International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 12 • Issue 1 pp. 13-17 Sept 2020 - March 2021 www.csjournals.com

A Study on Big Data Analytics

Dr. D.V. Honagannavar
Principal, KLE Society's J.G.College of Commerce, Hubballi.

Abstract: Often Data analysis and Data analytics are often known as interrelated or interchangeable terms holding minute differences in their meanings. At the outset, the basic difference between analytics and analysis is the scale, data analytics is a broader term of which data analysis is its subcomponent. Data analysis is the process of examination, processing and arrangement of a given data set in specific ways in order to study its individual components or entities and extract the useful information, wheras, Data analytics is an overhauling/overarching science or a discipline that covers/encompasses the complete management of data. This not only includes analysis, but also data collection, data organization, data storage, and all the tools and techniques required/used.

Introduction: Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by analytics systems. It is often a complex process of examining huge amount of data to unfold the information, viz, hidden patterns, correlations, market trends and customer preferences which can help the organizations to make informed business decisions. On a broader scale, data analytics technology and techniques provide a means to analyze data sets and provide new information which can help the organizations to take up informed business decisions.

The role of a data analyst is to collect, analyze, and transform the data into accessible information. The analysts help the organizations make better business decisions by identifying the trends and patterns of their functionality on par with their objectives. The ability of the analyst to describe, predict, and improve the organizational performance has given them a increasingly high demand and

recognition globally and across industries. Data analysis performs the evaluation of data through analytical and logical reasoning which leads to an outcome or conclusion within a stipulated frame/context. It is a multiphased process that involves a number of steps, approaches, and diversified techniques. The approach towards data analysis largely depends on the type of data available for analysis and the purpose of the analysis.

The importance of big data analytics:

Big data analytics through specialized systems and software can provide a positive business-related outcomes as follows:

- Opportunities for generating new revenue
- Effective marketing to the core
- Better customer service
- Efficiency in operations
- Competitive edge over rivals

Its applications allow the data analysts, data scientists, predictive modelers, statisticians and other analytics professionals to analyze growing volumes of structured transactional data including other forms of data that are often left untouched by conventional Business Information systems and analytics programs. It's a mixture of semi-structured and unstructured data where the internet of things (IoT) such as internet clickstream data, web server logs, social media content, text from customer emails and survey responses, mobile phone records, and machine data captured by sensors are counted upon.

Functionality of big data analytics:

The data clusters and NoSQL systems are basically used as launch pads and staging areas for the data. Before getting loaded into a data warehouse or analytical database for analysis it is usually transformed into in a summarized form i.e into a more conductive and relational structures. However, big data analytics users adopt the concept of "Data Lake" that serves as the primary repository for incoming streams of raw data. In such architectures, data can be analyzed directly in a data cluster or it can be run through a processing engine such as 'Spark'. As in data warehousing, sound data management is a first step in the process of big data analytics. Data being stored in the databases must be organized, configured and partitioned properly in order to get good performance out of both extract, in scenarios such as transformation and loading and analytical queries. Text mining and statistical analysis software will also play a big role in the process of big data analytics by mainstreaming business intelligence softwares and data visualization tools.

Once the data is ready, it can be analyzed with the software commonly used for advanced analytics processes such as :

- Data mining: which shift through data sets in search of patterns and relationships.
- Predictive analytics: which build models to forecast customer behavior and other future developments.
- Machine learning: which taps into algorithms to analyze large data sets.
- Deep learning: a more advanced machine learning.

Challenges and uses of Big data analytics :

Its applications often include data from both internal systems and external sources, such as weather data, or demographic data on consumers compiled by

the third-party information services providers. In addition to this streaming analytics applications are common in big data environments as users look to perform real-time analytics on data fed into the systems through stream processing engines. Early big data systems were mostly deployed on premises in large organizations that collected, organized and analyzed massive amounts of data. But cloud platform have made it easier to set up and manage Data clusters in the cloud. Thus supporting the distribution of the big data framework on the clouds. The users can now spin up the data clusters in the cloud, run them as long as they need and then take them offline with usage-based pricing which doesn't require any software licensing

Big data has become increasingly beneficial in the field of supply chain analytics. It uses big data and quantitative methods to enhance the decision making processes across the supply chain. Specifically, these big supply chain analytics expands datasets for increased analysis that go beyond the traditional internal data found on enterprise resource planning (ERP) and supply chain management (SCM) systems. They also implement highly effective statistical methods on new and existing data sources. The insights thus gathered facilitate better informed and more effective decisions that benefit and improve the supply chain. Potential drawbacks of big data analytics include lack of internal analytics skills and the high cost of hiring experienced data scientists and data engineers.

Conclusion: The term big data is often used to refer to increasing data volumes encompassing the increase in the variety of data being generated by organizations and the velocity at which that data is being created and updated. The three vital factors i.e. volume, velocity and variety known as the 3Vs of big data is a concept thus popularized. With a distributed processing framework and an open source projects it has clustered a platform built on commodity hardware and geared big data application softwares. It has began to take a firm hold in the or-

ganizations and the public sector, along with data clusters and various related big data technologies that have sprung up in and around it.

Their applications primarily province a large of internet and e-commerce structures as well as analytics and services providers. In the upcoming years, big data analytics can be increasingly embraced by retailers, financial services firms, insurers, healthcare organizations, manufacturers, energy companies, organizational administrations and other as such.

References:

https://www.getsmarter.com

https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

https://www.zapmeta.co.in