

Paper Review

“ CloudScale: Elastic Resource Scaling for Multi-Tenant Cloud Systems ”

1. Summary

Multi-tenant is common in cloud service providers, which partitioned a host machine and sell VM (virtual machines) to various customers. Host sellers usually have a couple of hosts and manage VMs. For example, they can use the KVM manager to migrate one VM from one host to another without shutting down the machine. The migration will be performed when host machines do upgrading.

This paper uses intelligent algorithms to migrate, for energy-saving and SLO providing. Cloudscale is a system that automates fine-grained elastic resource scaling for multi-tenant cloud computing infrastructures, and is focused on resolving scaling conflicts.

The previous work like Overbooking restricts the resource but cannot meet the elastic demands. The challenges CloudScale met are: over and underestimating demand, frequent migration takes time.

To solve online adaptive padding, Cloudscale will add a small extra value to the predicted resource demand. If the CPU usage is bursty, the padding will be large. It collects errors and does classification, and uses a weighted moving average (WMA) to calculate the weight of error to do Remedial padding.

Similarly, CloudScale will multiply the resource allocation by a ratio $\alpha > 1$, which is also computed by statics to avoid excessive resource waste and under-estimating.

To deal with scaling conflicts when the available resources are insufficient to accommodate all scale-up requirements on a host, CloudScale does conflict prediction instead of rejecting or migration. The scaling is first depending on the priority. Otherwise, CloudScale allocates the residual resources to all the remaining applications in proportion to their resource demands.

The migration is also done with the prediction by the metrics of CPU and MEM.

CPU frequency and voltage are also considered. CloudScale will slow the CPU as long as providing SLOs.

2. Advantages

- + CloudScale take CPU voltage and frequency into consideration.
- + The intelligent schemes are reasonable.

3. Disadvantages

- The metric is not such reasonable, I think a new classifier can be used to check whether the metric is reasonable.
- More architectural work can be done, such as the allocation of memory, vcpu.

4. Brainstorming

In our project, as a deadline predictor and a scheduler, we also do prediction by collecting the metrics. We would like to compute the throughput of the program and put the program into the host with proper residual resources.