17

What is the Attunity Stream CDC Solution

This section includes the following topics:

CDC Solution Overview

The Attunity Stream CDC Architecture

What Can Be Captured?

CDC Solution Overview

Attunity Stream captures and delivers the changes made to enterprise data sources in real-time. This enables you to move mainframe and enterprise operational data in real-time to data warehouses and data marts, improving the efficiency of ETL processes, synchronizing data sources, and enabling event-driven business activity monitoring and processing.

Attunity Stream provides agents that non-intrusively monitor and capture changes to mainframe and enterprise data sources. Changes are delivered in real-time or consumed as required using standard interfaces.

The Attunity Stream CDC solution provides the following capabilities:

Capture changes to data in real-time or near-real-time for applications that demand zero latency and require the most up-to-date data. Real-time data capture guarantees that a change event is immediately available at the consumer. Near-real-time data capture involves a configurable delay before a change is available at the consumer.

Using a near-real-time configuration, when there is significant capture activity, events are reported immediately. However, after the system has been idle, it may take a few seconds for the events to start flowing again.

Enable consumers of changed data to receive changes quickly, either by asking for the changes in high-frequencies (such as every few seconds), or by sending them the changes as soon as they are identified.

The consumer application periodically requests changes, receiving each time a batch of records that represent all the changes that were captured since the last request cycle. Change delivery requests can be done in low or high frequencies, for example, every 15 seconds or a set number of times a day. The extracted changes are exposed to enable the consumer application to seamlessly access the change records using standard interfaces like ODBC and JDBC or XML and JCA.

Using the Attunity Stream CDC solution enables ETL (extract, transform, and load) processes to run without bringing the site computer systems down. CDC enables the movement of only changes to data while the operational systems are running, without the need for a downtime window.

The CDC architecture consists of the following components:

CDC agents, which are located on the same computer as the changes to be captured. Each agent is customized for the specific data source on the specific platform.

A CDC agent provides access to the journal (or logstream), to read the journal for specific table changes. The agent maintains the last position read in the journal (the stream position or context) and starts at this point the next time it polls the journal. The context is stored in the repository where the agent adapter definition is stored. The adapter definition includes a single primary interaction which is used to access the appropriate journal and includes the list of tables to monitor for changes.

Depending on the agent, transactions are supported. If transactions are not used,then auto-commit is used.

A CDC staging area: Changes are stored in a staging area, enabling a single scan of a journal to extract the more information about changes to more than one table and to also enable filtering of the changes. A committed change filter is currently available and a redundant change filter (hot-spot optimization) is planned. Both of these filters are described in Tracking Changes - Auditing. The staging area is described in more detail in The Staging Area.

CDC data sources, to enable access to the changes using industry standards such as ODBC and JDBC.

A CDC data source uses Attunity metadata and holds metadata for tables marked for change data capture. The CDC data source points to either the CDC agent or the staging area to enable getting the changes, as events, using ODBC or JDBC.

Additionally, the CDC data source converts the data types in the table with changes to be captured to standard ODBC data types.

An audit of the changes that are captured. The audit is stored on disk and changes captured by both the CDC agent and written to the staging area are recorded.

Once a CDC solution has been defined, the changes can be pushed to the consumer, using a standard Attunity Connect event router. Setting up and using an event router is described in AIS User Guide and Reference.

The Attunity Stream CDC Architecture

The Attunity Stream CDC solution operates with or without a staging area. When a staging area is used, the change data capture agent reads the journal and channels requested changes from the journal to the staging area. The change data capture agent resides on the same computer as the data source with changed data to be captured.

The staging area can reside on any computer. If the consumer application is SQL-based, a change data capture data source is available to access the relevant changes written to the staging area. The change data capture data source can reside on any computer.

Note: In addition to this interaction, secondary interactions are defined during runtime, based on the table metadata of each table specified as part of the change data capture solution.

The CDC architecture is shown in the following figure:

Figure 17–1 Change Data Capture when a Staging Area is used

When a staging area is used, the change data capture agent reads the journal and channels requested changes from the journal to the staging area. The change data capture agent resides on the same machine as the data source with changed data to be captured. The staging area can reside on any machine.

If the consumer application is SQL-based, a change data capture data source is available to access the relevant changes written to the staging area. The change data capture data source can reside on any machine.

A context is stored both for the staging area and the agent which mark the last point where changes were captured. For more information about the context, see The Staging Area.

Figure 17–2 Change Data Capture when a Staging Area is not used

When a staging area is not used, the change data capture agent reads the journal and channels requested changes from the journal either to a JCA or XML-based consumer application or, if the consumer application is SQL-based, to a change data capture data source that is made available to access the relevant changes. The change data capture agent resides on the same machine as the data source with changed data to be captured. The change data capture data source can reside on any machine.

A context is stored for the agent. For more information about the context, see to The Staging Area.

The Staging Area

The staging area is an area used by Attunity Stream to store captured data from a journal.

When capturing changes from more than one table, without a staging area, the journal is scanned by the agent once for each table. When a staging area is used, the journal is scanned once and changes for every required table read during that scan are passed to the staging area, where they are stored. Thus, the journal is scanned once each time it is polled. Furthermore, once the changes have been written to the staging area, processing of these changes is performed independently of the journal.

Another benefit of using the staging area is when transactions are used. The changed data is not written to the change queue until the transaction is committed. Thus, if the transaction fails, there is no overhead of having to back out any processing done with the steps in the failed transaction.

The staging area can be on any Windows platform running Attunity server and not necessarily on the same server as the CDC agent. Once the information has been extracted from the journal and written to the staging area, processing of changes is performed on the staging area only. Thus, the staging area should be setup to consider the network configuration and where the consumer application runs.

The staging area maintains the last position read by the consumer application (the staging area context) and starts at this point the next time a request from the consumer application is received. The context is stored in the repository where the staging area is maintained.

The staging area is indexed. Thus, access to the staging area for a specific stream is quick.

Use of the staging area is recommended in the following situations:

When changes to data in more than one table needs to be captured. When transactions are used. The changed data is not written to the change queue until the transaction is committed. Thus, if the transaction fails, there is no overhead of having to back out any processing done with the steps in the failed transaction.

When repositioning the stream position (resetting the context) is planned to be performed often.

The staging area is cleared by default every 48 hours. All events that have been in the staging area for more than 48 hours are deleted.

Handling Before and After Images

Where applicable, when setting up a CDC solution, you can specify if you want to save the before image information of changed data that is recorded in the journal as well as the after image (undo and redo records for DB2 on the z/OS platform) information. The default is that only the after images are captured. The following table shows what is captured if before images are requested:

To capture before images, the journal must be set up to include before images, as described for each type of journal.

Tracking Changes - Auditing

When using a CDC solution, you can produce an audit trail of the captured data. The following audit levels are available:

None: Indicates no auditing is performed.

Summary (Statistics): The total number of records retrieved from the change queue and system and error messages are reported. In addition, header information about each record captured, such as the type of operation and the table name, is reported.

Detailed: The total number of changes retrieved are reported, along with system and error messages. In addition, header information and record information about each record captured is reported.

The audit entries can be viewed in Attunity Studio Runtime Manager perspective in the Event monitor. Entries include a direction as follows:

– Entries from the CDC agent: The audit entries show what data changes were extracted from the agent by the consumer application.

– Entries from the staging area: The audit entries show what data changes were extracted from the staging area by the consumer application and what entries were written to the staging area by the agent.

Security Considerations

In general, Attunity Stream relies on the security mechanisms implemented at a site.

For example, on az/OS system using RACF to manage security, the security rules implemented in RACF are also applied to Attunity Stream CDC.

Additionally, you can specify as part of the CDC setup who can access the CDC agent and staging area to extract changed data and who can write changed data to the staging area.

What Can Be Captured?

Attunity Stream includes the following data source CDC agents:

DB2 Journal CDC on z/OS systems: The change data capture agent monitors both the archived and active DB2 journals and captures changes made to specific tables, which are written to these journals. Since transaction information is also stored,

Table 17–1 What is Captured

Operation Before Image After Image

INSERT X P

UPDATE P P

DELETE P X

Note: Details of the header information is provided in the CDC agent specific chapters.

the committed change filter can be used to ensure that only committed changes are captured. For more information about the committed change filter, see Tracking Changes - Auditing.

DB2 Journal CDC on OS/400 platforms: The change data capture agent monitors a DB2 database journal and captures changes made to specific tables, which are written to this journal. Since transaction information is also stored, the committed change filter can be used to ensure that only committed changes are captured. For more information about the committed change filter, see Tracking Changes - Auditing.

DISAM on Windows platforms: The change capture agent monitors a journal for changes in DISAM tables and captures changes made to specific tables, which are written to this journal. This solution can only be used when the DISAM data is updated using Attunity Connect and not when updated directly by another program.

IMS on z/OS systems: The change capture agent monitors a system log for changes in IMS/DB tables and captures changes made to specific tables, which are written to the logstream.

Oracle on Windows and UNIX platforms: The change data capture agent monitors

Oracle REDO log files for changes in Oracle tables from Oracle version 9iR2. The CDC solution polls the Oracle Logminer with its archive mode to capture changes.

The staging area must be used when capturing Oracle data changes.

VSAM-Batch on z/OS systems: The change data capture agent monitors a system log for changes in VSAM tables. Transactions are handled at the program level of the program. If a program fails, a decision to rollback must be made independently of the VSAM-batch change data capture agent.

VSAM-CICS on z/OS systems: The change data capture agent monitors a CICS logstream for changes in VSAM tables. When transaction information is not available, rolled-back transactions appear as a set of backed-out changes, applied to the data.

Query-based CDC on all Attunity server platforms: A generic change capture agent that enables the capture of changes in any of the data sources supported by Attunity Connect. The query-based agent only captures changes based on changes to a specific field in the table. An initial value is specified for the field as the starting change data capture context. The query-based agent does not include the ability to specify a staging area, nor to reset the context.