
Big Data

Saranya T, Nivetha D

Assistant professor Department of BCA PSGR Krishnammal college for women Coimbatore, India
Student BCA 3rd year Psgr krishnammal college for women Coimbatore, India

Abstract: Big data is a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process the data with low-latency. And it has one or more of the following characteristics – high volume, high velocity, or high variety. Big data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media - much of it generated in real time and in a very large scale.

Keywords: big data, black box data, social media data, stock exchange data, power grid data, transport data, search engine data

I. Introduction

Due to the arrival of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year. If you pile up the data in the form of disks it may fill an entire football field. The same amount was created in every two days in 2011, and in every ten minutes in 2013. This rate is still growing enormously. Though all this information produced is meaningful and can be useful when processed,

II. Big Data

A. What is big data

Big Data refers to a collection of large datasets that cannot be processed using traditional computing techniques. It is not a single technique or a tool, rather it involves many areas of business and technology..

B. Benefits of big data

- By using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.
- By using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.
- By using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

III. Operational Big Data

Operational big data generally refers to systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored.

“NoSQL Big Data systems” are designed in such a way to take advantage of new cloud computing architectures that have emerged over the past to allow massive computations to be run inexpensively and efficiently. This makes operational big data workloads much easier to manage, cheaper, and faster to implement.

Some “NoSQL systems” can also provide insights into patterns and trends based on real-time data with minimal coding and without any need for data scientists and additional infrastructure.

IV. Analytical Big Data

Analytical big data generally refers to systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data.

MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

These two classes of technology are complementary and frequently deployed together.

	Operational	Analytical
Latency	1 ms - 100 ms	1 min - 100 min
Concurrency	1000 - 100,000	1 - 10
Access Pattern	Writes and Reads	Reads
Queries	Selective	Unselective

Fig :-1

A. Applied fields

Big data generally involves the data produced by different devices and applications. Given below are some of the fields that come under the umbrella of Big Data.

- **Black Box Data:** Refers to a component of helicopter, airplanes, and jets, etc. It captures voices of the flight crew, recordings of microphones and earphones, and the performance information of the aircraft.
- **Social Media Data:** Refers to any Social media such as Facebook and Twitter hold information and the views posted by millions of people across the globe.
- **Stock Exchange Data:** The stock exchange data generally holds information about the 'buy' and 'sell' decisions made on a share of different companies made by the customers.
- **Power Grid Data:** The power grid data similarly holds information consumed by a particular node with respect to a base station
- **Transport Data:** Transport data includes large amount of model, capacity, distance and availability of a vehicle.
- **Search Engine Data:** Search engines data retrieve lots of data from different databases.

B. Traditional enterprise approach

In this approach, an enterprise will have a computer to store and process big data. For storage purpose, the programmers will take the help of their choice of database vendors such as Oracle, IBM, etc. the user interacts with the application, which in turn handles the part of data storage and analysis.

V. Limitations

This approach works fine with those applications that process less voluminous data that can be accommodated by standard database servers, or up to the limit of the processor that is processing the data. But when it comes to dealing with huge amounts of scalable data, it is a hectic task to process such data through a single database bottleneck Authors and Affiliations.

A. Google Solution

Google solved the problem using an algorithm called MapReduce. This algorithm resolves the problem by dividing the task into small parts and assigns them to many computers, and collects the results from them which when integrated, form the result dataset.

B. Hadoop

Using the solution provided by Google, Doug Cutting and his team developed an Open Source Project called HADOOP. Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel with others. In short, Hadoop is used to develop applications that can perform complete statistical analysis on huge amounts of data.

Hadoop is an open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application generally works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

VI. Conclusion

For a smater experience in day today life , and the advent in new technology day by day , the impact of Big data is impeccable

References

- [1]. *Big Data: A Revolution That Will Transform How We Live, Work, and Think* Author: Viktor Mayer-Schonberger and Kenneth Cukier
- [2]. *The Big Data-Driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits* Author: Russel Glass & Sean Calahan
- [3]. *The Human Face of Big Data* Author: Rick Smolan