Introducing data warehouse design for university libraries Case study: Sri Palee campus library

Ven. Pathakada Rewatha Thera

prewatha@hotmail.com

ABSTRACT

Libraries maintain their information resources in a variety of forms, so their use may be limited. Library information resources are currently being digitized, but they are no longer used for data analytics. But it is to examine how effectively library data can be utilized by using big data technology for libraries, as in other fields. When creating a data warehouse, it appears that a library has two main types of data: collected information in the library and other data, such as reader transactions. All that data can be used for data warehouses. Sri Pali Campus library which has its own collection and other data like transactions, is the domain for this data warehouse. When processing data as appropriate for a data warehouse the challenges to be faced can be seen here as well. Data warehousing benefits the library to make decision-making data exhibits, monitor reader behavior as well as manage the collection accordingly, and integrate library data with the latest technology to expand access.

Keywords: *Big data, Data warehouse, Library data, University library*

Introduction

Books, journals, research articles and other materials both in physical and electronic media have been collected by libraries. Collection of a library is for researchers and public users to find their information needs. Because of their quantity and variety of format, library data might be effective for efficient use. Though some library's data these days have been digitized, most of them have not been used for big data or data mining technology. It is worth using big data technology for efficient use for users of that kind of isolated data in libraries. Discussions of data mining projects which could have been done using library data, are limited. Some subject specialists believe that database management systems are enough for libraries and that big data or data mining projects are not required. Hence, it is worth highlighting the importance of applying big data technology to the library in order to make the right decision in this regard. Today, Big Data technologies are being used to enhance their predictability and effective and comprehensive decision making in various areas such as education, trade, security and governance etc. This research paper seeks to consider the basic requirements for implementing a Big Data project in association with the Sri Palee University Library and to evaluate the benefits that can be derived from it. In addition, it provides a basic understanding of future library data mining projects.

Objectives

The main objective of this research paper is to point out that big data projects can be also used for libraries like many other fields to support top order management decision making.

The domain for the big data

The database of the Sri Pali University Library is used for this purpose. The library's daily transaction information and reader suggestions, book purchase lists and most importantly, the library collection is an adequate database for creating a data warehouse. The Sri Palee Campus library initiated its library and information services in July 1999 and currently has a collection of approximately 28,000 books mainly concentrated on subjects pertaining to Performing Arts and Mass Media. Currently, the library membership consists of 650 undergraduates, 78 Academic staff members, and 135 Nonacademic staff members of the Sri Palee Campus. The visiting academic staff is eligible for library membership with a refundable deposit for borrowing facilities. The library consists of the Permanent Reference section, Reference section, reading room, Lending section, Periodicals section, and the reading capacity is sixty. Library management system is Alice for windows.

Challenges for the project

Volume

The amount of data available in the library for a Big Data project is minimal compared to any other field for such a task. Everyday data collection in the library is not as massive as in a bank or a telephone company. The library has a database of about 28,000 and about a thousand transactions take place daily. Accordingly, the library's current collection of information has not been virtually translated, but it is a vast collection of data. So quantitatively, library data can be seen to be a volume that is described as the first challenge of the big data concept.

Velocity

It covers the speed at which data is created, speed of storing data, speed of analyzing data. According to that library has usually done more than a thousand transactions and at the same time data will save in a server.

Veracity

Data Veracity means the accuracy of the data. When thousands and thousands of pieces of information are collected, some may have problems with their accuracy. The same is true of library data. For example, the accuracy of some reader information may be questionable. In addition, there may be errors in the entry of bibliographic information and incorrect presentations of information by some authors.

Data warehouse design

Data warehouse is a collection of numerous data that are non-updatable, integrated and historical data having time variant, to support decision making. The process of building a data warehouse from the source system is extraction, transformation and loading (ETL). In this process, data must be clean in order to retrieve good data quality. When considering library day-to-day activities in relational databases, there may be invalid telephone numbers, invalid book codes etc. Clearing that kind of errors, ETL process will put data in a staging area, and

then transform and load it into a data warehouse. In this process is to make a dimensional data model in which each dimension table has a surrogate key. And the fact table for all of them.

In order to make a data warehouse there are few steps that consist of separate tasks to fulfill.

- Selecting the process- it means determining the subjects. Here it is library transactions that have 3 processes: book lending, return and procurements.
- Selecting the grains- what grains are to be represented by the fact table? Book lending, book procurement and book status will be represented here.
- Recognizing the dimension- Determination of dimensions that are related to the fact tables is the task in this step. Here we identified what they are. (fig.1)
- Round out the dimension- consideration of every dimension table is the task here. We had 9-dimension tables. (fig.2)
- Duration of database- in this step, it was selected last 22 years.
- Design the physical database- After completing all above steps, this is the final design data warehouse for the sri palee campus library.

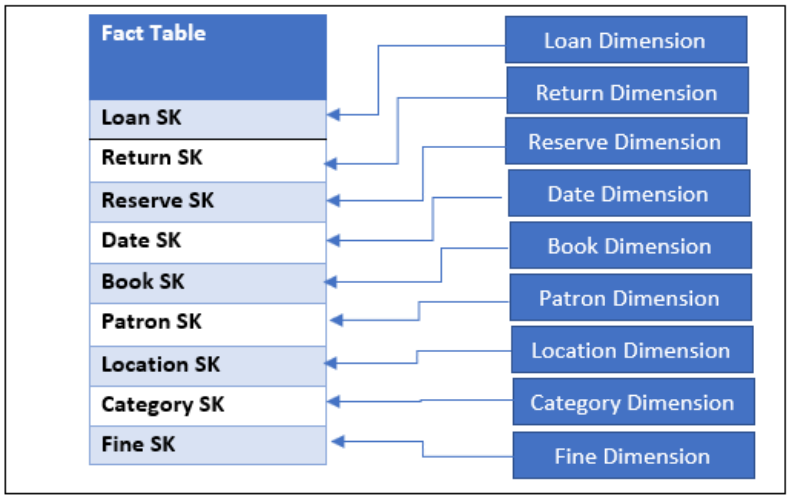

Fig.1 Fact table

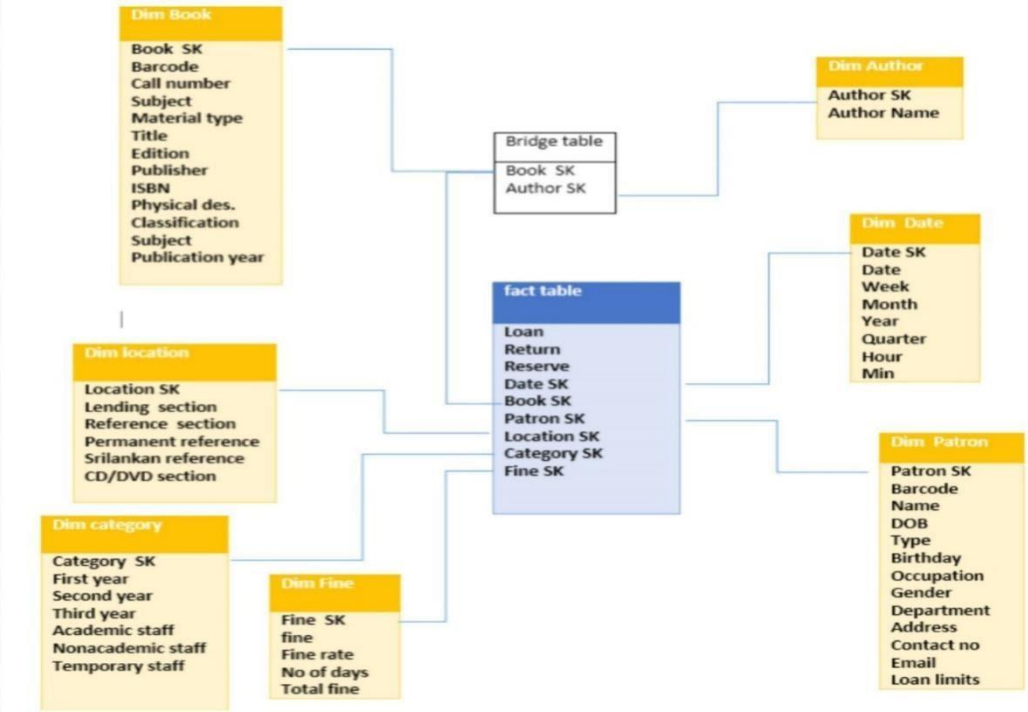

Fig.1. Data Warehouse design for Library system fact table and dimension tables

Advantages of using dw for libraries

As we mentioned previously, many establishments such as the government, hospitals, and businesses are analyzing big data for their improvement. For example, because of the prevention of illness, hospitals use big data, finding improved ways to sell products, and business bodies use it. So, what can it do for the library? The Following are some of them.

Decision making

This is the main purpose of using big data analytics. Decisions could be more useful when they are based on the data. For example, analyzing data on library transactions or searches can provide ideas for optimizing its collection. This approach would improve reader satisfaction and the effective use of library resources.

User Behavior study

For the library big data, user activities tracking data is possible in addition to information of library collection. After collecting a large amount of user behavioral data, the library could analyze it and improve overall user experience as well.

Data visualization

The library data could be selected and visualized as per the user's needs. That visualization will display various collections of the library as well as unbalance each of them. Moreover, such an approach can be considered when allocating funds and making plans in collection development.

New Data Format

Accessibility and sharing of the information are main goals of the library. In order to achieve that, the data medium must be changed. Ancient sources in particular need to be digitized. This will allow them to communicate and access information more efficiently. Another advantage from this is library data can be used with other online resources. It means library data is going to be linked data like google and facebook.

Conclusion

Although some argue that it is not yet necessary to create a data warehouse for libraries, Data Warehouse provides more effective access and many more benefits than a relational database. In addition to the information available in the library, daily transaction data and reader information can also be used for data warehouses. The library of the Sri Palee Campus has a collection of over **28,000** books and nearly a thousand readers, and data on their behavior was used to create the data warehouse. There were a number of different challenges. After overcoming the challenges, it could be done in a few steps. This research is a blueprint for the future library big data projects.

References

V. Rainardi. Building A Data Warehouse with Examples in SQL Server, Apress, New York.

P. Ponniah, Data Warehouse Fundamentals: A Comprehensive Guide for IT Professionals, J. Wiley, New York

Salo, D., 2010, "Retooling Libraries for the Data Challenge" Ariadne (64) (July)(http://www.ariadne.ac.uk/issue64/salo/).

Bell, S., 2013, Promise and Problems of Big Data,. https://www.libraryjournal.com/?detailStory=promise-and-problems-of-big-data-from-the-bell-tower

Dinesh, A., 2020, Think about Data Volume in Vs of Data, http://dbfriend.blogspot.com/2020/10/think-about-data-volume-in-vs-of-data.html