International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 3

ISSN: 2454-4248
251 – 256

Exploring Business Potential in Retail Segment Using Big Data Analytics

Sachin Sharma, Sandip Kumar Goyal, Kamal Kumar, Avinash Sharma

ABSTRACT: Retail segment in banking comprises of housing sector, automotive sector, education sector etc. The enormous data available from various sources like web portal, enterprise data can be analyzed to generate business and unveiling the potential in retail segment. Big data analytics can be very useful to process and analyze data from various commercial portals available viz. magicbricks.com, cartrade.com, education portals etc. The data can be processed and analyzed by banks to identify the potential customers and offer them the suitable products so that it can be turned up in business generating entity.

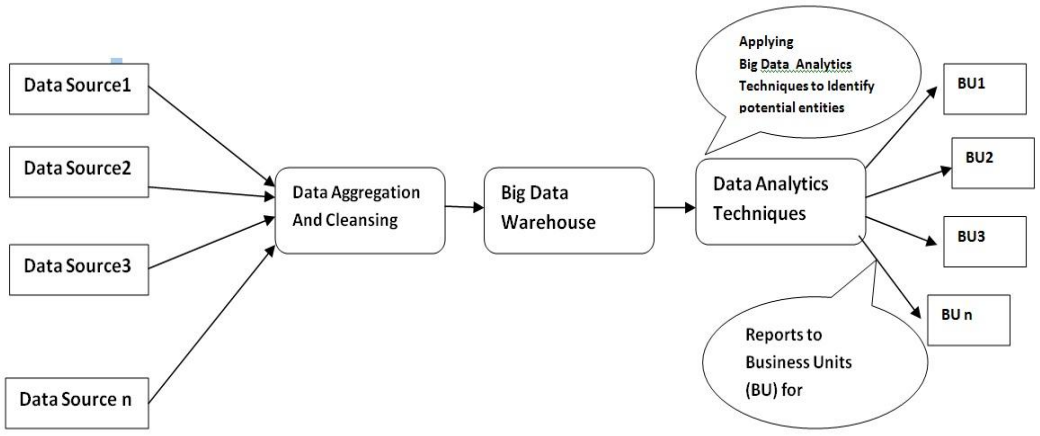
*****

1. Introduction

With the emergence and reach of internet and social media; people are looking for opportunities in housing, automobiles and education sector for their utilization. There is a lot of potential from the customers which are still unforeseen. There is very useful data available in and around us who needs to be taken into consideration by financial institutions. The aim of the research is to take the data generating from various forums including social media, web portals and other medium. The data collected will be processed using big data analytic techniques so that a trend pattern can be identified for potential customers. The ultimate data extracted in the form of reports / graphs can be analyzed and shared with different business units within the financial institution for tapping this untapped business.

2. Proposed Research Methodology

The data from various web portals, social media platforms, education websites etc. can be taken and pushed into the Data Aggregation unit. Thereafter, the data can be cleansed and brought into a common format which was previously aggregated from various resources. Now, the Big Data analytics techniqueswould be applied to Big data store to identify and analyzemeaningful data related to consumer behavior extracted from the choices/inputs they have exercised on different platforms. This analysis of data can then be presented in form of reports to different Business Units to tap the potential from the consumers by designing the customized products. The research methodology can be presented in the diagrammatic form as below:



Proposed architecture diagram for Big Data Analytics for Financial Institutions

3. Existing Measures to Access data warehouse potential:

The success and potential of a data warehouse should be measured and without metrics , the data warehouse is of no use and is not quantifiable to judge. Without the use of metrics , its difficult to get idea of turnaround time, user experience, data quality etc. The major parameters on which data warehouse potential can be measured on parameters :

a) Success measure conformance is a tool to measure data warehouse success or failure. The success and failure parameters vary organization to organization. The factors can be internal or external.
b) Metrics Types:

 ISSN: 2454-4248
251 – 256

i. Performance: Generally we get very limited time windows for data warehouse process as data has to be taken from transactional systems. Transactional systems work 24*7 mode and DWH activity can slow down the performance of the transactional system. So the data warehouse process needs to be completed in a time bound environment and performance is a major measurement tool for the data warehouse potential. Performance can be judged with the turn around time for processing data which includes extraction from source, staging, and transformation,cleansing, loading and reporting. There are many factors involved in it viz. query optimization, performance tuning, ETL workflow optimization, hardware optimization and many more.
ii. Usage: There is lot of efforts in terms of money and manpower to build a data warehouse. The potential of data warehouse can be utilized only when maximum users can be benefitted by the data warehouse. It is an important metrics to ensure the usability of the data warehouse and in case some arrangements are required , that should be materiealized in terms of awareness, knowledge transfers and trainings to boost the DWH utility.
iii. Data Availability: The time window for users to access data available in data warehouse is an important metrics for data warehouse. The technical architecture of the data warehouse design should be in such a way that the data availability should be as maximum time as we can.
iv. Utilization of Resources: Hardware utilization like RAID configuration , cloud cluster, operating system resources including memory plays an important role. Tools utilization should also be ensured as they have been purchased a a very high cost.
v. User Satisfaction: The user requirements should be adhere and reflect in the delivered data warehouse / reports.
vi. Quality of data: Data quality is an important aspect of the data warehouse. Data Cleansing and purging of unused and unwanted data should be ensured and covered under a policy.
vii. Costing: Generally data warehouse development and maintenance cost is too high that it is concentrated in big organizations only. The cost to utilization ration needs to be considered while evaluating the data warehouse potential.
viii. Organizational Benefits: The ultimate goal of an organization is to cut the expenditures and increase the revenue. The return on investment for data warehouse can be measured on the fact that how it increase in the organizations profitability by means of forecasting, reports, trend analysis etc.

c) Service Level Measures: Service level measures should be put in place for a potentially good data warehouse. Three major points to be included in the service level agreement is data availability, response time and mean time to failure / problem response time.

4. Related Work:

5. Methodologies published recently for Data Warehouse and Data Analysis

a. AXML Decision Support Methodology

The methodology given in [23] based upon the dynamically extraction of real time data based upon the query used for decision support system. AXML based database has been proposed for data integration from external sources. The approach is a complimentary to the existing data warehouse approaches.

b. Five dimensional data analytic approach

The methodology suggested in [24]has been introduced five dimensional approach. The New dimension has been introduced as scenarios-based dimension which is addition to Traditional four dimensional systems. Authors have stated the result for the better analysis of the large data warehouse than the existing approaches.The existing dimensions for analytic system are use dimension, governance dimension, asset dimension and integrity dimension. The newly introduced dimension is scenarios dimension which acts diagonally between dimensions.

c. Semantic data warehouse methodology:

Sharing of information between various organizations is having big challenge of heterogeneity among data. Various methods and ontologies have been proposed and a set of semantic data has been produced as stated in [25]. The growing data will require

 ISSN: 2454-4248
251 – 256

management of semantic data. To handle the situation, semantic database have been proposed which may be referred as SDB. These databases are extension to the traditional databases. Managing semantic data is not an issue at all , rather decision making data extraction from semantic databases is a major challenge. Data warehouse technology is a ideal for data analytics. Authors in [25] have given a framework to take benefit of data warehouse for analytical processing and at the same time using semantic database sources as the data source for semantic datawarehouse. To develop semantic datawarehouse for collaborative organizations, authors have suggested a) Generic integration process development <G,S,M> , b) generic ETL algorithm to deal with heterogeneous semantic database sources.

d. Kimball Data warehouse Approach in Health Care Research:

Authors in [26] have focused on the clinical research database and enlighten the weakness of Kimball approach when it comes at clinical research which is more complex than other areas. Health care analytics involves examination of cause and effect relationship. Kimball approach is very strong at individual business processes but its not suited for clinical research as process interrelation is required in the clinical research to establish cause and effect relationship. The Kimball approach consists of Fact and dimensions, where fact table represents business process and surrounding tables i.e dimensions tables represents business process definitions. The authors in [26] have put in their efforts to address the weakness in Kimball approach so that it can be utilized in the Health Care research analytics.

e. Decision making methodologies – A comparative analysis for sales data mart

Authors in [27] have presented decision making analysis among three approaches and arithmetic mean has been used to take appropriate decision in sales data mart. The result of analysis of three decision making techniques is the best analysis method. Arithmetic mean value method , decision matrix method and place wise decision analysis method has been studied and compared for the best results. For effective decision making ; after experimental analysis using arithmetic means among various methods , authors have concluded that decision matrix method can be used for effective decision making.

f. ATDM Methodology for Enterprise Data Warehouse

Authors in [28] have given a significant contribution in development lifecycle for enterprise data warehouse design and development. Authors have identified and pointed out that existing data warehouse methodologies do not support cohesive data warehouse development life cycle that should contain metadata documentation, feasibility analysis , logical and physical data model , implementation and maintenance. Authors have proposed an extension to the existing Triple-Driven data modeling (TDM) methodology. The Adapted Triple Driven data modeling (ATDM) methodology has been introduced which supports technical metadata generation and documentation.

g. Big Data Iterative Methodology

Authors in [29] have found that return on investment from big data projects is rarely achieved by organizations. The major challenges in big data is “5 V” which are volume, velocity, variety, veracity and value. Authors have suggested some of the reasons for failure of big data projects which are i. challenges in integration of multiple data sources, ii. Un-skilled staff for big data iii. Data quality issues iv. Design issues v. deficiency in tools integration. Authors have mentioned another study carried by Gartner Consulting firm which stated major issues in big data as – i. value extraction from data, ii. big data strategy defin ition, iii. Skill and capability requirements, iv multiple data sources integration and v. data quality issue. Authors in [29] tried to address above mentioned issues and proposed iterative methodology to be adopted in big data warehouse. Authors allowed developers to apply methodology in a systematic way to any domain and apply the methodology in a case study to validate the results.

6. Assessment Outputs from Representative Approaches

a) AXML Decision Support Methodology:

The assessment output that has been arrived from the AXML approach is the complimentary approach to the existing data warehouse approach. The combination of two approaches i.e active integration approach and passive integration approach has been materialized. The active integration approach has been introduced by authors in [23] whereas passive integration approach is currently in use and has Extraction, Transformation and Loading for data warehouse. The identification of decisive and indecisive data has been carried out so that priority among them can be decided. Decisive data should be in sync always as decision for the organizations depends upon the critical decisive data. The indecisive data generally does not have any impact on the decision

 ISSN: 2454-4248
251 – 256

taking capabilities of an organization. The methodology that has come into picture is the traditional ETL based approach for indecisive data and active integration approach has been applied to the decisive data.

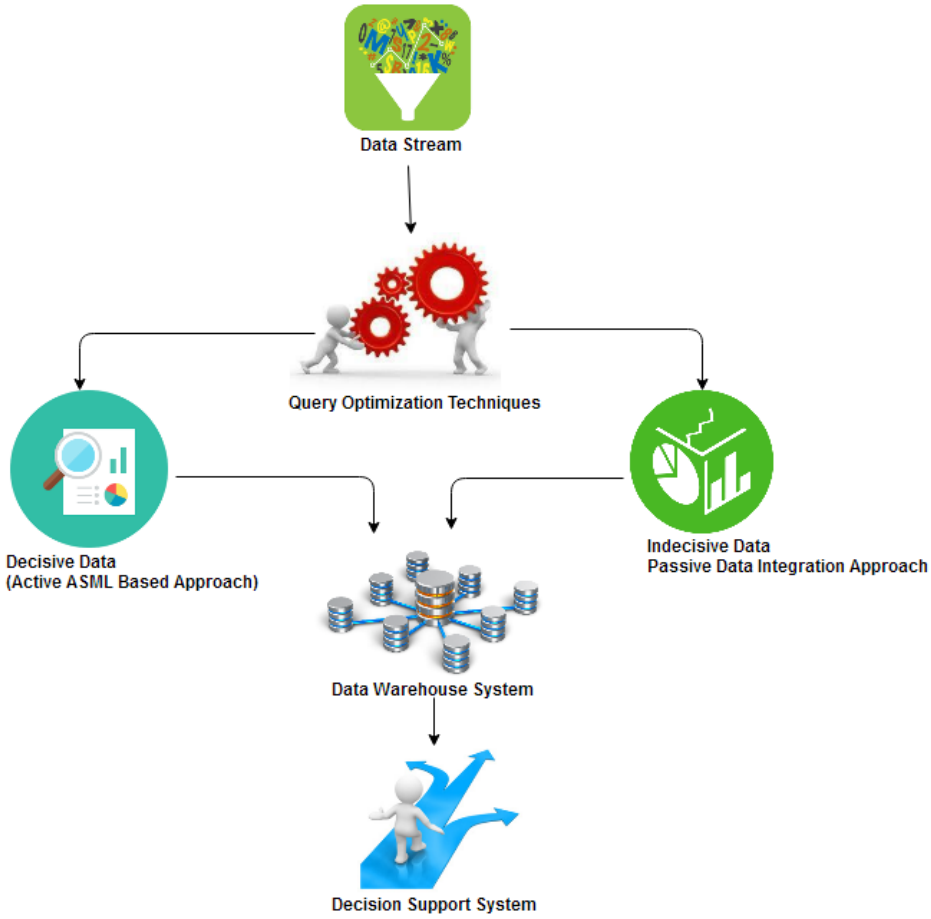


Fig. – AXML Based approach methodology outcome.

b) Five dimensional data analytic approach

The five dimensional approach as detailed in [24] focused on the problem related to open-end behavior of requirements which are vulnerable to change upon the organizational requirements. The enhancements are welcomed in growing organization and authors have identified data warehouse as a good opportunity for scenario/use-case based analysis. The scenario /use-case based analysis help an organization in gap identification which are enhancement centric.

The ultimate outcome of this approach is the fifth dimension i.e Scenario / Use-Case based dimension in addition to existing four dimensions (Governance Dimension, Asset Dimension, Integrity Dimension and Use Dimension). The use-case / scenario based dimension works in top down approach. Future enhancements and requirements are been listed based upon the use case scenario while building data warehouse. The fifth dimension i.e scenario based dimension act as a logical tool to identify the gaps , to identify organizational and technical deficiency.

c) Semantic Data Warehouse Methodology:

Authors in [25] have proposed a conceptual generic integration approach having framework and ETL algorithm on existing semantic database. Authors have termed Data warehouse as Data Integration Systems (DIS) and Data Integration system has been termed as <G,S,M> as referred by authors in [25].

"G" represents global schema for heterogeneous data sources

"S" represents Local Schema associated with each source.

"M" represents mapping between local and global schema.

The output from this approach is a data warehouse system having classes and integrated ontology interacting with Framework and ETL algorithm on existing semantic database. The generic integration approach is been applied to form a data warehouse system. The resultant architecture is the outcome of this approach which is as:

 ISSN: 2454-4248
251 – 256

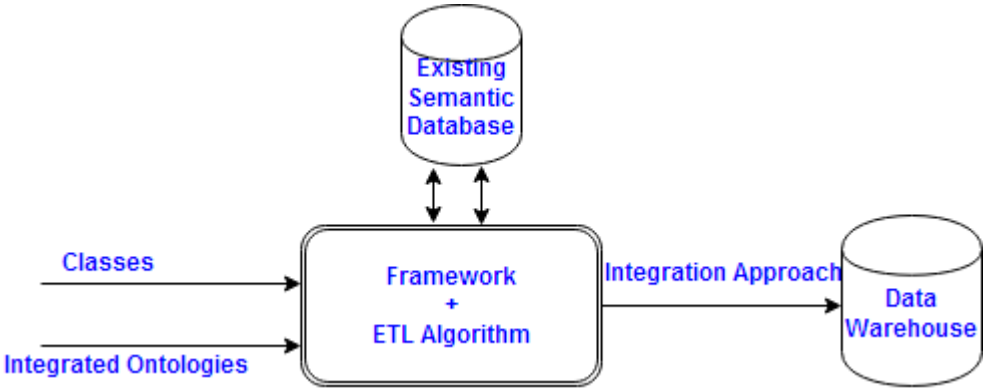


Fig: Assessment output from Semantic Data Warehouse Methodology

d) Kimball Data Warehouse Approach in Health Care Research

Authors in [26] have precisely identified the limitations in Kimball data warehouse architecture when it applied to health care domain. Kimball approach is based upon fact tables and dimensional tables. The concept of fact tables is to focus on a particular business domain / metrics and at the same time it ignores the correlation between the processes. Authors have identified that in case of new metrics requirement, that cannot be accommodated in the existing schema structure of Kimball data warehouse. Complete medical history of patient needs to be fetched from across the hospital data available. It is not only a single process, but multiple components needs to be fetched and at the same time correlation needs to be established. The assessment outcome of the research done in [26] is a extension to the Kimball approach which focus on identify unique identifier across the database (not on table wise). The identifier will not be unique to table only but is unique across database. The model provided by authors to overcome the limitation is as :

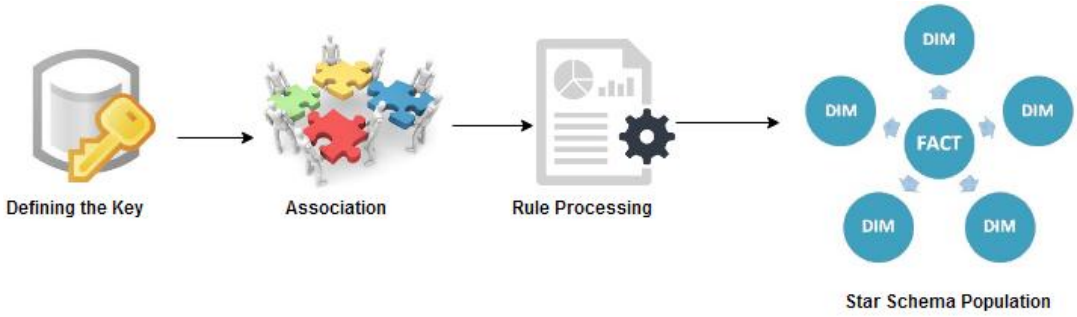


Fig: Interrelation centric DWH model for Health Care

e) Decision making methodologies – A comparative analysis for sales data mart

The outcome of the research carried out in [27] is the best approach for decision making in the sales data mart. Authors have used rank matrix method and arithmetic mean method in data mart and proposed an Extraction Transformation and Loading for effective decision making in sales data mart. Three methodologies for decision making have been analyzed for sales data mart and best methodology has been identified based upon quantitative techniques to be used in sales data mart. Based upon the research on decision matrix, decision analysis and markov decision model- authors have come to a conclusion that Decision making methodology is the best technique for decision making in sales data mart.

ATDM Methodology for Enterprise Data Warehouse

7. Conclusion:

The Big Data Analytics using the above Research Methodology can be used by a Financial Institution for tapping the unlimited business growth potential in Retail Segment. The Retail Segment is the upcoming business driver for the financial institutions as the Corporate, SME and Agriculture segments are witnessing single digit or negligible growth in times of uncertainty and protectionism. The retail segment has an added advantage that it has least non-performing assets among all segments.

ISSN: 2454-4248
251 – 256

References:

[1]. N. Almoqren, "The Motivations for Big Data Mining Technologies Adoption in Saudi Banks," 2016.
[2]. S. Alouneh, I. Hababeh, F. Al-hawari, and T. Alajrami, "Innovative Methodology for Elevating Big Data Analysis and Security," 2016.
[3]. T. Dong, B. Yang, and T. Tian, "Volatility Analysis of Chinese Stock Market Using High-Frequency Financial Big Data," *2015 IEEE Int. Conf. Smart City/SocialCom/SustainCom*, pp. 769–774, 2015.
[4]. A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. Van Der Voort, and M. Janssen, "Antecedents of Big Data Quality An Empirical Examination in Financial Service Organizations," pp. 116–121, 2016.
[5]. D. Kriksciuniene, M. Liutvinavicius, V. Sakalauskas, and D. Tamasauskas, "Research of customer behavior anomalies in big financial data," *2014 14th Int. Conf. Hybrid Intell. Syst. HIS 2014*, pp. 91–96, 2003.
[6]. X. Ma, Z. Fu, Y. Jiang, M. Yang, H. Stephen, L. Vegas, and L. Vegas, "On a Cyberinfrastructure Platform for Multidisciplinary , Data-intensive Scientific," 2016.
[7]. A. Mandloi, "Big Data analytics with case study on financial organization," *IT Business, Ind. Gov. (CSIBIG), 2014 Conf.*, p. 1, 2014.
[8]. P. C. Mondal, R. Deb, and M. N. Huda, "Transaction Authorization from Know Your Customer (KYC) Information in Online Banking," pp. 523–526, 2016.
[9]. S. O'Halloran, S. Maskey, G. McAllister, D. K. Park, and K. Chen, "Big Data and the Regulation of Financial Markets," *Proc. 2015 {IEEE/ACM} Int. Conf. Adv. Soc. Networks Anal. Mining, {ASONAM} 2015, Paris, Fr. August 25 - 28, 2015*, pp. 1118–1124, 2015.
[10]. S. Pang, "The Intelligent Control Model and Application for Commercial Bank Systems Emergence under Risk Status Based on Big Data," 2016.
[11]. S. Singh, A. Singh, and R. Kumar, "A constraint-based biometric scheme on ATM and swiping machine," *2016 Int. Conf. Comput. Tech. Inf. Commun. Technol. ICCTICT 2016 - Proc.*, pp. 74–79, 2016.
[12]. S. Sobolevsky, I. Sitko, R. T. Des Combes, B. Hawelka, J. M. Arias, and C. Ratti, "Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain," *Proc. - 2014 IEEE Int. Congr. Big Data, BigData Congr. 2014*, pp. 136–143, 2014.
[13]. N. Sun, J. G. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE: A framework for big data-based banking customer analytics," *IBM J. Res. Dev.*, vol. 58, no. 5/6, p. 4:1-4:9, 2014.
[14]. G. Xiao, "Research on Data Processing of Bank Credit System," *Commun. Syst. Netw. Technol. (CSNT), 2012 Int. Conf.*, pp. 905–908, 2012.
###################################### Methodology ############
[23]. A. Alrefae and J. Cao, "Web-based Real-Time Decision Support System," pp. 2–5, 2014.
[24]. E. Begoli, T. F. Chila, and W. H. Inmon, "Scenario-driven architecture assessment methodology for large data analysis systems," *2013 IEEE Int. Syst. Conf.*, pp. 51–55, 2013.
[25]. N. Berkani, S. Khouri, and L. Bellatreche, "Generic methodology for semantic data warehouse design: From schema definition to ETL," *Proc. 2012 4th Int. Conf. Intell. Netw. Collab. Syst. INCoS 2012*, no. i, pp. 404–411, 2012.
[26]. R. Hart and A. M.-H. Kuo, "Meeting Health Care Research Needs in a Kimball Integrated Data Warehouse," *2016 IEEE Int. Conf. Data Sci. Adv. Anal.*, pp. 697–705, 2016.
[27]. A. Prema, V. T. Clara, and A. Pethalakshmi, "A Comparative Analysis of Decision Making Methodologies in Sales Data Mart," *Comput. Commun. Technol. (WCCCT), 2014 World Congr.*, pp. 108–113, 2014.
[28]. B. Scholtz, C. Cilliers, and C. Ferreira, "The ATDM Methodology to Support the Design and Implementation of an Enterprise Data Warehouse," *Enterp. Syst. Conf.*, pp. 1–9, 2013.
[29]. R. Tardío, A. Mate, and J. Trujillo, "An iterative methodology for big data management, analysis and visualization," *2015 {IEEE} {International} {Conference} {Big} {Data} ({Big} {Data})*, pp. 545–550, 2015.
[30]. http://www.eiminstitute.org/library/eimi-archives/volume-1-issue-5-july-2007-edition/measuring-the-data-warehouse