BIG DATA DESIGN AND ITS CHALLEGES

CH.Sirisha[*1] **Y.Pratima**[*2]
[*1,2]Assistant Professor Department of Computer Science & Engineering
[*1,2]CMR College Engineering&Technology-Kandlakoya, Hyderabad

***Abstract*— Big Data is still a maturing and evolving discipline. Big data databases and files have scaled beyond the capacities and capabilities of commercial database management systems. Structured representations become a bottleneck to efficient data storage and retrieval. Gartner has noted four major challenges (the four Vs): increasing volume of data, increasing velocity (e.g., in/out and change of data), increasing variety of data types and structures, and increasing variability of data. A fifth V: value is suggested, which is the contribution big data has to decision making. Add to these the increasing number of disciplines and problem domains where big data is having an impact and one sees an increase in the number of challenges and opportunities for big data to have a major impact on business, science, and government, social networks.**

Keywords-BigData,DataQuality,Security,Architecture.

I. INTRODUCTION

I. Big Data

Big data refers to data sets that are too large or complex for traditional data-processing application software to adequately deal with. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value The term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem.

A. What are The 5'vs Of Big Data?

The five V's of big data. Volume, velocity, variety, veracity and value are the five keys to making big data a huge business. "Big data is like sex among teens.

- Volume refers to the vast amount of data generated every second. Just think of all the emails, Twitter messages, photos, video clips and sensor data that we produce and share every second. We are not talking terabytes, but zettabytes or brontobytes of data. On Facebook alone we send 10 billion messages per day, click the like button 4.5 billion times and upload 350 million new pictures each and every day. If we take all the data generated in the world between the beginning of time and the year 2000, it is the same amount we now generate every minute! This increasingly makes data sets too large to store and analyze using traditional database technology. With big data technology we can now store and use these data sets with the help of distributed systems, where parts of the data is stored in different locations, connected by networks and brought together by software.

- Velocity refers to the speed at which new data is generated and the speed at which data moves around. Just think of social media messages going viral in minutes, the speed at which credit card transactions are checked for fraudulent activities or the milliseconds it takes trading systems to analyze social media networks to pick up signals that trigger decisions to buy or sell shares. Big data technology now allows us to analyze the data while it is being generated without ever putting it into databases.

- Variety refers to the different types of data we can now use. In the past we focused on structured data that neatly fits into tables or relational databases such as financial data (for example, sales by product or region).

In fact, 80 percent of the world's data is now unstructured and therefore can't easily be put into tables or relational databases—think of photos, video sequences or social media updates. With big data technology we can now harness differed types of data including messages, social media conversations, photos, sensor data, video or voice recordings and bring them together with more traditional, structured data.

- Veracity refers to the messiness or trustworthiness of the data. With many forms of big data, quality and accuracy are less controllable, for example Twitter posts with hashtags, abbreviations, typos and colloquial speech. Big data and analytics technology now allows us to work with these types of data. The volumes often make up for the lack of quality or accuracy.

But all the volumes of fast-moving data of different variety and veracity have to be turned into value! This is why value is the one V of big data that matters the most.

Value refers to our ability turn our data into value. It is important that businesses make a case for any attempt to collect and leverage big data. It is easy to fall into the buzz trap and embark on big data initiatives without a clear understanding of the business value it will bring.

B. Big data can deliver value in almost any area of business or society:

- It helps companies to better understand and serve customers:

Examples include the recommendations made by Amazon or Netflix.

- It allows companies to optimize their processes: Uber is able to predict demand, dynamically price journeys and send the closest driver to the customers.

- It improves our health care: Government agencies can now predict flu outbreaks and track them in real time and pharmaceutical companies are able to use big data analytics to fast-trackdrugdevelopment.

C. Difference and Challenges between big data and real-time big data:-

- A.Big data is characteristic by multi-source heterogeneous data, widely distributed, dynamic growth, and ―data mode after the data‖ In addition to having all the characteristics with big data,

- real-time big data has its own characteristics. Compared with the big data, when it comes to data integration real-time big data has higher requirements in data acquisition devices, data analysis tools, data security, and other aspects. The following introduces from data integration, dataanalysis, data security, data management and benchmarking.

B. Data Collect

With the development of internet of things [10] and Cyber Physical System (CPS) [11], the real time of data processing requires higher and higher. Under the big data environment, numerous sensors and mobile terminals disperse in different data management system which makes data collection itself a problem. In RTDP system, its real time data collection faced makes data integration facing many challenges.

II. GRAND CHALLENGES IN BIG DATA

There are numerous difficulties in tackling the capability of enormous information today, running from the structure of handling frameworks at the lower layer to investigation implies at the higher layer, just as a progression of open issues in logical research. Among these challenges, some are brought about by the qualities of enormous information, a few, by its present examination models and strategies, and a few, by the restrictions of current information handling frameworks. In this area, we quickly portray the serious issues and difficulties.

- **Data complexity**

The study of data complexity metrics is an emergentarea in the field of data mining and is focus on theanalysis of several data set characteristics to extractknowledge from them. This information used tosupport the election of the proper classification algorithm.

- **System complexity**

Enormous information preparing frameworks appropriate for dealing with a assorted variety of information types and applications are the way to supporting logical research of enormous information. For information of tremendous volume, complex structure, and inadequate esteem, its handling is gone up against by high computational unpredictability, long obligation cycle, and continuous necessities. These necessities not just posture new difficulties to the structure of framework models, registering structures, and handling frameworks, however additionally force stringent limitations on their operational effectiveness and vitality utilization.

III. ARCHITECTURE OF BIG DATA

The design of system architectures, computing frameworks, processing modes, and benchmarks for highly energy-efficient big data processing platforms is the key issue to be address in system complexity. Solving these problems can lay the principles for designing, implementing, testing, and optimizing big data processing systems. Their solutions will form an important foundation for developing hardware and software system architectures with energy-optimized and efficient distributed storage and processing.

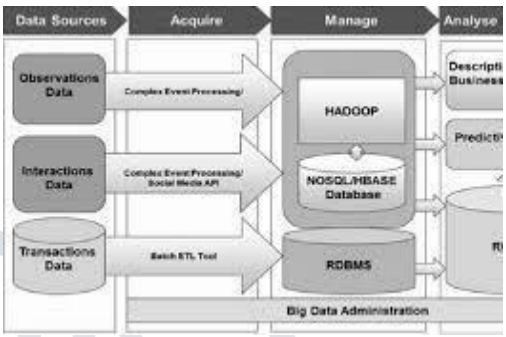

Fig:1 Architecture of Big data

3.1 Architecture of Big data

Data sources: The architecture of Big data system architecture the development of tables which are put away in the Cloud and of mobile infrastructures has prompted a noteworthy increment in the size and multifaceted nature of informational indexes. The information combination environments should therefore incorporate different systems for the entrance and capacity of a tremendous amount of shifted information. The accompanying characterization can be made

Data storage: the information gathered is put away in NoSQL/SQL databases, or Log Management frameworks for logs;

Data Transformation: so as to stack information into the preparing stage, it should initially be changed by utilizing: import/send out devices (SQL/NoSQL seller explicit apparatuses), Sqoop (information source to Hadoop information change device), Log the board instruments;

Data Processing: both organized and unstructured information are joined with the goal that bunch preparing or constant handling can be performed. Information Warehousing and Processing at that point produce usable information for information utilization.

Data Analysis: can can beperformed utilizing: DWs: guarantee the fundamental essential data. New usefulness must be included for the better joining of unstructured information sources and for fulfilling the dimension of execution required by investigation stages. So as to perform key choices, operational examination must be isolated from profound investigation, which makes utilization of authentic information.

Data Consumption: the aftereffects of the information examination must be exhibited is a meaningful and open shape to the last clients. Inquiry reports or representation diagrams can be utilized.

IV. CONCLUSION

Big data has made a strong impact in almost every sector and industry today. In this paper, we have briefly reviewed the grand challenges that big data brings us. We close by a few suggestions on how to make a big data project successful. It is no secret that in big data research and applications, industry is ahead of academia. The successful applications of big data in industry point to the following necessary conditions for a big data project to be successful. Firstly, there must be very clear requirements, regardless of whether they are technical, social, or economic. Secondly, to efficiently work with big data, we will need to explore and find the kernel structure or kernel data to be processed. Finding kernel data and structures, which are small enough and yet can characterize the behavior and properties of the underlying big data, is non-trivial because it is very domain-specific. Thirdly, a top-down management model should be adopted. Although a bottom-up approach may allow us to solve some niche problems, the isolated solutions often cannot be put together into a complete solution. Finally, the goal should be to solve the entire problem by an integrated solution, rather than striving for isolated successes in a few aspects.

REFERENCES

[1] David Gewirtz for DIY-IT | April 20, 2016 -- 12:47 GMT (18:17 IST) | Topic: Big Data Analytics

[2] E. Mcnulty, "Understanding Big Data: The Seven V's", Dataconomy,May 22, 2014.

[3] Gartner, "Big Data Strategy Components: Business Essentials", October 9, 2012.

[4] Gartner, "IT glossary: big data" [webpage on the Internet]. Stamford, CT; 2012. [5] Canada Inforoute, "Big Data Analytics in health", White Paper, Full Report, April 2013.