iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33597

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

A Survey paper on Applications of Data Analytics

Mr. Raghavendra M. Ichangi
KSRDPR University Gadag

Abstract: Data Science is the buzzword of of the present scenario. The exponential growth in Internet of Things (IoTs) results in huge amount of data. Individuals, Businesses and Organizations are the major contributors for the data aggregated at the Clouds and Data Storage. Data alone is of no use unless we analyze it. Discovering the Knowledge, Patterns and observing the trend in the data is the need of the hour. The term Data Analytics is very much useful in processing the Raw Data, Eliminating the inconsistencies and producing the valuable information along with Knowledge or the hidden insights. The insights generated by the Data Analytics is of very much use in Business Applications, Recommendation Systems and many more.

Index Terms: Data Analytics, Internet of Things (IoT), Knowledge Discovery, Insights.

I. INTRODUCTION

The smartphone era has made the life so simple. People can read the Textbooks, Magazines and Newspapers anywhere anytime from the mobile screen. The content can be downloaded to local drive and can be replicated any number of times. The process of Uploading and Downloading the data has become the routine work. The Selfies, Snaps, Video Lectures, Short videos etc. is uploaded every second. The Facebook, WhatsApp, Twitter, Instagram and YouTube has billions of users who update their status, content every now and then. The tweets tweeted per minute across world, Facebook uploads, WhatsApp messages exchanged between people and groups, Videos uploaded to YouTube per hour etc. is hard to predict. It is not that important to store the information, what is important is to retrieve the information in time and to find the insights from the large data sets. Data Scientist and Data Analyst are the most demanding professions today's time. A recent study on jobs and hiring trends in Analytics and Data Science domains conducted by Great Learning, an EdTech company there are more than 93,500 data science jobs were vacant in India by the end of 2020. Presently, data is plays a very important role in Business Intelligence (BI). Data when collected from various sources will be in raw form and processed as per the requirement of a company and then this data is utilized at the Managerial level for the decision making purpose. This process will help the businesses to grow & expand their operations in the market. This entire process is termed as Data Analytics. Data Analyst and Data Scientist are the professions who perform this process. Data mining is a process which particularly focuses on statistical modeling and knowledge discovery for predictive analysis. Predictive analysis uses the data along with statistical algorithms and Machine learning algorithms to find the likelihood of the events based on the collected data over the years. This is usually referred as finding patterns or trend in data. s intelligence covers data analysis that relies heavily on aggregation, focusing mainly on business information. In statistical applications, data analysis can be divided into descriptive statistics, Exploratory Data Analysis (EDA), and Confirmatory Data Analysis (CDA). EDA focuses on discovering new features in the data while CDA focuses on confirming or falsifying existing hypotheses.

Predictive analytics focuses on application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. Learning analytics can transform models and pedagogical approaches. In online mode of teaching one can analyze the viewer's retention, responses and number of views and number of engagements of the particular post or the content to student`s success. The user retention analysis will be more useful in creating the content and identifying the target audience. Predictive analytics can also be used in Cybersecurity to identify the frauds and real time abnormalities with credit card transactions.

II. LITERATURE SURVEY

The Organizations, Institutes, Companies and Industries have drastically changed the mode of working. Most of them started going online or digital. With change in mode of operations to Online they need to keep track of Information about the Stakeholders, Customers and the clients. These profiles contain the complete information which constitute a large amount of Data. Dealing with a huge amount of raw data, which has information and knowledge in their hidden layer, it becomes very important to analyze it. The ways of collecting information like Online Survey, Feedback forms, Enquiry form and interviews generate data in various formats, size, variety, and velocity. These various formats bring complexity for Analysts to apply them in an efficient and effective way.

So, complexity in data analysis may not guarantee the efficiency in the end results hence it is advised to deploy advanced tools and techniques to overcome the difficulties of managing raw data. Big data analytics is the better and advanced method that has the capability for managing data of various volumes and formats.

Big Data uses advanced machine learning techniques and deep learning methods to derive insights from gathered data. With advancements in the Processors and the Devices, it has become a big challenge for traditional architectures and infrastructures to process large amounts of data within an acceptable time and resources. As per reports more than 500Hours of Fresh video content is uploaded to YouTube per minute. It becomes quite difficult to figure out the amount of data uploaded per Day, Per Week or Month. In order to identify the pattern or knowledge from these data, one has to find new tools and methods specialized for big data processing. For this reason, big data analytics has become so important nowadays.

Facebook, WhatsApp and other companies want to analyze the data to come up with hidden information and achieve competitive advantages in the market. Data is gathered from various sources which may be relevant or irrelevant. In some aspects it is meaningless until it turns into useful information and knowledge which can be helpful in decision making. For this purpose, we have several top big data software available in the market. Most of the Big Data Analytics Tools use Extract Transform and Load (ETL) which is a type of data integration tool for data from various sources.

1) *Extract:* It is the first and foremost process of reading data from large data sets which may be databases, data marts or warehouses. In this stage, the data is collected from multiple and different types of sources such as forms, data entry operators and other known sources.
2) *Transform:* The data which is stored in the Databases may be structured or unstructured, Normalized, unprocessed or of any format. The data need to be converted as per end user requirement. Transform is the process of converting the extracted data into the form it needs to be in so that it can be used for another database which is used as input for processing. The process of Transformation usually consists of Rules and Look-up tables which combines the data from one source with another. The transformation phase is very important as it makes the data readily available for further processing. The Big Data Analytics tools and techniques mainly rely on the Transformation of the data. Efficiently consistent the transformation will be the results.
3) *Load:* Load is one the important aspect that deals with loading the transformed data to the Analytical Engines. The Analysis is performed at the Analytical Engines to find the hidden insights, information or the knowledge which is very hard to find with bare eyes. Most of the Businesses use their own Analytical Engines for analytics purposes. The Analytical engines used online are usually referred as Online Analytical Processing(OLAP) Engines.

OLAP is a category of database processing that is used by business intelligence. The OLAP allows data analysts, Managers, and executives with the insights required for effective decision making. They can have OLAP to decide the organization's strategic directions. OLAP can provide valuable information about the performance of their business, employees. With the help of OLAP the Managers can also decide the ways for improvements and even business expansions.

OLAP tools are effectively used for two important aspects 1) Querying and 2) Reporting. OLAP tools let its users analyze data interactively in multiple dimensions from multiple perspectives. This feature is not available with Database Management Tools. The Data Analysts can easily find the trends and patterns from the data set. The visualization helps them to easily convey the hidden information with ease to the end users. When used with the sales data of the supermarkets OLAP can be used to identify the Association Rules among the items purchased, buying patterns of their regular customers. This is usually referred to as Market Basket Analysis. The popular OLAP applications include business reporting for super markets, management reporting, business process management, budgeting and forecasting, financial reporting, and more.

III. DATA ANALYTICS TOOLS

In this section, we describe the well-known analytical tools for Data Analytics which are widely used by Data Scientists, Analysts and Academicians. When the data is large and cannot be handled manually or by database management systems you can try analyzing it using different tools. How to find the best tool? Which tool is more suitable? Start by considering your business needs and end users. Will it be used by professional data analysts and data scientists, by nontechnical end users who need an intuitive interface, or should it suit general users? Most of the platforms are intended to provide interactive experience for coding and analysis using SQL while others are mainly focused on point and click analysis for beginners and nontechnical users. The business establishment and expansion are mainly depending on the market analysis for customer behavior, rivals and product demand. Big Data Analytics plays a very vital role in decision making for Business Intelligence (BI). Pricing and licensing Some offerings are free, while others charge licensing or subscription fees. The most expensive tools are not necessarily the most feature-complete, and users should not ignore the many robust free solutions available.

A. *R Programming*

R is the most popular free, open-source programming language with a lot of enhancements to it in the form of user written packages. R is the preferred programming language for statistical modeling, visualization, and data analysis. It is majorly used by statisticians for statistical analysis, Big Data and machine learning. R has a steep learning curve and needs some amount of working knowledge of coding. However, it is a great language when it comes to syntax and consistency. R is a winner when it comes to EDA (By definition - In statistics, Exploratory Data Analysis(EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods). Data manipulation in R is easy with packages such as plyr, dplyr, and tidy. R is excellent when it comes to data visualization and analysis with packages such as ggplot, lattice, ggvis, etc. R has a huge community of developers for support. R is used by Facebook - For behavior analysis related to status updates and profile pictures. Google - For advertising effectiveness and economic forecasting. Twitter - For data visualization and semantic clustering. Uber - For statistical analysis.

B. *Python Programming*

Python was initially designed as an Object-Oriented Programming language for software and web development and later enhanced for data science. Python is the fastest-growing programming languages today. It is a powerful Data Analysis tool and has a great set of friendly libraries for any aspect of scientific computing. Python is free, open-source software, and it is easy to learn. Python's data analysis library Pandas was built over NumPy, which is one of the earliest libraries in Python for data science. With Pandas, you can just do anything! You can perform advanced data manipulations and numeric analysis using data frames. Pandas support multiple file-formats; for example, you can import data from Excel spreadsheets to processing sets for time-series analysis. (By definition - Time-series analysis is a statistical technique that analyses time series data, i.e., data collected at a certain interval of time). Pandas is a powerful tool for data visualizing, data masking, merging, indexing and grouping data, data cleaning, and many more. To know more about Pandas, checkout Python Pandas Tutorials. Other libraries, such as Scipy, Scikit-learn, StatsModels, are used for statistical modeling, mathematical algorithms, machine learning, and data mining. Matplotlib, seaborn, and vispy are packages for data visualization and graphical analysis Python has an extensive developer community for support and is the most widely used language Top Companies that use Python for data analysis are Spotify, Netflix, NASA, Google and CERN and many more.

C. *Tableau*

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format. Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets. The data that is created using Tableau can be understood by professional at any level in an organization. It even allows a non-technical user to create a customized dashboard. The best feature Tableau Are Data Blending, Real time analysis, Collaboration of data. The great thing about Tableau software is that it doesn't require any technical or any kind of programming skills to operate. The tool has garnered interest among the people from all sectors such as business, researchers, different industries, etc.you integrate data from many sources into one.

1) *Cloud Storage Services:* Out-of-the-Box Data Transformations: Without coding anything, Xplenty offers out-of-the-box data transformations. These help Third-Party Integrations: Xplenty integrates with a variety of tools. Especially for logging, visualization, and data analytics.
2) *Rich Connectivity Library:* Xplenty connects with a variety of data sources. Including structured query language (SQL) data stores, NoSQL databases, and Scheduling: Run your data processes when you want with the ability to schedule.

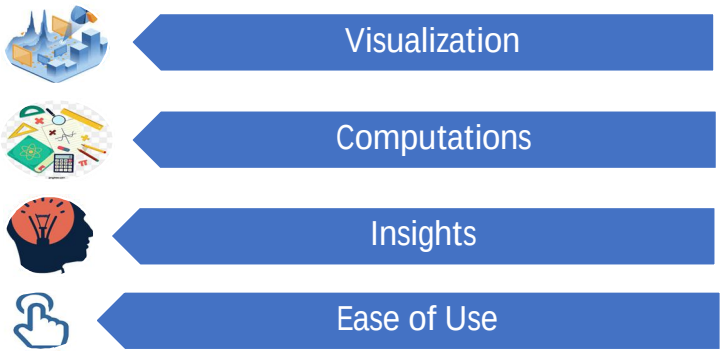


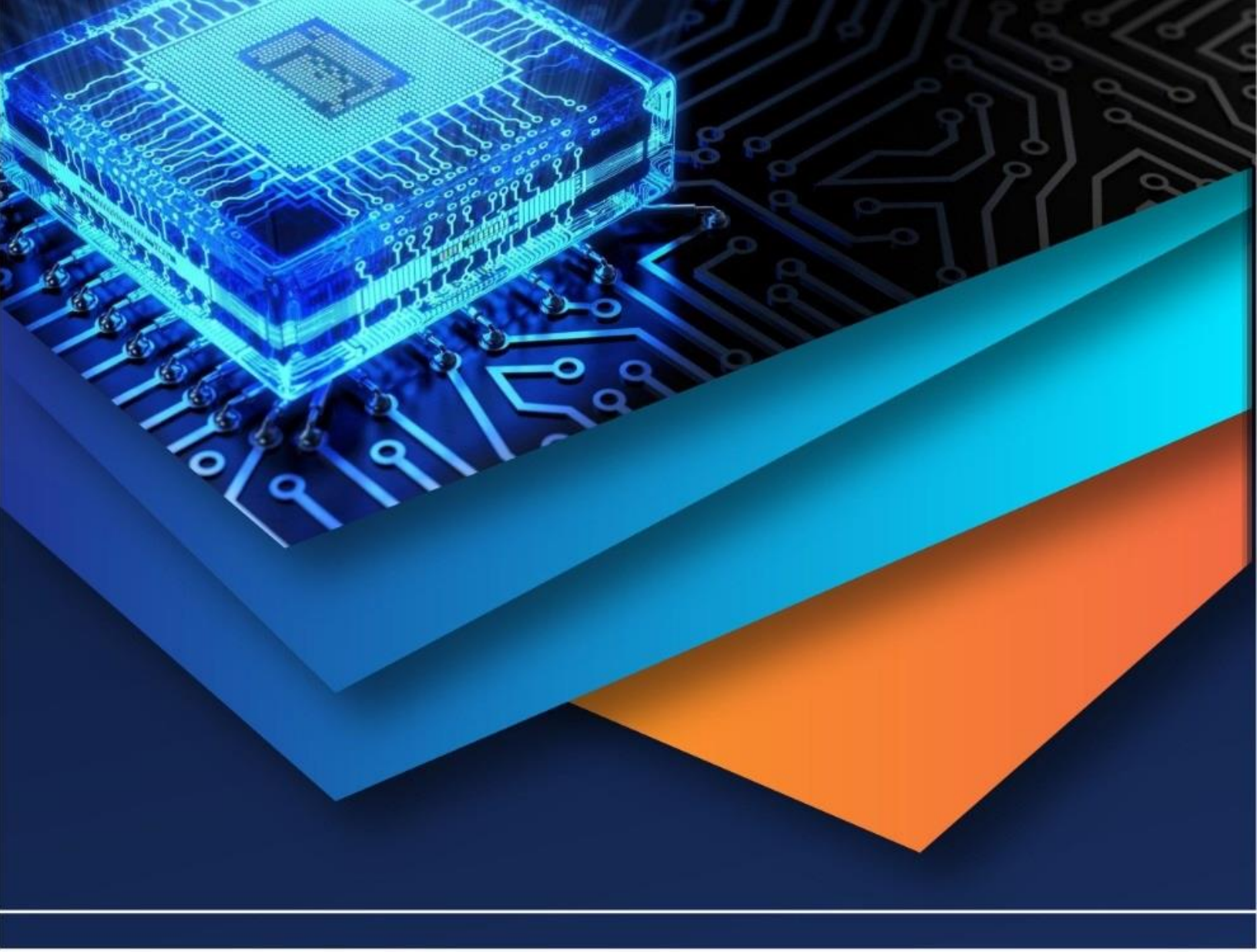
Figure 1: Advantages of Tableau

IV. CONCLUSION AND FUTURE WORK

This paper describes a cloud storage integrity verification techniques targeted at achieving integrity proof without retrieving whole data. To achieve public verifiability, we considered third party who can verify the proof received from cloud server. The proposed method encrypts and signs the state information to make the data unreadable by the server, since it contains information that would allow cloud server to generate fake proof of storage. The experimental analysis show that the method proposed is safe and provides integrity of data at cloud storage. Designing TPA to support multiple auditing tasks and verification technique that can accommodate dynamic data can be considered for future work.

REFERENCES

[1] Shuhui Jiang, Xueming Qian, Tao Mei, Yun Fu, Personalized Travel Sequence recommendation on Multisource Big Social Media, 2016, IEEE Transactions on Big Data,Vol.2, Issue:1
[2] Vallabh Dhoot, Shubham Gawande, Pooja Kanawade and Akanksha Lekhwani, Efficient Dimensionality Reduction for Big Data Using Clustering Technique, Imperial Journal of Interdisciplinary Research (IJIR), Vol-2, Issue-5, 2016, ISSN: 2454-1362
[3] Gantz J, Reinsel D, Extracting value from chaos.IDC iView, 2011, pp 1–12
[4] Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, Understandable Big Data: A survey, Mobile New Applications 2014, 171-209
[5] Mayer-Schonberger V, Cukier K, Big data: a revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt; 2013.
[6] Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Quart.2012; 36(4):1165–88.
[7] Kitchin R. The real-time city? Big data and smart urbanism. Geo J. 2014, 79(1), pp: 1–14.
[8] Katrina Sin and Loganathan Muthu, Applications of big data in education data mining and learning analytics – A literature Review, ICTACT Journal on soft computing special issue on soft computing models for big data, July 2015, Vol:05, Iss: 04, pp: 1035-1049
[9] Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, Understandable Big Data: A Survey, Computer Science Review, 2015, Vol: 17, pp: 71-80
[10] K. Krishnan, Data warehousing in the age of big data, in: The Morgan Kaufmann Series on Business Intelligence, Elsevier Science, 2013.
[11] H.V. Jagadish, D. Agarwal, P.Bernstein, Challenges and Opportunities in Big Data, The Community Research Association, 2015
[12] K. Davis, D. Patterson, “Ethics of Big Data: Balancing Risk and innovation”, O’Reilly Media, 2012.
[13] K. Krishnan, Data warehousing in the age of big data, in the Morgan Kaufmano series on Business Intelligence, Elsevier Science, 2013.
[14] Mike Barlow, Real-Time Big Data Analytics: Emerging Architecture, ISBN: 978-1-449-36421-2, 2013
[15] http://www.techrepublic.com/blog/big-data-analytics/10- emerging-technologies-for-big-data
[16] http://hortonworks.com/apache/hbase
[17] http://www.slideshare.net/infoDiagram/big-datacloudappsvisualiconpptinfodiagramtoolbox



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY