The Building Blocks to the Architecture of a Cloud Platform

Promise. Mvelase, Nomusa. Dlodlo, David. MacLeod, and Happy. Sithole

***Abstract*—** This research is on the identification of components that are general across cloud platforms. These components are classified under hardware, virtualisation and cloud management. These building blocks encompass nodes and clusters, hypervisors, virtual machines, billing and accounting, security and monitoring, to name but a few.

***Keywords*—**Cloud Platforms, Hardware, Cloud Management, Virtualization

I. Introduction

THERE are a number of definitions of the cloud. In [1] cloud computing is defined as follows:

"*Cloud computing refers to both the applications delivered as services over the Internet and the hardware systems and software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what is called the Cloud. When a Cloud is made available in a pay-as-you-go manner to the public, it is called a Public Cloud; the service being sold is Utility Computing. The term Private Cloud refers to internal datacenters of a business or other organizations that are not made available to the public. Thus Cloud Computing is the sum of SaaS and Utility Computing, but does not normally include Private Clouds"*

The National Institute of Standards and Testing (NIST) defines cloud computing under 5 identified characteristics as follows [2] o*n-demand self-service, which allows business units to get the computing resources they need without having to go through the IT department*

• *broad network access, which allows applications to be built in ways that align with how businesses operate today – mobile, multi-device, etc.*

• *resource pooling, which allows for pooling of computing resources to serve multiple consumers*

• *rapid elasticity, which allows for quick scalability or downsizing of resources depending on demand*

• *measured service, which means that business units only pay for the computational resources they use. Its costs match business success*

Reference [3] gives the operating definition of cloud computing as:

"Cloud computing provides on-demand network access to a computing environment and computing resources delivered as services. There is elasticity in the resource provision for users, which is allocated dynamically within providers' datacenters. Payment schemes are typically pay-as-you-go models."

Cloud computing, according to [3] is a combination of a technical architecture and a business model. The "computing environment and computing resources delivered as services" is most usefully disaggregated into applications, platforms and infrastructure all delivered as services. Applications are the software for users, platforms are the programming language-level environment for developing applications and infrastructure can include processing power, storage and user configurable 'virtual machines". Delivered as services, these elements are accessed via networks with users typically charged by the amount of usage – infrastructure elements are therefore virtualized.

Reference [4] defines cloud computing as the dynamic provisioning of IT capabilities (hardware, software or services) from third parties over a network. McKinsey on the other hand says that clouds are hardware-based services offering compute, network and storage capacity: where hardware is highly abstracted from the buyer; buyers incur infrastructure costs as variable operating expenditures; and infrastructure capacity is highly elastic.

Reference [5] conceives the cloud as:

"Clouds are a large pool of easily usable and accessible virtualized resources such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the infrastructure provider by means of customized SLAs'.

II. Nodes and Clusters

A cloud is made up of nodes and clusters. Nodes are physical servers. The nodes are connected to form a cluster. In

Promise.Mvelase is with the Meraka Institute (Centre for High Performance Computing), CSIR, Pretoria, South Africa: +2712 841 2569; fax: +2712 841 4720; e-mail: pmvelase@csir.co.za).

Nomusa. Dlodlo, is with Meraka Institute (Internet of Things Engineering Group), CSIR, Pretoria, South Africa.(e-mail: ndlodlo@csir.co.za).

David.MacLeod is with the Meraka Institute (Centre for High Performance Computing), CSIR, Pretoria, South Africa. (e-mail:dmacleod@csir.co.za).

Happy Sithole is with the Meraka Institute (Centre for High Performance Computing), CSIR, Pretoria, South Africa. (e-mail:hsithole@csir.co.za).

each cluster you have cloud software that manages the nodes in the cluster. Nodes in one and the same cluster have to be running the same software. Each cluster can operate independently of other clusters. Clusters in a cloud are managed by a central manager through one application programming interface (API). As shown in Figure 1, each cluster has a public IP and an internal IP. Machines inside the cluster are connected to the public interface via the internal machine IP.

The cloud maintenance software decides if a new customer coming in would like to submit a job into the cloud, that is, deploy a virtual machine. On each node sits the real data and data in the form of virtual machines. The data may not only be on the nodes, but it could also be in the storage area network (SAN). A virtual machine is a software-based, fictive computer. Virtual machines may be based on specifications of a hypothetical computer or emulate the computer architecture and functions of a real world computer. Virtualization is the process of creating logical computing resources from available physical resources. This is accomplished using virtualization software to create a layer of abstraction between workloads and the underlying physical hardware.

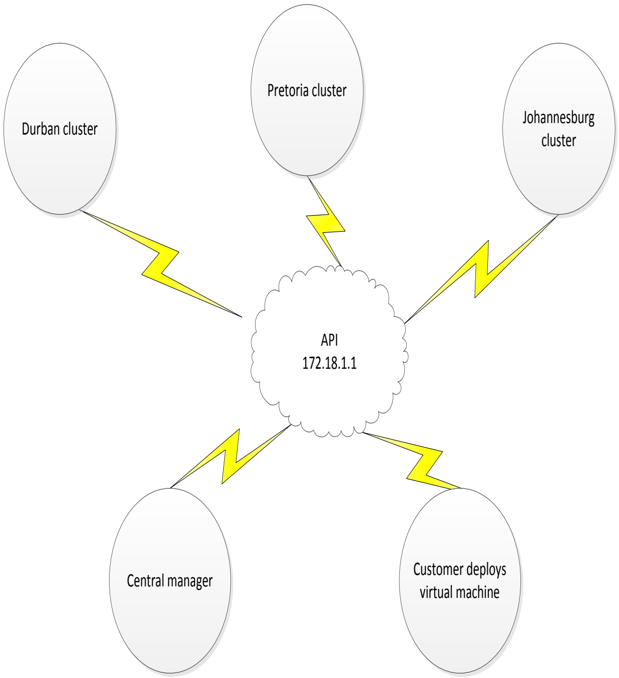

Fig. 1 Nodes and clusters

III. Hypervisors

A hypervisor or virtual machine monitor (VMM) is a piece of computer software, firmware or hardware that creates and runs virtual machines. A computer on which a hypervisor is running one or more virtual machines is defined as a host machine. Each virtual machine is called a guest machine. The hypervisor presents the guest operating systems with a virtual operating platform and manages the execution of the guest operating systems. Multiple instances of a variety of operating systems may share the virtualized hardware resources.

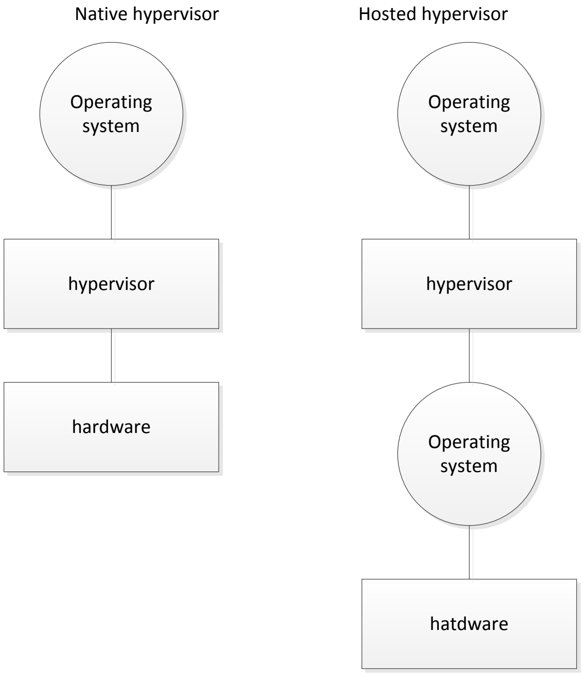

Fig. 2 Types of hypervisors

A hypervisor provides a virtual platform to run multiple operating system instances at once keeping each one of them unknown from one another. As said, a hypervisor partitions a physical server. There are two types of hypervisors: 1) the native or bare metal or hardware-based, and 2) the hosted or software-based (see Figure 2).

- The native or bare metal or hardware based hypervisor runs directly on the hardware of the physical server and has full control on that hardware so as to provide separate OS instances to the guest servers. An example of this type of hypervisor is VMware, XEN and Microsoft's Hyper-V. The so-created virtual server is separated from other VMs such that they are unaware that they are on the same 'parent server'. They are also called semi-dedicated servers and whatever the allocation is, in terms of hardware, its 'dedicated' allocation.
- With the hosted or software-based hypervisor, a pre-installed OS is required on the physical machine and so the control of the server is with the OS and not with the hypervisor. It includes VMware workstation, Microsoft Virtual PC and Parallels Workstation that are hosted on hypervisors.

The most used hypervisors are either VMware or Hyper-V, though VMware is suggested and opted for by many users and companies as VMware provides good features and high level of scalability. With the new VMware Vsphere (vCloud) the cloud computing has become more common nowadays as the user can scale his server in the run-time without affecting up-time.

Type 1 (or native, bare metal) hypervisors run directly on the host's hardware to control the hardware and to manage the guest operating systems (OS). A guest operating-system thus runs on another level above the hypervisor. Type 2 (or hosted) hypervisors run within a conventional OS environment. With the hypervisor layer as a distinct second software level, guest operating-systems run at the third level above the hardware. The hosted hypervisor runs on another OS from the host, while the native hypervisor runs on the same OS. In virtualization technology, a hypervisor is a software program that manages multiple operating systems (or multiple instances of the same OS) on a single computer system. The hypervisor manages the system's processor, memory, and other resources to allocate what each operating system requires. To write to storage on a physical device, the hypervisor talks to storage. Hypervisors are designed for particular processor architecture and may also be called virtualization managers. The native hypervisor is preferred over the hosted since the physical OS and hypervisor are built into each other. The hosted hypervisor has an additional OS which might not necessarily be optimized in all cases.

Various hypervisors have different features. Different hypervisors have other layers written on top of the hypervisor features which allow easy access. For example, a library of tools called LibVirt and VirtManager sit on top of the KVM hypervisor and also on the XEN hypervisor. These high level commands take away the specifics of particular hypervisor topography. The same commands in VirtManager and LibVirt can run on both KVM and XEN. KVM and XEN hypervisors don’t cost anything as they are open source software. Companies however, make money through controlling the hypervisors in the management of virtual machines including installing software.

In virtualization, a hypervisor is a software that servers as a mode of partitioning a physical server or host server into multiple OS based virtual servers or guest servers. A hypervisor is responsible to allocate resources (CPU, RAM, HDD) of the physical server or a dedicated Server among various virtual servers or virtual machines.

IV. Virtual Machine

A virtual machine (VM) is a software implementation of a machine (i.e. a computer) that executes programs like a physical machine. Virtual machines are separated into two major classifications, based on their use and degree of correspondence to any real machine:

- A system virtual machine provides a complete system platform which supports the execution of a complete operating system (OS). These usually emulate an existing architecture, and are built with the purpose of either providing a platform to run programs where the real hardware is not available for use (for example, executing software on otherwise obsolete platforms), or of having multiple instances of virtual machines leading to more efficient use of computing resources, both in terms of energy consumption and cost effectiveness (known as hardware virtualization, the key to a cloud computing environment), or both.
- A process virtual machine (also, language virtual machine) is designed to run a single program, which means that it supports a single process. Such virtual machines are usually closely suited to one or more programming languages and built with the purpose of providing program portability and flexibility (amongst other things). An essential characteristic of a virtual machine is that the software running inside is limited to the resources and abstractions provided by the virtual machine—it cannot break out of its virtual environment. The host OS runs on the hardware. The data on the hardware can only be manipulated by the host OS. On the host OS is the hypervisor software that mimics the hardware and becomes available to the virtual machine OS. This is associated with the type 2 hypervisor. This means that the virtual machine is not running on the physical hardware but on the hypervisor. Type 1 or native hypervisors run directly on the host's hardware to control the hardware and to manage the guest operating systems. A guest operating-system thus runs on another level above the hypervisor. This model represents the classic implementation of virtual-machine architectures.

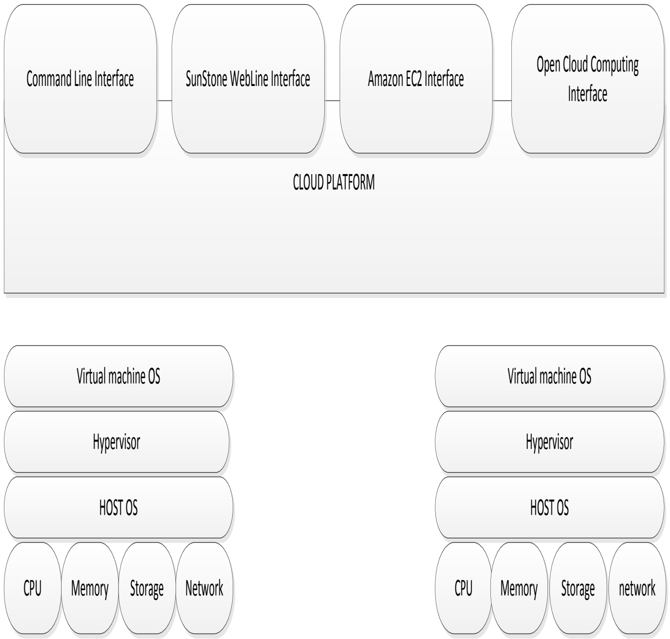

Fig 3 Cloud platform

The virtual machine OS takes space on the hardware when deployed, if type 2 is selected. Therefore space and random memory should be reserved on the node for the virtual machine to run, on type 2 hypervisors. The Cloud management infrastructure decides which physical node on the cluster to deploy the virtual machine (the hypervisor only provides virtualization). There is no human intervention as it is the software which makes decisions on which node and which cluster to deploy the virtual machine. Some cloud providers give the option for the customer to select where to deploy the virtual machine. They can only choose the cluster but not the node. This is decided by the software. If a cluster is overloaded, it is not made available to the customer. The interfaces such as Amazon EC2, Open Cloud Computing interface, Command line Interface and Sunstone WebLine

Interface afford communication with the cloud platform (see Figure 3).

A. Virtual Local Area Network

A virtual local area network (VLAN) is an isolated network on an Ethernet switch. Each VLAN in turn connects a number of physical machines. No unauthorized access is permitted to the VLAN except by the machines that are connected to the VLAN (see Figure 4).

Virtual LANs are connected to one another via a switch. InfiniBand is a low latency L2 protocol. InfiniBand is a switched fabric computer network communications link used in high-performance computing and enterprise data centers. Its features include high throughput, low latency, quality of service and failover, and it is designed to be scalable. The InfiniBand architecture specification defines a connection between processor nodes and high performance I/O nodes such as storage devices. InfiniBand host bus adapters and network switches are manufactured by Mellanox and Intel

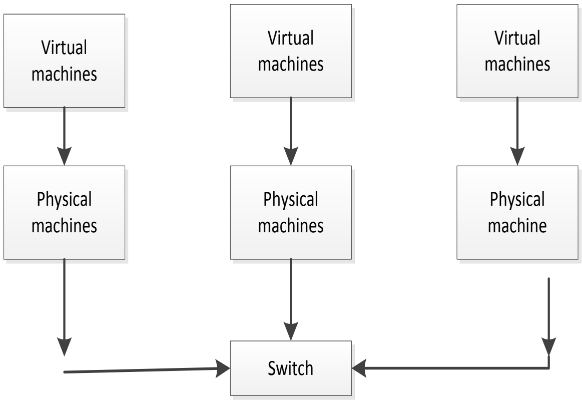

Fig. 4 Switch in VLANs

V. Virtualization

Virtualization is the process of creating logical computing resources from available physical resources. This is accomplished using virtualisation software to create a layer of abstraction between workloads and the underlying physical hardware. Once installed, the virtualised computing resources such as memory, CPUs, network and disk I/O and storage can all be pooled and provisioned to workloads without regard for physical location within a data centre.

Virtualization technology provides encapsulation that prevents workloads from accessing resources that are not assigned to it which allows a virtualised system to support multiple independent workloads simultaneously. Virtualisation is used to consolidate IT infrastructure, centralise system administration and management tasks, support workload scalability and agility and optimise the use of computing resources. This concept is the foundation of server consolidation, which has been the principal driver for virtualisation adoption.

Virtualization technology includes:

- Server virtualization - abstracts the server’s physical computing resources into logical entities that are provisioned to multiple workloads, allowing far more workload to exist independently on the same physical host and increasing the server’s utilisation.
- Storage virtualisation - abstracts and pools the storage resources available within the data centre, and allows storage to be provisioned to workloads that require it. This reduces the incidence of lost or “orphaned” storage and increases storage utilisation. Mainly on a storage area network (SAN).
- Network virtualization - allows a large physical network to be provisioned into multiple smaller logical networks, and conversely allows multiple physical LANs to be combined into a larger logical network. This behaviour allows administrators to improve network traffic control, organisation and security.
- Desktop virtualization - runs a complete desktop instance on a centralised server rather than the individual desktop PC, and provides that instance to a remote endpoint such as a simple thin client. Centralising desktop instances in the data centre allows better manageability and security.
- Application virtualization - used to virtualize specific applications rather than entire desktop instances. The virtualized application is operated on a centralised server and streamed across the LAN or WAN to client users that need access to that application. Application and desktop virtualization are often used in tandem.
- Mobile virtualization - allows multiple instances to exist on a smartphone, tablet or other mobile device.

VI. Billing/Accounting

Billing is about accounting for usage of physical hardware and virtual resources. It is not necessarily for payment monitoring usage. It is possible to cater for software that allows connecting to some other accounting system. are not italicized).

An excellent style manual and source of information for science writers is [9]. A general **PSRCENTRE** style guide, *Information for Authors,* is available at the web site.

A. Billing in the Cloud

One of the key attributes of cloud computing is the usage model. Customers consume resources as a service and pay only for what they use, rather than buying a license and annual maintenance. Regardless of whether the provider focuses on Infrastructure as a Service (IaaS), Platform as a Service (PaaS) or Software as a Service (SaaS), billing is the missing link. Cloud computing is based on a usage model where access to computing resources is delivered through Internet technologies. The user pays per usage, rather than buying a license and annual maintenance. Infrastructure costs (servers, DASD, network costs) are typically included in SaaS.

B. Infrastructure as a Service

Different hardware resources are provided through Infrastructure as a Service (IaaS).Charging model examples include the following:

• **CPUs:** CPUs are differentiated by power and number of CPU cores and, consequently, price. CPU power may be differentiated by time zone, e.g., static (based on peak and off peak resources) or dynamic, where price is determined by demand at the time. An extreme example of this concept is a two-way negotiated price between buyer and seller.

- **Server type:** Because the same CPU can be deployed either in a low cost server or in a top-of-the-range server with high availability and a significantly different cost point, the customer price must reflect this variation.
- **System administration:** The same server type resource may be charged at a different rate depending on the operating system (e.g., Windows or Linux). This is highly unlikely though because the client must administer the system in IaaS.
- **Storage (DASD):** Different storage capacity (including mirroring) is available, as well as different types of storage reflecting different price points from disk storage suppliers. The same variability exists as in the case of CPUs. For example, the cost for 1GB will vary depending on whether it’s provided in a low- or high-end unit.
- **Disaster recovery:** This involves the time window within which SaaS would need to be available should a disaster take out a data centre from which SaaS is provided. For short time window an active-active deployment in two data centres may be required.
- **Other:** Charges for space, power, network capacity, security, operating system and so on are built into the infrastructure pricing.

- **Service level agreements (SLAs):** If high availability is part of the agreement, SLAs may impact the price (e.g., refunds when contractual SLAs are not achieved).

Billing for IaaS may be done based on the quantity and quality of the infrastructure resources provided.

C. Platform as a Service

PaaS includes software frameworks and the necessary hardware in which to develop and deliver Software as a Service (SaaS). Examples of such frameworks include the following:

- Different hardware architectures with different server sizes—from small, Intel-based servers to mid- or top-range servers and mainframes—utilizing different chips
- Various software operating systems (e.g., Windows, Linux, MAC OS, Solaris, z/OS, and so on)
- Various development and application frameworks (e.g., Java, .Net)
- Solution stacks (e.g., LAMP, MAMP, WINS, and so on)

Billing must take into account Infrastructure as a Service (IaaS) costs, as well as software features and product offerings provided in the PaaS layer. Different frameworks have different prices and may include different infrastructures. All of this, together with usage, needs to be taken into account.

D. Software as a Service

Software as a Service (SaaS) may be delivered as a single or multi cloud offering. An example of a single cloud offering is Unified Communications (UC), which consists of different modules. An example of a multi-cloud offering is one including UC and ERP clouds. The assumption is that the cloud provider deploys all third party products necessary to run such offerings into the cloud.

VII. Security in the Cloud

This section looks at ways of implementing cloud security and the dimensions of cloud security.

A. Implementing Cloud security

Cloud computing security (sometimes referred to simply as "cloud security") is an evolving sub-domain of computer security, network security, and, more broadly, information security. It refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing. Cloud security is not to be confused with security software offerings that are "cloud-based" (a.k.a. security-as-a-service). Organizations use the Cloud in a variety of different service models (SaaS, PaaS, IaaS) and deployment models (Private, Public, Hybrid). There are a number of security issues/concerns associated with cloud computing but these issues fall into two broad categories: Security issues faced by cloud providers (organizations providing software-, platform-, or infrastructure-as-a-service via the cloud) and security issues faced by their customers. In most cases, the provider must ensure that their infrastructure is secure and that their clients’ data and applications are protected while the customer must ensure that the provider has taken the proper security measures to protect their information.

The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for customers or tenants of a public cloud service. Virtualization alters the relationship between the OS and underlying hardware - be it computing, storage or even networking. This introduces an additional layer - virtualization - that itself must be properly configured, managed and secured. Specific concerns include the potential to compromise the virtualization software, or "hypervisor". While these concerns are largely theoretical, they do exist. For example, a breach in the administrator workstation with the management software of the virtualization software can cause the whole datacentre to go down or be reconfigured to an attacker's liking.

1. Access Control

Cloud security architecture is effective only if the correct defensive implementations are in place. Efficient cloud security architecture should recognize the issues that will arise with security management. The security management addresses these issues with security controls. These controls are put in place to safeguard any weaknesses in the system and reduce the effect of an attack. While there are many types of controls behind cloud security architecture, they can usually be found in one of the following categories:

2. Deterrent Controls

These controls are set in place to prevent any purposeful attack on a cloud system. Much like a warning sign on a fence or a property, these controls do not reduce the actual vulnerability of a system.

3. Preventative Controls

These controls upgrade the strength of the system by managing the vulnerabilities. The preventative control will safeguard vulnerabilities of the system. If an attack were to occur, the preventative controls are in place to cover the attack and reduce the damage and violation to the system's security.

4. Corrective controls

Corrective controls are used to reduce the effect of an attack. Unlike the preventative controls, the corrective controls take action as an attack is occurring.

5. Detective controls

Detective controls are used to detect any attacks that may be occurring to the system. In the event of an attack, the detective control will signal the preventative or corrective controls to address the issue.

6. Dimensions of cloud security

Correct security controls should be implemented according to asset, threat, and vulnerability risk assessment matrices. Cloud security concerns can be grouped into any number of dimensions.

7. Security and privacy identity management

Every enterprise will have its own identity management system to control access to information and computing resources. Cloud providers either integrate the customer's identity management system into their own infrastructure, using federation or SSO technology, or provide an identity management solution of their own.

8. Physical and personnel security

Providers ensure that physical machines are adequately secure and that access to these machines as well as all relevant customer data is not only restricted but that access is documented.

9. Availability

Cloud providers assure customers that they will have regular and predictable access to their data and applications.

10. Application security

Cloud providers ensure that applications available as a service via the cloud are secure by implementing testing and acceptance procedures for outsourced or packaged application code. It also requires application security measures be in place in the production environment.

11. Privacy

Finally, providers ensure that all critical data (credit card numbers, for example) are masked and that only authorized users have access to data in its entirety. Moreover, digital identities and credentials must be protected as should any data that the provider collects or produces about customer activity in the cloud.

12. Legal issues

In addition, providers and customers must consider legal issues, such as Contracts and E-Discovery, and the related laws, which may vary by country.

VIII. MONITORING

Cloud monitoring can refer to the monitoring of the performance of physical and virtual servers, the resources they share, and the applications running on them. Cloud monitoring tools are often employed for this purpose and can aggregate data and show patterns that might be hard to identify otherwise. Cloud monitoring tools help an administrator keep cloud environments operating at peak efficiency. The dashboard tells about the network, speed, CPU utilization, allocated and used memory when provisioning memory or CPU.

Monitoring includes:

- Tracking virtual server instances
- Triggering event and notifications in case of server failures.
- Overseeing web servers, databases, mail servers, TCP ports and SSH access all based on user- rules for existing and automatically-launched new server instances.
- Adding monitors and notifications automatically for newly launched servers based on user-defined rules.
- Automatically deploying monitoring agents on new servers to monitor their performance and resource utilization.
- Eliminating time-consuming setups required with other uptime and end-user monitoring services.

IX. CONCLUSION

This research gives a review of the building blocks to the architecture of a cloud platform. These building blocks encompass nodes and clusters, hypervisors, virtual machines, networks and network isolation, billing and accounting, elasticity, cloud storage, security and monitoring, to name but a few.

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

[1] "Cloud computing as a business continuity plan," http://www.tectonic.co.za/2009/05/cloud-computing-as-a-business-continuity-plan/.

[2] "How cloud computing can transform business," http://blogs.hbr.org/cs/2010/06/business_agility_how_cloud_com.html

[3] K. E. Kushida, D. Breznitz, J. Zysman, "Cutting through the fog: understanding the competitive dynamics in cloud computing," The Berkeley Roundtable on the International Economy,(BRIE) Working Paper 190 (Beta), May 1, 2010.

[4] McKinsey & Co. "Clearing the air on cloud computing," Report presented at Uptime Institute Symposium. http://uptimeinstitute.org/content/view/353/319, April 18 2009.

[5] L. M. Vaquero, l. Rodero-Merino, J. Caceres, M. Lindner, "A break in the cloud: towards a cloud definition", ACM SIGCOMM Computer Communication Review, Vol. 39, No. 1, pp. 50 – 55. January 2009.