International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 2, Feb.-2015, pp: (242-245), Impact Factor: 1.252, Available online at: www.erpublications.com

Cloud Analytics - Path towards Next Generation

Kush Sehgal[1], Mr. Lokesh[2]

[1]Kush Sehgal, M.Tech. Student, RIEM, Rohtak, Haryana
[2]HOD, CSE Dept. RIEM, Rohtak, Haryana

ABSTRACT: Technology innovation and its adoption are two critical successful factors for any business/organization. Cloud computing is a recent technology paradigm that enables organizations or individuals to share various services in a seamless and cost-effective manner. Business Intelligence for organizations, on the other hand, is becoming a growing need to understand their business insights and trends. Currently organizations are trying to leverage BI to maximum extent; however, there is a big gap in turning BI outcome to aid their ROI expectations and further growth. This necessitates porting current data analytic applications on to the cloud due to its ability to process large datasets as well as extensive support for scalability at low cost. This article brings out the technology challenges and opportunities to enable analytics in cloud environment, which makes BI affordable for all organizations.

CLOUD PARADIGM

Cloud computing is an emerging computing paradigm that was innovated to deploy cost effective solutions over Internet. Companies such as Google, IBM, Amazon, Yahoo and Intel have already started providing computing infrastructures for its intend use. Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. It is a scalable service delivery platform build over Service Oriented Architecture (SOA) and Virtualizations concepts. The benefit of cloud can be stated in two perspectives: (i) from cloud service provider perspective and (ii) from cloud user perspective. Cloud service providers get benefited with the better utilization of infrastructure they own. Even, they can obtain an improvement in the processing power with minimal additional cost. Thus, giving way for new business opportunities and speed up existing business processes. Users get benefited by paying only for what they consume and even they need not have insights of the technology infrastructure used in the "cloud" that supports them. Data analytics applications which require processing extremely large datasets while support extreme scalability are attracting huge interest towards cloud.

The cloud computing generally incorporates combinations of the following:

- Infrastructure as a service (IaaS) - Providing servers, software, office space and network equipment as service.
- Platform as a service (PaaS) - Providing computing platform and solution stack as a service.
- Software as a service (SaaS) - Providing an application as a service.

Cloud consists of huge number of servers facilitating distributed computing platform. Each server can have multiple virtual machine instances (VMs) of the applications hosted in the cloud. As customer demand for those applications changes, new servers are added to the cloud or idled and new VMs are instantiated or terminated. A major advantage with cloud is the recoverability in case of disaster. Cloud data is mostly backed up regularly and stored in a different geographic location. This eases the recovery process which otherwise is a very costly feature for companies. Cloud computing infrastructure is totally different from the current data warehouse infrastructures. Unlike the high-end servers which constitute the current infrastructure, Cloud computing offers the same functionality with low-cost. Since most analytics for an organization uses huge number of database reads than writes, there is a requirement for new relational database software architecture to efficiently store and retrieve large volumes of data for BI on Cloud.

CLOUD ANALYTICS

Few decades back, the problem was the shortage in information or data. In recent past, this problem has been overcome with the advent of Internet and reduced Storage Memory cost. But a new challenge is how to analyze the data. Data is getting generated at a much faster pace than the speed at which it can be processed with the current infrastructure. Huge

and dedicated servers were developed to solve this problem. But the problem is with the cost of such an infrastructure which is not affordable to all the companies for Availability of data and accessing each and every specific purpose. So today, these companies are looking it is the key success factor for Cloud computing which makes feasible for all these companies to hire on a temporary basis, the computational power and storage space value-based analytics. Migrating for a specific purpose. This is termed as "Utility Computing" derived from required data in and out of Cloud the utilities like electricity, gas for which we only pay for what we use from a shared resource. With the growing interest in cloud, analytics is a challenging task over cloud has seen surge in interest. Recent reports illustrate this with various architectures and analytic services possible over cloud.

Analytics for cloud:

Analytics has its applications for cloud economics also:

i. Resource optimization: Helps in optimally scheduling resources available based on the nature of nodes waiting to be allocated and their cost factor.
ii. Demand Forecasting: Helps the Architects and Managers forecast the demand and act accordingly. Estimations can even help in pricing the service which will bring in noticeable profits while minimizing expenses in terms of new hardware or infrastructure to meet the forecasted demand.
iii. Billing strategy based on Seasonal analysis can even be learned into an analytics model which can automatically quote the price for any time period and computing service requirements.
iv. Identifying rarely accessed data and moving it to highly compressed and low cost devices.

Two-kinds of services which can be visualized for cloud analytics are:

(a) Analytics as a Service (AaaS)

Analytics as a Service provides clients with analytics on demand. They pay for the usage of the analytic solutions as a service by CSP. The idea here is that, CSP lists analytic solutions as a service and customers pick required solutions and leverage it for their specific purposes. For example, a customer can select sentiment analysis as a service, which helps him/her analyze their customer's sentiment on his/her products Vs competitor's products. This helps in drafting product roadmaps based on market analysis. The cost incurred in the traditional process is very high as a company has to plan its course from scratch. Instead providing AaaS will be highly cost-effective as the data is publicly available on web as blogs or review articles.

(b) Models as a Service (MaaS)

Models as a Service provide clients with building blocks to develop their analytical solutions by subscribing to the models available over a cloud. Various models which have extremely high CPU and memory usage tasks (e.g., Clustering models like SVM, Neural Nets, Bayesian models) can be ported to a cloud.

ISSUES AND CHALLENGES

Though cloud computing is on the verge of becoming a reality, there are several issues and challenges. Few of them are described below from analytics point of view:

(a) Tuning Knowledge models

Tuning existing models for a particular application being served over the cloud will enable the clients to leverage existing skill sets to customize the model to their particular requirement.

(b) Ensuring Privacy

The data or knowledge Analytic services may be processed outside the client's premises. So, one needs to ensure the privacy of the data as well as knowledge derived as a service.

(c) Supporting Column Based Indexing

Most of the current relational databases being used by the clients are row-based. However, the analytics over cloud which accesses data on column at a time motivate one to opt for column oriented databases in place of the row oriented databases. Existing models have to be augmented to support column-oriented databases.

(d) Ensuring Data Availability

Data analytics are data intensive applications and hence required new mechanisms when one or more nodes of a cloud fail. Cloud should have the capability to recover and progress in the event of multiple node failures. Specialized file systems are needed for cloud units to handle such failures.

(e) Embedding Analytical models inside Databases

Analytics are also computational intensive. Moving the analytical computations inside database engine paves way for better performance and reduced cycle time. These features also help in utilizing the parallel-processing capabilities of database engine and converting cloud databases into active cloud databases.

(f) Ensuring Data Quality

It is a reported fact that, in P2P systems, the output of a query is most of the times incomplete due to poor quality of data. Adopting similar situations over the cloud, the data residing on multiple and highly distributed processing units creates poor data quality that may not be suitable for analytics.

(g) Ensuring Data Currency

In Data Analytic applications, by the time models are generated from the data that is available, there might be more recent data on which the model built might stay invalid. Cloud can tackle this problem by reducing the cycle time between data availability and model generation. Specialized algorithms which enable fast incremental learning model generation over cloud need to be explored.

(h) Enabling Knowledge Process flow

Knowledge process can move from one active node in the cloud to the other, extracting enough knowledge from it. These knowledge instances learned need to be consolidated from time to time to infer Knowledge on the cloud.

PUBLIC vs PRIVATE CLOUD

Clouds, synonymous to Public Clouds, facilitate sharing available computing resources to multiple businesses. One of the major issues with cloud computing today is the trust in security and privacy of business data. To overcome this issue, organizations are building their own clouds, referred to as Private Clouds. Though Private Clouds are costlier since the business owns the resources, they make organization to relax from unnecessary stress of dealing with the unknown. As standards around security and privacy are becoming stronger, we may envisage that public clouds will slowly increase their impact and importance to businesses in terms of trust. We can visualize the following scenarios from analytics point of view:

Data Private, Model Private- Analytic services are used for company's internal data analysis and may form a secret to the outsiders. This introduces the need for Data to be private as well as the Model to be Private. This scenario can be handled purely with private clouds. Data Private, Model Private- Analytic services are used for company's internal data analysis and may form a secret to the outsiders. This introduces the need for Data to be private as well as the Model to be Private. This scenario can be handled purely with private clouds.

Data Public, Model Private- Service Applications like web crawlers depend on data over a cloud which gets data from public resources and leverages this data for analytics internal to the business. For example, a crawler can crawl the web to meet together various service requirements by different clients in different applications and provide the analytics or statistics requested as a service.

Data Public, Model Public - This is a trivial case of cloud computing, for instance, blogs are now being widely used by companies to advertise their products and ask for reviews, web blogs managed by Industry verticals or specific companies are usually public along with their content. Data Private, Model Public - This is a classic case for Model as a Service. In this scenario, private data (stored in a private cloud or enterprise) analysis is done using a public model residing over the cloud. For example, the standard financial models can be used from the public cloud to analyze the financial data which is confidential to the business.

CLOUD VIEWS

'Cloud Views', introduced in this article, is a solution that overcomes the limitations of security, privacy, consistency and compliance aspects of cloud management. The concept Cloud Views evolved from the Views which can be created over a relation database. A Cloud View similar to the database views, abstracts the physical schema of the cloud limiting the degree to which the actual infrastructure is exposed to the outer world. Cloud Views are either virtual or materialized views created in a cloud environment. Each cloud view abstracts the required functionality to share cloud infrastructure. Unlike database views and workflow views, cloud views abstracts the data, application, infrastructure and all other computing resources provided over a cloud. Moreover, like other views, cloud views enable data access, privacy and security.
The Cloud views, when developed with industry-specific standards, can create a logical infrastructure which eliminates the need for a Intranet based cloud for any company. Cloud views also enable semantic definition of the application requirements and also facilitate manifestation and support of application requirements in cloud environment.

Cloud Views facilitate the following:

i. Privacy and Security
ii. Performance
iii. Ease of application development

Cloud views can be created incorporating either a community-wide (such as cloud view for financial industry) or a corporate-wide model for adopting industry-specific standards as well as corporate-specific standards respectively. One point of view is the creation of a community-wide cloud is to provide cloud views for a specific industry/community at low-cost, high reliability and heavy security mode. One of the major limitation of adopting cloud computing is the lack of stringent standards around usage of cloud services. However, using cloud views, the existing standards can be adopted and deployed easily on respective cloud view. The second advantage of having these views is that the policies can be maintained at view level, rather than at cloud level as it is difficult to maintain different policies on the same cloud, which may show redundant and arising conflicts.

CONCLUSIONS

Cloud computing is a recent technology paradigm that most of the infrastructure and services industries are focusing to capture potential opportunities. Analytic over cloud enables organizations to realize their analytical needs in a more affordable manner. This paper explores various technological perspectives for cloud analytics and various cloud services that can be envisaged in future.

REFERENCES

[1] Ananthanarayanan R., Gupta K., Pandey P., Himabindu P., Sarkar P., Shah M., and Renu Tewari R., Cloud Analytics: Do We Really Need to Reinvent the Storage Stack?, (http://www.usenix.org/ event/hotcloud09/tech/full_ papers /ananthanarayanan.pdf).
[2] Armbrust M., Fox A., Griffith R., Joseph A. D., Katz R. H., Konwinski A., Lee G., Patterson D. A., Rabkin A., Stoica I., and Zaharia M., Above the clouds: A Berkeley view of Cloud computing, Technical report UCB/EECS-2009-28, Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, USA, February 2009.
[3] Cerri, D., Valle, E., Francisco Marcos, D., Giunchiglia, F., Naor, D., Nixon, L., Teymourian, K., Obermeier, P., Rebholz-Schuhmann, D., Krummenacher, R., and Simperl, E., Towards Knowledge in the Cloud, The OTM Confederated international Workshops and Posters on the Move To Meaningful internet Systems, Spain, 2008.
[4] Dean J and Ghemawat S., MapReduce: Simplified data processing on large clusters. Sixth Symposium on Operating System Design and Implementation, pp. 137–150, 2004.
[5] http://en.wikipedia.org/
[6] http://www.vertica.com/
[7] http://www.forrester.com/rb/Research/in-database_analytics_heart_of_predictive_enterprise/q/id/48289/t/2