ISSN (ONLINE) : 2395-695X
ISSN (PRINT) : 2395-695X
Available online at www.ijarbest.com
International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 9, December 2015

A Safety Data Migration of Argus Application

Aarthi M[1], Malathi K[2]

Student, Database Systems, Indian Institute of Information Technology, Srirangam, Tiruchirapalli, India [1] Faculty, Database Systems, Indian Institute of Information Technology, Srirangam, Tiruchirapalli, India [2]

***Abstract*—Many established organizations have their legacy information that are too expensive to maintain and difficult to modify. Legacy systems are those have some old and important information from decades to years and it needs to be maintained properly in low cost and in an effective way. Several hurdles may happen when conversion of legacy system to modern system where data extraction and data cleansing need to be done. Extracted data has to be cleaned only after analyzing the nature of the data.The metadata are to be analyzed by improving the quality of data during migration. Profiling phase eliminates the unwanted data and the data in the source table is mapped to the target table by either one to one mapping or many to one mapping or many to many mapping. SQL code has to be developed to migrate or move the data stored in source table to target table. By validating the scripts, errors can be detected and verified according to the requirements. The scripts are executed and the data are successfully migrated.**

***Index Terms*—Data Extraction, Data Cleansing, Data Migration, Legacy Information**

I. INTRODUCTION

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, and images. databases typically contain collections of data records or files, such as sales transactions, product catalogs and inventories, and customer profiles. Typically, a database manager provides users the capabilities of controlling read/write access, specifying report generation, and analyzing usage. Databases and database managers are prevalent in large mainframe systems, but are also present in smaller distributed workstation and mid-range systems and on personal computers. SQL (Structured Query Language) is a standard language for making interactive queries from and updating a database such as IBM's DB2, Microsoft's SQL Server, and database products from Oracle, Sybase, and Computer Associates.

A. Relational Databases

This is the most common of all the different types of databases. In this, the data in a relational database is stored in various data tables. Each table has a key field which is used to connect it to other tables. Hence all the tables are related to each other through several key fields. These databases are extensively used in various industries and will be the one of which most likely to come across when working in IT.

B. Operational Databases

In day to day operation, an organisation generates a huge amount of data. Think of things such as inventory management, purchases, transactions and financials. All this data is collected in a database which is often known by several names such as operational/ production database, subject-area database (SADB) or transaction databases.

C. Distributed Databases

Many organisations have several office locations, manufacturing plants, regional offices, branch offices and a head office at different geographic locations. Each of these work groups may have their own database which together will form the main database of the company. This is known as a distributed database.

D. Database Management Systems

DBMS are software applications that interact with the user, other applications, and the database itself to capture and analyze data. A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of databases.

Database management systems are often classified according to the database that they support; the most popular database systems since the 1980s have all supported the relational models represented by the SQL language. Sometimes a DBMS is loosely referred to as a 'database'.

E. Database Warehouses

Organizations are required to keep all relevant data for several years. In the UK it can be as long as 6 years. This data is also an important source of information for analysing and comparing the current year data with that of the past years which also makes it easier to determine key trends taking place. All this data from previous years are stored in a database warehouse. Since the data stored has gone through all kinds of screening, editing and integration it does not need any further editing or alteration.With this database ensure that the software requirements specification (SRS) is formally approved as part of the project quality plan.

ISSN (ONLINE) : 2395-695X
ISSN (PRINT) : 2395-695X
Available online at www.ijarbest.com
International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 9, December 2015

II. APPLICATION AREAS OF DBMS

A. Banking: For customer information, accounts, and loans, and banking transactions.

B. Airlines: For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner - terminals situated around the world accessed the central database system through phone lines and other data networks.

C. Universities: For student information, course registrations, and grades.

D. Credit card transactions: For purchases on credit cards and generation of monthly statements.

E. Telecommunication: For keeping records of calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about the communication networks.

F. Finance: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds.

G. Sales: For customer, product, and purchase information.

H. Manufacturing: For management of supply chain and for tracking production of items in factories, inventories of items in warehouses / stores, and orders for items.

I. Human Resources: For information about employees, salaries, payroll taxes and benefits, and for generation of paychecks.

III. DATA MIGRATION

Data Migration is the process of transferring data from one system to another while changing the storage, database or application. In ETL (Extract-Transform-Load) process, data migration always requires at least Extract and Load steps. Typically data migration occurs during an upgrade of existing hardware or transfer to a completely new system. Examples include migration to or from hardware platform upgrading a database or migrating to new software or company-mergers when the parallel systems in the two companies need to be merged into one.

A. Storage Migration

Storage migration can be handled in a manner transparent to the application so long as the application uses only general interfaces to access the data. In most systems this is not an issue. However, careful attention is necessary for old applications running on proprietary systems. In many cases, the source code of the application is not available and the application vendor may not be in market anymore.

B. Database Migration

Different data types can be handled easily by approximating the closest type from the target database to maintain data integrity. If a source database supports complex data formats (e.g. sub-record), but the target database does not, amending the applications using the database is necessary. Similarly, if the source database supports different encoding in each column for a particular table but the target database does not, the applications using the database need to be thoroughly reviewed.

When a database is used not just as data storage, but also to represent business logic in the form of stored procedures and triggers, close attention must be paid when performing a feasibility study of the migration to target database. Again, if the target database does not support some of the features, changes may need to be implemented by applications or by middleware software.

C. Application Migration

In the case of application migration process is straight forward and however, is extremely uncommon. The reason is that the applications, even when designed by the same vendor, store data in significantly different formats and structures which make simple data transfer impossible. The full ETL process is a must as the transformation step is not always straight forward. Of course, application migration can and usually does include storage and database migration as well. The advantage of an ETL tool in this instance is its ready-to-use connectivity to disparate data sources/targets.

IV. DATA EXTRACTION

Data Extraction is the act or process of retrieving data out of usually unstructured or poorly structured data sources for further data processing or data storage. Data extraction is where data is analyzed and crawled through to retrieve relevant information from data sources like a database in a specific pattern. Further data processing is done which involves adding metadata and other data integration another process in the workflow.

V. DATA LOADING

Data in a warehouse may come from different sources; a data warehouse requires three different methods to utilize the incoming data. These processes are known as Extraction, Transformation, and Loading (ETL). The process of data extraction involves retrieval of data from disheveled data sources. The data extracts are then loaded into the staging area of the relational database. Here

ISSN (ONLINE) : 2395-695X
ISSN (PRINT) : 2395-695X
Available online at www.ijarbest.com
International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 9, December 2015

extraction logic is used and source system is queried for data using application programming interfaces. Following Data loading is part of a larger and more complex component of the Data Warehouse architecture called Data Staging. Complex programming is often involved in data staging. This component also often involves analysis of quality data and filters which can identify certain patterns and data structures within the existing operational data.

VI. DATA PROFILING

Data profiling is the process of examining the data available in an existing data source (e.g. a database or a file) and collecting statistics and information about that data. Data profiling is an analysis of the legacy data sources for a data warehouse to clarify the structure, content, relationships and derivation rules of the data.

Profiling helps not only to understand anomalies and to assess data quality, but also to discover, register, and assess enterprise metadata. Thus the purpose of data profiling is both to validate metadata when it is available and to discover metadata when it is not. The result of the analysis is used both strategically, to determine suitability of the candidate source systems and give the basis for an early go/no-go decision, and tactically, to identify problems for later solution design, and to level sponsors' expectations.

Data profiling is performed several times with varying intensity throughout the data warehouse developing process. A light profiling process should be undertaken as soon as candidate source systems have been identified right after the acquisition of the business requirements for the DW/BI. The purpose is to clarify at an early stage if the right data is available at the right detail level and anomalies

VII. SCOPE OF THE PROJECT

Patient Safety has become a critical component among the many core processes of Life Sciences and Pharmaceuticals due to ever increasing Regulatory & Compliance expectations and continuously evolving guidelines. It is also becoming imperative to minimize risks associated with carrying out complex clinical trials involving huge budgets running into more than a decade. Since the number of data running on the system maintaining in an Argus application having huge memory for retrieval of the data from database will be an important issue.

For that, migration process has been carried out to make the patients safety a trustful one which will suit in Life Sciences requirements. As the database is upgrading in its new version every time, an effective approach is needed to make the migration as safety one.

this process, the data is now ready to go through the transformation phase of the ETL process.

ARGUS Application is a web-based data management System that automates and seamlessly manages data control process to ensure compliance with international regulatory authorities like FDA (USA), EMEA (European Union), Health Canada (Canada), TGA (Australia), and HSA (Singapore)and also FDA follows safety information and event reporting system which is carried out by Med Watch Form 3500 A and CIOMS I Form (EUROPEAN UNION).

Argus application is an web based data management system where everypatients details has been stored and the stored data is important. The Argus application module is designed for organizations maintaining patient records in accordance with United States Food and Drug Administration (USFDA) regulations and other federal laws concerning the use of computerized systems for data management.

A. Load and Cleanse

Loading of data is done by creating staging schema where the actual data is processed to migrate from source table to the target table. Input data is pulled from the dump database and placed into the server which the data is cleaned by the profiling process.

B. PROFILE

Profile helps to clear noisy data from the source database as the data input are pulled from the dump database where there is a chance that data may polluted with impure data. It is done by categorizing the data in an ordered way and using of SQL queries to extract those needed data.

C. MAPPING

Cleaned data is now analyzed and enters into a mapping process where the source and target table are mapped with one to one mapping. All the data in a source table are mapped with the respective tables in the target.

D. RULES

Mapped data are written in hardcore which is usually a draft query used for writing SQL query. Each and every table in target table is written as hardcore value. After that, queries are written and the query scripts are executed in staging table.

E. VALIDATION

In validation phase, the executed scripts are tested for back end and it is compared with the front end as the migrated data are checked as the data are correctly populated in the application.

ISSN (ONLINE) : 2395-695X
ISSN (PRINT) : 2395-695X
Available online at www.ijarbest.com
International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 9, December 2015

VIII. EXPERIMENTAL RESULTS

A. PROFILING

Profiling basically a cleaning of unwanted data which is collecting the specific data or values in source table comparing it with target table. Profiling can be done by taking several tables in both source and target analyzing it and know how exactly or accurately can map both tables without missing any important data or without any flaws.

If the source table has the column as Fetal Outcome which has the Congenital Information where its id is coded as Birth Defect in the target table. Congenital Information is the term referred as birth defect which means child born with defect.

Likewise all the source table columns are need to be checked and mapped according to their values. This type of mapping is called as code list mapping. Each and every table is checked in order to be coded or not. Code lists represents as the code values for the columns such as Countries, Pregnancy Outcome, Ethnic Origin, Cause of death. Each column has different code values.

B. CODELIST VALUES

Code values for the countries are designed as CODE_LIST_ID = 1. Within that, countries have large number of values and each value has unique id which means unique country id. The code list values are coded with their respective source and target tables and stored as LM_COUNTRIES. All the code list values are needed to be mapped and analyzed.

By the time code list values are analyzed and mapped, what are the tables as the source are available to migrate those in target table are to be found. Finding how many source table columns are needed to be migrating into the target tables. To find what the tables are need to be migrated and the records if available in each and every table are executed by the simple Sql Query in Sql Developer, where the number of records in each tables with or without null values are found.

In migration process, both source and target tables have parent table from which the other tables are classified. From the parent table, there is a child table for parent and child has a leaf node. One parent may have multiple child tables and child table may have multiple leaf nodes. The tables are processed in order so that no data are missed in any tables. Each child and leaf tables are linked with every other parent table.

C. TABLE COUNT

As the parent tables are processed with certain order, all the records need to be calculated to find records either available or not in the source tables by taking “Table Count” as the calculation. These can be executed as simple Sql Select Queries. Table count has been calculated based on the Frontend Application which is a Argus Console by creating dummy cases in the application. By doing this the values coded are monitored and stored in the correct form both in back end and frontend.

During the profiling of data, identified one scenario that several columns have no values in source tables but they have a space in the target table. A challenge may occur during the table count whether this table should move to the target or not. Since there are no values in source tables cannot be migrated in target table. So it is decided not to move the empty value to target table.

D. DATA MAPPING

After arranging those available data in source and target table, move for the next phase called as Data Mapping. Meanwhile a Query Tracker has been maintained if having any queries or customer have any queries; both can share their doubts and views in one platform.

Profiling the data from resources is splitted where the data are cleaned by removing unwanted values such as Null or Missing values. After Cleaning, collected data are taken for data mapping where the data have to undergo two processes. These processes are able to make the mapping process a better one. Rules are able to give idea for the writing of scripts which has been useful while writing the Sql Query which means they are written as hardcore values. Then the hardcore values are tested whether they works properly or not. After these two processes only mapping is done by comparing the source and target tables.

E. VALIDATION

Similarly, the legacy systems, (i.e.) data given by customers has been cleaned and mapped accordingly and placed it on the staging table. In this staging table, Validation of data has been performed by running the environment as Dry Run in Validation environment. For Validation, several scripts had to be written. These scripts have different type of scenarios that need to be checked on the validation environment. These written scripts must get passed in all scenarios. Else, the scenarios are treated as fail and issues have to be raised. These issues need to be fixed within particular period of time. If only all the scenarios get passed, the data can be moved to development environment where there is a real time environment and the production has to be done. For that, scripts in validation environment need to be passed without any issues.

IX. CONCLUSION

In this work, a source data are checked cleaned the unwanted impurities and trustful data are presented for the migration. Data are cleaned using manual process by arranging the specific tables which is important for migrating those data in a safety manner. Though this processing is difficult by checking the data manually, it is the best process when compared to schema migration tools which it produces bugs as the major error. This type of migration is trustful and

ISSN (ONLINE) : 2395-695X
ISSN (PRINT) : 2395-695X
Available online at www.ijarbest.com
International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST)
Vol. 1, Issue 9, December 2015

useful toimplement in migration having complex data sets.

Further, the cleaned data are used for mapping where they will create the hardcore values and those values are being checked. After that, the coding is developed based on the hardcore values and the developed scripts are compiled in a Staging table. After that errors are identified and detected. Debugging is done in bug-fix environment and the errors are cleared. Finally the data in the staging table has been moved into the real environment and the project will go lively.

X. REFERENCES

[1] Haikun. Liu, C.-Z.Xu, H. Jin, J. Gong, and X. Liao, "Performance and energy modeling for live migration of virtual machines," in Proc20th Int. Symp. High Perform. Distrib.Comput., 2011, pp. 171–182.

[2] **Tiago**.Ferreto, M. Netto, R. Calheiros, and C. De Rose, "Server consolidation with migration control for virtualized data centers," FutureGenerationComput.Syst., vol. 27, no. 8, pp. 1027–1034, 2011.

[3] Benjamin. Speitkamp and M. Bichler, "A mathematical programming approach for server consolidation problems in virtualized data centers," IEEE Trans. Serv. Comput., vol. 3, no. 4, pp. 266–278, Oct. 2010.

[6] V. Makhija, B. Herndon, P. Smith, L. Roderick, E.
Zamost, and J. Anderson, "VMmark: A scalable benchmark for virtualized systems," VMware Inc, Palo Alto, CA,
USA,Tech. Rep. VMware-TR- 2006-002 2006.Tong. Chen, J. Lin, X. Dai, W.-C. Hsu and P.-C.Yew."Data dependence profiling for speculative optimizations".In Compiler Construction, Lecture Notes in
Computer Science. 2004.

[7] Hillson, Dave, (2000, September) "Project Risks, Identifying Causes, Risks and Effects", PM Network.