Database Migration- How hard can it be?

Arun Swaminathan

School of Computer Science and Engineering, VIT University, Vellore

Abstract. This paper summarizes the tasks and issues of executing a database migration project. This is not a guide to a successful implementation. The characteristics, need and constituents of a migration project are explained. Finally, a 10 step procedure for a successful database migration is presented and justified. The ideas provided are independent of any platform.

Keywords: Database, Database Migration, Database Migration project, Schema

1. Introduction

Data migration is the process of moving data between storage units or entire computer systems. In order for this process to be efficient, powerful data extraction and data loading designs are critical. These designs help in mapping data, which is present on the current system, to the new system which is being implemented.

One of the categories of Data migration is Database Migration. Database migration is the process of moving the business logic, schema, physical data and database dependencies from a current system to a different/new system. Database Migration is used when it is required to shift from one database vendor to another. This may be because of various reasons such as cost, capabilities, functionalities, requirements etc. This paper deals with Database Migration exclusively and not Data Migration as a whole.

2. Characteristics of Database Migration

On first thought, it may seem that when two systems have to hold similar data they would relate to one another seamlessly. But this is most certainly not the case. The differences in their architecture, technology and implementation leads to innumerable issues during database migration. For this reason, database migration projects (and hence data migration projects) have a tendency to fail. According to Bloor Research, as many as 60 percent do not succeed [1].

Statistics show the following picture [2]:

- 16 percent of data migration portion of projects were successful
- 37 percent had budget overruns
- 64 percent were not delivered on time

Most organizations have centres worldwide which are operational 24x7. This calls for the migration to be carefully structured and planned. Some of the benchmarks to keep in mind during the process are:

- **Production Downtime-** The amount of time for which the database and its related application(s) remain 'down' and unusable during the migration process
- **Data Integrity-** The accuracy and consistency of the data stored in the database
- **Database performance**- The optimization of resource use to increase throughput and minimize contention, enabling the largest possible workload to be processed by the database (*http://datatechnologytoday.wordpress.com/2011/06/03/the-definition-of-database-performance/*)

3. Motivation for Database Migration

3.1. Upgradation

Organizations deem it fruitful to move the large amounts of data stored in legacy systems to newer and more reliable current systems. This may be done to make use of the advanced features and functionalities that current systems offer. Further, legacy systems are continually on the decline and moving to contemporary systems make sense when seen in terms of service support, use of latest technology etc.

3.2. Total Cost of Ownership (TCO)

Because legacy systems require experienced DBA's and Application Developer, the total cost incurred is high. Therefore to bring down cost, organizations move on to a single contemporary system with greater functionalities and low total cost of ownership (TCO).

3.3. Technology

One of the major reasons for organizations to decide to migrate is to capitalise on latest technology. There is always a risk with legacy systems that vendors may withdraw its support and make the system obsolete. This motivates organizations to move to new suitable systems. Another motivation could be business demand. The business may expand beyond the capabilities of the current system forcing the need to move to a new system. For example, a leading insurance company used a character based application driven by an 8 year old Informix database. The business demands changed and they decided to opt for an e-business and ERP integration, with a single vendor providing for all the requirements. To exploit the extant technology and its capabilities, the organization migrated from the Informix database to Oracle 8i [3].

3.4. Database Unification

It is most often the case that organizations have their data stored across multiple databases. Different applications run on different databases. But managing and tracking multiple databases is a logistical nightmare. Multiple licenses, multiple vendor support, source feed synchronizations etc. takes up the cost and effort many fold. Thus consolidating the data as far as possible makes sense. For example, an airline company had its data and corresponding applications running on 9 different databases. When running costs and data management came into question, they decided to migrate their data and applications to just two database platforms [3].

4. Constituents of Database Migration

- **Schema Migration-** Replicating the database schema of the current system in the new system
- **Data Migration-** Extraction of data from the current system and loading it on to the new system
- **Application Migration-** Making necessary changes in the applications associated with the current system so that its behaviour remains unchanged when made to interact with the new system

4.1. Role of the Database Administrator (DBA) During the Migration

The database administrator (DBA) plays a crucial role in the project. The DBA is responsible for the success of the migration. The DBA oversees every step of the project except the ones that involve application transfer and application testing on the new platform. [4]

The tasks of the DBA are:

- Assigning well-defined roles to each person involved in the migration project
- Scheduling the test and production migrations
- Performing test migrations
- Performing back-ups of the production database
- Completing the upgrade of the production database
- Performing back-ups of the new database

4.2. Role of the Application Developer During the Migration

The application developer's role is to make sure that applications associated with the database are not affected by the migration. The application developer has to make sure the applications designed for the current database work correctly with the new database once the migration is complete. For this purpose, the

DBA should first install a test database. The application developer can then test and modify the existing applications on the test database [4].
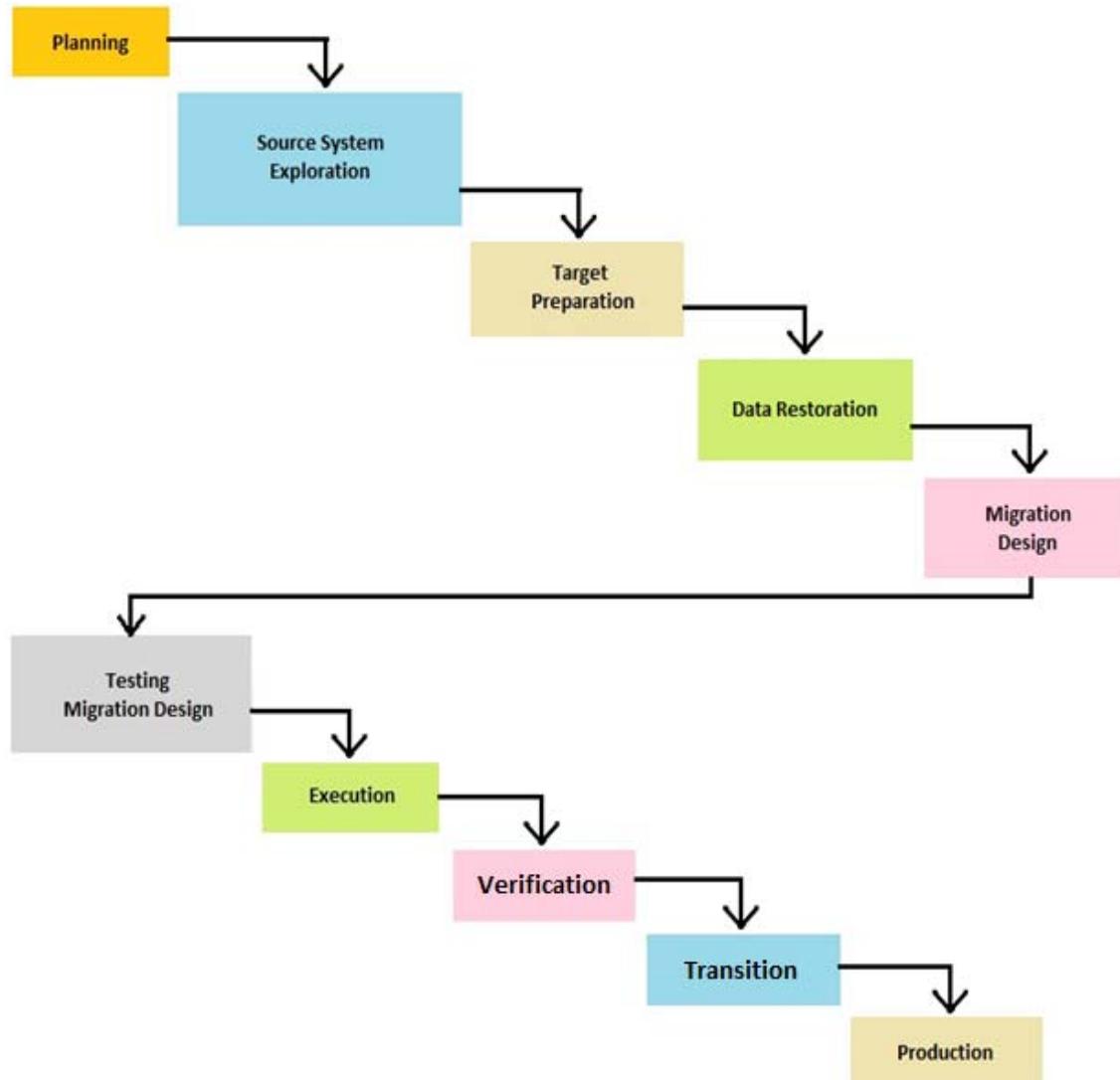
5. Steps in Database Migration Project

Fig. 1: Steps in Database Migration Project

5.1. Planning

At this phase, every aspect that affects the project is taken into account- amount of data, downtime available, time required, security, system vulnerabilities, and back-up/roll back. This plan has to be as detailed as possible.

One of the most common aspect that is underestimated is the time required to transfer the data. FTP manages a maximum rate of 100 megabits per second. When we have a database of hundreds of gigabytes, it would take too long to be transferred by FTP protocol. Moreover, these days almost all data is transferred by SSH protocol in order to make the process secure. This adds an encryption overhead that makes the process slower. The source and target platforms also affect the rate of data transfer. For instance, there are many ways of migrating from Oracle 10g to 11g but not so many for 9i to 10g. If the migration project is across different platforms, the issues only get complicated.

5.2. Source System Exploration [5]

This phase involves identifying and exploring the data that needs to be migrated. Some of the data fields from the source might be redundant. Such fields are not loaded into the target system and hence are not

required for migration. By the end of this phase, all the relevant and required data that will populate the target system is identified.

5.3. Target Preparation

This phase involves preparing the target system such that it can become the primary one. The target system has to be configured such that it can handle all the applications and its corresponding transactions once the migration process is complete. This configuration is done according to the business needs of the organization where the migration project is taking place.

5.4. Data Restoration

Sometimes during transfer of data between source and target systems, certain data inconsistencies creep-in. This phase involves absolving such inconsistencies so that the target system does not fail. In order to successfully carry out this phase, Data Profiling can be done.

"Data profiling is the process of systematically scanning and analysing the contents of all the columns in tables of interest. Profiling identifies data defects at the table and column level. Data profiling is integral to the process of evaluating the conformity of the data and ensuring compliance to the requirements of the target system.

Through the use of data profiling, you can:

- Immediately identify whether the data will fit the business purpose.
- Accurately plan the integration strategy by identifying data anomalies up front.
- Successfully integrate the source data using an automated data quality process." *(David Barkaway, SAS Global Technology)*

5.5. Migration Design

This is where the blueprint for the actual migration process is defined. Further, the test plans are also developed at this stage. This phase is critical for the success of the entire migration project. All the processes carried out after this phase will be based on the designs defined here.

5.6. Testing Migration Design

This is the phase where the design developed in the previous phase is tested. Migration has to be developed thoroughly. This is because migration can be executed just once, there is no second try. This is why database migration projects have such high failure rates[6].

Ideally, small portions of the source data are taken and tested when developing a migration. First one portion of the data is taken and all routines are checked for proper functioning. If the results are satisfactory, data volume is increased. After the test results are out, all necessary changes are made to the migration design to make it flawless. It may be required to carry out testing multiple times.

Final output after this phase gives a migration process that is reliable and rapid.

5.7. Execution [5]

After the test results are satisfactory, the migration is executed. Because generally the source system is shut down during the migration process, it is carried out on week-ends or during the time when data traffic is expected to be low. Sometimes, when the applications are to be run 24/7, a zero downtime migration process might be needed.

5.8. Verification

During this phase, the executed migration process is verified for faults or errors. The target system should meet all the requirements of system performance and data integrity.

5.9. Transition [5]

At this phase, the legacy system is decommissioned and all the applications are run on the new platform. The entire process is documented and logs are created to ease the transition of users to the new system.

5.10. Production [5]

Support has to be provided to the new system and data quality enhancements are continually made. Issues may creep up during operation that need rectifying.

6. Conclusion

Database migration cannot be overlooked as a simple step of retiring an existing platform and moving on to a new one. It is a complex process with many phases and every migration process is unique. This makes it prone to failures. Thorough knowledge of what is required out of the migration, proper migration design and predicting the possible issues that might come up during the migration can bring down the chances of failures drastically. Therefore, being aware of the current issues in database migration and following some simple steps can prove to be critical to successfully executing a migration project.

7. References

[1] Philip Howard, "Data Migration: A White Paper", Bloor Research, September 2007

[2] Philip Howard and Carl Potter, "Data Migration in the Global 2000- Research, Forecasts and Survey results", Bloor Research, September 2007

[3] The Great Migration, Oracle Magazine, May/June-2002 Volume XVI, Issue 3, page–57-65

[4] Tony Morales, "Oracle 9i Database Migration, Release 2 (9.2)", October 2002, Part No. A96530-02

[5] David Barkaway, "How to migrate your data in seven steps", SAScom Magazine

[6] Ken Paris, "10 Steps to Database Migration", Database Programming and Design Journal, Volume 7 Issue 7, July 1994

Arun Swaminathan was born in Hyderabad, India on 27th May 1992. He is currently in his final year of undergraduate study, majoring in Computer Science and Engineering, at VIT University, Vellore. He has worked as an INTERN at Oracle during the summer of 2011 and his current research interests are Database Systems and Information Extraction.