DS 7346 Section 402

Ballenger, Chris: cballenger@mail.smu.edu
Kasarla, Venkat: vkasarla@mail.smu.edu
Tom Gianelle: tom.gianelle@gmail.com
Blaine Brewer: mbbrewer@mail.smu.edu

Cloud Instance Automation

Problem Statement:

Cloud computing is one of the fastest growing segments of the tech industry. Gartner predicts that the worldwide public cloud service market will grow 12.6% year-over-year. Its also projected that IaaS and PaaS cloud computing usage will have grown 27.5% and 21.8% by the end of 2019, respectively. With this growth and the competitive nature of the market, the technology is constantly evolving. It requires full-time dedication to stay apprised of the constantly shifting technology and pricing. In fact, since the pricing remains so fluid, its virtually impossible to stay on top of current pricing and how the pricing compares between the various cloud providers and their services. Whether your team is working on a development project or promoting your process to production, it's difficult to consider the costs associated with setting up your environment and finding that sweet spot in the sliding scale of compute and storage.

In *The Resource Allocation Optimization Problem for Cloud Computing Environments*, a research paper that explores the process of walking the tightrope of cost versus performance while deploying a cloud model, the researcher creates each instance, installs anaconda, and initiates the models manually to calculate a dollars per compute hour. He uses this metric to evaluate and compare the varying pricing structures between the providers and the offerings from those providers.

In a business setting, a team of developers might need to quickly spin up an inexpensive test environment, report pricing to their executives, or build a production ready process with massive storage and compute but while taking pricing into consideration. Building a mechanism to deploy models while incorporating transparent pricing could save corporations a lot of money by only deploying what they need and only paying for what they use.

Project Proposal:

Our team aims to research and develop a way to automate the initialization of a cloud instance with compute and storage while incorporating some of the findings from the research team in the aforementioned paper.

We understand that this is a lofty goal, but were confident that exploring the different IaaS, Paas, and Saas options from the major cloud providers, along with some out-of-the-box solutions such as Docker and Kubeflow, we will be able to quickly deploy instances that store data and successfully test and train models for production. We will use the assistance of a competitive pricing matrix to factor in the decision on which service provider (e.g. AWS, Azure, Google) and service tier to use. After all, forcing cloud providers to compete not only promotes cost savings for the consumer, but also promotes industry-wide innovation.

Citations

Columbus, Louis (2017) “Roundup Of Cloud Computing Forecasts, 2017”

Available at: https://www.forbes.com/sites/louiscolumbus/2017/04/29/roundup-of-cloud-computing-forecasts-2017/#4e5caa0331e8

Yim, Victor and Fernandes, Colin (2018) "The Resource Allocation Optimization Problem for Cloud Computing Environments," *SMU Data Science Review*: Vol. 1 : No. 3 , Article 2.

Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/2