e-ISSN : 2349-9745
p-ISSN : 2393-8161
Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
www.ijmter.com

Survey of Big Data Applications

Sachin Choudhary[1], Amay Devwrat[2]
[1]Department of Computer Engineering, MGMCET, Navi Mumbai
[2]Department of Computer Engineering, MGMCET, Navi Mumbai

Abstract— Today "Big data" is a evolving term which denotes very large amount of data. Data is accumulated in various format it may be structured, semi-structure and un-structured. Big data has the potential to help companies improve operations and make faster, more intelligent decisions. All of this data creates new opportunities to extract more value in human genomics , health science, surveillance and other areas. We are entering into the age of "Big Data". With the right solutions, organizations can dive into all data and gain valuable insights that were previously unimaginable. Data visualization can result in analysis of data in more accurate manner for the right decisions made by the organizations for its success.

Keywords- Big data, Big data analysis, Solutions of Big data, Big data applications

I. INTRODUCTION

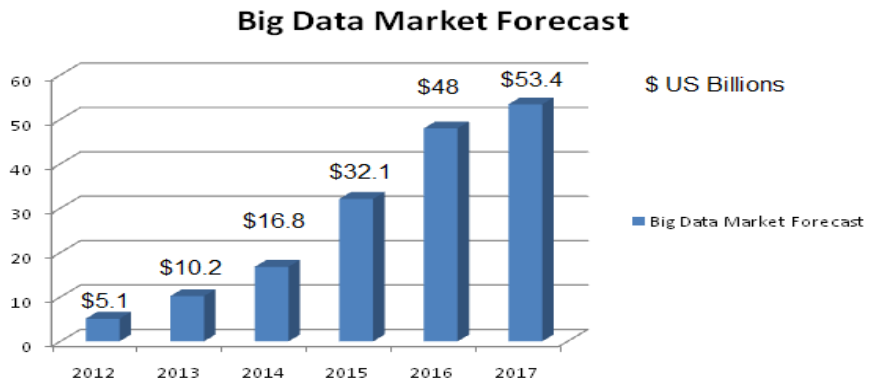
Big data is a term generally used and is penetrating still the notion gives rise to confusion. Big data gives the idea about the large quantities of data, social media analytics, effective data management, real-time data, and more. All the information the organization is going to explore and vast information in data analytics in real time system. While doing so, a small, but growing group of organization is achieving good outcome. In industries throughout the world, executives recognize the need to learn more about how to exploit big data. As data collected is day by day increasing. But despite what seems like not getting proper media attention, it becomes difficult to find in-depth information on what organizations are really doing.

Big data application gives the idea about how to see a large data in analytical form and based on that it gives appropriate result which will be helpful for organization.

Survey in Big data gives the idea about the data in large quantity different platform for the big data application and analysis of its infrastructure.

II. IMPORTANCE OF BIG DATA

III.

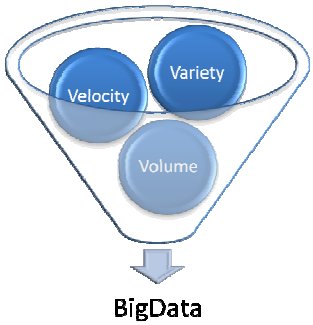


Big Data Market forecast gives the idea about the economy of big data.

Mainly the social networking and e-commerce sites are making use of the concept of Big Data. When big data is distilled and analyzed in combination with traditional enterprise data, enterprises can develop a more thorough and insightful understanding of their business, which can lead to enhanced productivity, a stronger competitive position and greater innovation – all of which can have a significant impact on the bottom line. For example, in the delivery of healthcare services, management of chronic or long-term conditions is expensive. Use of in-home monitoring devices to measure vital signs, and monitor progress is just one way that sensor data can be used to improve patient health and reduce both office visits and hospital admittance. Manufacturing companies deploy sensors in their products to return a stream of telemetry.

III. BIG DATA IN DIMENSIONS

Big Data has to deal with large and complex datasets that can be structured, semi-structured, or unstructured and will typically not fit into memory to be processed. When we talk to the people actually building Big Data systems and applications, we get a better idea of what they mean about 3Vs. They typically would mention the 3Vs model of Big Data, which are velocity, volume, and variety.



3.1. Volume:

Volume refers to the mass quantities of data that organizations are trying to improve decision-making across the enterprise. Data volumes continue to increase at an unprecedented rate. However, what constitutes truly "high" volume varies by industry and even geography, and is smaller than the peta bytes and zeta bytes often referenced.

3.2. Variety:

Variety is about managing the complexity of multiple data types, including structured, semi-structured and unstructured data. Organizations need to integrate and analyze data from a complex array of both traditional and non-traditional information sources, from within and outside the enterprise.

3.3. Velocity:

It means data in motion. The speed at which data is created, processed and analyzed continues to accelerate. Contributing to higher velocity is the real-time nature of data creation, as well as the need to incorporate streaming data into business processes and decision making. Velocity results the latency – it is the lag time between when data is created or captured, and when it can be accessible.

IV. NO SQL DATABASE

NOSQL Database is a distributed, highly scalable, key-value database based on Oracle Berkeley DB. It delivers a general purpose, enterprise class key value store adding an intelligent driver on top of distributed Berkeley DB. This intelligent driver keeps track of the underlying storage topology, shards the data and knows where data can be placed with the lowest latency. Unlike competitive solutions, NOSQL Database is easy to install, configure and manage, supports a broad set of workloads, and delivers enterprise-class reliability backed by enterprise-class support.

NOSQL databases are frequently used to acquire and store big data. They are well suited for dynamic data structures and are highly scalable. The data stored in a NOSQL database is typically of a high variety because the systems are intended to simply capture all data without categorizing and parsing the data into a fixed schema. This data has to handled carefully.

Difference between SQL and NOSQL

SQL	NO-SQL
1) Primarily called as Relational Databases (RDBMS). 2) Table based databases. 3) Databases have predefined schema. 4) Databases are vertically scalable 5) Databases uses SQL (structured query language) for defining and manipulating the data. 6) Examples: MySql, Oracle, Sqlite, Postgres and MS-SQL 7) SQL databases are good fit for the complex query intensive environment 8) SQL databases are not best fit for hierarchical data storage. 9) SQL databases are best fit for heavy duty transactional type applications, as it is more stable and promises the atomicity as well as integrity of the data. 10) SQL databases emphasizes on ACID properties (Atomicity, Consistency, Isolation and Durability)	1) Primarily called as Non-Relational or distributed database. 2) Document based, key-value pairs, graph databases or wide-column stores. 3) Databases have dynamic schema for unstructured data. 4) Databases are horizontally scalable. 5) Queries are focused on collection of documents. Sometimes it is also called as UnQL (Unstructured Query Language) 6) Examples: MongoDB, BigTable, Redis, RavenDb, Cassandra, Hbase, Neo4j and CouchDb 7) NOSQL databases are not good fit for complex queries 8) NOSQL database fits better for the hierarchical data storage 9) NOSQL for transactions purpose, it is still not comparable and sable enough in high load and for complex transactional applications. 10) NOSQL database follows the Brewers CAP theorem (Consistency, Availability and Partition tolerance)

V. BIG DATA PLATFORM

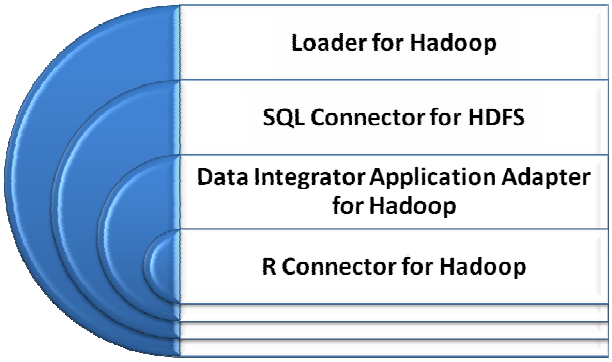
Big data platform is any platform which supports huge king of large datasets. As with data warehousing, web stores or any IT platform, an infrastructure for big data has unique requirements. In considering all the components of a big data platform, it is important to remember that the end goal is to easily integrate big data with enterprise data to allow to conduct deep analytics on the combined data set.

HADOOP:

Hadoop is an open source Java framework for processing and querying vast amounts of data on large clusters of commodity hardware. It is used for analysis of various dump data. Hadoop enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that both store and process the data, and can scale without limits.

Big Data Connectors

The Big Data Connectors consist of four components:



Where Big Data Appliance makes it easy for organizations to acquire and organize new types of data, Big Data Connectors tightly integrates the big data environment with Exadata and Database, so that we can analyze all of your data together with extreme performance.

VI. KEY CHALLENGES

Big data is set to offer companies tremendous insight. But with terabytes and petabytes of data entering in to organizations today, traditional architectures and infrastructures are not up to the challenge to face large data. IT teams are burdened with ever-growing requests for data, ad hoc analyses and one-off reports. It is seen that decision makers become frustrated because it takes hours or days to get answers to questions, if at all. As more users are expecting self-service access to information in a form they can easily understand and share with others.

To fully take advantages of big data services we need to address several challenges related to visualization and big data. Here we'come out some of those key challenges.

6.1. Meeting the need for speed:

In today's competitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly as possible. Visual-ization helps organizations perform analyses and make decisions much more rapidly, but the challenge is going through the volumes of data and accessing the level of detail needed, all at a high speed. The challenge only grows as the degree of granularity increases.

6.2. Understanding the data:

It takes a lot of understanding to get data in the right shape so that you can use visualization as part of data analysis. For example, if the data comes from social media content, you need to know who the user is in a general sense – such as a customer using a particular set of products – and understand what it is you're trying to visualize out of the data. Without some sort of context, visualization tools are likely to be of less value to the user.

6.3. Addressing data quality:

Even if you can find and analyze data quickly and put it in the proper context for the audience that will be consuming the information, the value of data for decision-making purposes will be at risk if the data is not accurate or timely. This is a challenge with any data analysis, but when considering the volumes of information involved in big data projects, it becomes even more pronounced. Data visualization will only prove to be a valuable tool if the data quality is assured.. It's always best to have a pro-active method to address data quality issues so problems won't arise later.

6.4.Displaying meaningful results:

On a graph plotting points for analysis becomes difficult when dealing with extremely large amounts of data or a variety of categories of information. For example, imagine we have 10 billion rows of retail data that you're trying to compare. The user trying to view 10 billion plots on the screen will have a hard time seeing so many data points. One way to resolve this is to collect and merge data into a higher-level view where smaller groups of data become visible. By grouping the data together, we can more effectively visualize the data.

6.5. Dealing with outliers:

The graphical representations of data made possible by visualization can communicate trends and outliers much faster than tables containing numbers and text. Users can easily spot issues that need attention simply by glancing at a chart.

VII. SOLUTIONS FOR BIG DATA

Current technologies -- developed long before big data come in to account -- limit our ability to analyze more than just a very limited portion of big data sets at any one time. Data tends to be stored in discrete data structures, or data silos, which are not easy to connect.

Although a single living will can be analyzed as a stand-alone entity, bringing together even just two living wills for a combined analysis -- essential to understand related impacts of bank failures -- would require a painstaking, time-intensive data-preparation process. More than 100 financial firms are required to have living wills -- linking them all together for analysis, to understand the myriad of interconnections and interdependencies, would simply be technically infeasible.

This challenge requires the ability to consolidate an organization's entire repository of information, so that it is all connected, and immediately available for analysis. Thanks to recent advances in data science, the task can be accomplished through a new approach that brings together multiple sources of information in what is known as the data lake. An industry-recognized term, the data lake is a massive repository of information that consolidates data -- including structured, unstructured, batched and streaming -- into a single table. The data lake eliminates the once-siloed, cumbersome data-preparation process, making information easily accessible to the analysts responsible for mining it.

Because the data lake can hold an almost unlimited amount of information, it would enable regulators to integrate all 100-plus living wills, and more if necessary. And with the advanced analytics that sit on top of the data lake, so to speak, the interdependencies of the banks would become transparent to regulators.

This transparency would also help regulators identify hidden risk that may be building across the financial system With the data lake, regulators would be able to see when an institutional practice might be creating collective risk -- even if individual banks cannot recognize it within their own organizations.

And because the data lake can be continuously updated, regulators would have the ability to see in real time, or near real time, exactly what is happening when a firm unwinds -- what impact its actions are having on other banks, and how that effect is rippling through the larger financial system. Regulators would be able to take quick action to prevent a firm's problems from spreading to others -- before it is too late -- and perhaps set the firm on a different path to keep it alive.

By building models and scenarios into the data lake, regulators would have the ability to quickly answer any number of "what if?" questions, whether they are stress testing living wills, or supervising their execution during an actual liquidation.

Moving to a new approach such as the data lake requires a new mindset about data analytics. Many organizations tend to focus on computer infrastructure and data storage -- how each might be

expanded, for example, as analytic needs change. But if we are to solve big data "problems" like the living wills, we must not focus on just on bigness, but on diversity, finding ways to harness, and extract value from all of the data that we are collecting.

VIII. CONCLUSION

Big Data is not a new concept but very challenging. It calls for scalable storage index and a distributed approach to retrieve required results near real-time. It is a fundamental fact that data is too big to process conventionally.. Nevertheless, big data will be complex and exist continuously during all big challenges, which are the big opportunities for us. The application of data lake in the data set will results in storing huge data in simplest form. In the future, significant challenges need to be tackled by industry and academia. It is an urgent need that computer scholars and social sciences scholars make close cooperation, in order to guarantee the long-term success of cloud computing and collectively explore new territory.

BIBLIOGRAPHY

Sachin Choudhary is currently studying in Mumbai University at MGMCET, Kamothe Navi Mumbai since 2011, currently pursuing B.E. in Computer Science and Engineering, with excellent Academics . He is having interest in Studying New Technologies, and Student Member of CSI.

Amay Devwrat is currently studying in Mumbai University at MGMCET, Kamothe Navi Mumbai since 2011,currently pursuing B.E. in Computer Science and Engineering, with excellent Academics. He is having interest in Learning New Technologies, and Student Member of CSI.

REFERENCES

1) Living will work- Big data solution by *Mark Herman, Booz Allen Hamilton*
2) Big Data Analytics with R and Hadoop-Vignesh Prajapati
3) *By Michael Schroeck, Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano* –Real use of Big Data
4) *SAS (The Power toKnow)*-Challenges of Big Data

