# Introduction

With the advent of Cloud technologies, there is access to almost infinite capacity. With expandable computation power and memory, application of cloud services have become limitless. With a flexibility to pay per use, variety of enterprise IT needs can be satisfied without need for dedicated servers. This has resulted in major IT players like Amazon, Microsoft and Google to provide their cloud services Amazon Web Services, Microsoft Azure and Google Cloud respectively [Dimitrios Sikeridis et al, 2017].

Out of their several cloud services, the two key ones are memory and computation. In case of Amazon, their cloud memory service is called S3 and cloud computation service is called EC2 (Elastic Computing). Computation services are generally costlier compared to memory. Typically cloud service providers look for options to minimize the cost for computation. One of the options available to reduce the cost is to use the left over computational instances which comes at a much cheaper rate.

Amazon Web Services provide different pricing options for instances like On Demand Instance, Reserved Instance and Spot instance. On-demand instances let users dynamically acquire compute capacity one hour at the time without any prior commitment, but come at a premium. In contrast, reserved instances call for a one or three year commitment, but enjoy a significant discount compared to on-demand instances. Finally, spot instances, like on-demand instances come with a one hour time granularity and do not require any type of commitment. They are left over instances which can be significantly cheaper than on-demand instances. However, they have a lot of price fluctuations and availability is not guaranteed [Jiayi Song, Roch Guerin, 2017].

Spot instances provide cost advantage but also poses challenges on meeting SLAs [Shaojie Tang et al, 2014]. There is no solution found in the literature which provides prediction of availability of spot instances which can help agree on SLAs. This project proposes a methodology to predict Availability of a spot instance which can be used to choose the right bid price which will help them meet SLAs and also minimize cost.

# Problem Statement

1. What is the correlation between Bid Price and Availability?
2. Is there is multi-collinearity between Spot Price and Availability?
3. Can regression techniques help predict Availability
4. What features will be needed for building models?

The project aims at building causal relationship between Availability, Spot price and Bid Price using visualization and correlation techniques. Having established the causal relationship, regression models will be built to estimate the empirical relationship between the 3 metrics and use it for predicting Availability of spot instances. Multiple Non Linear regression technique using Ordinary Least Squares algorithm is proposed to be used

# Value

EC2 spot instance is a virtual machine (VM) that is cheaper than its on-demand or reserved counterpart. Prices of Spot instances are generally 1/10th of that of the price of on-demand instances. However it has poorer availability; here the definition of availability being “What is the expected lifetime of an instance?” [Wang, Cheng et al, 2017].

Because the spot-instance market mechanism does not provide a way to guarantee how long an instance will run before it is terminated as part of an SLA, market prices are often significantly lower by up to an order of magnitude than fixed prices for the same instances with a reliability SLA. That is, because a user cannot determine a bid that will ensure a specific level of reliability in the spot market, this uncertainty generally leads to lower prices. However, users who wish to ensure that their instances will be reliable must submit large maximum bids [Wolski, Rich et al, 2017]. If there are 2 bidders and one spot instance available, the one with the highest bid will be given the instance. However, the bid will continue. In case, another bidder bids higher than the current user, the instance will terminate within short notice and transferred to the new bidder.

The research problem of predicting Availability of spot instance, when solved provides lot of value as they can save cost and at the same time help maintain SLAs. This problem is also not seen to have been solved in the literature and hence should be a value to research community.

# Reference:

[1] Dimitrios Sikeridis, Ioannis Papapanagiotou, Bhaskar Prasad Rimal, Michael Devetsikiotis: “A Comparative Taxonomy and Survey of Public Cloud Infrastructure Vendors”. CoRR abs/1710.01476 (2017)

[2] Jiayi Song, Roch Guerin: “Pricing and bidding strategies for cloud computing spot instances”, 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)

[3] Shaojie Tang ; Jing Yuan ; Cheng Wang ; Xiang-Yang Li; “A Framework for Amazon EC2 Bidding Strategy under SLA Constraints”; IEEE Transactions on Parallel and Distributed Systems; 2014

[4] Wang, Cheng and Liang, Qianlin and Urgaonkar, Bhuvan; “An Empirical Analysis of Amazon EC2 Spot Instance Features Affecting Cost-effective Resource Procurement”; Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering; 2017

[5] Wu, Hao & Ren, Shangping & Timm, Steven & Garzoglio, Gabriele & Noh, Seo-Young; “Experimental Study of Bidding Strategies for Scientific Workflows using AWS Spot Instances”; 2015

[6] Wolski, Rich and Brevik, John and Chard, Ryan and Chard, Kyle; “Probabilistic Guarantees of Execution Duration for Amazon Spot Instances”; Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; ACM; 2017

[7] Ashish Kumar Mishra and Dharmendra K. Yadav; “Analysis and Prediction of Amazon EC2 Spot Instance Prices”; International Journal of Applied Engineering Research; 2017

[8] Bahman Javadi, Ruppa K. Thulasiram, and Rajkumar Buyya; “Statistical Modeling of Spot Instance Prices inPublic Cloud Environments”; 2011 Fourth IEEE International Conference on Utility and Cloud Computing; 2011