BIG DATA

Ms. Niki Mehta
Assistant Professor,
Swami Sahajanand College of Computer Science,
Bhavnagar, India.

Received Feb. 16, 2015 Accepted March. 01, 2015

ABSTRACT

Big data is buzzword that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information. Although big data doesn't refer to any specific quantity, Big Data is the ocean of information we swim in every day- vast petabytes and exabytes of data. The process of research into massive amounts of data to reveal hidden patterns and secret correlations named as big data analytics. These useful informations for companies or organizations with the help of gaining richer and deeper insights and getting an advantage over the competition.

Key words **:** data

*A research paper presented at National Seminar held at Department of Commerce, M. K. Bhavnagar University on 14th February.

Introduction:

Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data.

We define Big Data as a cultural, technological, and scholarly phenomenon that rests on the interplay of:

(1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets.
(2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.
(3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy

What is big data?

Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques. In most enterprise scenarios the data is too big or it moves too fast or it exceeds current processing capacity. Big data has the potential to help companies improve operations and make faster, more intelligent decisions.

Why Bigdata?

when the size and performance requirements for data management become significant design and decision factors for implementing a data management and analysis system. For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration

While the term may seem to reference the volume of data, that isn't always the case. The term big data, especially when used by vendors, may refer to the technology (which includes tools and processes) that an organization requires to handle the large

amounts of data and storage facilities. The term big data is believed to have originated with Web search companies who needed to query very large distributed aggregations of loosely-structured data.

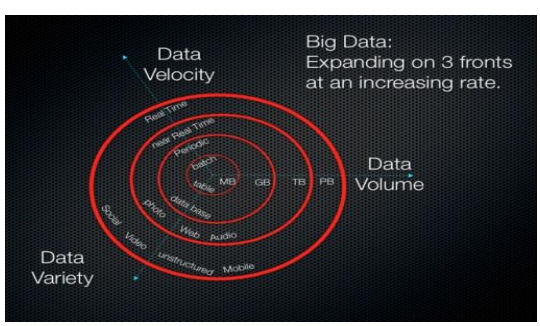
Example:- An example of big data might be petabytes (1,024 terabytes) or exabytes (1,024 petabytes) of data consisting of billions to trillions of records of millions of people—all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured data that is often incomplete and inaccessible.

Big Data is characterized by the following 3 Vs:

• Volume - the vast amount of data generated every second that are larger than what the conventional relational database infrastructures can cope with.

• Velocity - the frequency at which new data is generated, captured, and shared.

• Variety - the increasingly different types of data (from financial data to social media feeds, from photos to sensor data, from video capture to voice recordings) that no longer fits into neat, easy to consume structures.



BIG DATA FRAMEWORK

We have demonstrated a simple framework below to look at the key components of a Big Data system in order to work through many architectural decisions as you explore the world of big data. Big data often brings four new and very different considerations in enterprise architecture:

 Data sources have a different scale many companies work in the multi-terabyte and some in petabyte arena.

 Speed is critical nightly ETL (extract-transform-load) batches are insufficient and real-time streaming from solutions like s4 and Storm are required.

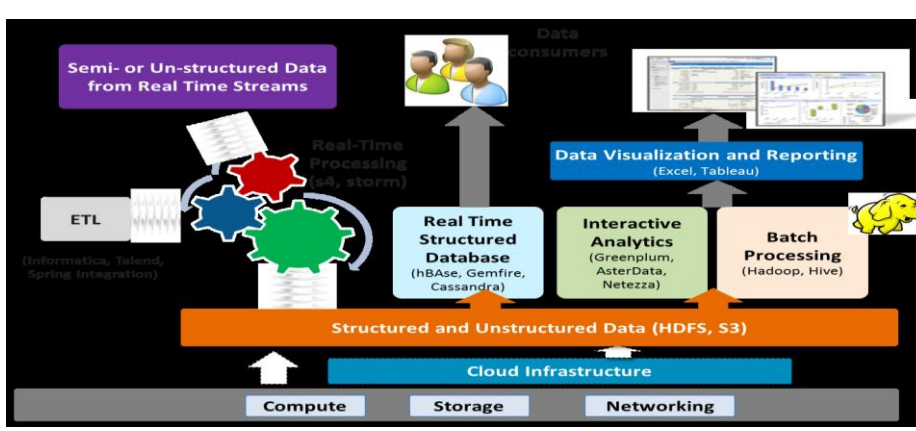
 Storage models are changing solutions like HDFS (Hadoop Distributed File System) and unstructured data stores like Amazon S3 provide new options.

 Multiple analytics paradigms and computing methods must be supported:

o Real-time database and analytics: These are typically in-memory, scale-out engines that provide low-latency, cross-data center access to data, and enable distributed processing and event-generation capabilities.

o Interactive analytics: Includes distributed MPP (massively parallel processing) data warehouses with embedded analytics, which enable business users to do interactive querying and visualization of big data.

o Batch processing: Hadoop as a distributed processing engine that can analyze very large amounts of data and apply algorithms that range from the simple (e.g. aggregation) to the complex (e.g. machine learning).



Holistic View of a Big Data Framework

Conclusion

The biggest challenge does not seem to be the technology itself – as this is evolving much more rapidly than humans – but rather how to make sure we have enough skills to make effective use of the technology at our disposal and make sense out of the data collected. And before we get to that stage, we need to resolve many legal issues around intellectual property rights, data privacy and integrity, cyber security, exploitation liability and Big Data code of conduct. Like in many other technological areas, customs and ethics around Big Data possibilities and excesses take time to develop. Promises of Big Data include innovation, growth and long term sustainability. Threats include breach of privacy, property rights, data integrity or personal freedom. So provided Big Data is exploited in an open and transparent manner, delivery of the promise of Big Data is not far ahead of us.

Reference

[1] World's data will grow by 50X in next decade, IDC study predicts http://www.computerworld.com/s/article/9217988/World_s_data_will_grow_by_50X_in_next_decade_IDC_study_predicts

[2] International Journal of Advanced Research in Computer Science and Software Engineering

[3] From Databases to Big Data by Sam Madden – Article published in IEEE Internet Computing magazine

[4] Considerations for Big Data: Architecture and Approach by Kapil Bakshi – Paper published in IEEE

[5] http://en.wikipedia.org/wiki/Big_data

Proper words in proper places make the true definition of a style.

-Jonathan Swift