

# Chapter 1

## Literature Review

### 1.1 Background

This chapter begins by providing an overall understanding of Cloud computing and its related research field. Then, it narrows down to the server consolidation problem in Section ??.

#### 1.1.1 Cloud Computing

1. What is Cloud computing ?
2. Service Model
3. Resource management

In order to understand Cloud computing, firstly we will illustrate the five essential elements of Cloud computing and their advantages.

Cloud computing has five essential elements:

1. On-demand self-service, it means a Cloud user can require computing resources (e.g CPU time, storage, software use) without the interaction with Cloud provider.
2. Broad network access, Computing resources are connected and delivered over the network.
3. Resource pool, a Cloud provide has a “pool” of resources which are normally virtualized servers. In IaaS, it provides predefined sizes of VMs. In PaaS, the resources are ‘invisible’ to Cloud users who have no knowledge or ability to control.
4. Rapid elasticity, from the perspective of Cloud users, computing resources are assigned and released in real time. In addition, the resources assign to their software is “infinite”. Therefore, Cloud users do not need to worried about the scalability of their applications.
5. Measured Service provides an accurate measure of the usage of computing resources. It is fundamental to the pay-as-you-go policy.

What is your purpose to describe the following content? I would like to discuss the differences, advantages of disadvantage of the resource management in different service models. Therefore, after illustrate how they are work. The point is to compare the resource management. And then, lead to a new service model. And the advantage of new service

model should be obvious.

Traditional Cloud computing has three service model as illustrated in Figure ??.

1. Infrastructure as a Service, Cloud provider offers the fundamental computing resources, often in the form of various sizes of VMs. Apart from the virtualized hardware and operating systems, Cloud users treat the remote servers as local and deploy their applications. In terms of resource management, Cloud users have the responsibility to estimate the quantity of resources, while Cloud providers have no knowledge and control inside VMs, resource management is based on VM.
2. Platform as a Service, Cloud providers establish the software development platform to enable the in-progress software to be developed in the platform. The main difference between PaaS and SaaS is that, PaaS supports the full life cycle of software development, whereas SaaS only host completed applications deployment. In terms of resource management, Cloud providers have the full control of resource allocation, auto-scaling and consolidation. Therefore, Cloud users can focus on software development.
3. Software as a Service (SaaS). Cloud users deploy their applications in Cloud which can be accessed by End Users. SaaS describes the relationship between Cloud provider and End users with the connection of applications.

Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service.

### 1.1.2 Resource Management

Scope of Cloud computing resource management.

1. Actors
2. Management Objectives
3. Resource Types
4. Enabling Technologies

Server Consolidation

# Bibliography