Data Warehouse Architecture – Leading the next generation Data Science

Rahul Reddy Nadikattu[#1]

University of the Cumberlands, Department of Information Technology
6178 College Station Drive, Williamsburg, KY 40769

Abstract

The present study emphasizes the importance of the data warehouse as an important tool to maintain both ancient and present data, The study provides insight into the significant components of data warehouse architecture with its specific usage in the information science domain. In recent year ears, there has been a rapid expansion of the applicative properties of data warehousing owing to its wage in mobile and other social media platforms, In the current scenario, there has been a gradual shift in big data science towards green computing to overcome the environmental issues which have resulted in establishing data centers. The centers are gaining impute importance across the global especially in countries like America which is technology0oriented and is serving to maintain long-lasting working environments..

Keywords *— Data warehouse, Big Data, Computing, Data Transformation, Meta data, Apple, Green Computing.*

Introduction

According to my Research, a data warehouse is a system designed to store both ancient and present data that is collected for either single or multiple sources. The data collected is used in organizations mainly for making the decision or predicting the future. Designing a data warehouse is essential because if poorly done, it could lead to the collection and usage of inaccurate data. The usage of data warehouses has gradually increased across the globe to benefit the different sectors. In America, you can find more beneficiaries especially in health care, tourism, banking and biomedical sectors (Khan et al., 2013). The implementation of data warehousing in America has effectively influenced the customer relationship, improved financial performance, uplifting the operational efficacies. The ongoing efforts to promote data warehouse in America has envisioned a new business intelligence tool which has contributed towards the growing economy of the nation (Chen and Edward, 2012). The data can, in return, lead to reduced productivity and slow down the growth of an organization. A data warehouse gets characterized by subject-orientation, integration, time-variation, and non-volatile. The significant components of data warehouse architecture include;
Data Warehouse Database

The basic need for a data warehouse is the centralization of the data in a single database unit. The relational database usually enhances the scalability and sharing of memory. However the relational database has some limitations which are addresses by the use of the multidimensional database. Besides, new structural indexes get used to increase operability speed.

Sourcing, Acquiring, Cleaning and transforming tools :
The acquisition and transformation tools are employed to convert, brief, and change data into a single unit in the warehouse. The functions of these tools include; discarding data that is unwanted in the warehouse before it is loaded, altering the names of data drawn from various sources, populating the missing information by use of defaults, carrying out summary analysis, and de-duplication of data.

Metadata :
Metadata is data that helps in the definition of the data available in the warehouse. Its purpose is to generate, maintain, and manage the warehouse data. As well, it helps in the specification of data depending on its origin, functionality, value, and characteristics. It also outlines the transformation and processing process of data.

Query tool :
The basic need for data warehousing in an organization is decision making; the query tools enable users to relate with the systems in the data warehouse. These tools get divided into tools for reporting, tools for developing an application, tools for mining data, and the OLAP tools.

Data buses :
The buses are used in a data warehouse to determine the data flow methodology. The flow of data is subdivided into meta flow, down-flow, in-flow, up-flow, and out-flow.

Forms of data transformation
Data transformation is a model of altering the structure and format of data. The commonly employed process of change is the ETL process. Data modification is the following form :

Extraction :

This is the process of acquisition of data for their various sources and duplicating them into their intended destinations. It aims to alter the format of data making it compatible with existing data.

Data translation :

Translation involves changing data format from their previous systems to suit the current methods of operation. The translation may be combined with mapping to increase the understandability and usefulness of data.

Data consolidation :

The process entails breaking down massive data into small, manageable units. It also requires the filtration, aggregation, and summarization of data

Data encryption :

If the data processed has uniquely identifiable information, or could lead to privacy alteration, then it should be strongly encrypted for security purposes.

Trends in the data warehouse

Following the continued centralization of data in the data warehouse systems and applications, many organizations are integrating their systems with new technologies to ease work overload. Some of these trends include;

- Increased warehousing capabilities due to the use of mobile applications, internet-enabled sensors, and social media platforms, which are generating big data. Organizations are reacting to these changes by increasing warehouse capabilities to tackle the data load.

- Consolidation of data both physically and logically to curb their costs. Compression of data and the introduction of servers that can handle significant volumes of data is also a means of data consolidation.

- The use of Hadoops has led to the optimization of data in the warehouses through processing.

- The use of systems engineered to optimize massive data, for instance, the Oracle Big Data Appliance, leads to data transformation.

- The use of database analytics has employed a wide range of tools for data analysis. These databases reduce the back to back data movement in the system, enhancing a smooth flow of data and optimization.

Big Data :

Big data is an accumulation of data in

large volumes, and that is growing exponentially faster as time goes by. The data growth becomes complex with time to a point the ancient means of managing data become less efficient. This data is categorized into structured, unstructured, and semi-structured. Structured information is the form of data that is kept, retrieved, and processed in a fixed structure. Unstructured data is a type of data with an unfixed format, and despite the considerable size, it has many challenges in processing and derivation of values. These data entail the integration of videos and images. Semi-structured data is data represented in either form of structured or unstructured.

Uses of Big Data :

The generation of big data has changed the way businesses operate. Traditional analysis tools have presented organizations with adverse challenges following the significant data generation.That has forced individuals and organizations to adopt technologies like Hadoop. Some of the applications of big data entail; In the education sector, big data has revolutionized changes through program customization for learning. The program is designed to improve the student results by the use of data gathered from the student's history. Also, adequately analyzed data is used for grading and predicting the careers of students.Big data has helped a big deal in the health sector by reduction of the treatment costs as some diagnoses are unnecessary. As well, big data has helped in forecasting the emergency of pandemics and coming up with measures to combat the same.

- In the entertainment sector, big data gets used to forecast the audience's interest and make adverts that directly focus on the target market.

- The data in the banks is generated at a very high rate and gets mostly used to detect banking malpractices such as money laundering. The use of big data majorly depends on how an individual or organization comes up with the data and how important it is to them

Demand of Big Data in Organizations :

The utilization of big data in the organization fully transforms the way its business operations get carried out. Analysis of big data helps in increasing the performance of an organization by discovering new data patterns. Most organizations now realize that utilization of every data that comes in, by use of data analytics, can profit them in a big way. The current technology has provided systems that enable the analysis of big data and get instant results. Every data evolution has led to organizations adapting new technologies to organize their data as the return profits are high. Another factor forcing organizations to embrace big data analytics technology is the ability to make fast and effective decisions. Again, data analytics is a chance for reducing operational costs due to the identification of efficient means of doing business.

Green computing

How organizations are making data center green. Many organizations are going green through the provision of healthy environments. They are taking the lead in the use of renewable energy to run their

data centers. They are also using natural means to ventilate their servers. Renewable energy is generated from natural sources such as wind by use of windmills and solar using solar panels. The power generated enhances less emission of pollutants in the atmosphere. Power Usage Effectiveness (PUE)is the ultimate determiner of how efficient a data center is. PUE drives a comparison of the energy used in the non-computing activities to the power that gets used to gear machines. Many data centers generally produce overhead energy of 70%, according to Colocation America, 2018. Therefore, many organizations are taking measures to make their data center environment friendly. The application of natural energy in the data centers has fostered its efficiency.

Apple

Apple is one of the leading organizations that has successfully implemented green computing despite facing considerable criticism of the degradation of their phones' batteries intentionally. The company clearly stated their intent on being conscious of energy. They have implemented programs for recycling energy needed by devices used besides mining. The use of clean energy mainly operates Apple data centers. The company has invested in solar energy, and according to the press release, they have almost reached 100% coverage with renewable energy in every one of their data centers. With such effort, it has led to many prominent organizations following the same route to achieving green computing.

Conclusion

Many data centers also act as the data warehouse, and out of the wide variety of operations carried out therein, the environment is unfriendly. Recently, data centers are taking measures to bring in the future close by going green to make their environment more adaptable and efficient. To maintain a long-lasting working environment, data centers are investing in clean sources of energy like wind, solar, and geothermal. The usage of data warehouses has greatly influenced the American economy by designing business intelligence tools.

Acknowledgments

References

[1] Amir G, Murtaza H, (2015) "*Beyond the hype: Big data concepts, methods, and analytics*".

[2] Chen, Edward T. (2012) "Implementation Issues of Enterprise data Warehousing and Business Intelligence

[3] Colocation America (2018) How data centers are going green. Retrieved from: https://www.colocationamerica.com/blog/how-data-centers-save-energy

[4] Jia R., Xu S., Peng C. (2013) Research on Real Time Data Warehouse Architecture. Communications in Computer and Information Science, vol 392. Springer, Berlin, Heidelberg. 2013. Elsevier.

[5] Khan A, Ehsan N, Mirza E, & Sarwar SZ (2013). "*Integration between Customer Relationship Management (CRM) and Data Warehousing*". Procedia Technology, 1, 239-249.

[6] Nathan M, & James W, (2015) Big Data: Principles and best practices of scalable realtime data systems.

[7] Philips Woods (2017) Why Big Data demands new technology. Retrieved from: https://www.itproportal.com/features/why-big-data-demands-new-technology

[8] Robert Laberge (2011) The data warehouse mentor: Practical data warehouse and business intelligence insights.