

Cloud for Big Data Analytics Trends

Affreen Ara¹, Aftab Ara²

¹ MTech in Computer Science and Engineering, Padmanava College of Engineering/BPUT, India

² Phd in Management, Ravenshaw College/Ravenshaw University, India

Abstract. Cloud can provide infrastructure, platform, and software resources as services to big data. Big Data gives substantial value to organizations who implement cloud in their infrastructure. Big data and cloud technology converge to make big data analytics in clouds a reasonable option. Big data analytics is growing to include data from multitude of sources. Most of the largest-volume, fastest-streaming, and most complex data can be source of big data. Cloud for Big Data Analytics is also known as Data Analytics as a Service. Enterprise of all types use cloud service to reduce capital cost as they provide resources required to run their applications. The objectives of my paper is introducing big data and cloud, giving an insight to interleaving big data and cloud technology, and giving brief explanation about Data Analytics as a Service trends and limitations. The paper describes the future or trends of Cloud for Big Data Analytics. The methodology of my study is from secondary source i.e. from journals, white papers, tutorials, conference proceedings, technological papers and websites.

Keywords: Big Data, Cloud Computing, Analytics, Big data Analytics, Data Analytics as a Service.

I. Introduction

The rise of cloud data stores and cloud technology has been pioneer to growth of big data. The volume of data being generated in enterprise is expanding. In future every sector will need to deal with exploding data volume which can fuel exponential growth of data in future. There is dire need of using a cost-effective scalable solution for storing and analyzing these data. Data Analytics can create unified set of solution beginning from collecting data, processing and analyzing and storing data. Big data Analytics is the amalgamation of Big Data and Analytics where advanced analytic techniques are applied on big data sets. Data analytics help enterprise allowing them to make better decisions based on real-time analysis. They allow analysts, researchers, and big business users to make quicker decisions using data that was previously redundant. By using superior analytics techniques such as text analytics, machine learning, predictive analytics and data mining enterprise can analyze previously unexploited data source. The objective of my paper is to explain how recent cloud –big data analytics trends that are occurring in cloud and big data universe which can be beneficial in future.

II. Big Data

Data becomes big when it difficult for enterprise to store, analyze and process it. Big data refers to huge set of data [16] that are of high order of magnitude larger (volume); more diverse including structured, semi structured and unstructured data (variety) and arriving fast than your organization to deal with. This large volume of data makes companies difficult for traditional relational databases to capture, analyze, and store the data. The key enablers for growth of big data are increase in computer processing power, storage capacity and easy availability of data. Big data has following characteristics – high volume, high velocity, or high variety. Big data comes from variety of sources such as sensors, devices, video/audio, networks, log files, transactional applications, web, and social media.

Four [15] Vs of Big data

- Volume-Scale of data
- Variety-Different form of data
- Veracity-Uncertainty of data
- Velocity-Analysis of streaming data

Data can be collected from variety of sources for example social networking, sensors, web, customer email, audio/video twitter feeds, call detail reports, network data, video cameras, and equipment sensors. Online social networking and shopping website collects customer information to understand customer behavior. Amazon, Twitter, Facebook, Google and LinkedIn use tools to collect data, analyze and study users behavior to provide better service and turn into profit.

III. Cloud Computing

According [13] to NIST “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage,

applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

It has following properties.

- **Rapid Elasticity:** Scale up or down quickly and easily to meet demand.
- **Metered service:** Pay as per use.
- **On demand self-service:** All the IT resources you need are available to the customer as self-service access.
- **Resource pooling:** It serves multiple clients, customers with provisional and scalable services.
- **Network Access:** All resources are available for access from a wide range of network devices, such as tablets, laptop ,computer and mobile

There are three types of Cloud deployment models which are adopted by enterprises.

- **Private Cloud:** These types of clouds are deployed on a private network which are managed by the enterprise itself or by third party. This cloud infrastructure is used to share the services and data across various departments of a large enterprise.
- **Public Cloud:** These clouds are deployed over the Internet and available to the general public for use. They offer analytics and data management service which are controlled by the provider itself. Organizations use these Clouds to carry out analytics or share results of public analytics.
- **Hybrid Cloud:** It merges both public and private Clouds where supplementary resources from a public Cloud is delivered as required to a private Cloud. Hybrid model can achieve elasticity and security and also provide cheaper base load and burst capabilities to cloud. Hybrid deployment can create and deploy analytics applications using a private environment, thus using security benefits of public cloud.

IV. Big Data Analytics Adoption In Cloud

Cloud computing[8] is a low cost model for big data analytics .There are increasing number of technologies that are adopting big data in cloud .Big data and cloud technology when converge make big data analytics in clouds a good option. Companies move the data to dedicated servers for getting better analytical results. The elastic property of cloud is ideal for big data analytics which is the process of decomposing large volumes of structured or unstructured data to identify patterns and make better business decisions. The Cloud Standard Customer Council underlines cloud architecture for Big Data and Analytics which describes reference architecture for analytic environment. .Hadoop [11] can deliver critical workloads beyond the web scale companies such as Yahoo, Spotify and TrueCar .The businesses like Yahoo can significantly leverage Hadoop as Hadoop can used by enterprises to extract valuable information from data under management plus it can deliver latest critical analytic applications .

4.1 Business Drivers:

Companies are making use of big data analytics to determine business trends and insights from the volume of data created. As cloud [4] provides elasticity and expertise for accessing data and it can derive value from it.Enterprise can use their private cloud infrastructure to improve risk and maintain control by analyzing load, cost and security etc. Analyzing big data has many advantages as it results in business growth, cost savings, revenue growth and also helps in better marketing for organizations. Analytics services when used by enterprise boost scalability of enterprise itself. Microsoft’s big data solutions run on Hadoop and can be used either in the cloud or natively on Windows. Business users can use Hadoop to operate on data using benchmark tools including Excel or Office 365 which can be integrated with core databases to analyze both structured and unstructured data and create sophisticated 3D visualizations .Microsoft’s solution enables users to analyze Hadoop data from within Excel, adding new functionality to software packages .



Fig-1-Interleaving Big Data and Cloud

Cloud model when interleaved with big data yields following benefits:

- **Speed and agility:** Private clouds can offer an effective method to use big data analytics and supplements internal resources through public cloud facilities. The agility provided by the cloud lets user to speed up their infrastructure as data volumes changes.

- **Extract value from big data:** Enterprise emphasis their budget on Analytics as a service (AaaS) which is supported by all three cloud model i.e. public, private and hybrid model.
- **Reduce expenditure.** Big data cloud-based analytics can bring substantial cost benefits when it comes to storing large volumes of data and provides resourceful ways of doing business.
- **Better decision making:** Memory analytics when combined with the ability to evaluate various sources of data helps to analyze information and make quick decisions based on data analysis..
- **New products and services:** Big data analytics enables companies to create more product according to customer needs and satisfaction by analyzing customer requirements and behavior.
- **Data is becoming more valuable:** Companies need to find advanced approaches to process, manage, and analyze their data whether data is structured or unstructured. Combining variety of data sources and types can uncover some of the most interesting unexplored patterns and relationships.
- **Enhancement in cloud security:** Security within a cloud environment has always been a prime concern, especially for organizations in highly standardized industries. Cloud providers now offer ways to enforce security at multiple layers.
- **Innovation:** The convergence of analytics and cloud can give innovative results .It lowers costs, increases speed, agility, and security that the cloud services offer.

V. Data Analytics As A Service (Daaas)



Fig-2 :Big Data Analytics as a Service is combination of big data analytics and cloud

Big Data Analytics[15] can be defined as a process of examining huge sets of data to discover hidden patterns, unknown correlations, market trends, customer preference and other useful information. Data Analytics as a Service (DAaaS) is an extensible analytical platform that is provided using a cloud-based delivery model, where various tools for data analytics are available and can be configured by the user to efficiently process and analyze huge quantities of heterogeneous data. The combination of Big Data with Cloud can improve the adoption of analytic capabilities over large volume of heterogeneous data sources that businesses cannot handle. Companies largely benefit from value derived from big data analytics. Value is in the data itself, and not in the expertise used in it. Companies need to access services that provide advance big data Analytics for the future. They also need to develop good infrastructure and the associated capital investment .Cloud can be good resource for analytics engineers due to elasticity property .Software as a Service (SaaS) model is adapted by analytics service providers to help cloud to solve big data problems. Emcien Corp provides pattern detection software as a service running on Amazon Elastic Compute Cloud (EC2) .It caters clients such as large retailers, telecommunications providers and intelligence agencies. In Software as a Service [14]model the user does not require to purchase all that hardware but only need to have a web browser. Big data are sophisticated platforms because to get cost effective service from analytics vendors the consumer needs to install additional software resources such Hadoop and Cassandra and other foundational components to get better performance.

VI. DAaaS Applications

Data analytics as a service is a business changing trends as it has lot of application in many field. Companies get competitive advantage by adapting and implementing DAaaS. Analytic approach in different field may vary. DAaaS application are as follows:

- **Social Media:** Cloud data analytics has huge importance in social media, website such as Facebook, twitter etc .These websites use analytics to interpret user social media activity and customer behavior. Cloud drivers allow simultaneous examination of social media data and result can be analyzed at real time.
- **Medical:** Big data, cloud and analytics technologies can help doctors to make better decisions by allowing them to conduct real time quick analysis of patient history, genome sequencing, blood test, investigate similar medical case and research articles throughout the world. Doctors can use analytic tools , techniques and make the best use of technology thereby reducing cost of treatment.
- **Smart Cities:** Cloud based analytic service can help in developing information intelligence and smart cities can benefit from big data. Real time information on transport, energy, smart lighting, noise pollution,

citizen information, waste management, water management etc. can be fed into cloud for analysis. City management can get information from data collected from various source to provide better governance, planning and management. Urban life aspects such as health, pollution, public, air pollution can improve and benefit from cloud- big data analytic solutions.

- **Tracking preference and products** Amazon uses data analytics on cloud drivers to track product across its warehouse. Websites provide user recommendations based on their previous search history. Most of user preference is stored in cloud for future reference .Social media analyst analyze data across network to provide markets standard to analyze their own success.
- **Keeping customer records;** Website store user profile in cloud for future and they record and process data to local servers real time. Most user information is stored on cloud so that when user changes computer his preference does not change.

VII. DAaaS Trends

Big data and analytics are evolving so rapidly that businesses need to adopt cloud, as cloud has replaced displaced legacy systems. Hadoop[11] and business intelligence (BI) platforms such as Birst have established new paradigm for enterprise analytics. Hadoop and Birst are offering data accessibility, analytic agility and application performance that big business consumer require. Data technologies have earned significance along with social, mobile and cloud, analytics as one of the core disruptors of the digital age. The year 2016 holds in future for the data and analytics space. Analytics provides the foundation for the business models of the future. It's not just essential for our decision makers, but also for customers. The following diagram shows Data Analytic as a Service trends followed by brief description.

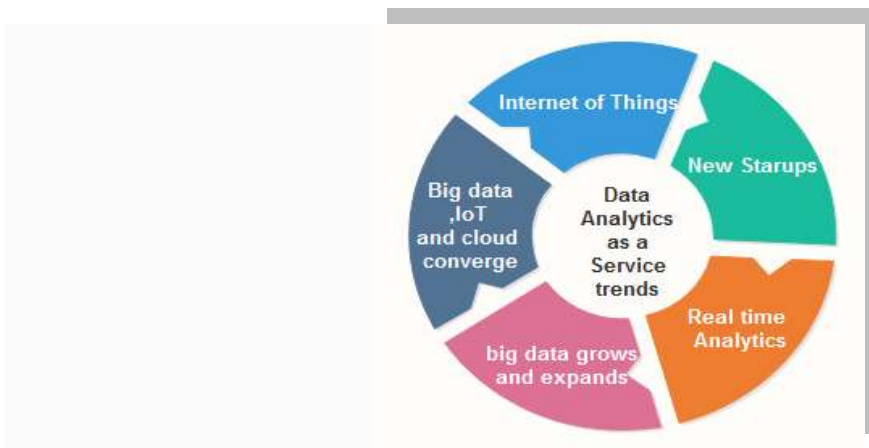


Fig-3:Data Analytics as a service trends

- **Internet of Things (IoT) :** Gartner has [14]placed IoT at peak of its so-called *hype cycle*, which means that IoT at present is the key target of inflated expectations .Statistical Analytical System for Internet of things covers full analytic cycle starting from data collection ,integration and deployment .IoT is future of data analytics whether data is in rest or motion and help in making better storage assessment and decisions Smart city is an intelligent system which when deployed with sensors could transmit real time data to cloud. Sensors would enable smart use of resource whenever required. Gartner has estimated that IoT acceptance will grow to 26 billion Internet-enabled devices installed by 2020.
- **Real time analytics:** Real-time analytics is just the beginning .Analytic Software delivers elastic, management decision and application development tools as enterprises need to make faster real-time decisions. Advanced platform tool can handle and assess real time streaming data close to its origination, transforming and understanding data. The platform would facilitate communications in multiple protocol languages. New tools like the Spark Cassandra Connector emerge, which allows an enterprise to turn to Spark to analyze data stored in Cassandra, companies will increasingly be able to collect business intelligence from stack of data at high speed. Big data future is monitoring integrations and alert systems. In future real-time analytics platforms will recognize and alert users of potential breach or network vulnerabilities the instant they occur. Artificial systems can be future of Big Data as they will resolve issues before you are even aware of them. Artificial system solutions will change network management from being reactive to proactive.
- **New start-ups: There** is huge demand for Hadoop deployment in cloud as it can deliver more critical workloads. Hadoop has sophisticated stack because building its stack involves correlating many different

open source components that do not share a common set of APIs and do not work together. It requires hiring or developing new sets of highly technically skilled analyst. Today there are new start-ups offering solutions supporting SQL. The Analytics market is growing of net technologies and new big players have moved in and started to buy those start-ups for their technology and skill sets.

- **Big data grows and become fast:** Hadoop[7] has been in market for long time and it has become a core part of the enterprise IT landscape and it will see investment in security and privacy in future .Apache Sentry provides service for implementing fine-grained, role-based authorization to data and metadata stored on a Hadoop cluster. In today's world security and privacy is critical for enterprise success .Users want security from enterprise-grade RDBMS platforms which is barrier to analytic adoption. There is increase in demand of Hadoop as volume of data is growing rapidly, so faster data analysis is expected from traditional data warehouses. Cloudera,,Actian Vector , AtScale, Impala and Jethro Data enable OLAP cube, for Hadoop thereby bridging the gap between traditional business intelligence and the big data.
- **The buzzwords converge: IoT, cloud and big data come together.** Internet of things, cloud and big have become priority for developers. Internet of Things (IoT), shows businesses how to take benefit of the fast Big Data that flows from the digital world.. Thingalytics, a composite of Things and Analytics, shows organizations how to use instantaneous analytics and algorithms to collect the data that flow from IoT while simultaneously minimizing threats. Big data analytics developers are spending most of time creating Internet of things. The technology is nascent but the data collected from devices used in Internet of Things is essential for the cloud. The volume for data collected from Internet of things can fuel big data explosion. Cloud and data companies such as Google, Amazon Web Services and Microsoft bringing Internet of Things services to life where the data can move impeccably to their cloud based analytics engines.

VIII. DAaaS Limitations

The DAaaS platform is built for horizontal platform that can be used in different vertical platform. It can be designed to be universal as well as extensible platform. As DAaaS is an universal platform that has data storage and modelling issues plus It does not support real time analysis of data. For example, the Map Reduce framework is based on batch execution modes of operation. This imposes workload onto the system if real time modes of operation are used. The fundamental property of big data is be able to represent and store different kinds of data and store huge volume of data in distributed manner.DAaaS limitations are as follows:

- **Storage:**Storage is affected by virtualization loads as a result it impacts performance in a highly virtualized and distributed cloud is risky .
- **Networking Issues:** Big data analysis in the cloud causes networking issues for service providers. It takes technical knowledge to efficiently get data to an analytics team.
- **Analytic Knowledge:** Analytics demands knowledgeable data scientist who is updated with latest advanced analytics techniques.
- **Inconsistency in data collection:** Tools used to gather big data sets are sometimes imprecise.
- **Security and privacy:** Businesses might be concerned that their data is stored in a virtual server passing their data over to a third party. Privacy can impact cloud analytics as data becomes public after analytics.
- **Data Volume:** Moving data to cloud for computation becomes strenuous task but vice versa is not true.

IX. Conclusion

Companies that embrace cloud-big data analytics will make more well-informed decisions and see financial gains. Many companies are developing tool for cloud-big data analytics. Companies will find big data tool and technologies effortlessly accessible and be able to deploy, administer, manage and secure .The companies that don't take advantage of big data business opportunities will fall behind the competition. In the best case scenario, companies will analyze information collected from mobile devices, storage arrays and management tools.Cloud- Big data Analytics can be used for network optimization and intelligence system to identify potential breaches beforehand. Internet of things when combined with analytics can help in diverging data sources and enables application of analytics in various areas such as monitoring sensors .But it's unclear whether today's companies are ready to take advantage of Internet of things potential. IoT will expand data sources and enable analytics to be applied to new areas, such as the monitoring of sensors attached to manufacturing devices. Internet of things could be a boon to analytics by serving as a major source of data. The introduction of new tools and technology will give chance to smaller firm to surpass larger ones in efficiency and productivity leading to growth of new start-ups. Growth in Internet of Things, data analytics and cloud would create more jobs for data scientists and analyst.

References

Journal Papers:

- [1]. Kanchan A. Khedikar, Role of Cloud Computing in Big Data Analytics, Using MapReduce Component of Hadoop, Maharashtra, India., *International Journal of Innovations in Engineering and Technology (IJJET)*, Vol. 4 Issue 1 June 2014 87 ISSN: 2319 – 1058
- [2]. Han hu1, Yonggang Wen. Tat-Seng Chua1 and xuelong , Toward Scalable Systems for Big Data Analytics: A Technology Tutorial *Komal V. Kumawat*
- [3]. Marcos D. Assunção a,* , Rodrigo N. Calheiros b, Silvia Bianchi c, Marco A.S. Netto c, Big Data computing and clouds: Trends and future directions, Elsevier, *M.D. Assunção et al. / J. Parallel Distrib. Comput. 79–80 (2015) 3–15*
- [4]. Celestino Güemes , Jordan Janeczko, Thierry Caminel and Matthew Roberts bill, Atos Data Analytics as a Service: unleashing the power of Cloud and Big Data
- [5]. booz, allen and Hamilton, cloud analytic playbook
- [6]. Dave Chappelle , An Oracle White Paper September 2013 Oracle Enterprise Transformation Solutions Series Big Data & Analytics Reference Architecture
- [7]. Big data analytics fourth QUARTER 2011 By Philip Russom tdwi best practices report
- [8]. Big Data in the Cloud: Converging Technologies:, 2015 Intel Corporation, 0415/LF/ME/PDF-USA 328762-001
- [9]. Big data as a Service, EMC2 Corporation, A market and technology perspective, July 2012

Website

- [10]. Rachel Shusher, Big Data initiative essentials: Adapting Analytic strategies , <http://searchnetworking.techtarget.com/feature/Big-data-initiative-essentials-Adopting-analytics-strategies>
- [11]. Thor Olavsrud , 21 data and analytic trends that will dominate 2016 <http://www.cio.com/article/3023838/analytics/21-data-and-analytics-trends-that-will-dominate-2016.html>
- [12]. Robert L. Mitchell, 8 big trends in big data analytics , <http://www.computerworld.com/article/2690856/big-data/8-big-trends-in-big-data-analytics.html>
- [13]. Peter Mell , Timothy Grance , The NIST Definition of Cloud Computing.
- [14]. SAS: Analytics, Business Intelligence and Data Management., http://www.sas.com/en_in/home.html
- [15]. Big data Analytics <http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>.
- [16]. Big data Analytics hub <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>