IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ICRAESIT - 2015

B V Raju Institute of Technology, Narsapur

Vol. 2, Special Issue 2, December 2015

Pursuance Model for Cloud Analytics (PMCA) – An Analysis

S.Hemanth Chowdary[1], Sujoy Bhattacharya[2], K.V.V.Satyanarayana[3]

Koneru Lakshmaiah Education Foundation (KLU), Vaddeswaram, Guntur, India[1]

Guide Lane, Hyderabad, India[2]

KL University, Vaddeswaram, Guntur, India[3]

Abstract: In today's digitally activated world, Analytics is an indispensable tool for most businesses to get insights from past business activity data and use it to drive the planning and decision making for the future. For this reason, in the last decade or so, data analytics has come to be recognized as an integral operation in business execution. As easily accessible Internet connectivity gave rise to the creation of huge amount of data, to the tune of ½ billion GB every day, the presence of specialized Analytics tools became necessary. However, the field of Big Data Analytics is still emerging and a number of nascent tools are available in the market to cater to various needs. This has created a confusing environment, where specific applicability of such tools is sometimes hazy. In this paper, we have attempted to bring in clarity in this field by analyzing some typical workflows of the analytics process, along with a comparison of different types of analytics like Business Analytics, Cloud Analytics and Real Time Analytics.

Keywords: Cloud computing; Big Data, Analytics; Cloud Analytics; Business Analytics; Real-Time Analytics.

I. INTRODUCTION

Cloud computing is one of the most widely used technologies in the present digital era. Cloud computing is used in one form or the other by most of the industries. Some companies use it directly (in the form of Private Clouds or through a provider like Google / Amazon) and some indirectly (in the form of applications like Twitter). Such usages of the cloud are slowly replacing traditional in-campus applications. The main reasons behind the popularity of the cloud in today's business could be any of the following: (a) Reduction in operating costs, (b) Universal access, (c) Up to date software, (d) Wider choice of Applications or (e) Increased flexibility [1]. As a cloud environment usually handles large volumes of data that cannot be analyzed using traditional Business Intelligence tools, Big Data Analytics plays a major role in revealing meaningful information.

There are different types of analytics requirements for analyzing different types of data in different environments. Some examples include Cloud Analytics, Data Analytics, Business Analytics, Real-time Analytics and so on. In the present environment, various tools are available in the market for each of these Analytics requirements [2]. However, all tools typically aim to automate the analytics process using various statistical algorithms.

Cloud Computing is internet based computing technology in which large groups of remote servers work within a network to allow centralized data storage and online access to computing services or resources. A cloud can be A) Public B) Private or a C) Hybrid Cloud. These cloud environments offer different types of services like Platform as a Service, Infrastructure as a Service, Software as a Service, Data as a Service etc.

In Software as a Service (SaaS) model, users are provided access to a centrally installed application software and its backend databases from a remote location through a secure internet connection. Examples for SaaS include CRM, E-mail and Virtual Desktop applications. Through this model, companies are spared the burden of buying and installing an in-house version of the software. Similarly, Infrastructure as a Service provides storage space or computing engines on a shared basis and so on.

II. ANALYTICS

Analytics is the process of discovering and transferring data in the form of meaningful patterns. Software analytics utilizes the data-driven approach to enable software professionals to explore and unearth the information content from the data and use this perceptive information for completing various tasks around software systems, software users and Software Development Processes (SDP). A subset of all data analytics, Cloud Analytics enables us to visualize, forecast and optimize costs across the entire portfolio of public and private clouds.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ICRAESIT - 2015

B V Raju Institute of Technology, Narsapur

Vol. 2, Special Issue 2, December 2015

Analytics as a Service (AaaS) provides customers with analytics in an on-demand basis. The users pay for the usage of analytic solutions provided as a service by Cloud Service Providers (CSP). Reputed CSP's have multiple analytics platforms at their disposal and provide analytic solutions as a service for a specific purpose.

The cost of analytics can be calculated based on the size of the business (small, medium or large). For a small scale (an example would be a company with 1 store, 1 market and 1 product) business it is easy to calculate and analyze the transactions. For medium scale businesses, there may be a million transactions taking place per month, Data analysis for such companies can be done using Microsoft Excel 2010 or traditional Databases / BI tools, which can store a few million rows of data. For large scale business like Walmart, which has close to 9000 stores across the world, it is very difficult to analyze the transactions using Excel or by using traditional databases and tools. In a typical approach, such companies used to productize the analysis and do it for 1 product at a time. The problem is more complex for Tesco, which has 18000 stores across the world and 35 million Club Card customers. Similar problems are there for Airtel (which has 261 million subscribers, 8 billion calls going through its systems everyday), Amazon and Flipkart (which has several million customers across the world and has billions of transactions) happening per month. Financial Services like VISA (which has 100 million transactions a day) and FICO (which is used for fraud detection, offers protection on over 2 billion accounts). Usage of RDBMS and BI tools make the task incredibly complex for such large organizations. Thus, they look for alternative ways to efficiently store and analyze their data.

Data flow statistics:

- 1 million GB of data is processed by Google every hour.
- 30 billion updates are shared by Face book users per month.
- 2.8 trillion Rows of data with sprint.
- 300 trillion entries in Google databases.

Analytics can help in all industry types including Banks, Retailers, Telecom Companies, Health Care and Sports to name a few.

III. ANALYTICAL MODEL

While constructing any analytical model we have to consider two important characteristics.

- Volume of data supported
- Pursuance of the model.

In the analytical modeling process, clients get connected to the applications (installed on servers located remotely at the provider's premises) using an internet-enabled cloud environment and the applications in turn access (usually) a distributed database system. These database systems are designed to handle multiuser interaction and provide a warehouse of relational data that can be read and modified concurrently by various users and applications. Different users access the database in parallel by spawning multiple client processes. The applications allow its clients to group data access into transactions and offers access control to prevent unauthorized access. Examples of traditional data systems are DB2 or ORACLE, but the new-age non-relational No*SQL databases are now gaining popularity with cloud applications. A typical cloud structure model for an analytics system is given in Fig.1.

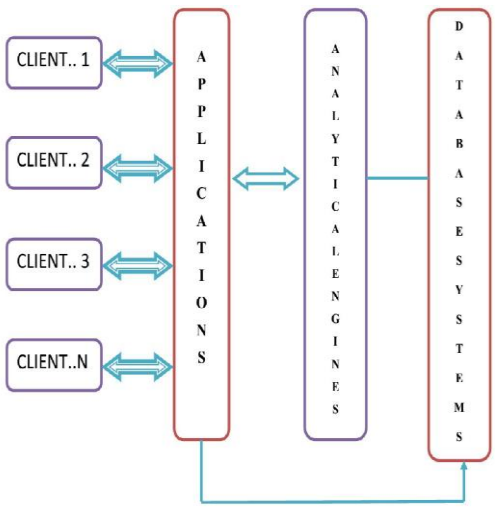

Fig. 1. Analytical Model For Data Analytics

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ICRAESIT - 2015

B V Raju Institute of Technology, Narsapur

Vol. 2, Special Issue 2, December 2015

The analytic engine provides the Online Analytical Processing (OLAP) functions and takes up services like aggregation, calculations, analysis process, caching, security and planning services (functions and sequences). Examples of the analytical engines are SPSS (IBM), SAS (SAS) and R (Revolution Analytics).

For any business analytics system, some of the characteristics that must be guaranteed for effective analysis of data are:

- Scalability
- Data I/O performance
- Fault tolerance
- Real-time processing
- Data size support
- Iterative task support

The various functions of such a generic analytic tool for Cloud Analytics would be:

Cloud Analytics: Cloud analytics enables us to visualize, forecast and optimize costs in both the public and private clouds. The expected result of such analytics is to minimize the cost associated with the upkeep of the cloud but maximize its efficiency at the same time.

Visualize: This function enables the cloud service provider to track budgets at the team, department and user levels. The cloud service administrator is then able to allocate, spend and chargeback based on the components that are related to the team or department or user. It also helps to analyze and save views in order to identify and understand cost drivers.

Forecast: An analytics system analyzes existing data and comes up with an intelligent forecast based on resources and information across the accounts, clouds and applications. The analytics also provides an insight to the patterns generated by changes, new apps, and actual business growth.

Optimize: The analytics system ultimately helps the provider / user the ability to identify unused resources and use them effectively to optimize resource usage.

Examples of cloud analytics products and services include hosted data-warehouses which is a centralized repository for enterprise data that is made available to users from a remote operation located by the service provider, as against being located on the local systems. Software-as-a-Service business intelligence (SaaS BI) involves delivery of business intelligence applications to end users from a hosted location and cloud based media analytics involves the remote provisioning of tools.

IV. TOOLS

Different tools for analytics are available based on data volumes and complexity. The following figure (Fig. 2.) illustrates different types of tools that can be used based on the calculation complexities and data volumes ranging from low to high.

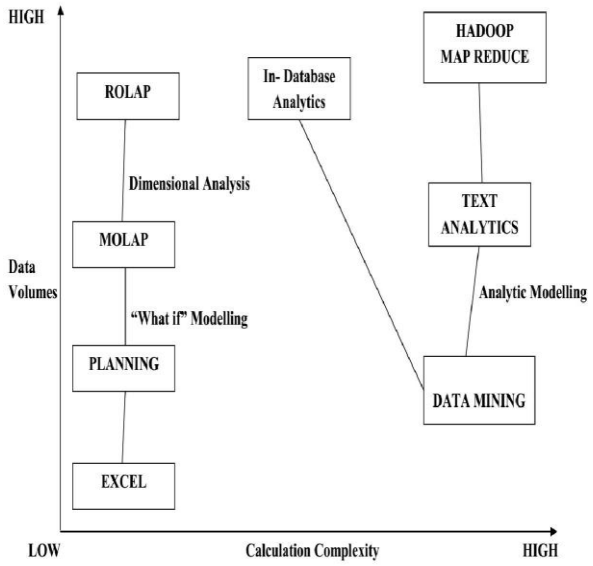

Fig. 2. Different Tools Used Based on Data Volumes & Calculation Complexity

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ICRAESIT - 2015

B V Raju Institute of Technology, Narsapur

Vol. 2, Special Issue 2, December 2015

V. REAL TIME ANALYTICS

As the name suggests, Real Time Analytics deals with the capacity to use and analyze data that is generated recently and the usage of resources as per current requirements. The analysis is typically dynamic in nature and updates itself in real time based on the actual use. Real time analytics is sometimes referred to as real time data analytics, real time data integration or real time intelligence.

Technologies that support real-time analytics include:
In-Database Analytics: Data processing and Analytic logic is inbuilt inside the database and the analysis is performed in the database environment without the use of external resources.
Data Warehouse Appliance: It is a kind of appliance which allows the client/customer to setup a high performance data-warehouse. Some products are designed particularly with the combination of software and hardware for analytical processing.

A typical example of real-time data analytics is the use of Cloud-based CRM (Customer Relationship Management) software, where the customer calls and the data generated needs to be analyzed then and there and business decisions need to be made based on the enhanced information. For example, a flurry on complaints on a newly launched product can be used as a trigger to stop selling that product and resolve the problem to keep goodwill.

VI. CONCLUSION

An established best practice for analytics users is to put in place processes and structures within the organization and choose the best tools to solve high-value business problems. It is also important to make continuous refinements to the existing processes and deploy new, efficient tools to support growing volumes of data. This paper provides an overview into different types of analytics and their importance along with different analytical tools that are used based on the data volumes and calculation complexity.

REFERENCES

[1] John W. Rittinghouse, James F. Ransome, "Cloud Computing Implementation, Management, and Security", 978-1-4398-680-7
[2] P. RadhaKrishna and Kishore Indukuri Varma, "Cloud Analytics A Path Towards Next Generation Affordable BI", Infosys White Paper
[3] Vignesh Prajapati, "Big Data Analytics with R and Hadoop", 978-1-78216-328-2
[4] Cruz, S., Campos, M., and Mattoso , M., 2009. Towards a Taxonomy of Provenance in Scientific Workflow Management Systems", in proceedings of IEEE Congress on Services– I, Los Angeles, CA, USA, pp.259-266,July 2009.
[5] Mike Barlow, "Real-Time Big Data Analytics: Emerging Architecture", 978-1-449-36421-2
[6] SAS, http://www.sas.com/en_us/software/cloud- analytics.htm
[7] Cognizant Technology Solutions, http://www.cognizant.com/enterprise-analytics/big-data

BIOGRAPHIES

S. Hemanth Chowdary is currently pursuing his PhD from Koneru Lakshmaiah Education Foundation (KLU) in the area of Big Data Analytics. He obtained his Masters degree from JNTUH. His current research interests include Big Data Analytics and Cloud Computing.

Dr. Sujoy Bhattacharya received his B Tech and PhD degrees from IIT Kharagpur in 1990 and 1997 respectively in the Department of Electrical Engineering. As a professional, he has worked with a reputed IT consultancy firm for a period of 16 years and in the field of education and consultancy for a period of 5 years. He currently works as an independent consultant on technical and vocational subjects for educational institutes. His current research interests include Cloud Computing, Image Processing and Education / Leadership.

Dr K. V.V. Satyanarayana is working as Professor in the Department of Computer Science & Engineering since 2012 with specialization areas in Bioinformatics and Cloud Computing. He has 17+ years of teaching experience both in UG and PG engineering courses. He obtained his Engineering Masters degree from JNTUK-Kakinada and Doctoral Degree from Acharya Nagarjuna University. He worked as Director and Head of the Department at other institutions.