**Machine Learning Engineer Nanodegree**

## Capstone Proposal

Samar Shaban haytamy  
8 May 2018

## Proposal

### Domain Background

Cloud computing is conducting business platform [1]. Big companies such as Amazon, Microsoft, Google and IBM are oﬀering cloud computing solutions in the market. A fast increasing number of organizations are already outsourcing their business tasks to the cloud, instead of deploying their own local infrastructures [2]. An advantage of cloud computing is its economic beneﬁts for both users and service providers.

Compared to traditional service composition, cloud service composition is usually long-term based and economically driven. Traditional quality-based composition techniques usually consider the qualities at the time of the composition [3]. For example, which composite service has the best performance at present?

This is fundamentally diﬀerent in cloud environments where the cloud service composition should last for a long period. For example, which composite cloud service performs best in the next few years, despite it may not be the best one at present?

### Problem Statement

the cloud environment has four actors: End Users, Composer, SaaS (Software as a Service) Providers and IaaS (Infrastructure as a Service) Providers. Platform as a Service (PaaS) layer is omitted as we assume that it is included in the IaaS layer. End Users are usually large companies and organizations, e.g., universities, governments. The composer represents the proposed composition framework. ***The composer acts on the behave of the end users to form composite services that contains services from multiple SaaS providers and IaaS providers.***

The problem is the selection of composition plans based on nonfunctional (Quality-of-Service, or QoS) attributes such as response time, throughput and cost [8]. The requirements of end users are modeled as a set of time series. Also, cloud service providers market their services (SaaS or IaaS) using a set of time series. Each time series represents the values of a corresponding QoS attribute over a long period. Hence, cloud service composition problem becomes a ***similarity search problem*** whose query is a set of desired time series.

Traditional techniques seldomly handle complex time series queries which require the correlation between the time series to be used during similarity matching. The correlation, however, are prevalent during service composition where each QoS attribute is correlated with several other QoS attributes.

### Datasets and Inputs

synthetic data sets:

### Solution Statement

### Benchmark Model

### Evaluation Metrics

### Project Design