International Journal of Applied Research 2017; 3(1): 323-327

International Journal of Applied Research

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2017; 3(1): 323-327
www.allresearchjournal.com
Received: 14-11-2016
Accepted: 15-12-2016

Jyoti Shekhawat
Assistant Professor Somany (P.G) Institute of Technology & Management, Rewari, Haryana, India

Correspondence
Jyoti Shekhawat
Assistant Professor Somany (P.G) Institute of Technology & Management, Rewari, Haryana, India

Cloud configuration, performance and capacity management

Jyoti Shekhawat

Abstract
Cloud Computing makes the dream of computing real as a tool and in the form of service. This INTERNET-based ongoing technology which has brought flexibility, capacity and power of processing has realized Service- oriented idea and has created a new ecosystem in the computing world with its great power and Benefits. Cloud capabilities have been able to move IT industry one step forward. Nowadays, large and Famous enterprise has resorted to cloud computing and have transferred their processing and storage. Due to popularity and progress of cloud in different organizations, cloud performance evaluation is of special importance and this evaluation can help users make right decisions. In this paper, we provide an overall perspective on cloud evaluation criteria and highlight it with help of simulation. For this purpose, we present different major factors in cloud computing performance and we analyze and evaluate cloud performance in various scenarios considering these factors.

Keywords: Cloud configuration; cloud format, performance, capacity management

Introduction

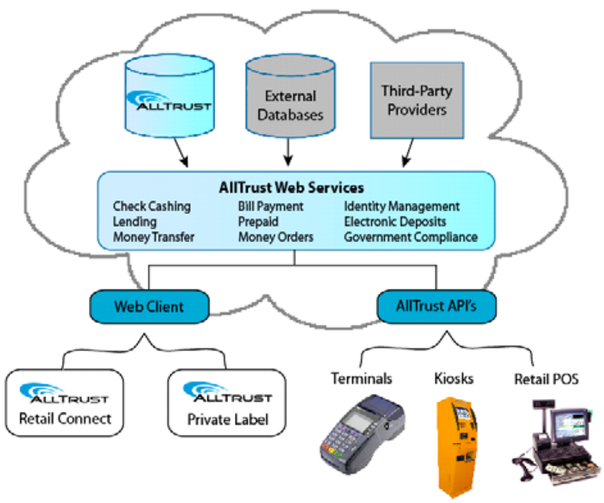
The term "cloud computing" is not a new concept for the users of computer's world and the Concept dates back the last decades and when John Mac Carty predicted that computers might be used one day as a public utility The confluence of technological advances and business development in Internet broadband, web services, computer systems and applications has created complete storm for cloud computing during the past decade Nowadays, cloud is the best solution for people who are looking for rapid implementation methods A more accurate and scientific definition of cloud computing, Cloud configuration, Cloud Performance And Capacity Management will be reviewed in other key considerations of this paper

Cloud configuration

Cloud configuration is the process of setting hardware and software details for elements of a cloud environment to ensure that they can inter operate and communicate. The complexities of a cloud environment, and hybrid cloud environment in particular, make configuration more challenging than would be the case in a single-location, homogeneous network. A cloud provider has to enable service for clients in various locations, using all different kinds of hardware and software. Just like the administrator of a single-premises network, the provider has to ensure that the service is reliable, performance is acceptable and communications are secure. application program interfaces sit on top of client hardware to enable interaction with the provider's environment. The APIs are supposed to deliver a unified programmable experience no matter what hardware is running underneath. However, that means it's not possible to use the same practices used on traditional hardware, simply because there are too many variations. Configuration specifics differ among the three major elements of cloud provision, sometimes referred to as the SPI model:

In the Software as a Service distribution model, applications are hosted by a vendor or service provider and made available to customers, usually over the Internet. Software configuration may be enabled for the user so that they can make the same types of changes they would to customize a locally-hosted application. In other cases, there may be a single configuration offered.

The provider's configuration responsibilities are typically laid out in the service-level agreement In the Platform as a Service model, operating systems and associated services are delivered over the Internet without downloads or installation. The platform for a developer could include, for example, an operating system, a programming language, an execution environment, a database and a web server. Configuration and management of all elements are the responsibility of the provider. In the Infrastructure as a Service involves outsourcing the equipment used to support operations, including storage, hardware, servers and networking components. Similarly to the PaaS service model, configuration of all elements is the responsibility of the provider.



Cloud computing is defined as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" In this section we introduce the constituents of cloud systems.

Compute Resource: The core processing capabilities that are used to execute software instructions. We define this as comprising of a CPU, typically in multi course configuration, CPU cache and primary storage memory. Data centers typically house many thousands observers containing these compute resources.

Storage Resource: Non-volatile secondary storage memory houses the data used by compute resources. As this resource is typically cheaper than primary memory, many operating systems are able to use it as an extension of main memory, to temporarily swap out unused memory state. Many data centers will have servers with access to internal storage as well as to a Storage Area Network that consolidate and abstract the complexity of accessing storage throughout the data center.

Network cards and storage are also visualized and presented as individual devices to VMs.

Service Management Resource (SMR): is a knowledge library where IPs store management objectives, policies, pricing and orchestration information.

Management Tools: are used by IPs to provision, monitor, and reconfigure, backup and restore the infrastructure. IPs typically builds the infrastructure and offer access to virtual resources, with a VM being the main component. VMs reside on physical nodes of heterogeneous capabilities where the performance characteristics of compute, storage and network vary. Demand for recourses varies over time as users consume and release these resources. As more resources are used, power consumption in the data centre increases and IPs may choose to optimize the allocation of VMs to physical nodes. In the next section, we will cover IPs objectives and approaches used to optimize this allocation.

Cloud Formats

The cloud environment is subdivided into public, private, hybrid and community clouds.

- Public clouds are those in which services are available to the public at large over the internet in the manner already described in this chapter.
- A private cloud is essentially a private network used by one customer for whom data security and privacy is usually the primary concern. The downside of this type of cloud is that the customer will have to bear the significant cost of setting up and then maintaining the network alone.
- Hybrid cloud environments are often used where a customer has required- mints for a mix of dedicated server and cloud hosting, for example if some of the data that is being stored is of a very sensitive nature. In such circumstances the organization may choose to store some data on its dedicated server and less sensitive data in the cloud. Another common reason for using hybrid clouds is where an organization needs more processing power than is available in-house and obtains the extra requirement in the cloud. This is referred to as 'cloud bursting'. Additionally, hybrid cloud environments are often found in situations where a customer is moving from an entirely private to an entirely public cloud setup.

Community clouds usually exist where a limited number of customers with similar IT requirements share an infrastructure provided by a single supplier. The costs of the services are spread between the customers so this model is better, from an economic point of view, than a single tenant arrangement. Although the cost savings are likely to be greater in a public cloud environment, community cloud users generally benefit from greater security and privacy, which may be important for policy *Network Resource:* includes the network cards that connect into servers as well as infrastructure components that include repeaters, load balancers, switches and firewalls. Networks can use different topologies and protocols, which influence the level of security, resilience and Quality of Service.

Virtual Resource: is an abstraction added onto compute, storage and network resources. It enables slicing of these resources into smaller chunks that can be scaled vertically or horizontally. Typically visualization is used in a data center to slice data center compute resource into Virtual machines, and potentially to present several logical processors by mapping these onto a single physical processor reasons.

Cloud Performance

Cloud computing resources must be compatible, high performance and powerful. High Performance is one of the cloud advantages which must be satisfactory for each service Higher performance of services and anything related to cloud have influence on users and service providers. Hence, performance evaluation for cloud providers and users is important.

There are many methods for performance prediction and evaluation; we use the following methods in our evaluation:

- Evaluation based on criteria and characteristics
- Evaluation based on simulation

Another category which can be considered for evaluating cloud performance is classification of three layers of cloud services evaluation.

Factors affective on performance

Nowadays, the term "performance" is more than a classic concept and includes more extensive concepts such as reliability, energy efficiency, scalability and soon. Due to the extent of cloud computing environments and the large number of enterprises and normal users who are using cloud environment, many factors can affect the performance of cloud computing and its resources. Some of the important factors considered in this paper are as follows:

- Security, the impact of security on cloud performance may seem lightly strange, but the
- Impact of security on network infrastructure has been proven. For example, D DoS attacks
- Have wide impact on networks performance and if happen, it will greatly reduce networks
- Performance and also be effective on response time too. Therefore, if this risk and any same risks threaten cloud environment, it will be a big concern for users and providers
- Recovery, when data in cloud face errors and failures or data are lost for any reason, the time required for data retrieval and volumes of data which are recoverable, will be effective on cloud performance. For example, if the data recovery takes a long time will be effective on cloud Performance and customer satisfaction, because most organizations are cloud users and have quick access to their data and their services are very important for them.
- Service level agreements, when the user wants to use cloud services, an agreement will be signed between users and providers which describes user's requests, the ability of providers, fees, fines etc. If we look at the performance from personal view, the better, more optima land more timely the agreed requests, the higher the performance will be This view also
- Network bandwidth, this factor can be effective on performance and can be a criterion for evaluations too. For example, if the bandwidth is too low to provide service to customers, performance will be low too.
- Storage capacity, Physical memory can also be effective on the performance criteria. This factor will be more effective in evaluating the performance of cloud infrastructure
- Buffer capacity: as shown in figure 2, if servers cannot serve a request, it will be buffered in temporary memory. Therefore, buffer capacity effect on performance. If the buffer capacity is low, many requests will be rejected and therefore performance will be low.
- Disk capacity, can also have a negative or positive impact on performance in cloud[
- Fault tolerance, this factor will have special effect on performance of cloud environment. As an example, if a data center is in deficient and is able to provide the minimum services, this can increase performance. Availability, with easy access to cloud services and the services are always available, Performance will be increase Number of users, if a data center has a lot of users and this number is greater than that of the rated capacity, this will reduce performance of services. Location, data centers and their distance from a user's location are also an important factor that can be effective on performance from the users' view. Other factors that can affect performance which are as follows:· Usability· Scalability, Workload, Repetition or redundancy· Processor Power· Latency

Performance Evaluation Criteria

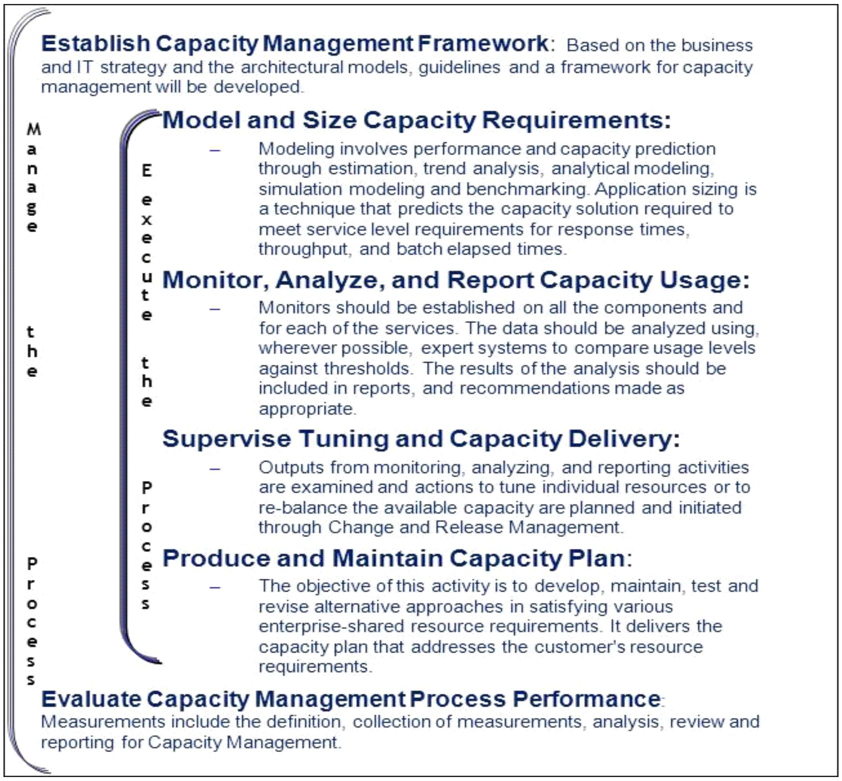
There is a series of criteria for evaluation of all factors affecting performance of cloud computing some of which will be used in this paper. These criteria are under development. Some of these criteria have been selected considering the importance and criteria in simulation. It should be mentioned that all of criteria listed in previous sections cover the factors mentioned in the previous section but some of the factors will be important in special criteria:

Average response time per unit time, this criterion will cover all factors completely, Network capacity per second (Mbps)or unit time, the most important factor associated with this criterion is network bandwidth, availability and scalability. Workload (requests) to be serviced per second (M bps) or a unit of time, Throughput this criterion will be recovered recovery, buffering capacity and processing power factors.

- The average time of processing
- Percentage of CPU utilization
- The number of requests executed per unit time
- The number of requests per unit time buffer
- The number of rejected requests per unit time ands per second (IOPS)or unit time

Capacity Management

The Capacity Management process is not an isolated, stand-alone process. Capacity Management has interdependencies with several other IT Service Management processes and some are even more important in the context of cloud computing. For example, Capacity Management has a strong dependence on well defined and managed Asset and Configuration management to provide and track the number, size, relationship, and location of the cloud components that are in scope for any given cloud computing environment. See Figure.



Capacity Management can drive the monitoring and measurement requirements useful to other IT Service Management processes, such as Event Management and Availability Management. IT resource utilization metrics used for capacity planning can also be used for usage-based cost recovery and billing. Metrics requirements details, such as a unified time base and relationship meta data, become critical to capacity management of cloud environment since many metrics and measurements from different tools and sources need to have a consistent time base, and a relationship method between data items to allow analytics And predictions to be performed. Automated data collection, summarization, pruning, Archiving and storage is essential. Reliability and integrity of the data elements and data Management controls are essential to ensure consistent, repeatable, reliable metrics. Elements of Demand Management are the main source of business demand that Capacity Management uses to develop capacity forecasts and solutions. Capacity Management Translates business demand through to the infrastructure component level. A common, standard set of resource usage data can be shared by performance, capacity, and Usage-based show-back/charge-back to forecast and manage IT demand.

These examples of interdependencies between Capacity Management and other IT Service Management processes are directly relevant to cloud services.

The process covers understanding service requirements, determining component capacities, and designing and deploying capacity to meet expectations. It collects and analyzes data Relevant to infrastructure utilization and performance to determine whether there are potential Problems and issues that need to be addressed.

ITIL defines the following three focus areas, or sub-processes, which are addressed by Capacity Management. Each uses the primary activities of the process decomposition in Differing ways, to differing end results.

Business Capacity Management (BCM): This focus area is responsible for ensuring that The impacts of future business requirements for IT services upon IT resources are Considered, planned, and implemented in a timely fashion (requirements (NFRs), drivers, Volumetric, use cases).

Service Capacity Management (SCM): This focus area is the management of the Performance of the IT services used by the customers. It is responsible for ensuring that the service performance is monitored, measured, reported, and meets business Requirements and agreements (workload by service/application/transaction and Corresponding stack Component Capacity Management (CCM): This focus area is the management of the Performance, utilization, and capacity of individual technical components possessing finite Resources (IT resource load/consumption on the server, such as CPU, I/O, memory, Storage, network infrastructure, and other hardware elements).

When applying Capacity Management to cloud computing we explore who is responsible for the activities, inputs, outputs, and controls. This must be fully examined and understood in Terms of roles and responsibilities, depending on the provider/consumer perspective and the Service depth. The two activities in Figure are not examined in this paper because these are fairly Standard process design and management activities. They are not significantly different for cloud computing, other than to underscore the need to clarify the objectives, scope, and roles and responsibilities for this process. They must be negotiated and agreed to between the Various business roles and layers of service. Capacity Planning involves tasks embedded within each of the four activities in Figure on. The next section of this paper summarizes those four activities and the elements that are relevant to cloud computing.

Conclusion

Performance implications for cloud computing in rather general terms, since the whole are still relatively new. With the benefit of additional industry experience, it was appropriate for this Red paper to spend more time on what we learned about deploying and managing cloud solutions that perform well while being cost-effective.

We are seeing two extremes in leading-edge computing. On one hand, we have purpose-built, highly customized systems such as Watson, in which every tier of the solution is optimized for the high-performance computing needs of a specific domain. On the other hand, we have utility cloud computing systems that must be able to rapidly allocate computing resources in response to new and diverse workloads and then characterize and manage those workloads on an ongoing basis. Can today's Watson become one of tomorrow's utilities? For cloud computing to take its proper place in the “IBM Smarter Planet” toolbox, we must build intelligent clouds that can handle increasingly diverse, variable, and challenging workloads while fully exploiting the emerging capabilities of new technology.

In this IBM Red paper, we discussed how contemporary approaches to testing, monitoring, and capacity management of cloud solutions are starting to address these concerns. For example, yesterday's performance testing was typically concerned solely with the responsiveness and utilization of a single application at a time. In today's cloud environments, we must also ensure that the cloud management solution components perform well, and that the overall solution exhibits sufficient elasticity to manage multiplication (and sometime multi-tenant) workloads. This places new demands not only on how we do performance testing, but also on how we monitor performance and manage capacity as well. Who takes on these responsibilities depends largely on whether a public, hybrid, or private cloud solution is chosen.

We also provided examples of how cloud computing is used to address tomorrow's problems today, but we understand that this only scratches the surface of what is being done. We want to hear from you about your journey into cloud computing, particularly the performance and capacity challenges you encountered along the way. Contact one of the authors to learn more.

References

1. Borko Furht, Armando Escalante. Handbook of Cloud Computing, Springer, 2010.
2. Abah Joshua, Francisca Ogwueleka N. Cloud Computing with Related Enabling Technologies, International Journal of Cloud Computing and Services Science (IJ-CLOSER). 2013; 2(1):40-49.
3. NIST Advisory Working Group. NIST Cloud Computing Standards Roadmap, NIST, 2011.
4. Onur E, Sfakianakis E, Papagianni C, Karagiannis G, Kontos T, Niemegeers I *et al*. Intelligent End-To-End Resource Virtualization Using Service Oriented Architecture‖, Delft Univ. of Technol., Delft, Netherlands, GLOBECOM Workshops, IEEE, 2009.
5. Buyya R. Market-Oriented Cloud Computing: Vision, Hype, and Reality of Delivering Computing as the 5th Utility, 9th IEEE/ACM International Symposium on.
6. Buyya R. Market-Oriented Cloud Computing: Vision, Hype, and Reality of Delivering Computing as the 5th Utility, 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009, 1.
7. Gruman G. What cloud computing really means, Info World, Jan. 2009.