Using Heartbeat Tables to Determine the Latency of Change Data Flow
Abstract

This article describes an end-to-end solution for monitoring the latency of PowerExchange change data flow from a source to a target. You can use this method to assess change data latency for individual PowerCenter real-time CDC sessions and to troubleshoot delays.

Supported Versions

- PowerExchange 9.0.1

Table of Contents

Overview ..................................................................... 2
Configuring Heartbeat Tables to Monitor Data Latency for CDC Sessions. ............................ 2
Analyzing Session Information in the Target Heartbeat Table. ................................. 7

Overview

To determine the change data latency of individual PowerCenter real-time sessions that have PowerExchange change data capture (CDC) sources, you can configure heartbeat tables on the source and target systems. Then include these tables in your PowerCenter CDC sessions to record time stamps that mark key points in the data flow from source to target. By analyzing these time stamps, you can identify and troubleshoot delays.

Include the source and target heartbeat tables within a mapping for a PowerCenter CDC session that you want to monitor. You can use the same source and target heartbeat tables for multiple CDC sessions.

The source heartbeat table contains a single column that is updated with the current system time stamp at a regular interval. You configure the update interval. The CDC session processes this update and adds other time stamps that mark key processing points. Then the rows with this time stamp information for the session are inserted into the target heartbeat table.

When configuring heartbeat table processing in PowerCenter, import the extraction map for the source heartbeat table so that the PowerExchange-generated DTL__CAPXTIMESTAMP column is available in the source definition. In the mapping for an existing CDC session, add the source and target definitions for the heartbeat tables, an Expression transformation, and an Update Strategy transformation. The Expression transformation adds columns for the session name and for the time stamp that indicates when the CDC session processed the source update. The Update Strategy transformation causes the data in all of the heartbeat columns to be passed to the target heartbeat table as inserts instead of updates. By using inserts, you maintain the target heartbeat table information for analysis across multiple updates and sessions.

The target heartbeat table contains a session-specific row for each source update that is captured. Each row contains a column for the session name and four different time stamp columns. You can compare the time stamp values to determine the data latency of a session and to troubleshoot any delays.

Caution: If you use a target database type that does not support row-level locking, such as Microsoft SQL Server, PowerCenter CDC sessions might encounter contention problems when trying to write data to the target heartbeat table. If you use a database type that supports row-level locking, such as Oracle or DB2, these contention problems should not occur.

Configuring Heartbeat Tables to Monitor Data Latency for CDC Sessions

Complete the following steps to configure source and target heartbeat tables and include them in a PowerCenter CDC session to monitor change data latency.
This procedure uses Oracle source and target heartbeat tables that are named PM_SRC_HEARTBEAT and PM_TGT_HEARTBEAT, as an example. If you prefer, you can create the heartbeat tables in other types of databases that PowerExchange and PWXPC support and use different table and column names. However, if you use a different database type, you might need to use different column datatypes.

1. In the CDC source database, create the source heartbeat table.

   Use a table name that implies the table function, such as PM_SRC_HEARTBEAT. The table contains one column named SRC_TSTAMP, which holds the system time stamp that marks when an update occurs.

   For example, use the following DDL to create an Oracle table that has a column with the datatype of TIMESTAMP:

   ```
   CREATE TABLE PM_SRC_HEARTBEAT
   (SRC_TSTAMP TIMESTAMP DEFAULT SYSTIMESTAMP NOT NULL,
   CONSTRAINT PM_SRC_HEARTBEAT_PK PRIMARY KEY (SRC_TSTAMP))
   );
   COMMIT;
   ```

2. In the PowerExchange Navigator, create a capture registration for the PM_SRC_HEARTBEAT table. Set the registration Status option to Active, and set the Condense option to Part.

   PowerExchange generates an associated extraction map. The extraction map includes the generated DTL__CAPXTIMESTAMP column. When you import the extraction map into PowerCenter to create a source definition, this time stamp column is included in the source definition.

3. Create a script that updates the SRC_TSTAMP column in the PM_SRC_HEARTBEAT table with the current system time stamp.

   The script should perform an update instead of an insert to prevent the table from becoming too large. You can use the following sample DDL:

   ```
   update PM_SRC_HEARTBEAT
   set SRC_TSTAMP = CURRENT_TIMESTAMP;
   COMMIT;
   ```

   Schedule the script to run at a regular interval, such as 5 seconds or 5 minutes. The frequency depends on your environment.

   **Tip:** For the script to run successfully, it must have an initial value to update. Before you run the script for the first time, manually perform an insert that populates the SRC_TSTAMP column with current system time stamp.

4. On the target system, create the target heartbeat table.

   Use a table name that implies the table function, such as PM_TGT_HEARTBEAT. The table contains five columns, including a SRC_TSTAMP column that corresponds to the SRC_TSTAMP column in the PM_SRC_HEARTBEAT table. The other columns are:

   - SESSION_NAME. Contains the session name from the PowerCenter CDC session.
   - CAPX_TSTAMP. Contains the DTL__CAPXTIMESTAMP date and time in a format that can be compared with the other time stamps. This value is the time stamp that was recorded by the source database when an update was written to the PM_SRC_HEARTBEAT table.
   - MAP_TSTAMP. Contains the current system time stamp on the PowerCenter Integration Service system when the source update was processed by the CDC session.
   - TGT_TSTAMP. Contains the current system time stamp on the target when a row for a source update is inserted in the PM_TGT_HEARTBEAT table for a session.

   You can use the following DDL to create an Oracle table that includes these columns:

   ```
   CREATE TABLE PM_TGT_HEARTBEAT
   (SESSION_NAME VARCHAR2(30) NOT NULL,
   SRC_TSTAMP TIMESTAMP NOT NULL,
   CAPX_TSTAMP TIMESTAMP NOT NULL,
   MAP_TSTAMP TIMESTAMP NOT NULL,
   TGT_TSTAMP TIMESTAMP DEFAULT SYSTIMESTAMP NOT NULL,
   CONSTRAINT PM_TGT_HEARTBEAT_PK PRIMARY KEY (SESSION_NAME, SRC_TSTAMP)
   )
   ```
Note: In the DDL for the TGT_TSTAMP, the clause "DEFAULT SYSTIMESTAMP not null" causes the target system time stamp to be written to this column. You do not have to create a script to perform this function.

5. In PowerCenter Designer, add the source and target heartbeat tables, an Expression transformation, and an Update Strategy transformation to an existing mapping for a workflow that you want to monitor:
   a. In the Source Analyzer, import the PowerExchange extraction map for the PM_SRC_HEARTBEAT table to create the source definition.
   b. In the Target Designer, create a target definition for the PM_TGT_HEARTBEAT table.
   c. In the Mapping Designer, drag the source and target definitions into a mapping that is associated with a workflow that you want to monitor.
   d. Create and configure an Expression transformation.

   In the Edit Transformations dialog box, click the Ports tab and complete the following actions:
   - To add the SESSION_NAME and MAP_TSTAMP columns, add the following built-in variables in the Expressions field:
     - $PMSessionName. This variable populates the SESSION_NAME column with a session name. Set the Port Type field to Output, use a datatype of string, and set the Precision field to a value that accommodates the length of your session names.
     - SYSDATE. This variable populates the MAP_TSTAMP column with the system time stamp at the point when the source heartbeat data flowed through the mapping. Set the Port Type field to Output, and use a datatype of date/time.
   - To convert the DTL__CAPXTIMESTAMP input values to a format that can be compared to the other time stamps, add the CAPX_TSTAMP as an output port. Then, enter the following TO_DATE function for the CAPX_TSTAMP port:

     TO_DATE(DTL__CAPXTIMESTAMP, 'YYYYMMDDHH24MISS')

   The following figure shows how the Ports tab in the Edit Transformation dialog box should appear when you are done:
e. Create and configure an Update Strategy transformation.

In the Edit Transformation dialog box, click the Properties tab. In the Expression Editor for the Update Strategy attribute, enter DD_INSERT. This function causes all rows to be treated as inserts to the PM_TGT_HEARTBEAT table.

The following figure shows how the Properties tab should appear when you are done:
f. In the mapping, place the Expression transformation and Update Strategy transformations between the source and target definitions. The Expression translation should precede the Update Strategy transformation. Then, link the ports between all elements in the flow, including the source definition, source qualifier, transformations, and target definition.

The following figure shows an example mapping that includes the heartbeat tables in the top flow:

6. In Workflow Manager, configure the connections and session for the workflow that you want to monitor.
   a. Create and configure connections for the source and target databases, if they are not available.
      For the source, use a PWX Oracle CDC Real Time application connection. For the target, use a PowerCenter relational connection.
   b. In Task Developer, double-click the CDC session in the workflow to edit session properties.
      In the **Edit Tasks** dialog box, set the following attributes:
      * On the **Properties** tab, set the **Treat source rows as** attribute to **Data driven**.
      * On the Mapping tab, under **Sources**, click the source qualifier for the PM_SRC_HEARTBEAT table, and then click the pencil icon for the source connection. In the **Connection Object Definition** dialog box, set the **Image Type** attribute to **AI**. This setting causes the workflow to write only after images of change data to the PM_TGT_HEARTBEAT table and to all other targets in the CDC session.

7. Start the workflow that includes the heartbeat tables.

8. Analyze the time stamp data that is written to PM_TGT_HEARTBEAT table for the session.
   As the PowerCenter CDC session moves change data of interest to the target, the PM_TGT_HEARTBEAT table concurrently accumulates time stamp information for the session.

   You can configure other CDC sessions and mappings to write time stamp information to the same PM_TGT_HEARTBEAT table. A table row provides information for a specific session.
Analyzing Session Information in the Target Heartbeat Table

After you start a real-time CDC session that uses the heartbeat tables, the session writes a row of time stamp information to the PM_TGT_HEARTBEAT table each time the SRC_TSTAMP column in the PM_SRC_HEARTBEAT table is updated. You determine the frequency of the source updates.

If you configured multiple CDC sessions to include the heartbeat tables, the PM_TGT_HEARTBEAT table contains a session-specific row for each source update. You can query on a session name to find all rows for a particular session.

To monitor session performance or to troubleshoot suspected data latency issues, compare the time stamp values in a row for a session. Calculate the time difference between two different time stamps to determine the time delays during different processing phases. The following table describes what you can determine by comparing columns and calculating the difference between their time stamps:

<table>
<thead>
<tr>
<th>Compare Columns</th>
<th>To Determine</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC_TSTAMP and TGT_TSTAMP</td>
<td>The total time that the PowerCenter CDC session took to propagate an update end-to-end. Use this time interval as an indication of overall session performance for an update.</td>
</tr>
<tr>
<td>SRC_TSTAMP and CAPX_TSTAMP</td>
<td>The date and time when an update was written to the PM_SRC_HEARTBEAT table. These time stamps should be the same or nearly the same. If they are different, a PowerExchange change capture problem might exist.</td>
</tr>
<tr>
<td>CAPX_TSTAMP and MAP_TSTAMP</td>
<td>The time that the PowerCenter CDC session took to process an update to the PM_SRC_HEARTBEAT table. A large time interval might indicate a problem in either PowerExchange or the PowerCenter CDC session and workflow.</td>
</tr>
<tr>
<td>MAP_TSTAMP and TGT_TSTAMP</td>
<td>The time that the PowerCenter CDC session took to process and insert the change into the PM_TGT_HEARTBEAT table. A large time interval might indicate a delay in workflow processing.</td>
</tr>
</tbody>
</table>

Author

Virginia Pfeifle
Principal Technical Writer

Acknowledgements

The author would like to acknowledge Dan Smith, Lead Technical Support Engineer, and Rickey Link, Technical Support Engineer, for their help in preparing this article.