18

Implementing a Change Data Capture Solution

This section contains the following topics:

Overview

CDC System Architecture

Setting up AIS to Create a Change Data Capture

CDC Adapter Definition

CDC Streams

Transaction Support

Troubleshooting

Overview

The CDC solution lets you identify changes to data. This is referred to as consuming changes. The consumed changes are reported in a special log file. With CDC, the data is consumed in real time, that is at exactly the time the INSERT, UPDATE, or DELETE operations occur in the source tables. Changes are stored in change tables. Attunity Stream uses CDC (Change Data Capture) agents to consume changes to data that they display when needed. The following agents are supported:

VSAM Under CICS CDC (z/OS)

VSAM Batch CDC (z/OS)

IMS/DB CDC on z/OS Platforms

Adabas CDC on z/OS Platforms

Adabas CDC on UNIX Platforms

Adabas CDC for OpenVMS

DB2 CDC (z/OS)

DB2 CDC (OS/400 Platforms)

Enscribe CDC (HP NonStop Platforms)

Microsoft SQL Server CDC

Oracle CDC (on UNIX and Windows Platforms)

Query-Based CDC Agent (for work tables)

SQL/MP CDC on HP NonStop

Setting up AIS to Create a Change Data Capture

You set up a change data capture in Attunity Studio. This provides a single, integrated GUI where you can easily configure a Change Data Capture. The workflow required for setting up a CDC solution is:

1. Install the System Components in the proper locations.

See Installing System Components for an explanation on what to install and the locations for installing the components.

2. Apply the license. Use Attunity Studio to apply the license to the machines defined in your system.

3. Configure the Change Data Capture solution. Use the CDC Solution perspective in Attunity Studio to add a new CDC solution and define its configuration properties.

This automatically configures all the necessary components for the Change Data Capture, including the data source driver and CDC agent. See Creating a CDC Solution with Attunity Studio for more information.

See Configuring the System for Change Data Captures (Using Studio) for additional information.

Installing System Components

To use AIS for a Change Data Capture solution, install the following:

AIS on the Back End: The back end of the system is where your data is stored. In an CDC solution, you consume changes from data sources such as Oracle, to create a log with the change data information. This is where your CDC agent resides. You must install the full (thick) version of AIS on the back end. For information on how to install AIS, see the Installation Guide for the platform you are working with. Data Source: The data source that you are working with is installed with AIS. For more information, see the installation guide for the platform you are working on. Attunity Studio: Attunity Studio can be installed on any Windows computer in your system. Attunity Studio provides a graphic interface for configuring the components in your system. For information on installing Attunity Studio, see the installation guide or the platform you are working on.

See also CDC System Architecture.

Configuring the System for Change Data Captures (Using Studio)

You must configure the following system components to set up your system for application access:

Machines: You must configure the machines used in the system. Make sure to configure the machine where your back-end application and application adapter reside. For information on how to add and configure machines, see Setting up Machines. You can also test the machine connection in Attunity Studio.

User Profiles: You must set up the users in your system. Setting up users is for security purposes. You can specify which users have access to various machines. For information on setting up user profiles, see User Profiles and Managing a User Profile in Attunity Studio.

CDC Agents: You need to set up the CDC agent that you want to use. Attunity Studio lets you set up a CDC agent by Creating a CDC Solution with Attunity Studio. This lets you use Attunity Studio to manually set up your own CDC.

Data Sources: If you create the CDC agent manually, you need to include the data source that your CDC agents supports. To set up your data source, you must:

– Prepare to use the data source by Adding Data Sources to Attunity Studio.

– Make sure that your data source is responding correctly by carrying out the procedures in Testing a Data Source.

– For non-relational data sources, you must define the metadata. Data source metadata defines the table structure for your data source.

All metadata is imported from relevant input files (such as COBOL copybooks) and maintained using the Attunity Studio Design perspective Metadata tab.

For more information on setting up data sources, see Data Sources and the data source Reference section for the adapter you are using.

Generated CDC Components

The following are generated when you set up a CDC using the CDC Solution perspective.

A new binding for the change data capture components on the machine where the data source with changes to be captured resides. CDC components are defined under this new binding, which is named name of project\_ag, if the agent and the staging are on different machines. If the agent and staging area are on the same machine, binding is called name of project\_router.

A CDC agent, which is defined as an adapter under the CDC binding and named name of project\_ag.

A CDC data source, which is on the same machine as the CDC agent. This data source is called name of project.

A staging area. This includes a binding and a data source. Both are called name of project\_sa.The staging area is defined on the machine where the CDC changes are stored.

A CDC router. This includes a binding, a data source, and an adapter. The binding and the adapter are called name of project\_router and the data source is called name of project\_sa. The router data source is a copy of the staging area’ s data source. For each of the bindings defined in the CDC Solution, a workspace with the same name is defined.

Handling Arrays Defined in the Source Data

The CDC solution stores changes in XML and does not enable capturing virtual tables.

To work with these tables, they need to be flattened. By default, the data source used in Attunity Stream is defined with the arrays flattened.

Note: Attunity metadata is independent of its origin. Therefore, any changes made to the source metadata (for example, the COBOL copybook) is not indicated in the Attunity metadata.

When using the CDC solution, the built-in flattening mechanism enables the following:

A view for the parent table records in the non-relational structure. This view does not include reference to the related child-arrays.

A view for each child array that includes a unique key from the parent table and an array row number. This enables adding further columns from the parent table to the virtual array view.

CDC System Architecture

The CDC solution system components include software that is installed in your system. These components are described in Installing System Components. In addition, various files are created, such as the change data tables that keep track of the consumed changes.

The CDC solution is usually deployed on three platforms. A platform can be one or more servers. The following figure shows the CDC system architecture.

Figure 18–1 CDC Solution System Components

This figure shows the three platforms used to deploy a CDC solution. These platforms are:

Database Platform: The database platform is where the database and the agent run. This platform can be any platform supported by AIS. The database platform is, in many cases, also a legacy application platform. This means that processing overhead on this platform should be minimized.

Staging Area Platform: The staging area platform is where the change tables are hosted. This platform also hosts the SQL based change router, which enters data into the change tables. This platform can be any Windows or UNIX platform.

ETL Platform: The ETL platform is where the ETL tool runs. This platform can be any platform supported by AIS as well as any platform that can run any of AIS thin clients (such as any standard Java platform). The ETL platform may be the same as the Staging Area platform depending on the ETL tool resource requirements. If the platform is strong enough, placing both the staging area and the ETL tool on the same platform can reduce network utilization and improve throughput.

CDC Adapter Definition

The CDC adapter definition is returned in a standard XML format. This format is based on the Attunity ACX Protocol. The CDC adapter definition is composed of two parts, the interactions, which are static and stored in the Attunity Studio repository and the schema, which is dynamic and is created at runtime. The following figure shows the parts of the CDC definition’s metadata.

Figure 18–2 CDC Definition

The figure above shows the adapter metadata as viewed in Attunity Studio. This shows the interactions only.

You can view the schema by Using the Attunity XML Utility. The following figure shows how to configure the XML utility to return the full definition including the schema.

Figure 18–3 XML Utility Metadata Button

CDC Agent Metadata Definition Description

This section describes the elements in the CDC agent metadata. This metadata has two elements.

Interactions

Schema

Interactions

The following are the outbound interactions in the CDC agent metadata:

getEvents: Gets the CDC events from the agent.

setStreamPosition: Moves the stream position back to a specified position to re-read events from that point.

The following are the inbound interactions in the CDC agent metadata. These events are not callable by a client application.

eventStream

table-name

Schema

The following are the schema elements in the CDC agent metadata:

header: This accompanies all events. It is the same for a specific CDC agent. A list of header fields is included in the CDC agent reference sections. For more information, see the Overview for a list of the CDC agents.

eventStream: An implicit union of all CDC event types. This includes all tables plus transaction events. \*table-name\* - A record for each table including all its fields.

CDC Streams

A Change Data Capture tracks or consumes changes in a data source using streams. A stream is a point in a data base’s change log that is used as a point of reference by the CDC agent. Each time you begin to start a Change Data Capture, a new stream, with a unique name is automatically created. You can have as many concurrent streams as necessary. The place where the changes are currently being consumed is called the stream position. The stream position is automatically saved in the data base.

You can filter a stream to consume for selected tables in the data base or for all of the tables.

The figure below shows the CDC Stream.

Note: The schema records of these interactions are important for client apps because they describe the metadata of the change events.

Figure 18–4 CDC Stream

The following are the parts of the CDC stream in the figure above.

Data Source: The native data source where you are consuming changes. For example, Adabas.

Change Logger: This is the native tool that logs changes for a data source.

Change Source: This is the data source log. It contains various information depending on the data source and the log configuration. This part has a different name depending on the data source you are working with.

For example, for Adabas this is called the PLOG, for DB2 it is called a journal, and for Oracle it is called REDO logs. For more information, see the Overview for a list of the CDC agents.

Change Capture Agent: This is the Attunity CDC Agent. The agent consumes the changes from the Change Source and outputs them into a standard XML format.

Transaction Support

CDC agents that support transactions can eliminate uncommitted changes from the stream service Most CDC agents support transactions. The reference section for each CDC agent indicates whether the agent supports transactions.

The following CDC agents support transactions:

DB2 and DB2/400

VSAM Batch

Adabas (all Adabas agents)

Oracle

Enscribe

SQL/MP

When a CDC agent handles a transaction, it reads the change record and then:

If the record is from a new transaction it creates an in memory transaction record cache for Txn-ID

If the record is a DML record, then it is added to the appropriate transaction record cache by the Txn-ID

If there was a Rollback, it deletes the Txn-ID transaction cache

If there was a Commit, it distributes the Txn-ID transaction records to the various change files

It the transactions is idled for too long it is timed out

This is written to a special Txn-ID.txe file

For information on troubleshooting, see Troubleshooting in AIS.

Troubleshooting

If changes are not written to the journal or other native Change Source, make sure that the journal and system (such as CICS) file definitions match (for example, the same logstream number in CICS is used in the journal).

To check that changes are written in the journal

1. From the Windows Start menu, point to Programs, Attunity and then Server Utilities, then click XML Utility. The XML Utility is displayed.

2. Select the relevant server, workspace and CDC agent adapter.

3. Click Events to open the Events listener.

4. Click Start Events in the Events listener. You can specify the event name and qualify it with an application name to provide a unique starting point for the change events.

5. Update a table with data to be captured (for example, using NAV\_UTIL execute). The captured event is displayed in the XML Utility Events listener window. If changes are written to the journal, set the Attunity server environment properties driverTrace and acxTrace in the debug section for the CDC binding. Make changes to the data source and analyze the resulting standard Attunity server logs.