Data Mapping and Required Tools

Dr. Mohini T. Bherwani

Librarian, Shri Binzani City College, Nagpur.

ABSTRACT: *Data mapping begins with the need to know where a data element is located in two distinct data models. Data mapping is done by comparing the entity attribute information from both the conceptual and the logical model with the corresponding element or elements in the database. This paper describes characteristics of big data, steps of data mining and visualization tools for big data.*

Key Words: *volume, variety, velocity, veracity, data integration , data migration, tableau, qlikview, datawrapper*

Data Mapping

In the process of research, when all the data is collected and needs to be integrated, we need a warehouse system to store our data. Data mapping tools are useful to keep track of the data stored in the warehouse. Data mapping translates between one source of information and another, essentially matching data source fields to the target fields in the data warehouse.

Data mapping is the process of creating data elements mapping between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks, including:

- Data transformation or data mediation between a data source and a destination
- Identification of data relationships as part of data lineage analysis
- Discovery of hidden sensitive data such as the last four digits of a social security number hidden in another user id as part of a data masking or deidentification project
- Consolidation of multiple databases into a single database and identifying redundant columns of data for consolidation or elimination

Big Data

Big data refers to data sets that are too large or complex for traditional data processing, application software to adequately deal with. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source.

Characteristics of Big Data

Big data can be described by the following characteristics:

Volume: The quantity of generated and stored data. The size of the data determines the value and potential insight, and whether it can be considered big data or not.

Variety: The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Big data draws from text, images, audio, video; plus it completes missing pieces through data fusion.

Velocity: In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time. Compared to small data, big data are produced more continually. Two kinds of velocity related to big data are the frequency of generation and the frequency of handling, recording, and publishing.

Veracity: It is the extended definition for big data, which refers to the data quality and the data value. The data quality of captured data can vary greatly, affecting the accurate analysis.

Data must be processed with advanced tools (analytics and algorithms) to reveal meaningful information. For example, to manage a factory one must consider both visible and invisible issues with various components. Information generation algorithms must detect and address invisible issues such as machine degradation, component wear, etc. on the factory floor.

Steps of Data Mapping

To leverage data and extract business value out of it, the information collected from various external and internal sources must be unified and transformed into a format suitable for the operational and analytical

processes. This is accomplished through data mapping, which is an integral step in various data management processes, including:

Data Integration

For successful data integration, the source and target data repositories must have the same data model. However, it is rare for any two data repositories to have the same schema. Data mining tools help bridge the differences in the schemas of data source and destination, allowing businesses to consolidate information from different data points easily.

Data Migration

Data migration is the process of moving data from one database to another. While there are various steps involved in the process, creating mappings between source and target is one of the most difficult and time-consuming tasks, particularly when done manually. Inaccurate and invalid mappings at this stage not only impact the accuracy and completeness of data being migrated but can even lead to the failure of the data migration project. Therefore, using a code-free data mapping solution that can automate the process is important to migrate data to the destination successfully.

Data Warehousing

Data mapping in a data warehouse is the process of creating a connection between the source and target tables or attributes. Using data mapping, businesses can build a logical data model and define how data will be structured and stored in the data warehouse. The process begins with collecting all the required information and understanding the source data. Once that has been done and a data mapping document created, building the transformation rules and creating mappings is a simple process with a data mapping solution.

Visualization Tools for Big Data

Data visualization is about how to present your data, to the right people, at the right time, in order to enable them to gain insights most effectively.

Tableau

Tableau is a business intelligence and data visualization tool that has a very intuitive user interface. It is very useful in drilling down data and creating reports. There is no need for any prior technical knowledge in order to use tableau. It can be connected to all kinds of data sources. Various advantages of tableau are:

- Fantastic Visualization
- In depth insights
- User friendly approach
- Adding data sets
- Switching between visualization

Qlikview

QlikView is a leading Business Discovery Platform. It is unique in many ways as compared to the traditional BI platforms. As a data analysis tool, it always maintains the relationship between the data and this relationship can be seen visually using colors. It also shows the data that are not related. It provides both direct and indirect searches by using individual searches in the list boxes.

QlikView's core and patented technology has the feature of in-memory data processing, which gives superfast result to the users.

Fusion Charts

This is a very widely-used, JavaScript-based charting and visualization package that has established itself as one of the leaders in the paid-for market. It can produce 90 different chart types and integrates with a large number of platforms and frameworks giving a great deal of flexibility. One feature that has helped make FusionCharts very popular is that rather than having to start each new visualization from scratch, users can pick from a range of "live" example templates, simply plugging in their own data sources as needed.

Highcharts

Highcharts is a pure JavaScript based charting library meant to enhance web applications by adding interactive charting capability. Highcharts provides a wide variety of charts. For example, line charts, spline charts, area charts, bar charts, pie charts and so on. This tutorial will teach you the basics of Highcharts. There are chapters discussing all the basic components of Highcharts with suitable examples.

This tutorial is designed for Software Professionals who are willing to learn Highcharts in simple and easy steps. This tutorial will give you an understanding of the Highcharts concepts and after completing this tutorial you will be at an intermediate level of expertise from where you can take yourself to a higher level of expertise.

Datawrapper

Datawrapper is increasingly becoming a popular choice, particularly among media organizations which frequently use it to create charts and present statistics. It has a simple, clear interface that makes it very easy to upload csv data and create straightforward charts, and also maps, that can quickly be embedded into reports.

Plotly

Plotly enables more complex and sophisticated visualizations, thanks to its integration with analytics-oriented programming languages such as Python, R and Matlab. It is built on top of the open source d3.js visualization libraries for JavaScript, but this commercial package (with a free non-commercial licence available) adds layers of user-friendliness and support as well as inbuilt support for APIs such as Salesforce.

Sisense

Sisense provides a full stack analytics platform but its visualization capabilities provide a simple-to-use drag and drop interface which allow charts and more complex graphics, as well as interactive visualizations, to be created with a minimum of hassle. It enables multiple sources of data to be gathered into one easily accessed repositories where it can be queried through dashboards instantaneously, even across Big Data-sized sets. Dashboards can then be shared across organizations ensuring even non technically-minded staff can find the answers they need to their problems.

References:

1. English, Bill, Alderman, Brian, Ferraz, Mark. Microsoft® SharePoint® 2010 Administrator's Companion. Redmond, Washington: Microsoft Press; 2010.
2. Ganci, John, Fox, Leonard, Hutton, James, Jensen, Thomas, Stec, Andrzej, Zhang, Eric. Document management using WebSphere Portal V5.0.2 and DB2 Content Manager V8. IBM Redbooks; 2004.
3. Kriegel, Alex, Trukhnov, Boris M. SQL Bible, Second Edition. Indianapolis: Wiley Publishing Inc; 2008.
4. Kriegel, Alex. Discovering SQL. Indianapolis: Wiley Publishing Inc; 2011. 15 Choate, Mark. Professional wikis. Indeanapolis: Wiley Publishing Inc; 2008.
5. Singh, Shio Kumar. Database systems: Concepts, design and applications, Second Edition. New Delhi: Dorling Kinderley; 2011.
6. The official introduction to the ITIL service lifecycle. Norwich: The Stationery Office; 2007.
7. Wikipedia. Isaac Asimov [online]. URL: http://en.wikiquote.org/wiki/Isaac_Asimov. Accessed 11 March 2013.