

Enterprise Cloud Analytics

K.Vasuki[#] S.M.Srihari Shankar^{*}

[#]PG Scholar, Department of CSE, Sri Shakthi Institute of Engineering and Technology

^{*}Assistant Professor, Department of CSE, Sri Shakthi Institute of Engineering and Technology.

Abstract: Cloud computing revolutionize IT and business by offering computing as a utility over the internet. The evolution from internet to a cloud computing platform, the emerging development paradigm and technology and how these will change the way enterprise applications should be architected for cloud deployment plays an important role but these enterprise technologies are critical to cloud computing. New cloud analytics and business intelligence (BI) services can help businesses (organizations) better manage big data and cloud applications. Analysing and gathering business intelligence (BI) has never been easy, but today BI is complicated further by overwhelming amounts of data loads and the number of data entry and access points. New cloud analytics advancements may offer BI relief and even profit-increasing predictability for enterprises. These new cloud analytics applications can deliver functional capabilities that can be easily, quickly and economically deployed, producing tangible and measurable benefits far more rapidly than in the past. Many organizations that recognized, effectively analysing their business needs and providing the data they require to make the right business decisions depends on a combination of internally generated data and externally available data.

Keywords: Enterprise clouds, Business Intelligence, Analytics, business models.

I. Introduction

The term “Cloud computing” is internet-based computing in which large groups of remote servers are networked to allow the centralized data storage, and online access to computer services or resources. Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility (like the electricity grid) over a network. The foundation of cloud computing is the broader concept of converged infrastructure and shared services.

The main focus is to maximize the effectiveness of the shared resources. Cloud resources are usually not only shared by multiple users but are also dynamically reallocated per demand. This approach should maximize the use of computing power thus reducing environmental damage as well since less power, air conditioning, rackspace, etc. are required for a variety of functions. With cloud computing, multiple users can access a single server to retrieve and update their data without purchasing licenses for different applications.

The term "moving to cloud" refers to an organization moving away from a traditional CAPEX model (buy the dedicated hardware and depreciate it over a period of time) to the OPEX model (use a shared cloud infrastructure and pay as one uses it). The availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing have led to a growth in cloud computing. If the business is concerned about having the required skills and resources for installing and managing the business analytics infrastructure and software, then a cloud alternative business analytics applications that are now available in cloud-based or software as a service (SaaS) editions can be considered. The business analytics deployments reduce the costs and risks of development and deployment [1, 14].

The business analytics software in general provides the following analytics capabilities: Customer Analytics, Predictive analytics, operations analytics, Risk analytics and financial analytics.

Strategic workforce and Talent management analytics creates competitive advantage through deeper workforce insights based on market-proven HR best practices, prebuilt dashboards, reports, and rich metrics. For instance, Data-Driven Companies trust the Scale and Speed of Wave to know their customer in a whole new way. The world’s most data-driven companies, including Equinix, GE Capital and TriCore Reference Laboratories, trust the scale and speed of Wave to know their customers in a whole new way. Any company can now build custom analytics apps for its customers to analyze patterns, discover new trends and deliver actionable insights. Salesforce, the Customer Success Platform and world's #1 CRM empowers companies to connect with their customers in a whole new way.

II. Architecture

The architecture of enterprise on cloud with analytics is as shown in the figure below, it has five major components/layers: layer1, the lowermost layer is the IaaS (Infrastructure as a Service) offered by any of the open source providers such as openstack, CloudStack or the public providers such as Amazon web services. Above this layer is the layer2, referred to as PaaS (Platform as a Service) offered by any of the service providers such as cloud Foundry, Heroku or Google App Engine. These two layers are integrated by the API referred to as IPaaS API.



Fig 1 Architecture of Enterprise on cloud with Analytics

Above this layer is the BPaaS (Business Process as a Service) which is used to perform business process workflow monitoring, either by programming in an eclipse IDE or by accessing it through the web Console. This Business process Services include JBPM, MuleESB, Activiti, Drools rules engine, Wso2 BPEL. The topmost layer is the SaaS (Software as a Service) through which the cloud consumers/users access the Business process Services. This Service is offered in web services or Sugar CRM.

Another important component of this architecture is the analytics of the data gathered during each of the business processes done by Cloudera, MapR, HortonWorks H2O, talend, or pentaho.

The IaaS (Infrastructure as a Service) provides the basic cloud-service model. The providers of IaaS offer computers – physical as well as virtual machines and other resources. The examples of hypervisors include Xen hypervisor, Oracle Virtual Box, KVM (Kernel Virtual Machines), VMware ESX/ESXi, or Hyper-V. These hypervisors runs the virtual machines as guests. IaaS clouds offer additional resources such as a virtual-machine disk image library, raw block storage, and file or object storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles.

In the PaaS models, cloud providers deliver a computing platform, typically operating system, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers. In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis. SaaS providers generally price applications using a subscription fee. Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user, who sees only a single access point.

In the UCaaS (Unified Communication as a Service) model, multi-platform communications over the network are packaged by the service provider. The services could be in different devices, such as computers and mobile devices. Services may include IP telephony, unified messaging, video conferencing and mobile extension etc.

At the Infrastructure as a Service (IaaS) layer, the advancements of companies like Amazon, Rackspace, Google and others bring us an exponentially more economical data capture and storage capability. That is needed to handle the sheer growth of data that has overwhelmed many organizations, organizations that can't afford the luxury of rolling out huge server farms of their own. Now, [they can] pay-as-they-go for IaaS offerings. These aggregated data sources that the SaaS vendors are capturing can be converted into very meaningful benchmark statistics and KPIs. Finally, on the Platform as a Service (PaaS) cloud, we now have the ability to plug in business intelligence (BI) or analytic capabilities as we develop new apps. This exciting new opportunity wasn't available before the emergence of the cloud.

III. Architecture And Technology Descriptions

Cloud infrastructure services, known as "Infrastructure as a Service" (IaaS), deliver computer infrastructure (such as a platform virtualization environment), storage, and networking. Instead of having to purchase software, servers, or network equipment which increases the CAPEX, users can buy these as a fully outsourced service that is usually billed according to the amount of resources consumed. There are multiple IaaS service providers like openstack, amazon, CloudStack, etc.

A. Infrastructure as a Service

Apache **CloudStack** is open source software designed to deploy and manage large networks of virtual machines, as a highly available, highly scalable Infrastructure as a Service (IaaS) cloud computing platform. CloudStack is used by a number of service providers to offer public cloud services, and by many companies to provide an on-premises (private) cloud offering, or as part of a hybrid cloud solution. Users can manage their cloud with an easy to use Web interface, command line tools, and/or a full-featured Restful API. CloudStack provides an API that's compatible with AWS EC2 and S3 for organizations that wish to deploy hybrid clouds.

OpenStack is another free and open-source software cloud computing software platform. Users primarily deploy it as an infrastructure as a service (IaaS) solution. The technology consists of a series of interrelated projects that control pools of processing, storage, and networking resources throughout a datacentre—which users manage through a web-based dashboard, command-line tools, or a Restful API. OpenStack.org releases it under the terms of the Apache License.

OpenStack lets users deploy virtual machines and other instances which handle different tasks for managing a cloud environment on the fly. It makes horizontal scaling easy. OpenStack lets users to quickly add new instance, upon which other cloud components can run. Typically, the infrastructure then runs a "platform" upon which a developer can create software applications which are delivered to the end users.

B. Platform as a Service

The most complex of the three, cloud platform services or "Platform as a Service" (PaaS) deliver computational resources through a platform. PaaS makes the development, testing, and deployment of applications quick, simple, and cost-effective, eliminating the need to buy the underlying layers of hardware and software. There are multiple PaaS providers available like Google App Engine, Heroku, Windows Azure Cloud Services, Amazon AWS, Red Hat OpenShift.

Google App Engine is a Platform as a Service (PaaS) that lets users build and run applications on Google's infrastructure. App Engine applications are easy to build, easy to maintain, and easy to scale as the traffic and data storage needs change. With App Engine, there are no servers to maintain. Simply upload the application and it's ready to go.

Heroku is a cloud platform as a service (PaaS) supporting several programming languages. Heroku was acquired by Salesforce.com in 2010. Heroku, one of the first cloud platforms, has been in development since June 2007, when it supported only the Ruby programming language, but has since added support for Java, Node.js, Scala, Clojure, Python and PHP and (undocumented) Perl. The base operating system is Debian or, in the newest stack, the Debian-based Ubuntu. Heroku lets app developers spend their time on their application code, not managing servers, deployment, ongoing operations, or scaling

C. Software as a Service

In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis. A completely transformed, individualized CRM user experience that is immersive, engaging, and intuitive.

SugarUXfuses the straightforward simplicity, mobility, and social aspects of a consumer app with the business process optimization of conventional CRM.

D. Business Process as a Service

JBPM is a flexible Business Process Management (BPM) Suite. It makes the bridge between business analysts and developers. JBPM has a dual focus: it offers process management features in a way that both business users and developers like it. The core of jBPM is a light-weight, extensible workflow engine written in pure Java that allows users to execute business processes using the latest BPMN 2.0 specification. It can run in any Java environment, embedded in your application or as a service.

JBPM supports adaptive and dynamic processes that require flexibility to model complex, real-life situations that cannot easily be described using a rigid process. Complex business logic can be modelled as a combination of business processes with business rules and complex event processing. BPM is a flexible Business Process Management (BPM) Suite. It is light-weight, fully open-source (distributed under Apache license) and written in Java. It allows modelling, executing, and monitoring business processes throughout their life cycle. The **WSO2 Business Process Server** enables developers to easily deploy business processes written using the WS-BPEL standard, and also serves as the business process management and hosting environment for SOA. Powered by the Apache Orchestration Director Engine (ODE) BPEL engine, the WSO2 Business Process Server provides a complete web-based graphical console to deploy, manage and view processes. The key advantage is all WSO2 products are built on a common foundation - WSO2 Carbon, a modular, reconfigurable, elastic, OSGi-based architecture.

Drools are a Business Rules Management System (BRMS) solution. It provides a core Business Rules Engine (BRE), a web authoring and rules management application (Drools Workbench) and an Eclipse IDE plugin for core development. The term **Rule Engine** can be referred to any system that uses rules, in any form that can be applied to data to produce outcomes.

Activiti is a light-weight workflow and Business Process Management (BPM) Platform targeted at business people, developers and system admins. Its core is a super-fast and rock-solid BPMN2 process engine for Java. It's open-source and distributed under the Apache license. Activiti runs in any Java application, on a server, on a cluster or in the cloud. It integrates perfectly with spring; it is extremely lightweight and based on simple concepts. **Alfresco Activiti** provides a highly scalable, java based workflow and Business Process Management (BPM) platform targeted at business people, developers and administrators.

IV. Analytics

New cloud analytics and business intelligence (BI) services can help businesses better manage big data and cloud applications. Gathering and analysing business intelligence (BI) has never been easy, but today BI is complicated further by overwhelming data loads and the number of data entry and access points. New cloud analytics advances may offer BI relief and even profit-increasing predictability for enterprises. These new cloud analytics applications can deliver functional capabilities that can be easily, quickly and economically deployed, actually producing tangible and measurable benefits far more rapidly than in the past. More and more organizations that recognize that effectively analysing their business needs and providing the data they require to make the right business decisions depends on a combination of internally generated data and externally available data.

Cloudera Inc. is an American-based software company that provides Apache Hadoop-based software, support and services, and training to business customers. Cloudera's open-source Apache Hadoop distribution, CDH (Cloudera Distribution Including Apache Hadoop), targets enterprise-class deployments of that technology.

HortonWorks H2O is the open source in memory solution from Oxdata for predictive analytics on big data. It is a math and machine learning engine that brings distribution and parallelism to powerful algorithms that enable you to make better predictions and more accurate models faster. H2O brings better algorithms to big data. H2O is the open source math & machine learning platform for speed and scale. With H2O enterprises can use all of their data (instead of sampling) in real-time for better predictions. Data Scientists can take both simple & sophisticated models to production from the same interactive platform used for modelling, within R and JSON. H2O is also used as an algorithms library for Making Hadoop Do Math. Our earliest customers have built powerful domain specific predictive engines for Recommendations, Pricing and Outlier detection in Fraud & Insurance. H2O is nurturing a grassroots movement of math, systems and data scientists to herald the new wave of Discovery with Big Data Science. Oxdata is a provider of H2O.

The key to monetization of Big Data is not only the ability to capture and process information quickly but also to analyze the data to derive meaningful insights[3]. Big Data can be complex, and often the expertise of a programmer is needed to create focused and targeted queries. Oxdata, a provider of open source machine learning and predictive analytics for Big Data, helps to facilitate the manipulation and extraction of data with the

use of its H2O prediction engine for statisticians. The H2O tool helps reduce the need for programming and coding to model the data and supports the complete end-to-end analytical workflow.

MapR is a complete enterprise-grade distribution for Apache Hadoop. The MapR Distribution for Apache Hadoop has been engineered to improve Hadoop's reliability, performance, and ease of use. The MapR distribution provides a full Hadoop stack that includes the MapR File System (MapR-FS), MapReduce, a complete Hadoop ecosystem, and the MapR Control System user interface. The MapR Control System (MCS) provides a graphical control panel for cluster administration with all the functionality of the command-line or REST APIs. The MCS provides job monitoring metrics and helps troubleshoot issues, such as which jobs required the most memory in a given week or which events caused job and task failures.

V. Conclusion

Enterprise IT infrastructure largely predates the emergence of cloud computing as a viable choice for hosting data-driven applications. Large organizations are now showing real signs of adopting cloud computing for certain applications and a few forward-thinking enterprises are formulating the concept of data as a service, based on performing big-data analytics in the cloud[7]. However, exactly when big-data analytics will move to the cloud remains an open question. Compatibility, security, and performance concerns have kept enterprise organizations from being completely comfortable with the idea of moving their complex core applications to the cloud. Organizations which leverage analytics and testing to understand their customers' needs and behaviour, and transform their offering accordingly, are fast gaining market share and popularity[8]. Any organization or enterprise being data-driven is the art and ability to leverage all business assets to exercise judgment in the decision process. Large organizations are now showing real signs of adopting cloud computing for certain applications and a few forward-thinking enterprises are formulating the concept of data as a service, based on performing big-data analytics in the cloud. The report will give guidance to CxOs, IT and business leaders, and decision-makers at software-as-a-service (SaaS) companies and cloud service providers.

References

- [1]. George Mironescu, Nick Mayes, "Managing Enterprise Applications Workloads in the cloud", Analyst IT Services and cloud, IBM, UK, White Paper, June 2014.
- [2]. IBM Center for Applied Insights, "Raising the game: The IBM Business Tech Trends Study," August, 2014.
- [3]. Gartner, "Predicts 2014: Big Data," Gartner #G00258154, Nick Heudecker, Mark A. Beyer, Douglas Laney, Michele Cantara, Andrew White, RoxaneEdjlali, Andrew Lerner, Angela McIntyre, November 20, 2013.
- [4]. QuinStreet, "2014 Big Data Outlook: Big Data is Transformative – Where is Your Company?" 2014.
- [5]. Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018," February 5, 2014.
- [6]. IDC, "EMC Digital Universe Study, with Data and Analysis by IDC," April 2014.
- [7]. Gartner, "Predicts 2014: Infrastructure Protection," Gartner #G002147953, Ray Wagner, Kelly M. Kavanagh, Mark Nicolett, Anton Chuvakin, Andrew Walls, Joseph Feiman, Lawrence Orans, Ian Keene, November 25, 2013.
- [8]. IBM, "The IT infrastructure conversation: New content, new participants, new tone," July 2014.
- [9]. Gartner, "Mobile Device Proliferation Is Forcing Network Leaders to Redesign Enterprise Wireless LANS," Bjarne Munch and Christian Canales, May 19, 2014.
- [10]. EMC Digital Universe Study, with data and analysis by IDC, April 2014.
- [11]. Gartner, "Predicts 2014: Big Data," Nick Heudecker, Mark A. Beyer, Douglas Laney, Michele Cantara, Andrew White, RoxaneEdjlali, Andrew Lerner and Angela McIntyre, November 20, 2013.
- [12]. IDC, "Infrastructure and Cloud Services: Datacenter Rationalization and Cloudification—A Much-Needed Strategy Reset," IDC #247423, March 2014.
- [13]. Gartner, "Predicts 2014: Infrastructure Protection," Ray Wagner, Kelly M. Kavanagh, Mark Nicolett, Anton Chuvakin, Andrew Walls, Joseph Feiman, Lawrence Orans and Ian Keene, November 25, 2013.
- [14]. IDC, "Key Forces Shaping Datacenters in the 3rd Platform Era," IDC #240270, March 2013.