Article can be accessed online at http://www.publishingindia.com

Quick Ready Steps to Deal in "Data" Forest

Madhumita Ghosh*

Every organisation has some degree and magnitude of 'chaotic' culture, especially when it comes to data organisation and data management. Some of them breed chaos and unconsciously operate in chaos. Data management is designed to operate with structure. However, reality has always contained a dose of 'wonderland' as well. Data resettlement is typically the most neglected or forgotten component of any IT project. It is a process of migrating from one system to another, preferably without disrupting or disabling active business processes.

On some occasions, it is not easy to understand or probably viable to consider that the data migration is needed in the project and most of the times data migration is not seen as an item whose requirements need to be captured during analysis phase. That's why, migration related problems begin either during development phase or in testing phase when data need is identified or when the data from the old system refuses to fit properly into your new user interfaces or business rules despite transportation of the old data.

For a successful project, and to carry out an insightful analysis the need of data migration and its requirements must be identified early during beginning of analysis phase and further actions should be reflected to project plan accordingly, but how? Capturing the essentials of data nature is a life-saver activity which is the targeted idea of this work.

Is Data Migration Really a Need?

Understanding the business needs and thereby translating those into analytical needs are utmost essential which should be identified during elicitation phase. If not possible during elicitation phase, the same can be done during analysis phase, but one needs to be careful and should identify the needs as earlier as possible.

One needs to enquire data related simple questions:

- Need the requirements mentioned from the systems be shut down? If yes, does this system keep any data? What type of information is it?
- Are the requirements mentioned from entities already used in company's systems? In order to use that data, need the new system's data model keep them as well? If yes, should this be kept in the same format or otherwise?

If either of the answer is yes, one will need a potential data migration. Moreover, it is advisable to get more details. Fig. 1 is a reference to remember the building blocks for data management

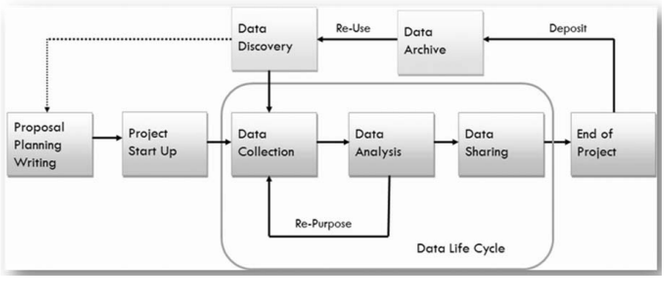

Fig. 1: Building Blocks for Data Management

How to Capture Requirements and Define Business Rules?

If the group or department is determined that it needs a data migration, be careful and specify at least following details in the situation analysis.

To Identify if Company has Multiple Data Sources for that Entities

If yes, determine the master and the fate of other systems' data. You will see how many systems you are dealing

* Practice Leader - Big Data & Advanced Analysis BA & Strategy - Global Business Services IBM, India.

with. It is important to know how and where the data is stored, backed up, and if it is archived.

Data Profiling

Data profiling is a 'must have' activity to understand what are specs of the material the team/ department will be working on. They need to conduct profiling activities and classify the data, such as data with missing unique IDs or missing name or false data etc. Results of the profiling guides to design user interfaces and capture business rules.

Fig. 2: The Cycle of Pre-Analytics Data Processing

Data Cleansing

According the data profiling, one might need cleansing activities for the final structure of the data. Also, it is required to clarify cleansing related activities on the project timeline to have a clearer understanding for the project timeline. If the answer is no, definitely providence of missing or dirty data to be checked.

Data Structures

Data structures must be well understood if the project requires design of user interfaces or data forms. Consideration is must for the data structures while determining dynamic user interfaces or form context.

For example, old system may store the name and its salutation in one field by separating data by " – ". This means two information are stored in one field and one should consider such constraints while you are designing a new interface with multiple fields on screen validation and rules or determine ways to separate salutation and name effectively while moving the data as mentioned on section on data mapping.

Providence of the Historical Data

It is needed to take into account if historical data will be migrated or not. It may affect user interfaces or business processes.

Providence of the Missing or Dirty Data

After profiling, most probably it will be seen that some of the data is not clean or adequate to be used in further actions. For example, study is with sensible customer data and in that case unique identity number is mandatory but some records do not have identity numbers. It will cause problem to pinpoint the customer or it will be further problems if this information is mandatory to display customer on the screen. Even worse, if one has validations based on the identity number such as debt control or billing, the system will not be able to conduct such validations.

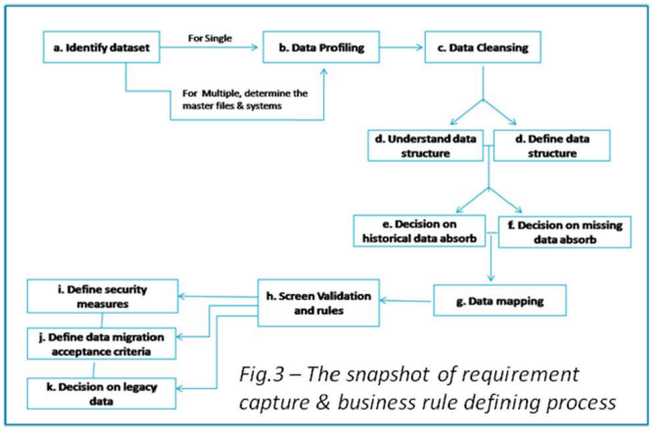

Fig. 3: Snapshot of Requirement Capture and Business Rule Defining Process

Always check whether data ownership belongs to a specific business unit. If yes, let them decide to the fate of data.

- It is advisable to decide whether the data is just enough to use or will it cause problems to conduct business processes. To migrate such data, sections on screen validations & rules and migration acceptance criteria will be highly important.
- Is it possible to clean or enhance problematic data somehow? If yes, determine the ways and related requirements.
- If one decides not to move such data, he/ she should always consult the respective business unit for possible further actions. They may need to find the customer and inform him/ her legally according his/ her account status or they may need another manual/ automation processes.

Data Mapping

Data mapping is basically the activity of creating a map of the existing data model by matching each entity & field with the future data model. Each entity should be mapped correctly and in details to be able to move data successfully. The map is an essential item of *data migration strategy*. Based on the mapping, one can see the gap between legacies and future and one can use the information on screen validations and rules.

Screen Validations and Rules

Results of earlier sections will provide clear information about user interface validations and potential need of new business processes.

Screen Validations

- To define entity specs such as type and length based on the profiling results, the information will guide to design potential data forms.
- To define the rules for the gap: If entities are not matching neatly, need to define UI standards and validation rules accordingly.

New Processes are Needed, in Case One Decided to Transport Problematic Data

- One may need to create new processes to correct such data. For example if the identity number is mandatory and your customer update process originally does not allow to update the number, in order to enable the correction of the customer information, one may need the define rules such as displaying identity number field editable.
- One may create new processes to alert the system or trigger different actions.

Define Security & Security Measures Such as Encryption

If data needs to be migrated in encrypted form, general rules shall be set during analysis phase.

Define Migration Acceptance Criteria

Define migration acceptance criteria such as data quality, migration duration etc. if they cause termination on your services.

Define the Fate of the Legacy Data

Is it advisable to keep the data on the legacy systems? If yes, determine whether synchronisation is needed with the new system or not? How long the data should be kept on legacy system? What are maintenance rules?

What Next?

Of course, one should plan the future analysis steps! All answers will guide for further activities which need to be added to the project plan in details. Identifying these steps at the beginning will prevent from future unexpected surprises and definitely will help to close the project on time. One is rehearsals and the other is testing scenarios.

Conclusion

A core belief in the traditionalist viewpoint of IT project is that business analysis focuses on 'what' the client wants while design concentrates on 'how' to deliver the solution. As a business analysis practitioner, one can discover and describe a requirement that specifies the ability of a user to capture their state (or province) of residence. The main purpose of business analytics IT projects is to develop systems that turn large and often highly complex data sets

into meaningful information from which decisions can be made.

Putting together a comprehensive data migration plan now can help firms avoid serious downtime later, according to experts. Data migration is almost never done simply for the sake of data migration. It has to be tied back to something in the business, and IT managers need to generate their case around those business anchors.