International Journal of Research in Advent Technology, Vol.7, No.5S, May 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Big Data Pattern With Special Reference To Business Management

A.Tharma Jeyaseeli
Department of Computer Application
Ayya Nadar Janaki Ammal College,Sivakasi,India
jeyaanbuchelian93@gmail.com

Abstract: Innovations in technology and greater affordability of digital devices have presided over today's Age of Big Data, an umbrella term for the explosion in the quantity and diversity of high frequency digital data. These data hold the potential—as yet largely untapped—to allow decision makers to track development progress, improve social protection, and understand where existing policies and programmes require adjustment. Turning Big Data—call logs, mobile-banking transactions, online user-generated content such as blog posts and Tweets, online searches, satellite images,etc.—into actionable information requires using computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets. New insights gleaned from such data mining should complement official statistics, survey data, and information generated by Early Warning Systems, adding depth and nuances on human behaviours and experiences—and doing so in real time, thereby narrowing both information and time gaps. With the promise come questions about the analytical value and thus policy relevance of this data—including concerns over the relevance of the data in developing country contexts, its representativeness, its reliability—as well as the overarching privacy issues of utilizing personal data. This paper does not offer a grand theory of technology-driven social change in the Big Data era. It is important to recognize that Big Data and real-time analytics are no modern panacea for age-old development challenges.

1. INTRODUCTION

The Big Data Solution provides an approach for storing, managing and accessing data of very high volumes, variety or complexity. Storing large volumes of data from a large variety of data sources in traditional relational data stores is cost-prohibitive. And regular data modeling approaches and statistical tools cannot handle data structures with such high complexity. This solution offering discusses new types of data management systems based on NoSQL database management systems and MapReduce as the typical programming model and access method.

2. BIGDATA ETHICS

Data Ethics is concerned with the following principles: 1. **Ownership** - Individuals own their own data. 2. **Transaction Transparency** - If an individuals personal data is used, they should have transparent access to the algorithm design used to generate aggregate data sets 3. **Consent** - If an individual or legal entity would like to use personal data, one needs informed and explicitly expressed consent of what personal data moves to whom, when, and for what purpose from the owner of the data. 4. **Privacy** - If data transactions occur all reasonable effort needs to be made to preserve privacy. 5. **Currency** - Individuals should be aware of financial transactions resulting from the use of their personal data and the scale of these transactions. 6. **Openness** - Aggregate data sets should be freely available.

Big Data Governance

One of the major challenges for data integration in the context of big data is to estab-lish and sustain the right level of governance. And it's not all about technology. Key issues like data quality, data privacy and security, relevance and meaningfulness must be considered at the enterprise level.Let's look at this a little deeper. Linking to new data sources, especially for external sources and unstructured data, will put data out of reach for typical data governance programs. In other words, standards and data quality will no longer be controlled at the source. All the same, trying to enforce traditional levels of quality for big data might annihilate the anticipated benefits of big data initiatives related to rapid data integration and handling data streams in real time. There is clearly a balance to be found between the data quality imperative and the benefits of big data velocity.

Big Data Trend

Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.

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 has one or more of the following characteristics: high volume, high velocity or high variety. Artificial intelligence (AI), mobile, social and the Internet of Things (IoT) are driving data complexity through new forms and sources of data. For example, big data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media — much of it generated in real time and at a very large scale.

Analysis of big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable. Businesses can use advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics and natural language processing to gain new insights from previously untapped data sources independently or together with existing enterprise data.

Big data that is very small

Modern machines such as cars, trains, power stations and planes all have increasing numbers of sensors constantly collecting masses of data. It is common to talk of having thousands or even hundreds of thousands of sensors all collecting information about the performance and activities of a machine.

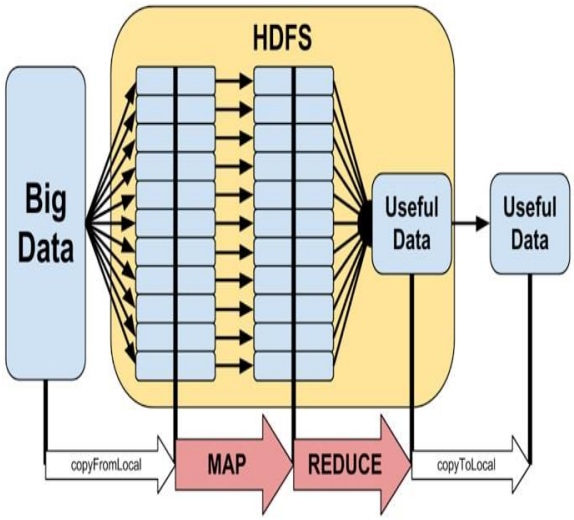
A plane on a regular one hour flight with a hundred thousand sensors covering everything from the speed of air over every part of the airframe through to the amount of carbon dioxide in each section of the cabin. Each sensor is effectively an independent device with its own physical characteristics. The real interest is usually in combinations of sensor readings (such as carbon dioxide combined with cabin temperature and the speed of air combined with air pressure). With so many sensors the combinations are incredibly complex and vary with the error tolerance and characteristics of individual devices.

The data streaming from a hundred thousand sensors on an aircraft is big data. However the size of the dataset is not as large as might be expected. Even a hundred thousand sensors, each producing an eight byte reading every second would produce less than 3GB of data in an hour of flying (100,000 sensors x 60 minutes x 60 seconds x 8 bytes).

3. HADOOP IS A MASTER

HDFS runs on a large cluster of commodity nodes. Whenever a node is placed in the IP range as specified by a name node, one of the necessary Java virtual machines, it becomes available for data storage in the file system and will report a heartbeat henceforth to the name node. All data is unsequenced, stored in 64 MB block sizes (although records can span blocks) and is replicated three times to ensure redundancy (instead of using RAID).

The first block is written to the node creating the file. The second copy is written to a node within the same rack to minimize cross-rack network traffic. The third, and final, node is written to a different rack to tolerate switch failure. All MapReduce accesses to the data store are full scans and choose one of the three placements to read for each data block. Hadoop has full control over record placement.



This approach is highly scalable to thousands of nodes. The large block sizes maximize the I/O

operation. It is also very fault tolerant, which is necessary when dealing with commodity class nodes. Hadoop offers the ability to quickly analyze massive collections of records without forcing data to first be modeled, then transformed, then loaded.

The ability of Hadoop to handle schema on-read (as opposed to schema on-write) provides the needed agility to quickly bring new data sources on board without having to shoehorn inadequate formats into a predefined data model. Hadoop can serve as:• The next-generation data warehouse, to augment or supplement the traditional RDBMS.• A new data store for new data types – particularly the unstructured data that an RDBMS cannot handle – and for new data sources, such as web, social network and sensor data.• A data lake, to stage all the organization's available data in a minimally processed state.Data access is traditionally conditioned to predefined data models, predefined data integration flows and predefined reporting models. Any change requires IT involve-ment, which often means long turnarounds on design, implementation and testing. But to keep pace with competitors, businesses need access to data in real time. Only then will they have the flexibility to extract valuable insight from the data, when it's needed. Techniques like self-service data preparation make this possible. Organizations need to be able to use data as soon as it is produced (or available), so employees can make decisions in real time and take action as soon as an event occurs. To do it, they must be able to analyze data streams on the fly, before the data even hits a data store. Event stream processing addresses this need by streaming millions of records per second and providing the most up-to-date information possible.

Big data types of Data Stores

NoSQL database systems are schema-less and often store data as a key followed by a value (i.e., FIRSTNAME: William). As for many solutions this form of storage is their only one, they are often categorized as Key-Value stores.

In **Key-Value stores**, the schema can differ from row to row. Key-Value stores are great for stock quoting, parts lists and other forms of high-volume data storing.

Column stores are, again, key-value, but supercolumns or column families are declared in the schema. For example, a super column could be Name, which is broken down into First Name and Last Name columns. This slight deviation to the schema-less concept makes column stores very useful for Big Data with its mix of known and unknown in every row. Column stores are great for time series data.

Graph Databases primarily store the aforementioned relationship information. They therefore handle a workload very different from other NoSQL data stores. Edge is the term for a relationship and the edges can connect any nodes (which are like rows in a table) that have a relationship. For example, customers have social relationships with other customers, they send invitations and information to certain other customers, they place orders, they have addresses, etc. All of these (customers, orders, addresses) are nodes and edges refer to the social relationships, the sending of certain communications like invites, the placing of orders and having addresses. Graph database navigation excels over SQL for relationship navigation. Product catalogs are another common use of Graph Databases.

More commonly, Graph Databases can depict the social graph of relationships between entities (customers, employees, suppliers, etc.). Graph databases are also useful for recommendation engines and network analysis. Graph databases are further described in Graph databases.

Finally, there are **Document Stores**. Document is the term these NoSQL stores use to refer to a JSON row. Document Stores have the richest set of possibilities for the values stored. These include scalars, arrays, attachments and, importantly, links to other documents – so documents can be embedded. Document Stores provide developers with the most flexibility of all.

Document Stores are great for complex modeling and accessing actual documents, such as the workload for electronic health records.

4. ENTERPRISE INFORMATION MANAGEMENT AWARENESS

An effective big data addition to the information management environment means recognition of the differences between big data/NoSQL approaches and traditional SQL-oriented approaches. The vast majority of the users – either applications or end users – will be used to the SQL approach. Once NoSQL is decided upon, it is important for the team to continue to socialize the differences, so the implications are well understood as the technology blueprint and other phases are developed. Enterprise Information Management Awareness for big data creates a mutual understanding in the enterprise between the development team and the using community of the unique aspects that big data solutions must contend with.

While the lack of complete ACID compliance may cause a slight, immeasurable risk, the status of the ACID compliant nature of the big data solution should be well understood and fit with the risk profile of the application being supported or the application(s) being developed for big data.

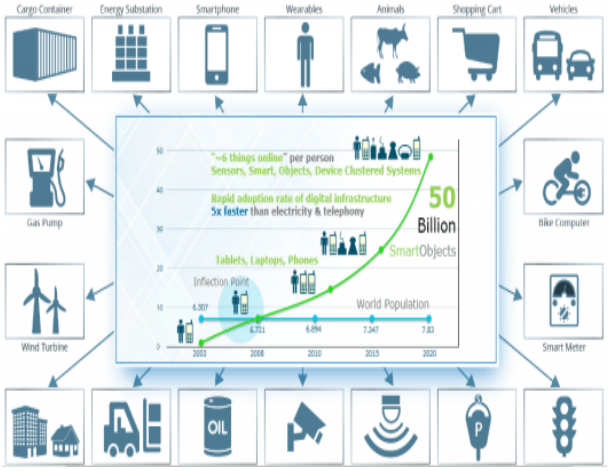
Organizational QuickScan for Information Development

Big data solutions are for either an operational or analytical purpose. Furthermore, there could be many NoSQL stores in the environment. Therefore, they can fit into the overall architecture in several places. An Organisational QuickScan for Information Development around information usage and information stores across the enterprise, combined with how the new solutions fill gaps and line up other data stores for eventual replacement, should be done.

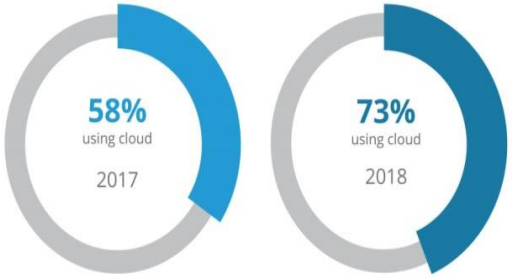
Big Data is Only Going to Get Bigger

With the growth of devices and transactions that generate increasingly complex data streams, effectively using that data is rapidly becoming a significant competitive advantage for many companies. In fact, some companies consider data to be one of their most valuable assets. Therefore, big data should only get bigger as organizations look for more and better ways to tap into existing data and gather new and emerging types of data to make critical decisions, answering questions that were previously considered beyond reach.

Big Data Driving Factors



The quantity of data on planet earth is growing exponentially for many reasons. Various sources and our day to day activities generates lots of data. With the invent of the web, the whole world has gone online, every single thing we do leaves a digital trace. With the smart objects going online, the data growth rate has increased rapidly. The major sources of Big Data are social media sites, sensor networks, digital

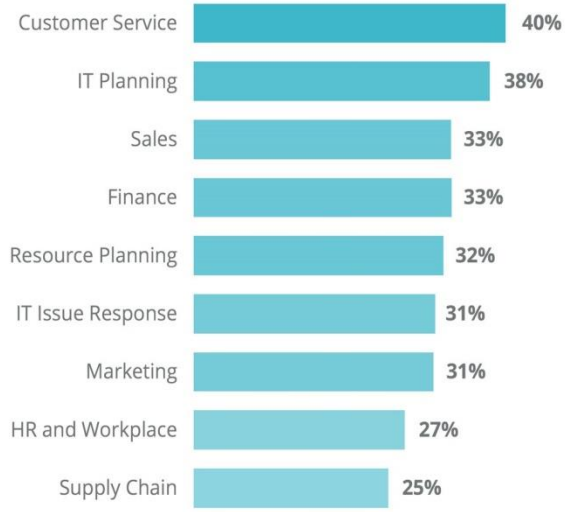


images/videos, cell phones, purchase transaction records, web logs, medical records, archives, military surveillance, eCommerce, complex scientific research and so on. All these information amounts to around some Quintillion bytes of data. By 2020, the data volumes will be around 40 Zettabytes which is equivalent to adding every single grain of sand on the planet multiplied by seventy-five.

Enterprise Value from Big Data

Big data is being used across a wide and growing spectrum of departments and functions in modern organizations. The insights and intelligence provided by big data translate directly into operational efficiencies and competitive advantage.

Business Processes Receiving Value from Big Data



Big Data Processing is Moving to the Cloud

The overwhelming majority of organizations are moving their big data processing to the cloud to take advantage of its convenience, cost, integration, and performance benefits.

Applications area:

Banking

With large amounts of information streaming in from countless sources, banks are faced with finding new and innovative ways to manage big data. While it's important to understand customers and boost their satisfaction, it's equally important to minimize risk and fraud while maintaining regulatory compliance. Big data brings big insights, but it also requires financial institutions to stay one step ahead of the game with advanced analytics.

Education

Educators armed with data-driven insight can make a significant impact on school systems, students and curriculums. By analyzing big data, they can identify at-risk students, make sure students are making adequate progress, and can implement a better system for evaluation and support of teachers and principals.

Government

When government agencies are able to harness and apply analytics to their big data, they gain significant ground when it comes to managing utilities, running agencies, dealing with traffic congestion or preventing crime. But while there are many advantages to big data, governments must also address issues of transparency and privacy.

Health Care

Patient records. Treatment plans. Prescription information. When it comes to health care, everything needs to be done quickly, accurately – and, in some cases, with enough transparency to satisfy stringent industry regulations. When big data is managed effectively, health care providers can uncover hidden insights that improve patient care.

Manufacturing

Armed with insight that big data can provide, manufacturers can boost quality and output while minimizing waste – processes that are key in today's highly competitive market. More and more manufacturers are working in an analytics-based culture, which means they can solve problems faster and make more agile business decisions.

Retail

Customer relationship building is critical to the retail industry – and the best way to manage that is to manage big data. Retailers need to know the best way to market to customers, the most effective way to handle transactions, and the most strategic way to bring back lapsed business. Big data remains at the heart of all those things.

Future work:

Machine learning concepts and tools are gaining popularity among researchers to facilitate meaningful results from these concepts. Research in the area of machine learning for big data has focused on data processing, algo-rithm implementation, and optimization. Many of the machine learning tools for big data are started recently needs drastic change to adopt it. We argue that while each of the tools has their advantages and limitations, more efficient tools can be developed for dealing with problems inherent to big data. The efficient tools to be developed must have provision to handle noisy and imbalance data, uncertainty and inconsistency, and missing values.

5. SUMMARY

Big data is here and it is here to stay. Despite the hype, big data does offer tangible business benefit to organizations. It enables enhanced insight, decision making, and process automation. The characteristics of big data is the three V: Volume, Velocity and Variety. The big in big data is not just about volume. While big data certainly involves having a lot of data, big data does not refer to data volume alone. What it means is that you are not only getting a lot of data. It is also coming at you fast, it is coming at you in complex format, and it is coming at you from a variety of sources. Data comes from variety of sources, and can be used in various industry applications. Often it is the combination of data sources that counts. Along with big data, there is also a so-called paradigm shift in terms of analytic focus. That is a shift from descriptive analytics to predictive and prescriptive analytics.Big data necessitates a new type of data management solution because of its high-volume, high-velocity and/or high-variety nature. This new type of data management solution bears the trademark of highly scalable, massively parallel, and cost-effective. New technologies, such as Hadoop, are not replacing other technologies, such as relational database, but rather are being added alongside them.

6. CONCLUSION

Business and IT leaders need to ask themselves whether their industrial enterprise is maximizing the full potential value of their process data and using that insight to drive real-time improvements. As data volumes continue to expand, information-driven strategies will only become more pervasive as a source of competitiveness—making the use of big data in the industrial space ever more imperative. A

closer look at advanced historians demonstrates how such technologies can help enterprises leverage their time-series process data by providing the ability to efficiently run real-time analytics within massive sets of historical data. These solutions have the potential to revolutionize the way enterprises do business by providing critical insights for timelier operational decisions while also enabling continuous improvements across the enterprise. Going forward, as information increasingly empowers enterprises to understand their businesses better and to foresee what is possible, those that capitalize on the value of big data will gain insights to improve performance beyond their competitors. They will be positioned to better innovate, compete, and drive value—all of which will significantly accelerate business growth and continuously drive optimized performance for long-term success.

REFERENCE

[1] Survey on Big Data Using Data Mining *(PDF)*. International Journal of Engineering Development and Research. 2015. *Retrieved 14 September 2016*.

[2] IBM Research - Zurich *(PDF)*. Zurich.ibm.com. *Retrieved 8 October 2017*.

[3] Grimes, Seth. Big Data: Avoid 'Wanna V' Confusion. InformationWeek. *Retrieved 5 January 2016*.

[4] Digital Technology & Social Change. Canvas.instructure.com. *Retrieved 8 October 2017*.

[5] Siwach, Gautam; Esmailpour, Amir (March 2014). Encrypted Search & Cluster Formation in Big Data *(PDF)*. ASEE 2014 Zone I Conference. University of Bridgeport, Bridgeport, Connecticut, US.

[6] Resources on how Topological Data Analysis is used to analyze big data. Ayasdi.

[7] Daniele Medri: Big Data & Business: An on-going revolution. Statistics Views. 21 October 2013.

[8] Degrees in Big Data: Fad or Fast Track to Career Success. Forbes. *Retrieved 21 February 2016*.

[9] Pelt, Mason. Big Data is an over used buzzword and this Twitter bot proves it. siliconangle.com. SiliconANGLE. *Retrieved 4 November 2015*.