A Review of analytics platform architecture in a Diversified Approach for Big Data Techniques

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

[1] Research Scholar, Department of Computer Science & Informatics, Maharishi Arvind University, Jaipur, Rajasthan, India

[2] Professor & Supervisor, Department of Computer Science & Informatics, Maharishi Arvind University, Jaipur, Rajasthan, India

Abstract—The rapid climb of the latest technologies, modernization of application, and therefore the dimensionless growth of communication industries produces a multivolume of data every year. Data comes from various sources at different rates. It's impossible to use traditional data management strategies because of the complexity of big data. The new age of big data is a result of this. Big data is here to stay. On the other hand, Traditional data Analytics, are incapable of dealing with such enormous data volumes. The challenges that arise right now are how to establish a high-efficiency system for big data analysis & how to design an appropriate mining algorithm to discover important objects in vast data. The development of big data applications has risen in importance over the last several years. Data mining is being used by an increasing number of firms across a broad variety of sectors to obtain meaningful insights. This study focuses on the analytical perspectives expressed by several academics in the field of huge data. This article begins with a basic overview of data analytics and then moves on to a discussion of big data analytics. The majority of its focus is on the platform's key elements and applications, as well as a few huge data concerns. Several crucial unsolved challenges and research objectives will also be presented for the next step in big data analytics.

Keywords—Big data, Big data analytics, Issue and Challenges of big data, Applications, Characteristics, big data tools, and technologies.

- *x* -

I.   INTRODUCTION

Without data storage, it's difficult to envision a world in which every piece of information about a person or organization, every transaction, and every conceivable aspect is permanently lost.As a result, firms lose the ability to gather critical data & knowledge, conduct in-depth analyses, & generate new benefits or opportunities. As a result, everything from customer information to product availability has become crucial for day-to-day operations. A company's success is based on its ability to collect and analyze data.

Think about how much information & data is available now because of advances in technology and the internet. Huge volumes of data have been more readily accessible because of advances in data storage and gathering technologies. To retrieve value from all of the informationwhich is being generated, it must be kept and examined. Furthermore, the cost of storing data has decreased, therefore enterprises need to extract the most value out of their vast volumes of stored data.[1].

Big data analytics must be adapted to deal with the volume, diversity, & quick change of this data. The sheer volume of big data necessitates careful analysis & extraction of relevant data.[1].

Many people throughout the globe create enormous amounts of data in the digital age, making it difficult to keep track of everything. There are several different types of data in this set, all of the varying quality. As the need for storing and analyzing ever-increasing volumes of complex data grows, relational database suppliers have responded by creating a variety of specialized analytical systems, ranging from software alone to analytical services which function in third-party hosted environments. Data from web traffic, social media material, or machine-generated data like sensed data & global positioning system data is increasing in volume and complexity. For ad hoc searches against semi-structured information, new non- relational databases integrate text indexing & NLP using classic database technology. Platforms like this one may be available on the market for the analysis of complex, structured, and unstructured data.[2].

www.ignited.in

Figure 1: Big Data

Big Data is often defined as having the following 3 primary features (called the 3Vs)[3]:

•   **Volume:** More and more data is being gathered and processed from a variety of sources (ICTs, mobile phones, product codes, social networks, sensors, logs, etc). In 2012, an estimated 2.5 exabytes of data were produced per day. Every 40 months or so, this figure doubles. It was projected that 4.4 Zettabytes of digital data had been generated, duplicated, and consumed in 2013 by the International Data Corporation (a research report publishing organization) (ZB). It's growing at a rate of double every two years. Data generated digitally reached 8 ZB in 2015. By 2020, IDC predicts that the amount of data will exceed 40 zeta bytes, an increase of 400 times above the current level.

•   **Velocity:** A lot of data is created quickly, therefore it's important to analyze it quickly to get the most out of it. Customers' purchases create more than 2.5 PB of data per hour for Walmart (a bargain retail company), for example. Another excellent illustration of Big Data's lightning-fast speed is the video-sharing website YouTube.

•   **Variety:** Massive volumes of data are generated in several ways and from various sources (e.g., videos, documents, comments, logs). This includes both organized and unstructured data, as well as public and private data that is both local and distant.

To describe datasets that have become so vast that typical database management systems can no longer handle them, the phrase "Big Data" has recently been coined. Those are data sets that are too large to be processed in a reasonable amount of time by frequently used software tools and storage methods[4]. The volume of big data is constantly increasing, with a single collection now containing anything from a few hundred to several petabytes (PB). As a result, gathering, storing, searching, distributing, analyzing, and displaying large amounts of data may be problematic. Today, organizations are studying vast amounts of very comprehensive data to find things they didn't know before[5]. As a result, complex analytic methods are used on large data sets in big data analytics. Using massive data sets, analytics may expose and influence corporate transformation. The more data there is, the more difficult it is to keep track of.[1].

Analysis of existing literature on big data analytics is the paper's main contribution. Thus, some of the numerous big data tools, methodologies, & techniques that may be employed & their applications and potential in various decision-making areas are reviewed.

II.   BIG DATA ANALYTICS

The goal of big data analysis is to unearth patterns, correlations, as well as other insights buried within the enormous amount of data that has been collected.[6]. With today's technology, you can evaluate your data and get answers from it quickly. Slower and less effective than standard business intelligence systems. In the realm of Big data analytics, which deals with data sets ranging from terabytes to zeta bytes in size and originating from a wide variety of sources, sophisticated analytical approaches are applied to big, heterogeneous data sets.[7].

2.1 Big Data

Large or complicated data collections that are beyond the storage, administration and processing capabilities of traditional relational databases are referred to as "big data". A dataset must fulfill at least one of the following three requirements to be deemed "big data": volume, velocity, or variety.With the development of new data kinds & knowledge sources brought on by mobile, social, & the Internet of Things, data complexity is rising across the board. Real-time & large- scale big data are created by sensors & devices that capture video or audio, network logs & transactional apps on the web, as well as social media[7].

Analysts, academics, & business users can now leverage previously unavailable or inadequate information to create better & more timely choices. With the use of modern analytic methods such as text analysis and machine learning as well as statistical and natural language processing tools, organizations can acquire new insights from previously untapped data sources.

2.1.1 Big Data Architecture

Big Data architects need to have the ability to design and implement reliable, scalable, and automated data pipelines. A comprehensive understanding of the whole stack is required for this job, from cluster design through Hadoop tuning and on up to the setup of the top chain responsible for processing the data. Using the following picture, you can see how the data pipeline affects every aspect of the stack.
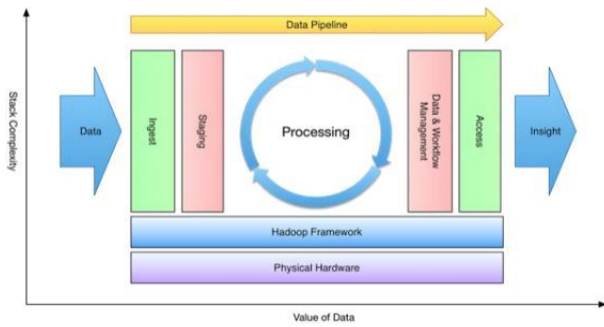
Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

Figure 2: Big Data architecture

Here, the most important thing to remember is that data pipelines take raw data and transform it into actionable information (or value). An important choice for the Big Data engineer is how the data will be handled as it moves through the system: where it will be kept, how it will be accessed internally, what tools will be used to analyze the information, and how it will be made available externally. Tools like Impala or Apache Spark are likely to be used for data processing, whereas business intelligence (BI) or other analytic tools are used for the latter. Big Data engineers are those who create and execute this architecture.

2.2 Big Data Analytics

An ever-increasing amount and variety of data are being processed through big-data analysis techniques, which are typically streamed in nature. Because of the enormous dimensionality, heterogeneousness, unstructured incompleteness, erroneous and noisy nature of big data it has to be polished from a variety of angles. To avoid losing the power of large data, we must use newer, more advanced techniques to analyze it. NLP, DL, AI, ML, & other innovative tools & methodologies have been developed & utilized to increase the accuracy, speed, and precision of big data analysis. These techniques and processes reveal the underlying pattern, find the previously unknown relationship, & retrieve important information from the data avalanche. For example, data analysis can reveal which areas of a city have high housing prices, analysis of patient reports can reveal early recognition of the disease in a patient and prompt decision-making at an early stage[8], and sales trends can help management formulate better policies to keep customers coming back.[9].
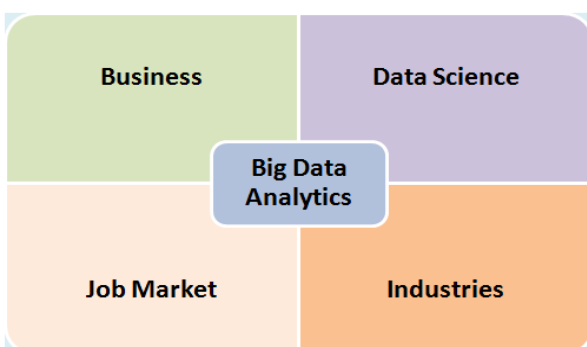

Figure 3: Big Data Analytics

A new set of problems arises when there are unknown and faulty data at all stages of the big-data analytics process. The usefulness and accuracy of the output data may be harmed by big data analytics. Data analysis methods include the following.[10].

(i) Text Analytics: Data from unstructured sources like blogs, company papers, & online forums may be mined using the text mining technique.

(ii) Audio Analytics: It's a method for locating and extracting audio data from large, unstructured datasets. Listening devices like smart speakers, customer service centers, etc. are the most common uses of audio data.

(iii) Social Media Analytics: Facebook, LinkedIn, blogs, micro-blogs, and Instagram owned by Facebook are only a few of the internet platforms that make up this data. It is the study of numerous kinds of data. The 2 kinds of graphs that make up the framework of a social network are social graphs and activity graphs.[11].

(iv) Predictive Analytics: Using previous & present data, it can forecast what will happen next. Market basket consumer behavior, staff attrition, etc. may all be predicted with predictive analytics. SVM, neural networks, decision trees, & linear regression are some of the forecasting methods.

III.   TECHNOLOGIES USED FOR BIG DATA

The term "Big Data" is used to describe a class of software intended to analyze, process, and extract information from massive data volumes that standard data processing tools simply cannot handle.Many methodologies & methods from big data analytics are combined in the technology [14].

Big data technologies are driving the fastest increases in data volume in banking, healthcare, insurance, securities & financial services, and telecommunications. Financial institutions have an abundance of compelling use cases for big data analytics, such as customer service and fraud detection, which makes it even more exciting.Big data solutions are available from a wide range of technology providers.[12]Many of the prominent big data solutions nowadays fall into one of the following groups:[7]:

(i) The Hadoop Ecosystem: Many services are provided by the Hadoop Ecosystem to tackle large data concerns. Different open & commercial projects are all included. MapReduce, YARN, & Hadoop Common are 4 of the most important components of Hadoop.

(ii) Spark Apache: To do computations quickly, the Spark cluster computing technology was created. As a Hadoop MapReduce extension, it

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

462

makes use of the MapReduce paradigm for a wider range of computations such as interactive queries and real-time data processing on the cloud. Using Spark's in-memory cluster computing, an application's processing performance may be significantly boosted Spark is built to handle a broad variety of tasks, including batch processing, iterative algorithms, interactive queries, and real- time streaming. With a single system, you don't have to worry about managing a variety of different tools.

(iii) R: R is a premier data science programming language with a robust set of functions for dealing with all aspects of processing large amounts of data. Apache Hadoop ecosystem, HDFS, and MapReduce frameworks will be added in the future. (iv) Python It helps programmers to write code with fewer lines of code and in a more understandable format. For scientific computing, it provides scripting capabilities and a wide range of sophisticated libraries like NumPy and Matplotlib. Hadoop and big data are synonymous, and Python and big data are also interchangeable. Since big data requires Hadoop, Python has been enhanced to be natively compatible with the platform. Hadoop MapReduce programming may be done using Python's Pydoop module, which provides access to the HDFS API and the ability to write code.

(v) Data Lakes: In the context of Big Data & real-time analytics, a data lake is a central repository for storing large volumes of as-is, high-velocity, high-variety data. Companies may import massive volumes of organized, semi-structured, & unstructured information. From everywhere, in real-time, into a data lake.

(vi) NoSQL: Databases For large, dispersed data stores, NoSQL databases are the way to go. The usage of NoSQL in large data sets & real-time online applications is on the rise. Data may be stored in a variety of ways: as organized or semi-structured, as unstructured or polymorphic as needed.

(vii) Predictive Analytics: Big data is made possible via Predictive Analytics. Predictive analytics makes use of previous data & consumer knowledge to make predictions based on the large volumes of real-time customer data collected.

(viii) In-Memory Databases: Database management systems that keep their data collections directly in the working memory of one or more computers are referred to as "in-memory databases," or simply "DBS". In-memory databases may be accessed more quickly thanks to the use of RAM.

(ix)Big Data Security: NoSQL databases like MongoDB and Couchbase, as well as Hadoop infrastructures, can all be secured with big data security solutions without interfering with the analytical capabilities which make these systems so valuable

(x) Big Data Governance Solutions: All aspects of data availability, security, usefulness or integrity may be managed by a centralized management system called Data Governance.

(xi) Self-Service Capabilities: Self-Service Analytics, or ad hoc reporting, empowers users to evaluate their data promptly. Users may start generating their reports right away thanks to JReport's self-service BI functionality.

(xii) Artificial Intelligence: Big data analytics & Internet of Things systems benefit from the use of artificial intelligence algorithms. Using AI and sophisticated big data analytics, raw data may be transformed into valuable information for making decisions.

(xiii) Streaming Analytics: With streaming analytics, you can keep tabs on everything that is occurring in your company at any given time and take immediate action as a result of that knowledge. As Streaming Analytics happens instantly, organizations have a limited window of time to act on the data before it loses its value.

(xv) Blockchain: The "open ledger" function of blockchain is to store and handle transactions across a distributed database system. Information in the database is stored in blocks, each of which carries a timestamp and a link to the one before it.

(xvi) Prescriptive Analytics: To predict future outcomes, predictive analytics employs the most advanced technology, such as machine learning and artificial intelligence.

IV. COMPARISON AMONG BIG DATA ANALYTICAL TOOLS

As a consequence of our empirical inquiry of some of the most popular big data analytics tools, we have come up with the comparison shown in Table I.

Table I: Comparison Among Big Data Analytical Tools [13]

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

Big Data Analytical Tool	Pentaho BigDataAnalytics	Alteryx Designer	Datameer BigData Analytic	GoogleBigQuery
Vendor	Pentaho	Alteryx	Datameer	Google
Distinguished Feature	Provides blending of operational data sources with big data sources to create an on-demand analytical view	Offers Analytics Gallery that allows single-click consumption, sharing, and publishing of analytic applications	Offers Instant Visualization at every stage of analysis to gain insights	Offers cloud storage with no charges for loading data and Google"s own infrastructure
Offerings	Provides visualization, integration, and analytics in a single tool	Provides data Integration and Analytics	Integration, analytics, and visualization	Cloud Server Storage, data analysis
Runs as	Standalone software	Standalone software	Standalone software	Cloud-based service
Interface	Web-Based interactive reporting view	Graphical user interface	Spreadsheet like interface	Web-browser or command-line tool-based graphical web interface.
Data Import from Hadoop	Not all distributions are supported Only a select number of distributions may be used	Supports all major distributions	Supports all major distributions	Not supported
Data Storage	On-premise	On-premise	Both in the cloud and on-premises	Storage in the cloud provided by Google
Pricing (Basic)	Personalized price for licenses	$3,995 Per User, Per Year	$2,999 for 10 Users.	The First 100 GB of data processing per month is included in the monthly subscription fee.

Figure 4: Some challenges of big data analytics

(i) The Uncertainty of Data Management: Their designs are focused on enabling operational or analytical processing. NoSQL is short for SQL systems aren't the sole option for large data applications that need high speed. NoSQL techniques include key-value storage using hierarchical object representations like JSON, XML, and BSON. Because of the large variety of NoSQL tools, developers, and market conditions, data management is becoming more & more unclear.

V.   CHALLENGES OF BIG DATA ANALYTICS

It is impossible to handle huge and complicated datasets with typical data processing programs. Incorporating enormous amounts of data is no small

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

feat. There are a variety of issues to contend with. Data curation, collecting, sharing, search, visualization, information privacy, & storage may all be problematic throughout the integration process. Big data management accuracy will lead to more confident decision- making if the data is handled correctly.[14]. The integration of large data and the many difficulties which might be encountered throughout the process.[7][10][15][16].

(ii) Talent Gap in the Big Data: Conventional relational database tools using some alternate information layouts meant to raise access speed while lowering storage footprints, NoSQL data management systems, in-memory analytics, & the vast Hadoop ecosystem are some of the new tools that have emerged in this industry. Currently, the market lacks qualified personnel to work with big data technology.

(iii) Getting Data into Big Data Structure: The goal of a big data management system is to analyze and handle massive amounts of data. Data and information are sent, accessed, and delivered from a broad variety of sources and then loaded into a big data platform, which is complicated. The complexity of data transport, access, and loading is just a small fraction of the difficulty. Traditional relational data sets aren't the only ones that have to deal with transformation & extraction.

(iv) Syncing across: Sources of data It is possible for data that has been imported into big data platforms and has been moved from a variety of sources at varying speeds to quickly go out of sync with the original system. The commonality of data definitions, ideas, metadata, and the like is meant. In conventional data management and data warehouses, there is a danger of data becoming unsynchronized due to the sequence of data transformation, extraction, & transfer.

(v) Insufficient Awareness of the Possibilities of Big

Data: People with management abilities are in charge of running businesses, while IT takes a backseat. Managers are happy with the conventional approaches of data warehouse- based analytics and business intelligence. They don't know enough about the power of big data to see how it might broaden their horizons and improve their decision-making processes. The answer is that senior management must take the initiative and devote significant attention to using big data. It's in everyone's best interest for top management to learn about and implement programs to teach their subordinates about the foundations of big data and how it may benefit their firm. Big data development companies may serve as consultants & demonstrate the value of their services today & into the future.

(vi) Volume and Complexity: The sheer volume and complexity of big data analytics seem to be a barrier too high for corporations to overcome. A wide range of data kinds and origins are available. Additionally, data from non-traditional sources including the Internet of Things, web searches, & social media were included. Because of the wide range of information sources, there is also a wide range of data kinds. These days, data are more than just plain old text. All of these things are included in the package: audio, video, and graphic design. Because the data is unstructured, it's considerably more difficult. Data is readily available, but it must first be organized so that analytics tools will make sense of it. Hadoop MapReduce, Cassandra, or HBase are the best options. Using any of these instruments requires the assistance of an expert. The quickest answer is to hire big data application development expertise.

(vii) Stupendous Infrastructure and a High Cost: Adopting big data is costly, regardless of how you look at it. It is essential to have a top-notch IT hardware infrastructure, which is pricey on its own. Big data management requires a lot of power, space, and personnel. You may be dealing with a high-priced group of data scientists and analysts. You'll need a specialized big data solution, which will cost you a lot of money. Businesses may choose between on-premise and cloud-based systems, both of which have their advantages and disadvantages. Data lakes may save costs by storing data that does not need fast processing. You may avoid these infrastructure expenditures by outsourcing the whole process for creating a bespoke big data solution and analytics. You get the rewards of predictive analysis while others do the effort.

(viii) Trained Manpower: To properly gather & analyze large amounts of data, even if a business is ready to spend a lot of money on technology and software, qualified personnel are still required. If you're an IT professional who has some experience with big data but lacks the competence to be a data scientist, it's just a fact of life. Scientists are difficult to find and difficult to keep. Simply said, the finest big data analysis software is all about numbers. A data scientist is needed to make sense of the analytics & provide relevant outcomes. For the most part, you can avoid having to deal with any of the challenges associated with big data analytics and the generation of useful insights by just hiring a big data development business.

ix) Big Data Is About Real-Time Insights and Predictive Capabilities: Only when you may get real-time insights & predictive capabilities can you reap the advantages of big data. A high level of veracity, a large amount of data, and a high level of accuracy are all required to handle this data rapidly and efficiently. Using real-time analytics, firms may see emerging patterns and adjust their product specifications to meet changing customer demands. If you don't have these skills, then the big data solution isn't offering your company 100%. Real-time analytics and reports which reveal emerging patterns are best left in the hands of a big data

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

development business.

(x) Data Security: Data governance and security issues arise as a result of the proliferation of data. According to Big Data's very nature, information comes from a variety of sources. Having a large number of nodes makes a system more susceptible to attacks, which might result in financial losses. These sources need professional governance methods to manage them & ensure their integrity and security. If you decide to run huge data operations in your own home, you'll face difficulties. Outsource your big data needs to Smart Sight Innovations & you won't have to worry about a thing.

(xi) Up-Scaling: Upscaling is a must for a business that has implemented a system to manage big data. In addition to storage capacity, this refers to processing power and the ability to manage rising demand. Upgrades to the system architecture may be the best option, but this is a never-ending process. Allowing bespoke big data application development solutions to handle system scalability & upgrades while you reap the benefits of their work is the best approach. Big data, on the other hand, is something entirely new. In terms of both financial and operational resources, as well as human capital, the difficulties might be overwhelming. It is more cost-effective and more strategic to outsource the management of large amounts of data to a third party & then use their specialized big data solution to get an advantage over the competition.

VI. BIG DATA APPLICATIONS

Analytics and predictive approaches provided by big data analytics assist businesses and entrepreneurs make better- educated business decisions. Using big data analytics, the following are some of the possible outcomes:[10]:

Figure 6: Big data application in different fields

(i) Health care: An enormous amount of data is being created by electronic health records[17]. In a hospital or clinic, there are 3 different sorts of data generated: clinic data, patient information, and sensor data. There is a mix of quantitative and qualitative information in the vast majority of medical records. Structured and unstructured data are handled by the big data analytics approaches, which outperform the old methods.

Clinical issues may be answered with the use of big data. Analytical skills shown by big data may assist in putting customized medicine into practical practice. In this way, the system biology and health records may be integrated. There are numerous instances of how big data may be used to enhance healthcare, such as anti-cancer medicines, monitoring of patient vitals, improving hospital administration and promoting commercial development for health insurance firms, etc.

(ii) Educational data mining and learning analytics: During the epidemic, online[18]education has seen a significant increase in popularity on a global scale. Students' online actions generate a lot of untapped data. The utilization of big data methods in educational contexts is becoming increasingly prevalent. Big data learning analytics technology can be utilized for several objectives, including data visualization, intelligent feedback, course ideas, assessing student competence, & detecting student behavior patterns.

(iii) Process safety and risk management: The use of big data in process safety and risk management is becoming more commonplace in organizations[19].As a result of giving statistics, big data may aid in the examination of quality and risk management. Management can thus make swift decisions when they are needed.

(iv) Smart agriculture: Smart agricultural processes may benefit from the usage of big data[20]. Using new technology, external big data sources, like market data & weather data, may be integrated with farms in a way that contributes to the development of smart farms. Agriculture is being transformed by big data, which improves production, predicts yields, manages risk, and ensures food safety. A strong instrument like big data may be used in a variety of industries. Many additional industries, including government, social media analytics, and fraud detection as well as the contact center analytics and banking and marketing sectors are also using these technologies.

VII. LITERATURE REVIEW

Many types of information sets make up big data. Since the early days of computers, the notion of "big data" has permeated the fields of digital communication and information science. From mobile devices to contact centers to online servers, data is being gathered from all corners of the world every single day. The problem is that conventional databases & present technology can't manage data this huge, rapid, & hard to handle.

According to our study, the literature was chosen depending on its uniqueness & discussion of relevant big data- related themes.

Many firms collect massive volumes of data created by high-volume transactions such as call centers,

sensors, online logs, and digital photographs. Meeting big data difficulties while continuously enhancing operational efficiency is critical to the fulfillment of their enormous company. The majority of the researchers have contributed their research to properly managing vast amounts of data. The study of big data and its properties is examined at numerous levels, and so a full literature evaluation of various scholars' works is provided below:

This study[13], shows that Google Big Query, Alteryx Designer, and Pentaho Big Data Analytics are the best Big Data Analytical tools. Big Data Analytical techniques have also been compared in this research. The results of an industrial survey demonstrate that Pentaho Big Data Analytics is more extensively employed in the software sector than other Big Data Analytical solutions.

The present work[21] is part of an engineering design of surfaces project, which proposes a model of analysis and visualization of data using temperature and pressure sensors at plastic injection machines. The objective is to monitor the data collected from the manufacturing process of plastic molds. The result was achieved by developing a model for analyzing and visualizing data in real-time, with sensors that collect temperature and pressure data, and the possibility of viewing the application's data history. It is also possible to monitor the alerts that the model can trigger based on the collected data. Through the developed dashboards, production managers can have at their disposal a decision support system that aggregates information on productivity and the need for preventive maintenance of equipment.

This study[22]presents a novel ZSL approach for reducing the cost of training workloads for big data analytic workloads. We show that multi-user big data workloads can be effectively identified without explicitly training on multi-user workload sample instances by considering them as hybrids of simpler single-user workload classes. Utilizing the same classifier, we can accurately detect both unseen multi-user workloads & seen single-user workloads. They achieve an accuracy of 83% for unseen multi-user workload classifications & 92% for seen single-user workload classifications.

In this context, [23]an OLAP cube is often investigated using several aggregations that choose distinct subsets of cube dimensions to study patterns or uncover unexpected outcomes. Unfortunately, such analytic processes are often manual and do not provide statistically significant explanations for outcomes. Based on these concerns, they present in this study a unique OLAP-shaped visual big data analytics systemthat combines a cutting-edge statistical approach for assisting with the exploration and visualization of OLAP data cubes. An experimental study using a medical data set yields statistically meaningful findings as well as interactive visualizations that relate risk variables with illness severity.

This work,[24]describes an ML & large data analysis tool for processing and analyzing COVID-19 epidemiological data. To put it simply, the tool uses taxonomy and OLAP to generalize a few specialized features into several generalized attributes for successful big data analytics. For unknown or undeclared properties, the tool gives users the option of incorporating or removing them, based on their preferences and applications. Additional patterns and related patterns discovered by the program help disclose vital information such as the absolute and relative frequencies of certain patterns.

This research[25]provides a revolutionary big data strategy for the investigation, prevention, and rehabilitation of power system cascading. It is hoped that the methodologies created would be able to perform cascade analysis, better estimate the extent of system risk and provide possible solutions. This new big data approach's benefits in terms of accurate prediction and quicker and more effective corrective measures have been shown in experiments with IEEE 118- and 563-bus systems. Other power system applications might benefit from the newly established methodologies.

In this session, [26]describe the design of a cross-sector Big Data platform for process industries. The main goal was to create a scalable platform for data analysis that can gather, store, and interpret enormous amounts of data from many industries. Predictive functionalities for the manufacturing processes should thus be included in such a platform. It will provide a programming environment for data scientists to design these services and a stimulatingenvironment to test the models in the analytical platform. Several sites from various industries will use the platform. The exchange of information across diverse fields will be made possible via cross-sectoral collaboration. The architecture was put to the test in two different process industries: aluminum manufacturing and plastic molding.

VIII. CONCLUSION AND FUTURE WORK

Considering that the universe will have grown by a factor of 50 by the year 2022, we must learn how to effectively manage all of the data that will be generated. Big data is created from both internal and external sources, & traditional systems are unable to deal with it, necessitating high-performance, highly scalable systems with new approaches. As the finding of the vast volume of big data analytics papers, practitioners and academics have a tough time finding relevant subjects and keeping up with the latest developments. The goal of this paper is to offer an overview of the content, breadth, and conclusions of big data analytics, and the prospects that may be gained by employing big data analytics. Data growth necessitates a revision of the present

Suman Choudhary[1]* , Dr. Mahaveer Sain[2]

technology and tools, according to one study.

It's possible to discover a solution to the numerous problems that are arising on the way to Big Data Analytics in the future. Designing new methods and algorithms for the processing of large amounts of data may be done in the future. A wide range of innovative strategies may be used to enhance the collection of data from several sources. It's time to pay more attention to data security concerns that haven't been addressed in this study, but they're a very important topic. We haven't come up with a way to deal with the problems that Big Data Analytics faces. Researchers may go on with their future efforts to address the issues raised in this document. Analyzing real-time data rather than static historical data should be the focus of the strategies that are developed in the future.

REFERENCES

1. N. Elgendy and A. Elragal, "Big data analytics: A literature review paper," 2014, doi: 10.1007/978-3-319-08976-8_16.

2. S. Srimathy, "A Review on Big Data Analytics with Hadoop Technology," vol. 13, no. 10, pp. 10–13, 2017.

3. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," J. King Saud Univ. - Comput. Inf. Sci., vol. 30, no. 4, pp. 431–448, 2018, doi: https://doi.org/10.1016/j.jksuci.2017.06.001.

4. W. R. Kubick, "Big Data, Information, and Meaning," Appl. Clin. Trials, 2012.

5. D. Singh and C. K. Reddy, "A survey on platforms for big data analytics," J. Big Data, 2015, doi: 10.1186/s40537-014-0008-6.

6. Sas, "Big Data Analytics," sas.com, 2021. https://www.sas.com/en_in/insights/analytics/big-data- analytics.html.

7. Journal, B. Umadevi, D. Sundar, and B. Sakthi, "An Analytical Review on Big Data in a Diversified Approach," Int. Res. J. Eng. Technol., pp. 2548–2557, 2020, [Online]. Available: www.irjet.net.

8. P. Galetsi, K. Katsaliaki, and S. Kumar, "Big data analytics in health sector: Theoretical framework, techniques and prospects," International Journal of Information Management. 2020, doi: 10.1016/j.ijinfomgt.2019.05.003.

9. N. Khan et al., "Big data: Survey, technologies, opportunities, and challenges," Scientific World Journal. 2014, doi: 10.1155/2014/712826.

10. R. Rawat and R. Yadav, "Big Data: Big data analysis, issues and challenges and technologies," 2021, doi: 10.1088/1757-899X/1022/1/012014.

11. T. Shah, F. Rabhi, and P. Ray, "Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions," Cluster Comput., 2015, doi: 10.1007/s10586-014-0406-8.

12. Quora, "upcoming technologies in the area of big data," quora.com,2021. https://www.quora.com/Which-are-the-upcoming- technologies-in-the-area-of-big-data.

13. H. Singh, G. S. Matharu, A. K. Dardi, and J. S. Matharu, "Empirical investigation of big data analytical tools: Comparative analysis," 2019, doi: 10.1109/ICOEI.2019.8862739.

14. B. A. Dearmer, "CHALLENGES OF BIG DATA ANALYTICS," integrate.io, 2020. https://www.integrate.io/blog/big-data-problems- and-solutions/.

15. N. Khan, M. Alsaqer, H. Shah, G. Badsha, A. A. Abbasi, and S. Salehian, "The 10 Vs, issues and challenges of big data," 2018, doi: 10.1145/3206157.3206166.

16. I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," Information Systems. 2015, doi: 10.1016/j.is.2014.07.006.

17. T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," JAMA - Journal of the American Medical Association. 2013, doi: 10.1001/jama.2013.393.

18. Sin and L. Muthu, ""APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS – A LITERATURE REVIEW "," ICTACT J. Soft Comput., 2015, doi: 10.21917/ijsc.2015.0145.

19. R. H. Hariri, E. M. Fredericks, and K. M. Bowers, "Uncertainty in big data analytics: survey, opportunities, and challenges," J. Big Data, 2019, doi: 10.1186/s40537-019-0206-3.

20. S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big Data in Smart Farming – A review," Agricultural Systems. 2017, doi: 10.1016/j.agsy.2017.01.023.

21. G. Ferreira, P. Alves, and S. De Almeida, "Platform for real-time data analysis and

visualization based on Big Data methods," 2021, doi: 10.23919/CISTI52073.2021.9476628.

22. M. Genkin, "Zero-Shot Machine Learning Technique for Classification of Multi-User Big Data Workloads," 2020, doi: 10.1109/BigData50022.2020.9378023.

23. C. Ordonez, Z. Chen, A. Cuzzocrea, and J. Garcia-Garcia, "An Intelligent Visual Big Data Analytics Framework for Supporting Interactive Exploration and Visualization of Big OLAP Cubes," 2020, doi: 10.1109/IV51561.2020.00074.

24. C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang, and A. Cuzzocrea, "Machine Learning and OLAP on Big COVID-19 Data," in 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 5118–5127, doi: 10.1109/BigData50022.2020.9378407.

25. Y. Chen, T. Yin, R. Huang, X. Fan, and Q. Huang, "Big Data Analytic for Cascading Failure Analysis," in 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 1625–1630, doi: 10.1109/BigData47090.2019.9005593.

26. M. Sarnovsky, P. Bednar, and M. Smatana, "Big data processing and analytics platform architecture for process industry factories," Big Data Cogn. Comput., 2018, doi: 10.3390/bdcc2010003.

Corresponding Author

Suman Choudhary*

Research Scholar, Department of Computer Science & Informatics, Maharishi Arvind University, Jaipur, Rajasthan, India