Big Data Analytics

Abstract

There is a proverb which is accepted in today's reality is "Everything is data and Data is everything". Information is a key for any business to be successful. Companies like Twitter, Facebook , Microsoft which are known to be the Big bulls of IT generates more than 2.5 quintillion data bytes of data everyday which mostly contains unstructured data such as emails, Images, Videos etc. This is Big data. Although there are many ways to store and process some huge data with clusters grouped together acting as a distributed fashion but Big Data isn't that simple. The computational power and Infrastructure required will be more for it. In this report my Aim is to talk about How organizations deal with Big data analytics and the challenges that many companies are facing currently regarding the use of Big Data Analytics.

Introduction

In the modern era, Data or Information is the primary source of everything. Since past couple of years there were many challenges with regards to storing and processing it. In the olden days people use to store the data on handwritten papers. As we started progressing, In the year 1956,IBM launched it's first computer with magnetic disc Storage(HDD) and It was able to store only five Megabytes of data. As we started evolving further with all the advancements in the latest technology today we have many storage devices like pen drive, hard disk, memory card, databases, clusters etc. The storage capacity also increased very well that today we can store terabytes of data in minimal devices which can even fit our pocket but when the data started flowing in the Increasing volumes the measures to store and process it has been devised.

The major challenges w.r.t data includes the below three factors:

- Volume.
- Variety
- Velocity.

Let’s discuss each of them in detail and various challenges of Data.

Volume:

Companies like Facebook produces 4 petabytes(approx equal to million gigabytes),Twitter produces 500 million tweets of data everyday and our well known Google which is a famous search engine processes more than 20 petabytes of data everyday. By looking at these figures we can understand that day by day there is huge amount of data that is being generated and the storage has also been an underlying issue. Following are the main problems or challenges related to data storage:

- **Infrastructure**: We just can’t store data. We need Infrastructure which needs to be setted up to store it which often means Investing in high tech servers that will occupy significant space.
- **Cost**: Having our own data center is also an expensive operation and we need people to set it up and maintain it on the long run.
- **Data Security**: Securing the data is the important challenge for any firm or an organization. There are many layers of security that can help us prevent unauthorized access, but there’s a limit on that as well on how well they can protect us.
- **Data Corruption**: Every form of data has a chance to be corrupted. Data will be degraded over time and the best way to protect it would be having multiple backups.
- **Scalability**: We should be able to scale the data. storage. Our storage plan might work efficiently for the data limit we have but what if those needs change suddenly?

Variety:

The data we get in the real world can be classified into three main categories :

1. **Structured Data :** The name itself depicts that the data will have a proper structure. Structured data will be usually refereed to data that is stored in rows and columns i.e. our traditional relational databases and all the rows in the table has the same set of columns and the users can quickly input, search and manipulate it. There are many structured data tools which are widely used mainly OLAP,SQL,ORACLE etc and used in many domains like CRM,Online booking systems and Accounting firms. The problems with structured data includes Limited usage i.e. As it's structure is predefined we can use it only for specific purposes with limited flexibility.

2. **Unstructured Data:** The data which doesn't have a proper structure is called as unstructured data. It can be categorized as qualitative data and our conventional data tools cannot process it. It dosen't have a proper model so we have NoSQL databases for storing this type of data. Even they cannot guarantee us the proper storage and processing of it so people tend to go for Data Lakes which is built with a concept of bigdata.

3. **SemiStructured Data:** The combination of Structured and unstructured can be treated as semi structured data. Eventhough there are few organizational properties like semantic tags or meta tags it's still fluidic. The best example of semi structured data is email. Eventhough the actual data is unstructured in an email It has different things like sender email ID, subject, Time etc which are structured attributes. Another example would be an image where it's unstructured but It'll still have attributes like date and time & even organized properly on our mobile phones in a file manager.

Velocity:

The term velocity means how fast we can process it. Several Petabytes of data is being generated by companies these days and the processing ways have still been devised and improving day by day. Initally we had Map reduce framework which was built on

java helped in processing huge amounts of data. It was good for Linear processing of huge data sets and if Intermediate results are not expected. Then Spark came work with a huge revolutionary change by Introducing In memory processing with a 100 times faster in memory and It was ten times faster on disk and can process 100 TB of data 3 times faster than Hadoop Map reduce framework. Later on many frameworks like Apache Flink, Presto, Heron etc came into the market and they're unique in their own way.

What is Big Data?

The data which is beyond the storage and processing capabilities of our systems and is beyond the ability of traditional relational databases to capture, manage and process the data with low latency. Data sources are becoming more and more complex as they are driven by several components like AI , mobile devices, sensors, satellites, micro controllers, transaction applications, World wide web and social media. With Big Data we can gain business insights and data driven insights which will help an organization to improve their business and understand what people need in a long run.

Nowadays, companies are investing time in finding ways to deal with real-time insights to compete with other competitors. Companies and market researchers strongly believe that Big Data will bring a huge change, huge value, Huge return on investment,huge competition, and huge impact at a big scale in all business domains, including manufacturing, logistics, health care, retail, banking, insurance, financial services, government, etc.

The recommendations given by Youtube that I would be interested in certain types of videos based on my past plays and also pattern of people's past plays with "similar taste" was enabled by analysing massive amount of data which Google's computers had gathered from millions of users.This is a typical recommendation system. The recommendations that amazon gives us based on our earlier purchases is also by using Big Data and using it effectively to understand people tastes. Discovering patterns in

massive data of different varieties to formulate new frameworks is the essence and also it was the biggest challenge.

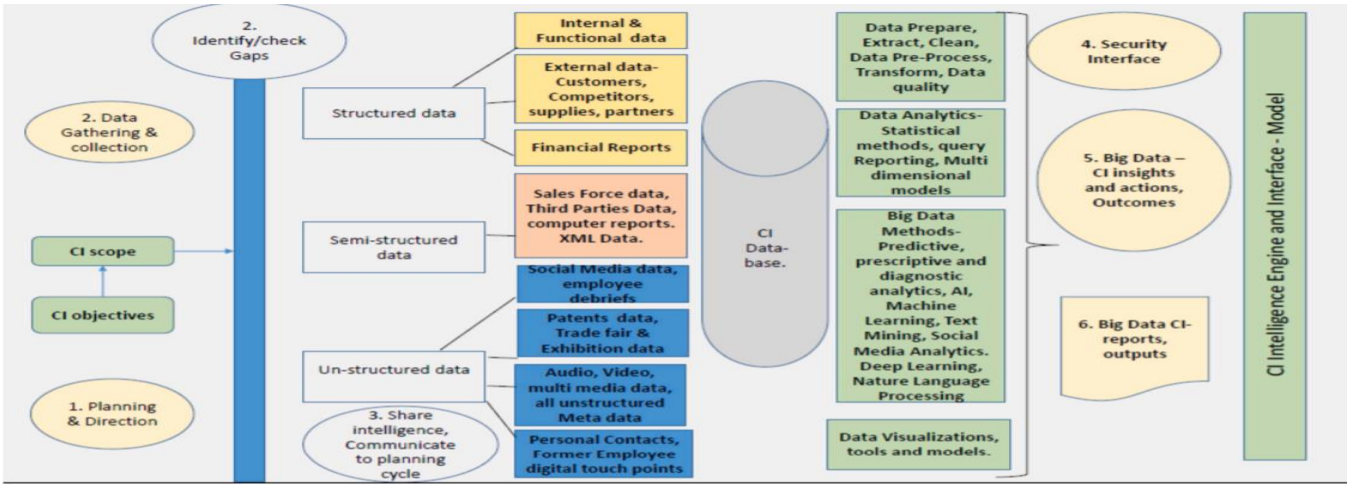
Literature Review

According to Frizzo-Barker, Chow-White, Mozafari, and Ha (2016: 412), who prepared a useful systematic literature review on Data Engineering with Big Data across the business between 2009 and 2014, "*although the field is in its earliest stages of scholarly development, we found clear evidence of the energy and increasing interest focused on Big Data studies in business.*" Various insights have examined initiatives of Big Data in various companies. These include revolution of Big Data in corporate strategies and management (George, Haas, & Pentland, 2014); the impact of Information technology in the productivity of a Company (Chang & Gurbaxani, 2012); Life cycle of Big Data (Khan, Liu, Shakil, & Alam, 2017); Applying Machine Learning libraries on Big Data (Zhou, Pan, Wang, & Vasilakos, 2017); the evolution of big data (Lee, 2017); 3 tier-based strategies for Big Data in various Companies (Matthew & Mazzei, 2017);

Practically, Various companies have used different statistical techniques and advanced database methods such as data mining & few processes required extensive trial-and-error mechanisms using various qualitative and quantitative data for modelling. It has observed that customer loyalty is in rapid decline, and there are companies that sell data to various organizations on what users see, tag, post, listen to, comment on, link, read, like, etc. Big Data gives companies the ability to look at things in a different way that enables them in their business Insights. Due to evolving nature of Big Data and Its tools, companies will utilize such huge, more competitor data sets to find impacts, gain Insights and work with irregularities.

Even a small firm or a company with a positive attitude towards Big Data will be able to take a huge advantage against a much bigger huge firm or a company simply by understanding their patterns and the way they deal with the data in the market. As long as companies have access to data – from both internal and external sources – analysing that data correctly is what will produce competitive advantage.

Nowadays, companies are interested in finding ways of engaging more with real-time data dealing with competitors. Companies and market researchers strongly believe that Big Data will bring huge change,huge value, huge return on investment, huge competition, and huge impact in a long run for all business domains, mainly manufacturing, logistics, health care, retail, banking, insurance, financial services, government,etc.



The above figure demonstrates the flow of data that happens in an organization or firm. It all starts with data gathering and collection. In this phase data is collected from various sources and formats. After that we define a scope and Objective for our analysis. Then we identify gaps and categorize all the data we got into structured, unstructured and semistructured data. Internal and functional data, Customer & Competitor data,financial reports data goes under structured data. XML data, Salesforce data, Json data goes under semi structured data and all the social media data, Patents data,Audio and video data will be under unstructured data.

Then In the next stage we will have a CI database where all this data will be stored which we are calling it as Hadoop or any Cluster which will be holding these vast amount of data. From here our Processing starts. First step will be data cleaning, data extraction and data preprocessing. We use data analytics, Statistical models, Multidimensional models to fit our data. Then we use this preprocessed data with our Big data processing tools like Pyspark and It’ll take care of all the transformations and actions after that we get the data in a proper format such as a dataframe or dataset.

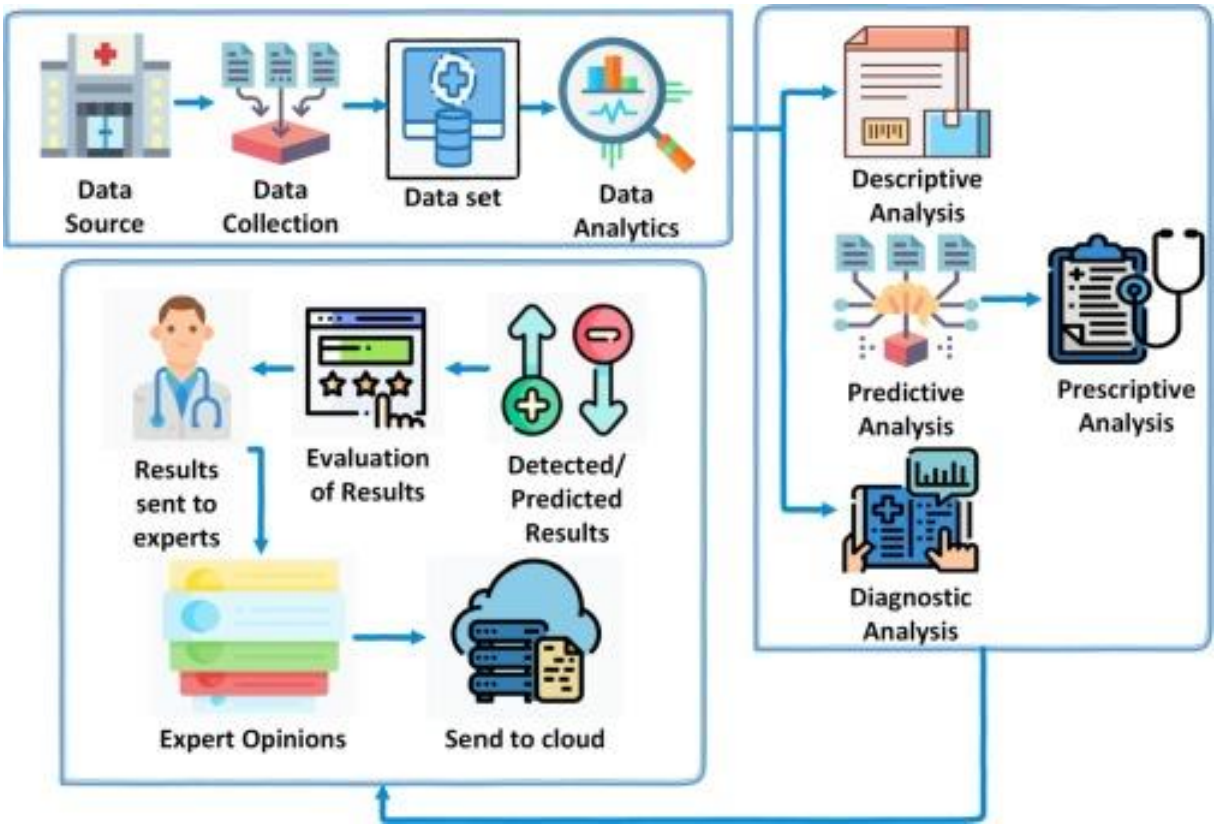
Now we can perform Predective and descriptive analytics with Machine Learning models like Text mining, Social media Analytics etc and gain some business insights about the data. We can also use Artificial Intelligence frameworks for digging more into data and understanding it proactively. After this is done we can use Data visualization tools or models and reporting tools to create some graphs or heat maps which will give us the insights in a statistical way and helps us further in decision making.

After this stage we get CI insights and actions and outcomes with some statistical reports and this entire architecture forms as a CI Intelligence Engine and Interface model. This process repeats on a regular basis and all these processing part is being taken care by Cloud providers like AWS, IBM , Microsoft, GCP etc. On IBM cloud we will have analytics engine that helps us to run our pyspark jobs. On AWS we have an EMR service. On Google cloud platform we have DataProc and Oracle cloud platform also has some infrastructure that let's us deploy these big data applications and helps us to achieve this functionality.

During the Covid-19 pandemic Big data analytics played a major role in pandemic prediction. IoT devices and sensors emit huge amount of data that might be useful for healthcare sectors. if analysis is the theme, then big data analytics comes into the picture. Recently, the novel coronavirus pandemic (COVID-19) outbreak had a serious impact on healthcare industry and the human life. In this scenario, the IoT and big data together have played a major role in fighting against covid-19. The applications majorly include the huge collection of vast amounts of data, visualization of pandemic information, tracking information of covid cases, and adequate assessment of COVID-19 prevention and control.

A framework was designed by Imran Ahmed that takes advantage of Big data analytics and IoT. We deal with descriptive, diagnostic, predictive, and prescriptive analysis and apply big data analytics on original covid realtime data set, which was focussing on different symptoms in panademic. The main aim was integrating Big Data Analytics and IoT to analyze and sense the covid-19 panademic. Neural networks was used and helped to diagnose and predict the pandemic, which can help

medical staff. We predict pandemic using neural networks and then compare the results with other machine learning algorithms. The results state that the neural network was performing effectively at an accuracy rate of 99%.

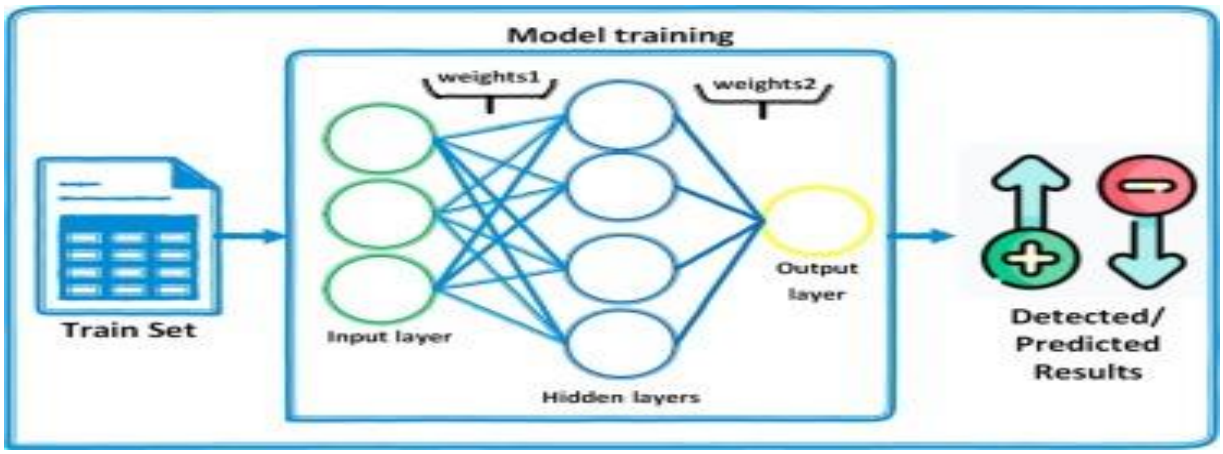


The above figure explains how IOT and Big data analytics were successful in panademic analysis and prediction. The dataset which was used is structured data that was obtained from different hospitals. It contains huge amounts of different patients data (males and females) of various age groups. Several attributes were used for analysis, prediction, and detection of a pandemic. The target attribute of the data set was the Lab Results and patients survived after the diagnosis of the virus.

In the descriptive analysis, detailed information about every attribute present in the data set is provided, including the number of attributes, features, and the size of attributes in the data set. Using Data preprocessing libraries like seaborn, the summary or detailed description of the raw data is made and displayed as visual graphs that is

understandable by humans. The diagnostic analysis which is also an advanced form of data analytics was used to examine some insights about the data that answers the question “Why did it happen?”. It uses different set of attributes and features in order to determine the relationships. We can call it as, data mining, and correlation techniques. It helps to understand the causes and behaviour of the data. In health care, diagnostic analytics explore the data and make correlations using different attributes information. Diagnostic analysis of data set involves in data discovery and applying data mining algorithms. Different attributes and their relationships are analyzed i.e. relationship between different symptoms of COVID-19.

Further in Predictive Analysis The data is sent to a deep learning model that focuses on the data's key patterns and trends. The model is then trained with current data for prediction. In healthcare, predictive analysis is used for disease forecasts, Survival rates etc. Individual features are selected from the data set. The information is further separated into train and test data. The training data is given as an input to our machine learning or deep learning model and trained. For prediction purposes, the trained model is tested using other sets of data set. The results are verified using different parameters and sent to experts for opinions.

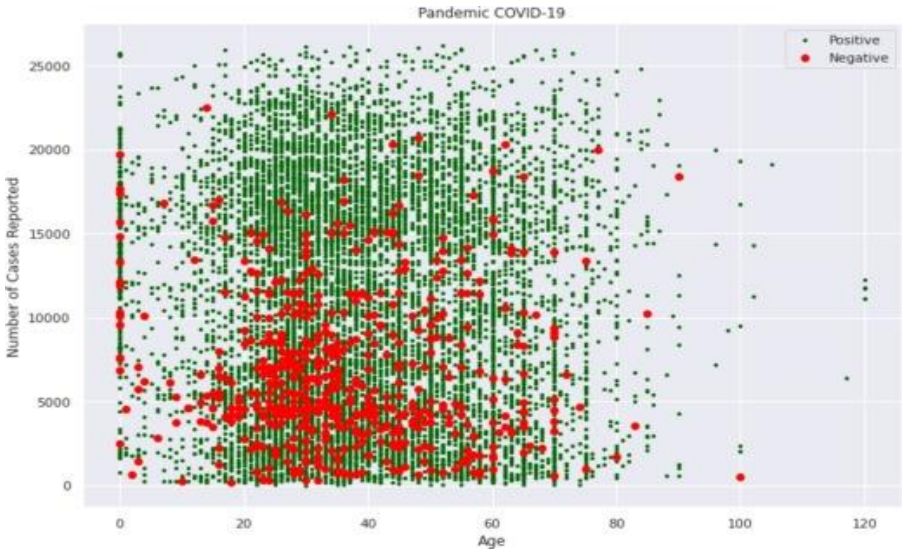


The above figure illustrates the Deep Learning model. A Neural Network Algorithm or model was used. It contains an input layer x, output layer y, and few hidden layers containing 4 neurons. Each layer will have a set of biases and weights represented as b

and W except the output layer. A sigmoid function which is also known as an activation function. The fine-tuning of biases and weights from the input data is done in order to enhance the prediction accuracy of the model. In training the model, each iteration will have the following steps:

- Feed-forward operation which helps in calculating predicted output y.
- Back-propagation technique that updates the weights and biases.

Further In Prescriptive analysis of data set the results of our model are sent to medical experts for further conclusions or opinions after applying different data analytics techniques. For example, what do we do if a person is diognosed. Decisions like hospitalizing patients with positive results, medications, and do they need to self quarantine? are answered by these exprets. The prescriptive analysis also helps in preventing and controlling covid-19 panademic spread.



Further as a result we can see that the number of positive cases which was detected by our neural network model is greater than the number of negative tests. These results states that the patients with symptoms are having more chances of getting infected. A medical expert can easily analyze and conclude that there are higher chances of positive cases in patients who are dealing with any of the discussed symptoms, and they can start their treatment earlier before waiting for test results.

Trends in Big Data Analytics

Let’s discuss on the topic which was written by Karthik Kambatla on Trends in Big Data Analytics. Author mainly focused on Trends in scale and application scope of Big Data Analytics and discussed on Software Techniques that are currently used and future trends to address different applications.

The work load characteristics of different big-data analytics applications and scaling them individually, provides meaningful insights on how our future hardware and software applications will be.Recent hardware advancements have played a major role in understanding the distributed form of software platforms needed for big-data analytics.

Future hardware innovations like In processor technologies, Latest memory or storage hierarchies, Newer Software defined networks or architectures will continue to drive software innovations. Following newer and practical appproaches in design of these systems will be majorly on reducing the time required to move the data from storage to the processor or between storage/computational nodes in a distributed system.

Parallel and distributed processing applications which are current trend in the market comes under big-data analytics. Data storage for such applications is beyond our capacities and are growing at a higher rate in size everyday. These datasets and it's corresponding applications are providing significant challenges for the IT industry and software development. Datasets follow distributed patterns and often placed on the platforms with huge computational power and processing capabilities.

Features like fault-tolerance, high availability, and security are the challenges for many applications. Data Analysis have deadlines, and data quality is a major concern in some of the applications. For the modern applications, data models and processing mechanisms, capability of operating at a particular scale, are as-yet unknown. On top of all of the above challenges, data validation and output is a major issue. The functions of hardware platforms and the software stack are having a major impact on big data analytics.

The mechanism of accessing data and in specific at what frequency the data is being accessed (cold versus hot data) can drive or help us in optimizing future storage applications: data is usually hot nothing but raw and can be in any format; however as we move on, it becomes archival, i.e. cool so that we can store in our naive databases. However, there are exceptions for accessing hot massive datasets (Genomic statistical calculations) which should be taken into consideration.

For Instance A single video, which is in multiple formats or different language subtitles, the result of it will be in many versions. These can either be produced offline and stored (ample storage) or they should be generated on the go (Transcoding and translation) with a pressure on the data center infrastructure, or alternatively on the user's system which is client-side computing.

Better Understanding of the workloads can also create or provide oppurtunities for implementing special purpose processing elements which are directly fed into hardware. Graphics processing units, Field Programmable Gate Arrays, component specific integrated circuits, and Encoders and decoders.Such hardware components will reduce the consumption of energy on the hardware. These can be integrated on-chip, which can be further placed into data-centric asymmetric multiprocessors

Software systems and It's associated data storage, and processing power need to provide solution to a large problem space resulting from scaling the data, workload's nature, and other requirements like data consistency, high availability, and fault tolerance. Huge data scales demand highly-scaled distributed storage systems which can process huge volumes of data, with efficient ingress and egress operations.

Apache Flume, Apache Kafka, Kubernetis are few examples of systems that facilitates data movement. Parallely, these storage systems should support in-memory caching like Spark for querying efficiently and other Online Transaction Processing workloads; even our Hadoop Distributed File Systems recently added support for caching.

Big-data analytics allows good accuracy constraints on its quantitative results which can play a major role in algorithm design. The volume of data operated upon by latest

applications is growing at a huge rate, thereby giving challenges for parallel and distributed computing platforms. These challenges range from building storage systems that can accommodate these large datasets to collecting data from various sources across the globe into storage systems which can run advanced computations on the data.

Few resource constraints, like Brewer's CAP theorem, require problem handling as per application requirement, exploiting application-specific functions and it's heuristics. Recent efforts on working with these problems or challenges led to scalable distributed storage systems precisely file systems and execution engines that can handle a variety of computing paradigms. In the future, as the size of data grows and the applications starts diverging, these systems should be focussing on application specific optimizations. To tackle this highly distributed data , future systems might leave some of the computations to the source itself which will decrease the computational costs.

Big Data Analytics – A layer between Security Experts and Data Scientists

Big Data Analytics-as-a-Service fills the gap between Data security experts , Machine Learning engineers and Data scientists. Both can work and contribute to the deployment of Cyber security analytics process. Additionally They reduce the impact of the lack of labor who have the knowledge about both Cyber Security and Data Science. We live in an interconnected world where large amounts of data is emitted and collected every second.

To use the data at it's full potential by performing advanced analytics, applying machine learning algorithms and building artificial intelligence models, are important for businesses these days, from small scale firms to large organizations, resulting in a main advantage (or shortcoming) in the ever

growing market for business analytics solutions. This situation is deeply changing the security landscape, new risks and threats are introduced everyday which is effecting security and privacy of systems, on one side, and exploiting user data, on the other side. Many firms or organizations that are relying on solutions based on Big data analytics have strict security requirements to fulfill.

The Energy domain's Smart Grid is one of the real time example of systems which worked between security layer and Business intelligence. The Smart Grid plays an important role on the infrastructure build for enerygy. However, it has two major challenges that are related to data security namely managing front-end intelligent components such as power assets and smart meters, and securing the the large amount of data received from these components.

On the other hand proper analytics setup is a complex problem because security controls may have an other side which may decrease the analytics accuracy. This is even more crucial and complex when the security configuration controls are given to the security expert, who just has basic knowledge of big data analytics and data science. To solve this issue We use the concept of Model-Based Big Data Analytics-as-a-Service (MBDAaaS) that fills the gap between Cyber security experts and data scientists. The solution acts as a middleware allowing a security expert and a data scientist to work together collobaratively.

MBDAaaS methodology as a middleware was capable to orchestrate the users that are involved in the business process i.e. Cyber security experts and data scientists in carrying out a data analytics. This approach has been taken in a security scenario typical of Smart Grids, where security experts and data scientists work together and build some data analytics which will be used to detect security breaches and incidents by preserving privacy with the help of log analysis.

Applications of Big Data Analytics

Big data applications can help different organizations in making smart business decisions by analyzing large volumes of data and discovering hidden patterns. These data sets might be coming from various data sources mainly from social media, sensor data, website logs, feedbacks of the customers, etc. Organizations are investing alot on big data applications and data engineers to discover hidden patterns, understanding market style, prefences of the consumer, and other valuable business information.

There are various domains where Big Data Analytics is used for business insights which majorly include following:

- Manufacturing and Natural Resources.
- Communications, Media and Entertainment.
- Banking and Financial Sector.
- Security Management
- Internet of Things
- Retail and Wholesale trade
- Government
- Insurance Domain.
- Education Sectors
- Transportation domain
- Energy and Utilities

Big Data Analytics is used pretty much everywhere and all the domains because In every field companies or firms will definitely compete with their competitors and In order to do that they need the data to be tuned or analyzed correctly for a long run profits. Now let’s discuss about all of the above applications of Big Data in detail and how they are used in real time to make business decisions.

Big Data Applications in Government

The ability to handle and produce value from enormous streams of data from various sources and in various formats (structured/stored, semi-structured/tagged, and unstructured/in-motion) appears to be a new kind of competitive difference for elected representatives, administrations, and people. Many governments that are implementing or developing big-data projects must take a step-by-step approach to develop suitable objectives and reasonable expectations.

Their capacity to combine and analyse data (using emerging technologies like Hadoop), construct supporting systems (such as big-data control towers), and assist decision-making through analytics is critical to their success.

An analytical firm will handle and integrate massive data that crosses departmental boundaries, a top-down approach is required. Governments should consider establishing big-data control towers to combine gathered datasets from departmental silos, whether organised or unstructured.ss

Furthermore, governments must create an advanced analytics agency to develop strategies for managing big data through new technological platforms and analytics, as well as how to secure experienced professional staff.

Collaboration on a global scale The Group on Earth Observations (GEO), for example, is a collaborative multinational intergovernmental effort to integrate and distribute Earth-observation data. Its Wide Earth Science System of Systems (GEOSS), a greater public architecture that creates accurate, near-real-time data from the environment, aims to give data and analysis to a broad variety of global users and judgments.

Data about security dangers, fraud, and criminal activities must also be shared by governments. Not only does big data necessitate translation tools, but it also necessitates an international cooperative process to share and integrate information.

Applications of Big Data in Transport

Huge volumes of data from location-based social networks, as well as elevated information from telecommunication, have recently influenced travel behaviour. Regrettably, investigation on travel behaviour has still not evolved at the same rate. Transport demand models are still largely based on poorly understood new social media structures in most regions.

The use of Big Data by organizations includes traffic conditions, route planning, intelligent transportation systems, and congestion management (by predicting traffic conditions. Revenue management, technology advances, logistics, and competitive advantage are all examples of private-sector use of Big Data in transportation (by consolidating shipments and optimising freight movement). Individual applications of Big Data include route planning to save money and time, as well as travel plans in tourism.

APPLICATIONS OF BIG DATA IN HEALTH CARE

The medical industry has accessibility to enormous amounts of information, but it has struggled to use it to manage healthcare costs, as well as outdated systems that inhibit quicker and more effective healthcare coverage all across the board.

A few clinics, such as Beth Israel, are leveraging information from millions of patients obtained through a mobile phone app to help clinicians to practice evidence-based medicine rather than administering multiple medical/lab tests to all patients that visit the hospital. A series of testing can be useful, but they can also be highly inefficient.

The University of Florida used free healthcare data with Google Maps to create visual data that enables faster analysis and identification of healthcare information, which is used to track the spread of chronic disease. Obamacare has also made extensive use of Big Data. Heterogeneous Data, Humedica, Explorys, and Cerner are some of Big Data providers in this industry.

APPLICATIONS OF BIG DATA IN EDUCATION

The ability to significantly enhance education. We manage to create a contemporary, innovative educational system from which all single school can benefit to the fullest extent possible. Moreover, teachers now have vital resources that they did not already have, allowing them to make more detailed selections to choose from a wide range of innovative learning approaches.

As a result, Data Analysis is actively involved in changing the way businesses, especially education, operate. Traditional challenges will no longer be relevant inside the new phase of data, although excellent procedures will be maintained.

New learning methods will be added to the education system, making it more efficient and targeted. However, the journey into this new era has only just begun, and there are other challenges ahead, including a scarcity of competent employees in the fields of Big Data and Data analytics. Finally, teachers and academics must be trained and involved in using these new technologies, and students must embrace and use them.

APPLICATIONS OF BIG DATA IN MANUFACTURING AND NATURAL RESOURCES

Big Data has been used for absorbing and integrating vast volumes of data from geographical information, graphical data, textual, and spatial analysis in the natural resources industry, allowing for predictive modeling to enhance decision making. Seismic interpretations and reservoirs characterization are two areas where this has been applied. Among other things, big data has been used to solve today's manufacturing difficulties and create a competitive advantage.

APPLICATIONS IN MEDIA AND ENTERTAINMENT

The Media and Entertainment Industry likewise integrates and collects the same type of data from numerous sources in order to better understand viewer behavior and improve in a way that would allow them to thrive and be the viewers' favorite among all of them.

It's a well-known advertising and potential revenue fact that the better you know your consumer, the better you can anticipate their preferences and adjust price, content, and UI accordingly.

Big data gives all of the knowledge that every industry requires to attract and retain loyal customers. Companies can channel and improve client behavior by retrieving data based on different criteria such as age, location, language, and so on because they could fetch data based on numerous criteria such as age, location, language, and so on.

E-commerce enterprises can benefit from the same algorithms because these platforms can serve as one of the greatest remarketing and advertisement platforms. This is obviously a backdrop and a complicated process, but it plays a critical part in the formation of media and entertainment organizations, as previously said.

Gone are the days when viewers had no choice but to watch a single station with no integration or consultation. However, these dynamics are shifting; there are now millions of viewing alternatives to pick from, and they can also be streamed across several devices, making them much more user-friendly.

Big Data, which is a combination of real-time data, network information, time series, and other types of data, has undoubtedly played a key part in the creation and successful implementation of these concepts and alternatives in the film industry.

APPLICATIONS OF BIG DATA IN RETAIL AND WHOLESALE INDUSTRY

From conventional brick-and-mortar shops and wholesalers to today's e-commerce merchants, the industry has amassed a wealth of information throughout the years.

This data, obtained through loyalty rewards cards, Checkout scanners, RFID, and other sources, is not being used effectively to improve overall customer experiences. Any modifications or enhancements have been gradual.

Involves a broad range of possible applications and is slowly but steadily being used, particularly by brick and mortar companies. Social networks are used for new customers, retaining customers, promotion of the product, and other purposes.

Conclusion

There is always need for the data. Be it in any Business from a small scale industry or a firm to huge organizations there is always a compettetion among businesses for the right data which can give them proper response about the market trends and generate value out of it.

Big Data or Data Engineering plays a crucial role in achieving this. Be it Association rule chaining where it helps to discover interesting correlations in a huge database. It's one of the most important big data analysis technique where it identifies whether a combination of two items or sets is likely to be taken by people with any product or not. For Instance It can tell us whether people who buy milk and butter are likely to buy a tissue or sugar or latte powder with it.

Techniques like Genetic algorithms help us derive solutions based on mechanisms such as Inheritance and mutation for the problems which requires optimization. For an Instance Scheduling doctors for hospital emergency rooms, generating different content based on people's tastes on a Television etc.

Techniques like Sentimental analysis helps business to know about the sentiments of users who use their service. For an Instance Companies like Zomato or Uber eats, they take feedback from the customers how the ride was or how the food tasted like and based on those feedbacks they improve their service.

Big Data is pretty much everywhere. It's all in the data and business need data for every single decision they make. Indeed there is a proverb which states that "Everything is Data and Data is Everything".

References

Ranjan, Jayanthi, and Cyril Foropon. "Big data analytics in building the competitive intelligence of organizations." International Journal of Information Management 56 (2021): 102231.

Ahmed I, Ahmad M, Jeon G, Piccialli F. A framework for pandemic prediction using big data analytics. Big Data Research. 2021 Jul 15;25:100190.

Kambatla, K., Kollias, G., Kumar, V. and Grama, A., 2014. Trends in big data analytics. Journal of parallel and distributed computing, 74(7), pp.2561-2573.

Ardagna, C.A., Bellandi, V., Damiani, E., Bezzi, M. and Hebert, C., 2021. Big Data Analytics-as-a-Service: Bridging the gap between security experts and data scientists. Computers & Electrical Engineering, 93, p.107215.

Awan, U., Shamim, S., Khan, Z., Zia, N.U., Shariq, S.M. and Khan, M.N., 2021. Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. Technological Forecasting and Social Change, 168, p.120766.

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., et al.(2020). Big Data analytics and artificial intelligence pathway to operational performance, 226, Article 107599

Morabito, V. (2015). Big Data and analytics: Strategic and organizational impacts. Business & Economics, 100–140.

Li, W., Chai, Y., Khan, F., Jan, S.R.U., Verma, S., Menon, V.G. and Li, X., 2021. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mobile Networks and Applications, 26(1), pp.234-252.

Iftikhar, R. and Khan, M.S., 2022. Social media big data analytics for demand forecasting: development and case implementation of an innovative framework. In Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 902-920). IGI Global.

Mehta N, Shukla S. Pandemic Analytics: How Countries are Leveraging Big Data Analytics and Artificial Intelligence to Fight COVID-19?. SN Computer Science. 2022 Jan;3(1):1-20.

Shah, T.R., 2022. Can big data analytics help organisations achieve sustainable competitive advantage? A developmental enquiry. Technology in Society, 68, p.101801.

Manogaran G, Thota C, Lopez D. Human-computer interaction with big data analytics. InResearch Anthology on Big Data Analytics, Architectures, and Applications 2022 (pp. 1578-1596). IGI global.

Do Nascimento, I.J.B., Marcolino, M.S., Abdulazeem, H.M., Weerasekara, I., Azzopardi-Muscat, N., Gonçalves, M.A. and Novillo-Ortiz, D., 2021. Impact of big data analytics on people's health: Overview of systematic reviews and recommendations for future studies. Journal of medical Internet research, 23(4), p.e27275.

Rahmani, A.M., Azhir, E., Ali, S., Mohammadi, M., Ahmed, O.H., Ghafour, M.Y., Ahmed, S.H. and Hosseinzadeh, M., 2021. Artificial intelligence approaches and mechanisms for big data analytics: a systematic study. PeerJ Computer Science, 7, p.e488.

Dahiya, R., Le, S., Ring, J.K. and Watson, K., 2021. Big data analytics and competitive advantage: the strategic role of firm-specific knowledge. Journal of Strategy and Management.

Biesialska, K., Franch, X. and Muntés-Mulero, V., 2021. Big Data analytics in Agile software development: A systematic mapping study. Information and Software Technology, 132, p.106448.

Ahmed, H.M., Javed Awan, M., Khan, N.S., Yasin, A. and Faisal Shehzad, H.M., 2021. Sentiment analysis of online food reviews using big data analytics. Hafiz Muhammad Ahmed, Mazhar Javed Awan, Nabeel Sabir Khan, Awais Yasin, Hafiz

Muhammad Faisal Shehzad (2021) Sentiment Analysis of Online Food Reviews using Big Data Analytics. Elementary Education Online, 20(2), pp.827-836.

Lee, M., Kwon, W. and Back, K.J., 2021. Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. International Journal of Contemporary Hospitality Management.

Awotunde, J.B., Jimoh, R.G., Oladipo, I.D., Abdulraheem, M., Jimoh, T.B. and Ajamu, G.J., 2021. Big data and data analytics for an enhanced COVID-19 epidemic management. In Artificial Intelligence for COVID-19 (pp. 11-29). Springer, Cham.

Stehel, V., Bradley, C., Suler, P. and Bilan, S., 2021. Cyber-physical system-based real-time monitoring, industrial big data analytics, and smart factory performance in sustainable manufacturing Internet of Things. Economics, Management, and Financial Markets, 16(1), pp.42-51.