Big Data -the Next Frontier for Innovation, Competition and Productivity

Ms.S.Rekha[1], Ms.N.Prameela[2], Dr.M.Inbavalli[3]

[1](III MCA, perumal manimekalai college of engineering, hosur, ANNA University, Tamilnadu)
[2](II MCA, perumal manimekalai college of engineering, hosur, ANNA University, Tamilnadu)
[3](Associate Professor, MCA, perumal manimekalai college of engineering, hosur, ANNA University, Tamilnadu)

Abstract: *Collection is realization a tipping point for main technical changes that can fetch new ways in decision making, managing our health, cities, finance and education. whilst the data complexities are increasing counting data's volume, variety, velocity and veracity, the real impact pivots on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics pose a impressive challenge on Big data is a new driver of the world economic and public changes. The world's data the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthrough include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data competently and effectively. Big Data Analytics is relevant to Hong Kong as it move towards a digital financial system and society. Hong Kong is already among the best in the world in Big Data Analytics, taking up such management position as chairs and editor in chiefs of imperative conference and journal in Big Data linked areas.*

Keyword: *big data, volume, velocity, veracity, analytics, technologies, application, educations.*

I.   Introduction

Big data is a new driver of the world financial and public changes. The world's data collection is attainment a tipping point for major industrial change that can bring new ways in management organization our health, cities, finance and education. whilst the data intricacy are rising counting data's volume, variety, velocity and veracity, the real impact hinges lying on our ability to expose the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a impressive challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large secreted values from datasets that are diverse, complex, and of a huge scale. Potential break throughs include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data competently and effectively. Big Data Analytics is relevant to Hong Kong as it moves towards a digital country and society. Hong Kong is already among the best in the world in Big Data Analytics attractive up such leadership positions as chairs and editor in chiefs of important conferences and journal in Big Data related areas., Hong Kong universities, government and industry must act quickly in addressing a number of major challenges.These challenges includes "rationality which concern new algorithms, theory and methodologies in knowledge discovery from large amount of data and "system and applications," which concerns inventive application and systems useful for beneath Big Data practice. Big data analytics must also be team crack harsh across scholarly the social arrange management and the social order and industry, and by researchers from many discipline including computer science and engineering, health, data science and social and policy areas.

The history of big data

Although the perception of big data itself is comparatively new, the genesis of large data sets go back to the 1960sand'70s when the world of data was just getting started, with the first data center and the expansion of the relationaldatabase. Around 2005, people began to understand just how much data users generate through Facebook, YouTube, and other online services. Hadoop (an open-source support created specially to store and analyzbig data sets) was developed that same year. NoSQL also began to gain reputation during this time. The development of open-source frameworks, such as Hadoop (and more recently, Spark) was vital for the escalation of big data because they make big data easier in the direction of work with and cheaper to store. In the years since then, the volume of big data has skyrocket. user are still generating huge amounts of data but it's not just humans who are doing it.With the advent of the Internet of Things (IoT) , extra objects and devices are connected to the internet, congregation data on purchaser usage patterns and product manceperfor. The look of machine learning has produced still more data.While big data has come far, its usefulness be only just beginning.
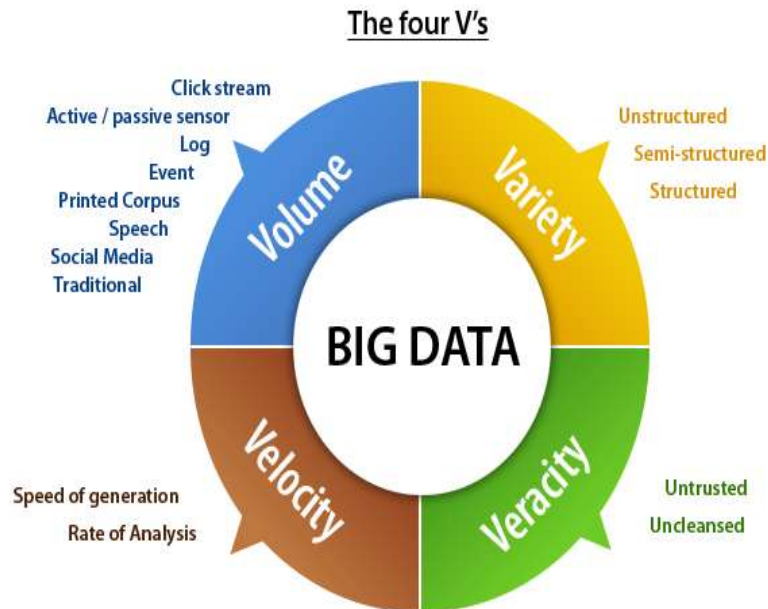
Fig.1.1 history of big data

What is big data

   The BIG DATA is data whose scale, diversity, and convolution require new architectures, algorithm, techniques, and analytics to supervise it exact value unseen knowledge from it. an plan to solve new problems or old problems in a better way actually know big data, it's helpful to have some past background. Here's Gartner's characterization circa 2001 (which is still the go-to definition): Big data is data that contains greater variety inner in rising volume and with ever-higher velocity.

Fig.2.2 big data

Three Vs of Big Data
Volume

   The amount of data matter. With big data, to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, click streams on a web page or a mobile app, or sensor-enabled equipment.

Velocity

   Velocity is the fast rate at which data is received and (perhaps) acted on.the maximum velocity of data streams directly into memory versus being written en route for disk. Some internet-enabled smart products .real-time evaluation and action.

Variety

   Variety refers to the a lot of types data that are available. habitual data types were structured and fit neatly in a relational database. Unstructured and semi structured data types, such as text, audio, and video require extra preprocessing to derive meaning and support metadata.

The four V's

Fig 3.3 three vs of big data

Benefits of big data
Why Big Data is so important:

The collected statistics and not in how much data they have been able to collect. an organization is look to get benefitted by Big Data then being able to efficiently use Big Data is what that matters the most.

Prediction and Decision making:

Big Data to reduce the risks about the decisions of the organization and manufacture prediction as become one of the many benefits coming from big data in industry

Cost-effectiveness:

Big Data the way of make relevant predictions and smart decisions also make the organizations cost-effective. Big Data to manage and study the data also bring cost advantages to businesses particularly a huge amount of data is to be stored and processed.

Marketing effectiveness:

Big Data with organism to help businesses and organizations in making smart decisions also radically increases the sales and marketing effectiveness of the businesses and organizations thus highly improving their performances in the industry.

Big Data Analytics

The "Big Data" have recently been realistic to datasets that raise so large that they become uncomfortable to work with using traditional database regulation systems. The are data sets whose size is gone from the ability of commonly used software tools and storage systems to capture, store, manage, as process the data within a tolera-ble forgotten time of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing.

Application of the big data and analytics

Fig 5.1 application of big data and analytics

Insurance:

Manufacturing is data-dependent. Today shipper and mediators are in a meeting in humanizing data capture to hel p them to better manage their business, manage their risk and know their customers.

Big Data Contributions to Education:

Big data has influence in the education world. It able to provide solution to one of the biggest pitfall in the education system, the one-size-fits-all manner of academic set up, by contributing in e-learning solutions..

Big data contributions to Healthcare:

The big data is in comprehensive use in the field of medication and healthcare. Big data **reduces costs of treatment** since there is less chances of having to pe rform unnecessary diagnosis.

Big data contributions in Public sector:

It provides a large range of benefits to public sectors most of which are the services provided by Big Data to the government sectors that includes power intake deceit recognition.

Big Data Media and Entertainment:

The platforms that utilizes the benefits of big data, big time. For example:Spotify, user.Amazon Prime, that offers, videos, music and Kindle books in a one-stop shop is also big on using big data.

Big Data Contributions to Transportation:

Big data can be used to understand and estimate the user's needs on different route and on multiple modes of transport

Data Contributions to Banking Zones and Fraud Detection:

Big data is massively used in the fraud detection in the banking sectors. The misuse of credit cards

Misuse of debit card
Venture credit hazard treatment
Business clarity
Customer statistics alteration
Money laundering
Risk Mitigation

Banking and security:

The commonly associated with equal impact on the banking and securities industry. The global recession in sweeping changes for the entire banking industry, with banks having to comply with more stringent capital requirements.

Big Data Analytics Technologies

big data analytics tools and technologies. Design data and system architecture, data integration, exploration and visualization, descriptive analytics, data science, predictive analytics, machine learning, and AI. we have partnerships with Microsoft, Cloudera, AWS, Qubole, Snowflake, Profisee, Tableau, and MapR and work with the following technologies: Hadoop

Apache Hive and Pig
Apache Spark and Mahout
Apache HBase and Cassandra
Amazon Web Services (Data Pipeline, EMR, Redshift)
Apache Impala and SQL on Hadoop

big data technologies and analyse data:

Apache Hadoop is a java based free software framework that can effect store large amount of data in a cluster .Microsoft HDInsight.It is a Big Data solution from Microsoft power-driven by Apache Hadoop which is available as a service in the cloud.

Big data analytics tools and technologies
VIEW RESOURCE

Fig 6.1 big data tools and technologies

Hive:
Apache Hive is a data storehouse system for data summarization, analysis and query of large data systems in open source Hadoop platform.

pig :
Apache Pig is a platform that is used to analyze large datasets. writing data to and reading data from the Hadoop Distributed File System (HDFS) and performing processing by means of one or more MapReduce jobs.

Hbase:
HBase is a column-oriented database administration system of Hadoop Distributed File System (HDFS).

Zoo keeper:
Zookeeper is Hadoop can be viewed as centralized repository where distributed applications can put data and get data out of it.

Avro:
Avro facilitate the substitute of big data between programs written in any language. With the entertainment in installments service, programs can efficiently serialize data into files or into messages.

Spoof:
Spoofing is often the way a bad actor gains access in order to execute a larger cyber attack such as an advanced determined threat or a man-in-the-middle attack.

Flume:
Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the HadoopDistributed File System (HDFS).

Oozie:
Oozie is a workflow scheduler scheme to manage Apache Hadoop jobs. OozieWorkflow jobs are intended for Acyclical graph (DAGs) of actions.

II. Conclusion

Big data provides enterprise with more because of its lots of related technologies and tools, will continue to be developed and become innovative in the future, Hadoop distribution, the next generation of data storehouse advanced data visualization, academia pays more attention to cloud computing. Big data focuses on "data", data service, data acquisition, analysis and data mining, pays more attention on ability of data storage. Cloud computing focuses on computing architecture along with practices. Big data and cloud computing are two sides of the same issue. It is more accurate to analyze and estimate big data by using cloud computing and release more hidden value of data; in order to meet the service demand of big data, find even better practical application to the cloud computing.

References

[1]. Kaisler, S., Armour, F., Espinosa, J.A., Money, W.: Big **data: Issues and challenges moving forward. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on. pp. 995-1004. IEEE (2013).**

[2]. Katal, A., Wazid, M., Goudar, R**.: Big data: Issues, challenges, tools and good practices. In: Contemporary Computing (IC3), 2013 Sixth International Conference on. pp. 404-409. IEEE (2013).**

[3]. Jabin, S., & Zareen, F. J. **Biometric Signature Verification. International Journal of Biometrics, 7(2), 97-118 (2015).**

[4]. Fan, J., Han, F., Liu, H**.: Challenges of big data analysis. National science review 1(2), 293-314 (2014).**

[5]. Beyer, M.A., Laney, D.: **The importance of big data: A definition. gartner (2012).**