1.125
Team Members: Robin Jha
Team Name: PI (Predictive Insights)

Project Description:
Operating a data center (DC) is a challenging task as customers expect it to be continuously available. Applications running on the cluster are dependent on it running smoothly; any outages will cause service disruption (ex- Netflix getting impacted due to Amazon Web Services (AWS) outage) eventually leading to revenue loss. Since customer demands are dynamic and depends on multiple factors, operations and maintenance become unpredictable and tedious.

In this project I will be exploring the possibility of streamlining infrastructure management process by monitoring behavior pattern of important metrics (disk utilization, disk I/O, network traffic, etc.) and running it through an analytics engine to predict future trends. Using this data, organizations can plan maintenance and disaster management better.

Approach:
For this project I will be using collected data from monitoring agents running on all nodes of a cluster. This agent collects multiple data points for network, CPU, memory, disk, etc. every given polling interval (usually between 2-5 sec). I will then use Spark to explore this data and find trends which will be written to a ElasticSearch cluster for storage and visualization.

Technologies:
Currently, this is what I think I will need, list might change depending on my needs/progress:
1. Apache Spark (ML)
2. Mesos (For Orchestration)
3. Docker
4. ElasticSearch (Datastore)
5. Apache Kafka (Data spout)
6. Kibana (Visualization)