

Discrete Optimization COMP 567-Semester

Problem Formulation

Students: Hanqiang Cheng

Optimal Cloud Purchasing on Amazon EC2

Problem Introduction:

Cloud Computation refers to both the software and hardware in the large data center delivered as commodity. Cloud computing users could customize their computing resources according to what they really need therefore reduce their capital expenditure. In addition, cloud computing allow company to start small and dynamically increase their computing capacity according to their business increase. Amazon EC2 is the leading cloud provider in the world and is widely used. EC2 cloud provides lots of pricing options for different kinds of users to choose, such as data center location, on-demand instance or reserved instance; small size instance, medium size instance or large size instance. More details could be acquired from <http://aws.amazon.com/ec2/#pricing>.

Problem statement:

Suppose we have an online business service and want to deploy it on EC2. The workload of this service varies from hour to hour every day and would possibly increases at a constant rate every 3 months. Based on the dynamic characteristics of the workload, we want to make smart choice on what kinds of instances and how many instances we need to minimize our capital expenditure on cloud purchasing in one year, at the same time meet the requirements of business workload.

Constraints:

1. To ensure service availableness and reduce the impact of potential downs of one cloud, we want the service to be deployed in two locations and every location should at least hosts 30% of the total workloads.
2. The total cloud we purchase should be able to host more workloads than that of our prediction.
3. To ensure quality of service, every type of computing instance in the cloud has its limit on how many workloads they could support. For example, we could define the maximal workload for a small size instance is 900 service requests concurrently.