
IBM InfoSphere Data Replication's CDC DDL Change Management for Oracle

This document describes the new simplified procedure to follow after changes are made to the source table structure. The procedure is applicable for IIDR version 10.2.1 and above. **Note, this procedure does not apply for Rule Based Mappings.**

This document describes the steps to follow to resume replication following a DDL change which results in a table structure change.

CDC maintains and depends on metadata describing source and target tables and columns being replicated. When DDL changes occur which results in the table representation in the log to differ versus the image currently in the CDC metadata (usually due to a change to the source table structures), the metadata will no longer contain the correct information with which to read source database log records to capture change data during mirroring.

For that reason, following a DDL change which results in table structural changes, some action is required to update the metadata before replication can be performed or resumed.

This document describes how to update table definitions to accommodate a DDL change so that replication can proceed.

Planning DDL changes

Ideally, a DDL change that will affect replication of a table or tables that are being replicated should be known and planned in advance. In this case you should ensure that the entire procedure described below is executed during a quiescent maintenance period for both source and target systems where no DML changes occur to the *in-scope* tables (including tables that will not have DDL changes).

DDL Awareness

In the event that DDL occurs in an unplanned manner, CDC has a feature called "DDL Awareness" which identifies DDL in database log entries that affect replication, and ends mirroring at that point in the log using a controlled shutdown.

Examples of DDL which affect replication and will cause replication to shut down due to DDL awareness:

- **Adding/Dropping columns:** The row image in the log file will change as a result, and the log parser would fail if it tried to process a record with the old definition.
- **Adding or dropping out of scope columns:** The engine will detect DDL changes for all table structure modifications. If the columns are not required for replication the above procedure still needs to be followed as well as de-selecting the added or modified column.
- **Modifying column formats** (for example, data type, length, precision, and so on): Changing a column type (CHAR to VARCHAR) will cause the target to receive data differently. In the example, CHAR to VARCHAR, data will no longer be blank padded, so this may affect derived expressions. Additionally, source and target tables will now have data in mixed formats. Older data is blank padded, new data is not. Also, if a column width on the target was reduced (say varchar2(20) to varchar2(10)), it would likely result in a target apply error ORA-12899: value too large for column.
- **TRUNCATE PARTITION** is something of a special case DDL with its own procedure and is covered separately in Appendix B at the end of this document.

In order to support “DDL Awareness” the database must log full "image" data for all columns in a table, including newly added columns. The minimum logging requirements can be found in the CDC Oracle users guide: http://www-01.ibm.com/support/knowledgecenter/SSTRGZ_11.3.3/com.ibm.cdcdoc.cdcdfororacle.doc/concepts/enablingdatabasesupplementallogging.dita?lang=en

DDL Statements that Do Not Affect Replication

DDL statements which do not change the "image" or "log record format" of the log entries for the table will not cause CDC to shut down. Examples include analyzing table statistics and other maintenance procedures such as table re-org and a table re-index which will have no effect on CDC replication.

Table Partition Changes

By default Drop table partition will be treated as a DDL change and detected by CDC's DDL Awareness, after which you would have to follow the DDL recovery steps below. However, if your business need is such that you wish to ignore the drop partition DDL and keep the data on the target, then you can set the following CDC source system parameter:

for_all_table_mappings_history_is_managed_independently_on_target = true

By setting this parameter, CDC will ignore the drop partition DDL and continue replication. However, be aware that from that point forward your source and target data are no longer the same.

Other Table partition changes (such as add partition) do not affect replication for CDC because they do not affect the row image. CDC recognizes new or modified table partition object Ids and continues to replicate changes with no downtime for recovery. CDC does not replicate the DDL changes. To avoid errors on the target, it may be necessary to modify the target table before affected records are replicated.

Behavior of InfoSphere CDC when DDL is encountered

1. The engine detects when DDL that changes the structure of an in-scope table has occurred, and will initiate a normal (controlled) shutdown for a subscription which includes the table. A normal (controlled) shutdown means that mirroring for the affected subscription will shutdown once the target has applied all transactions up until the point of the DDL.
2. The normal (controlled) shutdown must be allowed to run to completion - do not shutdown immediate or shutdown abort or otherwise stop the product during the controlled shutdown sequence. With a normal shutdown, all 'in-flight' operations are allowed to complete and will be applied on the target, allowing the bookmark to advance to the point where the DDL operation was executed. With a shutdown immediate or abort, some or all of these operations may be discarded. On restart, the bookmark from the target may then result in the log reader processing operations on the changed table that occurred before the DDL and a failure will result.
3. If the controlled shutdown does not complete normally, then a Refresh must be performed to bring the table contents back in sync.
4. In order to reduce the risk of a controlled shutdown failure, ensure that InfoSphere CDC is running with little or no latency, and in the ideal case perform the DDL when InfoSphere CDC has no latency and the database is idle, following the steps in *Planned DDL Steps* below.
5. When the subscription has stopped mirroring, the event log will show an error indicating that a DDL change has been detected and shutdown initiated.

To avoid having to do a Refresh, note that you cannot have successive DDL operations intermingled with DML. See the notes in Appendix A for a further description.

Procedure:

The order of the steps is critical for the success of the execution.

Planned DDL Steps:

For a planned DDL, you would first perform the following three steps (A,B,C), and then perform the Unplanned DDL steps.

Step A: For a planned DDL execution, end all DML activity on all of the table(s) that will be affected by DDL.

Step B: Ensure there are no open transactions involving the table that will be affected by DDL.

Here is how you can check open transactions in your system:

The easiest way is to check if there are any transactions is

```
SELECT * FROM gv$transaction
```

This query shows any open transactions in your Oracle system.

However, if your system is complex and you have no control over out-of-scope tables, you may never see gv\$transaction empty. You can run the following query to check if any of the in-scope tables are involved in open transactions and close the transactions, (adjusting the subquery to return object_id values for your in-scope tables.)

```
SELECT t.start_scnw,t.start_scnb,t.start_time,s.username,
o.object_name, o.owner FROM gv$transaction t, gv$session s,
gv$locked_object l, dba_objects o
WHERE l.object_id IN (
select object_id from dba_objects where owner='SCHEMA' and
object_name='TABLENAME'
)
AND t.ses_addr = s.saddr AND t.xidusn = l.xidusn AND t.xidslot =
l.xidslot AND t.xidsqn = l.xidsqn AND l.object_id = o.object_id;
```

Step C: End replication for all running subscription(s) in the instance. Stop all subscriptions at head of log using the scheduled end controlled shutdown

Command line: ./dmendreplication -I <instance> -se <affected subscriptions>

If the shutdown is interrupted, you may have to perform a refresh of the tables affected by the DDL change.

Step D: Follow the 'Unplanned DDL Steps' to complete the procedure

Unplanned DDL Steps:

If the subscription stops because CDC detected a DDL change that altered the structure of the table, it will start a normal (controlled) shutdown, and must be allowed to complete this or a refresh may be required.

When a DDL change has been detected by InfoSphere CDC, the following event will be recorded in the Event Log.

[Event 9505] IBM InfoSphere Change Data Capture has encountered a critical data definition (DDL) change for source table <SCHEMA>.<TABLE> and will shutdown. In order to resume replication for this table you must do the following before restarting the subscription.

- 1) Update the Source Table definition and verify that the resulting table mapping is correct
- 2) Provide a new replication synchronization point by either refreshing the table via the product (Flag for refresh) or setting the table active (Mark Table Capture Point for Mirroring) and manually ensuring that the target table is in synch

Note: If you have a situation where a table is shared in multiple subscriptions and at least one of those subscriptions is using a private scraper, then you would have to wait for the private scraper to hit the same DDL detected error before following the below steps (step 2 onwards). If you do not do this, you would be required to do a refresh of the table for those subscriptions that were using a private scraper.

Step 1. For all the target tables that will be affected by DDL, apply DDL changes before the next step

e.g. alter table <table name> add <colname> <datatype>

Step 2. On Source, update all table definitions, this must be done using the command line

Command line: ./dmreadddtable -I <instance> -t <list of tables> -a

Step 3. On Source, using the command line do a describe

Command line: ./dmdescribe -I <instance> -s <affected subscriptions>

Step 4. On Target, update table definitions for target (only if DDL changes occur on the target side)

Command line: ./dmreassigntable -I <instance> -s <list of affected subscriptions> -t <list of affected tables>

Management Console: Configuration → Datastores view, right click the target datastore and update target table definition. This step will update the target table definition, as well as remap the new columns if the names match.

Step 5. End replication for all subscriptions within the instance, and clear the staging store

Command line: ./dmclearstagingstore -I <instance name>

Step 6. Start mirroring for all subscriptions

Command line: ./dmstartmirror -I <instance> -s <list of subs> or -A (for all subscriptions)

Management Console: Monitoring or Configuration Tab → right-click subscription, choose Start Mirroring

Appendix A: Additional Considerations

Requirement to issue DDL without interleaved DML

DDL has to be controlled so the "log record format" changes do not happen interleaved with DML statements. Meaning, the pattern of operations cannot be a quick succession of DDL DML DDL DML.

All DDL statements must be executed at once with no DML in-between for the duration of controlled shutdown, metadata reconfigure and restart.

When DML occurs in-between DDL statements, a refresh must be performed to get tables back in-sync as there is no other path to recovery.

Example:

A DDL was performed. Product detected the DDL and ended replication.

While replication was down, additional DML operations were performed on the table and then additional DDL.(for example, Insert row, add another column, insert row, then drop a column).

The table definition is now updated within CDC and would be for the latest structure of the table. When CDC encountered the DML in between the DDL changes it would not match to the current table definition and it would fail. There is no other way to recover from this except for a refresh, following which mirroring could resume.

Conditions when a Refresh must be performed

- If there were interleaved DDL/DML as mentioned above
- In the event that logs need to be re-scraped from the time before the DDL change was performed, i.e. the restart position is before the DDL change
- Dropping and recreating tables
 - Dropping and recreating tables would also create a condition where a refresh is required. During mirroring replication, a table's object_id value is used. When a table is dropped and recreated, even though it may have the same name, it will have a different object_id and to CDC is considered a different table from the original.

Table structure changes on the target table only

If there are target only table structural changes InfoSphere CDC will behave as follows:

- Add new column: The new column will not be mapped but be populated with the database default value.
- Delete previously mapped column: CDC will encounter an apply error since CDC attempts to write to a non-existent column. Action: User needs to update table definition on target side and unmap the column mapping
- Alter column definition: CDC may encounter mapping issues because of changed target column definition. Update column mapping as necessary

DDL changes to out of scope columns

The CDC log reader will read and detect DDL changes for all table structure modifications. If the columns are not required for replication the above procedure still needs to be followed as well as de-selecting the added or modified column.

Replicating to a Flat File

Note that if you are replicating to a flat file as the target there are no CDC DDL considerations with respect to the target side of course but you should take into account the effect of the change on the flat file consumer.

When the source table definition is updated, all columns will automatically be selected for replication including any new column that may have been added to the table. The output row in the flat file will include all selected rows when replication resumes.

If the consumer, for example a DS job, depended on the old format of the row it would encounter a problem in processing the output file. If columns have been added, you would need to manually deselect any new columns for replication, or else modify the DS job to accommodate the new row format.

Appendix B: Change Procedure for Oracle Truncate Partition

Oracle's TRUNCATE PARTITION is a special case DDL that has its own procedure for CDC.

The procedure outlined below is a simplified procedure specifically to handle truncate partition DDL. This procedure should only be used if customer can ensure data consistency outside CDC.

Truncate partition DDL change will be logged by Oracle as a simple DDL operation, without any additional information that CDC can use to identify what data has been removed from the source table.

The default recovery procedure for CDC will be to Refresh the affected table.

The following steps indicate how truncate partition operations can be planned in such a way that CDC can skip over the DDL so long as data consistency is ensured outside of CDC.

Procedure:

The order of the steps is important for the success of the execution.

Step 1. Find an appropriate quiet period for the source table on which the truncate partition DDL must be executed.

Step 2. Add an exclusive lock on the table to ensure no DML operations can be executed.

Step 3. Execute the truncate DDL, but do not release the exclusive lock on the table.

Step 4. Subscription(s) where that table is present will stop with an error.

Step 5. Follow Unplanned DDL procedure steps 2, 3, 4, and 5.

Step 6. Release the lock on the table and resume mirroring for the affected subscriptions.