iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 9 **Issue:** XI **Month of publication:** November 2021

**DOI:** https://doi.org/10.22214/ijraset.2021.38888

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 XI Nov 2021- Available at www.ijraset.com*

# Analyzation and Comparison of Cloud Computing and Data Mining Techniques: Big Data and Impact of Blockchain

Adarsh Bhandari[1], Sanyog Bardia[2]
[1, 2]*Department of Computer Science and Engineering, Medi-Caps University, Indore*

***Abstract:*** ***With the rapid escalation of data driven solutions, companies are integrating huge data from multiple sources in order to gain fruitful results. To handle this tremendous volume of data we need cloud based architecture to store and manage this data. Cloud computing has emerged as a significant infrastructure that promises to reduce the need for maintaining costly computing facilities by organizations and scale up the products. Even today heavy applications are deployed on cloud and managed specially at AWS eliminating the need for error prone manual operations. This paper demonstrates about certain cloud computing tools and techniques present to handle big data and processes involved while extracting this data till model deployment and also distinction among their usage. It will also demonstrate, how big data analytics and cloud computing will change methods that will later drive the industry. Additionally, a study is presented later in the paper about management of blockchain generated big data on cloud and making analytical decision. Furthermore, the impact of blockchain in cloud computing and big data analytics has been employed in this paper.***

***Keywords: Cloud Computing, Big Data, Amazon Web Services (AWS), Google Cloud Platform (GCP), SaaS, PaaS, IaaS.***

## I. INTRODUCTION

### A. *Big Data Analytics for Enterprises: Challenges, Needs and Solution*

In the evolving 21st century there is tremendous volume, variety and velocity of data and in order to manage that there is need of strong cloud computing techniques in order to make data an asset. There is extensive bulk of data i.e. petabyte and Exabyte of data generated daily from online social media platform like Facebook, WhatsApp, Instagram and various other online enterprise applications. In order to manage that data companies, pay abundant amount for data warehousing therefore in order to reduce and substantiate the cost involved by these enterprises, are getting encumbered, resulting in consequential processing load. In order to eradicate this bottleneck, Big Data Analytics (BDA) comes into play, wherein several big data tools, techniques and methodologies are acquired by organizations for extracting the Right Data from these huge sets of unstructured data. Top companies like Amazon, Google, IBM and Microsoft has already set up their powerful cloud processing technology to handle the problem of big data management. One such tool that is being opted by enterprises for bringing offloading solution in warehouse data and processing functions is Apache Hadoop when amalgamated with various data warehouses, can turn out to very cost effective and highly business-like. Other frameworks being adopted by enterprises are Pig, Hive, Scala, R-programming, and many more.

### B. *Cloud Deployment Technology for Big Data:*

Cloud deployment technologies are used to place workloads into a cloud computing environment. They provide the foundation to facilitate execution of applications and services. These technologies range from bare metal to serverless computing and enable organizations to realize the benefits of cloud computing. The Definition of Cloud Deployment Technology Options section of this Practical Guide provides insight into the available technologies. Cloud deployment models indicate how the cloud services are made available to users. Cloud deployment is traditionally defined in terms of service models and deployment models. The “classic” definition of cloud service models comes from the National Institute of Standards and Technology (NIST) Special Publication 800-145 [2], which defines three service models:

• Infrastructure as a Service (IaaS) • Platform as a Service (PaaS) • Software as a Service (SaaS)

The four deployment models associated with cloud computing are as follows:

*1)* Public
*2)* Private
*3)* Hybrid
*4)* Community

©IJRASET: All Rights are Reserved

*C. Cloud based Big Data Architecture:*

As noisy data is generated from online platforms and enterprise applications. There are certain steps in order to manage this data and use this data for further examination by data analysts or data scientists. Moreover, cloud is now even used to handle traffic and live video streaming.

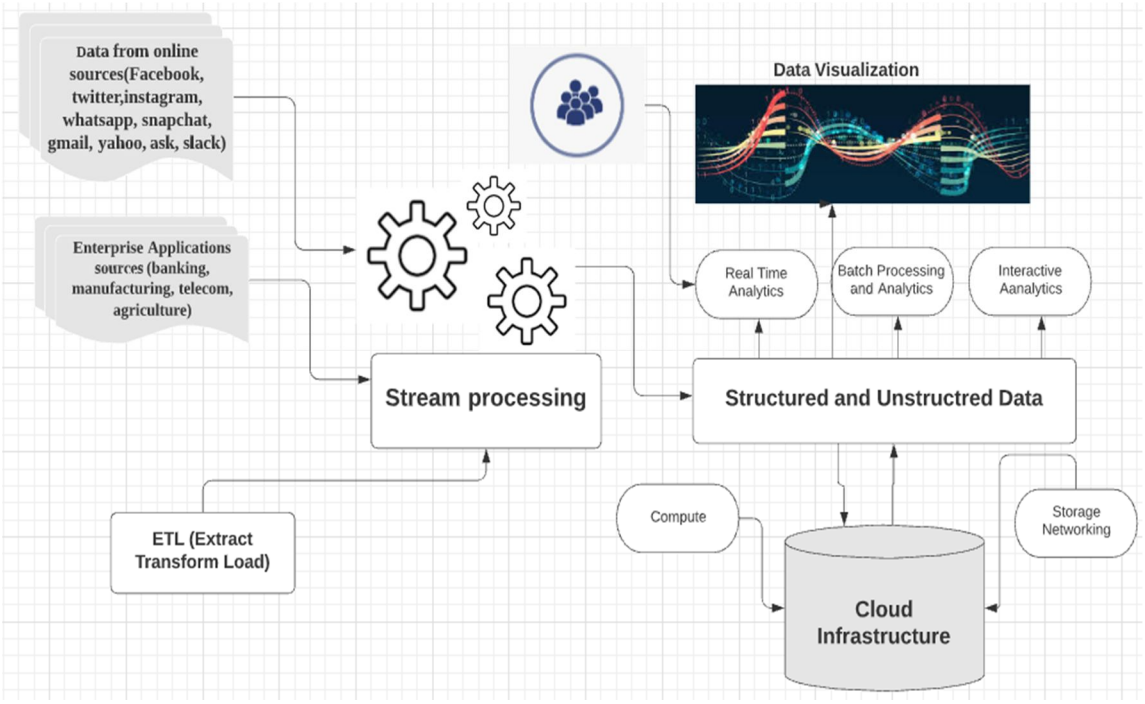



Fig 1. Architecture of cloud based big data architecture

*D. Big Data Mining*

Due to rapid research in big data field there are various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., used for knowledge discovery from databases and find insights from big data.

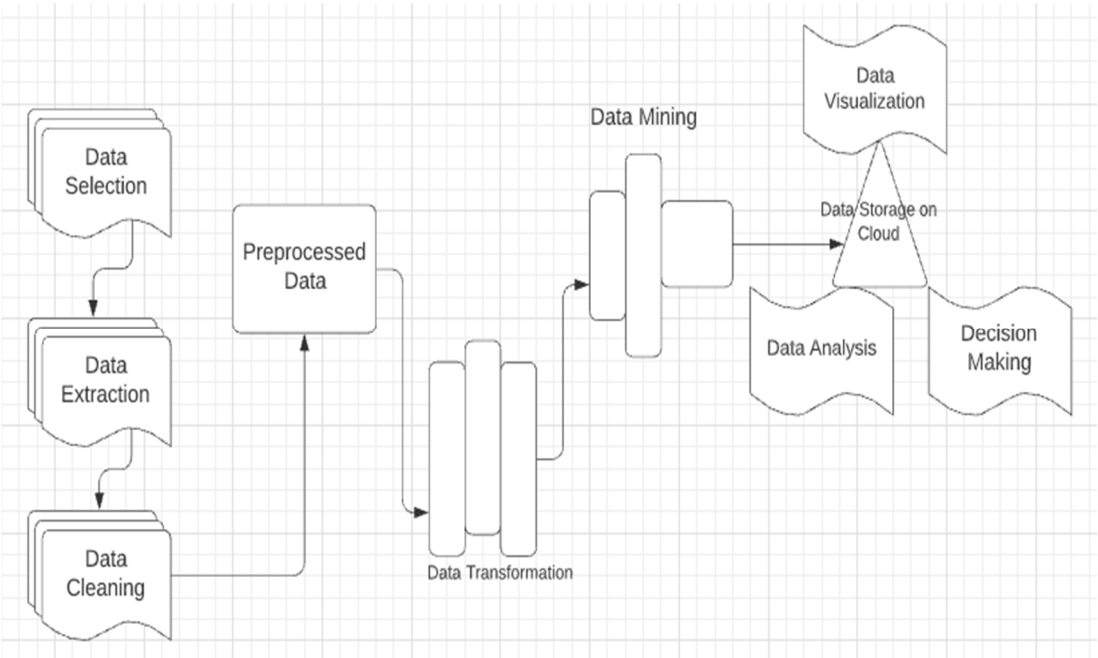



Fig 2. Data Mining Process

©IJRASET: All Rights are Reserved

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue XI Nov 2021- Available at www.ijraset.com*

## II. DIFFERENT CLOUD BASED BIG DATA TOOLS

1) *Amazon Web Services (AWS):* AWS is widely and most renowned cloud based tool for application deployment and handling big data. It deploys big data applications in easy and quick manner to scale up application software. There are certain AWS framework for handling big data:

a) *Amazon Elastic Search Service:* Amazon ES has benefits of real time application monitoring, log analytics, and clickstream analytics. It is easy to use, integrates well with other AWS services and support for Open-Source APIs and Tools.

b) *Amazon S3 and Lake Formation:* These are most popular big data storage framework with vast scalability, scalability, security and robust big data storage with a broad spectrum of engines. Customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as data lakes, websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics. Amazon Athena is used to query S3 data with standard SQL expressions.

c) *Amazon Redshift:* It is used analyze all types of data that is stored across AWS warehouses and S3 resources. It is used as delivering analytics as a service to perform business intelligence operations. It is relational database management based on PostgreSQL.

d) *EMR:* It easily run and scale Apache Spark, Hive, Presto and other big data workloads. Amazon EMR processes big data across a Hadoop cluster of virtual servers on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3). The Elastic in EMR's name refers to its dynamic resizing ability, which enables administrators to increase or reduce resources, depending on their current needs.

2) *IBM Cloud:* The IBM cloud offers diversified cloud based Big Data services, and delivers the accurate tool for the accurate job, ranging from Big Data Analytics Framework, networking, monitoring and so on. Some of the big data IBM Cloud techniques include:

a) *IBM Watson:* IBM Watson is AI for business the helps organization to predict future outcomes, automate complex processes and optimize employee's time. IBM Watson is built to apply advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering.

b) *IBM Cloud Object Storage:* IBM Cloud Object Storage supports exponential data growth and cloud-native workloads with built-in high-speed file transfer capabilities, cross-region offerings and integrated services. This cloud based storage tool is developed by IBM for the purpose of storing, managing and accessing data through REST based APIs and IBM's self-service portal.

c) *IBM DB2 Warehouse:* This is a highly organized, IBM CLU Acceleration driven business-class cloud based data warehouse service for incomparable query performance. This tools are highly-organized and fully secured, compatible with Oracle for the hybrid cloud, a data store for data sciences, and unburdens analytics workload. With built-in machine learning, automated scaling, built-in analytics, and SMP and MPP processing, Db2 Warehouse enables you to bring AI to your business faster and easier.

d) *IBM Analytics Engine:* IBM Analytics Engine provides an architecture for Hadoop clusters that decouples the compute and storage tiers. Instead of a permanent cluster formed of dual-purpose nodes, the Analytics Engine allows users to store data in an object storage layer such as IBM Cloud Object Storage and spins up clusters of compute notes when needed. Some of its features are open-source, scalable clusters, configurable environment.

3) *Microsoft Azure*: Azure is a public cloud computing platform with solutions including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and **Software as a Service** (SaaS) that can be used for services such as analytics, virtual computing, storage, networking, and much more. It can be used to replace or supplement you're on premise servers. Azure is helpful in automating patch management for virtual service, auto-scaling and on premise integration. Azure has following big data handling and analytical solutions:

a) *Azure HDInsight:* This Big Data Analytics solution offers 99.99% SLA for a single instance of virtual machine unlike other industrial services which offer SLA on the critical virtual machines only. For instance, it offers provision for optimized clusters creation for Spark, Hadoop, Kafka, HBase Storm, and Microsoft R Servers supported by 99.9% SLA.

©IJRASET: All Rights are Reserved

b) *Azure Blob Storage:* Blob Storage is an easy and cost-effective storage for exabytes of varied unstructured data like audio, videos, images and a lot more, in tiers classified as hot, cool and archive based on the frequency of data access. Its features are strong data integrity confirming availability of its latest version everywhere, flexibility to perform modifications for application enhancement and bandwidth usage reduction, various blob types like page, block and append blob providing flexibility for storage optimization, and automatic geo-replication for easy empowerment of improved local and global access.

c) *Azure SQL Warehouse:* It is one of the most used services in Microsoft Azure. It is basically SQL server in the cloud but fully managed and more intelligent. SQL Data Warehouse is a framework for massively-parallel processing with SQL analytics enabling elastic and independent scaling of compute and storage, with the capability of its effortless integration with big data stores for building a hub for cubes and data marts.

4) *Google Cloud Platform(GCP):* This is offered by Google which delivers an array of some powerful tools for diverse purposes ranging from big data analytics, data warehouse, to database and storage, and a lot more. Some of these standard tools, each for big data analytics, big data storage, and big data warehouse framework, are summarized below.

a) *Google Cloud Storage (Big Data Storage Framework):* Designed specifically for enterprises and developers, Google cloud storage is an incorporated object storage which performs numerous tasks, ranging from real-time data processing to data archiving (with Coldline and Nearline storage solutions) to data analytics. Some of the benefits of its usage are high availability and low pricing, refined archiving and storage, seamless and effortless data transition, cost-effectiveness, enhanced security for enterprise-critical resources, and partnership with leading vendor solutions.

b) *Google BigQuery (Big Data Warehousing Framework):* This is a highly scalable, swift, and low-priced fully-organized data warehouse for data analytics in production. Some of the advantages of using this product are quick infrastructure set-up, seamless scaling, effective analysis and quick insights, and business data and investments protection.

5) *Alibaba Cloud :* Alibaba cloud is also known as Aliyun and it offers cloud services that are available on a pay-as-you-go basis, and include Elastic Compute, Data Storage, Relational Databases, Big-Data Processing, Anti-DDoS protection and Content Delivery Networks (CDN). It is the largest cloud computing company in China, and in Asia Pacific according to Gartner.

a) *AnalyticDB:* AnalyticDB for MySQL is a real-time data warehousing that can process petabytes of data with high concurrency and low latency. It uses relational model to store data and provide SQL statements to compute and analyse data quickly.

After you get started, you can read the following topics to understand features of AnalyticDB for MySQL.

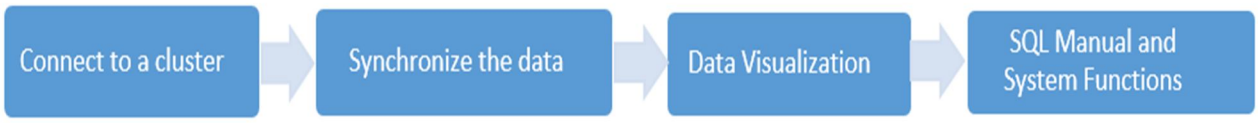



Fig 3. Working of AnalyticDB

b) Alibaba Block Storage: Block Storage is a block-level data storage services attached to Elastic Compute Service(ECS) instances. It has features like low latency, persistence and high reliability. It also enables automatic replication within the same zone to avoid data loss caused by hardware failures and guarantee the stability and continuity of your workloads. Alibaba Cloud provides a variety of EBS devices for ECS, such as cloud disks based on a distributed storage architecture and local disks located on the physical machines that host ECS instances like ESSD, SSD, Ultra Disk, etc.

c) Alibaba Max Compute: MaxCompute (formerly known as ODPS) is a fast and fully managed computing platform for large-scale data warehousing. It can process exabytes of data. MaxCompute is available in 16 countries and regions around the world, with customers in the finance, Internet, biomedical, energy, transportation, and media industries. MaxCompute provides large-scale data storage and computing services for global users. Several MaxCompute customer cases won the 2017 Big Data Outstanding Product and Application Solution Cases Award that was issued by the Ministry of Industry and Information Technology of China. In addition, Alibaba Cloud, represented by MaxCompute, DataWorks, and AnalyticDB, was positioned as a contender in Cloud Data Warehouse (CDW) in the Forrester Wave™: Cloud Data Warehouse, Q4 2018 report.

©IJRASET: All Rights are Reserved

6) *Oracle Cloud:* Oracle Cloud is the first public cloud built from the ground up to be a better cloud for every application. Oracle Cloud Infrastructure is built for enterprises seeking higher performance, lower costs, and easier cloud migration for their applications. Some of the services provided by oracle to be the competitor are: Oracle Big Data Cloud Service: Oracle big data services help data professionals manage, catalog, and process raw data. Oracle offers object storage and Hadoop-based data lakes for persistence, Spark for processing, and analysis through Oracle Cloud SQL or the customer's analytical tool of choice.

*a)* Autonomous Data Warehouse: Oracle Autonomous Data Warehouse is a cloud data warehouse service that eliminates all the complexities of operating a data warehouse, dw cloud, data warehouse center, securing data, and developing data-driven applications. It automates provisioning, configuring, securing, tuning, scaling, and backing up of the data warehouse. It includes tools for self-service data loading, data transformations, business models, automatic insights, and built-in converged database capabilities that enable simpler queries across multiple data types and machine learning analysis.

*b)* Oracle Cloud Advertising

*c)* Oracle Big Data Service

## III. COMPARISON AMONG DIFFERENT CLOUD BASED BIG DATA ENTERPRISE SOLUTIONS FRAMEWORKS

1) *Big Data Analytics:* For big data analytics AWS uses Amazon Elastic Search Service, IBM- Analytics Engine, MS Azure-Azure HDInsight, Alibaba-Data Lake Analytics and GCP- Google Cloud Dataproc

| Framework types and feature | IBM Cloud | MS Azure | GCP | Alibaba | AWS |
|---|---|---|---|---|---|
| Mode of Software | Open - Source | Open – Source | Open – Source | Open-Source | Open - Source |
| Types of Data | Unstructured | Unstructured | Structured, semi-structured and unstructured | Structured, semi-structured and unstructured | Structured,semi-structured and unstructured |
| Data Sources | IBM Cloud Object Storage | Blob Storage | GoogleBigtable, Google Cloud Storage, and Google BigQuery | Object Storage Service | Amazon S3, Amazon Kinesis Firehose, and Amazon DynamoDB |
| Supported Operating System | CentOS, Ubuntu, Linux, RHEL | Ubuntu , Windows Server 2012, Windows 10 | Debian 8, AWS Linux, RHEL | Alibaba Cloud Linux, CentOS, CoreOS, Debian, Ubuntu | CentOS, Ubuntu, and Amazon Linux, RHEL |
| Applications | Data analytics, enterprise solution for various Big data problems, and analytics applications development and deployment | Stream and Batch data analytics | Batch processing, querying, streaming, and machine learning | Data analytics, enterprise solution for various Big data problems, and analytics applications development and deployment | Logs analytics, real-time applications monitoring, and clickstream analytics |
| Deployment | Regional | Regional | Zonal | Regional | Zonal |
| Deployment Unit | Cluster | Cluster | Cluster | Cluster | Cluster(EC2) |
| Compute Scaling | Manual | Manual | Manual | Manual | Manual |
| Compute Node Billing Model | Pay– Per -Hour | Pay-Per-Minute | Pay-Per-Second | Pay-Per-Byte | Pay– Per -Hour |

Table 3.1 Comparison based on big data analytics

©IJRASET: All Rights are Reserved

2) *Big Data Storage:* For big data storage AWS uses Amazon S3, IBM- Cloud Object Storage, MS Azure- Azure Blob Storage, Alibaba Cloud- Block Storage and GCP- Google Cloud Storage.

| Framework types and feature | IBM Cloud | MS Azure | GCP | Alibaba | AWS |
|---|---|---|---|---|---|
| Storage Type | Distributed Object Storage | Distributed Object Storage | Distributed Object Storage | Distributed Object Storage | Distributed Object Storage |
| Object Types | Object | Append blobs, block blobs | Object | Object | Object |
| Deployment Unit | Bucket | Container | Bucket | Bucket | Bucket |
| Deployment Identifier | Universally Unique Identifier | Account-level Unique Key | Globally Unique Key | Globally Unique Key | Globally Unique Key |
| Storage Scaling | Yes, Manual | Yes, Manual | Yes, Automatic | Yes, Automatic | Yes, Automatic |
| Deployment Locality | Regional and multi-regional | Zonal and Regional | Regional and multi-regional | Regional | Regional |
| Object Metadata | Yes | Yes | Yes | Yes | Yes |
| Data Store Format | Any | XML | CSV, JSON (Newline Delimited Only), Google Cloud Datastore Backups, Avro | Any | Any |
| Data Store Size Limit | Unlimited | Limited | Limited | Unlimited | Unlimited |
| Object Lifecycle Management | Yes | No | Yes | Yes | Yes |
| Storage Classes | Accessibility: Standard, Vault, and Cold Vault, Flex Class (Dynamic Data) | Tiers Classes: Hot, Cool Redundancy Classes: LRS, GRS, ZRS, RA-GRS | Nearline, Coldline, Regional, Multi-regional | Standard, Infrequent Access, Archive, Cold Archive | Standard, Infrequent Access, Amazon Glacier |
| Object Versioning | Yes, Manual | Yes, Manual | Yes, Automatic | Yes, Manual | Yes, Automatic |

Table 3.2 Comparison based on Big Data Storage

©IJRASET: All Rights are Reserved

3) *Big Data Warehousing:* For big data warehousing AWS uses Amazon Redshift, IBM- Analytics Engine, MS Azure- Azure HDInsight, Alibaba- MaxCompute and GCP- Google Cloud Dataproc

| Framework types and feature | IBM Cloud | MS Azure | GCP | Alibaba | AWS |
|---|---|---|---|---|---|
| Deployment Unit | Cluster | Massively Parallel Processor (MPP) | Fully Managed | Cluster | Cluster (EC2) |
| Compute Scaling | Yes, Automatic | Yes,Manual | Yes, Automatic | Yes, Automatic | Yes, Automatic |
| Permissible range of Compute Nodes | 1-60 (1 head node, 59 data nodes) | 1-60 (excluding1 control node for MPP Engine) | NA, Fully Managed | 1-128 | 1-128 |
| Deployment Locality | Region | Region | Region | Availability zone, Region | Availability zone, Region |
| Supported Data Format | Delimited format such as CSV | Parquet, Orc, flat delimited text, RC | CSV, Avro, JSON (Newline delimited only) | CSV, Avro, JSON | CSV, Avro, Parquet, SequenceFile, TSV, Grok, RCFile, TSV, ORC and RegexSerDe |
| Storage Format | Columnar | Columnar | Columnar | Columnar | Columnar |
| Data Sources | Data file on network, data stores like Amazon S3 or IBM Cloud Object Storage, and Db2® server | Azure Blob Storage | Google Cloud Storage, Google Cloud Dataflow, readable data sources | Object Storage Service | Amazon S3, Amazon Dynamo DB, Amazon EMR, AWS Data Pipeline |
| Data Loading Methods | Load from Cloud, Load from File, Load Geospatial Data, Load Twitter Data, Load Public Data | Data Load via PolyBase, SQLBulkCopy API, Bulk-Load data with SSIS (SQL Server Integration Service) and BCP Command, Azure Data Factory | Streaming Upload, Bulk Upload, Google Analytics PremiumData | Load from Cloud, Load from File, Load Geospatial Data, Load Twitter Data, Load Public Data | COPY from S3, Streams from Amazon Kinesis Firhose |
| Query Language | SQL Reference, CLPPlus, SQL PL, PL/SQL | PolyBase T-SQL | Standard SQL (Beta), Legacy BigQuery SQL | SQL, PostgreSQL, PL/SQL | PostgreSQL |
| Integration with ETL and BI Tools | Yes | Yes | Yes | Yes | Yes |
| Backup Retention Policy | 2 Days | 7 days | 7 days | 10 days | 1-35 days |

Table 3.3 Comparison based on Big Data Warehousing

©IJRASET: All Rights are Reserved

## IV. IMPACT OF BLOCKCHAIN ON CLOUD COMPUTING AND BIG DATA ANALYTICS

With the rapid evolution of virtual currency on numerous platforms there is expansion of data generated from these sources and soon it will be billion-dollar revenue generating industry. Big Data is one of the fastest growing sectors in the world. Every business wants to get insights into usage patterns of their consumers. Massive datasets are thereby analyzed using advanced statistical models and data mining. These Big Data sets will become even more prevalent over the coming years. While blockchain is focused on recording validating data (data integrity), data science analyses data for actionable insight, making predictions from large amounts of data (prediction). While blockchain is changing data management, the latter is transforming the nature of transactions. Or said in another way: "If Big Data is the quantity, blockchain is the quality". Securing and interpreting such large amounts of information is not an easy task. Blockchain could be seen as an ideal solution to address many of the challenges of Big Data management and analytics. Blockchain will make Big Data more valuable because it ensures data quality, accessibility and security.

### *A. Growth of big data with blockchain and uses:*

More quality data with more insights means more value. This enables better management of huge volumes and variety of information that keeps flowing in for businesses. Blockchain developers are building decentralized data marketplaces that are now starting to emerge. These marketplaces are platforms that use the peer-to-peer connectivity possible through blockchain technology to link data sellers with data buyers. Also there would be less fraud as now transactions would be much more transparent and secure. Supply chain is a notable use case where Blockchain can be leveraged to manage and sign contracts and audit product provenance. It could also be leveraged for votation platforms, titles and deed management - amongst myriad other uses. As the digital and physical worlds converge, the practical applications of Blockchain will only grow. With this inclined growth the need for cloud computing will rise as more data will be managed to gain analytical insights.

Five main features that cloud computing would help to manage blockchain:
1. Immutable 2. Distributed 3. Robust 4. Robust 5. Secure

With this digitization the number of cryptocurrency has increased with approx 10000 new bitcoins are mined per day. The rate and value too has seen a huge monetary gaps with exponential increase in last few years. The blockchain has also unexpectedly improved big data integrity. The banks and financial organization can now leverage the power of cloud and blockchain to perform real time transaction processing and detect fraudulent transactions retrospectively. The rate at which cryptocurrency has increased by more than 1000% in last 5 years is too fascinating with clearly predicting how plethora of data is generated from these and organize this on cloud. This significant jump has paved the way for new analytical techniques in order to gain insights from data and even using power of machine learning to make future decisions. The graphical information from 2013 to 2021 is clearly a symbol that this would be the field in business apart from stock market.

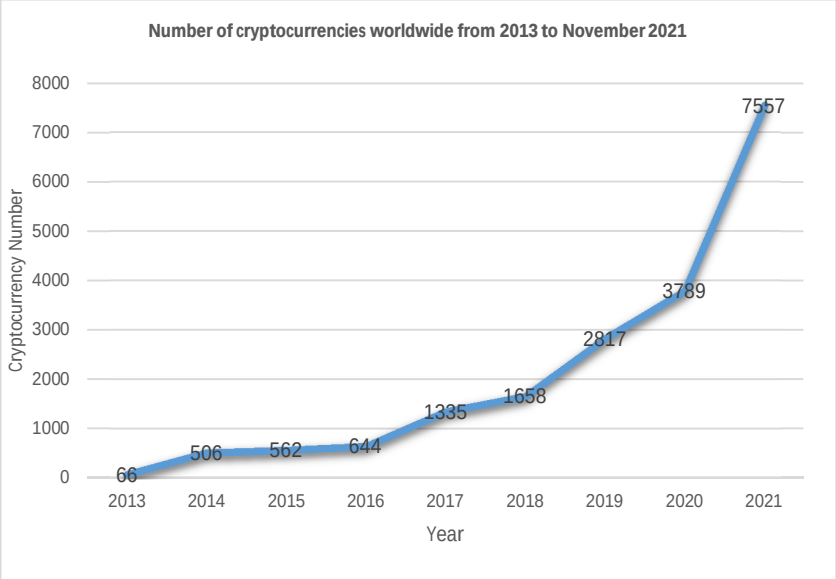



Fig.4 Increase in cryptocurrency from 2013-2021

©IJRASET: All Rights are Reserved

*B. Data Analytics on Blockchain (Cryptocurrency)*

The use of Big Data Analytics in cryptocurrency trading has proven beneficial for everyone. Further, Big Data Analytics has become a guided network for several crypto traders around the globe but has also enabled large-scale organizations to invest in the crypto space.

The data from bitcoin can be analyzed by designing a database scheme in order to create a central repository for data storage and gathering. The schema is structured in such a way to load the data into i.e. relational or non-relational database. Further after the database design the data can be stored on cloud for real-time data processing and used for further visualization and decision making. While multiple tools are available for visualization there are certain tools which can connect directly with cloud. The various such analytical tools are:

1) *Tableau:* Tableau is an AWS partner which provides SaaS platform. The tableau server standlone (single-node) deployment installs Tableau Server on an Amazon Elastic Compute Cloud (Amazon EC2). Tableau integrates with AWS services to empower enterprises to maximize the return on your organization's data and to leverage their existing technology investments. It all starts with direct connections to Amazon data sources including Amazon Redshift (including Redshift Spectrum), Amazon Aurora, Amazon Athena and Amazon EMR.
2) *Power BI:* Amazon redshift is connected to Power BI for insights and analytics. Adobe Analytics is connected with Power BI for web analytics and reporting. We can associate Power BI with Azure SQL Database, Azure SQL Data Warehouse, Azure Analysis Service Database, Azure HDInsights, and so on.
3) *Python:* It provides API and other connectivity drivers to connect to particular cloud. Through python we can also deploy large scale applications and host them on cloud. It has analytical and visualization libraries like matplotlib, seaborn, ggplot, bokeh, etc. to perform big data analysis.

In order to gain insights from cryptocurrency data there are certain steps in order to query the data and perform preprocessing:

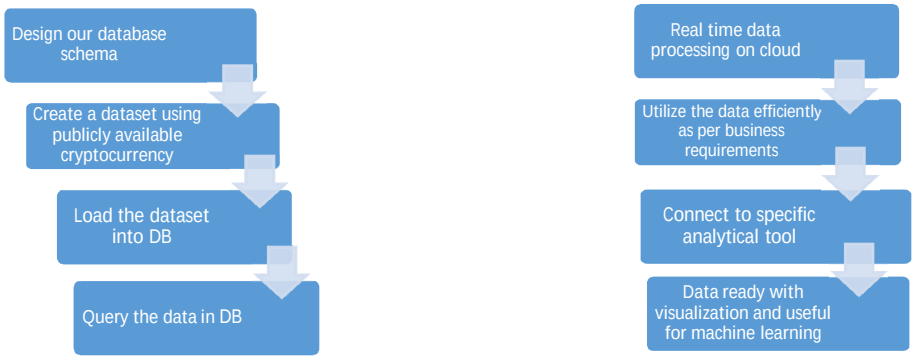



Fig 4. Data processing and steps for cryptocurrency visualization

## V. CONCLUSION

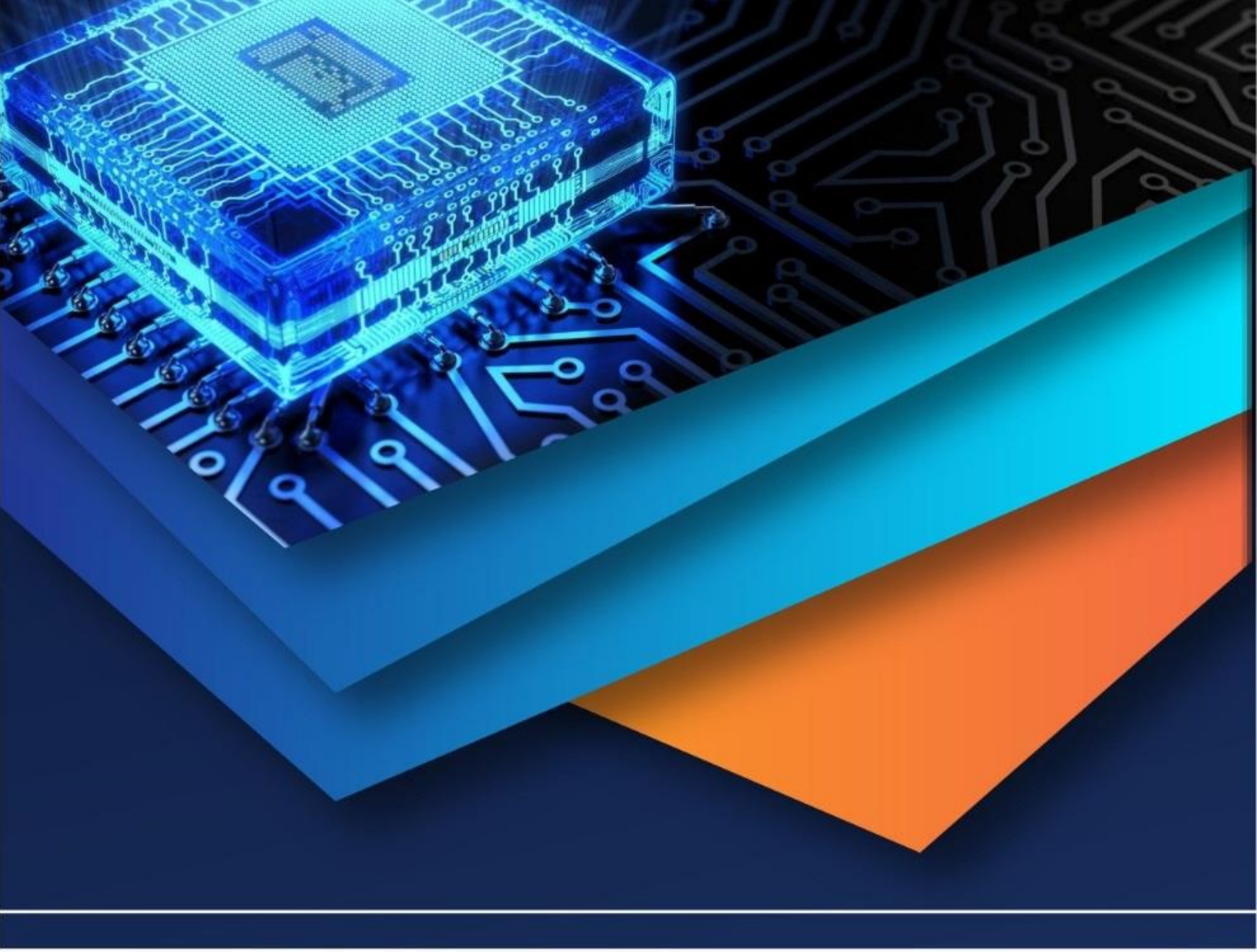
With the fast moving dynamic internet world the blockchain market is estimated to accumulate 20 billion dollars in revenue by the year 2024. The cloud computing is soon going to a key area to manage blockchain and simultaneously blockchain would help to perform big data analytics. The paper demonstrated specific knowledge and comparison among multiple cloud computing tools to be used for big data analytics. There is significant description about cloud tools like AWS, GCP, IBM, Alibaba etc. and how they can flourish even more with this tremendous growth of big data. There are distinctions among services provided by AWS, IBM, GCP, Alibaba and oracle and how they are managed and supported on multiple platforms. With globalization, in the foreseeable future, the amount of data generated will be beyond our realm of imagination. Although Big data technologies are prevailing in use, seeing the current trend of velocity of data gathered, this technology may soon start to discern issues with memory and speed, but again it's shrewd. The world must prepare itself for a more advanced future teemed with innovations in cloud computing and automation of data processing. Big data along with Machine Learning to make decision making more in-depth to find trends within the data. Moreover, different types of big data analytics are taken into account with real world examples to mage ingenious business decisions. Lastly, research is carried out by deemed companies like Google, Meta, Amazon, etc. to understand real world use of big data analytics in business and integrate it well with emerging blockchain technology to solve the monetary problems and generate new sources of income and payments which can all happen on cloud on a single go with enhanced security.

©IJRASET: All Rights are Reserved

## REFERENCES

[1] Jiawei Han and MichelineKamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed. 2. Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM
[2] B. P. Rao, P. Saluia, N. Sharma, A. Mittal, S. V. Sharma, Cloud computing.
[3] Rajesh Math. "Big Data Analytics: Recent and Emerging Application in Services Industry. Part of the Advances in Intelligent Systems and Computing book series (AISC, volume 654)" SpringerDoi: 978-981-10-6620-7_21
[4] Nicolas Serrano, Gorka Gallardo, JosuneHermantes."Infrastructure as a Service and Cloud Technologies"IEEEDOI: 10.1109/MS.2015.43
[5] K.S. Sangeetha, P.Prakash. "Big Data and Cloud: A Survey"SpringerDoi: 978-81-322-2135-7_81
[6] Samiya Khan, KashishAraShakil, MansafAlam."Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions"SpringerDoi: 978-981-10-6620-7_57
[7] Rahul Vora, KunjalGarala, PriyankaRaval."An Era of Big Data on Cloud Computing as Utility: 360° of Review, Challenges and Unsolved Exploration Problems"SpringerDoi: 978-3-319-30927-9_57
[8] Jabbari and P. Kaminsky, "Blockchain and supply chain management", 2018.
[9] M. K. R. Ingole and M. S. Yamde, "Blockchain technology in cloud computing: A systematic review", 2018.
[10] Deepa, N., Pham, Q.-V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., Maddikunta, P. K. R., Fang, F., & Pathirana, P. N. (2021). A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions.
[11] Towards Data Science and Vladimir Fedak. 2018. Blockchain and Big Data: the match made in heavens. https://towardsdatascience.com/blockchain-and-big-data-the-match-made-in-heavens-337887a0ce73
[12] Buyyaa R, Yeoa CS, Venugopala S, Broberg J, Brandic I. Cloud Computing and Emerging IT platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. Future Generation Computer Systems. 2009; 25: 599-616p
[13] IBM. "IBM Cloud and Solutions" https://www.ibm.com/in-en/cloud
[14] Google Cloud Platform. "Google Cloud: Cloud Computing Services" https://cloud.google.com/
[15] AWS "Amazon Web Services" https://aws.amazon.com/solutions/?nc2=h_ql_sol
[16] Alibaba. "Alibaba Cloud" https://in.alibabacloud.com/?utm_key=se_1007710386&utm_content=se_1007710386&gclid=CjwKCAiAm7OMBhAQEiwArvGi3MCrtj-YIfQ2FTI3UTTTyoUkD_YpVrMQawk8hCcatFZYWEcZdO3UPBoCoQUQAvD_BwE
[17] Azure "Microsoft Azure: Cloud Computing Services" https://azure.microsoft.com/en-in/

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)