

A Survey on Automated Data Migration Tools and Techniques for Heterogeneous Relational Databases

Dr. Dhaval Joshi¹

¹*Department of ICT, Veer Narmad South Gujarat University, Surat, Gujarat, India
(E-mail: joshi.dhaval@hotmail.com)*

Abstract—Data migration is now a crucial task for wide variety of domains to fulfill different industrial needs. It can be used for data synchronization, data warehousing, application migration from one platform to other or to migrate data from legacy application to new application. It is obvious a tiresome and lengthy task to migrate data between heterogeneous data stores as they may contain heterogeneity at different levels. In this paper, some of those known tools and techniques are discussed which are designed or developed for database oriented data migration between heterogeneous data stores. From the review of those tools and techniques, observations are described in research gap analysis which shows that there is lot of work still needed to automate data migration process between heterogeneous databases.

Keywords—Data Migration, Schema Matching; Heterogeneous Relational Database

INTRODUCTION

Data migration is a process of migrating data between two different data storage types. It is highly critical task for various real world requirements like data synchronization, data warehousing, application migration from one platform to other or to migrate data from legacy application to new application. To perform migration from one data storage to other, proper mapping between data storage is essential.

Generally, application's data is stored either database or XML or any other flat file storage. Hence, there is a need to migrate data between two data stores. Furthermore, it is a tedious and time consuming task to perform schema matching and/or data migration manually to migrate large scale data and it also become cumbersome task when both data storage type is heterogeneous in nature.

In order to provide automated solutions for data migration, various tools and techniques are evolved to ease this process. In this paper, some of existing automated or semi-automated data migration tools and techniques are discussed in the literature review section.

LITERATURE REVIEW

There have been several tools and techniques exist for data migration between heterogeneous data storages. In this paper, special focus is given to some of those tools and techniques which are designed and/or developed for relational databases.

ESF [1] is a database migration toolkit from EasyFrom which can migrate database to/from any of the database

formats like MySQL, PostgreSQL, SQL Server, MariaDB, Oracle, IBM DB2, IBM Informix, Microsoft Access, Microsoft Excel, Visual Foxpro, InterSystems Caché, SQLite, FireBird, InterBase, Teradata, Paradox, Lotus, dBase, CSV/Tex, etc. All of the table structures and data along with its constraints can be migrated from source to target database using step by step wizard by manual selection.

SwisSQL [2] is a complete Data Migration Tool from AdventNet that helps the migration of database structures and data across leading databases such as MS-SQL Server, Oracle, PostgreSQL, IBM DB2, Sybase, SAP DB and MySQL. It offers an open, user-friendly, and extensible migration process ensuring reliability and data integrity. It also offers comprehensive Database Migration Solutions to migrate Stored Procedures, SQLs, schema, and data across multiple databases. User need to select source database and target database. Once both databases are connected, user needs to select objects from source database to migrate them to the target database. After selection of objects from source database, it migrate data by creating objects at target database.

CUBRID Migration Toolkit - CMT [3] is a software tool which allows migrating data from MySQL, Oracle or CUBRID (previous versions) to CUBRID Database Server. The data and schema in the source database can be as sophisticated as possible. CMT provides the default settings to cast native MySQL and other DBMS data types to native CUBRID data types, though most of them will overlap as CUBRID provides over 90% MySQL compatibility. If it is necessary to have the destination column data type different from the default settings, it can be easily customized before the migration process starts.

OpenDBcopy [4] is a utility to migrate data from and to any JDBC driver supported database. It is an application framework providing everything required for running a plugin or a series of plugins for migration process. It contains GUI based migration steps where user need to give source and target database information and then validate source and target database tables - fields subsequently to start migration process.

Data Migration Assistant - DMA [5] provides features to upgrade old data platform to a modern data platform by identifying compatibility issues that can impact database functionality on user's new SQL Server version. Performance and reliability improvements for target environment are recommended by it while migration. It supports source database from any version out of SQL Server 2005, SQL Server 2008, SQL Server 2008 R2, SQL Server 2012, SQL Server 2014 and SQL Server 2016. It supports target database from any versions out of SQL Server 2012, SQL Server 2014,

SQL Server 2016 and Azure SQL Database (assessment only). It helps user to migrate schema of databases, data, users, server roles and SQL/Windows logins from source to target environment.

Oracle SQL Developer Migration Workbench [6] is product from Oracle helps migrating database from third party to Oracle database. For migration from third party database, first user needs to select source database with its required credentials. Then the tool will capture database metadata and convert it to Oracle format objects. From that DDL statements are generated to create new Oracle target database with identical schema using Oracle compatible types. After availability of target database, it starts migration of data from source database to target database.

The AWS Schema Conversion Tool [7] by Amazon can automatically convert the source database schema and makes heterogeneous database migrations easy by formatting maximum custom code (like views, stored procedures, and functions) compatible with the target database. Using this tool, existing database schema can be converted from one database engine to another. Converted schema is suitable for an Amazon Relational Database Service (Amazon RDS) an Amazon RDS PostgreSQL DB instance, an Amazon Aurora DB cluster, MySQL DB instance or an Amazon Redshift cluster. It provides a project-based user interface to convert the database schema of user's source database automatically into a format compatible with target instance of Amazon RDS. This tool also provides schema conversion guidance to create equivalent schema in target Amazon RDS database when it cannot convert automatically.

Flyway [8] is a popular open source database migration framework for Java. It brings structure and confidence to the evolution of user's database schema. It is really easy to use, yet powerful and both developer and DBA-friendly. Flyway supports the industry's most common databases including Oracle, SQL Server, DB2, MySQL, PostgreSQL, Hsqldb, H2 and Derby. Migrations can be written in plain old SQL or Java and can be executed through the API, the Maven plug-in, the Ant tasks or the Command-Line tool.

Bogdan et. el. [9] designed a fuzzy expert system based tool for processing data migration between different relational database management systems (RDBMS). The expert system contains a knowledge base which is composed of If-Then rules and based on the input data suggests appropriate data types of columns of database tables. The user need to select datatype of each filed of tables manually from the suggestions provided by expert system. After completion of selection process, the tool migrate entire database to the target environment by creating identical or compatible schema.

Shinde et al. [10] presented a semi automated data migration system for heterogeneous database. In this system, user need to select source database and target database first and then select all required information of source database like tables, columns, constraints, views, etc. After selection from source database, system starts automatic migration to target database by creating all tables, columns, constraints, views, etc.

as per target database type. Here, schema matching is not required as target database do not have any designed schema.

Maatuk et. el. [11] designed a solution which takes an existing RDB as input, enriches its metadata representation with as much semantics as possible, and constructs an enhanced Relational Schema Representation (RSR). Based on the RSR, a canonical data model is generated, which captures essential characteristics of the target data models that are suitable for migration. Their prototype of solution successfully migrates RDBs into object-oriented, object-relational and XML databases using the canonical data model.

Bhatt et. el. [12] proposed a model to migrate data between cloud and non-cloud data stores. In their model, they used RDF as an intermediate model to migrate from one data store to other. The source and target data stores can be of any type from cloud based or non cloud based environment. For experiment they took any relational database type for non-cloud based data store and Google App Engine – Bigtable as a cloud based data store. They performed successful migration from cloud to non-cloud and vice-versa by keeping the same schema structure from source to target.

Bogdam et al. [13] presented an expert system based data migration system between different database management systems. By taking metadata of source database and getting type of target database, using knowledge base of datatypes suggestions are provided to user for finalizing datatype for target database. Once target database datatypes are conform by user, tables are created using selected datatypes and migration is performed from source to target database.

Elamparithi et. el. [14] analyzed various database migration toolkits and listed them as shown in Table-1.

RESEARCH GAP

After reviewing various existing tools and techniques for data migration between heterogeneous relational databases, their derived summary is listed in Table-2. Here, they are reviewed with six different characteristics like type of tool or technique, type of schema matching, data model(s) supported for source type, data model(s) supported for target type, designed for source Vs target data model and is there any target schema exist or not for migration. Type of tool or technique column contains values like semi-automated (SA) or automated (A). Schema matching column contains values for type of schema matching like automated (A), manual (Mu) or not available (Na). Data model(s) supported for source type column and target type column contain values like multiple (M), single (S) or dual (Du). Designed for source Vs target data model type column contains values like different version of same type of database (D) or heterogeneous database types (H). The last column "Is there any target schema exist for data migration" contains values like yes (√) or no (x). As per the review contained here for above maintained tools and techniques, if any column value in Table-2 for specific tools or technique is not identified then it is defined with value undefined (U).

TABLE-1 : LIST OF EXISTING DATABASE MIGRATION TOOLKITS BY ELAMPARITHI

Name	Company	Source	From	To	Operating System
OSDM Toolkit	Appti-lity	Open	Oracle, SyBase, Informix, DB2, MS Access, MS SQL	PostgreSQL & MySQL	Windows, Linux, Unix & Mac OS
DB Migration	Akcess	Closed	Oracle & MS SQL	PostgreSQL & MySQL	Windows
Mssql2 Pgsq	OS Project	Open	MS SQL	PostgreSQL	Windows
MySQL Migration Toolkit	MySQL AB	Open	MS Access & Oracle	MySQL	Windows
MySQL Migration Toolkit	Intelligent Convertors	Closed	MS Access, MS SQL, Dbase & Oracle	MySQL	Windows
Open DBcopy	Puzzle ITC	Open	Any RDB*	Any RDB*	OS Independent
Progression DB	Versora	Open	MS SQL	PostgreSQL, MySQL & Ingres	Linux & Windows
Shift2Ingres	OS Project	Open	Oracle & DB2	Ingres	OS Independent
SQLPorter	Real Soft Studio	Closed	Oracle, MS SQL, DB2 & Sybase	MySQL	Linux, Mac OS & Windows
SQLWays	Ispirer	Closed	All Relational Databases	PostgreSQL & MySQL	Windows
SwisSQL Data Migration Tool	AdventNet	Closed	Oracle, DB2, MS SQL, Sybase & MaxDB	MySQL	Windows
SwisSQL SQLOne Console	AdventNet	Closed	Oracle, MSSQL, DB2, Informix & Sybase	PostgreSQL & MySQL	Windows
MapForce	Altova	Closed	SQL Server, DB2, MS Access, MySQL & PostgreSQL	SQL Server, DB2, MS Access & Oracle	Windows, Linux & Mac OS
Centerprise Data Integrator	Astera	Closed	SQL Server, DB2, MS Access, MySQL & PostgreSQL	SQL Server, DB2, MS Access, MySQL & PostgreSQL	Windows
DBConvert	DB Convert	Closed	Oracle, DB2, SQLite, MySQL, PostgreSQL, MS Access & Foxpro	Oracle, DB2, SQLite, MySQL, PostgreSQL, MS Access & Foxpro	Windows

TABLE-2 : ANALYSIS OF DATA MIGRATION TOOLS AND TECHNIQUES FOR HETEROGENEOUS DATABASES

Sr. No.	Tools / Techniques	Type	Schema Matching	Source Type	Target Type	Source Vs Target	Is Target Schema Exists?
1	ESF [1]	SA	Na	M	M	H	X
2	SwisSQL [2]	SA	Na	M	M	H	X
3	CMT [3]	SA	Na	M	S	H	X
4	OpenDBCOPY [4]	SA	Mn	M	M	H	√
5	DMA [5]	A	Na	S	S	D	X
6	Oracle SQL Developer Migration Workbench [6]	A	Na	M	S	H	X
7	AWS Schema Conversion Tool [7]	A	Na	M	M	H	X
8	Flyway [8]	SA	Na	M	M	H	X
9	Bogdan [9]	SA	Na	M	M	H	X
10	Shinde [10]	SA	Na	M	M	H	X

Sr. No.	Tools / Techniques	Type	Schema Matching	Source Type	Target Type	Source Vs Target	Is Target Schema Exists?
11	Maatuk [11]	SA	Na	M	M	H	X
12	Bogdan [12]	SA	Mu	M	M	H	X
13	OSDM Toolkit [13]	U	Na	M	M	H	X
14	DB Migration [13]	SA	Na	Du	Du	H	X
15	Mssql2Pgsq [13]	SA	Na	S	S	H	X
16	MySQL Migration Toolkit [13]	A	Na	M	S	H	X
17	Progression DB [13]	SA	Na	S	M	H	X
18	Shift2Ingres [13]	A	Na	Du	S	H	X
19	SQLPorter [13]	U	Na	M	S	H	X
20	SQLWays [13]	SA	Na	M	M	H	X
21	MapForce [13]	SA	Na	M	M	H	X
22	Centerprise Data Integrator [13]	SA	Na	M	M	H	X
23	DBConvert [13]	U	Na	M	M	H	X

SA – Semi-Automated, A – Automated, Mn – Manual, Na – Not Available, H – Heterogeneous, M – Multiple, S – Single, Du –Dual, D – Different Version of DB, U – Undefined, √ - Yes, X – No.

It can be seen from the review of above mentioned tools and techniques for data migration between heterogeneous relational databases that they have one or more issues from the following:

- Require human intervention for data conversion or query refinement.
- Manual schema mapping is required.
- Absence of existing target schema.
- Not able to migrate to multiple database types.

CONCLUSION

In this paper, several tools and techniques for data migration between heterogeneous relational databases are reviewed and their characteristics are listed. It shows that most of them are semi-automated tools which need user input at various stages. Some of them are also automated tools but they are not capable of mapping schemas of different types of databases or require manual schema matching. Furthermore, as per review of various tools and techniques for data migration, derived knowledge states that none of the solution is complete for automated data migration along with automated schema matching of heterogeneous relational databases. So, there is lot of scope to design automated solutions for data migration along with schema matching of heterogeneous relational databases.

REFERENCES

- [1] ESF Database Migration Toolkit. (2013, October). easyfrom.net. [Online]. <https://www.easyfrom.net/>.
- [2] SwisSQL. (2014, December). info.swissql.com. [Online]. <http://info.swissql.com>.
- [3] CUBRID Migration Toolkit. (2014, October). cubrid.org. [Online]. http://www.cubrid.org:8080/wiki_tools/entry/cubrid-migration-toolkit.
- [4] openDBcopy. (2013, March). opendbcopy.sourceforge.net. [Online]. <http://opendbcopy.sourceforge.net/>.
- [5] Microsoft. (2016, December). microsoft.com. [Online]. <https://blogs.msdn.microsoft.com/datamigration/dma/>.
- [6] Oracle. (2014, October). oracle.com. [Online]. <http://www.oracle.com/technetwork/database/migration/omwb-getstarted-093461.html>.
- [7] Amazon. (2015, January). aws.amazon.com. [Online]. <https://aws.amazon.com/documentation/SchemaConversionTool/>.
- [8] Flyway. (2015, April). flywaydb.org. [Online]. <https://flywaydb.org/>.
- [9] Walek Bogdan and Klimes Cyril, "Expert system for data migration between different database management systems," *Advances in Data Networks, Communications, Computers and Materials*, pp. 167-172, 2012.
- [10] Anita Vitthal Shinde, Vaishali Baban Thite, Roshni Warade, and Krupali Chaudhari, "Data Migration System in Heterogeneous Database," *International Journal of Engineering Science and Innovative Technology (IJESIT)*, pp. 88–92, 2013..
- [11] Maatuk Abdelsalam, Ali Akhtar, and Nick Rossiter, "An Integrated Approach to Relational Database Migration," in *International Conference on Information and Communication Technologies*, Bannu, Pakistan, 2008.
- [12] Bogdan Walek and Cyril Klimes, "A methodology for data migration between different database management systems," *International Journal of Computer and Information Engineering*, pp. 85-90, 2012.
- [13] M. Elamparithi and V. Anuratha, "A Review on Database Migration Strategies, Techniques and Tools," *World Journal of Computer Application and Technology*, pp. 41-48, 2015.