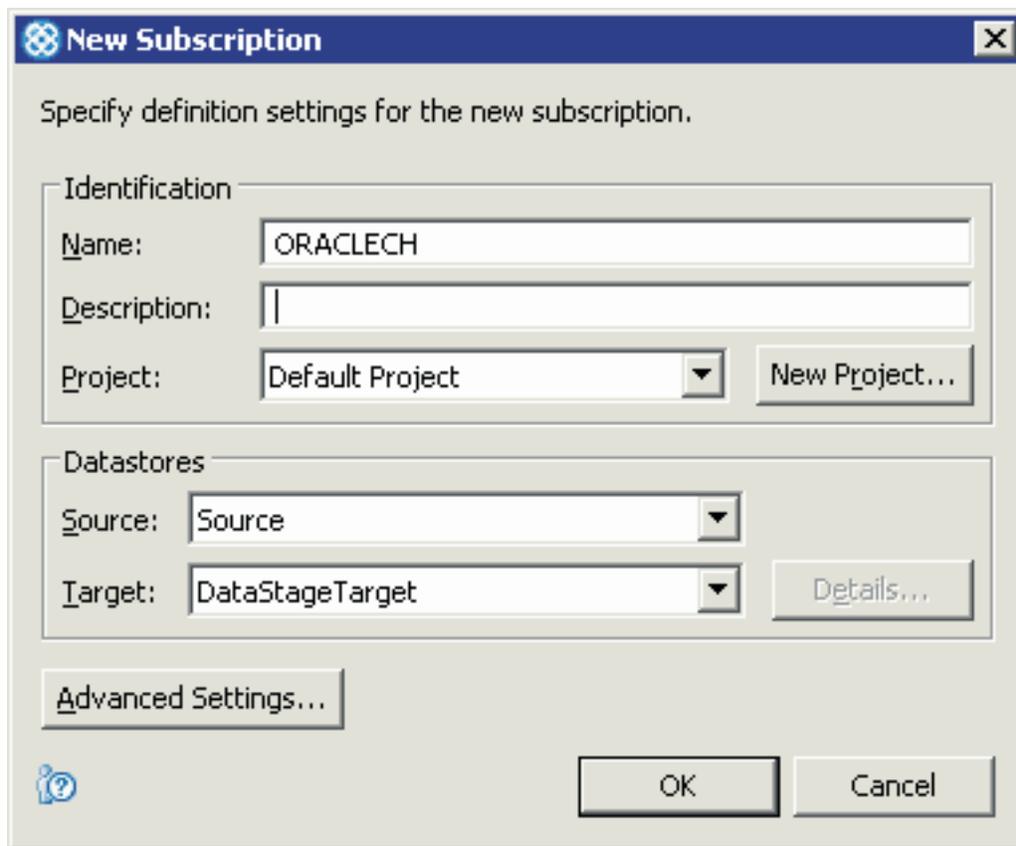


- c. In the **Connection Management** tab, assign a user to each data store. In the following example, the user Admin is assigned to both data stores:



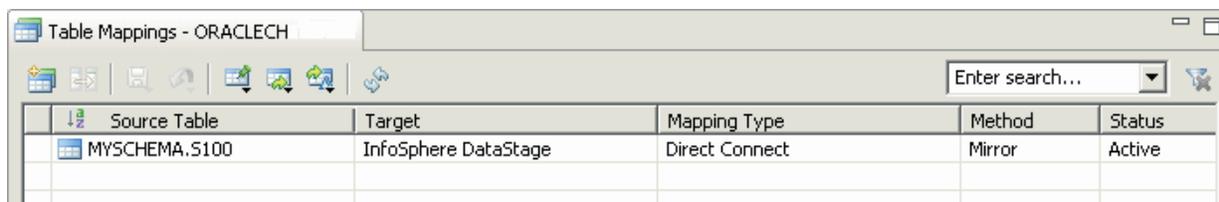
- 3. In the Configuration perspective of the InfoSphere CDC Management Console, add a new subscription called ORACLECH. In the **Source** field, select **Source**. In the **Target** field, select **DataStageTarget**. For example, the following image shows values that are specified for the new

subscription.



4. In the **Subscriptions** tab, click the subscription named ORACLECH to edit it. Right-click the subscription, and select **Map Tables** to map the table S100 to the target of the subscription, InfoSphere DataStage. When you run the wizard, make sure that you select the following options:
 - a. In the Select Mapping Type page, select **InfoSphere DataStage**.
 - b. In the Select InfoSphere DataStage Connection Method page, select **Direct Connect**.
 - c. In the Select Source Tables page, select the table named S100 that you created earlier.
 - d. In the InfoSphere DataStage Direct Connect page, select **Single Record**.

When you finish, the table mappings for the subscription are displayed in the **Table Mappings** tab.



5. In the **Subscriptions** tab, right-click the subscription, ORACLECH, and select **InfoSphere DataStage > InfoSphere DataStage Properties**. Specify the following properties:
 - a. In the Direct Connect area, specify the following values for the InfoSphere DataStage job. The subscription must be open for editing when you define these properties, or your changes are not saved.
 - In the **Project Name** field, specify the name of your InfoSphere DataStage project.
 - In the **Job Name** field, specify ORACLECH.
 - In the **Connection Key** field, specify myconnectionkey.

- b. Optional: If you want to configure the job to start automatically when the subscription starts, select **Auto-start InfoSphere DataStage Job**. If the check box is grayed out, the InfoSphere CDC for InfoSphere DataStage server instance is disabled for auto-start. For auto-start to be enabled, the PATH environment variable must contain the path to the **dsjob** executable file at the time that the instance is created.
- c. Click **OK**.

Generating a template for the job

After creating a subscription in the InfoSphere CDC Management Console, you generate a template for the InfoSphere DataStage job.

Procedure

1. In the **Subscriptions** tab of the Configuration perspective, right-click the subscription, and click **InfoSphere DataStage > Generate InfoSphere DataStage Job Definition**. This option is available only after the subscription is correctly set up.
2. In the dialog box that is displayed, browse to select the file named **ORACLECH.dsx**, and click **Save**.
3. Copy the **.dsx** file to a location that is accessible by the computer where InfoSphere DataStage and QualityStage Designer is installed.

Importing and configuring the job

When you generate the template job, some of the stage properties are set to values that are accurate for the subscription. However, you must update the job to specify settings that were not known when the job was generated.

Procedure

1. In the Designer client, click **Import > DataStage components**, and specify the path to the .dsx file, for example: C:\ORACLECH.dsx. Click **OK**.
2. In the **Repository** area, expand **Jobs**, and double-click ORACLECH to open the job.
3. Configure the CDC Transaction stage.
 - a. On the Job canvas, double-click the CDC Transaction stage.
 - b. In the upper left corner of the stage editor, verify that the stage is selected in the diagram.
 - c. In the **Properties** tab, update the following ODBC properties. These properties are used to retrieve bookmark information.

Bookmark DSN

Enter the ODBC DSN that you created for bookmark information when you configured InfoSphere CDC.

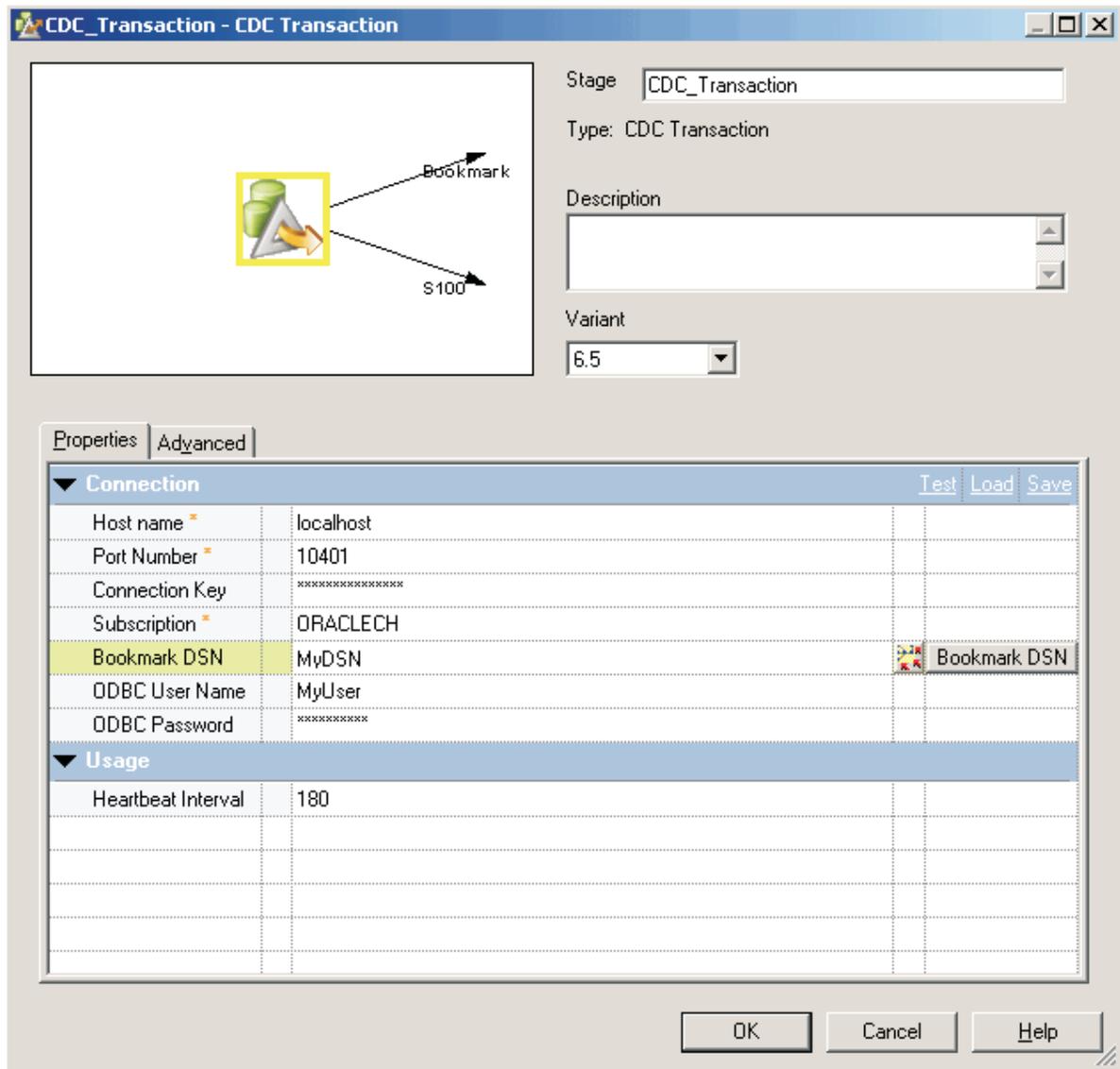
ODBC user name

Enter the user name that you specified for the bookmark DSN.

ODBC password

Enter the password for the ODBC user that you specified for the bookmark DSN.

The values for other properties are already set based on properties that are specified in the subscription. For example, the following image shows values that are specified for CDC Transaction stage properties.



d. Click **OK**, and then save the job.

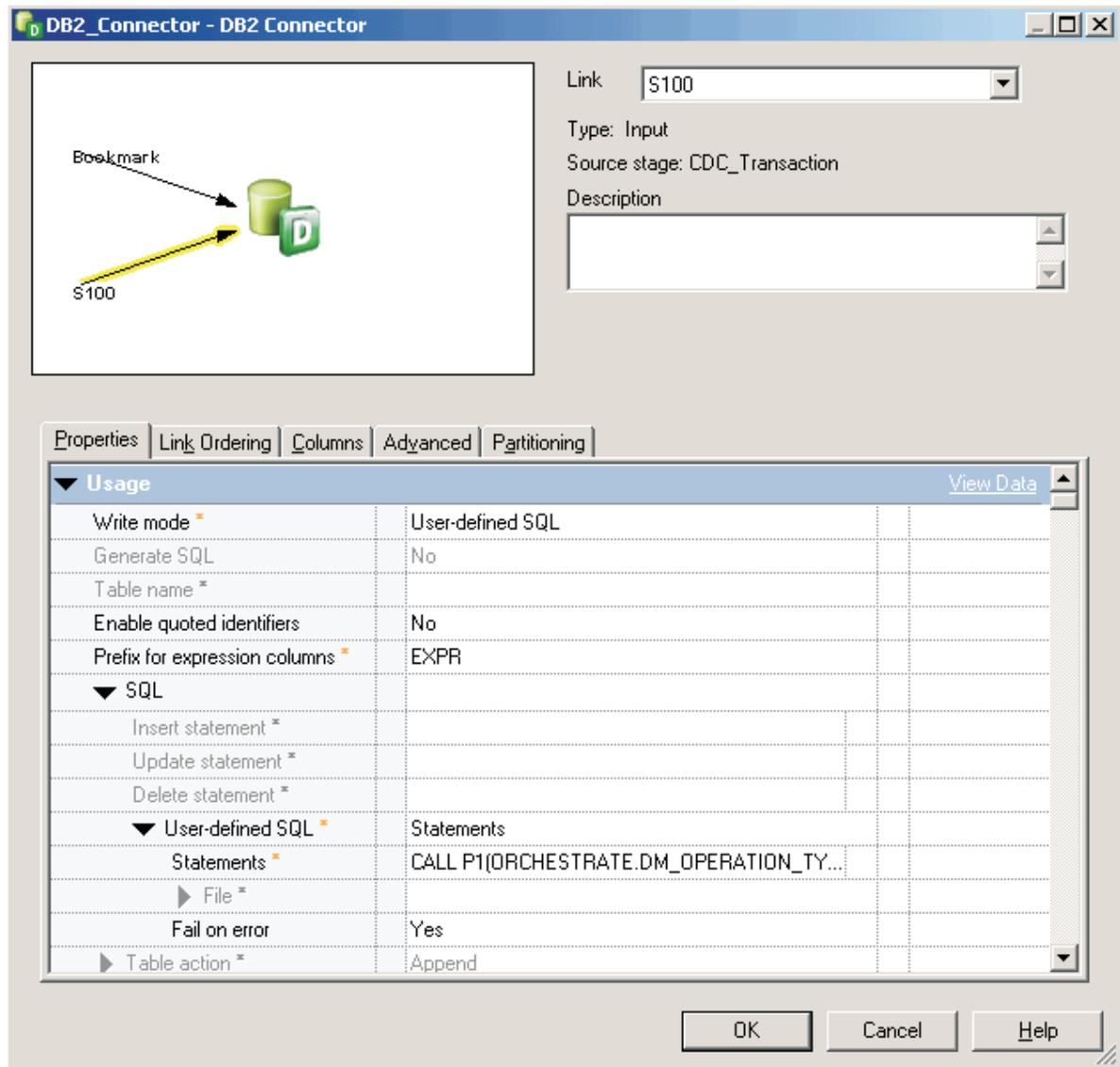
For this example, you do not configure the CDC Transaction stage output links because they are already configured as part of the job template. Also, you do not configure the bookmark link because the table that you set up for bookmarks uses the default name that is specified in the job template.

4. Configure the DB2 connector input links.

- a. On the Job canvas, double-click the DB2 connector stage.
- b. In the upper left corner of the stage editor, select the link named S100, and specify usage properties.
- c. In the **Write mode** field, select **User-defined SQL**.
- d. In the **SQL > User-defined SQL** field, select **Statements**.
- e. In the **Statements** field, enter the following procedure call:

```
CALL P1(ORCHESTRATE.DM_OPERATION_TYPE,
        ORCHESTRATE.BEFORE_C1,
        ORCHESTRATE.BEFORE_C2,
        ORCHESTRATE.C1,
        ORCHESTRATE.C2)
```

For example, the following image shows values that are specified for the DB2 connector input link.



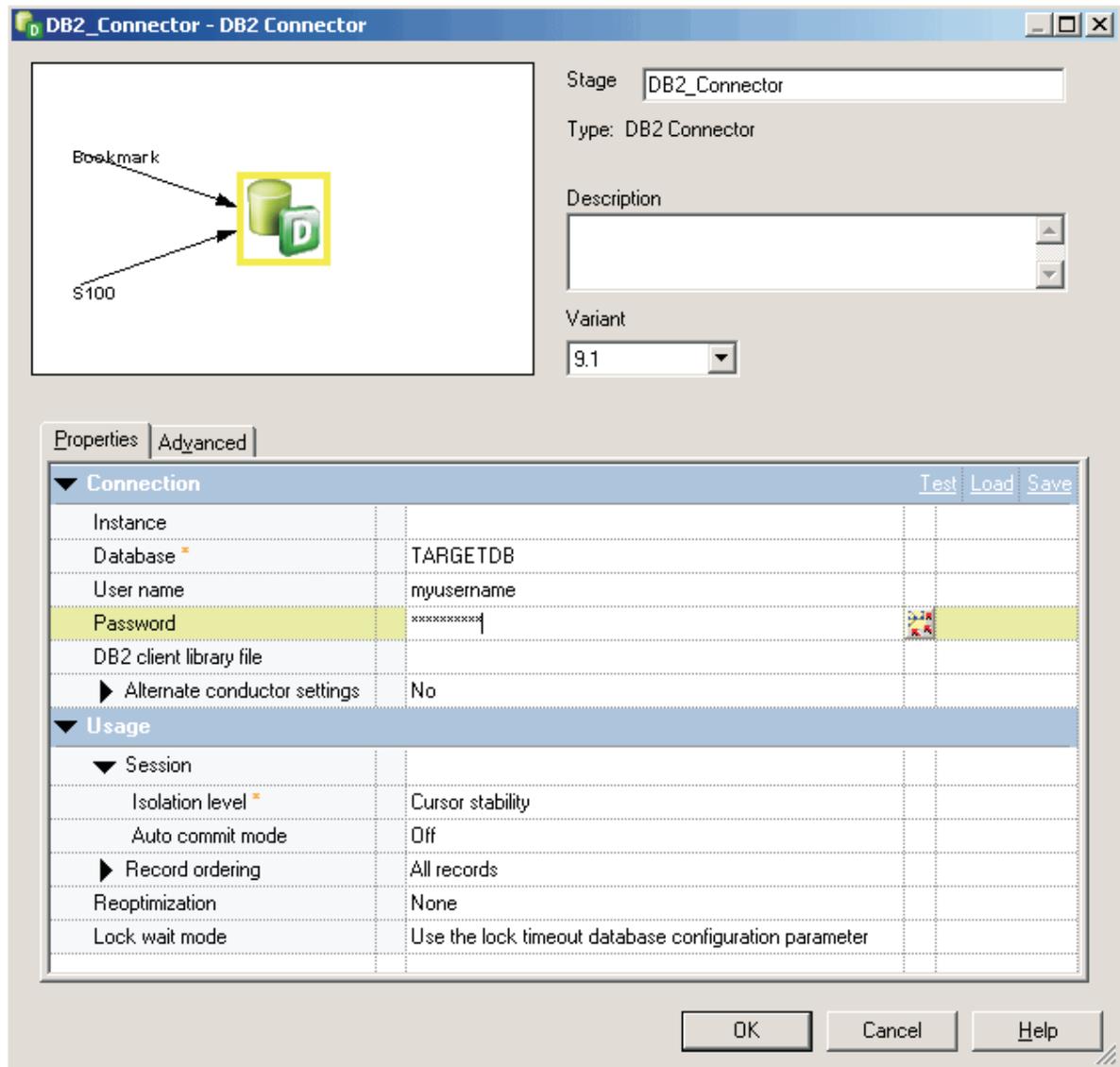
f. In the **Link Ordering** tab, specify the following link order:

Bookmark
S100

You do not configure the bookmark link because the table that you set up for bookmarks uses the default name that is specified in the job template.

5. Configure the DB2 connector stage.

- a. In the upper left corner of the stage editor, select the DB2 connector stage in the diagram.
- b. In the **Properties** tab, specify connection details for the target database. For example, specify the instance, database, user name, and password for connecting to the target database.



- c. Click the **Test** link to test the connection. If the connection is not successful, enter the correct connection information, and try again. Click **OK**.
6. Click **OK** and then save the job.
7. Click the **Compile** toolbar button (or press F7). If the Compilation Status area shows errors, edit the job to resolve the errors. After resolving the errors, click **Re-compile**.

Starting the subscription and running the job

You start the subscription in the InfoSphere CDC Management Console. If auto-start is selected, starting the subscription automatically starts the InfoSphere DataStage job. If auto-start is not selected, you must start the job manually.

Procedure

1. Log on to the InfoSphere CDC Management Console.
2. Open the Monitoring perspective.
3. In the **Subscriptions** tab, right-click the subscription named ORACLECH, and select **Start Mirroring**.
4. Select **Continuous** to replicate continuously until the subscription is stopped.

5. If the job is not configured to start automatically, start the job. In the Designer client, open the compiled job, and click **Run**. When the job is running, the links turn blue, and you can see that the row count for the bookmark link increments.
6. In Monitoring perspective of the InfoSphere CDC Management Console, verify that the state of the subscription is **Mirror Continuous**.

Writing change data to the target database

When you update the source database tables, InfoSphere CDC sends the change data to the CDC Transaction stage, which sends the change data to the target DB2 connector stage. The target DB2 connector stage then calls the stored procedure to replicate the change data to the target.

Procedure

1. Connect to the source database and issue the following query to update the source table, S100:

```
insert into S100 values (1, 'FIRST');
insert into S100 values (2, 'SECOND');
update S100 set C2='THIRD' where C1=1;
delete S100 where C1=2;
```

2. Issue the following select statement against the source database to view the data in the S100 table:

```
select * from S100;
```

The select statement returns the following data:

```
      C1 C2
-----
      1 THIRD
```

3. After changes are committed to the source database, connect to the target database, and issue the following queries to view the data in the target table:

```
select * from T100
```

The select statement returns the following data:

```
      C1 C2
-----
      1 THIRD
```

4. Issue additional queries to update, insert, and delete data. Issue select statements on the target tables to see how the change data is replicated to the target.

Troubleshooting

These topics describe how to resolve problems that you might encounter when you use the CDC Transaction stage in an InfoSphere DataStage job.

Value of DM_TIMESTAMP column in the target table is incorrect

When the CDC Transaction stage inserts a timestamp into the DM_TIMESTAMP column of the target table, the value of the column might not be consistent with the expected value. The value is converted to UTC time zone before it is passed to the CDC Transaction stage.

Symptoms

An unexpected timestamp value is inserted in the DM_TIMESTAMP column of the target table.

Causes

The value of DM_TIMESTAMP is converted to UTC time zone before the InfoSphere CDC for InfoSphere DataStage server passes the value to the CDC Transaction stage.

Resolving the problem

In the InfoSphere CDC Management Console, add a system parameter to the InfoSphere CDC for InfoSphere DataStage data store that specifies the following values:

- In the **Parameter name** field, specify `ds_output_timestamp_utc`.
- In the **Value** field, specify `false`.

For more information about adding a system parameter to a data store, see the InfoSphere CDC documentation.

Job fails to start automatically

After starting a subscription in the InfoSphere CDC Management Console, the job fails to start automatically if auto-start is not correctly configured.

Symptoms

When you start a subscription in the InfoSphere CDC Management Console, the InfoSphere DataStage job that is associated with the subscription does not start.

Causes

There are multiple possible causes for this problem:

- The instance of InfoSphere CDC for InfoSphere DataStage is not running on the same computer as InfoSphere DataStage.
- The subscription might not be configured for auto-start.
- The InfoSphere CDC for InfoSphere DataStage process does not have the path or library path to the **dsjob** command and dependent files.
- The job is not ready to run because it has not been compiled or is in a bad state, such as DSJE_BADSTATE or DSJE_JOBLOCKED.

Resolving the problem

In the InfoSphere CDC Management Console, verify that auto-start is specified for the subscription:

1. In the **Subscriptions** tab, right-click the subscription, and select **InfoSphere DataStage > InfoSphere DataStage Properties**.
2. Verify that **Auto-start InfoSphere DataStage Job** is selected.
3. Verify that the correct project name and job name are specified.

If the **Auto-start InfoSphere DataStage Job** check box is grayed out, you need to recreate the instance of InfoSphere CDC for InfoSphere DataStage, making sure that the PATH environment variable contains the path to the **dsjob** command. For auto-start to work correctly, the instance must be installed on the same computer as InfoSphere DataStage.

If the check box is selected, but the job fails to start, issue the **dsjob** command manually to verify that the command starts the job successfully. Also verify that the InfoSphere CDC for InfoSphere DataStage process has the library path to the dependent shared library files for the **dsjob** command.

If the job does not start because it is in a bad state, perform the following steps to resolve the issue:

1. In the InfoSphere DataStage and QualityStage Director client, select **Jobs > Cleanup Resources**.
2. Stop any processes that are accessing the job.
3. Either recompile the job in the InfoSphere DataStage and QualityStage Designer client, or reset the job in the Director client. To reset the job, in the Status view, select **Job > Reset**.

Job fails while processing large records

While running an InfoSphere DataStage job, the job fails if a record is larger than the size that is specified for a transport block in the InfoSphere DataStage parallel engine.

Symptoms

The job fails with an error message that indicates that the record is too big to fit in a block.

Causes

The job needs to pass each record in a single transport block. However, the job is attempting to pass a record that is larger than the size of the transport block that InfoSphere DataStage can pass between the stages of a job. The size of the transport block is specified by the `APT_DEFAULT_TRANSPORT_BLOCK_SIZE` environment variable. By default, the value of this environment variable is 128 KB.

Resolving the problem

Increase the value of the `APT_DEFAULT_TRANSPORT_BLOCK_SIZE` environment variable to a size that is larger than the largest record that you need to transfer between the stages of the job. Note that increasing the value of this environment variable affects all links and might increase overall memory and disk usage. For more information, see the *IBM InfoSphere DataStage and QualityStage Parallel Job Advanced Developer's Guide*.

Tip: If the source tables have LOB data type columns, consider specifying a larger value for the **Large Object Truncation Size** property in the InfoSphere DataStage Properties window of the InfoSphere CDC Management Console.

Job fails with authentication failure message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage cannot authenticate with the InfoSphere CDC for InfoSphere DataStage server.

Symptoms

The CDC Transaction stage fails with an error message that indicates that authentication has failed.

Causes

The CDC Transaction stage properties that are used for authentication are not specified or are specified incorrectly.

Resolving the problem

In the **Properties** tab of the CDC Transaction stage, specify the correct values for the **Subscription** and **Connection key** properties.

Job fails with bookmark table missing message

While running an InfoSphere DataStage job, the job fails if the InfoSphere CDC for InfoSphere DataStage server cannot write bookmark information to the bookmark table.

Symptoms

The CDC Transaction stage fails with a message that indicates that the bookmark table is missing.

Causes

The InfoSphere CDC for InfoSphere DataStage server cannot write bookmark information to the bookmark table because the table does not exist. The table might have been deleted after the subscription started.

Resolving the problem

Create the bookmark table on the target database, and then restart the subscription and job.

Job fails with connection error message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage is unable to connect to the InfoSphere CDC for InfoSphere DataStage server.

Symptoms

The job fails with an error message that indicates that the connection to the InfoSphere CDC for InfoSphere DataStage server failed.

Causes

The CDC Transaction stage failed to connect to server's listening port that is specified in the **Host name** and **Port number** stage properties for the CDC Transaction stage.

Resolving the problem

Verify that the InfoSphere CDC for InfoSphere DataStage server is running on the host name and port number that are specified in the stage properties for the CDC Transaction stage.

Job fails with multiple bookmark links or no bookmark link error message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage finds multiple bookmark links or no bookmark link.

Symptoms

The job fails with an error message that indicates that either multiple bookmark links exist, or that no bookmark link is defined.

Causes

The CDC Transaction stage requires one and only one bookmark link. The job fails if more than one link is specified as the bookmark link, or if no links are specified as the bookmark link.

Resolving the problem

In the stage editor for the CDC Transaction stage, modify the link properties to ensure that one and only one link is specified as the bookmark link. To specify that a link is the bookmark link, set the **Bookmark link** property to **Yes**.

Job fails with table name for subscription not specified on link error message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage cannot find the link that is required to transfer the data for a source table.

Symptoms

The job fails with an error message that indicates that the table name for the subscription is not specified in the link property.

Causes

The CDC Transaction stage cannot find the link to transfer the data for the source table because the **Table name** property for one or more output links is not correctly specified.

Resolving the problem

In the stage editor for the CDC Transaction stage, select the output link, and specify the correct value for the **Table name** property.

Job fails with no response from server error message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage does not receive a response from the InfoSphere CDC for InfoSphere DataStage server.

Symptoms

The job fails with an error message that indicates that no response was returned from the InfoSphere CDC for InfoSphere DataStage server.

Causes

The CDC Transaction stage and the server were able to establish a session, but the CDC Transaction stage failed to receive data within the timeout period. If this error occurs at the beginning a job, the subscription might not be started correctly in the InfoSphere CDC Management Console. This error might occur when the job is started manually, but the subscription is not started, and no data flows from the server. This error might also be caused by a server-side issue or network issues.

Resolving the problem

In the InfoSphere CDC Management Console, view the subscription status and the event log. If the subscription is not started, start it. Also check the health of the network connection between the server computer and the computer where the CDC Transaction stage runs.

Job fails with ODBC access failure message

While running an InfoSphere DataStage job, the job fails if the CDC Transaction stage cannot use ODBC to access the target database.

Symptoms

The CDC Transaction stage fails with a message that indicates that ODBC access has failed.

Resolving the problem

In the **Properties** tab of the CDC Transaction stage, verify that the correct values are specified for the following stage properties: **Bookmark DSN**, **ODBC user name**, and **ODBC password**. Click the **Test** link to verify that these connection parameters are correct.

Output data is truncated

The data that is output from InfoSphere CDC is truncated before it is written to the CDC Transaction stage.

Symptoms

The data that is output from InfoSphere CDC is unexpectedly truncated at some length.

Causes

If you are replicating LOB data, the data is truncated at the value specified for the **Large Object Truncation Size** property in the InfoSphere DataStage Properties window of the InfoSphere CDC Management Console.

Resolving the problem

Define a larger truncation size for LOB data. Specify a large enough value so that the total length of the output record does not exceed the size of the transport block that can pass between the stages of a job. The size of the transport block is specified by the `APT_DEFAULT_TRANSPORT_BLOCK_SIZE` environment variable. By default, the value of this environment variable is 128 KB. You can increase the size of the transport block. However, increasing this size affects all links and might increase overall memory and disk usage.

Subscription stops while still in "Starting" state

After starting a subscription in the InfoSphere CDC Management Console, the state does not progress as expected if the CDC Transaction stage fails to get a connection to the InfoSphere CDC for InfoSphere DataStage server.

Symptoms

The status of the subscription in the InfoSphere CDC Management Console does not progress beyond the "Starting" state.

Causes

The InfoSphere CDC for InfoSphere DataStage server is waiting for the connection from the CDC Transaction stage.

Resolving the problem

In the **Properties** tab of the CDC Transaction stage, verify that the values specified for **Host name** and **Port number** are correct. Also verify that the connection details specified for the bookmark DSN are correct.

Frequently asked questions about the CDC Transaction stage

Find the answers to frequently asked questions about using the CDC Transaction stage in an InfoSphere DataStage job.

“Does the bookmark table need to be created for each subscription?”

“How do I see all the properties in the stage editor?”

“What are the columns that are prefixed with "DM_" in the generated job?”

“What do I do if the generated job does not compile successfully?”

“What is the heartbeat interval, and do I need to configure it?” on page 50

“What is the Variant field in the stage editor? ” on page 50

“Why can't I change the execution mode from Sequential to Parallel in the stage editor? ” on page 50

“Why does the Test link succeed even though the wrong host name or port number is specified? ” on page 50

“Why doesn't the Bookmark DSN button show the correct DSN?” on page 50

“Why is the subscription name truncated in the stage property value when InfoSphere CDC Management Console generates the job template?” on page 50

Does the bookmark table need to be created for each subscription?

Yes. The bookmark table should be accessed from only one job. Other applications should not access the bookmark table while the subscription is running.

How do I see all the properties in the stage editor?

The properties that are displayed vary depending on whether you select the stage or the link in the stage editor. Click the link or the stage icon to see properties for each category.

What are the columns that are prefixed with "DM_" in the generated job?

These columns contain replication-related metadata that is provided by InfoSphere CDC or the CDC Transaction stage. You can use the data in these columns, or you can drop the columns if the design of your job or business logic does not require them. For more information about these columns, see the topic that describes the record format for change data records.

What do I do if the generated job does not compile successfully?

Edit the job and specify values for all required properties before you compile the job. You must specify values for the following properties in the **Properties** tab of the CDC Transaction stage:

- **Host name**
- **Port number**
- **Subscription**

When you generate the job, the InfoSphere CDC Management Console sets predefined default values for the following CDC Transaction stage properties. You usually need to update the values for these properties:

- **Bookmark DSN**
- **ODBC user name**
- **ODBC password**
- **Table name** for the bookmark link.

What is the heartbeat interval, and do I need to configure it?

The answer to this question depends on your computing environment, especially the network that used by InfoSphere CDC and the CDC Transaction stage to communicate. A heartbeat is a virtual message between InfoSphere CDC and the CDC Transaction stage. The heartbeat is used to keep the network connection alive. You specify a value in seconds for the heartbeat interval that accurately reflects the time interval during which you are sure that the network connection is kept alive. During this time interval, InfoSphere CDC and the CDC Transaction stage attempt to keep the session alive by sending a packet.

What is the Variant field in the stage editor?

The **Variant** field indicates the version number of the InfoSphere CDC for InfoSphere DataStage server. For this release, the only value that you can select is 6.5.

Why can't I change the execution mode from Sequential to Parallel in the stage editor?

To ensure transactional integrity, the CDC Transaction stage runs in sequential mode only.

Why does the Test link succeed even though the wrong host name or port number is specified?

When you click the **Test** link for testing connection parameters in the CDC Transaction stage, only the parameters for accessing the target database through ODBC are tested. These connection properties are specified in the **Bookmark DSN**, **ODBC user name**, and **ODBC password** fields.

Why doesn't the Bookmark DSN button show the correct DSN?

The ODBC DSN that you want to use is not correctly configured. For more information, see the topic about setting up the ODBC data source for bookmark information.

Why is the subscription name truncated in the stage property value when InfoSphere CDC Management Console generates the job template?

The CDC Transaction stage and InfoSphere CDC for InfoSphere DataStage use only the first eight characters of the subscription name that is specified in the InfoSphere CDC Management Console. Therefore, the name is truncated when the job is generated.

Links to non-IBM Web sites

This information center may provide links or references to non-IBM Web sites and resources.

IBM makes no representations, warranties, or other commitments whatsoever about any non-IBM Web sites or third-party resources (including any Lenovo Web site) that may be referenced, accessible from, or linked to any IBM site. A link to a non-IBM Web site does not mean that IBM endorses the content or use of such Web site or its owner. In addition, IBM is not a party to or responsible for any transactions you may enter into with third parties, even if you learn of such parties (or use a link to such parties) from an IBM site. Accordingly, you acknowledge and agree that IBM is not responsible for the availability of such external sites or resources, and is not responsible or liable for any content, services, products or other materials on or available from those sites or resources.

When you access a non-IBM Web site, even one that may contain the IBM-logo, please understand that it is independent from IBM, and that IBM does not control the content on that Web site. It is up to you to take precautions to protect yourself from viruses, worms, trojan horses, and other potentially destructive programs, and to protect your information as you deem appropriate.

Notices and trademarks

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

Notices

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

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

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

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

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

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

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

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

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

Trademarks

IBM, the IBM logo, and ibm.com are trademarks 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.

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.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium 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, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office

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

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

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^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS^{Link} licensee of the United States Postal Service.

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

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 11. 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 www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
Software services	You can find information about software, IT, and business consulting services, on the solutions site at www.ibm.com/businesssolutions/
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 www.ibm.com/account/
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 http://www.ibm.com/software/sw-training/
IBM representatives	You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/

Providing feedback

The following table describes how to provide feedback to IBM about products and product documentation.

Table 12. Providing feedback to IBM

Type of feedback	Action
Product feedback	You can provide general product feedback through the Consumability Survey at www.ibm.com/software/data/info/consumability-survey
Documentation feedback	To comment on the information center, click the Feedback link on the top right side of any topic in the information center. You can also send comments about PDF file books, the information center, or any other documentation in the following ways: <ul style="list-style-type: none">• Online reader comment form: www.ibm.com/software/data/rcf/• E-mail: comments@us.ibm.com

Index

Special characters

- .dsx files
 - generating for CDC Transaction stage jobs 17

A

- authentication
 - troubleshooting failures 45
- autostart
 - troubleshooting 43

B

- bookmarks
 - adding data source names (DSN) 14
 - CDC Transaction stage properties 22
 - creating tables for 18
 - record format 21
 - troubleshooting failures 45, 46

C

- CDC Transaction stage
 - configuration 22
 - overview 5, 6
- CDC Transaction stage jobs
 - additional stages 24
 - compiling 26
 - data flow 6
 - data type conversions 23
 - development steps 17
 - example 33
 - frequently asked questions 49
 - generating templates 17
 - guidelines for job design 24
 - limitations 8
 - monitoring 29
 - software prerequisites 8
 - starting 27
 - supported installation topologies 8
 - target database stage
 - configuration 25
 - target table setup 18
 - troubleshooting 43
- change data capture
 - CDC Transaction stage questions 49
 - configuration for CDC Transaction stage 13
 - InfoSphere DataStage job
 - development 17
 - overview 5, 6
 - record format 18
 - setup 15
 - target table setup 18
- connections
 - troubleshooting failures 45
- customer support
 - contacting 57

D

- data source name (DSN)
 - adding for bookmark access 14
- data types
 - supported conversions 23
- DM_BOOKMARK column 21
- DM_KEY column 21
- DM_OPERATION_TYPE column 18
- DM_SORTKEY column 18
- DM_TIMESTAMP column 18
- DM_TXID column 18
- DM_USER column 18

E

- encryption
 - support in CDC Transaction stage 8
- examples
 - CDC Transaction stage jobs 33

F

- frequently asked questions
 - CDC Transaction stage jobs 49

I

- InfoSphere CDC
 - CDC Transaction stage questions 49
 - configuring instances 13
 - ending subscriptions 31
 - overview 5, 6
 - software installation topologies 9
 - starting subscriptions 27

L

- large records
 - troubleshooting failures 44
- legal notices 53
- LOB data
 - troubleshooting truncation 47
 - truncation in CDC Transaction stage 8

M

- multiple record format 18

N

- non-IBM Web sites
 - links to 51

O

- ODBC
 - CDC Transaction stage failures 47
 - specifying access for bookmarks 14
- ODBC password
 - CDC Transaction stage 22
- ODBC user name
 - CDC Transaction stage 22
- overview
 - InfoSphere CDC 5, 6

P

- product accessibility
 - accessibility 1
- product documentation
 - accessing 3

R

- record format
 - bookmarks 21
 - change data capture columns 18
 - multiple record 18
 - single record 18
- replication
 - CDC Transaction stage installation topologies 9
 - CDC Transaction stage
 - prerequisites 8
 - CDC Transaction stage questions 49
 - ending 31
 - InfoSphere DataStage job
 - development 17
 - setting up 15
 - starting 27
 - supported installation topologies 8
 - target table setup 18

S

- single record format 18
- software services
 - contacting 57
- subscriptions
 - ending 31
 - setting up 15
 - starting 27
 - troubleshooting 48
- support
 - customer 57

T

- tables
 - mapping for InfoSphere CDC 15
 - troubleshooting table name errors 46

- templates
 - generating for CDC Transaction stage jobs 17
- trademarks
 - list of 53
- transactional integrity
 - CDC Transaction stage 8
- troubleshooting
 - authentication failures 45
 - autostart fails 43
 - CDC Transaction stage jobs 43
 - connection failure 45
 - missing bookmark tables 45
 - multiple or no bookmark links error 46
 - ODBC access failures 47
 - table name not specified error 46
 - too big for block error 44
 - truncated LOB data 47
- types
 - supported conversions 23

W

- Web sites
 - non-IBM 51



Printed in USA

SC19-3432-00

