Project Proposal

Team Members: Rutwik Ghag, Puneet Shetty

Team Name: Ice and Fire

Topic: Cloud Capacity Planning

Reasoning:

Capacity Planning is to make sure there will be enough resources available when they are needed. That is, it should be done so that the system is neither under- nor over-capacity. Every aspect of a service's infrastructure, such as central processing units, memory, storage, server instances, network throughput, switch ports, console connections, electricity, cooling, datacenter space, and so on, is considered a resource.

Capacity planning aims to accomplish two main things. The primary goal is to eliminate capacity-related service disruptions. Secondly, to limit future expenditures to necessary expansions of existing facilities. Capacity planning that shows the requirement for resources and caps resource utilization at a level that ensures good service delivers a measurable return on investment (ROI).

Large, rapidly expanding services necessitate the use of complex mathematical models for capacity planning. It is common practice for businesses to employ a statistician full-time to create and update such models. An expert statistician with a technical background is hard to come by, but well worth the search. Keeping the performance of servers at the same level after expansion is difficult but something that everyone strives for.

We aim to study the various approaches to Cloud Capacity Planning and summarize them in a comprehensive research paper.