

Handler De Datos

Urmil Lodaya^{*1}, Dharmil Fofariya^{*2}, Prem Shah^{*3}, Prof.Swapnil Molke^{*4}

^{*}Department of Computer Technology,

[#]Lecturer, Department of Computer Technology,
Shah & Anchor Kutchhi Polytechnic,
Mumbai, India

*dancerurmil@gmail.com, dharmilfofariya@gmail.com,
premshah935@gmail.com, swapnil.molke@gmail.com*

Abstract:-Big data is the term for data sets so large and complicated that it becomes difficult to process using traditional data management tools or processing applications. This paper reveals most recent progress on big data. It revolves around four important areas of analytics and Big Data, namely:-

- (i) Data management and supporting architectures;
- (ii) Model development and scoring;
- (iii) Visualization and user interaction; and
- (iv) Business models.

Lastly, Organizations are defining new initiatives and reevaluating existing strategies to examine how they can transform their businesses using Big Data. In the process, they are learning that Big Data is not a single technology, technique or initiative. Rather, it is a trend across many areas of business and 'technology.

I. INTRODUCTION TO PROJECT

Today, Society is becoming increasingly more instrumented and as a result, organization are producing and storing vast and large amounts of data. Maintain and gain insights from the produced data is a challenge and key to aggressive advantage. Analytics solutions that mine structured and for disorganized data are important as they can help organizations gain insights not only from their privately collected data, but also from large amounts of data publicly available on the Web. The ability to cross-relate private instruction on consumer preferences and amount with information from tweets, blogs, product evaluations, and data from social networks opens a wide range of possibilities for organizations to understand the needs of their customers, predict their wants and demands, and advance use of resources. This archetype is being popularly termed as Big Data.

II. BACKGROUND AND METHODOLOGY

Large volumes of data are generated by Organisations increasingly day by day as result of instrumented business processes, finance, accounting, web site tracking, sensors, monitoring of user activity, among other reasons. With the arrival of social network Web sites, Records are created by users of their lives by daily posting details of activities

they perform, events they attend, places they visit, pictures they take, and things they enjoy and want. This data deluge is often referred to as Big Data. A term that conveys the

challenges it poses on existing infrastructure with respect to storage, management, interoperability, governance, and analysis of the data..

III. KEY CHARACTERISTICS

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

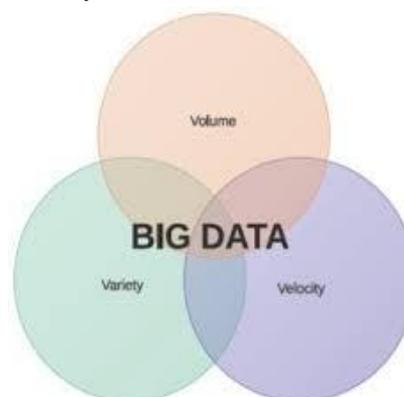


Figure 1. Big Data

1.1 Volume

A typical PC might have had 500 gigabytes to 1 terabytes of storage in 2000. To day Facebook ingests 500 terabytes of new data every day. a Boeing 737 will generate 240 to 250 terabytes of flight data during a single flight across the US. Use of the smart phones and the data they create and consume. And the sensor use into everyday object will soon result billions and more. Constantly updation of data feeds containing location, other information and including video also [3].

1.2 Velocity

Using internet millions of events done per second like clickstream and ad impression etc. high-frequency stock trading reflect market change within a microsecond. Now days machine to machine exchange the high-ended data between billions of device. Sensor generate massive log data in real-time.on-line gaming and social sight system supports millions of concurrent user and each producing multiple input per second [3].

1.3 Variety

Big Data is not just a date or number or string it's also combination of geospatial data,3D data, images ,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. Traditional database systems are also designed to operate on a single server, making increased capacity expensive and finite. As applications have evolved to serve large volumes of users, and as application development practices have become agile, the traditional use of the relational database has become a liability for many companies rather than an enabling factor in their business.Big Data databases (<https://www.mongodb.com/big-data-explained>)[3]

IV. BIG DATA FOR THE ENTERPRISE

With Big Data databases, enterprises can save money, grow revenue, and achieve many other business objectives, in any vertical.



Figure 2. Big Data[2]

1.4 Build new applications

Big data might allow a company to collect billions of real-time data points on its products, resources, or customers – and then repackage that data instantaneously to optimize customer experience or resource utilization. For example, a major US city is using MongoDB to cut crime and improve municipal services by collecting and analyzing geospatial data in real-time from over 30 different departments.

1.5 Improve the effectiveness and lower the cost of existing applications

Big data technologies can replace highly-customized, expensive legacy systems with a standard solution that runs on commodity hardware. And because many big data technologies are open source, they can be implemented far more cheaply than proprietary technologies. For example, by migrating its reference data management application to MongoDB, a Tier 1 bank dramatically reduced the license and hardware costs associated with the proprietary relational database it previously ran, while also bringing its application into better compliance with regulatory requirements.

1.6 Realize new sources of competitive advantage

Big data can help businesses act more nimbly, allowing them to adapt to changes faster than their competitors. For example, MongoDB allowed one of the largest Human Capital Management (HCM) solution providers to rapidly build mobile applications that integrated data from a wide variety of disparate sources.

1.7 Increase customer loyalty

Increasing the amount of data shared within the organization – and the speed with which it is updated – allows businesses and other organizations to more rapidly and accurately respond to customer demand. For example, a top 5 global insurance provider, MetLife, used MongoDB to quickly consolidate customer information from over 70 different sources and provide it in a single, rapidly-updated view.

V. CHALLENGES IN BIG DATA MANAGEMENT

In this section, we discuss current research targeting the issue of Big Data management for analytics. There are still, however, many open challenges in this topic. The list below is not exhaustive, and as more research in this field is conducted, more challenging issues will arise.

Traditional vs Big Data

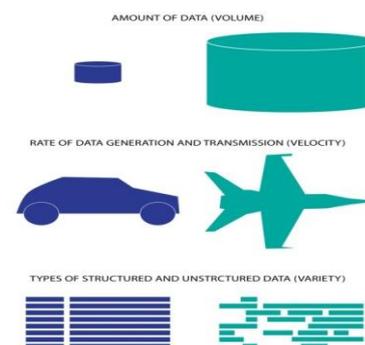


Figure 3. Traditional Vs Big data[4]

1.8 Data variety

How to knob an always increasing volume of data? Especially, The data is structured and unstructured, But when the data is unstructured we should know how to quickly extract meaningful content out of it? How to aggregate and correspond to streaming data from multiple sources?

1.9 Data storage

How to conveniently recognize and store important information extracted from unstructured data? How to store large volumes of data in a way it can be timely retrieved? Are current file systems optimized for the volume and variety demanded by analytics applications? If not, what new capabilities are needed? How to store information in a way that it can be easily migrate ported between data centers/Cloud providers?

1.10Data integration

New protocols and compound for integration of information that are able to manage information of different nature (structured, unstructured, semi-structured) and sources Data Processing and Capital Management: New programming models optimized for streaming and/or multidimensional data; new backend engines that manage optimized file systems; engines able to fuse applications from multiple programming models (e.g. Map Reduce, work flows, and bag-of-tasks) on a single solution/abstraction. Ways to optimize resource usage and energy consumption when executing the analytics application.

VI. CONSIDERATIONS FOR DECISION MAKERS

Although many big data technologies are capable enough to be used for any critical circumstances, in any condition, but it still lags somewhere in some conditions use big data is not possible .while developing big data strategies there are certain protocols/dimensions which should always be considered they are as follows.

1. Online VS Offline Big Data.
2. Software Licensing Model.
3. Community.
4. Developer Appeal.

1.11Online vs. Offline Big Data

Big Data are of two forms. It can be also online or offline process. In the Online Big Data process the data is created, transformed, and analysed in current time to support all the current applications and their users. . Latency for these applications must be very low and availability must be high in order to meet SLAs and user expectations for modern application performance. Online Big Data can be used in social networking sites, for storing Database, etc.

In offline Big Data the data, they do not create new data. Big data holds within applications that transform analyse and manage all the big data in a batch context Response time is very slow. Examples of offline Big Data are hadoop workloads, modem Data warehouses, and also Business intelligence tools.

1.12Software License Model

There are 3 general types of licensing for Big Data technologies are as follows:

1. Proprietary
 2. Open source.
 3. Cloud source
1. Proprietary;-The words proprietary means ownership or a product which can be used only by the owner itself. This source cannot be used be licenses customer. Customers take license product through never ending license that allows them to use indefinite data. This is all about proprietary type of licensing.
 2. Open source:-The meaning of open source is that the data is available to all the users for free of cost. Companies express the software product by selling subscription with some added components like management tools and also value added services.
 3. Cloud Service:-In this service the data is made available in a cloud based environment and delivered over the public internet. All the people connected in a cloud can use Big data services.

1.13Community

In these early days of Big Data, there is an opportunity to learn from others. Organizations should consider how many other initiatives are being pursued using the same technologies and with similar objectives. To understand a given technology's adoption, organiza- tions should consider the following:

- The number of users
- The prevalence of local, community-organized events
- The health and activity of online forums such as Google Groups and Stack Overflow
- The availability of conferences, how frequently they occur and whether they are well-attended.
- The prevalence of local, community-organized events

1.14Developer Appeal

There are many talented engineers in the world who in a group join the multinational companies like Google Facebook etc. because these companies are called like heaven .its dream of all the engineers to work in the Google or many multinational companies, as they get high stipend which is depend upon their work& progress. Developers are ready to

give whatever they want but the employee has to prove himself.

VII. CONCLUSIONS

With the need of the decisions it is necessary to convert the data into interactive format for the better decision making. For this best use is visualization of the data in various forms to represent the data and discovering the hidden patterns. This Big Data trend is being seen by industries as a way of obtaining advantage over their competitors: 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. The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. Thank you for taking the time to read our book and we hope you enjoyed reading it as much as we did writing it.

REFERENCES

- [1] "IBM What is big data? – Bringing big data to the enterprise". www.ibm.com.
- [2] Big Data Visualization tool with Advancement of Challenges By Shilpa and Manjit(Volume 4, Issue 3, March 2014)
- [3] "Manufacturing: Big Data Benefits and Challenges". By J. Parallel Distrib. Comput.
- [4] Paper by CSA (Cloud Security Alliance) on bigdata.
- [5] http://www.webopedia.com/TERM/B/big_data.html
- [6] <https://www.mongodb.com/big-data-explained>.