An Enterprise Cloud Model for Optimizing IT Infrastructure

Rabi Prasad Padhy*, Manas Ranjan Patra**
*Senior Software Engg, Oracle Corp., Bangalore, Karnataka, India
**PG Departement of Computer Science, Berhampur University, Odisha, India

ABSTRACT

In cloud computing, end users are not required to buy software or devices as they are provided by service providers on a rental basis. Cloud Computing is the new IT platform as it promises dramatic reductions in cost and time-to-market due to pay-per-use and scalability. A large number of enterprises are shifting their computing from in-house infrastructure to the cloud infrastructure to optimize their infrastructure. The new-coming generation of employees and customers expect technological proficiency to be part of the customer-centered business models. On their turn, organizations need to apply new forms of cooperation with their employees, customers, suppliers and partners. As the enterprise business system surrounding Cloud Computing develops, it is critical to have a conception of how cloud services pulling, the diverse types of service offerings, and the way those services fit together. To overcome this problem we proposed a service selection layered method for cloud based enterprise. In this paper, we propose an approach to select or pull multiple cloud services from cloud market place for an Enterprise. Our approach provides best service selection process which will suit for an Enterprise’s requirement. This approach also improves governance for the service provider while offering maximum flexibility for the consumers. Our aim is to provide a better understanding of the cloud service selection challenges in a Enterprise cloud computing systems.

1. INTRODUCTION TO CLOUD APPROACH

Today, more and more industry companies and organizations recognize the value and benefit when using cloud computing services. Reducing cost and maintaining scale and high availability are essential for the business to keep its continuity. Although cloud computing has been widely adopted by the industry, the research on cloud computing is still at an early stage. Cloud Infrastructure is the most important component in a cloud. It may comprise tens of thousands of servers, network devices and disks, and typically serve millions of users globally. Such a large-scale data center will consume considerable amount of energy. Cloud computing provides a pool of highly scalable and easily accessible virtualized resources (such as hardware, development platforms and services) capable of hosting end-user applications exploited in a pay-as-you-go model.
Datacenters in which these services are hosted, are sets of interconnected servers which generate huge computing power and virtualized environment that allows the dynamically deployment of applications at runtime. Generally, service-level agreements guard the provisioning of the cloud services.

The key attributes that distinguish cloud computing from traditional computing solutions have been identified in and generally comprise the following:

- Underlying infrastructure and software is abstracted and offered as a service.
- Build on a scalable and flexible infrastructure.
- Offers on-demand service provisioning and quality of service (QoS) guarantees.
- Pay for use of computing resources without up-front commitment by cloud users.
- Shared and multitenant.
- Accessible over the Internet by any device.

Cloud computing has become a significant technology trend, driven by big players like Amazon, Microsoft, Google, Salesforce.com, and transforming our current IT industry. Cloud computing delivers large-scale utility computing services to a wide range of consumers. Within cloud computing, users on various types of devices access programs, storage, processing and applications over the Internet, offered by cloud computing providers, resulting in a previously unprecedented elasticity of resources.

In general, a cloud computing system can be coarsely divided into three layers corresponding to the three concepts, IaaS, PaaS and SaaS. The IaaS layer is responsible for managing the physical machines, creating a pool of virtual computation or storage resource by using virtualization technologies, such as Xen and VMWare, and providing an elastic resource service to the upper layer. The PaaS layer is built on top of the infrastructure layer, and the platform layer consists of operating systems and application frameworks. The SaaS layer is at the highest level of the hierarchy, the application layer consists of the actual cloud applications. According to the cloud architecture, the IaaS layer is further composed of three layers, namely, physical resource, virtual resource, and management tool.

The physical resource layer consists of traditional data center and it comprises servers, network devices and disks, etc., or non-IT equipment (such as cooling equipment, lightening, air conditioner, etc.). The virtual resource layer holds a pool of unified virtualized computation or storage resources which are transformed from physical resource by using virtualization technique, such as VMWare, Xen, and Hyper-V, etc. Virtualization is one of the biggest differences between a traditional data center and a cloud data center and it makes the resources much easier to access and much more flexible to provide elastic service to the upper layer. The management tool layer is responsible for virtual resource management, accounting, and monitoring, such as OpenNebula and Eucalyptus.

Cloud computing involves the following three basic components: (1) Clients: Clients and their needs remain the same as in a common Local Area Network (LAN). They may use desktop computers, laptops, tablet computers or other mobile devices, to access internet data or services. (2) Datacenter: The datacenter is a set of servers where the requested applications are hosted. The growing trend in IT, servers virtualizing, facilitate multiple virtual servers to run on one physical server. The number of virtual servers that can run on a physical server depends on its size and speed and on the nature of the applications will be running on the virtual server. (3) Distributed servers: The structure of cloud computing allows cloud providers to host physical servers in disparate geographical locations without affecting the interaction of cloud end-users. This increases the flexibility and security options of the service provider. In case of a problem in a datacenter, a service will be still accessible through another distributed server. In addition, in case that cloud needs more hardware devices to support its workload, it is not necessary to attach more servers onto the primary datacenter but can be set up at another group of distributed servers and to be automatically embedded to the cloud.

Cloud Computing’s technical, business, and policy issues play out across three layers of technical architecture. The Infrastructure layer: It encompasses the hardware, networks and operating systems responsible for managing fundamental resources such as data storage, computation and network bandwidth. A critical element of the Cloud Infrastructure layer is the ability to virtualize the connection between physical resources and the services that consume them. Put simply, virtualization decouples applications and software platforms from the underlying physical hardware with software mimics hardware, “tricking” applications into thinking that they are interfacing with physical servers when they are in fact interfacing with software-created “virtual machines.” There may be several virtual machines residing on a particular physical server, or there may be multiple physical servers running one particular virtual machine. Virtualization enables greater flexibility in how workloads are managed, and how datacenters are constructed, since providers can
dynamically add, remove or modify hardware resources without having to reconfigure the services that depend on them.

The Platform layer: This layer serves two purposes. It provides a set of common services, such as databases, messaging, and business rules engines, that are shared by applications. It also insulates application developers from the complexity of the underlying infrastructure through a set of higher level Application Programming Interfaces (APIs). The Application layer: It provides the mechanism through which users interact with the Cloud applications—often through a web browser. In the Cloud datacenter the application layer is where the business logic for the application is run.

2. CLOUD COMPUTING ELEMENTS
   - Managed Services: A managed service is aimed at delivering an application to an enterprise, rather than to end customers directly.
   - Web Services: Web service providers offer APIs that application developers can use in developing applications.
   - Software as a Service (SaaS): The SaaS vendors run a single application in a data center and deliver the functionality via the internet to the users.
   - Infrastructure as a Service (IaaS): Many players have recently started to offer computing resources as virtual servers and storage as utility computing service.

3. CLOUD BASED ENTERPRISE ARCHITECTURE
   Enterprise cloud computing is a controlled, internal place that offers the rapid and flexible provisioning of compute power, storage, software, and security services. Cloud enables enterprises to unleash their potential for innovation through greater intelligence, creativity, flexibility and efficiency, all at reduced cost. Today, cloud computing gives businesses more control and flexibility over the technology they deploy and the way they deploy it. It helps companies reduce costs and focus resources on gaining strategic advantage. While deployment strategies differ, it is critical that an organization’s infrastructure is managed as a utility made up of secure, scalable and standards-based building blocks of integrated IT resources from storage to servers and network management tools. Companies need to look beyond this hype and seriously consider the real value of incorporating the Cloud in their own business.

Enterprises is growing and expanding their businesses globally where different people from different geographical locations connect, communicate and collaborate for achieving their business goals.
this scenario, enterprises require reduction in the cost of IT infrastructures without compromising their business values. Cloud computing assists enterprises on-demand resources provisioning where enterprises can exploit different cloud computing models such as Platform as a Services (PaaS), Infrastructure as a Services (IaaS) and Software as a Service (SaaS) according to their conditions for reducing the IT costs and increasing the productivity. Microsoft, IBM and Google are notorious cloud computing providers not only providing data and network infrastructure to the enterprises but also providing software and applications for ease of business work. For instance, word processing, document management, content management and spreadsheets are delivered to enterprises on-demand without buying and installing into their enterprise environment. Cloud services are defined hierarchically by using Cloud computing ontology. The layered approach represents inter-dependency and composure between the different services in the Clouds.

![Cloud Based Architecture of an Enterprise](image)

**Figure 2. High Level view of Architecture of a Enterprise Cloud**

The globalization of businesses, enterprises are producing large volume of disparate data with a different format, which are located on different geographical locations. In the larger interest of enterprises, such data require to be exposed to different trusted partners and co-workers. The exposure can be done on the basis of the relationship between enterprises and people. The semantic web technologies can be used as the glue that helps in providing meaning and linking to enterprise data, services and user profiles. With such semantic enhance descriptions it is possible to employ vertical search on a predefined topic to get relevant and precise search results. The in-built reasoning capabilities of semantic web enables the system to deduce new facts from the existing facts. Today, many enterprises are adopting semantic web technologies into their software development life cycle to bring intelligence and smartness in the decision-making process. The semantic web technologies are being implied in many areas such as enterprise information integration, content management, life sciences and e-government.

4. ARCHITECHTURE OF THE PROPOSED MODEL

In this case we proposed layered architecture:

- **Data Center Layer**: Physical Servers/ storages /networks
- **Virtualization Layer**: Virtual Machines / Virtual Infrastructure
- **Service Delivery Layer**: SaaS, PaaS, IaaS Models
- **Cloud Services Marketplace Layer**: It is called as a service pulling layer where cloud providers meets with cloud consumers for fulfilling their business requirement. Important factors influencing the cloud customers includes location of the Cloud Data Center, location of customer’s business units, Service Level Agreement, number of services, type of services, Complete cloud stack (Example: SaaS+PaaS+IaaS), durability of the services, resell of services, service package (Example: Amazon AWS ), promotion service
- **Access Layer**: VPN/ Firewall / Proxy Servers

An Enterprise Cloud Model for Optimizing IT Infrastructure (Rabi Prasad Padhy)
• Web Browsers: Google Chrome, Microsoft Internet Explorer, Mozilla Firefox, etc
• Enterprise Layers: Users (Developers / Testers / Admin’s / Management / 3rd party users

4.1 Scenario Description

A Large Enterprise named as ABC is a manufacturing company with a huge datacenter having various business units around the globe. Recently ABC acquired a SMB named as ABC-IN and it was reorganized to four business units called as BU1, BU2, BU3 and BU4. BU1 focusing on recruiting people for their internal IT as well as outsourcing to different service provide companies. BU2 focusing on sales and marketing of their products. BU3 focusing on product support called as a call center. BU4 focusing on market research and strategies development with respect to their competitors.

ABC looking for multiple cloud service for their new acquisition company ABC-IN which will fulfill their requirement i.e. use latest technologies with minimizing operation cost, best user of resources and manpower also minimize the installation and configuration time period as compare to existing traditional enterprise model application to new optimized and cost effective enterprise cloud model.

BU1 Requirement details: BU1 needs fresher candidates as well as in-services candidates for their upcoming projects as well as requirement for their customer’s. So to minimize the recruitment process or innovate the existing process, they require cloud services which will access multiple fresher recruitment databases, in-service recruitment databases and social media databases. Typically a HR application fulfills these needs.

BU2 Requirement details: BU2 needs a supply chain management application services for their new product for a new region. So in place of going back and installing servers and software’s, BU2 realizes from the past experience that its always better to go for a cloud application as following the upcoming market commutative and competing to their competitors.

BU3 Requirement details: BU3 needs a CRM application for their new region call centers for support new products. BU4 Requirement details: BU4 needs a ERP Application for their market research and strategic development.

4.2 Overview of the Layered Approach

Create a Virtual Machine for a user.

Figure 3. Proposed System Layered Architecture
Summary: The Provider creates a Virtual Machine for a user.
Actors: User
Precondition: The user must have an account with the Provider

Description:
- The User requests to the Provider to create a virtual machine. He specifies the physical location and the type of the instance.
- The Provider checks if the user has an account and redirects the request to the Hypervisor.
- The Virtual Machine Monitor creates an instance of the Virtual Machine and assigns it to a server and to the user.
Postcondition: A Virtual Machine is created in the specified location and assigned to a server and to the user.

4.3 How the model is applied to handle the problem described in the Scenario:
Initialization of a service request to fulfillment of the service request then the retirement of the service request Example: HR Application, Abc-in company looking for suitable candidates for recruitment for its call center at different levels as per below requirement.
- Freshers
- Experience
- Reference
- Campus Hiring

Cloud Services Marketplace Layer: In place of order a cloud service request directly to a cloud service provider its always good for a consumer to compare between the offering of the services by the various service providers. In our Layer approach we are provide a Layer 4 (shown in the figure 3) as cloud service market place layer where a consumer can compare and choose a best service which will fit to the requirement for their enterprise.

Cloud Service: A cloud service in cloud service repository is defined as a five tuple: $ws=\{\text{ServiceKey, wsName, wsDesp, QP, OprSet}\}$
- ServiceKey is the unique identifier;
- wsName represents cloud service name; wsDesp is service functional description;
- QP is published QoS information that is denoted as $QP=QN\_QD$. Where QN represents necessary quality criteria set for all web services and QD represents domain-specific quality criteria set for specific web services.
- OprSet is web operation set denoted as $OprSet=\{opr1,opr2,\ldots,opr_s\}$. Where each $opr_i(1\leq i \leq s)$ can be executed for a specific function task. Similarly, for the requirements of service requester, a corresponding service request description is given.

Service Request: A service request is defined as a four tuple: $sq=\{\text{wsName, InSet, OutSet, QR}\}$, Factors to be consider while selecting a Cloud Service:
- Identity of the requestor
- Nature of the workload
- Capacity of different elements in the environment
- Applicable compliance policies
- Required service levels
- Organization-specific policies

User Portal: The portal gives users a multitude of options from which to select and customize their cloud service to suit their needs. Options are presented based on a user’s role within the organization, and can range from different resource sizes, service tiers, and operating systems through application stacks and higher-level services, such as compliance and monitoring. The options presented are configured by IT through the service catalog, enabling both highly controlled and highly configurable cloud service requests.

Service Catalog: a service catalog is a listing of services from which a user can drive the cloud service provisioning process. The challenge lies in the natural tension between users, who want to completely customize their offerings, and the IT group, which has to maintain tight controls on the services in the environment. The role of the service catalog is to bridge that gap. The service catalog enables IT to define the
areas of configuration and choice that users can select, according to their role. Users then feel some measure of customizability of their cloud services. Each service offering has attributes that IT defines, including who can see and select this service, what service levels or constraints are important to this service, and what the internal costs are (for calculating chargebacks). Attributes are often defined in the service catalog:

- Resource configurations
- Operating systems
- Middleware stacks
- Application alternatives
- Networking options
- Compliance packages
- Monitoring tools
- Service levels
- Prices

Service Decommissioning: When a cloud service is requested, a retirement date is assigned to it. Cloud services are typically out-of-sight and thus out-of-mind, so the remnants of past cloud services, if not placed on a termination schedule, will often linger indefinitely. When the retirement date approaches, the system automatically notifies the service owner and IT. The owner and IT can jointly make an intelligent and informed decision about whether to extend the service or to decommission it, therefore reclaiming unused disk and CPU resources. Service decommissioning or retirement is a very important function, completing the lifecycle.

4.4 Sequence Diagram of the proposed Model

![Sequence diagram for the use case Create a Virtual Machine](image-url)

Figure 4. Sequence diagram for the use case Create a Virtual Machine
4.5 Flow Diagram of the proposed Model

Figure 5. Flow Chart of processing incoming request

5. KEY CONSIDERATION BY ENTERPRISE TO MIGRATE TOWARDS CLOUD COMPUTING

Some of the key characteristics of the resources possessed by a company to be considered are:

5.1 Size of the IT resources

Some of the factors that are taken into account, while determining the size of the IT resources of a company are:
1. The number of servers the company maintains in its centers,
2. The size of the customer base of the company,
3. The annual revenue from IT,
4. The number of countries across which the company spread over.

5.2 The utilization pattern of the resources

Average Usage:
- The type of services offered by the company.
- The number of users using the serviceumber services, i.e., same as customer base.
- The number of projects undertaken.

Peak Usage:
- The duration of peak usage/year.
- The number of times the peak value is of the average value.

Amount of data handling/transactions done.

5.3 Sensitivity of the data they are handling

Sensitivity of data is divided into 5 categories,
- Extremely sensitivity: Information and data of government federal and intelligence agencies like CIA, FBI. Blue prints and information regarding weapon systems, artillery, aircrafts, etc of the defense forces. Matters concerning national security.
• Very sensitivity: Bank related data like bank accounts, asswords, pin, transactions and balances, etc. Financial data of companies, quotations for various tenders, etc. Company databases. Ongoing confidential research. Trade secrets, drug formulas, early research findings. Source codes. E-mail accounts.

• Sensitivity: Personal information such as name, contact details, email-ids, and profiles. Patient records. Health records.

• Less sensitivity: Pictures, videos in social networking websites. Click stream data. Service usage details.


5.4 Criticality of work done by the company

Highly critical work requires the most stringent of resources, platforms, applications and security. The more critical the work gets, the more demanding gets the requirements. Thus very critical work may not find suitability to the cloud as it would require very stringent Service-Level Level Agreements (SLAs) with the cloud provider. If the company is not very big, then the cloud service providers may not be very willing to provide the highly customized Service-Oriented Oriented Architecture (SOA) and Application Programming Interfaces (APIs). They may not even be able to deliver the Quality-of-Service (QoS) in processes requiring very low latencies owing to constraints of the system or lack of profitability and feasibility on their part.

Criticality of services can be categorized into:
Highly critical : May not be suitable for the cloud
Critical : May be suitable if company is large
Less critical : Suitable
Standard : Easily Suitable

CONCLUSION

Selecting some suitable services as per the requirement for an Enterprise Cloud Computing is not easy. Managing distributed heterogeneous resources causes some problems: difficulty of resource information management, no standard definitions of resource requirements, and difficulty of guaranteeing compatibility of resource allocation. To solve these problems, we propose a Layer architecture method to select best services. Cloud computing has advanced as a model for delivering Internet-based information and technology services in real time. Our work was motivated by the challenges in building a Enterprise Cloud system that provides best service selection for an Enterprise. In this paper, we proposed a Cloud Enterprise Layered architecture and showed the criteria for how to select a best cloud service which will increase the trust between cloud consumer and service provider with respect to SLA and QoS.

REFERENCES


BIOGRAPHY OF AUTHORS

Rabi Prasad Padhy is currently working as a Senior Software Engineer - Oracle India Private Ltd. Bangalore, India. He has achieved his MCA degree from Berhampur University. He carries 8 years of extensive IT Experience with MNC’s like EDS, Dell, IBM and Oracle. His area of interests include IT Infrastructure Optimization, Virtualization, Enterprise Grid Computing, Cloud Computing and Cloud databases. He has published several research papers in national and international journals. He is a certified professional for Oracle, Microsoft SQL Server database and also ITIL certified.

Dr. Manas Ranjan Patra holds a Ph.D. Degree in Computer Science from the Central University of Hyderabad, India. Currently he is an Associate Professor in the Post Graduate Department of Computer Science, Berhampur University, India. He has about 24 years of experience in teaching and research in different areas of Computer Science. He had visiting assignment to International Institute for Software Technology, Macao as a United Nations Fellow and for sometime worked as assistant professor in the Institute for Development and Research in Banking Technology, Hyderabad. He has about 90 publications to his credit. His research interests include Service Oriented Computing, Software Engineering, Applications of Data mining and E-Governance. He has presented papers, chaired technical sessions and served in the technical committees of many International conferences.