IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Cloud Analytics and its Implementation Challenges

G. Kalpana*, NV.Chaitanya

[1] MCA, CBIT, Hyderabad, India
[2] MCA, CBIT, Hyderabad, India

Available online at: www.ijcseonline.org

Received: 16/Mar/2018, Revised: 24/Mar/2018, Accepted: 12/Apr/2018, Published: 30/Apr/2018

Abstract- Cloud computing is "a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.". Cloud computing is capable of enhancing business productivity and agility while allowing for greater efficiency and reducing costs. It is also possible to migrate big data to cloud. Increasingly, the focus of organizations has been on the ways that cloud will power data analytics. Cloud analytics enable organizations to move business intelligence, data warehouse operations and Online Analytical Processing (OLAP) to cloud. Data management, security, integration, compliance issues, poor software architecture and design are some of the issues that motivates with cloud inclusion for analytics solution. This paper aims in outlining the data analytics, cloud computing, cloud Analytics, and its Architecture, applications of cloud analytics and implementation challenges along with suggested solutions.

Keywords: ***Data analytics, cloud computing, cloud analytics.***

I. Introduction

The advent of disruptive technologies such as, data warehouses, social media, audio-video streams, blogs and tweets is responsible for the generation of huge volume and variety of data. Transforming these huge datasets into business intelligence by correlating, synthesizing and aggregating useful insights from them is the prime concern of analytics solutions.

Data Analytics have been the major area of focus for Businesses today to help direct their strategy to maximize profits, minimize the tremendous guesswork involved in trying to understand Behaviour clients, and help in systematic tracking of behavioural patterns of the customers to best construct business tactics and operations to minimize uncertainty[1]. Analytics helps in determining the requirement of new customers, and also helps in recognizing existing patterns in data to help better serve existing customers, which is typically more cost effective than establishing new business. In an ever-changing business world subject to countless variants, analytics provides an edge to the organizations in recognizing changing scenario so that they can initiate appropriate action to stay competitive [2].The clouds and analytics consolidation could help to improve businesses store, interpret, and process their big data to better meet their clients' needs. Migration of data and applications to cloud for convenience and cost saving has triggered, analytic solutions to follow the same path[3].

Big Data Analytics offers valuable insights that can spark innovations, create competitive advantage, and drive increased revenues .Big Data is used to handle large and complex type of data[4]. There is an immense requirement of handling this complex type of data, because today all the data from and internet and IOT devices are connected to each other in big manner. Big data describes the large volume of data – both structured and unstructured [4]. But it's not the amount of data that's important. It's what organizations do with the data that matters. When we deal with Big Data we need to handle multiple and variety of fields like business models, visualization, interaction, model development, data management, architecture management, etc[5]. There is an immense reason to use Big data for handling large, voluminous amount of data which will be created from various part of the different organizations.

This paper Organized as follows I.Introducation, II.Cloud omputing, III.Cloud Analytics, IV.Cloud Analytics Architecture, V.Applications of Cloud Analytics, VI.Implementation Challenges Of Cloud Analytics, VII.Conclusion

II. Cloud Computing

Many researchers have defined cloud computing differently. One mostly accepted definition is given by the United States Institute of Standards (NIST).

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of

configurable computing resources like networks, servers, storage, applications, and services, which can be provisioned and released with minimal management effort or service provider interaction[5]. This cloud model is composed of five essential characteristics, five service models, and four deployment models".

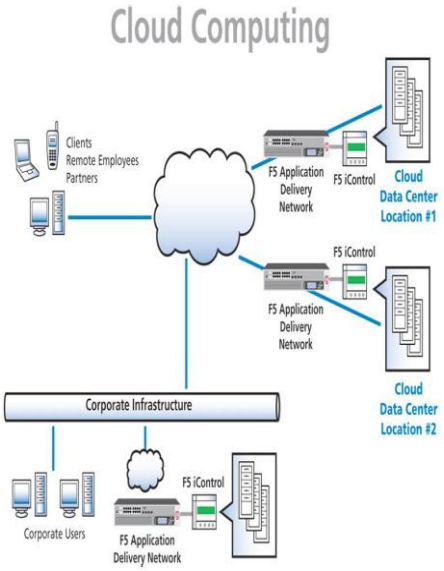

Figure 1: Cloud Computing [7]

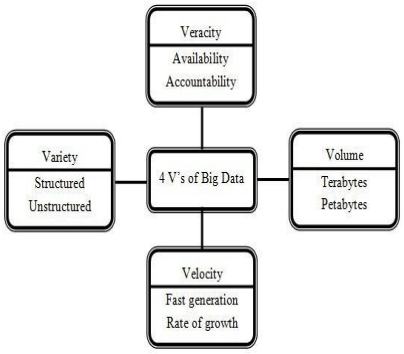

Fig 1: Big Data Characterstics[6]

III.Cloud Analytics:

"Cloud analytics," is referred as software that is delivered on a software as a service (SaaS) basis [6]. According to Gartner, "Analytics has emerged as a catch-all term for a variety of different business intelligence (BI)- and application-related initiatives. 'Analytics' is used to describe mathematical and statistical data analysis through which one can do clusters, segments, predicts and scores what scenarios are most likely to happen." Other terms used to describe the same software include cloud-based BI, SaaS BI or SaaS analytics [8].However, "cloud analytics is also referred to as analysis of data related to cloud computing. These are usually cloud management solutions that monitor the performance of cloud infrastructure and applications.

This article will focus primarily on the first definition of cloud analytics (big data analytics delivered through the cloud) rather than the second (analysis of cloud-related data).

IV.Cloud Analytics Architecture

Cloud analytics, which refers to the utilization of cloud services for a single or all components of an analytic solution, enable organizations to leverage social data, Internet data and third party data by using pay per use model . Integration of these data with the enterprise data will provide more insight about customer preferences and demands. On-premise analytics solution which is built on restricted storage space and compute power infrastructure when replaced by cloud analytics will enhance the business intelligence capacity of the organization. The need for new relational database model such as NoSQL, which efficiently stores and access huge volume of data from cloud evolved as the analytics applications utilize datasets with more reads than writes. The compute intensive module of the analytics is the primary candidates of the analytics application to be moved on to cloud for the advantage fast query responses and reduced decision making time .

Four main layers of cloud analytics architecture as given in figure 3 are a) infrastructure layer b) data storage and management layer, c) analytics layer and d) visualization layer . A common layer is present that includes security, information governance and regulatory compliance. Security management deals with access management, user authentication, protection of data and application in multi-tenant architecture. Information governance deals with the standard policies that are to be followed while transferring data in the analytics environment. Regulatory compliance is the compliance measures that are to be followed while handling data storage across the globe. Domain specific compliance measures such as HIPAA, PCI-DSS/ US-EU Safe Harbor are to be applied based on the application or user demand [9].

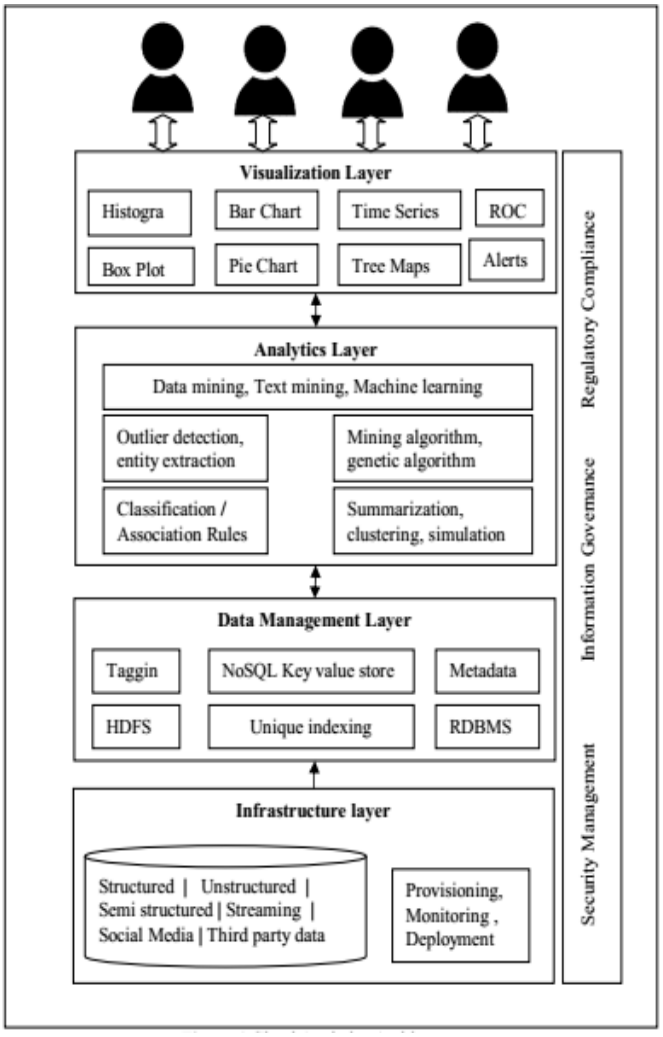

Fig 3: cloud analytics Architecture[8]

A. Infrastructure Layer

This layer is responsible for storing and managing data. The layer consists of both structured and unstructured data gathered from various data sources. The social data, sensor data, video streaming data and third party data are available on pay per use basis. The infrastructure layer handles the activities of dynamic provisioning, monitoring and automatic deployments. The rapid growth of the data base and the requirement spikes are handled by the vertical and horizontal scalability feature of cloud infrastructure. This is also the foundation layer for the cloud analytics architecture which enables organizations to leverage fast and low-risk infrastructure deployments .

B. Data Management Layer

Data Management layer is a data storage repository layer which is also known as "data lakes". Huge amount of raw data in semi-structured , structured and unstructured format is stored in flat files for analytics purposes . Until a query is raised the schema and data format requirements are not defined. Each data element of this layer is attached with a unique ID and is tagged using metadata tags. This layer is maintained using Hadoop-oriented object storage, where the queries are applied to retrieve intelligence [10]. It is different from the Enterprise Data Warehouses (EDW) as it employs schema-on-read processing, and has a greater reusability along with low cost storage. Node failure and data movements between the nodes are routine operations in the data management using cloud setup and maintains fault tolerance feature of cloud analytics using automatic replication of data across nodes in cloud cluster of this layer.

C. Analytics Layer

This is one of the important layers of the cloud analytics as it holds the business intelligence applications. It contains pre-analytic or filtering tools and analytics tools. A pre-analytic tool cleanses, chooses and organizes data that were retrieved from the data management layer and analytics tools to identify the patterns and retrieve actionable insights from the data. Various data and text mining algorithms such as association rules, frequent pattern matching algorithms, clustering algorithm, genetic algorithm etc. are utilized in this layer to retrieve intelligence. Supervised machine learning methods such as classification and regression are used on the existing set of data to perform predictive analysis. Various analytics tools and software used by this layer are R, SAS, Mathematica, Mat Lab, Map Reduce, Pig, Hive and ETL tools, etc.

D. Visualization Layer

Visualization layer provides user interface for accepting user queries and displaying analytics results. One of the desired feature of this layer is Quick and effective response to gain fast and deeper insights. This layer helps users to customize the data analytics visualizations . Visualization layer enables subject matter experts (SME) to explore the data without any assistance of the IT experts. This layer provides the flexibility to explore business query with various parameters and aspects, which will assist in making accurate decisions . The elimination of assistance to extract insights enhances the quality of information retrieval as the SME directly works on the data without the interference which eliminates the requirement specification degradation.

V.Applications Of Cloud Analytics

Following are the few real time applications where one can apply Cloud analytics and reduce their valuable time and difficulties in processing.

A. Social Media: The most important use of cloud data analytics is interpreting social media activity. Through the cloud drives one can get quick results in a short period of time for the simultaneous examination of social media site data

B.Tracking Products: companies such as Amazon.com use data analytics for tracking their products among their warehouses in different places and ship the items as per the needs of the customer across the globe.

C.Tracking Preference:

All users information is remotely stored on cloud drives. With the use of cloud analytics tools one can easily determine user preferences.[11]

VI.Implementation Challenges of Cloud Analytics

Traditional Business Intelligence applications maintain transaction data, enterprise data warehouse and its analytic solutions in the premises, which provides greater security and customizability but lack in handling increase in volume of data and computing on huge amount of datasets [12]. Cloud analytics offers various challenges in its implementation and provides effective solution for the traditional Business Intelligent issues. Most often Organizations fail to devise a collaborative plan involving different departments, prior to the implementation of cloud analytics solutions [13] . This results in a piece-meal approach of each department having their own analytics solutions which stubs the cost reduction benefit of cloud analytics.The challenges also arise due to geographically distributed storage and involves various third-parties for the provision of intermediate services, cross border compliance issues and network latency. The challenges are addressed in this research work in the viewpoint of organization, provider and governance.

A. Marketing Hype

This is also one of the main challenges as organizations often get carried away by the marketing claims. All companies that offer one or more components of the analytics solution are tagged as cloud analytics companies. Some companies offer complete Business Intelligent solutions and the others may offer either the analytic part or the data storage component. Lack of clear understanding of the scope required and the scope of the cloud analytic solution will result in wrong selection of the product that may result in performance degradation of the decision making process.

Suggested Solution: Six elements of the analytics solutions and their scope of operations are to be clearly understood by the person responsible for the selection of the analytic product. Depending on the nature of data and process sensitivity either SaaS BI or analytics as a service has to be selected.[14]

B. Integration

Integration may be considered a benefit if handled properly and also pose a major challenge for the organizations that maintain on-premise applications and opt out Cloud Business Intelligence for analytic operations. The format compatibility of the data that are to be used between the on-premise and cloud analytics solutions may lead to inconsistent decision making. Data originating from the cloud such as social media data or the SaaS application data are comparatively easier to handle than on-premise data. Cloud analytic solutions comprise of disparate and distributed data and often require extensive integration specifications.

Suggested Solution: Maintaining the data in the open data format or the standard data format is the optimal solution for the integration issues. The usage of data format conversion module for the purpose of integration must be minimum to reduce the software development and maintenance cost.[14]

C.Security

Security is a prime concern for the cloud users due to the loss of control on data. Scalability of the analytics data will be affected when maintained in on-premise storage servers.

Suggested Solution: Identification and segregation of sensitive and less sensitive data and using hybrid approach of on-premise and cloud storage will be a solution for this security challenge .If the organization's complete data collection is sensitive then the analytics as a service component, which provide modeling and visualization tools as service that works on the on-premise data can be utilized. Possession of valid security certificate by the provider will increase customer trust factor.

D.Lack of skill set

There is a growing demand for the data scientist as more and more organizations are getting into the bandwagon of analytics. Recruiting, training and employing data scientist is a challenging task. As Cloud Analytics is an emerging discipline getting expertise from other different fields is very difficult. The cloud analytics solutions are provided by numerous companies and each product is designed to provide domain specific solutions. The choice of the best analytic tool to assist the organization in making decision is very essential as it is the base for generating accurate business predictions and suggestions.

Suggested Solution: Organizations must focus on employing people with critical and analytical thinking skills. With the shortage of data scientist skill set, organizations must develop analytics skills of their internal talent pool. This is an optimal solution as they are already aware of the business processes. [14]

Deregulatory Compliance

Data governance provides standards and context specific policies regarding its storage and processing. Data regulation policies vary across locations.Personal data is protected in EU countries where as US patriot law allows access to data. In spite of the regulation variation, geographical distribution of data is essential to maintain fault tolerance. The restriction on data movement will force storage at a single data center which will affect optimal resource utilization, fail over facility and cost of the resources.

Suggested Solution: Product specifications are to be screened properly to check for the compliance certifications. The providers of cloud analytics services have to acquire compliance certificates such as HIPAA, PCI-DSS, US-EU Safe Harbor depending on the domain of the serving customers. These certificates are to be renewed periodically .

F.Customization

Based on their business requirements traditional BI applications are developed for a particular customer. SaaS BI on the other hand is developed for a wide range of customers and hence customization as per the business needs is not possible. Organizations are either forced to mould the analytics needs as per the SaaS BI specifications or have to tolerate presence of unwanted module. This will pose a great challenge for the organizations who maintain traditional BI along with analytics taken as service.

Suggested Solution: According to their requirement choosing an appropriate SaaS BI matching will provide a solution to the challenge. If the organization is very particular about their BI needs, then the data storage can be done in cloud and analytic solution can be developed as on-premise application. This will eliminate the tolerance of unwanted modules.[14]

G. Latency

Cloud inclusion brings in the latency challenge due to its Internet based working . This is a long standing challenge due to its lack of proper accountability. Data localization can be opted by the organization to reduce the latency but it will increase the cost of storage. Cloud analytic possesses automatic transfer of the process nearer to the data location in order to reduce latency issues. Organizations, who switch from traditional BI to cloud analytics, will also face the latency issues during initial data migration.

Suggested Solution: Third-party auditing of the network will attempt to solve the accountability issues. Based on the criticality of the data a tradeoff between the time delay and cost has to be carried out. The initial bulk data transfer delay can be reduced by sending the data in storage devices physically.[14]

VII. Conclusion

Today it is widely accepted that cloud computing and big data technologies are two dominant technologies that may shape up the business world. Scalability, flexibility, fast delivery of analytics solution, ubiquitous access, limitless computing and storage are the benefits of cloud analytics inclusion. Big data technologies along with cloud computing help in allowing businesses to make proactive, knowledge-driven decisions through which future trends and behaviors are predicted. Businesses will be able to store their data remotely and access data and services from anywhere and anytime. Adoption of cloud analytics has few inherent challenges that need to be addressed prior to its implementation for attaining deeper business insights. In spite of the challenges mentioned in this research work cloud analytics is sought after for analytics solution due to its ease of usage and cost reduction. This brief overview of Cloud Analytics may help the researchers who want to make further research in this field.

Acknowledgment

This Cloud Analytics Research work is a collaborative work under department of Master of Computer Applications and Computer Science Engineering for identifying and optimizing implementation challenges for effective Cloud Analytics.

References

[1] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. A. Netto & R. Buyya. “Big Data computing and clouds: Trends and Future Directions,” 79 -80 (2015), pp. 3-15. For Journal of Parallel and Distributed Computing,

[2] X. Sun, B. Gao, Y. Zhang, W. An, H. Cao, C. Guo, W. Sun, “Towards delivering analytical solutions in cloud: Business models and technical challenges,” IEEE Computer Society, Washington, USA, 2011, pp. 347-351for IEEE International Conference on e-Business Engineering (ICEBE 2011)..

[3] D. Talia. “Toward Cloud-based Big-data Analytics,” IEEE Computer Science, (2013), pp. 98-101.for Conference.

[4] C. Elena. “Business Intelligence”, 1(2), 2011, pp. 1-12, for Journal of Knowledge Management, Economics and Information Technology,

[5] J. Ranjan. “Business Intelligence: Concepts, Components, Techniques and Benefits, 9(1), 2009, pp. 60-70,for Journal of Theoretical and Applied Information Technology

[6] Dp.Acharja,Kauser Ahmed p title "Suvey on Big Data Analytics:challenge,open research issues and tools", volume 7,issue 2,2016 for International Journal of Advanced Computer Science and Applications
[7] Bala M Balachandran,sivika Prasad title"challenges and benefits of deploying Big Data Analytics in Cloud for Business Intelligence" for jounal elsevier
[8] R.VidhyaLakshmi, Dr.VikasKumar tilte "Challenges in implementation of cloud analytics" volume 4,issue 6,june 2016 for International journal of Information technology
[9] H. Demirkan, D. Delen. "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," Decision Support Systems, 55(2013), pp. 412- 421.
[10] C. W. Tsai, C. F. Lai, H. C. Chao and A. V. Vasilakos. "Big Data Analytics: a survey," A Springer Open Access Journal of Big Data, (2015)pp. 2-21 for Journal
[11] J. Roski, G. W. Bo-Linn, T. A. Andrews. "Creating value in health carethroughbigdata: Opportunities and Policy Implications," Health Affairs, 2014, 33(7), pp. 1115-1122.for Journal
[12] P. Vidhyalakshmi, V. Kumar. "Cloud Computing Challenges and Limitations for Business Applications," 2013, 1(1), pp. for Global Journal of Business Information Systems7-20.
[13] J. Myers. "Analytics in the Cloud", an Enterprise Management Associates (EMA™) End-User Research Report. Jan, 2015.for journal
[14]R.VidhyaLakshmi,Dr.VikasKumar tilte "Challenges in implementation of cloud analytics" volume 4,issue 6,june 2016 for International journal of Information technology

Authors Profile

G.Kalpana,Asst.Professor,Dept.of M.C.A,CBIT,Hyderabad G.Kalpana is currently working as an Assistant Professor in MCA Department, Chaitanya Bharathi Institute Of Technology, Hyderabad, A.P., India. She has pursued her M.Tech. in Computer Science and Engineering from HITAM Affiliated Engineering College under Jawaharlal Nehru Technological University, Hyderabad, India in 2010. She did her M.C.A from Annamalai University, Tamil Nadu,India in 2007. She has 13 years experience of teaching for undergraduate students and for post graduate students. She has published/presented 8 research papers in National/international journals and Conferences. Her research interests are in the areas of Image Processing , Data Mining and Data Analytics.

NV.Sai Chaitanya,MCA,CBIT, Hyderabad Chiaithanya has currently completed his MCA in Chaitanya Bharathi Institute of Technology ,Telangana,Hyderabad,India.He has recently been placed in multi national company "corecompete". He has presented 7 research papers in international/National journals/Conference. His research interests are data analytics,Block Chain,Cloud Analytics,Business Analytics and IOT.

