**Database design challenges for Big Data Applications**

**Seminar Report**

**Introduction:-**

Advances in the technology bound have resulted in major disruptions and transformations in the enterprise-wide information technology infrastructures. For the past 30 years, classical database management systems have maintained an excited pace in realizing significant efficiencies in dealing with the huge amount of information that needs to be maintained to model the operational characteristics of large-scale enterprises. Database research and development advances have a primary part to focus on advanced data models, declarative query languages, high throughput transaction processing and database reliability, etc. In the years especially in the 1990s, data warehousing and data analysis emerged as a major research and technology achievement.

While enterprises are struggling with the problem of poor database scalability, a new challenge has emerged that has further deformed the capability of modern IT infrastructures. This challenge has been labelled as the ‘big data’ problem. While earlier DBMSs focused on modelling operational characteristics of enterprises, big data systems are now expected to model huge amounts of heterogeneous and complex data. Classical approaches to data warehousing and data analysis are no longer of use to deal with both the scale of data and the sophisticated analysis that need to be conducted often in real time for e.g. online fraud detection. None of the commercial DBMS and Data Warehousing technologies provide a fair solution in this regard which is evident from the efforts led by companies such as Facebook, Google and Baidu to build appropriate solutions. Clearly, scalable data management and complex data analytics in the context of big data have emerged as a new research bound in the upcoming future. As an accepted challenge in the context of ‘big data’, since enterprises maintain huge amounts of sensitive user interaction data for its clients, it is compulsory that adequate mechanisms are provided to ensure security and privacy of user data.

**2- Motivation:-**

In today’s competitive business world the various aspects of business are fused. Change in one part has direct or indirect effect on the other part. Within an organization, this complication makes it problematic for business leaders to depend only on experience to make decisions. They need to depend on data - structured, unstructured or semi-structured - to back up their decisions.

Current tools don’t lend themselves to classy data analysis at the scale the user requires. Tools like SAS, R, and Mat lab support the significant analysis but they are not considered for the large datasets and neither DBMS can handle the data that are arrived at high rates. To channel this gap the “Big Data” came into the scene. Big Data can analyze and visualize their data effectively.

Big data helps to store all information related to business which comprises of Customer Feedback, vendor, trends etc.

With the help of Big Data health care system store all the information of a patient to get more effective view for insight into care direction, health management & outcome.

Big data can also be the key to actually deploying condition based maintenance program and improve estimating and scheduling of assets.

**3- Challenges:-**

**Understanding the Big Data and its proper Utilization:**

It is a challenging task in most industries that deal with big data just to understand the data that is available to be used and shaping the best use of that data based on the industry’s, strategy, and tactics. Also, these types of studies need to be performed on a constant basis as the data landscape changes at an ever increasing rate.

**High Scalability:**

Data is scaling at an extraordinary rate and it is a challenging issue as data volume is increasing very faster than calculated resources, and CPU speeds are fixed. In the past, this challenge was diminished by processors getting faster, by following Moore’s law, to deliver us with the resources needed to deal with increasing volumes of data. But, there is a vital shift underway now.

**Timeliness:**

The other side of speed is size. The larger the data is to be managed, the longer it will take to examine. The design of a system should be efficient so that it can deals with size is likely also to result in a system that can process data set faster.

**Privacy and Security:**

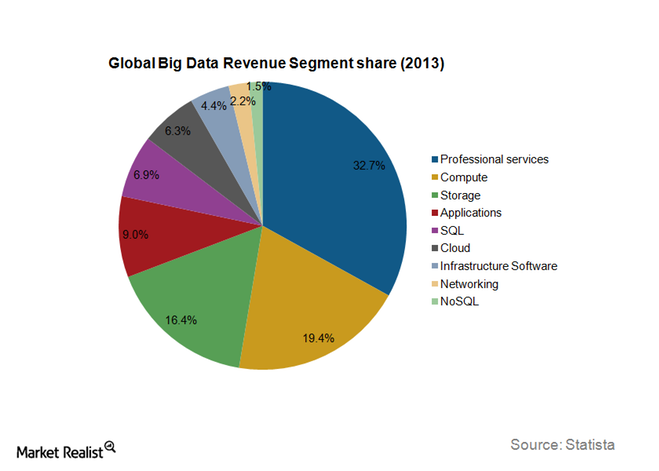
The privacy of data is another major and important challenge in the framework of Big Data. There is also a public fear regarding the wrong use of private data, particularly through relating of data from many sources. Handling privacy is both a technical as well as a sociological problem, which must be realized to take advantage of Big Data

**Data Integration:**

Big Data Integration is a multidimensional and multidisciplinary and requires multi-technology method which poses a big challenge.

**4- Related Work:-**

Big data as the name shows, means a collection of data or datasets that are so large and Complex that it’s outside the range of database management tools to process. It’s usually unstructured in the sense that data is not organized in a predefined manner. This is the reason why organizations prefer help from organizations that possess domain expertise in big data technologies. Big data technologies are better because they enable organisations to handle huge datasets and generate useful information and insights from it with minimal delay time. Teradata (TDC) and EMC (EMC) are the companies playing an integral role in providing data backup and storage to the big data industry.

Chart 21.png

Professional services made up the largest segment, contributing approximately ~33%, followed by computing and storage at 19% and 16% of the total big data revenues, respectively. Social media represent the multi-directional flow of data that has to be important, secure, scalable, crucial, and able to be stored. Mobile devices are the platforms where more than 40% of social media updates are happening. This explains why leading social media players like

Facebook (FB), Google+ (GOOG), Twitter (TWTR) and Linkedin (LNKD) have developed applications for Android and iOS. Users have started to move from Yahoo (YHOO) and Gmail services to Facebook, hangouts, and Whatsapp showing social media and mobile applications interdependence. Cloud lets the data like photos, videos be accessed across multiple devices from anywhere at any time provided there’s an internet connection. All this data generated is huge. It’s fast moving, complex and valuable. It can be made useful through big data or NoSQL databases. Going forward, SMAC is expected to be the foundation of enterprise applications. As a result, it would impact all applications used by organizations. It seems highly unlikely to think of an enterprise resource planning (ERP) product with no cloud presence or mobile capabilities as in current scenario, each organization is making mobile compatible applications.

**5- Loopholes/deficiencies/Left out issues-**

**Scale-**

Today, Facebook ingests 500 terabytes of new data every day, a Boeing 737 will generate 240 terabytes of flight data during a single flight across the U.S. the worldwide use of smartphones, the data they create and consume, sensors fixed into everyday objects will soon result in billions of new, constantly updated data feeds containing environmental, location, and others information, including audio, text and video. The data size has increased at a very high rate ranging to petabytes (1 petabyte= 1,024 terabytes). Our database management system (DBMS) finds it difficult to handle such a huge volume of data. To sort this, RDBMS adds more central processing units (CPUs) or more memory to the database management to scale up vertically.

**Behavior of data-**

Big Data isn’t just numbers, dates, and string. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media. Traditional database system was designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure. They are designed and structured to accommodate structured data such as weblog sensor and financial data. As applications have evolved to serve large volumes of users, and as an application development practice has become active, the traditional use of the relational database has become the liability for many companies rather than an enabling factor in their business. Big Data databases, such as MongoDB, solve these problems and provide companies with the means to create tremendous business value.

**Velocity-**

Also, big data is generating at a very high velocity. Here DBMS lacks in velocity because it’s designed for steady data use rather than rapid growth data storage. Even if DBMS is used to handle such a big volume of data it will turn out to be very expensive.

As a result, the inability of relational databases to solve this big data problem led to the emergence of new technologies.

**Complexity and retrieval of big data-**

With the continuously growing of data, its storage has become very complex. Hence require a different technique to retrieve data where the RDBMS lacks because a strong analysis will be required for data retrieval. Since relational databases use the logical and physical independence data schemes for data retrieval, which is not required for big data.

**6- Suggested approach for solving the problem:-**

Relational databases such as Oracle ORCL -0.71%, IBM IBM -0.54%’s DB2 and Microsoft MSFT -7.22%’s Access, form the backbone for data storage and management in most organizations today. While relational databases provide good structure and accessibility for most data, they also have limitations which have given rise to a new class of databases that address specific needs for dealing with extremely large or complex data resources.

These new databases don’t use the tables, fields and rows found in relational databases, and they don’t require establishing a schema (a highly-ordered database plan) to set them up called NoSQL (not-only SQL) they are designed to overcome specific data management challenges such as providing rapid data access to power real-time applications, bringing order to data in non-traditional formats, or avoiding the costs and turnaround time required to develop a conventional database schema.

The rise of NoSQL databases presents challenges for established database providers, and new options for data owners.

Do you need a NoSQL database? Today, probably not. Tomorrow, that may change. How will you know? You don’t need to learn the ins and outs of all the new databases now available, but you should get familiar with key types and the situations best suited to each.

Five major classes of NoSQL databases have emerged: column families (also known as

wide-column stores or columnar databases), document, graph, key-value and XML (also known as native XML). Here are the basics of each type, with an eye to the kinds of data analysis that each fits best.

**Softwares:-**

Key-value: These are designed for simple and easy development of applications.

They are good for situations where you need a working application developed fast, and all other considerations come second. Big names: Basho Technologies’ Riak, and Redis

Column Families: These are the NoSQL databases that most resemble conventional relational databases. They store structured data in individual, columns (rather than tables). In place of tables, these databases use groups of columns. They are good for machine-generated data, structured data sources too big to fit on a single computer, and for rapid data queries. If you are thinking about rapid precision analytics on machine data. Big names: Apache APA +2.59% Cassandra and Apache HBbase.

**7- Conclusion:-**

We are living in an era of Big Data. Proper and effective analysis of large volumes of data will lead to faster advances in many scientific disciplines and improve the profitability and success of many enterprises. While the future benefits of Big Data are real and significant, and some initial growth has been achieved in some of the projects, there remain many technical challenges that must be solved to fully realise the hidden potential of Big data. The larger size of the data is a major challenge however, there are others. The technical challenges faced by the relational database are found in a large variety of application domain, and therefore impose a huge cost. Furthermore, these challenges will require a transformative solution, and will require a wide range of tools, methodologies and applications to deal with. In order to get the future benefit of Big Data, these things have to be taken into strong consideration so that full potential can be derived to gain a competitive edge. Big data and its applications have not removed the need for good database design. Indeed, the database designer has more things to consider like backup and recovery, index management, multiple methods of data access, and SQL limitations. The best news is that advances in database hardware and software can speed up data queries to a remarkable degree.

**8- References:-**

<http://www.forbes.com/sites/metabrown/2015/06/30/next-generation-databases-take-on-big-data-management-challenges/#1571a9cc2d23>

<https://www.purdue.edu/discoverypark/cyber/assets/pdfs/BigDataWhitePaper.pdf>

<http://marketrealist.com/2014/07/traditional-database-systems-fail-support-big-data/>

<http://nsr.oxfordjournals.org/content/1/1/27.full#sec-9>