Google Cluster: Machine & Job Level Analysis

April 9, 2021

School of Information Studies

Research Symposium

Presenter: Prathamesh Pradip Datar

Faculty Supervisor: MD Tariqul Islam “Pavel”

Introduction

- Industries and workloads are shifting towards cloud
- Modern cluster management systems are evolving to cope with managing diverse cloud applications on heterogeneous clusters
- Vast amounts of data available for analysis

Objectives

- Conducting exploratory analysis to reveal underlying trends required to build an efficient cloud scheduler
- Taking a top-down approach to decode complex decisions taken by cloud schedulers
- Focusing on machine and collection level aspects to uncover valuable insights that will ensure high service reliability and availability

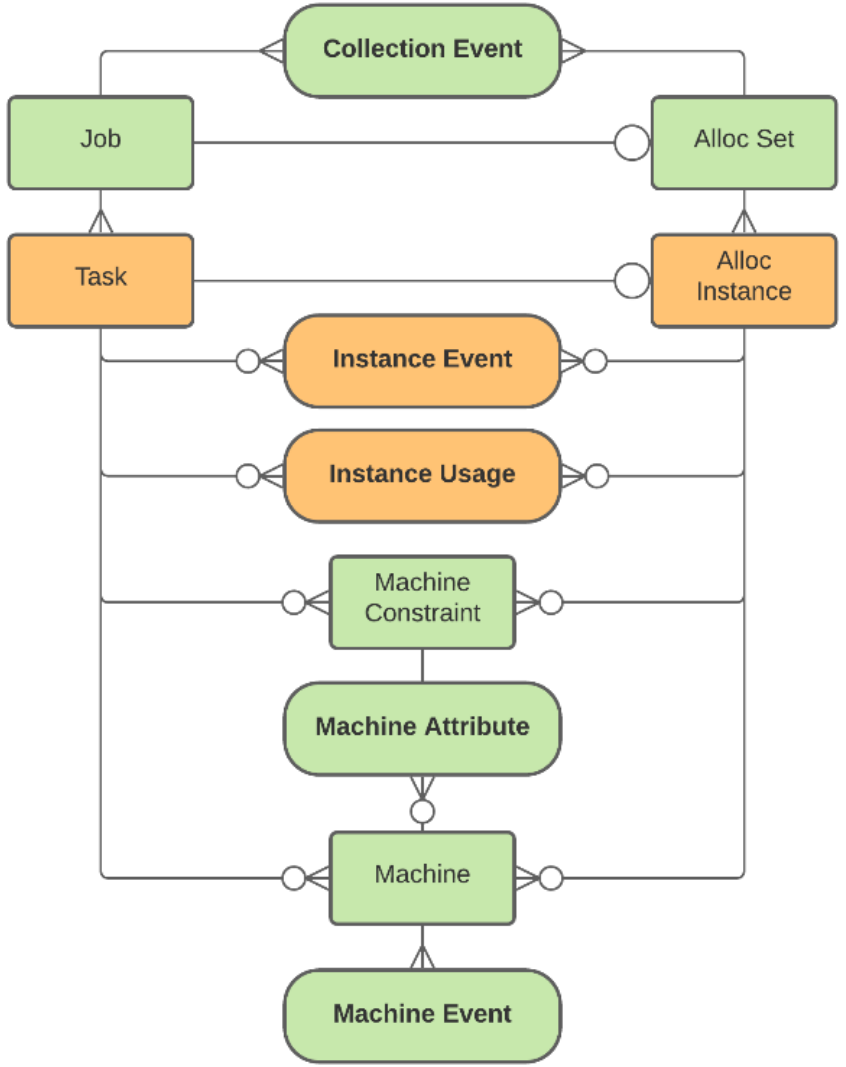
Cloud Components

- **Clusters** – A set of machines, packed into physical enclosures, and connected by a high bandwidth cluster network
- **Cell** – A set of machines, typically all in a single cluster, that share a common cluster-management system that allocates work to machines
- **Jobs** – A collection of one or more tasks
- **Tasks** – Linux program, consisting of multiple processes, to be run on a single cluster

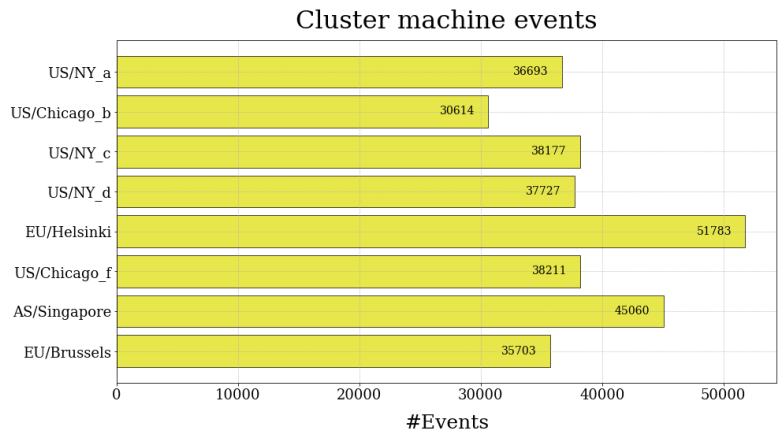
Data Components

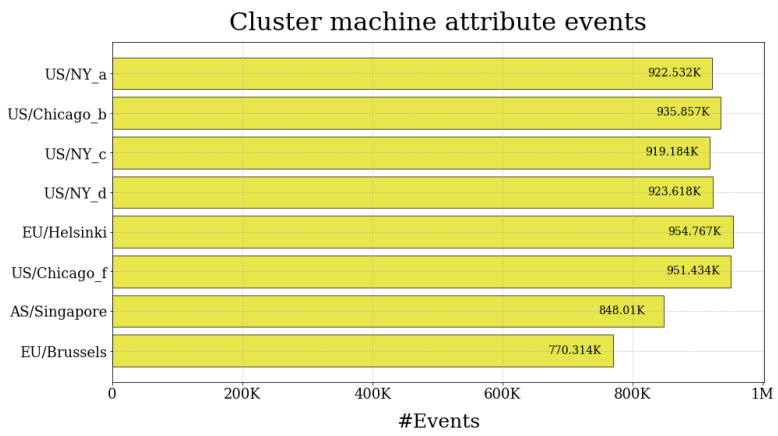
- **Machine Events** – Records ADD, REMOVE, UPDATE Events
- **Machine Attribute Events** – Records Machine Property changes
- **Collection Events** – Records Job lifecycle Events
- **Instance Events** – Records Task events
- **Instance Usage** – Resource Usage stats

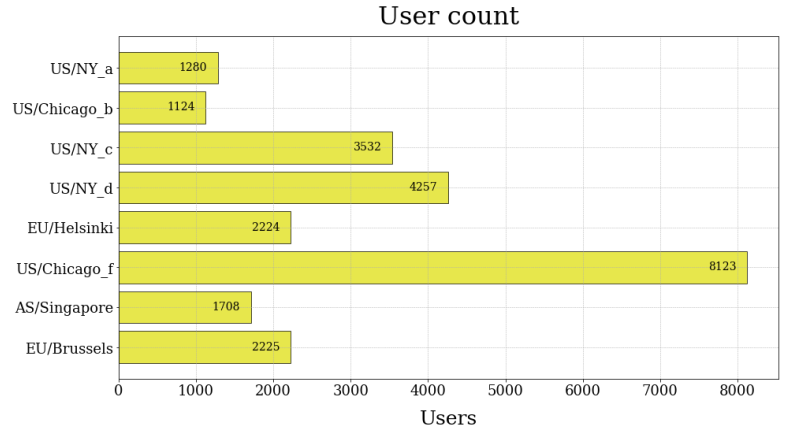
Entity-Relationship Model

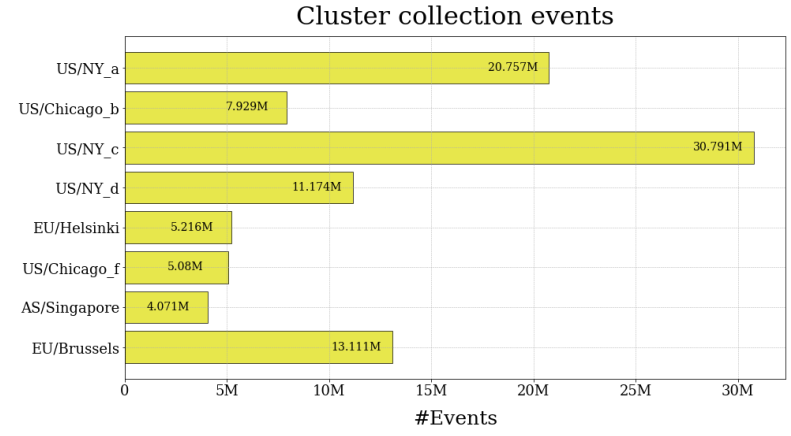


Data at Scale









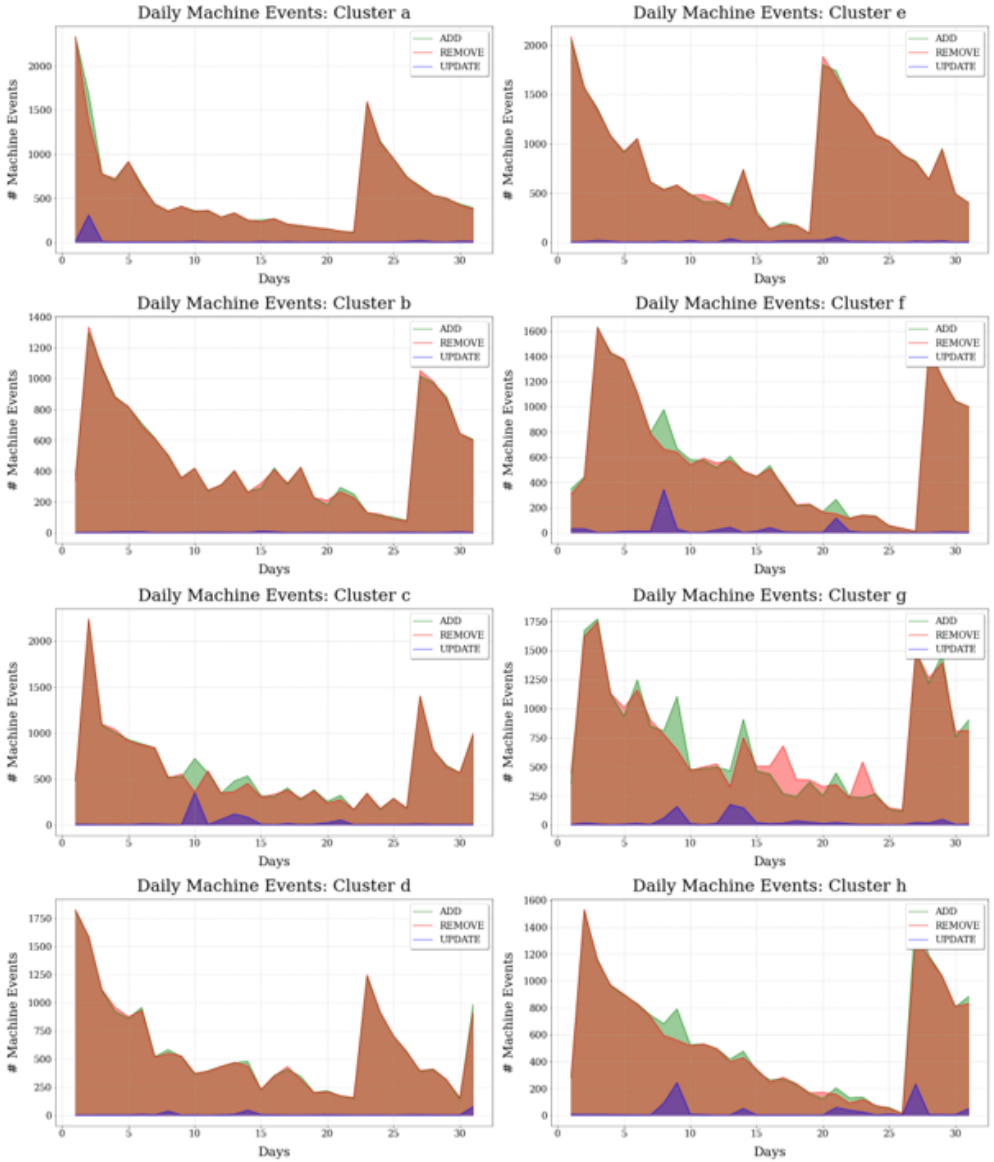
Machine Events

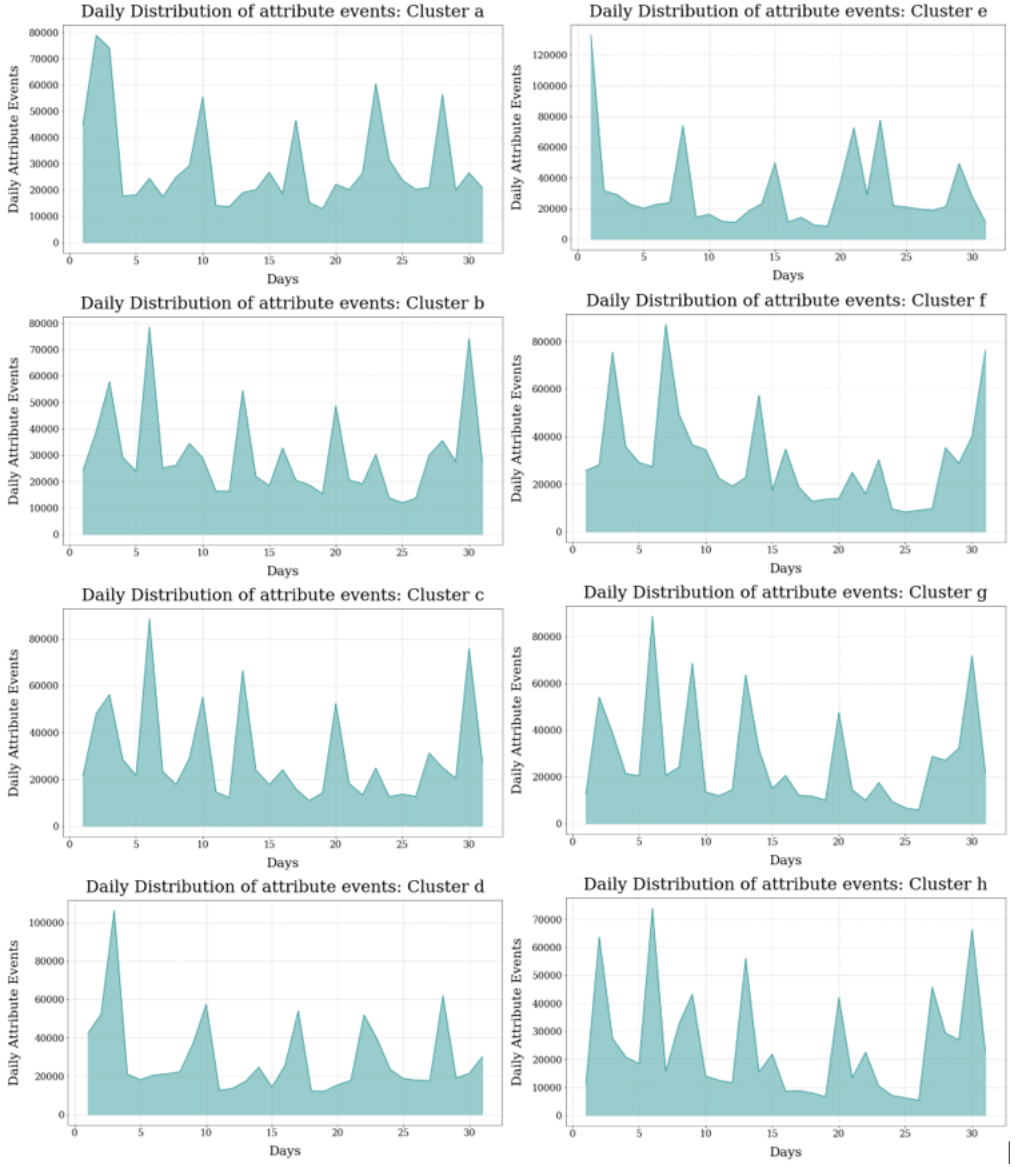
There are three types of machine events:

- **ADD**: Machine becoming available to the cluster.
- **REMOVE**: Machine removed from the cluster. Machine removal could indicate system failure or scheduled maintenance.
- **UPDATE**: The machine has its resources changed when it was available to the cluster.

A typical machine lifecycle is as follows:

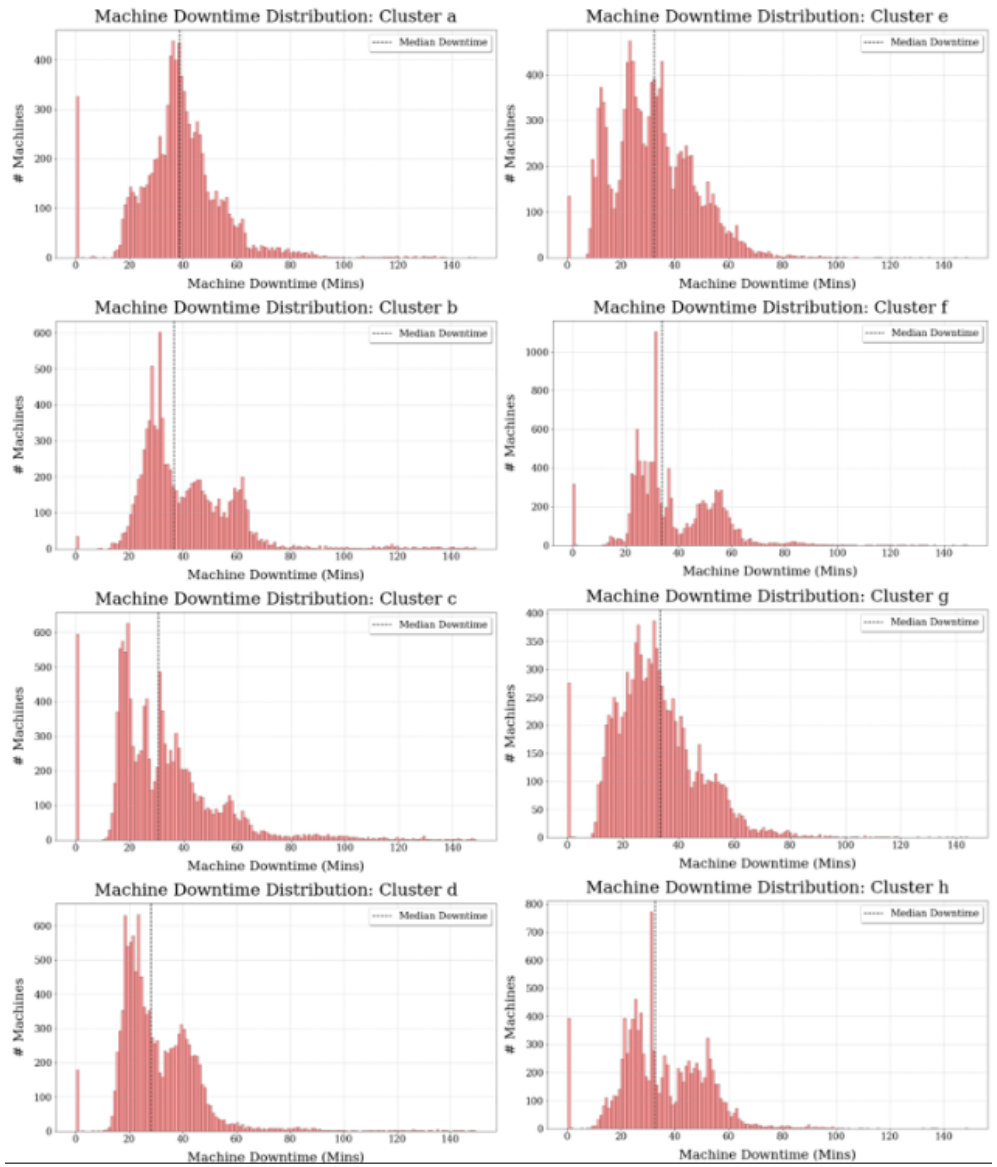
Machines become available to the scheduler to host tasks and alloc instances. While being operational, they may experience changes in resource availability. They might get removed and become unavailable due to a system failure or scheduled maintenance. They become available to the programmed after resolving the problem.

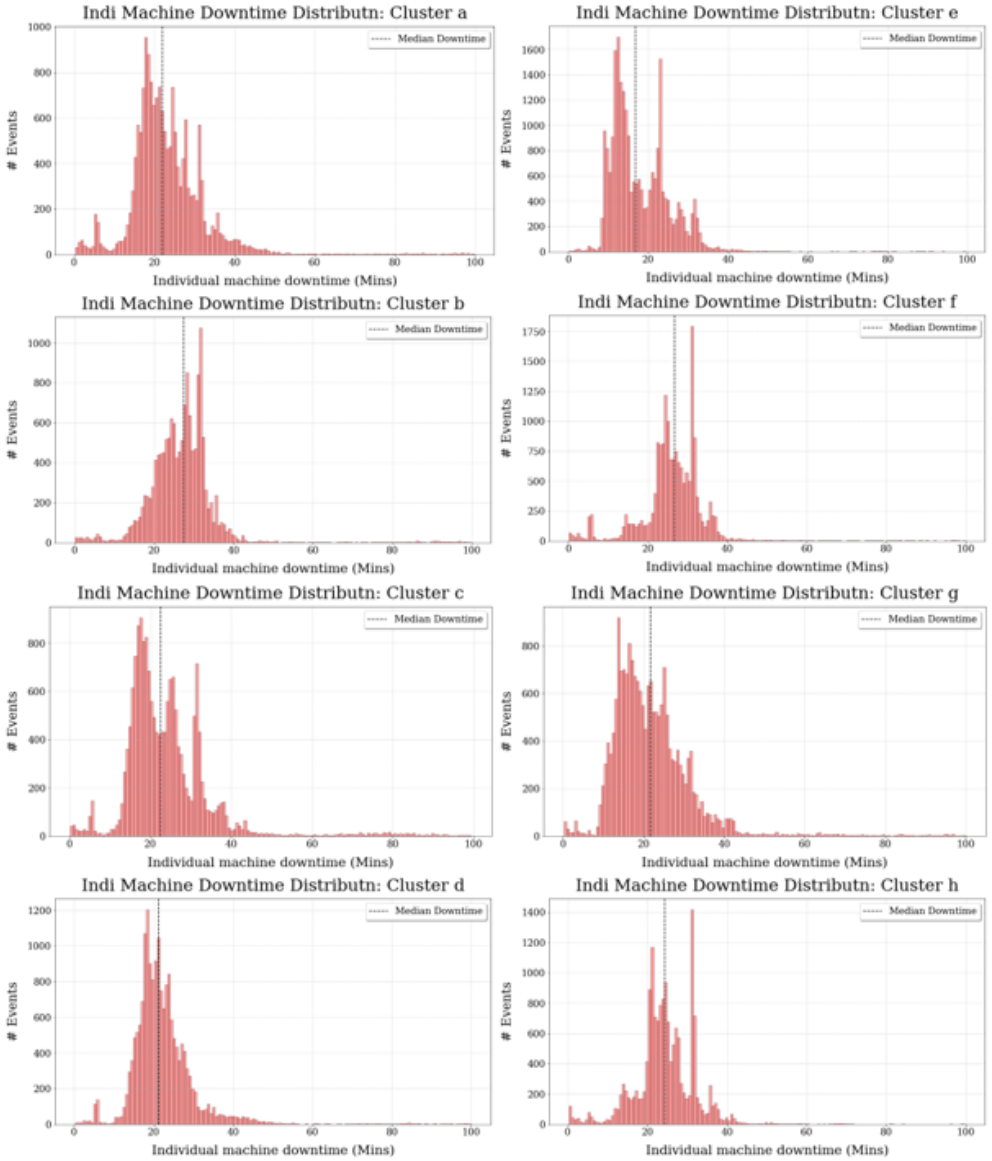

Daily Machine Events: Cluster a
Daily Machine Events: Cluster e
Daily Machine Events: Cluster b
Daily Machine Events: Cluster f
Daily Machine Events: Cluster c
Daily Machine Events: Cluster g
Daily Machine Events: Cluster d
Daily Machine Events: Cluster h
ADD
REMOVE
UPDATE
Machine Events
Days


Daily Distribution of attribute events: Cluster a
Daily Distribution of attribute events: Cluster e
Daily Distribution of attribute events: Cluster b
Daily Distribution of attribute events: Cluster f
Daily Distribution of attribute events: Cluster c
Daily Distribution of attribute events: Cluster g
Daily Distribution of attribute events: Cluster d
Daily Distribution of attribute events: Cluster h
Daily Attribute Events
Days

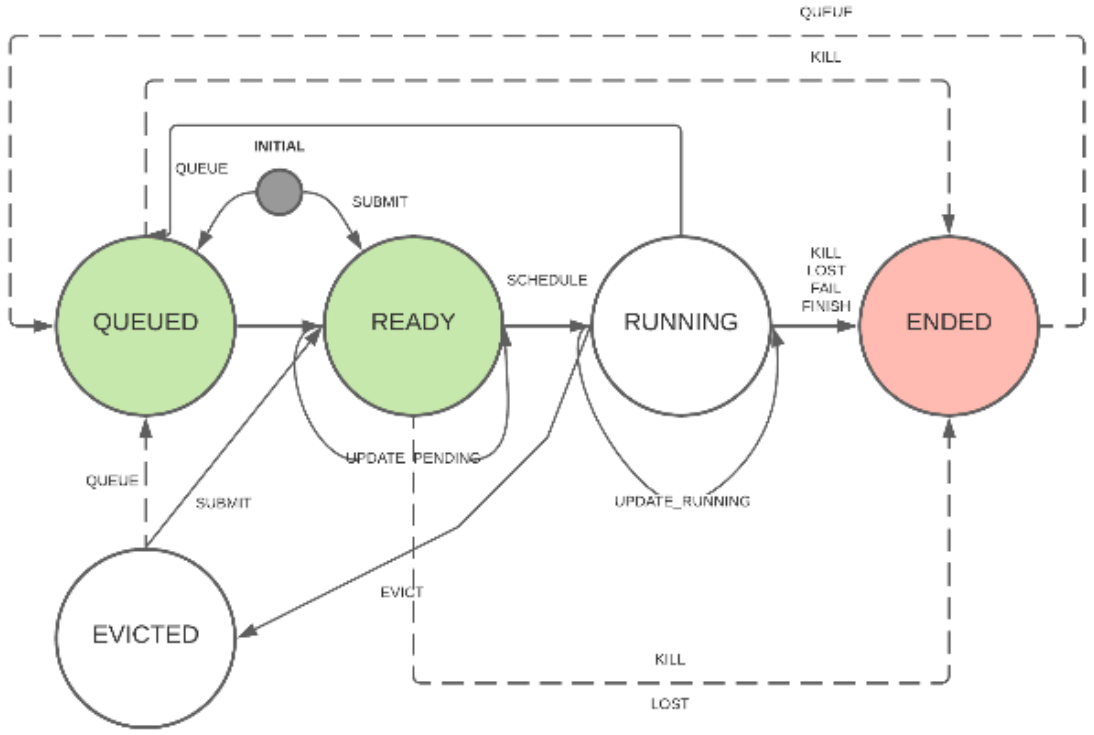
Machine Downtime Analysis

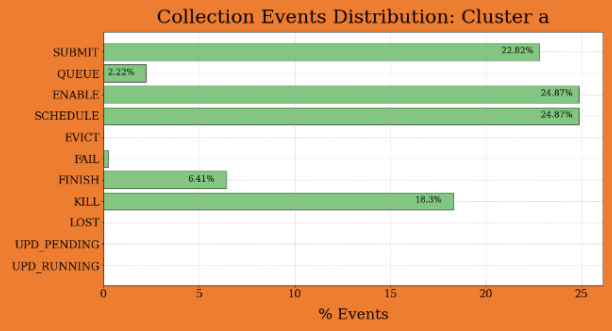
- Machine downtime is the time a machine was unavailable for task scheduling.
- Machine downtime is the time difference between REMOVE and subsequent ADD events.
- For the cases where we do not have a subsequent ADD event, we have considered the end of the trace time.

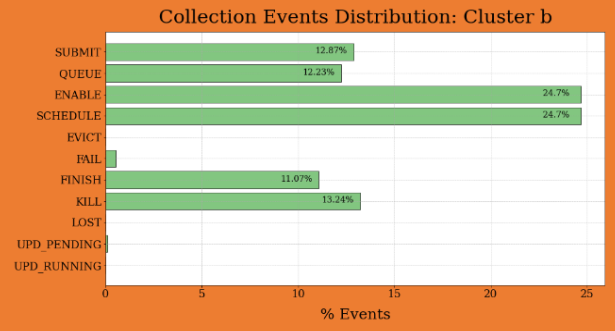

Machine Downtime Distribution: Cluster a
Machine Downtime Distribution: Cluster e
Machine Downtime Distribution: Cluster b
Machine Downtime Distribution: Cluster f
Machine Downtime Distribution: Cluster c
Machine Downtime Distribution: Cluster g
Machine Downtime Distribution: Cluster d
Machine Downtime Distribution: Cluster h
Median Downtime
Machines
Machine Downtime (Mins)

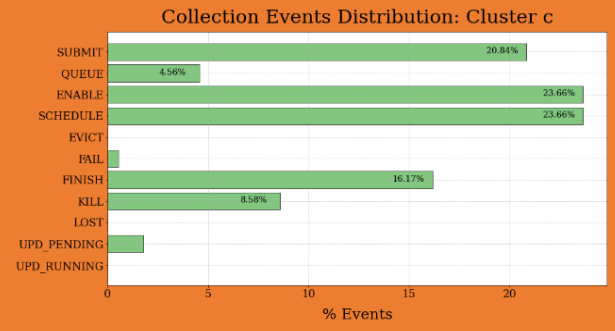

Indi Machine Downtime Distributn: Cluster a
Indi Machine Downtime Distributn: Cluster e
Indi Machine Downtime Distributn: Cluster b
Indi Machine Downtime Distributn: Cluster f
Indi Machine Downtime Distributn: Cluster c
Indi Machine Downtime Distributn: Cluster g
Indi Machine Downtime Distributn: Cluster d
Indi Machine Downtime Distributn: Cluster h
Median Downtime
Events
Individual machine downtime (Mins)

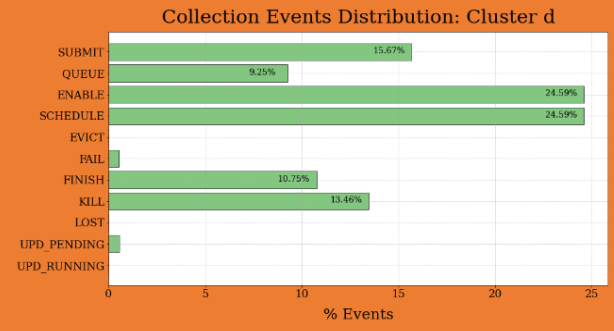
Collection Events

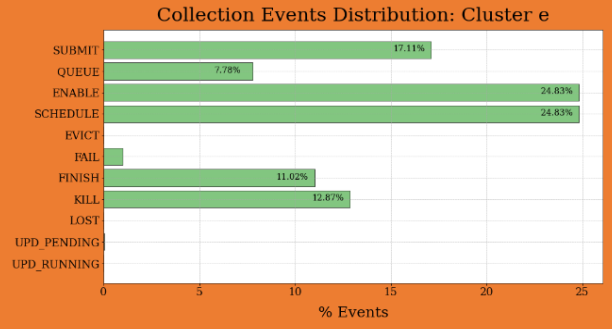


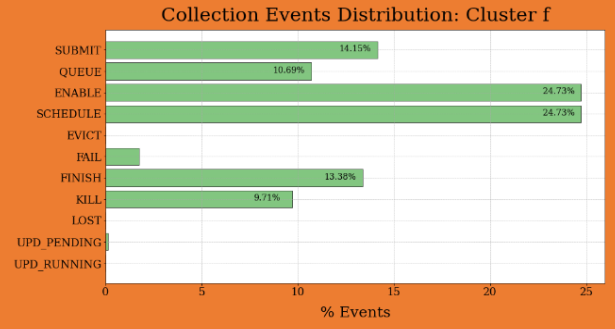


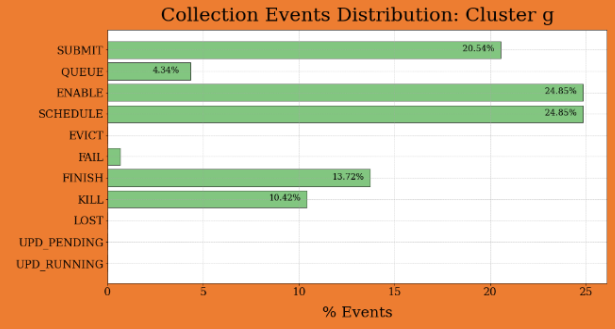


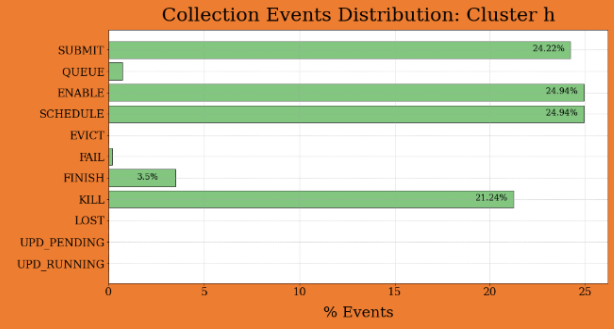




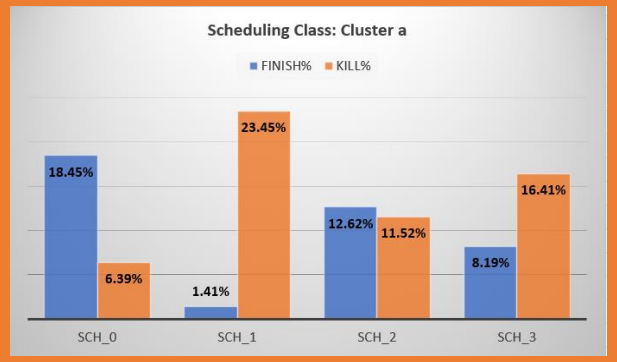


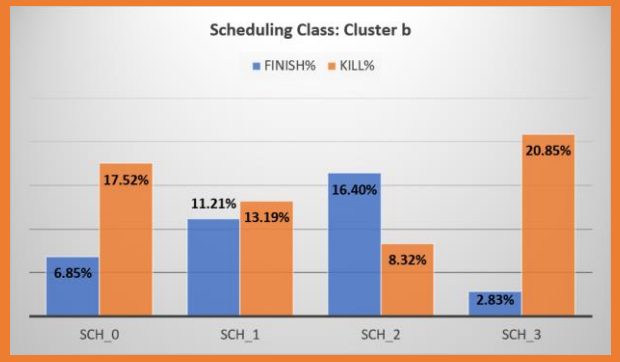


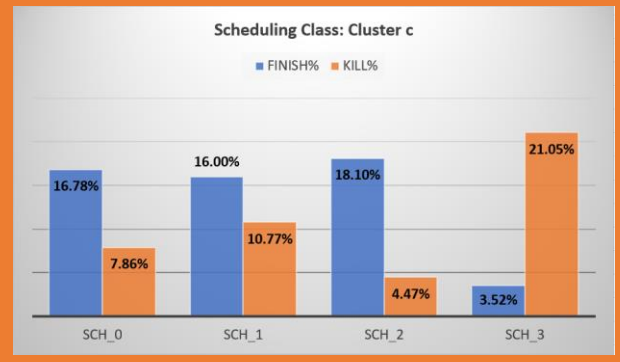


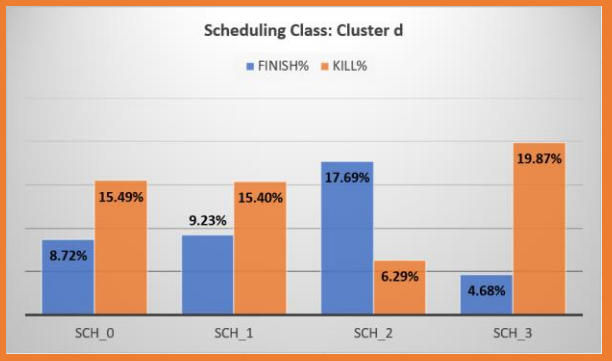


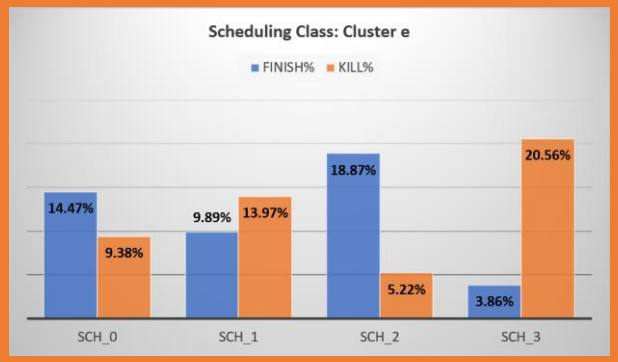
Collection Events Distribution

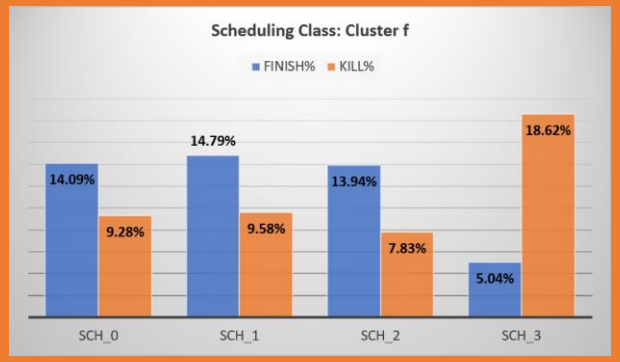


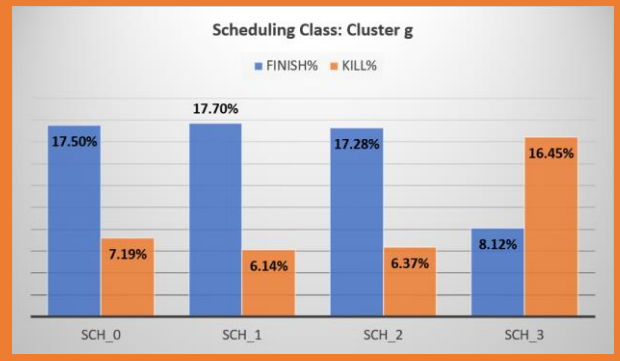


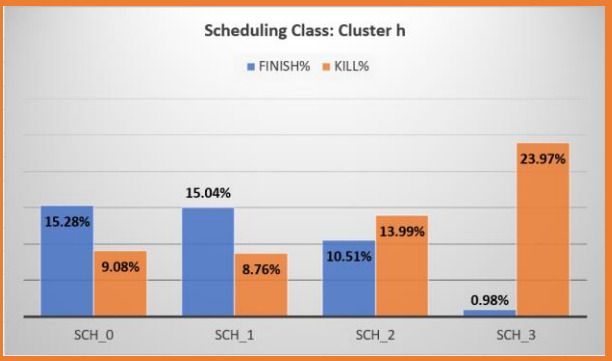




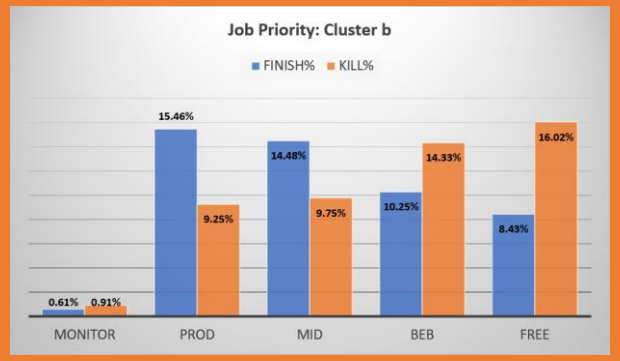


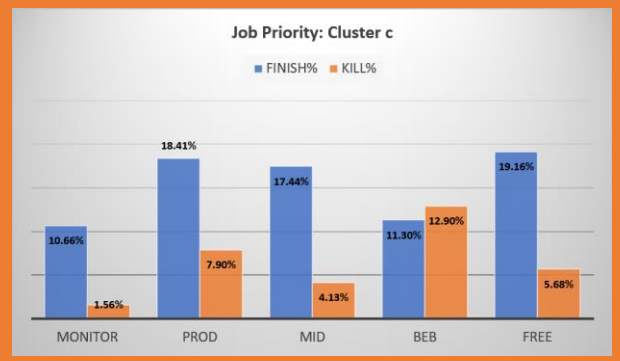


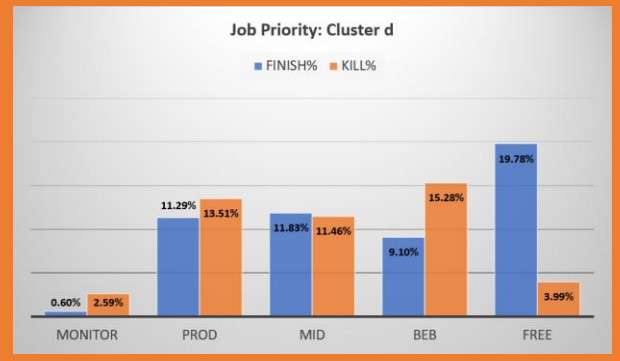


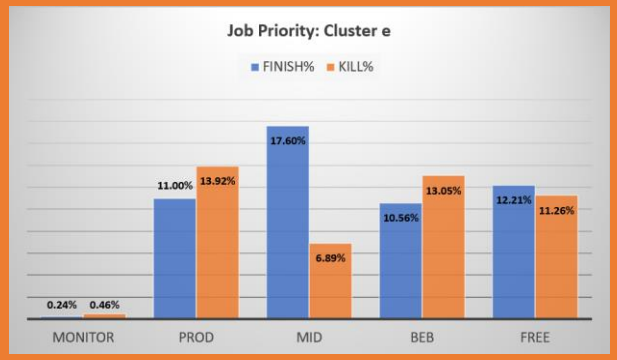


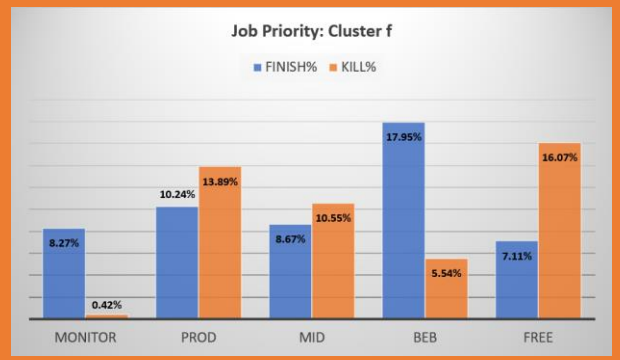
Scheduling Class FINISH and KILL Rates

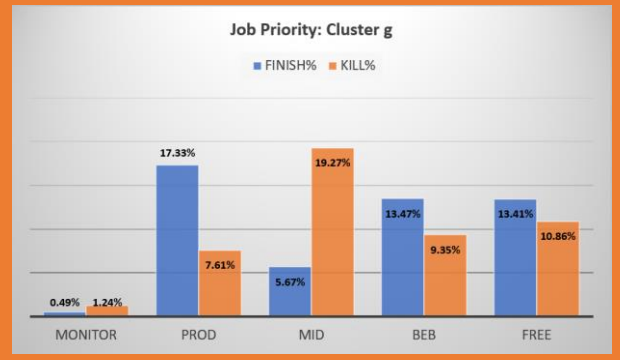


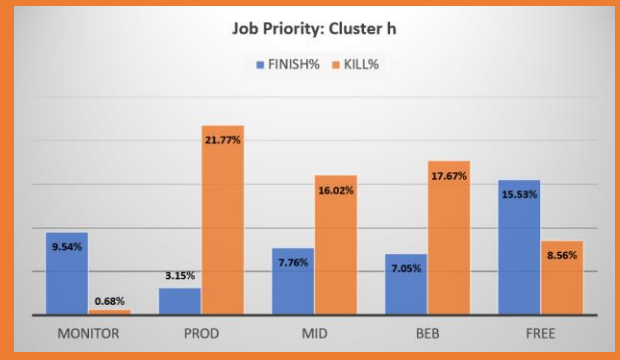






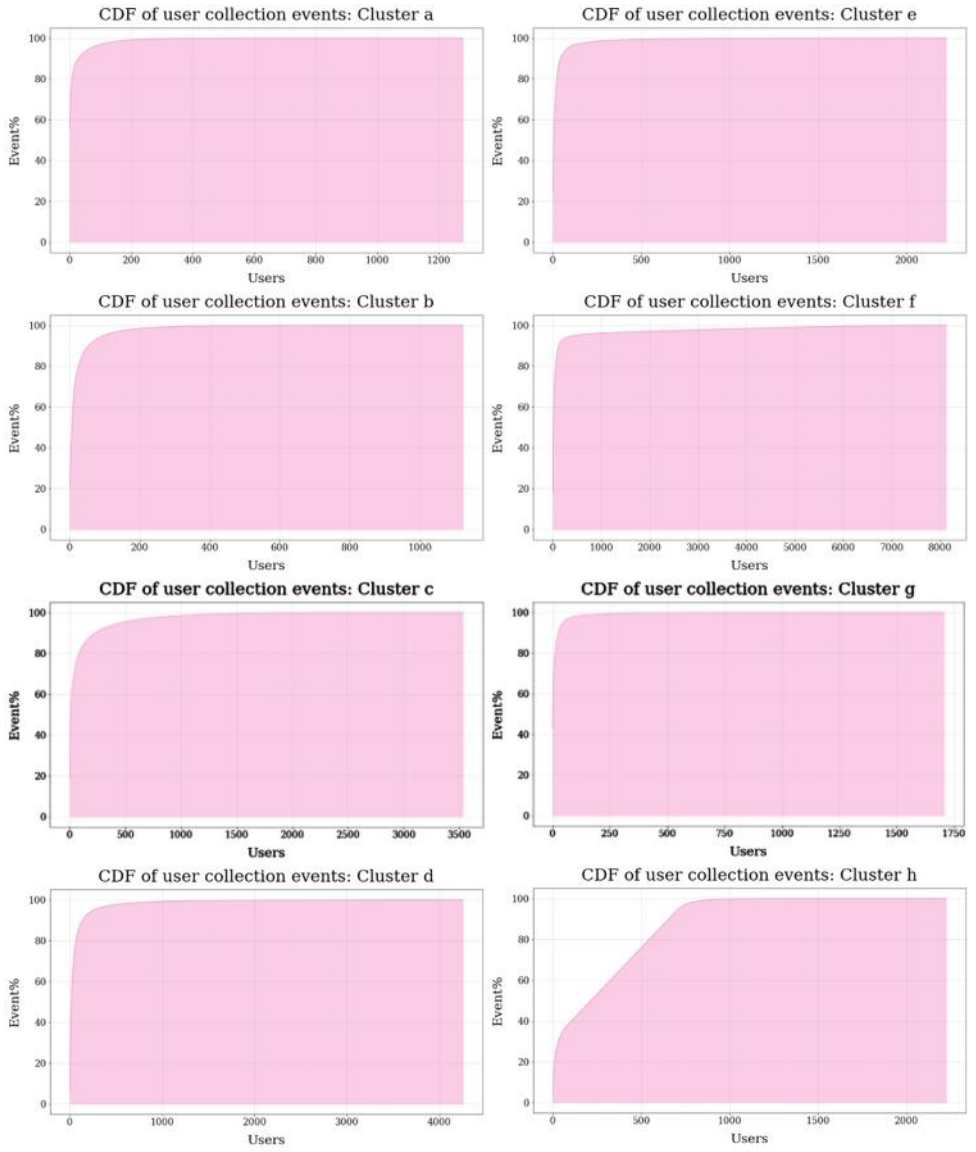






Job Priority FINISH and KILL Rates

User Event Rates



Journey Ahead

- We wish to dig deeper in understanding the remarkable shift of free tier jobs to best effort batch jobs that will help cater to the users' requirement needs
- We plan to employ a parallel research to combine queuing delay for collections and scheduling delays for tasks to better visualize task life cycle inside best effort batch tier jobs
- We plan to investigate quantitative contributions of vertical scaling to reduce slack resources.
- We would like to draw strong conclusions in how heterogeneity of machines is affecting machine downtimes

Questions