International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 12
ISSN: 2454-4248
402 – 408

A Literature review on Balancing Workload in Cloud Computing

[1] Pala Priyanka*, [2] Vamsi Krishna, [3]R V Ramana
[1] Assistant Professor, [1]AITS, JNTUA University
[3] Assistant Professor, [3] SVCE, JNTUA University
[1]*palapriyanka511@gmail.com*
[2]*vamsikrishna.chenna@outlook.com*
[3]*rvramana.r@gmail.com*

Abstract— Cloud computing is the realistic evolution of information technology in a world that is learning to be more and more based on the division of work. Cloud computing offers many principles that are long accomplished in other industries to the IT. We focus on basic characteristics of cloud computing help you understand them. A pattern language is used to interconnecting set of cloud patterns. A cloud pattern is a small human readable document of a well-defined format describing a good solution to a cloud related problem. We studied and captured such patterns describing different types of clouds, the offerings they provide and how to build application with them. In this paper we mainly focus on different types of application workload patterns. Pattern for application workloads describe different user behavior resulting in changing utilization of IT resources hosting an application. Having motivated the need for cloud offerings to handle different workloads we introduce common cloud service models that describe different styles to offer IT resources on different levels of an application stack. Furthermore we also discuss how the corresponding service models and cloud deployment models enable the cloud computing properties.

Keywords: *Cloud computing, Pattern, workload.*

*****

I. INTRODUCTION

Cloud computing is a model for enabling Omni-persistent, suitable, on-demand service access to a typical cluster of configurable computing resources(e.g., networks, servers, storage and applications) that can be conveniently available and released with least management struggle or cloud provider dealings. The vision of key characteristics of cloud environment is shown in Figure 1.

Figure1: Vision of Key characteristic of cloud environment.

The various elements of cloud computing are summarized below:

Economical: Cloud computing is cost effective because of utilization based billing model, no need of infrastructure.

Larger storage space: With the huge infrastructure, storage & management of large space is possible. Sudden Cloud workload fluctuations are also managed successfully, since the cloud can scale dynamically during the situation of overloading and under loading.

Elasticity: Cloud computing challenges on experiencing cloud workloads or applications to promote very rapidly, by using the most suitable building blocks essential for deployment.

Trustworthiness: It's ability to validate continuous working of computers without disruption i.e. no damage of data, no code change throughout execution etc.

Geographically Distributed: Cloud individuals can use resources through a web browser regardless of their site from where user are accessing.

Dynamic Scalability: The data and application resources can be rapidly provided when and where wanted.

Availability: With the right Cloud provider it will be guaranteed that available resources remain continuously accessible 24*7.

Less Management: The hardware, applications and bandwidth are maintained by the cloud provider.

Expert Service: At convenience, cloud computing features are continuously supervised and maintained by on-site staff of professional data center specialists.

Various Business **Advantages** to develop applications through cloud computing are:

 ISSN: 2454-4248
402 – 408

No Direct Infrastructure
JIT infrastructure
Maximum Resource Utilization
Consumption Based Charges
Reduce Execution Time

Elements of Cloud Computing

There are seven elements of cloud computing classified on economic, architectural and strategic elements based as shown in Figure 2.

Utility pricing: Cloud computing is well defined by its usage based billing model.

Elastic Resource Capacity: Cloud computing scales computing and storage resources up and down, consumers can add or remove resources instantly and make payment for the resources a cloud consumer is consuming. Resources immediately and make payment for the resources a cloud consumer is consuming.

Virtualized Resources: Without virtualization cloud computing is impossible not for unknown technical causes, but for one established business requirement: the requirement of multi-tenancy**.**

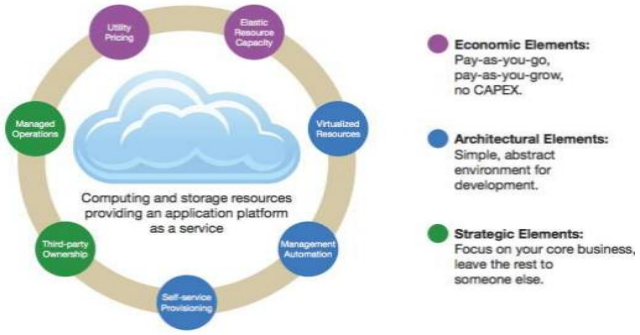


Figure 2: cloud computing elements and their broad classifications.

Management Automation: Standardization makes cloud computing different from conventional corporate data-centers by dramatically reducing operational costs through intense management automation.

Self Service Provisioning: Application Service provider model that became famous for short time is compared with cloud computing and software as a service in context of self service provisioning.

Third Party Ownership: Users demanding to emphasize the distribution of occasional principal resources to their primary businesses quickly recognize the reimbursements of moving IT infrastructure off their balance sheet.

Managed Operations: Assuming that the human resources that will openly affect the business, rather than handling the infrastructure in cloud computing this advocates a model according to the IT infrastructure which is retained and maintained by the third party.

II. RELATED WORK

Fundamental Cloud computing Patterns

Patterns for application workloads describe user behavior resulting in changing utilization of IT resources hosting an application. This workload can be measured in the form of user requests, processing load on servers, network traffic, amount of data stored etc.

In this paper we can cover different types of application workloads i.e., static workload that only changes minimally over time, periodic workload that has recurring peaks, once-in-a-lifetime workload that has a peak once, unpredictable workload that changes frequently and randomly, and continuously changing workload that grows or shrinks over time.ted in applications.

Once the workload experienced by an application can be described and categorized it is important to understand the cloud service models used by a cloud provider. It affects the pricing model of providers and therefore how workload should be measured and evaluated.

Infrastructure as a service describes how servers are offered by cloud providers. Platform as a service covers cloud offerings provide complete execution environment for a specific type of applications, i.e., those developed in a certain programming language. Software as a service describes how the complete applications can be offered to customers. Cloud deployment models describe the cloud environments hosting these resources, especially regarding the group of customers they are made available to. Therefore a combination of cloud service model and cloud deployment type characterizes the environment of a cloud provider.

A public cloud is generally available to everyone. A private cloud is hosted exclusively for one company. A community cloud is a cloud environment between these extremes and is made accessible only to a certain group of companies or individuals that trust each other and often wish to collaborate. A hybrid cloud provide means to interconnect clouds of the other deployment models to distribute applications among various hosting environments.

III. RESEARCH WORK

Application Workload

We use the term workload to refer to the utilization of IT resources on which an application is hosted. Workload is the consequence of users accessing the application or jobs that need to be handled automatically.

 ISSN: 2454-4248
402 – 408

Workload becomes imminent in different forms, depending on the type of IT resource for which it is measured: servers may experience processing load, storage offerings may be assigned larger or smaller amounts of data to store or may have to handle queries on that data In scope of the abstract workload patterns, we merely assume this utilization to be measurable in some form.

These measurements form the basis to increase or decrease the number of IT resources assigned to an application during elastic scaling, one of the cloud application properties introduced.

As customers desire to pay for used resources only, providers must employ the principle of rapid elasticity to elastically grow or shrink the resources assigned to a customer based on that customer's demand. Therefore, at least two of the essential cloud properties – pay-per-use and rapid elasticity – result from the demand to cope with non-static application workloads. In the following we examine some common utilizations of IT resources over time. This workload is the result of user requests to an application or cloud offering resulting in processing load, communication traffic, or data to be stored.

Figure 2 shows a general problem that arises in scope of workload changes to which a scaled out application has to react by changing resources numbers. Whenever workload is predicted as shown by the predicted workload curve it may be experienced slightly different as shown in the experienced workload curve.

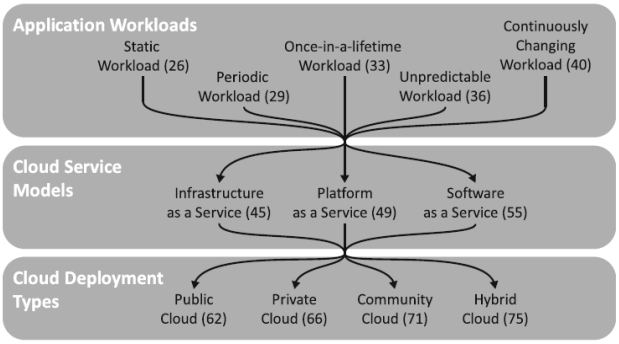


Figure 2: Workload Change

In case of static scaling where physical servers are provisioned the time it takes to order setup and start them may not be reactive enough to handle the faulty prediction. Therefore the necessary resources become available too late resulting in an un-provisioned application. To cope with the inflexibility of such resources, they have to be provisioned for the predicted peak-load right from the beginning and are hard to decommission once the workload decreases. This results in an "overprovisioned" application after the peak.

This over- and under provisioning has a direct impact on the properties of the hosted applications. Over provisioning has a lesser impact on the user of the application but leads to higher costs as resources are provisioned but remain unused.

Elastic scaling depicted on the right of Figure can provision and decommission resources much more flexible and thus is not as dependent on workload predictions. . Once the increase is detected, new resources are provisioned in small intervals and this provisioning is stopped even though the predicted workload peak has not been reached. Therefore, elastic scaling allows a much tighter alignment of IT resource numbers to experienced workloads, but has to be respected by the application architecture.

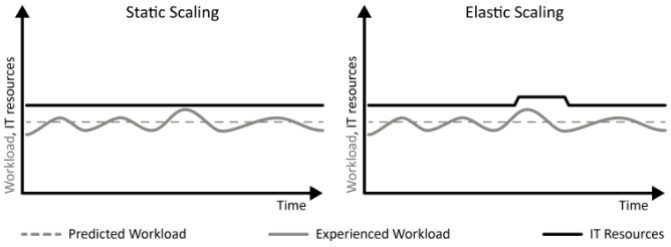
1. Static Workload:

IT resources with an equal utilization over time experience static workload.

Q: How can an equal utilization be characterized and how can applications experiencing this workload benefit from cloud computing?

Context: Static workloads are characterized by a more-or-less flat utilization profile overtime within certain boundaries.

Solution: An application experiencing static workload is less likely to benefit from an elastic cloud that offers a pay-per-use billing, because the number of required resources is constant.



Result: AN elastic cloud may be beneficial even for static workloads because elasticity does not only provide costs savings. As clouds can provision new resources very quickly, often, within minutes, elasticity also simplifies provisioning and decommissioning tasks that are necessary for other reasons, for example, to address resource failures or for maintenance purposes.

In conclusion, the cost benefits of a cloud offering might be limited or nonexistent in case of static workload.

2. Periodic Workload

IT resources with a peaking utilization at reoccurring time intervals experience periodic workload.

 ISSN: 2454-4248
402 – 408

Q: How can a periodically peaking utilization overtime be characterized and how can applications experiencing this workload benefit from cloud computing?

Context: In our real lives, periodic tasks and routines are very common. For example monthly paychecks, monthly telephone bills, yearly car checkups, weekly status reports, or the daily use of public transport during rush hour all these tasks and routines occur in well-defined intervals.

Solution: From a customer perspective the cost saving potential in scope of periodic workload is to use a provider with pay per use pricing model allowing the decommissioning of resources during non-peak times.

This has the effect that the customer does not pay for the resources during these times.

Result: The benefits for customers result from unneeded resources being decommissioned during non-peak times and thus these resources not generating costs during these times.

3. **Once-in-a-lifetime workload**

If resources with an equal utilization over time disturbed by a strong peak occurring only once experience once-in-a-lifetime workload.

Q: How can equal utilization with a one-time peak be characterized and how can applications experiencing this workload benefit from cloud computing?

Context: As a special case of Periodic workload the peaks of periodic utilization can occur only once in a very long time frame. Often this peak is known in advance as it correlates to a certain event or task.

Solution: The elasticity of a cloud is used to obtain IT resources necessary. The Provisioning and decommissioning of IT resources can often be realized as a mutual task, because it is performed at a known point in time.

Result: As provisioning and decommissioning is only performed once, the benefits of an automated alignment of IT resource numbers to the experienced workload are reduced possibly making the additional effort to automate them unreasonable.

4. **Unpredictable workload**

IT resources with a random and unforeseeable utilization overtime experience unpredictable workload.

Q: How can random and unforeseeable utilization be characterized and how can applications experiencing this workload benefit cloud computing?

Context: Random workloads are a generalization of Periodic workloads as they require elasticity but are not predictable. Such workloads occur quite often in the real world.

Solution: Unpredictable workloads require the unplanned provisioning and decommissioning of IT resources hosting applications.

Result: As with periodic workload providers have to be able to dynamically add and remove resources to customers during peak workload times and remove them when workload intensity is lower

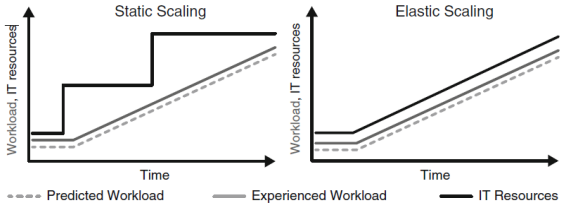
Continuously changing workload

IT resources with a utilization that grows or shrinks constantly overtime experience continuously changing workload.

Q: How can a continuous growth or decline in utilization be characterized and how can applications experiencing this workload benefit from cloud computing?

Context: Many applications experience a long term change in workload. Increasing workload often corresponds to the successful growth of a business after it was launched impacting the supporting applications.

Solution: Continuously changing workload is characterized by an ongoing continuous growth or decline of the utilization. This change can be linear, non-linear, exponential etc. but in any case the change in utilization is consistent towards one direction.



Result: If the rate of workload change is known and not very intense, the same effects apply as with planned once-in-a-lifetime workload.

Cloud Service models

Cloud service models describe the style how IT resources are offered. The following patterns compare and categorize different cloud service models according to the layers of the application stack for which they provide IT resources.

From bottom to top the six layers comprising the stack are:

Physical hardware: Perceptible physical infrastructure. This infrastructure contains for example, servers, storage, networks connecting servers and holders including the servers as well as the generating housing the data center, power lines etc.

 ISSN: 2454-4248
402 – 408

Virtual Hardware: Physical hardware components can be abstracted and mapped to virtual equivalents by a hypervisor and virtual networking. The aim of this mapping is to communicate physical hardware between multiple and virtual equivalents.

Operating system: Software installed immediately on the physical or virtual hardware. Operating systems abstract hardware by offering functions to applications, processes and data.

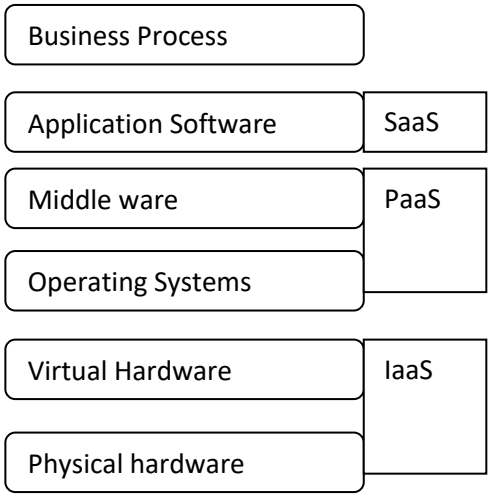


Figure 3: Layers of the comprising Stack

Application Software: Custom applications delivering features to human users or other applications are associated with this layer.

Business Processes: The processes of a company that are supported by a set of programs are associated with this layer. These processes are domain specific and subsume for example order processing, credit approval processes, billing etc.

Infrastructure as a service (IaaS)

Providers share physical and virtual hardware IT resources between customers to enable self-service, rapid elasticity and pay-per-use pricing.

Q: How can different customers share a physical hosting environment co that it can be used on-demand with a pay-per-use pricing model?

Context: Applications often experience varying workloads that lead to different utilizations of IT resources on which these applications are hosted.

Solution: A provider using the infrastructure as a Service model offers physical and virtual hardware, such as servers, storage and networking infrastructure that can be provisioned and decommissioned quickly through a self-service interface.

Result: IaaS clouds in detail offer infrastructure It resources such as servers, storage and networking.

Platform as a Service (PaaS)

Providers share IT resources providing an application hosting environment between customers to enable self-service, rapid elasticity and pay-per-use pricing.

Q: How can custom applications of the same customer or different customers share an execution environment so that it can b used on-demand with a pay-per-use pricing model?

Context: If many customers require similar hosting environments for their applications, there are many redundant installations resulting in an inefficient use of the overall cloud.

Especially small and medium sized businesses may not have the manpower and skills to perform these task efficiently and thoroughly.

Solution: a cloud provider using the PaaS service model offers managed operating systems and middleware. Customers may host individual application software supporting their business processes in this environment.

Result: PaaS subsumes the layers above physical and virtual hardware and below complete software applications thus it contains operating systems as well as middleware products.

Software as a Service

Providers share IT resources providing human-usable application software between customers to enable self-service, rapid elasticity and pay-per-use pricing.

Q: how can customers share a provider-supplied software application so that it can be used on-demand with a pay-per-use pricing model?

Context: Small and medium enterprises may not have the manpower and know-how to develop custom software applications or this purpose. Other applications have become commodity and are used by many companies.

Solution: A provider using SaaS service model offers a complete software application to customers who may use it on-demand via a self-service interface. Customers perform their individual business processes, but do not have to install and manage an application required to support these processes.

Result: SaaS is the established cloud service model in which the advantage of the cloud principles become most obvious. Instead of buying and installing hardware. Paying for licenses, handling and configuration of the necessary middleware and software products, training and paying for

 ISSN: 2454-4248
402 – 408

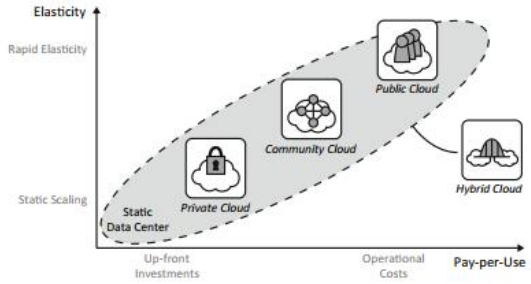
system administrators on the hardware, middleware and software level, customers can simply obtain the required software from the cloud.

Cloud Deployment Model

Regardless of the service model followed by a cloud provider a cloud can be hosted in different forms. The following patterns cover these cloud these cloud deployment models and different ways to host them in detail. In this scope we use the term "tenant" for companies or individuals that act as a customer of a cloud.

Each tenant may have multiple users associated with it i.e a company may act as the customer of a cloud that is then used by employees of that company.

The restriction of the number of tenants accessing a cloud and sharing IT resources has a significant impact on the cloud properties displayed by different cloud deployment models, especially regarding resource pooling, rapid elasticity and subsequently metered service.



Public Cloud

IT resources are provided as a service to a very large customer group in order to enable elastic use of a static resource pool.

Q: How can the cloud properties on-demand self-service, broad network access, pay-per-use, resource pooling and rapid elasticity be provided to a large customer group?

Context: A provider offering IT resources according to one of the cloud service models, IaaS, PaaS and SaaS has to maintain physical data centers that are limited in capacity. The capacity of the data center, however has to be planned statically and cannot be adjusted with the same elasticity even though this behavior shall be displayed to customers.

Solution: The public cloud is the cloud deployment model that best meets the desired cloud computing properties, because it serves a large number of customers and is thus large enough for customer diversity to level out peak workloads of individual applications.

Security mechanisms are employed to isolate customers from each other. Often this involves monitoring accesses and data entering and leaving the cloud in order to identify unlawful behavior. Such mechanisms are of vital importance to the success of the cloud provider because trust is the major asset for the acceptance of the public clouds.

Result: By sharing resources between a large number of customers and because of customer diversity. The size of the public cloud enables dynamic and elastic resource usage while ensuring a leveled utilization of the static physical data center hosting the cloud. The capacity of the data center may be adjusted with much less dynamicity.

Private Cloud

IT resources are provided as a service exclusively to one customer in order to meet high requirements on privacy and trust while enabling elastic use of a static resource pool as good as possible.

Q: How can the cloud properties on-demand self-service, broad network access, pay-per-use, resource pooling and rapid elasticity be provided in environments with high privacy, security and trust requirements?

Context: Many factors such as legal limitations, trust and security regulations, motivate dedicated, company-internal hosting environments only accessible by employees and applications of a single company.

Solution: A private cloud enables the cloud computing properties in a company-internal data center, thus only one tenant accesses the cloud. Alternatively the private cloud may be hosted exclusively in the data center of an external provider then referred to as outsourced private cloud.

Result: The main difference between a private cloud and other cloud deployment models is that the it resources hosting the cloud that are shared with other customers are reduced drastically upto the point where no resources are shared.

A private cloud therefore may provide a high level of security and privacy, which may, however reduce its elasticity and the ability to provide a pay-per-use pricing model.

Community cloud

IT resources are provided as a service to a group of customers trusting each other in order to enable collaborative elastic use of a static resource pool.

Q: How can the cloud properties on-demand self-service, broad network access, pay-per-use, resource pooling and rapid elasticity be provided to exclusively to a group of customers forming a community of trust?

 ISSN: 2454-4248
402 – 408

Context: Companies may have to collaborate for various reasons. For example a company may be a supplier of another company. Furthermore a group of companies or public institutions such as university or hospitals may have to exchange information or may have to share personnel for cost reduction.

Solution: IT resources required by all collaborating partners are offered in a controlled environment accessible only by the community of companies that generally trust each other. This cloud contains all shared data and functionality that the participating companies need to do their business.

Result: A community cloud hosted by one company is commonly used if this company has a central role in the collaboration. Outsourced community clouds are especially suitable if there is no company trusted enough by all the other companies to maintain the data center.

Virtual community clouds are commonly easiest to establish as they depend on resources of a public cloud to which access is only granted to collaborating partners.

Hybrid Cloud

Different clouds and static data centers are integrated to form a homogeneous hosting environment.

Q: How can the cloud properties on-demand self-service, broad network access, pay-per-use, resource pooling and rapid elasticity be provided across clouds and other environments?

Context: The cloud deployment model used by a company in a specific use case is often determined by the required level of accessibility, privacy, security and trust because private, public, community clouds significantly differ on these assurances.

Solution: A hybrid cloud integrates different hosting environments that can be accessed by different number of tenants and share underlying IT resources between different amounts of tenants.

Result: Through the establishment of a hybrid cloud the applications and their components required by a company may be hosted in multiple hosting environments. It can be used to add some of these properties to a static data center by integrating it with a cloud environment.

IV. CONCLUSION

The fundamental patterns form the basis for the understanding of the remaining patterns and should always be considered completely when designing cloud applications. Patterns for application workloads describe different user behavior resulting in changing utilization of IT resources hosting an application. This workload can be measured in the form of user requests, processing load on servers, network traffic, amount of data stored etc.

References:

[1] Cloud design patterns: prescriptive architecture guidance for cloud applications (Microsoft patterns & practices). MSDN Library. A Homer, J Sharp, L Brader, M Narumoto, T Swanson.

[2] Cloud computing patterns fundamentals to design, build and manage cloud applications Fehling, C:Leymann, F:Retter R: Schupeck w:Arbitter P. 2014 XXVI, 367p Hardcover ISBN:978-3-7091-1567-1

[3] Cloud computing patterns of expertise. IBM RedPaper. C Brandle, V Grose, My Hong, J Imholz, P Kaggali, M Mantegazza. 2014.

[4] Cloud computing design patterns. T Erl, R Cope, A Naserpour.

[5] http://www.cloudcomputingpatterns.org/static_workload/

[6] M. F. Barnsley, A. N. Harrington, The calculus of fractal interpolation functions, Journal of Approximation Theory 57 (1) (1989) 14–34.

[7] H. Wu, Y. Ding, C. Winer, L. Yao, Network security for virtual machine in cloud computing, in: 5th International Conference on Computer Sciences and Convergence Information Technology, 2010, pp. 18–21.

[8] R. Bobba, H. Khurana, M. Prabhakaran, Attribute-sets: a practically motivated enhancement to attribute-based encryption, in: Computer Security ESORICS, Springer, Berlin, Heidelberg, 2009, pp. 587–604.

[9] D. S. Kim, F. Machida, and K. S. Trivedi, "Availability modeling and analysis of a virtualized system," in *Proc. 15th IEEE Pac. Rim Int. Symp. Depend. Comput.*, Shanghai, China, 2009, pp. 365–371.

[10] L. Gomes and A. Costa, "Cloud based development framework using IOPT Petri nets for embedded systems teaching," in *Proc. 2014 IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Istanbul, Turkey, pp. 2202–2206.

[11] P. A. Dinda, The statistical properties of host load, Scientific Programming 7 (3) (1999) 211–229.