Reduce the source footprint of CDC by querying databases for archive logs

Benefits of InfoSphere CDC Oracle Redo physical standby database configuration

Ranganath Dhanasekaran (rdhanase@in.ibm.com)  
Advisory Software Engineer  
IBM  
21 August 2014

Learn how the Change Data Capture (CDC) Oracle Redo physical standby configuration in InfoSphere® Data Replication allows CDC to read logs off the standby device regardless of the log shipment method. This configuration reduces the source footprint by querying the standby database for archive log information.

Introduction

The InfoSphere Data Replication suite of products includes CDC replication, Q Replication, and SQL Replication. CDC captures database changes as they happen and delivers them to target databases, message queues, or an ETL solution, such as InfoSphere DataStage®, based on table mappings configured in the Management Console GUI application. CDC for Oracle databases supports multiple configurations, including local log reading, remote log reading, and log-shipping configuration. In log-shipping configurations, CDC uses copies of complete archive logs that ship to a secondary system accessible to CDC. The logs are shipped using custom scripts or Oracle’s Data Guard component. However, customers using a combination of custom scripts and Oracle Data Guard are forced to use the CDC manual log shipping configuration, which requires that the log files be registered with CDC.

In an Oracle Data Guard log-shipping configuration, Data Guard is sending and applying archived log files to a real physical standby database. To replicate data, Change Data Capture (CDC) requires information about the archive logs, which are present in the primary database and the physical standby database when the logs are shipped using Data Guard. CDC supports both configurations, wherein it can query the information about archive logs from the primary database or the physical standby database.

This tutorial discusses a few use cases in which CDC with physical standby configuration is suitable.
Prerequisites

Check out InfoSphere
Be sure to explore the available InfoSphere trials.

Basic knowledge about the following is assumed:

- CDC basic replication and configuration of various log-shipping modes such as Data Guard log shipping and manual log shipping.
- Knowledge of CDC commands, such as `dmconfigurets` and `dmts32/dmts64`. (See Resources for more information about CDC commands.)
- Knowledge of the InfoSphere Data Replication Management Console and Access Server.

To follow the step-by-step configuration instructions, the following components are required:

- CDC Oracle Redo 10.2.1 for Linux®, AIX®, HP-UX, or Solaris
- Oracle physical standby database configured with Data Guard log shipping from a primary database
- InfoSphere Data Replication Management Console 10.2.1
- InfoSphere Data Replication Access Server 10.2.1

Use cases

CDC with physical standby database configuration would be the most appropriate configuration for the following use cases:

- When a customer is using a combination of Data Guard log shipping and manual log shipping. For example, when Data Guard services are temporarily suspended (no logs are sent to the physical standby database), the logs generated during this time are moved and registered against the standby database manually. In this case, the primary database has no information about the archive logs registered on the physical standby database. The information in these archive logs is present in the physical standby database. CDC provides two approaches to handle this scenario, as follows:
  - Configure CDC with the manual log-shipping option, which will require registration of the log files with CDC using CDC commands. With manual log-shipping configuration, CDC connects to the primary database only and not to the physical standby database. The information about the log files is provided to CDC by registering the files with CDC manually. However, the manual log-shipping approach requires external scripts to handle the registration and de-registration of log files with CDC.
  - Configure CDC with the physical standby option so that CDC reads the information about the archive logs from the physical standby database. This way, customers can use a combination of shipping logs to the standby using Data Guard log shipping and custom scripts (when Data Guard is down). If Oracle Data Guard log shipping is used, the logs are registered with the database by Data Guard. If custom scripts are used, customers need to register the logs with the database. There is no need to register the logs with CDC in this configuration; CDC reads the information from the standby database directly, which is an advantage of using this...
configuration. However, this is a complex deployment with a requirement that primary and physical standby database instances need to be up for CDC replication to work.

- When a customer needs to reduce the source footprint on the primary database in terms of the number of control file queries executed by CDC. CDC with the physical standby option queries the standby database for information about archive logs, and the CDC source footprint is reduced on the primary database.

**High-level architecture**

CDC Data Guard log shipping with physical standby configuration connects to the primary database and the physical standby database, as shown below. The queries executed against the primary database are minimal. The control file queries CDC uses to process the log files execute against the physical standby database. Because CDC connects to both databases, you must provide the DB credentials such as username, password, and the Oracle SID of the primary and physical standby databases during configuration. The physical standby user needs to be a SYS user to connect to a physical standby database, which is in mounted mode.

**Figure 1. High-level architecture showing CDC and DB interaction**

The physical standby database must be in the mounted mode and not open. The physical standby database must have a TNS name entry defined in tnsnames.ora. The installation of CDC must be on the same machine as the physical standby database.

**Configuration**

This section provides the steps to configure CDC Oracle Redo Data Guard log shipping with physical standby configuration. It is applicable to the GUI version of the dmconfigurets command, but is not applicable to the command-line version.
In this configuration, CDC Oracle Redo is installed on the same server where the physical standby database is configured.

1. Invoke the `dmconfigurets` command in the `cdc_install_dir/bin` folder, which will bring up the UI as shown below. In the Database section, provide the DB TNS name, DB username, and password of the primary database.
   In the Configuration Mode section, select **Log shipping with Data Guard** and click **Advanced configuration**.

   **Figure 2. Configure CDC physical standby configuration**

   ![Configuration UI](image)

2. In the Data Guard section of the Advanced Configuration window, as shown in Figure 3, in the **Archive log directory** field, provide the path where the archive log files are stored on the physical standby database server.
   In the Physical Standby section, select **Connect to physical standby database** and provide the TNS name, SYS username, and password of the physical standby database. Click OK to create the instance.
3. When CDC is configured with physical standby configuration, CDC tries establishing a test connection to the physical standby database with the credentials and Oracle database details you provided. You will be notified of any connection error during configuration with the message "Error connecting to the physical standby using bequeath connection. DB Error:db_error." The connection error can occur due to one of the following reasons:
   - User specified a non-SYS username or password.
   - Physical standby database is not up.
   - Physical standby TNS name entry is not correct.
   - ORACLE_HOME specified is not the right path.
4. After configuring the instance as specified, start the instance using the `dmts64 -I instance_name` command.
5. Create a test subscription using the Management Console with a simple table mapped between the source and the target.
6. Start replication, and CDC will refresh and sync the contents of the target table with the source table and start mirroring. Insert a few rows on the source (primary database) and perform a log-switch operation so the log transfers to the physical standby. CDC will replicate the rows and the target table will contain the rows.

**Conclusion**

In this tutorial, you learned about situations where CDC Oracle Redo physical standby configuration is applicable. This CDC configuration reduces the source footprint by querying the standby database for archive log information. The example configuration is most applicable when a combination of Data Guard and custom scripts are used to ship the archive logs.
Resources

Learn

- Learn more about InfoSphere Change Data Capture, including how InfoSphere Data Replication CDC integrates information across heterogeneous data stores in real time.
- Visit the Knowledge Center to learn about Commands for InfoSphere CDC for Oracle databases. Learn how to use the commands available with InfoSphere CDC to control replication, manage your tables for replication, monitor replication, and perform various other tasks.
- Browse other developerWorks tutorials about Change Data Capture.
- Visit the developerWorks Information Management zone to find more resources for DB2 developers and administrators.
- Stay current with developerWorks technical events and webcasts focused on a variety of IBM products and IT industry topics.
- Follow developerWorks on Twitter.
- Watch developerWorks on-demand demos ranging from product installation and setup demos for beginners, to advanced functionality for experienced developers.

Get products and technologies

- Be sure to explore the available InfoSphere trials.
- Evaluate IBM products in the way that suits you best: Download a product trial, try a product online, or use a product in a cloud environment.

Discuss

- Get involved in the My developerWorks community. Connect with other developerWorks users while exploring the developer-driven blogs, forums, groups, and wikis.
About the author

Ranganath Dhanasekaran

Ranganath Dhanasekaran, who has 15 years of software development experience, has worked on IIDR replication technologies for more than five years. He has worked on several customer issues and is experienced in customer application use cases.

© Copyright IBM Corporation 2014
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)