International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 02 | Feb-2018 www.irjet.net p-ISSN: 2395-0072

Applications of Big Data: Review Paper

Ms. Swati S.Tawade[1]

[1] *Lecturer, Computer Engineering, V.P.M's Polytechnic, Thane, Maharashtra, India.*

Abstract - *For organizations of all sizes, data management has shifted from an important competency to a critical differentiator that can determine market winners and has-beens. Big Data refers to technologies and initiatives that involve data that is too diverse, fast-changing or massive for conventional technologies, skills and infra- structure to address efficiently. Said differently, the volume, velocity or variety of data is too great. Analyzing big data allows analysts, researchers, and business users to make better and faster decisions using data that was previously inaccessible or unusable. Using advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyze previously untapped data sources independent or together with their existing enterprise data to gain new insights resulting in better and faster decisions. This paper examines the challenges facing big data today and going forward including, but not limited to: data capture and storage; search, sharing, and analytics; big data technologies; data visualization; architectures for massively parallel processing; data mining tools and techniques; machine learning algorithms for big data; cloud computing platforms; distributed file systems and databases; and scalable storage systems.*

Key Words: *Big data, data mining, cloud computing.*

1. INTRODUCTION

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.

'Big Data' describes data sets so large and complex they are impractical to manage with traditional software tools.

Specifically, Big Data relates to data creation, storage, retrieval and analysis that is remarkable in terms of volume, velocity, and variety:

Volume. A typical PC might have had 10 gigabytes of storage in 2000. 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 US; the proliferation of smart phones, the data they create and consume; sensors embedded into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including video. 1 2 3

Velocity. Click streams and ad impressions capture user behavior at millions of events per second; high-frequency stock trading algorithms reflect market changes within microseconds; machine to machine processes exchange data between billions of devices; infrastructure and sensors generate massive log data in real-time; on-line gaming systems support millions of concurrent users, each producing multiple inputs per second.

Variety. Big Data data isn't just numbers, dates, and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media. Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.

2. ARCHITECTURE OF BIG DATA

Lambda Architecture is a useful framework to think about designing big data applications. Nathan Marz designed this generic architecture addressing common requirements for big data based on his experience working on distributed data processing systems at Twitter.

λ has three layers:

1. The Batch Layer manages the master data and pre computes the batch views
2. The Speed Layer serves recent data only and increments the real-time views
3. The Serving Layer is responsible for indexing and exposing the views so that they can be queried.

Some of the key requirements in building this architecture include:

- Fault-tolerance against hardware failures and human errors
- Support for a variety of use cases that include low latency querying as well as updates
- Linear scale-out capabilities, meaning that throwing more machines at the problem should help with getting the job done
- Extensibility so that the system is manageable and can accommodate newer features easily

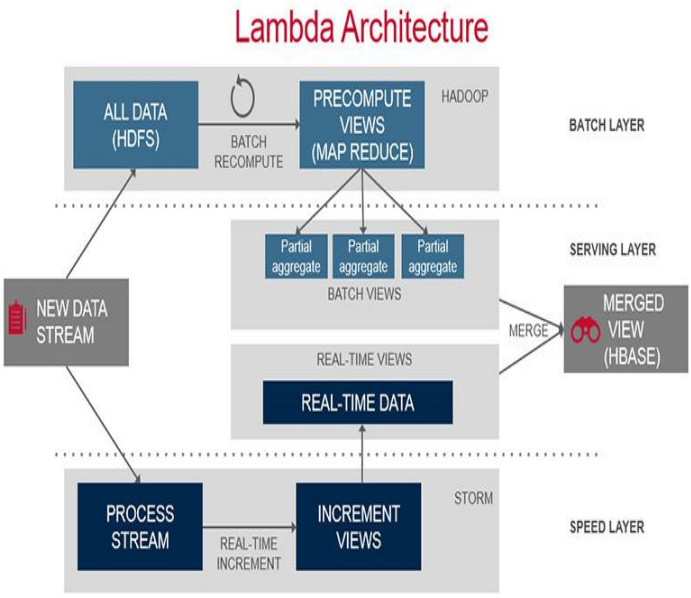

Fig -1: Lambda Architecture of Big Data

2.1. When to use this architecture

Consider this architecture style when you need to:

- Store and process data in volumes too large for a traditional database.
- Transform unstructured data for analysis and reporting.
- Capture, process, and analyze unbounded streams of data in real time, or with low latency.
- Use Azure Machine Learning or Microsoft Cognitive Services.

3. BIG DATA TECHNOLOGIES

- The bubble around Big Data has certainly started to burst and the coming year awaits reasonable developments in the applications of the Big Data world. Some of the trending big data technologies are discussed below:
- Hadoop: Doug Cutting, Mike Cafarella and team took the solution provided by Google(Mapreduce) and started an Open Source Project called HADOOP in 2005 and Doug named it after his son's toy elephant. Now Apache Hadoop is a registered trademark of the Apache Software Foundation.Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel on different CPU nodes. In short, Hadoop framework is capable enough to develop applications capable of running on clusters of computers and they could perform complete statistical analysis for a huge amounts of data.
- Spark: Hadoop and Spark are both Big Data frameworks – they provide some of the most popular tools used to carry out common Big Data-related tasks. Spark is reported to work up to 100 times faster than Hadoop . Spark handles most of its operations "in memory" – copying them from the distributed physical storage into far faster logical RAM memory. This reduces the amount of time consuming writing and reading to and from slow, clunky mechanical hard drives that needs to be done under Hadoop's MapReduce system.
- Data Lakes: A data lake is a large storage repository that holds a vast amount of raw data in its native format until it is needed. An "enterprise data lake" (EDL) is simply a data lake for enterprise-wide information storage and sharing.
- NoSQL Databases: NoSQL databases support dynamic schema design, offering the potential for increased flexibility, scalability and customization compared to relational software. That makes them a good fit for Web applications, content management systems and other uses involving large amounts of non-uniform data requiring frequent updates and varying field formats. In particular, NoSQL technologies are designed with "big data" needs in mind.
- Blockchain: Blockchain is a shared immutable ledger for recording the history of transactions. blockchain is an incorruptible digital ledger of economic transactions that can be programmed to record not just financial transactions but virtually everything of value.

4. APPLICATIONS OF BIG DATA

Big Data Contributions to Public Sector

In the public sectors, the major confrontations are the amalgamation and ability of the big data from corner to corner of various public sector units and allied unions. Big data provides a large range of facilities to the government sectors including the power investigation, deceit recognition, fitness interconnected exploration, economic promotion investigation and ecological fortification.

Big data is even used to examine the food based infections by the FDA. Big data results are fast which outputs to quicker well-being. Also in the investigation of a huge volume of communal complaints uses the big data analytics. This same analytics are utilized in the course of health check statistics in urgency and resourcefully for

quicker pronouncement manufacture and to become aware of mistrustful or falsified declarations.

Big Data Contributions to Learning

Big data has great influence in the education world too. Today almost every course of learning is present online. Along with the online learning, there are many examples of the use of big data in the education industry. Applications named as the Bubble Score allow teachers to convey multiple-choice assessments through mobile devices and notch up paper tests through the cameras of the mobile phones. Equipment like this usually assists teachers to send out the outputs to rank books and trail development all along distinct characteristics.

Big Data Contributions to Industrial and Natural Resources

The high demand of the natural sources on this earth is challenging the high volume as well as the velocity of big data. Similarly, a great quantity of data commencing the built-up industry is unexploited. The unused data avoids advanced eminence of merchandise, power competence, dependability, and improved income boundaries. In the natural wealth industry, big data enables for analytical modelling to sustain judgment creation that is used to consume and incorporate huge amounts of information from geographical information, graphical information, manuscript and chronological statistics. Big data has as well been worn in finding the solution to the development of confrontations and to grow aggressive improvements in the middle of former settlements.

Big Data Contributions to Transportation

In current times, vast volumes of statistics from area-oriented community networks and towering speediness statistics from telecoms have influenced journey policies a lot. Unfortunately, investigation of the appreciate voyage policy has not developed yet. Usually, transportation requirement representation is again oriented on defectively unstated fresh social media architectures.

Big Data Contributions to Banking Zones and Fraud Detection

Big data is hugely used in the fraud detection in the banking sectors. In banking sectors as the big data is implemented, it finds out all the mischief tasks done. It detects the misuse of credit cards, misuse of debit cards, archival of inspection tracks, venture credit hazard treatment, business clarity, customer statistics alteration, public analytics for business, IT action analytics, and IT strategy fulfillment analytics. The SEC uses this big data in order to keep a track of all the commercial market movements.

3. CONCLUSION

Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. With today's technology, it's possible to analyze your data and get answers from it almost immediately – an effort that's slower and less efficient with more traditional business intelligence solutions. The Age of Big Data is here, and these are truly revolutionary times if both business and technology professionals continue to work together and deliver on the promise.

REFERENCES

[1] Nada Elgendy, Ahmed Elragal,Big Data Analytics: A Literature Review Paper

[2] J.Archenaa E.A. Mary Anitb,A Survey of Big Data Analytics in Healthcare and Government, Procedia Computer Science

Volume 50, 2015

[3] A De Mauro, M Greco, M Grimaldi ,What is big data? A consensual definition and a review of key research topics, AIP Conference Proceedings 1644, 97 (2015)

[4] Yuri Demchenko, Cees de Laat, Peter Membrey, Defining architecture components of the Big Data Ecosystem, IEEE.

[5] AT Hashem, I Yaqoob, NB Anuar, S Mokhtar, A Gani,The rise of "big data" on cloud computing: Review and open research issues, Information Systems, 2015 – Elsevier.

[6] https://mapr.com/developercentral/lambda-architecture

[7] https://intellipaat.com/blog/7-big-data-examples-application-of-big-data-in-real-life/