

Comparative Study On Data Warehouse And Big Data

Ms.AshwiniS.Rupnawar
Author
ashwinirupnawar27@gmail.com

Mr.Yogesh D. Shinde
Guide
yshinde147@gmail.com

Abstract:-Data warehouse is a large store of data accumulated from a wide range of sources within a company and used to guide management decisions. We have comparatively study about data warehouse and big data.storing data in big data there are more complexity as compare to data warehouse. Big Data is a new term used to identify the datasets that due to their large size and complexity. A data warehouse is a special type of database. It is used to store large amounts of data, such as analytics, historical, or customer data, and then build large reports and data mining against it.In data warehouse using hadoop complexity, structured and unstructured data easily stored. The Big Data challenge is becoming one of the most exciting opportunities for the next years. A DataWarehouse is a structured relational database, where you aim to collect together all the interesting data from multiple systems. When putting data into a warehouse, you need to clean and structure it at the point of insertion. This cleaning and structuring process is usually called ETL – Extract, Transform, and Load. The data warehouse approach is helpful because then your data looks clean and simpleBig data and your data warehouse can be a powerful team, providing many new analytic applications that enterprises need to stay competitive. But you will need to make some changes to your existing infrastructure, tools, and processes to integrate big data into your current environment.

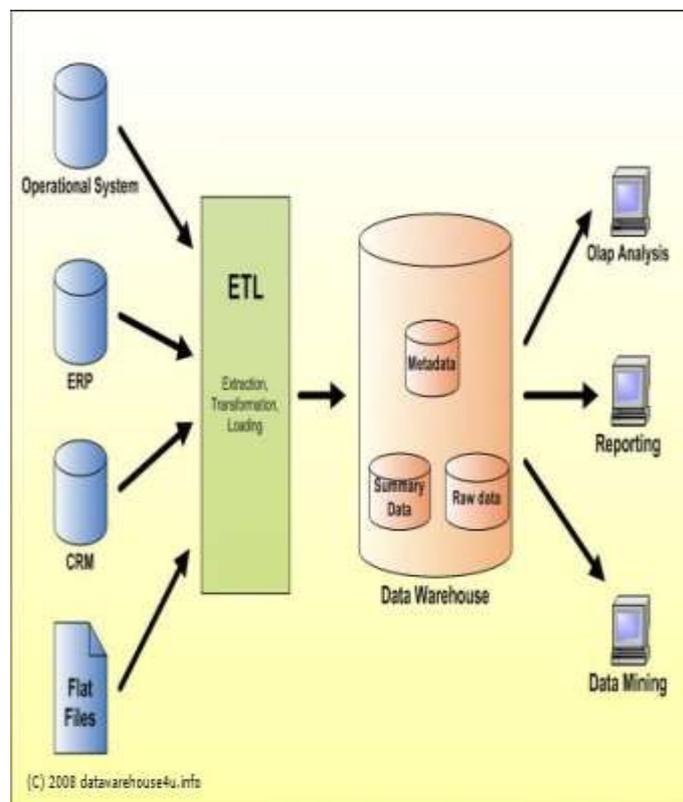
I. INTRODUCTION

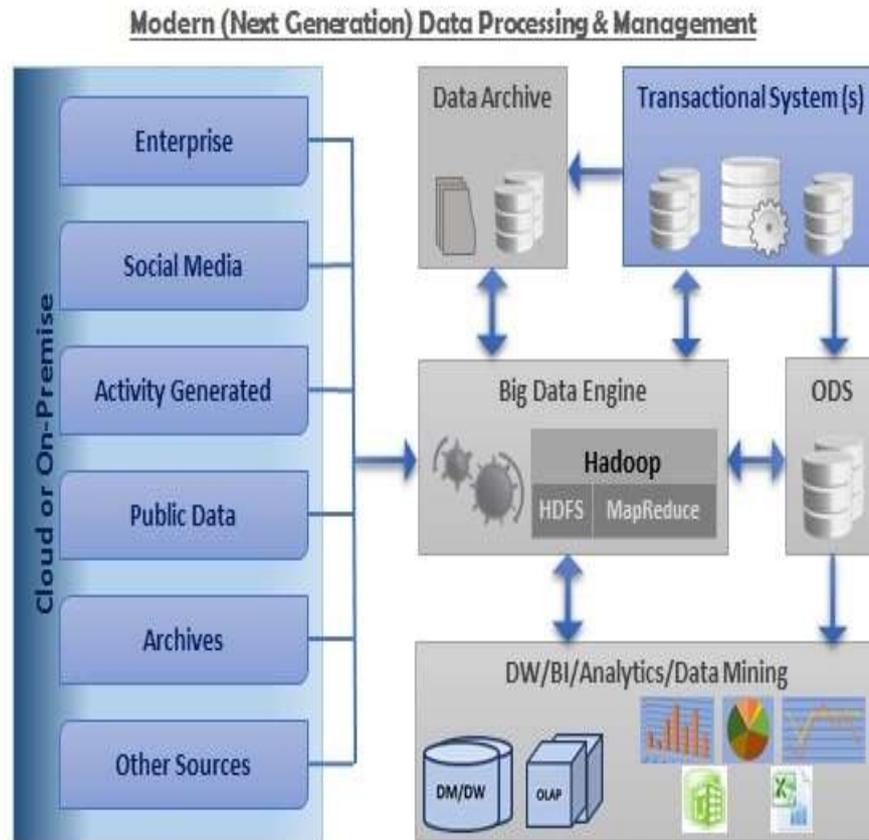
The Data Warehouse is typically used to store historical data (partially sourced from the ODS) and queries are usually subject based and deep analytics are used to help gain business value out of integrated data sources and possibly external data sources as well (competition data, weather data,etc).

I'm assuming that you know what Big Data is, as it's not just necessarily big in volume only, because your data

warehouse can certainly span multiple petabytes of data. Big data is usually unstructured, or semi-structured data -- meaning not structured in a way that is stored and processed easily in a relational database; more specifically processed in SQL. It's important not to mix data warehousing databases with transactional databases in the same instance, whether you are dealing with MySQL or Oracle.

Architectural difference between Data warehouse and big data





Big Data is a term applied to data sets whose size is beyond the ability of commonly used tools to capture, manage and process the data within a tolerable elapsed time. But Data-warehouse is a collection of data marts representing historical data from different operations in the company.

It means Big Data is collection of large data in a particular manner but Data-warehouse collect data from different department of a organization. However Data-warehouse require efficient managing technique. Conceptually these are same only at one factor that they collect large amount of information.

So when we compare a big data solution to a data warehouse, what do we find? We find that a big data solution is a technology and that data warehousing is an architecture. They are two very different things. A technology is just that – a means to store and manage large amounts of data. A data warehouse is a way of organizing data so that there is corporate credibility and integrity. When someone takes data from a data warehouse, that person knows that other people are using the same data for other purposes. There is a basis for reconcilability of data when there is a data warehouse.

While an immediate response to a query isn't an absolute requirement, a response time of five seconds or fewer is often desired and can be delivered by a distributed computing cluster running Hadoop, MapReduce, for example. Traditional data warehouses are burdened by the perception that a batch process runs overnight to produce

results that are available to decision makers the following morning. Another concept popular with new analytics platforms is the ability to do reiterative queries -- i.e., run a query, get results, then run a second query against those results and/or converge with others, for example. Data warehouses aren't known for ad hoc querying. However, data warehousing vendors are catching up in these areas.

II. MAIN BENEFITS OF DATA WAREHOUSE

Benefits from a successful implementation of a data warehouse include:

- Enhanced Business Intelligence
- Increased Query and System Performance
- Business Intelligence from Multiple Sources
- Timely Access to Data
- Enhanced Data Quality and Consistency
- Historical Intelligence
- High Return on Investment

1. Enhanced Business Intelligence:-

Insights will be gained through improved information access. Managers and executives will be freed from making their decisions based on limited data and their own “gut feelings”. Decisions that affect the strategy and operations of organizations will be based upon credible facts and will be backed up with evidence and actual organizational data. Moreover, decision makers will be better informed as they will be able to query actual data will retrieve information based upon their personal needs. In addition, data warehouses and related business intelligence can be used to

can be applied directly to business processes including marketing segmentation, inventory management, financial management, and sales.

2. Increased Query and System Performance:-

Data warehouses are purposely designed and constructed with a focus on speed of data retrieval and analysis. Moreover, a data warehouse is designed for storing large volumes of data and being able to rapidly query the data. These analytical systems are constructed differently from operational systems which focus on creation and modification of data. In contrast, the data warehouse is built for analysis and retrieval of data rather than efficient upkeep of individual records (i.e. transactions). Further, the data warehouse allows for a large system burden to be taken off the operational environment and effectively distributes system load across an entire organization's technology infrastructure.

3 Business Intelligence from Multiple Sources:-

For many organizations, enterprise information systems are comprised of multiple subsystems, physically separated and built on different platforms. Moreover, merging data from multiple disparate data sources is a common need when conducting business intelligence. To solve this problem, the data warehouse performs integration of existing disparate data sources and makes them accessible in one place. Consolidating enterprise data into a single data repository alleviates the burden of duplicating data gathering efforts, and enables the extraction of information that would otherwise be impossible. Additionally, the data warehouse becomes the "single view of truth" for the enterprise rather than the multiple truths that can come from reporting on individual subsystems.

4 Timely Access to Data:-

The data warehouse enables business users and decision makers to have access to data from many different sources as they need to have access to the data. Additionally, business users will spend little time in the data retrieval process. Scheduled data integration routines, known as ETL, are leveraged within a data warehouse environment. These routines consolidate data from multiple source systems and transform the data into a useful format. Subsequently, business users can then easily access data from one interface. Further, consumers of data will be able to query data directly with less information technology support. The wait time for information technology professionals to develop reports and queries is greatly diminished as the business users are given the ability to generate reports and queries on their own. The use of query and analysis tools against a consistent and consolidated data repository enables business users to spend more time performing data analysis and less time gathering data.

5. Enhanced Data Quality and Consistency:-

Datawarehouse implementation typically includes the conversion of data from numerous source systems and data files and transformation of the disparate data into a common

format. Data from the various business units and departments is standardized and the inconsistent nature of data from the unique source systems is removed. Moreover, individual business units and departments including sales, marketing, finance, and operations, will start to utilize the same data repository as the source system for their individual queries and reports. Thus each of these individual business units and departments will produce results that are consistent with the other business units within the organization. Subsequently the overall confidence in the organization's data is substantially increased.

6. Historical Intelligence:-

Data warehouses generally contain many years worth of data that can neither be stored within nor reported from a transactional system. Typically transactional systems satisfy most operating reporting requirements for a given time-period but without the inclusion of historical data. In contrast, the data warehouse stores large amounts of historical data and can enable advanced business intelligence including time-period analysis, trend analysis, and trend prediction. The advantage of the data warehouse is that it allows for advanced reporting and analysis of multiple time-periods.

7.High Return on Investment:-

Return on investment (ROI) refers to the amount of increased revenue or decreased expenses a business will be able to realize from any project or investment of capital. Subsequently, implementations of data warehouses and complementary business intelligence systems have enabled business to generate higher amounts of revenue and provide substantial cost savings. According to a 2002 International Data Corporation (IDC) study "The Financial Impact of Business Analytics", analytics projects have been achieving a substantial impact on a business' financial status. Additionally, the study found that business analytics implementations have generated a median five-year return on investment of 112% with a mean payback of 1.6 years. Of the businesses included in the study, 54% have had a return on investment of 101% or more.

III. CONCLUSION

Data warehouse is powerful as compare to Big Data .data warehouse as a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.it is handle any kind of Data more easily as compare to Big Data . The data warehouse must be able to control who can access what data within the data warehouse

IV. ACKNOWLEDGEMENTS

We are very thankful to Dr. RanjitPatilSir , Vice Principal of Computer Sciences Department, D.Y.Patil ACS college,Pimpri Pune for his valuable comments and suggestion throughout writing of this paper. This helped us in improving the quality of this paper.

V. REFERENCES

- [1] The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling Jul 1, 2013 by Ralph Kimball and Margy Ross.
- [2] Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners Hardcover – 12 Aug 2014 by Jared Dean (Author)
- [3] <http://bi-insider.com/portfolio/benefits-of-a-data-warehouse/>
- [4] <https://www.youtube.com/watch?v=KtfPAeu1OMM>
- [5] Wolfgang Hummer, Andreas Bauer, Gunnar Harde: XCube – XML For Data Warehouses, DOLAP'03, November 7, 2003, USA
- [6] J.NorbertoMazón, J.Trujillo and J. Lechtenbörger.” A Set of QVT Relations to Assure the Correctness of Data Warehouses by Using Multidimensional Normal Forms.”, 25th International Conference on Conceptual Modeling, Tucson, AZ, USA, November 6-9, 2006, Proceedings. Volume 4215 of Lecture Notes in Computer Science, pages 385-398, Springer, 2006.
- [7] R. Kimball and M. Ross, The Data Warehousing Toolkit, 2nd edition: John Wiley, 2002