Data Analytics as a Service (DAaaS): An Arriving Technology in Cloud Computing

Prof. Anup R. Nimje[1]

[1](Sant Gadge Baba Amravati University, Amravati, M.S., India)

Abstract:- An emerging technology in computer science is the cloud computing that has become very popular day to day in corporate as well as in researchers. The cloud computing provides on-demand various services "as-a-service", such as SaaS, PaaS, IaaS, Storage as a service etc. The paper describes the model of cloud services given by Ascent, used for big data analytics, called "data analytics as a Service". It analyses the big data and generates the business oriented information. This information allows the enterprises for continuous innovation and more real-time decisions. The cloud services are on demand as a service in which the users gain the access to utilities services and also the enforcement of security, privacy and access controls. The "data analytics as a Service" technology gives data scientists and knowledge workers the personalized access to cloud containing information data sets. The performance can be improved by more interactive and rapid data discovery. Analysts request the personalized tools or utilities for analyzing the cloud managed data sets or the data warehouse. The "analytics as a Service" has become beneficial for the IT enterprises by reducing the costs significantly and risks. There is no need to physically replicate content for every single user who wants to view it, using cloud systems. After finishing the work, users are simply disconnected from their access to the data.

Keywords: - analytics, big data, cloud computing, data warehouse, hybrid cloud, information data sets.

I.INTRODUCTION

The cloud computing is an online service that includes Software as a service, Platform as a service, Infrastructure as a service etc. It involves dedicated servers. The Cloud technology consists of interconnected computers and servers, in the sense that it allows the centralized data storage and online access to computer services or resources to its clients. Clouds can be classified as public, private or hybrid depending upon its topology. [1]

Big data is as the name suggests, it is a collection of "big" datasets which is very large and complex. Big data requires the different techniques that consist of integrated solutions to analyze and discover the hidden values. It is difficult to analyze the Big Data using the traditional data processing applications. There are various challenges to analyze the Big Data, such as analysis, capture, search, sharing, storage, transfer, visualization and privacy preservations etc. For example to capture the business trends form the huge collection of data. Analyzing the big data includes some techniques, such as predictive analysis, data mining, text analytics and statistical analysis, BI analysis, data visualization etc. [2] To analyze the big data, the new class of technologies has been introduced that is "Hadoop" with related supplementary tools such as YARN, Map-Reduce, Spark, Hive and Pig as well as NoSQL databases.

The Big data analytics is the process of analyzing the big data, to visualize hidden patterns, statistical analysis, values etc. This can be used for effective strategic planning, more effective marketing, operational efficiency, competitive advantages and other business benefits. [3] The main purpose of this analytics is for providing the data scientists, modelers and professionals to analyze the big data that includes the large volumes of transaction data and some BI data.

While discussing the different services by the cloud computing, there is new service has been introduced, that is "analytics as a service" that simplifies the big data analytics that can be done online. The "Analytics as a Service" is a combination of on demand cloud service with the big data analytics. This is an era of businesses where decision making is highly increasing. This technology allows the organizations to analyze, visualize and discover the values and smooth real time decisions. The "as-a-service" replaces the tasks in information technology that are done manually by business users. The introducing different technologies "as-a-service" requires enforcing the security, privacy and access controls. The "Analytics-as-a-Service" allows the data scientists the personalized access to the centrally managed large data sets. [4] This reduces the various delays that data scientists, business analysts, and other information consumers face, while exploring the information data sets. This is more interactive and faster.

II.DATA ANALYTICS AS A SERVICE

Data Analytics as a Service (DAaaS) or Analytics as a Service (DAaaS) uses the cloud based delivery model and provides the extensible analytical tools. [5] In this service, different analytical tools are available and able to analyze the heterogeneous data. The generalized data consists of the large amount of data collected from the transactions or operations. The service users uploads their enterprise data over the cloud and gets the analyzed data useful for business purpose via cloud which consists of the analytical algorithms of machine learning concepts. The DAaaS platform is designed to be extensible, to analyze the heterogeneous data; there are series of tools available.
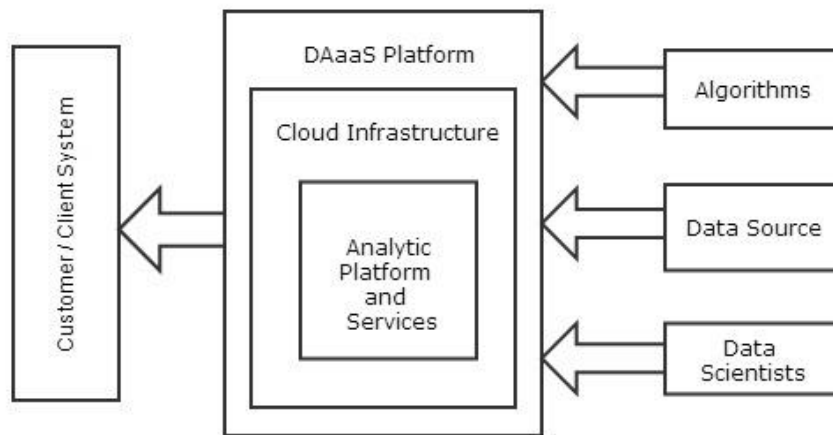

Fig. 1: Basic DAaaS model

The fig. 1 shows the customer system, data sources and DAaaS platforms containing analytics programs and tools and third parties consisting external resources.

III.COMPONENTS OF DAaaS SOLUTION

The architecture for DAaaS can be implemented as (given in Ascent white paper) shown in fig. 2 which consists of all components involved in service discussed in Ascent white paper. [5]
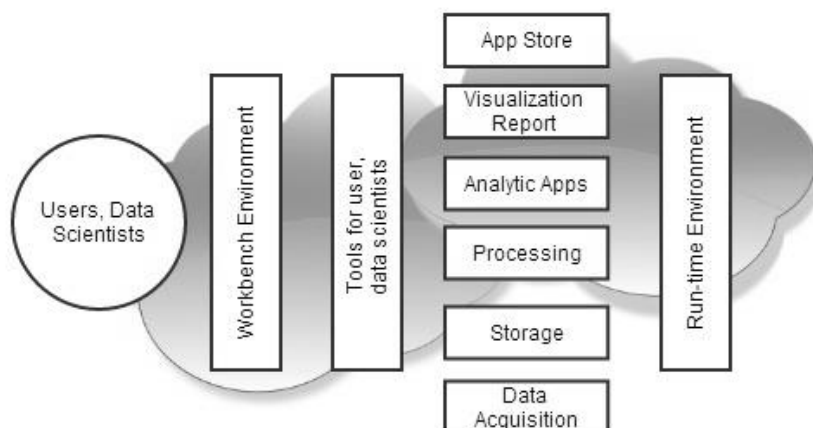

Fig. 2: DAaaS functional elements

The components that perform processing of the data and some components control the interaction with the users, including configuring the systems or workbench environment.

i. **DAaaS Runtime Environment Component:** There are components that are used as runtime, the execution of DAaaS solution. The logical data flow in the DAaaS from input to output.

ii. **Data Acquisition:** The component is associated with the data acquisition through web services and messaging interface. It works as the data source. This component should be flexible to work with large data sources and protocols.

iii. **Storage Component:** This component stores the information in the system and act as the core data repository. There is scope of research to store the peta-byte size data. It should be flexible with different data models.

iv. **Processing:** To process the big data, there is distributed system that operates in parallel, so that the various kinds of processing algorithms can be executed. This component is the pool between data storage and analytical services. There is popular solution for the processing of data that is Hadoop and Map-Reduce. It is distributed processing and supported by NoSQL data sources, different programming languages and analytical tools.

v. **Analytics Component:** It is the component of the system where the analytical processes are taken place. It consists of analytic services and the analytic framework. The Analytic Services are consisting of the main data analysis algorithms and perform very important functionality. According to the technologies used, dataset classes involved, the analytic component can be varied. The Analytic Framework that combines different components to get the desired business solution.

vi. **Analytic Apps:** The analytic apps are associated with the end user. It is business oriented and consists of the whole bundle of complete analytical services in the analytic framework.

vii. **Visualization / Reporting:** To simplify the access to information through web services, there is virtualization or reporting involved. It is user-oriented and provides integrity for end user in the form of virtualization and reports. Some commercial solutions are available (e.g. Tableau, QlikView) and some open source (e.g. Pentaho, Jasper Reports).

viii. **App Store:** The app store solution will provide Apps to end-users. An App-Store will provide the mechanisms to control the lifecycle of an app.

ix. **Workbench Environment Component:** To customize the sloutions from Analytic App there should be specific utilities provided to the end-user, called as Workbench Environment. It will consist of different kinds of functionalities:

a) **Data Integrator:** It is associated with integrity of data acquisition. It enables ETL process (extraction, transformation and load) of this information. Thus, it is properly introduced into the DAaaS environment.

b) **Data Scientist and Subject Matter Experts:** These can be emploees of the enterprise or service provider for the specialized services. They will work on the checking of the models, Analytic Apps, validity of data sets. The subject matter experts (SME's) provide the business context to the statistical outcomes. It helps to define the Business Pattern Rules that are based on the feedback from analytic patterns. These Business pattern Rules can be implemented in the Real-Time Analytics capabilities.

IV.BENEFITS OF DAaaS

There are various challenges to implement the DAaaS as it requires the large internal infrastructure and human resources. As suggested in the article [5], there are certain simpler steps for the data scientists that can be practice in the organization.

1) Explore the AppStore for an Analytical App that fits the problem.
2) Rent the Analytical App for a required time or quantity of data in GB/TB.
3) Configure the Analytical App to its needs including, for example, the usage of external data sources provided by the DAaaS.
4) Import the data to the Analytical App from the internal systems.
5) The SMEs in the company examine, validate the results with the business context.

6) The business solution is available for the organization.

As compared with the typical Big Data Project, there are the data scientists that need the additional resources to design and implementation of the solution. The complete Big Data infrastructure based in some complex technology like Hadoop is more expensive. More efforts are required for integrity. Using the DAaaS is not like using other kinds of services. It requires some initial requirements such as

1) To explore initial data,
2) To define analytical processes,
3) To implement and validate results using test data and
4) To optimize it as new data comes.

The advantages of the cloud delivery model are taken into consideration, these efforts are diminished. The cost includes only pay per usage, may allow temporal usage.

V.CHALLENGES OF DAaaS

There are various challenges in introducing the Analytic Services solutions such as it should support Big Data while using cloud model. [5]

1) Information Lifecycle: The flow of the information is very complex including the steps such as
   a. Data acquisition: It includes the processes of accessing of data, setting parameters, transformation, data cleansing, data quality etc.
   b. Data modeling: It includes the logical model, linking with other data etc.
   c. Data mining: It includes the operations of mining and variable identification, algorithm selection, validation etc.
   d. Visualization: It includes the customization of business oriented information, customized reporting, advanced graphics etc.
2) Variety of Data model: The data models are very diverse and are tightly coupled with the type of analytics required for business.
3) Advanced and complex analytics: The business requirements may be very specific, thus there is demand of highly complex and advanced analytics (such as Machine learning) with data scientist in the organization.
4) Data volume: The DAaaS is working with big data, obvious that there will be challenge with data volume. Processing the big data in tha cloud is very difficult. So, computation is brought to the data instead of moving the data in cloud.
5) Analytics in Real-time: The analytics is more useful for business when used in realtime. There is research going on the real time analytics in the cloud environment.
6) Security in DAaaS: The security is a very complex issue in cloud computing. For organizations who don't want to put their data in cloud, can use private cloud instead. So that the security and confidentiality of the data will be maintained.
7) Infrastructure: The existing Analytics and Big Data platforms that use the local infrastructure need to change their technology to adopt a cloud infrastructure.

VI.DAaaS IN HYBRID CLOUD ENVIRONMENT

There are various complex issues in the implementation of a functionally of complete Big Data Analytics solution. This DAaaS solution minimizes this technical complexity when properly deployed to manage a hybrid cloud model. [6] When the public cloud is used for DAaaS, there are some issues:

   a) As the data is very huge and rapidly growing, it is not possible to move data from clients to the DAaaS infrastructure.
   b) Security and confidentiality issues are there while using public cloud. So, to impose such customization or requirement, we introduce the private cloud approach into it. Thus, it will become Hybrid Cloud which will fulfill the client-user as well as DAaaS requirements.

The approach is based in the work being done internally by Atos [5] Scientific Community in the concept of Cloud Services Brokerage6 and can be summarized in the following figure:
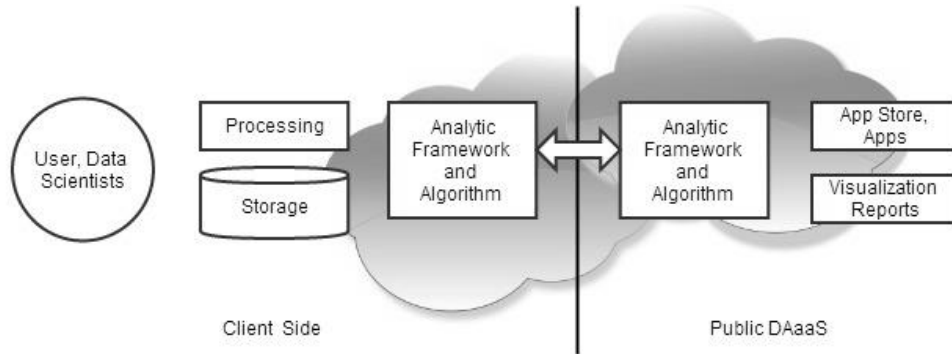
Fig. 3: DAaaS working in Hybrid Cloud model by Atos

In this model as shown in fig. 3, the architecture consists of customer premises and Public DAaaS. The customer premises consist of the Data Storage and Processing Layers. The data processing takes place on customer's premises such as Hadoop Map-Reduce processes. They are using higher level elements of DAaaS available and the Analytic Services present in the Public DAaaS solution. So, this Hybrid cloud provides flexibility [7] of deployment, integrity and data transfer and it is kind of Platform as a service of cloud.

VII. CLOUD SERVICE TYPES FOR DAaaS

The DAaaS can be deployed in the cloud depending on various service types, depending upon the user needs and resource availability such as public or private cloud environment. Using the basic cloud services types for analytics as a service include such as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) [8].

1) **Infrastructure as a Service (IaaS)**

It provides the on demand access to computing resources in a virtualized or cloud environment. It may include the virtualized servers, storage, and networks etc. i.e. computing infrastructure. [9] It is deployed on-user premise or via a cloud provider, Iaas can be virtualized, to handle the computing and storage needs for big data analytics. The Cloud operating systems manage high-performance servers, network, and storage resources. However, IaaS also requires greater investment of IT resources in the context of implementing big data analytics. Companies for example, Amazon Web Services, Citrix Cloud Platform are using this technology.

2) **Platform as a Service (PaaS)**

It delivers the computing platform. It may include operating system, compilers, database, and web server etc. Application developers can develop and run their software solutions on a cloud platform. [10] The PaaS provides the tools and libraries to build, test, deploy, and run applications on cloud. It doesn't need configure and scale elements of your implementation. It provides the development platform for advanced analytics applications. Companies for example, Google App Engine, Red Hat are using this technology.

3) **Software as a Service (SaaS)**

In this model, the application software in deployed in the cloud and users are allowed to use it. The platform and infrastructure is not need to be managed by the user. [11] The SaaS provides the specific applications for cloud-based big data analytics. According to the business scenarios and requirement, more than

one application can be used. SaaS can be a standalone solution. For example, Karma sphere offers a pay-as-you-go application that analyzes data stored with Amazon* S3 using Amazon Elastic Map Reduce.

VIII. CONCLUSION

Big Data is very fast growing, huge volume of data. It is a business requirement to analyze the data. The data is so huge that there are concerns to analyze the data at enterprise level for some organization. The solution called Data Analytics as a Service has been introduced, which is more flexible and uses cloud delivery model. In this paper, we discussed the concept of Data Analytics as a Service, various components, the benefits of DAaaS, Challenges and the DAaaS using hybrid cloud (Atos Model). The model given can provide flexible analytic solutions and taking advantage of Cloud technology in Big Data Analytics. The Big Data analytics with cloud models could be adopted by many companies as it is cost savings, with business operations and competitive advantage.

REFERENCES

[1] Peter Mell, Timothy Grance, "*The NIST Definition of Cloud Computing*", National Institute of Standards and Technology. Retrieved 24 July 2011.

[2] Lee, Jay; Lapira, Edzel; Bagheri, Behrad; Kao, Hung-An (2013). "*Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment*". Manufacturing Letters 1 (1). doi:10.1016/j.mfglet.2013.09.005

[3] "*Data, data everywhere*". The Economist. 25 February 2010. Retrieved 9 December 2012.

[4] "*The Road To Analytics As A Service*", Forbes Brand Voice, Oracle Voice, 26 September 2014

[5] "*Data Analytics as a Service: unleashing the power of Cloud and Big Data*", Ascent white paper, March 2013.

[6] Alycia Sebastian, Dr. L. Arockiam, "*A Study On Data Security Issues In Public Cloud*", *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 5*, May 2014

[7] "*The Bumpy Road to Private Clouds*". Retrieved 2014-10-08.

[8] "*Solution Brief Big Data in the Cloud: Converging Technologies*" Intel IT Center, September 2014 0914 / RPC / ME/ PDF - USA 328762-001

[9] Amies, Alex; Sluiman, Harm; Tong, Qiang Guo; Liu, Guo Ning (July 2012). "*Infrastructure as Service Cloud Concepts*". Developing and Hosting Applications on the Cloud. IBM Press. ISBN 978-0-13-306684-5.

[10] Boniface, M. et al. (2010), Platform-as-a-Service Architecture for Real-Time Quality of Service Management in Clouds, 5th International Conference on Internet and Web Applications and Services (ICIW), Barcelona, Spain: IEEE, pp. 155–160, doi:10.1109/ICIW.2010.91