Data Migration In Heterogeneous Databases (ETL)

Tribhuvan Shweta M.¹, Rasane Pratiksha V.², Prof. M. R. Bendre³
Department of Computer Engineering

Abstract—Development of economic systems presents rapid growth trend and requires establishing a rapid trans-regional clearing system in banks. We are studying the approaches and processes of data migration and take a bank’s database as an example. The process of data migration has three methods. Data migrated by tools beforehand, data migrated manually beforehand, data generated by new system afterwards. It converts the data from source database to destination database and then migrate it successfully. The database migration is the lack of a well defined low risk and low cost strategy for moving enterprise data from one database to another over time. A high amount of data is being managed by databases and applications in companies today. The process of moving data from one database to another database is the data migration. The data are extracted from different databases and stored to another databases.

Keywords-- Database, Database Conversion, Data Migration, Data Propagation.

I. INTRODUCTION

When a company starts as a small scale industry, MS-Access database is quite capable of handling the records and databases. But when the company expands its database has to expand to more efficient databases. "Database Migratory" is the software aimed at making the conversion of an Ms-Access database to Oracle database in easy and efficient way. It does not require any manual effort and the technical knowledge. A "Database Migratory" is usually developed for individuals and organizations to save time to convert database if a database already exist by creating the tables, views etc. are already in existing database. Data Migration is of transfer the data of storage types, formats and computer systems. This tool is useful when the organizations or individuals change computer systems or upgrade to new systems. It is used by organizations to deal with complex data import, export and migration issues. Importing, exporting and migrating data from different sources is very complicated and time consuming. The data sources store the data in different formats. The existing tool for migration available in the market is oracle migration workbench.

II. IMPORTANCE OF DATA MIGRATION

Database technology is the core technology of information system in each organization. Database technology has been promoted with development of information technology, which includes database management system technology. Due to advancements in information technology, old systems have been replaced by more powerful systems. Data migration is an important issue when it is concerned with the process of updating data. Data which are migrated into new system are huge in size and are prerequisite for new system starting up and basis of decision making. Therefore data migration is the process of data cleaning, transforming and loading into new system. It is applied to switching of old systems to a new system when existing data are to be migrated into new system.

III. DATA MIGRATION TECHNIQUES

Data migration system is uniform in terms of functions and it is a mature technology. Data migration system is considered as a component database system. Migration is divided into three methods, data migrated by tools beforehand, data migrated manually beforehand and data generated by new system afterwards.

A. Data migrated by tools beforehand

The migration tool called ETL (Extract Transform Load) can extract, transform and load data from existing systems to new systems, before switching. There are many tools available in the market for migration- Informix’s Info mover, Oracle’s Oracle Warehouse and Microsoft SQL Server7’s DTS.

This method has high efficiency and it is used most widely. The main precondition for this approach is data should be able to map into new system.

B. Data migrated manually beforehand

There are data migrated into new system manually, before switching. This method has very high cost and error ratio. These data are those, which are required for new system and old system cannot provide.
C. Data generated by new system afterwards

These data refers to the data generated by designed programs and functions after switching.

IV. DESIGN AND EXECUTION OF DATA MIGRATION

A. Selection of platform and language

The procedure of data migration of a bank is to be implemented and this procedure for design and implementation requires a manual work.

We are taking Java as a development platform for development and multiple platforms run because data migration must be able to apply on multiple platforms.

B. Data migration process design

Operation log should be tested when the result error rate is more than expected rate. By adjusting the mapping table and regulating database type compatible with Java, we can solve many errors. Following contents should be checked along with the data:

Format of data: - It is checked for usability and consistency.

Length of data: - Data length should be effective. For example, data field width is checked when data is converted from char type to archer type.

Interval range: - Ensure that data is allocated between predefined minimum and maximum interval.

Null value and Default value: - Different databases have different null and default value definitions. Null value and default value in old system may be inconsistent with the corresponding values in new system, so it should be checked.

Integrity: - Data association integrity is checked.

Data consistency: - Data that violate logical inconsistency should be eliminated.

Different function modules are designed including JDBC drive module, User connection and test module, source/object database and JDBC mapping module, Extraction of database list structure module, SQL sentence preparation module and data migration module. Following diagram shows relationship between these modules.

Source/object database type and JDBC mapping module: It includes the relationship between data type of JDBC interface. Purpose of this module is to find matched type in JDBC or to limit the errors in an allowable range and finally meet user’s requirements. Embedded data type of JDBC with one of Access, Oracle, Microsoft SQL Server 2000, DB2 and MYSQL, Data Types of JDBC are divided into numeric, character, big object, date and special.

A. Query and Extraction of database structure

Purpose of this module is to load the source database pattern and table structure information into middle database and then generate SQL sentence of source database. For creation migration sentence, this information of middle database is used. To load source database information into middle database extraction adopts the preorder traversal method.
B. Database Query

This module provides standard database query functions, including database information of active connection, database pattern and table information; field information of table, sentence that generates table and data within the table. Also the mapping relationship between data and JDBC, sending and implementing SQL sentence.

This module functions similar to database browser. Interface module controls all related action logic. Related parameters, initiation and synchronization operation are provided by background.

Private static synchronized database cache();
Synchronized Connection get();
Public static synchronized Database JDBC Driver get();
Synchronized Thread get();
Protected synchronized Object get Value();
Synchronized void start Timer();

Synchronization aims at global variable Multi-Migration, migration and inquiry can operate in parallel for multithreading system. Global variable’s atomicity must be preserved. Global variables are Database, Connection, JDBC Driver, Thread and Timer. Synchronization is achieved by the keyword ‘synchronized’ that is provided by JAVA.

C. Migration sentence preparation module

This module generates SQL sentence. Data types of middle database are converted into those of object database according to mapping table and it can be spliced into SQL sentence of object database. SQL sentence can be of three types: table creating sentence of object database, ‘insert’ sentence for inserting data and ‘select’ sentence for reading data. Splicing of table has three steps:-

- Create basic ‘create’ sentence.
- Add ‘alter’ sentence of primary key.
- Add ‘alter’ sentence of foreign key.

One-time splicing sentences are ‘select’ and ‘insert’. Splicing of ‘create’ sentence requires more steps in order to achieve more flexibility on the migrations of primary key and foreign key. Users may suffer from poor usability due to inconsistency of SQL syntax. In reference to standard SQL and comparison of the SQL syntaxes of all required databases, minimum intersections are found.

CREATE TABLE name (column_name COLUMNTYPE [NULL/NOT NULL], column_name COLUMNTYPE [NULL/NOT NULL], …)

ALTER TABLE table_name ADD PRIMARY KEY (column_name, column_name, …)

ALTER TABLE table_name ADD CONSTRAINT column_name FOREIGN KEY table_name(column_name)
SELECT(column_name, column_name, …)
INSERT INTO table_name VALUES(?, ?, …)

Data migration is divided into two steps. First, tables to be migrated along with primary key and foreign key are created in source database. Then data from this source database are migrated into object database.

When table is created in object database, all table names should be written into newly createdTables. Before data are migrated to object database, names of migrating modules will be read from createdTables.

private List createdTables= new ArrayList();
Migration of tables is similar to that of primary key and foreign key. The migration begins after creating all tables, primary key and foreign key in object database.

V. DATABASE MIGRATION STRATEGY

Database technology in the information system is of the data and code conversion to assist present applications running on new database generation .The relational databases are mostly used for end-user computing and developing new business functions. Most of the applications are used for several database management systems such as hierarchical network or relational databases. The transition from one database to another database, system generation is important. Few promising migration strategies exist.

The database migration means the process of moving from one database to another database technology without manually rewriting existing applications. Two technical problems need to be solved. First, it is decided which database system is target and how data are transferred to it; and Second, how application are converted smoothly without affecting availability or performance. It is important not to forget that the significant investments already put into both data and application.

This paper gives the practical solutions of the database migration described by three important software strategies: data and code conversion language transformation and data propagation. Data propagation is to maintain the consistency on the database of different systems. The evolution of database management systems. Suppose system E and system F is the identical database technology. In this software strategies used are data and code conversion as well as data propagation. For example, in relational technology delta changes from a relation of system A should be propagated to the suitable relation of system B, either synchronously or asynchronously.
Source code conversion is used for application programs. The benefit of migration strategy is that it needs only single copy of data. It is converted to source code, which is not very easy to maintain and has been known to exhibit performance difficulties. Language transformation allows the mapping of one database to another database language and vice-versa. For Ex. From EQL to FQL or from FQL to EQL. The language transformation direction is important. Mapping from EQL to FQL leads the target database system F which is in opposite direction. Then the existing database technology E. Data propagation targets to maintain the compatibility between two or more data copies of different way of database system by propagating changes from one copy to another. It avoids the determined attempt and involved risks converting and interfacing with good applications.

For the three migration strategies are pertaining to the products or partial solution can be found to the market which is the importance of database migration for the occupation.

VI. SYSTEM TESTINGS AFTER DATA MIGRATION

The system should check the migration quality whose results are important for deciding whether to startup the new system [7].

After the data migration, following system tests should be performed [8]

Completeness Check: The existence of referenced foreign keys is checked.
Consistency Check: To check same meaning data have consistent values in bit.
Total Balance Check: It is same as the depositors’ balance check.
Records Count Check: Consistency of records counts in new system and that in old systems is checked.
Check for Special Sample Data: Consistency of same sample data in new system and that in old system is checked.
System Integrity Check: CPU speed, memory capacity and migration time is checked.

The investigation of the data difference between new system and old system is to investigate the same data by every inquiry tool and compare the conclusion [9].

VII. CONCLUSION

A “Database Migration Suite” is usually developed for individuals and organizations to save time for converting to a new database if a database already exists.
The purpose of our project is to develop tool that helps in migrating database structures and data across various relational databases. It will also facilitate migrating data from MS-Access, Excel Worksheet to SQL server. This paper is based on database migration project and it introduces technique issues of data migration that involve manual work.

REFERENCES