DATABASE OF MIGRATION AND REPLICATION WITH ORACLE GOLDEN GATE

Suharjito[1]; Raymond[2]; Ida Ratna[3]

[1, 3]Master of Information Technology, Binus Graduate Programs, Bina Nusantara University
Jl Kebon Jeruk Raya 27, Jakarta 11530, Indonesia, harjito@yahoo.com
[2]Computer Science, Faculty of Computing and Media, Bina Nusantara University
Jl Hang Lekir I No. 6, Jakarta 10270, Indonesia

Abstract: The main goal of this research is to analyze and design a database configuration of migration and replication in PT Metro Batavia. Research methodologies used in this research are data collecting, analysis and design model. Data collecting method is conducted with library research and direct survey in the company. Analysis method is conducted by analyzing hangar system, migration and reflection process and the available problems. Design method is conducted by designing a prototype for migration process with the implementation of Oracle SQL Developer and replication process with implementation of Oracle Golden Gate. The result of this research is a prototype for configuration of migration and replication process by using Oracle Golden Gate, which can produce two sets of identical data for the purpose of backup and recovery, and also design a simple tool that is expected to help active-active or active-passive replication process. The conclusion of this research is migration process of MySQL database to Oracle database by using Oracle Golden Gate hasn't been conducted, because Oracle Golden Gate still has bug related to binary log, so database of migration is conducted by using Oracle Golden Gate. However, replication of bi-directional in between database of Oracle by using Oracle SQL Developer can guarantee data availability and reduce work burden from primary database.

Keywords: Analysis; Oracle Golden Gate; Migration; Replication; Bi-Directional

INTRODUCTION

In this globalization era, business is developing very rapidly and more competitive. As with this business development, information technology is also expected to develop and innovate so it can balance development in business. Company and organization will more depend on the most effective and efficient information technology in running their business process including process of storage and processing business information.

Integration between information technology and business process will produce very important information for a company or organization in process of decision-making concerned interests of business of that company. Then, from that guarantee about safety and reliability of an information technology, it becomes an important factor for a company to use offered information technology. However, there isn't any information technology provider companies that give one hundred percent guarantee that by using their information technology, system will run without any obstacles either planned or unplanned like difficulty in integrating data that is using different database, possibility of downtime when integrating the data, high performance from server because of continuous access so there is possibility of damage occurrence

to the server, or there is possibility of unexpected natural disasters. Those disturbances may cause bad impact to process of storage and processing business information resulting loss in a company. Therefore, company wants a system that can keep the continuity of business process, integrity and safety of data storage.

PT. Metro Batavia is a company moving in air transportation field or usually called as airline by using Batavia Air as brand name. In its business process, not all of systems from PT. Metro Batavia are integrated in the same database. Hangar system in this company still uses MySQL database. However, because of the big amount of data and information that must be provided by this system, then company decides to migrate the database from MySQL to Oracle to run this system.

In the development of business in PT. Metro Batavia, data availability and business information produced by hangar system in the company get a big portion in success of PT. Metro Batavia as one of the airlines that is the most interested by society. Then, from that, it needs a hangar system that runs real time so needed data and information can be distributed on time. In addition, PT Metro Batavia also wants a system that provides continuous and availability,

where that system is free from potential disturbance for causing data loss and delay in business process. Because of this awareness, PT Metro Batavia wants precise design of data replication in hangar system, so it can keep the continuity of their business process, integrity and safety of data storage.

The goals of this research are to analyze, to do database migration from MySQL to Oracle by using Oracle SQL Developer and to do active-active configuration of data replication in Oracle database of the result of replication to provide data continually (continuous data availability) in hangar system of PT Metro Batavia by using Oracle Golden Date. Furthermore, this research aims to do analysis of migration business by using Oracle Golden Gate and give solution of configuration of replication that can keep data availability.

While, the expected benefits in the end of this research are: (1) database of hangar system is not MySQL database anymore, but Oracle database, (2) to prevent data loss by doing active-active replication between two databases (primary and backup database). (3) The result of research can be used to develop other systems for application of high availability system.

Literature Review

Oracle Golden Gate allows change and manipulates data in transaction level in middle platform that is more than one and different. Oracle Golden Gate uses modular architecture that finally gives flexibility in extracting and replicating data notes that are chosen, change in transactional and changes in DLL (Data Definition Language) through various topologies, as can be seen on figure 1.
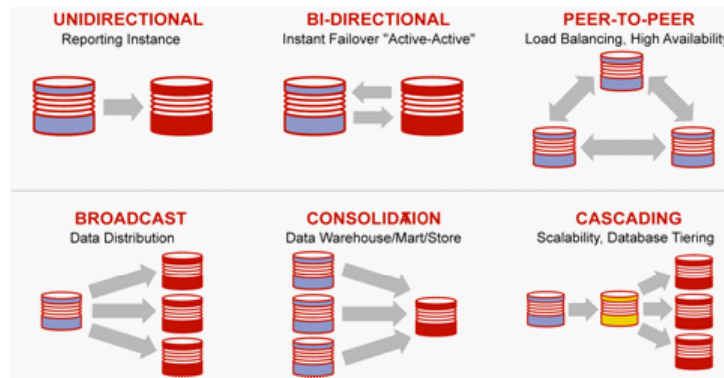

Fig.1: Supported topologies by Oracle Golden Gate

Oracle Golden Gate consists of these components: Extract, Data pump, Replicat, Trails or extract files, Checkpoints, Manager, and Collector, as can be seen on figure 2.
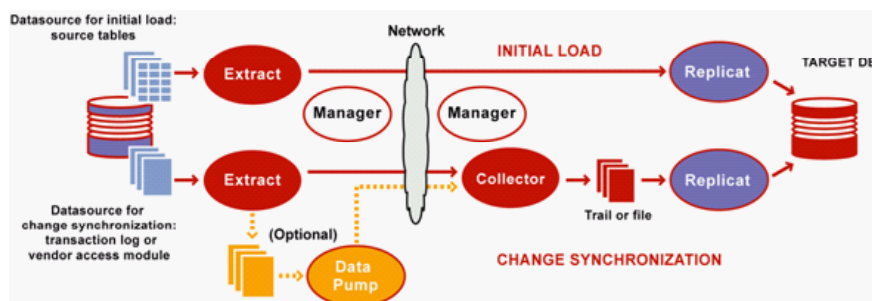

Fig. 2: Components of Oracle Golden Gate

Extract process runs in source system and is extraction mechanism from Oracle Golden Gate. Configuration of extract can be done by: (1) Initial load: for initial data loads, extracts a set of data directly from source object. (2) Changing synchronization: to keep synchronized data source with other sets of data, Extract catches changes done in data (especially insert, update and delete transactional) after initial synchronization has been done, change in DDL and sorting is also extracted if it is according to type of used database.

Extract process catches all changes that is done in configured objects to synchronize. Extract process saves all changes till stage has accepted commit records or rollbacks. When rollback is accepted, Extract throws away data for that transaction. When commit is received, Extract sends data for that transaction to trail process in target system. All log notes for a transaction is written on trail as a transaction unit arranged sequentially. This design holds responsible of data speed and integrity.

Data pump is additional extract configured in source system. If a data pump isn't used, Extract process must send data into remote trail in target system. If data pump is configured, group of primary Extract will write in local trail which is available in target system. Data pump add flexibility of storage and also serve isolation process of primary Extract from TCP/IP activity.

Replicat process runs in target system. Replicat reads change in data extracted and change in DDL (if there is change) specified in Replicat configuration and then replicates back to target database. Configuration of replicat is done by: (1) Initial load: for initial data loads, Replicat can apply data to target object or send it to high-speed bulk-load utility. (2) Changing synchronization: to keep synchronization, Replicat applies change in extracted data in target object using native database interface or ODBC, depends on the types of database. DDL and orders replicated are also applied, if it is according to used database. Replicat applies changes that are replicated by the same order when those changes are committed in source database.

To support extraction and replication process continually from change in database, Oracle Golden Gate saves changes caught to disc in series file that is called as trail. A trail can be in source or even target system, or also in intermediate system, spends on how the configuration of Oracle Golden Gate. In the local system, trail is known as extract trail or local trail. In remote system trail is known as remote trail.

In using trail for storage, Oracle Golden Gate support accuracy of a data and error tolerance. Usage of trail also allows extraction and replication activity appearing freely with other trails. By this separated process, there is an opportunity to arrange how data is sent. For example, rather than extracting

and replicating change continually, we can extract changes continually but save it in trail to replicate in target later, whenever target application needs it.

Checkpoint saves position that is recently read and copied from a process to a disc for recovery. This checkpoint ensures that any changes in marked data for synchronization is extracted through Extract process and replicated by Replicat process, and also prevent repetitive or redundant process. Checkpoint provides error tolerance by preventing data loss, and requiring system, network, or a process of Oracle Golden Gate to start again. For complex configuration of synchronization, checkpoint allows Extract and Replicat process to be more than one to read set of same trail.

Checkpoint works with inter-process acknowledgement to prevent data loss in network. Oracle Golden Gate has technology of incision of guaranteed message. Extract creates Checkpoint in its position in source data and in trail.

A checkpoint system is used by Extract and Replicat process operating continually, but it doesn't need Extract and Replicat process running in batch mode. A batch process can be run back from its start point, where the process continually needs support to interrupt planned or unplanned that is provided by checkpoint.

Manager is a process controller from Oracle Golden Gate. Manager must run in both configured system by Oracle Golden Gate before Extract or Replicat can be started, and Manager must still run while those processes running till management function of resource is done. Manager does functions like these: (1) Monitor and restart process of Oracle Golden Gate. (2) Publishing beginning reports, like when throughput runs slowly or when hidden synchronization increases. (3) Keeping trail and log files. (4) Allocating data storage room. (5) Reporting mistake and incident. (6) Accepting and sending demand from user of interface

Collector is a process running in background in target system. Collector accepts change in extracted database sent through TCP/IP network, and writes it into trail file or extract. Specially, Manager starts Collector automatically when a network connection is needed. When Manager starts Collector, process is known as dynamic Collector and users of Oracle Golden Gate also can be run manually. This is known as static Collector. Not all configurations of Oracle Golden Gate use Collector process.

When a dynamic Collector is used, Collector can received information only from an Extract process, then it must be a dynamic Collector in each Extract process that is used. However, the comparison of one and one is optimal. Collector process is finished when combination of Extract process is finished.

For default, the Extract process starts TCP/IP connection from source system to Collector in

target, but Oracle Golden Gate can be configured so Collector starts connection from target. Initiation of connection from target may be needed if target is in network area that can be trusted, but the source is in the area of network that can't be trusted.

According to Ref. [1], replication is a creation of a or more copy "file system". Replication is used to duplicate all changes that happen in a server called as master server or master, to other server or called as slave server or slave. Two important things from replication is creating backup from main server to avoid data loss if master is damaging and having copy from main server to run reporting and work analysis without disturbing the course of business. Replication like migration or data synchronization is done in database, between source and target.

According to Ref. [1] migration is movement of a "file system" from a server to the other server. According to Ref. [2], database of migration refers to collection of processes and procedures to convert data from a server database to other server database.

The method for data replication between a source and a target consists of: (1) Defining physical data model saved from source and target, each physical model represents plural data structure. (2) Defining logical data model from source and target, each logical model consists of plural nodes and based on data structure from physical related model

METHOD

In this research, writer collects data based on theories supporting research comes from several literatures and interview with company party to know condition and need of company. Based on data that has been collected, writer analyzes company condition and faced problems, decides precise solution to overcome those problems and does design system prototypes that are suggested.

Research that is done by reviewing theories and other scientific data from several literatures that will be made as reference of this thesis writing and others source references like text book and internet website related to research topic.

Research directly conducted to obtain main direct data from company. Survey analysis is conducted in form of direct interview with related parties to get information needed for this thesis writing. Writer interviews directly the Database Administrator of PT Metro Batavia about business process running especially hangar system and company needs.

Analysis is conducted on hangar system running in pt metro Batavia. Writer does also analysis comparison of migration process by using Oracle Golden Gate and Oracle SQLDeveloper and analysis active-passive (live standby) and active-active (bi-directional replication) replication process in Oracle database of migration result [3–17].

Design method that is used in design of this migration and active-passive replication is: (1) Design and implementation of migration using Oracle SQLDeveloper to change MySQL database to Oracle. (2) Design and implementation of active-passive by using Oracle Golden Gate to provide data continually (continuous data availability). (3) While, to ease replication process, then designed simple by using program language Microsoft Visual Basic .NET 2008 Express Edition.

RESULTS AND DISCUSSION

In hangar system, there are three important parts that related to each other, which is Land Maintenance (LM), Base Maintenance (BM), and PPC & MMC Login. LM and BM are in charge to repair plane, check plane condition, repair plane and maintenance regularly for each plane.

Procedure in hangar system starts by making Taskcard by PPC/MMC. Then, from available taskcard, engineer/mechanic will choose and print several taskcard to be done. That taskcard consists of works that must be finished by technician. Before starting doing taskcard, technician must earlier check stock of part and material that will be used. Part and material that will be used listed on printed taskcard. If part and material that will be used is not provided, then technician must request from logistic division. If goods that are needed provided in logistic division then that goods will be directly sent to hangar system. If they aren't, logistic division will do request purchasing good in purchasing division. Purchasing division will process that request and part and material that are requested wull be sent to logistic division that will be sent to hangar division later.

After all parts and materials are available, technician will immediately start to work. Each task that has been finished will be marked. Then after all is done, technician will do scan and save task card that have been finished. Inspector will check whether all works that have been finished are correct. Set of task card that has been done will be made in a work order that consists of works that have been done, start time, finish time as the part of salary calculation of a technician.

For hangar system, three servers fulfilling need of logistic data, login function and hangar data are used. In addition, each of available server has its own database. Database that is used is MySQL. The faced problems are because MySQL doesn't provide dblink facility, then when database is in 2 different servers, it needs to build 2 connections earlier, other than that, MySQL doesn't support inter join and left join facility.

Database of migration process is a process to convert data from a database to other database either in the same environment or the different environment. Environment that is meant is hardware

and software that is used altogether with database. In this process, not only data that is moved, but also available objects in that database like table, view, trigger, stored procedure, function and other objects. There are reasons why a company wants to do migration process in its database: (1) A company wants to upgrade database that they are using with the newer version. (2) A company wants to change their database with other database that is more reliable and can do better process of processing data. (3) Hardware and software that are being used will be changed, so company is required to change their database to higher version or other vendor.

Database of replication process is process of doubling all changes that happen in a server or called as master server or master to other servers called as slave server or slave. In this process, a company can make replica from database that is being used by them as a backup if failure happens on system either intentionally or unintentionally so that company still has copy from main server to run reporting and work analysis without disturbing the course of business.

Several reasons of a company want to do configuration of replication in the main database, are: (1) By doing configuration of replication, data availability so business process is more guaranteed. (2) Configuration of replication also allows to reduce burden of a database because target database of replication can be used to handle reporting and read-only queries that are in basic are not heavy. (3) Company will have backup database that can be used if failure happens planned or unplanned outage.

Problems that are being faced by PT Metro Batavia related with migration and replication process especially for hangar system are: (1) The size of data storage in MySQL database isn't enough to contain the amount of data that must be saved. (2) Processing data process becomes slow. (3) It doesn't support DB-Link feature. (4) It needs tough, cost-effective and can do migration in many platforms. (5) Accessing database excessively (overload transaction) causes processing data to be slow. (6) Hangar system hasn't had backup database

Here are configurations conducted in source database adjusted with requirement of environment adjustment before using Oracle Golden Gate: (1) Do configuration of TCP/IP and DNS on ,etc/hosts. This is done so secondary system can know its hostname. (2) Make sure that 7809 port hasn't been used by command of netstat-na. (3) Create user oraclegg as system user that will be given right to access and used Oracle Golden Gate. (4) Edit .bash_profile from user oraclegg like this, then type ../.bash_profile to activate profile of oraclegg that has been changed. (5) Editing file /etc/my.cnf. (6) Create directory as the source for Oracle Golden Gate. (7) After copying Golden Gate zipin directory /ggs/11.1.1/, do unzip. (8) Give right of directory ggs to user oraclegg. (9)

Give right to read and execute my.cnf file in oraclegg. (10) Give right to read and execute directory, where the binary log is stored in oraclegg. (11) Give right to read and execute tmp directory in oraclegg. (12) Create user oraclegg in mysql. (13) Deactivate firewall and selinux. (14) Then, we can resume GGSCI configuration. Extract tar in /ggs/11.1.1, the result of unzip has been done in the sixth stage. (15) After tar is done, enter as oraclegg and type ggsci. (16) Type create subdirs in ggsci. (17) Then, it is continued to manager configuration. Type edit params mgr and next enter parameters. (18) Manager can be started by command: START MANAGER. (19) Manager status can be seen with command: info manager, info all, or status manager.

In the previous stage, we have created user oraclegg to access database. Next, we will continue the other configurations in target database. (1) Do configuration of TCP/IP and DNS on ,etc/hosts. (2) Make sure that 7809 port hasn't been used by command of netstat-na. (3) Create user oraclegg. (4) Edit .bash_profile from user oraclegg then type ../.bash_profile. (5) Create directory as the source for Oracle Golden Gate. (6) After copying Golden Gate zipin directory /u01/app/oracle/ggs/, do unzip. (7) Give right of directory ggs to user oraclegg. (8) Give those rights to user oraclegg to access database. (9) Deactivate firewall and selinux. (10) Then, we resume GGSCI configuration. Extract tar in /u01/app/oracle/ggs/, the result of unzip has been done in the fifth stage. (11) After tar is done, enter as oraclegg and type ggsci. (11) Type create subdirs in ggsci. (12) Then, it is continued with manager configuration. Type edit params mgr and next enter parameters. (13) Manager can be started by command: START MANAGER. (14) Manager status can be seen with command: info manager, info all, or status manager.

Extract configuration in source database (MySQL): (1) Start MySQL database so Oracle Golden Gate can access database. (2) Add primary extract for primary system. (3) Add ex-trail for primary extract. (4) Edit extract parameter.

Replicat configuration in target database (Oracle): (1) Alter database in archivelog mode. (2) Add supplemental log data. (3) Open database if database hasn't been in open condition. (4) Add replicat for secondary system. (5) Edit replicat parameter

Binary log problem

Until the previous stage, we have done environment adjustment with Oracle Golden Gate, initialization of extract and replication, and its parameter. The next thing that we should do is START EXTRACT ext_1. However, when we run this, then we will get that extraction isn't running and the status is ABENDED, by giving error as shown in Fig. 3.

Error states that Oracle Golden Gate finds that

mysql-bin index binary log isn't in enabled condition. However, in fact in /etc/my.cnf, the configuration on binary log has been done (see Fig. 4).

Fig.3: The error message

Fig. 4: The configuration file

The result of analysis

Writer concludes that there is problem between MySQL or Oracle Golden Gate. There are two possibiities that writer can give: (1) MySQL database does have problem with binary-log, in middle of condition that is still in development stage. Oracle Golden Gate version 11.0.0 was just launched on previous October 2010, so there is also possibility that Oracle Golden Gate is still in development stage. Because of too new software, then there is possibility of bug that hasn't been resolved.

CONCLUSION

Analysis and design of data migration and replication In hangar system is meant to increase performance of hangar system database as the effect of many available transactions to build a useful active-active configuration when outage happens in that database.

After doing analysis and design of this data migration and replication, then it can be concluded that: (1) After doing research of usage of Oracle Golden Gate to migrate database from MySQL to Oracle, writer finds out that there is problem happening during migration, which is problem on binary log in MySQL database. Error happens from Oracle Golden Gate, when accessing status from binary log that has been in enabled condition as disabled. Evaluation of the result of migration research is by using Oracle Golden Gate. (2) Active-active replication (bi-diractional replication) is done to build a configuration between two databases, so those two databases are synchronized to each other and produces two databases that have identical data set. This replication can guarantee data availability in hangar system and reduce work burden to replace primary database. In addition, data set of replication result (secondary database) can be used to replace primary database if outage happens in that database.

Several suggestions that can be given for further development is that the company is recommended to document tables in each available database and business process lane in that company itself, so it can ease developer to design configuration of migration and replication in the future. This thing is based on the writer's difficulty to get structure of hangar system database that will be migrated and replicated.

REFERENCES
[1] J. Munoz, N. Syracuse, *Replication and Migration*, 2000, retrieved from: https://www.ietf.org/ proceedings/49/slides/NFSV4-4.pdf
[2] T, Morales, *Oracle9i Database Migration, Release 2 (9.2) Tony Morales 2002*, United States of America: Oracle Corporation, 2002.
[3] Anonymous, *Oracle SQL Developer 2.1.*, United States of America, Oracle Corporation, 2009
[4] Anonymous, Oracle GoldenGate 11g: *Real-Time Access to Real-Time Information*, United States of America, Oracle Corporation, 2010
[5] Anonymous, Oracle GoldenGate, *Workshop Module*. United States of America, Oracle Corporation United States of America, Oracle Corporation, 2010
[6] Anonymous, *Oracle Migration Workbench*, United States of America, Oracle Corporation, 2010
[7] Anonymous, Oracle® GoldenGate Director *Administrator's Guide 11g Release 1 (11.1.1)*, United States of America, Oracle Corporation, 2010
[8] Anonymous, Oracle® GoldenGate MySQL *Installation and Setup Guide 11g Release* 1 (11.1.1), United States of America, Oracle Corporation, 2010
[9] Anonymous, Oracle® GoldenGate Oracle *Installation and Setup Guide 11g Release* 1 (11.1.1), United States of America, Oracle Corporation, 2010
[10] Anonymous, Oracle® GoldenGate *Troubleshooting and Tuning Guide*, United States of America, Oracle Corporation United States of America, Oracle Corporation, 2010
[11] Anonymous, Oracle® GoldenGate Windows *and* UNIX *Administrator's Guide 11g Release* 1 (11.1.1). United States of America, Oracle Corporation, 2010
[12] Anonymous, Oracle® GoldenGate Windows *and* UNIX *Reference Guide 11g Release* 1 (11.1.1), United States of America, Oracle Corporation, 2010
[13] Anonymous, Oracle® GoldenGate Windows *and* UNIX *Troubleshooting and Tuning Guide 11g Release* 1 (11.1.1), United States of America, Oracle Corporation, 2010
[14] A. Silberschatz, H. Korth, and S. Sudarshan, *Database System Concepts*, 4th ed., New York: McGraw-Hill Companies Inc., 2002.
[15] C. Zeis, C. Ruel, M. Wessler, *Oracle® 11g For Dummies*, Indianapolis: Wiley Publishing Inc., 2009.
[16] D. K. Burleson, *High-Performance Oracle Database Applications*, New York: The Coriolis Group, 1996.
[17] E. Turban, Efraim, R. K. Rainer, and R. E. Potter, *Introduction To Information Technology,* United States of America, John Wiley & Sons Inc., 2005.