ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

Review On Big Data, Cloud Computing and Data Science Analytics : Issues and Challenges

[1]Namrata Gadgil, [2]Vanita Patil
Assistant Professor at
PES's Modern College Of ACS, Business Administration campus, Shivajinager Pune 05

Abstract:
information technology and various sectors produces big amounts of data. This data requires processing and storage. Big data processing embody a new challenge in computing, especially in cloud computing. Data processing involves data acquisition, storage and analysis. Big data, Cloud Computing and Data Science are currently trending in organizations across the world. And many questions arrives, including, what is the relationship between big data, cloud computing and data Science ? how is big data processed in cloud computing? In this paper, we find out the answers of these questions, the big data and cloud computing will be studied, also to getting the relationship between them in terms of safety and challenges. This paper discusses the conjunction of big data, cloud and data science. It also identifies various problems in Big Data, Cloud, Data Science and their convergence.

Keywords: *Big Data, Cloud, Data science*

Introduction:
Big information could be a radiance topic with ampleness of scope for analysis. With the arrival of social media, the information has began to cross the bounds of a system, server and even an information center. On the opposite hand, Cloud Computing is another space within the IT field wherever totally different services like package, Infrastructure, storage etc. area unit offered as services on-line. the most advantage here is that tiny business corporations needn't accumulate information storage as a physically existing 'thing'. At this juncture, it's to be realised that huge information in Cloud isn't just for storage however conjointly for analyzing. the large amounts of information generated by devices and Internet-based sources at a daily basis forms, big data. This information may be processed and analyzed to develop helpful info and applications for explicit domains. several mathematical and information analytics techniques have found use during this sphere. This has given rise to the event of computing models and tools for giant information computing. However, the storage and process needs area unit overwhelming for ancient systems and technologies. Therefore, there's a requirement for infrastructures that may regulate the storage and process capability in accordance with the dynamic information dimensions. Cloud Computing offers a possible resolution to the current drawback. Though , huge information computing within the cloud has its own set of challenges and analysis problems.

Big data:
The word 'BIG DATA' was 1st mentioned by michael Cox and David Ellsworth. massive data may be a data analysis methodology enabled by recent advances in technologies and design. massive information refers to technologies and techniques that involve data that's large, heterogeneous and fast-changing for standard technologies, skills and infra-structure to handle with efficiency. Big data is usually outlined by 5 V's (Variety, Velocity, Volume, Veracity, and Visibility). massive information square measure classified into completely totally different categories to higher understand their characteristics. Fig. 1.1. shows the numerous categories of big knowledge. The classification is extremely vital as a results of large-scale information at intervals the cloud. The classification depends on five aspects: (i) knowledge sources, (ii) content format, (iii) knowledge stores, (iv) knowledge staging, and (v) processing.

ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

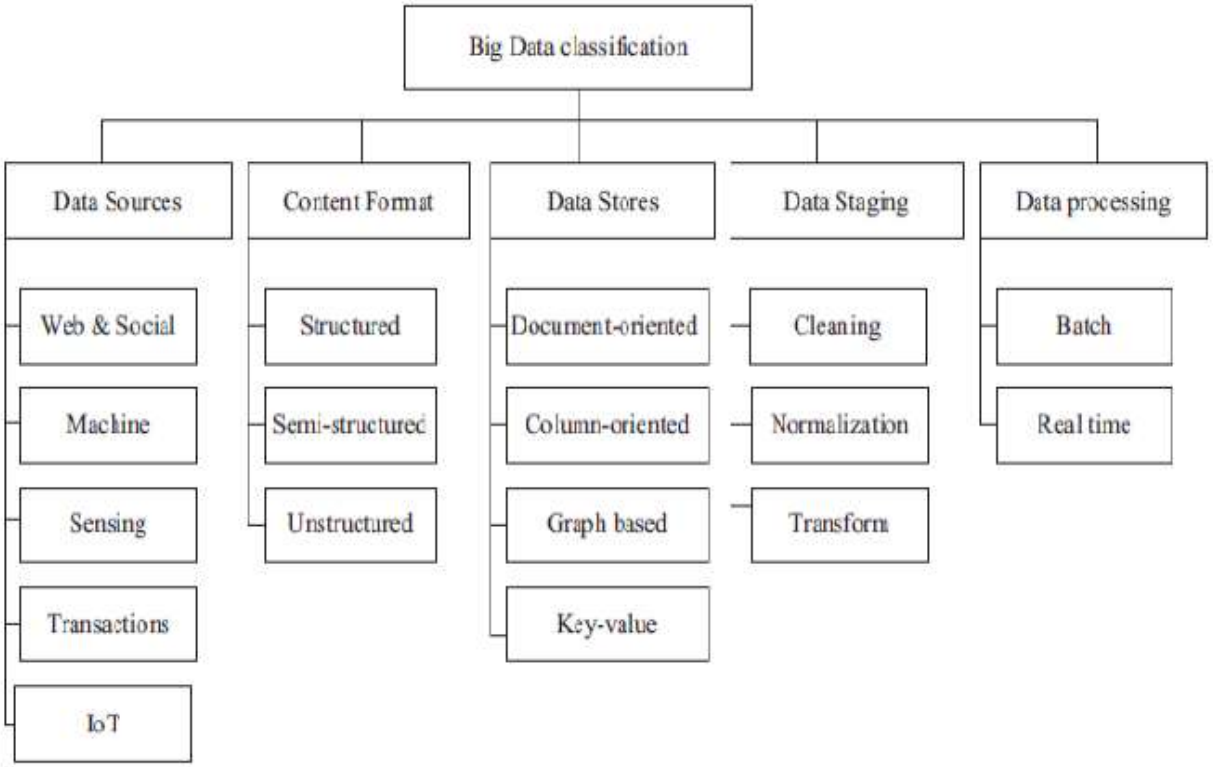

Figure: Big Data Classification

Big data refers to managing, analyzing and capturing different data sets where size, complexity and rate of growth varies for each of them. The benefits that can be accomplished with big data analytics are products can be redeveloped effectively, maintenance costs can be reduced, offering deeper insight from enterprise perspective, customizing websites in real time, creating new revenue streams and analysis of risks can be performed effectively.

Cloud:

The 3 kinds of cloud computing ar the general public cloud, the private cloud, and the hybrid cloud. A public cloud is that the pay- as-you-go services. a private cloud is internal information center of a business not available to the general public however supported cloud structure. The hybrid cloud may be a combination of the general public cloud and private cloud. Cloud Computing may be a paradigm that provides dynamically climbable and virtualized resource as a service over the net. the necessity to store, process, and analyze massive amounts of data is making enterprise customers adopt cloud computing at scale. Cloud allows users to perform advanced analytics with big data.

Cloud Computing Service Models:

Common models for cloud computing include platform as a service (PaaS), software as a service (SaaS), infrastructure as a service (IaaS), and hardware as a service (HaaS). Cloud computing is also an emerging research area where different services can be provided to the users on demand. The different services that can be provided are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). In SaaS, software is deployed over Internet is delivered as on demand service and its main characteristics are easy access to commercial software, no handling of software upgrades and patches and provides an API for integration between different pieces of software. In PaaS, platform for creation of software over the web is delivered as on demand service and its main characteristics are integration with

web services and databases via the common standards, providing web based user interface (UI) tools that help to create, modify, test and deploy different UI scenarios, providing support for multiple concurrent users utilizing same development application and support for development team collaboration. In IaaS servers, network, storage and operating systems are delivered as ondem and service and its main characteristics are including multiple users on single piece of hardware, resources are distributed as a service and providing support for dynamic scaling.

Data Sciences:
Data Science may be a field that comprises of everything that associated with data cleansing, preparation, and analysis. it's the umbrella of techniques used once making an attempt to extract insights {and information|and knowledge|and data} from data. huge data Analytics the science of examining huge data with the aim of drawing conclusions and inferences. it's a set of data science. huge data analytics is impossible while not cloud within the current situation.

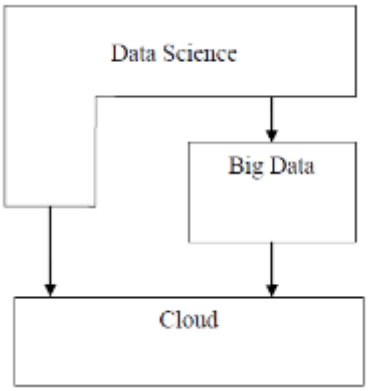

Fig. : Relation between Big Data, Cloud and Data Science

Relation between Big data, cloud and data Analytics :
Three major reasons for tiny to medium sized businesses to use cloud computing for large data technology implementation are hardware value reduction, process value reduction, and ability to check the worth of big data. the foremost considerations relating to cloud computing are security and loss of control. The term big data is derived from the very fact that the datasets are therefore massive that typical database systems aren't able to store and analyze the datasets. The datasets are massive as a result of the info isn't any longer traditional structured data, however information from several new sources, together with e-mail, social media, and Internet-accessible sensors. The characteristics of huge information gift information storage and information analysis challenges to businesses.

ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

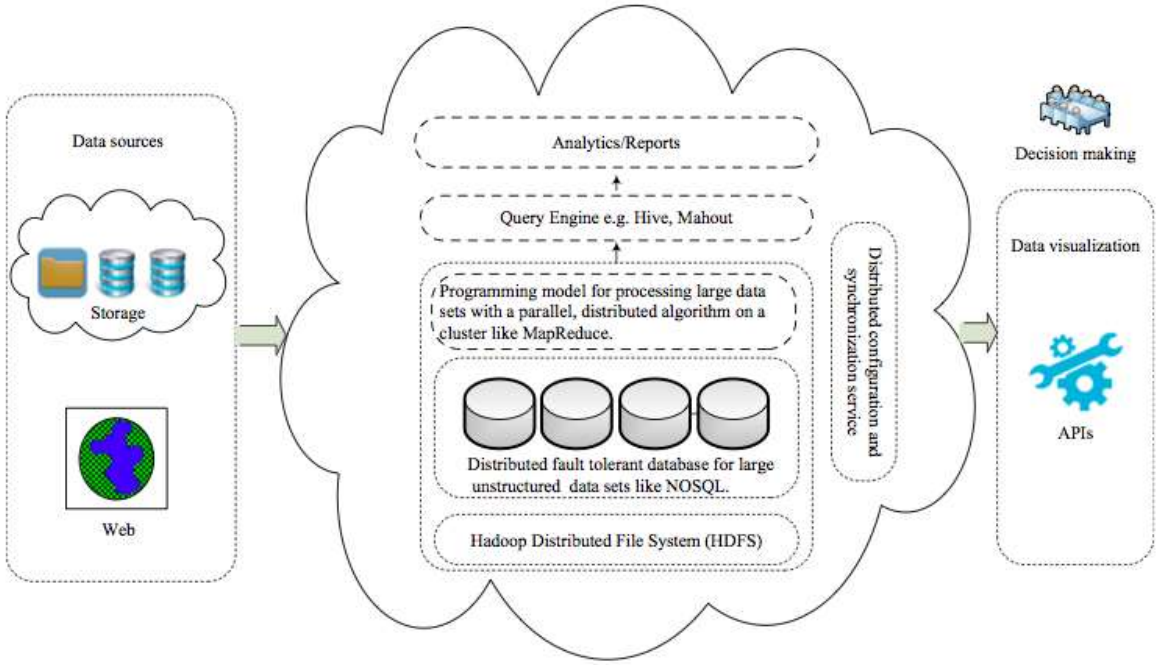

Fig : Usage of Cloud Computing in Big Data Analytics

Security issues:

Security is becoming major issue for data storage in cloud based networks. Cloud computing technology comes with security issues which include networks, databases,operating systems, virtualization, resource scheduling and allocation, transaction management, load balancing and
memory management. The security issues associated with cloud computing environment can be categorized into several levels such as:

1. Network level

The issues and challenges associated with the network level includes network protocols and security in networks such as distributed nodes, distributed data etc

2. User Authentication level

The issues and challenges associated with this level includes encryption/decryption techniques, authentication methods which includes authentication of distributed applications, access rights for nodes, logging etc

3. Data level

The issues and challenges associated with this level include integrity of data and availability issues with data such as protection of data and distributed data.

4. Generic level

The issues and challenges associated with this level includes different usage of security tools and usage of different technologies.

ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

Challenges:

Figure 4: Classification of the top ten challenges

Data management is changing into an progressively important component of analytics in the age of BigData. due to the data generated by the number of firms is increasing in huge size and greatest flexibility provided by cloud service provider is increasing currently a day. The additional the flexibility the additional the challenges exist in global internet market.Big data provides significant opportunities to service providers by creating info additional valuable. However, policies, principles, and frameworks that strike stability between risk and value within the face of skyrocketing data size and deliver better and quicker data management technology will create immense challenges.

Future Work:

As clouds become safer, reliable, and reasonable, the use of data analytics in cloud computing will still grow. it might not be unbelievable that soon all of a company's information can be saved to clouds and accessed anyplace by those in need of the information. local servers and personal computer onerous drives might make to all or any data being hold on remotely in data warehouses far away from the physical location of a business. whereas some still shudder at the potential security risks of cloud devices, likely they're going to become as efficient and secure as any typical drive or server.

References:

[1] Dr. M. Bhanu Sridhar, A. Koushik(2017). "A Study of Big Data Analytics in Clouds with a Security Perspective." International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 IJERTV6IS010010 Vol. 6 Issue 0.1

[2] Bernice M. Purcell. "Big data using cloud computing." Journal of Technology Research.

[3] Cloud Security Alliance CSA (2013) Expanded Top Ten Bid Data Security and Privacy Challenges. Retrieved from http//websevices.thesba.com accessed Sep 14, 2015
[4] Nabeel Zanoon, Abdullah Al-Haj, Sufian M Khwaldeh(2017). "Cloud Computing and Big Data is there a Relation between the Two: A Study." International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, pp. 6970-6982.
[5] Data Analytics in Cloud Computing, technologyadvice.
[6] Neha Mathur, Rajesh Purohit (2017). "Issues and Challenges in Convergence of Big Data, Cloud and Data Science." International Journal of Computer Applications (0975 – 8887) Volume 160 – No 9.