Survey of the State-of-the-Art of Cloud Computing

Sanjay P. Ahuja, University of North Florida, USA

Alan C. Rolli, University of North Florida, USA

ABSTRACT

Cloud computing as a computational model has gathered tremendous traction. It is not completely clear what this term represents though it generally is thought to include a pay-as-you model for computation and storage. This paper explains what Cloud Computing is and contrasts it with Grid computing. It describes the major cloud services offered, discusses architectural details, and gives details about the infrastructure of Cloud Computing. This paper surveys the state of cloud computing and associated research and discusses the probable directions to the future of this evolving field of computing.

Keywords: Benefits, Cloud Computing, Cloud Services, Computational Model, Grid Computing

1. INTRODUCTION

In the early 1990s, Ian Foster and Carl Kesselman came up with the idea of "The Grid"; applying the same concept of plugging into a grid for metered utility service to computing (Wallis, 2008; Douglis, 2009). This concept has transformed into what has been coined Cloud Computing. With Cloud Computing, data is stored in the "cloud" of the internet where web-based applications are utilized to access the data and perform various tasks. Amazon was one of the first to provide service, the Elastic Compute Cloud (EC2). EC2 charged for their resources by the hour, like the electric company charges their clients per kilowatts an hour of usage. Presently, there are three major categories of Cloud Computing systems offered: software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) (Viega, 2009). Each of the aforementioned services as well as the emerging Database-as-a-Service provides customers a variety of choices to meet their specific needs.

1.1. Comparing and Contrasting Grid and Cloud Computing

On the outside there is little difference between a Grid and a Cloud environment. They both are a collection of machines such as servers, network devices, and computers that, to the user, appear as a single resource. This resource provides services such as Web infrastructure, databases, application support and more. So what truly sets these two environments apart?

Grids were originally designed to help provide a way for researchers to allocate large

amounts of resources in order to perform some complex computation that would take a significant amount of time longer to accomplish with a single machine. In the Grid, scientists could allocate as many nodes in the environment as necessary to complete their task, and leave the remaining nodes available for other such computations. The nodes are allocated only for as long as it takes to complete the task. Once the task is completed, the nodes are leased back into the Grid to be allocated to another process.

In contrast to Grids, Cloud environments are geared towards smaller requests for resources. Instead of requesting 2000 nodes in order to perform some difficult task, one might request enough resources in order to meet their individual, business, or personal, needs. The resources in a Cloud are more semi-permanent as opposed to a Grid. In other words, the allocated resources in a Cloud will remain persistent until the client decides to cancel the service.

Another difference is that the environment in a Cloud will generally be a virtual environment. What this means is that if somebody were to request a resource such as a Web server, they would not be given an entire server in which to host their environment. Instead they would be allocated a virtual environment, there could be several such environments running on each physical server, where the user's Web server would run.

While research on Grid systems has been expanding on working towards being able to tie resources from different administrative domains, Cloud Computing systems are currently restricted to a single domain.

An additional key difference in cloud computing is how data is stored and distributed. Relational databases, as typically used in an Enterprise environment, are not very scalable because the data cannot easily be distributed across the domain. A relational database takes a huge performance hit if the files are distributed across a large number of nodes. If the workload doubles or triples overnight, you would have to upgrade the server quickly, which is not an easy task.

By using a distributed hash table you can distribute your data across thousands of nodes without having to have one or even several centralized database servers. Not only is this good for redundancy, it also allows for faster insertion and retrieval for things such as Web services.

1.2. Benefits of Cloud Computing

As you can see from the comparing and contrasting the Grid to Cloud Computing, there are clear advantages to Cloud Computing (Figure 1). One of the immediate benefits is that of time-to-application deployment. Cloud platforms allow you to develop and deploy new applications on existing infrastructure as quickly as you can write the code. Companies are also benefiting from the ability to scale their IT budget as demand grows, freeing up equity.

1.2.1. Operational Benefits

Server uptime and availability are very important to any company, regardless of the size of the business. The cloud-based infrastructure provides better uptime and availability by using technologies that have been developed for distributed systems for redundancy of resources. As discussed earlier, a Cloud environment is very similar to a grid in that the processes are generally distributed throughout the network. A properly configured Cloud environment will be fault tolerant and have the ability to recover from lost resources by reallocating new ones. In order to accomplish this Amazon allows users to place their work in what they call "Availability Zones" (Thorsten, 2009). An Availability Zone is essentially a separate data center where a user can have their system replicated for failover. If one data center goes (or simply one of the servers crash), then there is a replicated server in another location in which traffic can be routed to.

Besides having reliable and stable servers, the ability for deployment teams to launch new projects and bring new virtual servers up almost instantaneously is a largely attractive

Related Content

V2V Communication Protocols in Cloud-Assisted Vehicular Networks

Amal Ahmed Eltahirand Rashid A. Saeed (2018). *Vehicular Cloud Computing for Traffic Management and Systems (pp. 125-150).*

www.irma-international.org/chapter/v2v-communication-protocols-in-cloud-assisted-vehicular-networks/206613

Addressing Privacy in Traditional and Cloud-Based Systems

Christos Kalloniatis, Evangelia Kavakliand Stefanos Gritzalis (2015). *Cloud Technology: Concepts, Methodologies, Tools, and Applications (pp. 1631-1659).*

www.irma-international.org/chapter/addressing-privacy-in-traditional-and-cloud-based-systems/119924

Cloud Computing and Cybersecurity Issues Facing Local Enterprises

Emre Erturk (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications (pp. 947-969).*

www.irma-international.org/chapter/cloud-computing-and-cybersecurity-issues-facing-local-enterprises/224616

Software Engineering for Developing a Cloud Computing Museum-Guide System

Hadeel Al-Obaidy, Aysha Ebrahim, Ali Aljufairi, Ahmed Meroand Omar Eid (2024). *International Journal of Cloud Applications and Computing (pp. 1-19).*

www.irma-international.org/article/software-engineering-for-developing-a-cloud-computing-museum-guide-system/339200

Multi-Cloud Adoption Challenges for the Cloud-Native Era: Best Practices and Solution Approaches

Pethuru Raj Chelliahand Chellammal Surianarayanan (2021). *International Journal of Cloud Applications and Computing (pp. 67-96).*

www.irma-international.org/article/multi-cloud-adoption-challenges-for-the-cloud-native-era/274339