CMPSCI 677 Operating Systems **Spring 2015**

Lecture 23: April 22

Lecturer: Guest Lecture: Tian Guo *Scribe: Jing Xie*

23.1 Case Study on Data Center s and Cloud Computing

23.1.1 Data Centers

Data centers are collections of servers and storage farms for supporting applications that require distributed storage and computations. They are used by Enterprises and Internet companies for running their website, business applications and other application which require data processing. Spire is an example of an application that can be run on a data center. Two kinds of Data Center Architecture include Traditional and Modern Architecture. In the Traditional architecture there is a static mapping between the applications and the servers they run on. A system administrator decides the mapping and has to manually manage the servers. However with the fluctuations in the workload this static mapping might not the most effective technique. Modern Data centers provide a flexible mapping and also have increased automation thus handling the needs of these applications.

23.1.1.1 Components

- Racks of servers
- Storage arrays
- Networking components like routers
- Cooling infrastructure to handle the heat generated by the servers
- Electricity to power the servers
- USB/battery backup/generators for backup power

Another way to handle the fluctuating workloads is to make use of Modular Data Centers. In this technique the containers contain servers and are shipped based on demand. They are much effective since these servers and preinstalled and configured. The costs involved in installation is avoided and the data centers can also expand easily.

One major advantage of Data center is the use of virtualization. A single physical 16 core machine can work as 4 quad core machines working with different platforms and with different functionality. With the VMs adjustments in resource allocations can happen on the fly. The slide includes an animation which depicts this. As discussed in previous lectures VM migration can achieved within a LAN .

The data centers can be used either as Virtual servers or a virtual Desktop. The main advantages of using a data center as a virtual server include faster deployment and easier maintenance. In virtual Desktops, a low end machine is given to the user and a virtual machine on a remote server. The user using the low end machine connects to the desktop on the virtual machine . This is called thin client computing. This is much easier

to manage and maintain. Two major challenges that the data center face include Resource Management and Energy Efficiency. Resource Management includes handling unpredictable workloads, working towards higher performance, achieving efficient usage of servers and storage resources. Energy efficiency is one major concern since servers consume huge amounts of energy. Hence use of renewable sources of energy is preferred. However there are issues with the solar or the wind energy since they are not reliable and the system irrespective of the conditions has to cater the needs of the user .

23.1.2 Cloud Computing

Cloud is a data centers were the user can request machine on demand and pay for what is being used. It uses a shared infrastructure were a every server is shared among different users.

23.1.2.1 Types of cloud

- Software as a Service

  The user's application is hosted on a cloud. They are managed by the provider. Some of the applications include Gmail, Google Docs , Sales Force. In case of gmail, the mail server is provided by google, the user logs in and access the server. The user is abstracted from the hardware or the software used.

- Platform as a Service

  The user gets hardware and some software. Applications can be developed and deployed by the user. The applications are managed by the provider. Based on the needs of the application the provider assigns resources and takes care of the needs of the application. Some examples include Azure and Google App Engine.

- Infrastructure as a Service

  The user rents a bare bone machine and storage. No software or application is available. The user can use any software, OS and applications. Some examples include Amazon web services and at&t.

Google App Engine It is an example of a PaaS. The user is provided the hardware, a set of softwares and can build the application in python or Java. The scalability and performance are managed by google. Google figures out how many threads to allocate for an application based on the need of the application. Its not based on Virtualization.

Amazon EC2 The user rents a bare bone machine and storage. The User decides on the kind of machine needed to serve his/her application. The machines can be small/medium/large based on the need. The machines can be provisioned in minutes unlike buying machine which take quite sometime.

23.1.3 Cloud Computing

- Private Cloud: Enterprises use clouds for their internal usage . For example the IT department of the company or a University like UMASS might build a cloud just for its user base.
- Public Cloud: Anybody can be an end user. The cloud providers like amazon rent their services to any end user.
- Hybrid Cloud: In some cases when the enterprises use the private clouds and have resource needs that cannot be handled by the private cloud they dynamically pull resources from a public cloud. **Cloud bursting** is the process of moving to the public cloud when the workload goes beyond the private clouds Capacity.

23.1.3.1 Programming Models

The programming models include traditional Client server programming models or use of Batch processing or Map reduce models. The map reduce models are generally used for data intensive applications. It is a data parallel model where multiple servers are used to process the data. For example facebook applies mapreduce to

find the most popular pages and decide the rates of ads based on popularity.

23.1.3.2 Challenges in Cloud

One major concern with Cloud is Privacy/Security. Multiple users share a server and the provider has to be make sure that the users are isolated and the performance of an user is not affected by another user. With Extreme Scalability, management becomes a problem. Methods to automatically manage resources have to be incorporated. Appropriate Programming models have to be used to make optimum use of the large number of available servers.