

Analysis of E-Commerce Backend Operations Data

Amol Modi
UG Student,
Department of
Computer Engineering,
Pune Institute of
Computer Technology
Pune, Maharashtra,
411040 India

Anjana Pradeep
UG Student,
Department of
Computer Engineering,
Pune Institute of
Computer Technology
Pune, Maharashtra,
411040 India

Rohan Jain
UG Student,
Department of
Computer Engineering,
Pune Institute of
Computer Technology
Pune, Maharashtra,
411040 India

Payal Rathod
UG Student, Department of
Computer Engineering,
Pune Institute of Computer
Technology
Pune, Maharashtra, 411040
India

ABSTRACT

Data analysis is the process of finding the analyzing data to find useful results for decision-making, understanding fundamental the data, discovering the significant pattern in the data. Analysis of e-commerce data (which includes guidelines details, record details, stuffing details, delivery details, etc.) involves scanning through the database and extracting data and probing pattern that can be helpful to the company in classify to get better their competence in the backend behavior like selection, superpriority checks, packing and shipping. Such information obtained from analysis can provide insight into the various operations and can help the company to make informed decisions to speed up activities and ensure delivery in shortest possible time. The raw data from the various sources needs to be extracted, transformed and loaded into the data warehouse before beginning the proper analysis on the data. This consolidated data is available to run a series of patterns for knowledge discovery.

In this paper, ETL process, architecture of a system involving Pentaho ETL tool for analysis of large e-commerce data sets and the final visualization of the analysis are discussed.

Keywords

Database Management, Information Storage and Retrieval, Information Systems Applications

1. INTRODUCTION

In the current era, with massive technological growth and development of the Internet and online world, a vast volume of information and data is faced day by day from many different resources and services. Massive quantities of information are created by and about community, things, and their interactions. Those new data sources generate daily a huge quantity of unstructured data, to deal with such complex data, the use of big data analytic tools becomes an obligation. [5] Various organizations are not concerned with this entire volume of data rather they are interested in the useful information that can give them insight about the company operations and interactions.

Organizations gather increasing large and complex datasets each year. These organizations rely on data analysis to model customer engagement, streamline operations, improve productions, inform sales and business decisions, and combat fraud. [7] E-commerce sites, mobile apps produce huge amount of data. This mountain of data must be processed and analyzed in order to find useful patterns, to explain business phenomena, to make decisions. Such large amount of data can be produced by various different sources and various different formats. Such a variety of data can be successfully analyzed only if they can be stored together in a common format so that the relations between them can be established. Then this data can be the center for making data driven decision-making.

This paper proposes a method of analyzing large amounts of data, stored in a relational database, produced by the e-commerce backend operations. It starts with integrating an open source tool, Pentaho, which provides with a set of business intelligence products like data integration, OLAP services, ETL tools and reporting. It involves steps to extract the data, transform it into relevant formant and then loading it as per required. Then the data extracted is analyzed and the result is displayed as charts, bar graphs and pie charts, in a web application.

2. ETL: EXTRACT, TRANSFORM, LOAD

ETL, are three processes, that is, extract, transform and load, which is an integral part of data warehousing. It is widely used in managing databases. ETL is can be used to obtain sets of useful data from large data stores which can be useful for analysis. Raw data can be made useful by applying the three process of ETL. On large data sets the process of ETL can be speed up by paralleling executing all three processes.

2.1 Extract

The data required for analysis can be obtained from various sources. The first step of ETL involves extracting data from these different sources. It is important to extract the data correctly in order to avoid loss or corruption of data as the data extracted will determine the result of the analysis. Once the required datasets are obtained, the next step is to transform it.

2.2 Transform

Transformation of data includes cleaning and transforming the data. Cleaning involves handling missing data, removing redundancies, standardizing values. Making use of transform functions, rules or lookup tables, the acquired data is transformed into the target format. The conversion step also requires mixture data from several sources, generating aggregates, generating surrogate keys; categorization, deriving innovative intended values, and applying advanced validation rules.

2.3 Load

After the required data is extracted and transformed, the last step is to load the data. The commonly used target is a database. This data can then be used as per requirement.

3. PENTAHO: INTEGRATION WITH MYSQL

There are many ready-to-use ETL tools on the market. The main benefit of using these ETL tools is the fact that they are optimized for the ETL process by providing connectors to general information sources like databases, flat files, mainframe systems, xml, etc. They provide a means to implement data transformation without doubt and without fail transversely various data sources.

One such tool is Pentaho. Pentaho business intelligence server edition is installed on the Linux server along with required tools like JDK and JRE. Once the installation is done, the Pentaho server can be started and accessed using local web browser. Using the user interface provided by Pentaho,

connection to the MySQL database can be done using port 3306. Once this connection is established, required data can be extracted using SQL queries and joins and stored in the format as required.

4. OVERVIEW OF ANALYSIS PROCESS

Analysis of data involves inspecting and modeling the data in order to find useful patterns and conclusions that help in decision making. In the e-commerce backend analysis, as mentioned MySQL relational database is being used. The needed data is extracted using queries and joins.

5. RUBY ON RAILS: VISUALISATION OF ANALYSIS

Ruby on rails is a web application framework written in Ruby, a high level programming language. HTML, Hyper Text Markup Language, is a markup language used for creating web pages. With the combination of Ruby on Rails and HTML the visualization of the analysis is displayed.

6. ARCHITECTURE DIAGRAM:

The design architecture used for the problem statement shows the extraction of data from the database using ETL tool and a connector. Using SQL query and Query executor, data can be extracted for the parameters and the result is displayed on Web App (Dashboard) using Ruby on Rail integration as the backend.

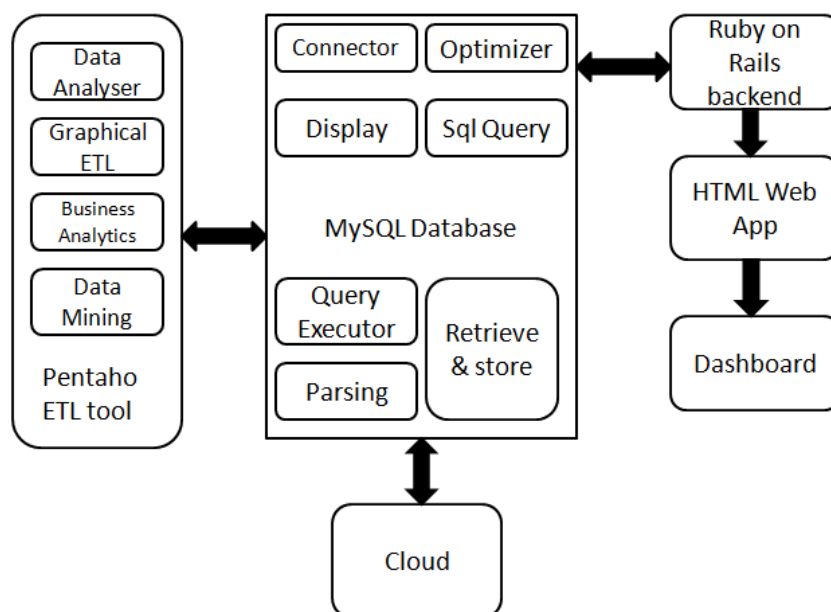


Fig 1. Architecture Diagram

7. CONCLUSION

As e-commerce has become such an integral part people's daily life, the e-commerce market is becoming more and more competitive. As a result, the requirement of efficiency and high quality service is becoming more and more important. This paper suggests how the data generated by the e-commerce backend be used to take important decisions in order to improve efficiency and quality of service. Using an

open source tool the data to be analyzed is extracted, transformed and loaded. This data is then analyzed to get an insight into the operations of the organization. Using visualizations in the form of graphs and charts informed decision can be taken in order to improve the efficiency of the organization.

The future scope of the project involves analysis of e-commerce data of various types at a macro and micro level of granularity and finding relevant connections between the data sets for less risk prone business decisions.

8. ACKNOWLEDGMENTS

Our special thanks to our internal guides Prof. A. D. Londhe and Prof. M. S. Chavan for giving us all the help and guidance. We are grateful for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. G. P. Potdar, Head of Computer Engineering Department, for his indispensable support and suggestions. Also special thanks to Mr. Abhimanyu Kashikar for providing various resources for the project.

9. REFERENCES

- [1] Yoshinori Fukue, Kessoku Masayuki, Kazuhiko Tsuda, 2004. "Extracting Purchase Patterns in Convenience Store E-Commerce Market Using Customer Cube Analysis".
- [2] Marin Fotache, Catalin Stimbei, 2015. "SQL and data analysis. Some implications for data analysis and higher education".
- [3] Thi Thi Zin, Pyke Tin and Takashi Toriu, Hiromitsu Hama, 2012. "A Big Data Application Framework for Consumer Behaviour Analysis".
- [4] Sachchidanand Singh, Nirmala Singh, 2012. "Big Data Analytics".
- [5] Abdelkarim Ben Ayed, Mohamed Ben Halima, Adel M. Alim, 2015. "Big Data Analytics for Logistics and Transportation".
- [6] YANG Hao, Song Hongwei, ZHANG Zili., 2011. "The application of e-commerce System based on data warehouse".
- [7] Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer, 2012. "Enterprise Data Analysis and Visualisation: An Interview Study".
- [8] D. Gotz and M. X. Zhou., "Characterizing users' visual analytic activity for insight provenance". Information Visualization, 8:42-55, 2009.
- [9] Wang Hongding, Yu Bo, Tang Shiwei., "An Effective Approach to Design Data Warehouse" [J]. COMPUTER ENGINEERING AND APPLICATIONS, 2004 40(9) 12 82.
- [10] Borkar, V.R., Carey, M.J., Li, C. (2012), Inside "Big Data Management": Ogres, Onions, or Parfaits, Proc. of the 15th International Conference on Extending Database Technology (EDBT '12), ACM, New York, USA, pp.3-14 .
- [11] T.K.Das, P.Mohan Kumar, "BIG Data Analytics: A Framework for Unstructured Data Analysis," Intl. Journal of Engineering and Technology (IJET), vol. 5, pp. 153-156, Feb-Mar 2013.
- [12] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization". In Proc. IEEE Information Visualization (InfoVis), 2005.