International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 12
ISSN: 2454-4248
409 – 411

Big Data

Jushmita Mishra
Lecturer, MCA & BCA Deaprtment
MPC(A) College, Baripada ,Orissa, India
mishrajushmita@yahoo.co.in

Abstract: Big data implies performing computation and database operations for massive amounts of data, remotely from the data owner's enterprise .Since a key value proposition of big data is access to data from multiple and diverse domains, security and privacy will play a very important role in big data research and technology. Making effective use of big data requires access from any domain to data in that domain, or any other domain it is authorized to access.

Big data to date has been all about the technologies-NOSQL databases, HADOOP in memory processing etc. However at the end of the day, Big data is about how to create value from data.

*****

I. INTRODUCTION

Big data is a term for data sets that are so large or complex. The term "Big Data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data.

II. WHAT IS BIG DATA

Big data is a term that describes the large volume of data structured, semistructured and unstructured.

Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets.

Structured data first depends on creating a data model of the types of business data that will be recorded and how they will be stored, processed and accessed. Structured data is often managed using Structured Query Language(SQL) a programming language Created for managing and querying data in relational database management systems.

Unstructured data is all those things that can't be so readily classified and fit into a near box: Photos and graphic, images, videos, streaming instrument data, web pages, PDF files, power point presentations, emails, and word processing documents.

Semi-structured data is a cross between the two. It is a type of structured data, but lacks the strict data model structure. With semi-structured data, tags or other types of markers are used to identify certain elements within the data, but the data does not have a rigid structure. For example word processing software now can include metadata showing the author's name and the date created. With the bulk of document just being unstructured text.XML and other markup languages are often used to manage semi-structured data.

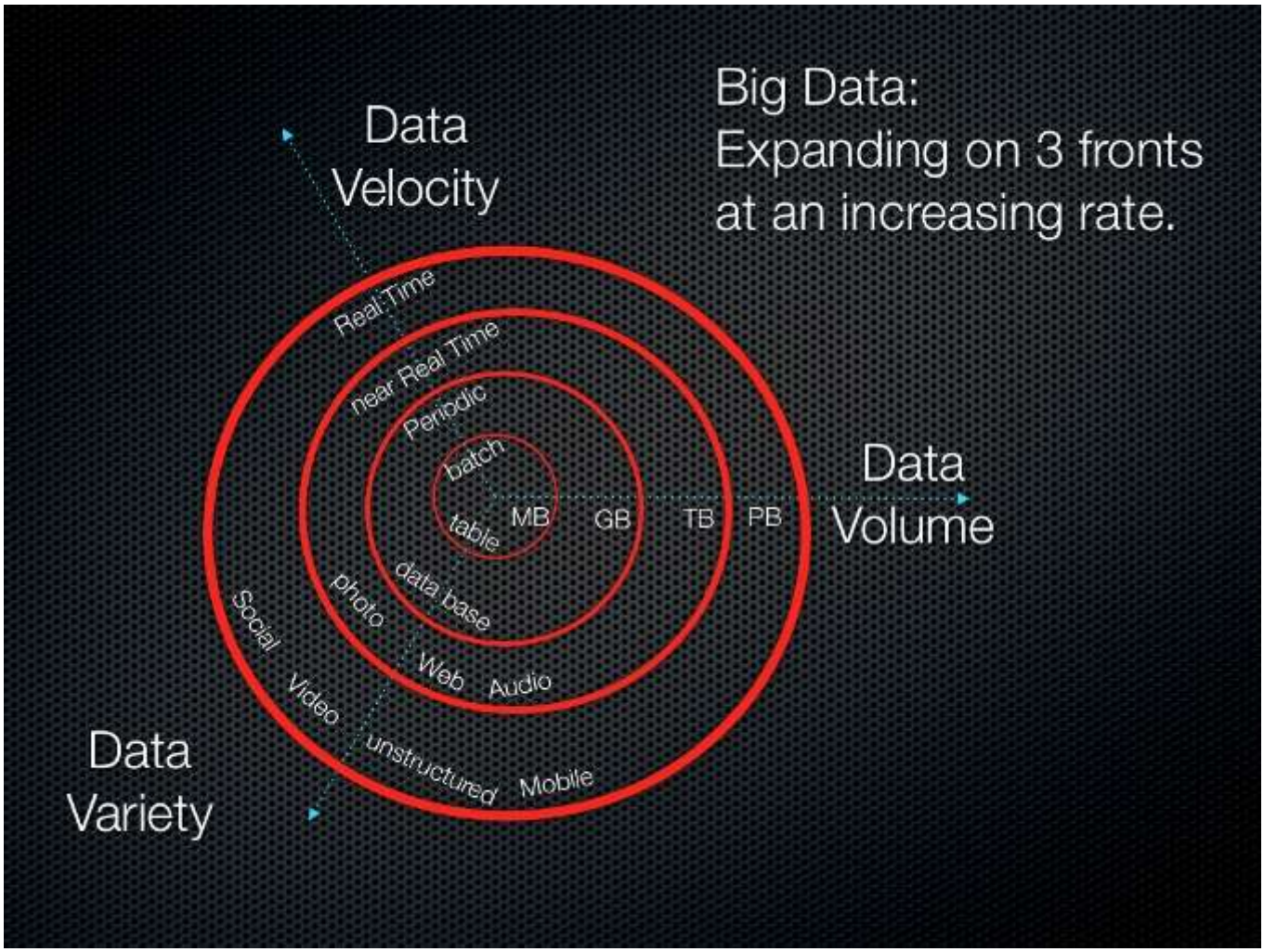
The term has been in use since1990's, with some giving credit to John Mashey for coining or at least making it popular. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture , manage and process data within a tolerable elapsed time. The main focus of big data is on unstructured data.

III. CHARACTERISTICS OF BIG DATA

Big data is often characterized by 3Vs: the extreme volume of data, the wide variety of data types and the velocity at which the data must be processed. Although big data doesn't equate to any specific volume of data, the term is often used to describe terabytes, petabytes and even exabytes of data captured over time.

Velocity is also meaningful, as big data analysis expands into fields like machine learning and artificial intelligence.

 ISSN: 2454-4248
409 – 411



IV. WHY WE NEED BIG DATA

Many organizations don't realize they have a big data problem or they simply don't think of it in terms of big data. In general an organization is likely to benefit from big data technologies when existing databases and applications can no longer scale to support sudden increases in volume, variety and velocity of data.

Failure to correctly address big data challenges can result in escalating costs, as well as reduced productivity and competitiveness. On the other hand, a sound big data strategy can help organizes reduce costs and gain operational efficiencies by migrating heavy existing workloads to big data technologies.

V. HOW DOES BIG DATA WORK

With new tools that address the entire data management cycle ,big data technologies make it technically and economically feasible, not only to collect and store large data sets, but also to analyze them in order to uncover new and valuable insights. In most cases, big data processing involves a common data flow from collection of raw data to consumption of actionable information.

Collect:- Collecting the raw data transactions, logs ,mobile devices and more is the first challenge many organizations face when dealing with big data. A good big data platforms makes this step easier allowing developers to ingest a wide variety of data-from structured to unstructured at any speed from real time to batch.

Store:- Any big data platform needs a secure scalable and durable repository to store data prior or even after processing tasks. Depending on the specific requirements, we may also need temporary stores for data in-transit.

Process & Analyze:- This is the step where data is transformed from its raw state into a consumable format-usually by means of storing, aggregating , joining and even performing more advanced functions and algorithms. The resulting data sets are then stored for further processing or made available for consumption via business intelligence and data visualization tools.

Consume & Visualize :- Big data is all about getting high value, actionable insights from the data assets.

 ISSN: 2454-4248
409 – 411

VI. WHY IS BIG DATA IMPORTANT

The importance of big data doesn't resolve around how much data we have, but what we do with it. We can take data from any source and analyze it to find answers that enable

1)cost reductions 2)time reductions 3)new product development and optimization 4)Smart decision making. When we combine big data with high powered analytics we can accomplish business-related tasks such as:

- Determining root causes of failures, issues and defects in near real time.
- Generating coupons at the point of sale based on the customer's buying habits.
- Recalculating entire risk portfolios in minutes.
- Detecting fraudulent behavior before it affects the organization.

VII. CONCLUSION

Big data with predictive analytics, high performance computing systems machine learning, and other strategies have been used in the past and will continue to be used heavily in the future of computational physics. By using these big data related systems, engineers and scientists have been able to more easily design cars, airplanes and other vehicles. They have also been able to more accurately predict daily weather as well as natural disasters. Big data analytics has affected the field of computational physics almost since computational physics was created. Computational physics with big data will continue to improve the quality of everyday life even though there will always be challenges.

REFERENCES

[1] WIKIPEDIA.COM
[2] SAS.COM
[3] AMAZONE.COM