Data Mining Project Proposal

Kaan Dandin & Erdi Ölmezoğulları

Selecting Correct Resource from AWS Spot Instances

23/11/2015

Data Mining Project Proposal

In this proposal, we will briefly explain our project details.

In our project, we will use public data, which was collected by third party people and released through some specific websites. Since our data will be mainly related to Amazon Web Services' (AWS) Elastic Computing (EC2), it will be consisting of some different fields. EC2 is a kind of virtual machine in the AWS's cloud.

A virtual machine can be created just in time either on private or public cloud over AWS whenever you need it. A new virtual machine can be picked with respect to different specs and configurations in terms of CPU, RAM, storage, and network band limit before creating it once from scratch. EC2 machines also are separated and managed by AWS on different geographical regions (US East, US West, EU, Asia Pacific, South America) and zone to increase availability of virtual machines across the world. AWS has different segmentations, which were classified with respect to system specs by AWS for based on different goals (macro instance, general purpose, compute optimized, storage optimized, GPU instance, memory optimized). Payment options are dedicated, on-demand and spot instance. Since they make different cost to customer's operation, customers may prefer different kinds of virtual machine according to their goals and budgets. In general, spot instance is cheaper than the rest of the options. However, spot instance may be interrupted if market price exceeds our max bid.

In our research, we will focus on spot instance payment. Our aim in this project will be selecting correct AWS instance from the Spot Instance Market according to the requirement of the customer. We plan to perform Decision Tree on streaming data to make a decision on the fly. It may be implemented as an incremental version of decision tree since data is changing continuously. (Figure 1.) At the end of this project we consider to make a decision to the customer to pick the optimal option from the data.

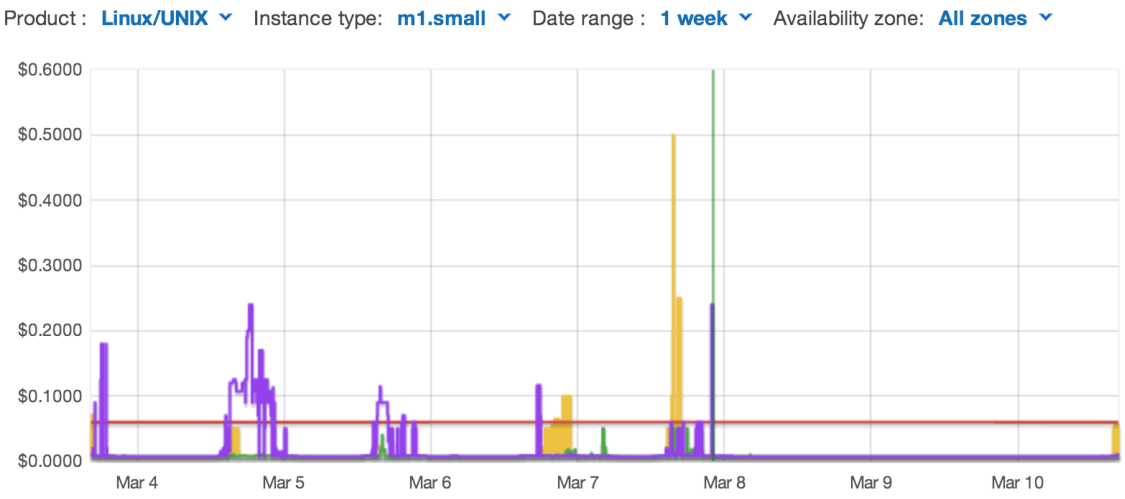

Figure 1) Spot instance price change for a specific machine type m1.small

Data Mining Project Proposal
Data Mining Project Proposal

Amazon Cloud Services

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable computer capacity in the cloud. It is designed to make web-scale cloud computing easier for developers.

Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios. (Ref.1)

Amazon Instance Types

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload. Instance type examples are t2.micro, t2.small, t2.medium, t2.large…

Using Spot Instances

Cloud providers possessing large quantities of spare capacity must either incentivize clients to purchase it or suffer losses. Amazon is the first cloud provider to address this challenge, by allowing clients to bid on spare capacity and by granting resources to bidders while their bids exceed a periodically changing spot price. (Ref. 2)

Amazon EC2 Spot instances are spare EC2 instances that you can bid on to run your cloud computing applications. Since Spot instances are often available at a lower price, you can significantly reduce the cost of running your applications, grow your application's compute capacity and throughput for the same budget, and enable new types of cloud computing applications. (Ref.3)

Ref:1) https://aws.amazon.com/ec2/

Ref:2) Deconstructing Amazon EC2 Spot Instance Pricing Orna Agmon Ben-Yehuda Muli Ben-Yehuda Assaf Schuster Dan Tsafrir Technion – Israel Institute of Technology {ladypine, muli, assaf, dan}@cs.technion.ac.il

Ref:3) https://aws.amazon.com/ec2/spot/

Data Mining Project Proposal
Data Mining Project Proposal