Capacity Planning

14 november 2014

Identify the best solution from the point of view of the costs for a web server that must be hosted on a public cloud. There is the possibility to use one of two providers.

The first provider rents resources at day, while the second rents resources at hour.

In both cases, it is necessary ensure the following SLA:
- a response time of less than 0.2 seconds
- a loss of requests less than 2% and
- an availability of 99%.

The site receives 10 requests per second from 0:00 to 12:00, while it receives 20 requests per second from 12:00 to 24.

These requests must be served by a cluster of identical servers. A load balancer divides the load equally among the servers.

The first provider provides resources that are able to meet the service demand with the following parameters:
- 10 ms of the CPU
- 30 ms of I / O disk

and each server is able to serve not more than 3 clients at a time.
The cost for this type of server is 12 euro per day.

The second provider provides resources that are able to meet the service demand with the following parameters:
- 20 ms of CPU
- 20 ms of I / O disk

and each server is able to serve not more than 4 clients at a time.
The cost for this type of server is 0,5 euro per hour.

Moreover, in both cases, the CPU have an MTTR of 2 years and a MTTF of 1 month.