BigPanDA Technical Interchange Meeting
September 23-24, 2018

Modeling utilization of allocation time

Status report

A. Poyda, M. Titov, S. Jha

Problem statement

The load on a resource is defined as a number of busy nodes at a certain time and it is determined by the number and parameters of jobs:

- the number of required nodes per job;
- required execution time called wall time per job;
- jobs generation rate.

The concept of an execution strategy is defined as the set of values of denoted parameters that uniquely define the group of jobs to be executed.

The goal

- find execution strategies that maximize the probability of utilizing a certain allocated resources;
- find the execution strategy that will optimize the utilization of a given number of core-hours on a resource.

Methods

Quantitative model

A quantitative model to estimate the probability of a given number of core-hours being utilized

- trained by the previous processes of utilization of the resource;
- represented by the equation which calculates the probability that utilization $\mathbf{U}$ during the time interval $\mathbf{T_0}$ will reach or exceed the predefined value $\mathbf{U_0}$.

$$P\,(U > U_0) = \sum_{n=100}^{\infty} \left[\int_{U_0}^{\infty} f\left(x, n\mu_U, n\sigma_U^2\right) dx \left(\int_{-\infty}^{T_0} f\,(x, n\mu, n\sigma^2)\, dx - \int_{-\infty}^{T_0} f\,(x, (n+1)\mu, (n+1)\sigma^2)\, dx \right) \right]$$

Probability that utilization will reach the defined utilization value during the defined time period (cumulative distribution function).

- $\mathbf{f(x, \mu, \sigma^2)}$ is a function of probability density of the normal distribution $\mathbf{N(\mu, \sigma^2)}$;
- μ and σ^2 are expected value and variance of a random variable describing duration of waiting time in the queue for jobs correspondingly;
- μ_U and σ_U^2 - the same as previous, but for a random variable describing utilization of one job.

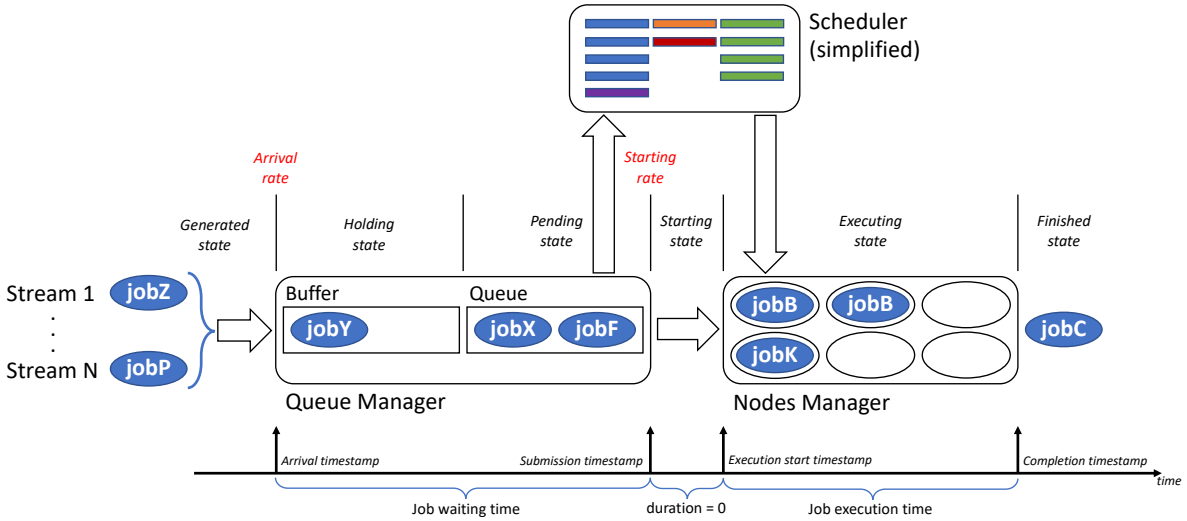
Simulator

Simulates the load on a supercomputer and produces job traces for a given workload

- used for the quantitative model validation and adjustment;
- based on queueing theory (M/M/total_num_nodes);
- provides a job state model with the following states: generated, holding, pending, starting, executing, finished;
- supports job priorities and backfill mode.

Scheduler (backfill mode)

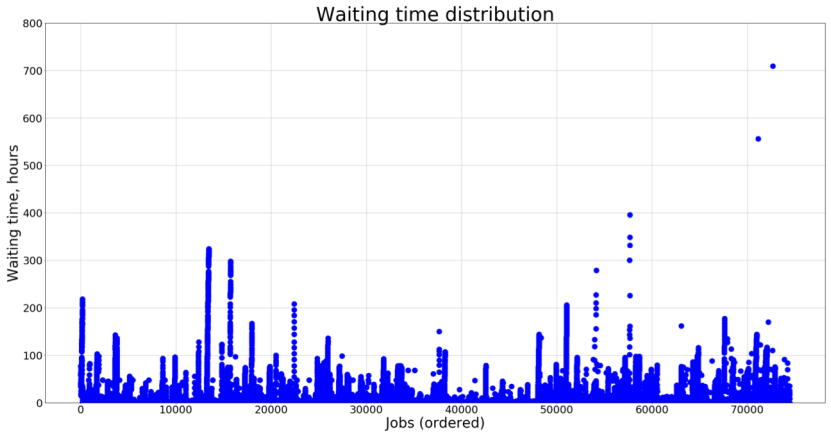
- gets info about job sizes and assign corresponding nodes
- gives a schedule when each job starts to be executed



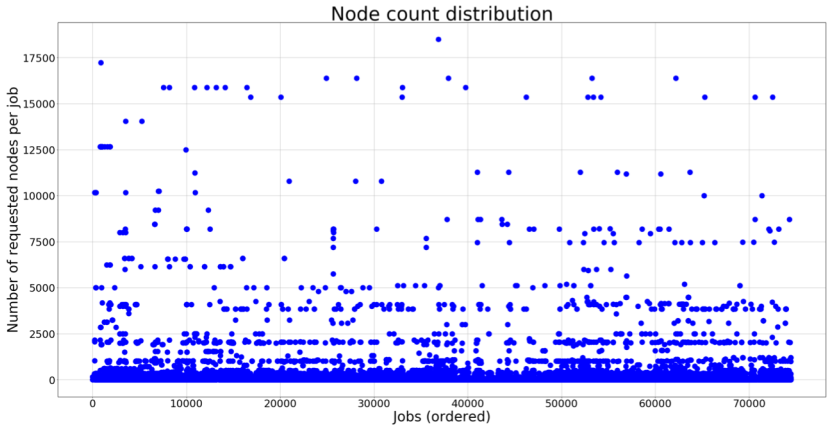
Job state transitions

Waiting time distributions

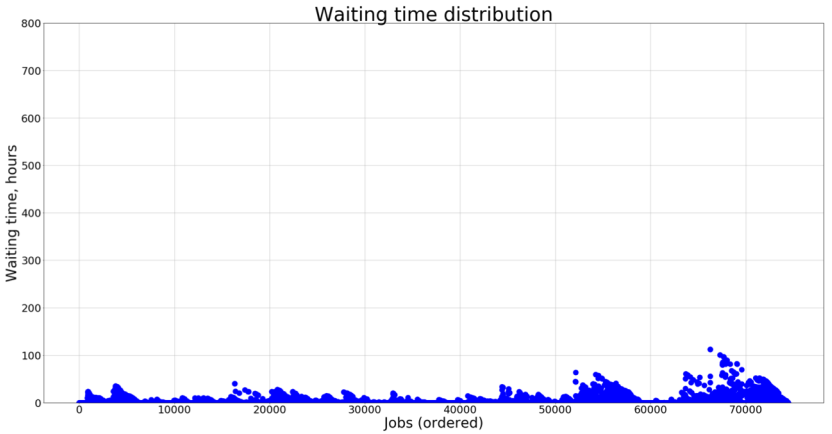
Titan log data



waitingTime (hours) - mean=5.54, std=20.70



Simulator using Titan log data

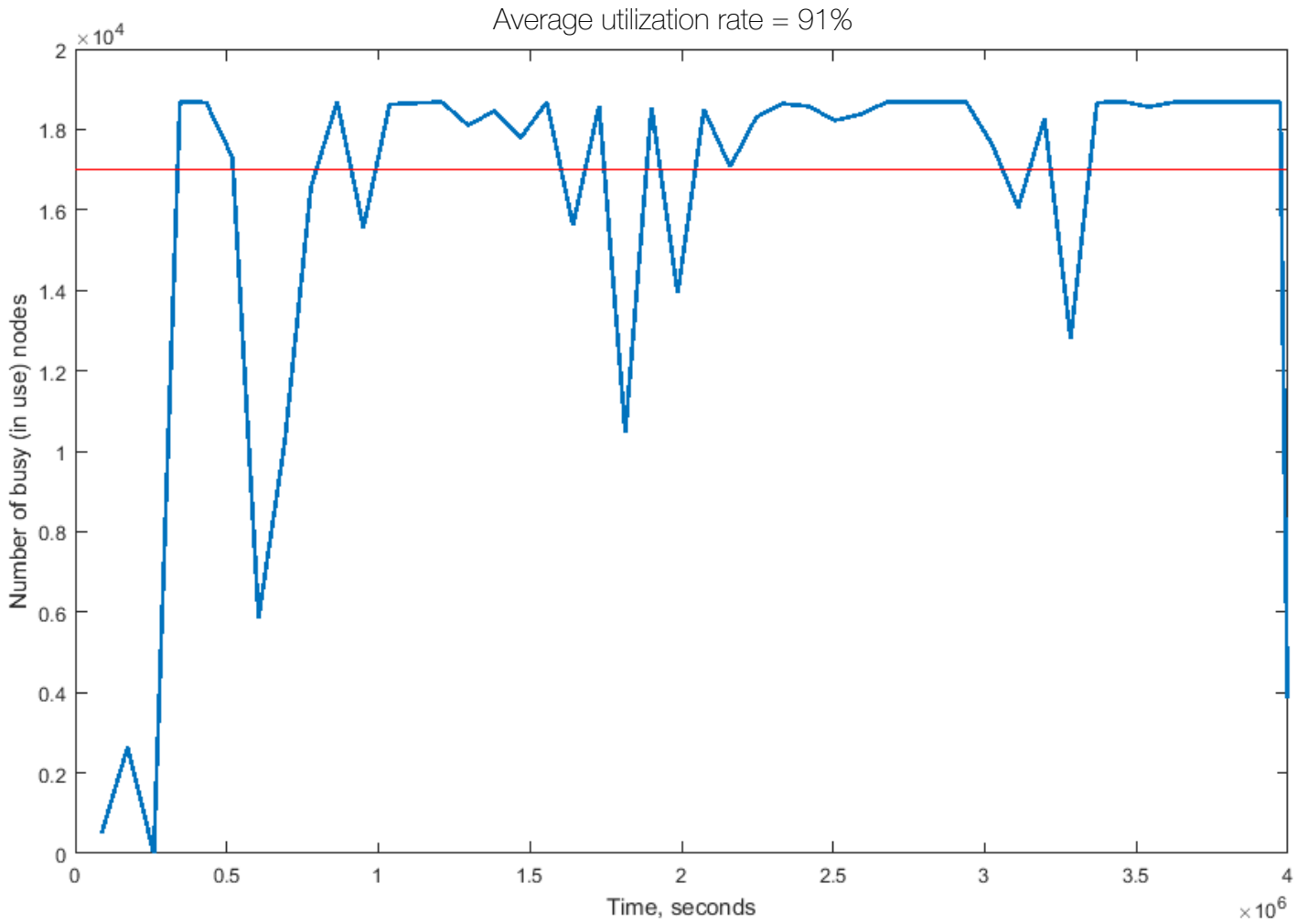


waitingTime (hours) - mean=2.06, std=5.68

Simulation parameters:

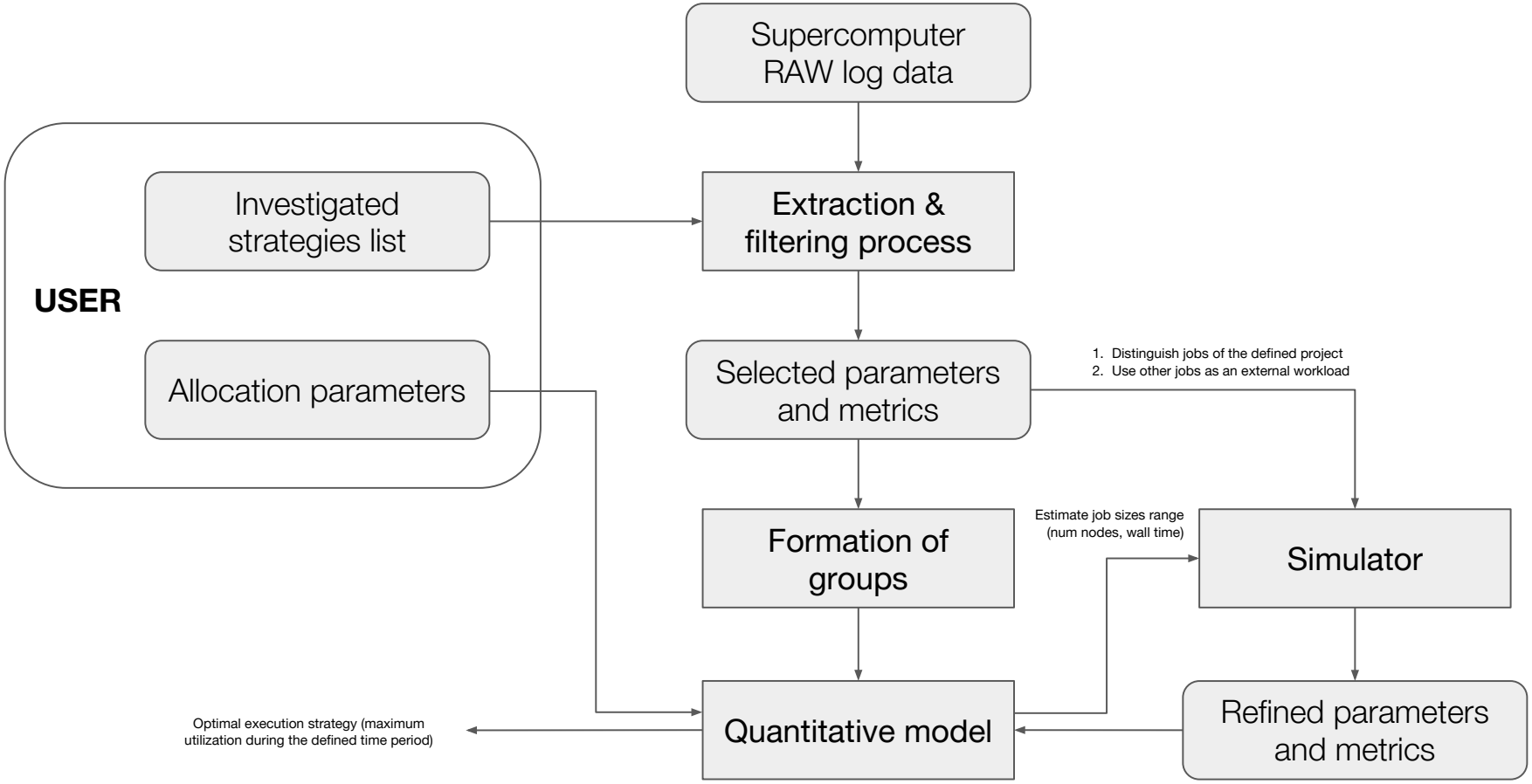
- Job characteristics from log
  - arrival timestamp
  - number of nodes per job
  - real execution time
- Queue characteristics
  - priority discipline
  - initial priorities (for “big” jobs)
  - no limitation by stream

Simulator load using Titan log data



45 days

Analysis workflow



Experiments

Quantitative model testing with synthetic data (I)

Common parameters for the quantitative model and for the simulator

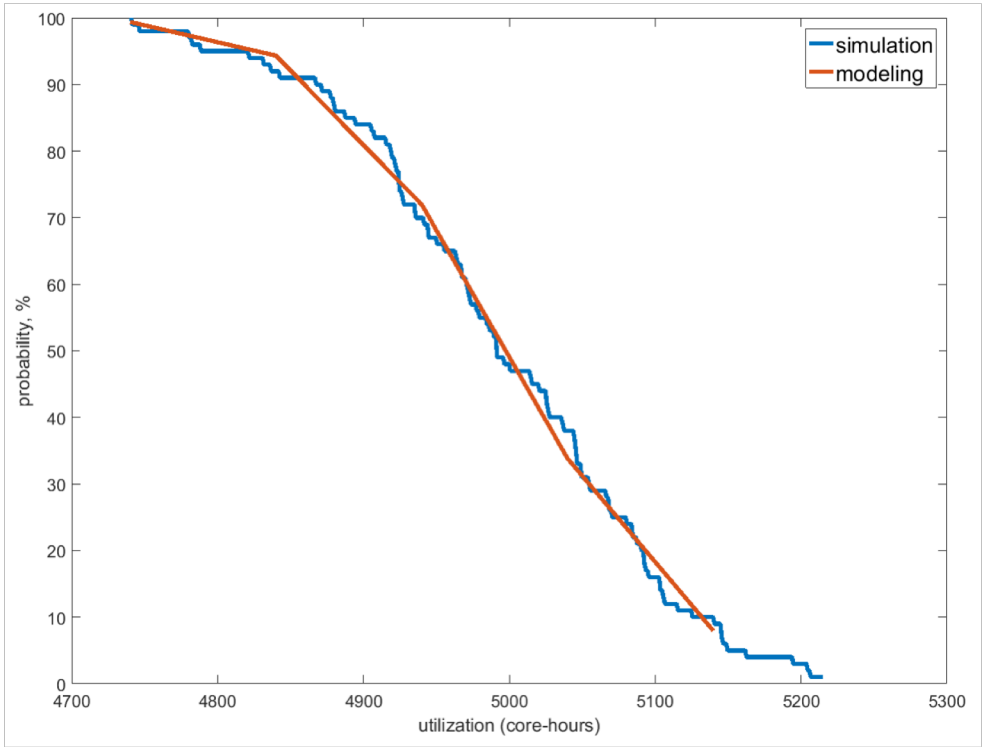
- number of nodes / cores per job = 1
- job waiting time and execution time characteristics (same values for both parameters) - expected value, variance = 1, 1
- the total processing time = 5000 (time units / hours)

Specific parameters for a simulation process

- job waiting time is defined according to the Poisson distribution
- job execution time is defined according to the Normal distribution
- job launching scheme: one stream and there is always one job in the queue
- the total number of simulation runs = 100

Quantitative model testing with synthetic data (II)

Figure shows the plot with two lines that represent the probability that a given utilization will be achieved in a given time interval. The blue line corresponds to the results obtained on the simulator, while the red line corresponds to calculations with the quantitative model.



Probability (axis Y) that utilization will reach the corresponding utilization value (axis X) during the time of 5000 hours.

Log data analysis overview

Log data characteristics

- contains information about job processing
  - job arrival timestamp (to the queue)
  - execution start timestamp
  - completion timestamp
  - the number of required nodes (1 node = 16 cores at Titan)
  - requested walltime

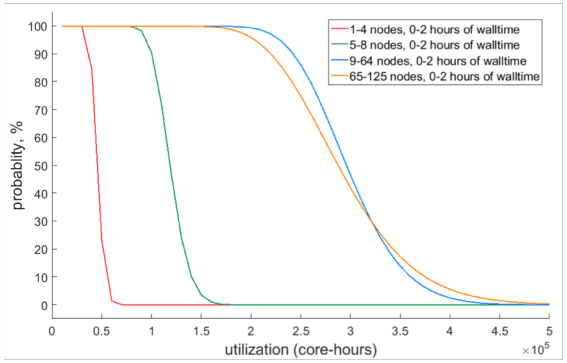
Analysis actions

- all jobs are divided into categories according to the number of required nodes and the volume of walltime requested (every category corresponds to a particular Titan's bin, where bin is a group of jobs that are treated equally)
- for each category the following values are calculated: the expected value and variance of the random variables describing waiting time in the queue and the utilization achieved by a single job
- obtained values were used as input data in equation for the quantitative model to calculate the probability that jobs of a given category will be able to utilize provided allocation in 3 months
- job launching scheme: one stream and there is always one job in the queue

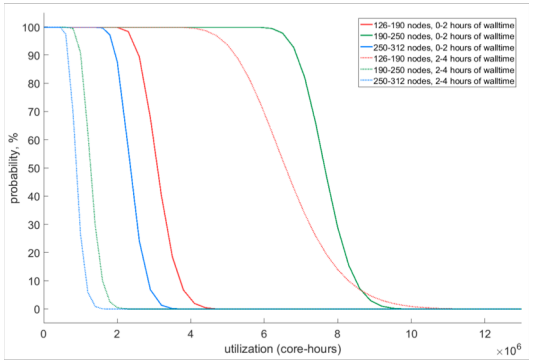
Titan log data analysis

Probability distribution of utilization of the resource during 3 months

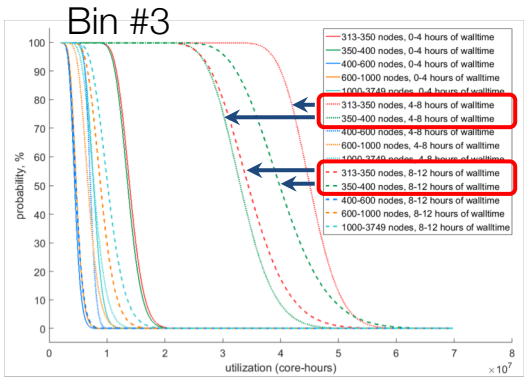
Bin #5



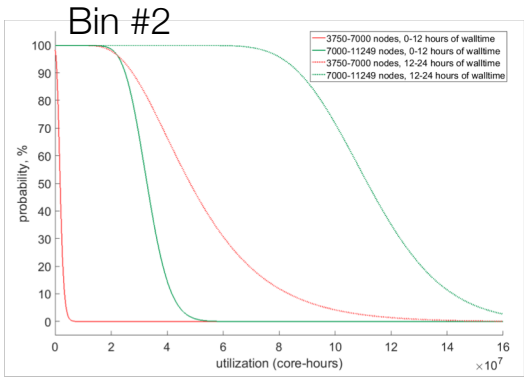
Bin #4



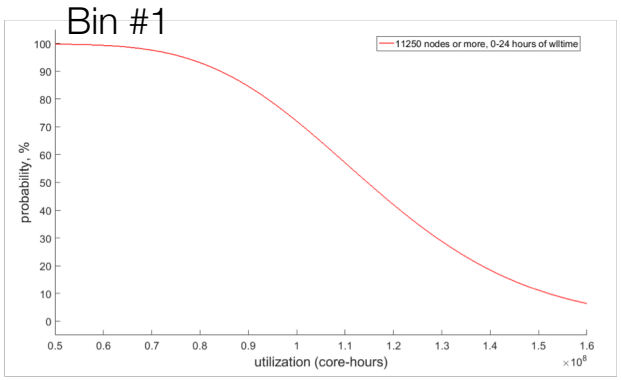
Bin #3



Bin #2



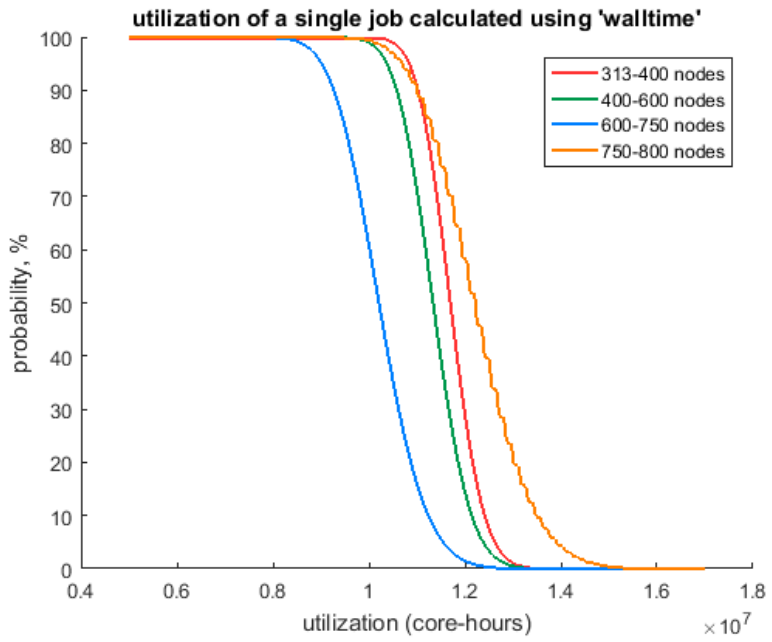
Bin #1

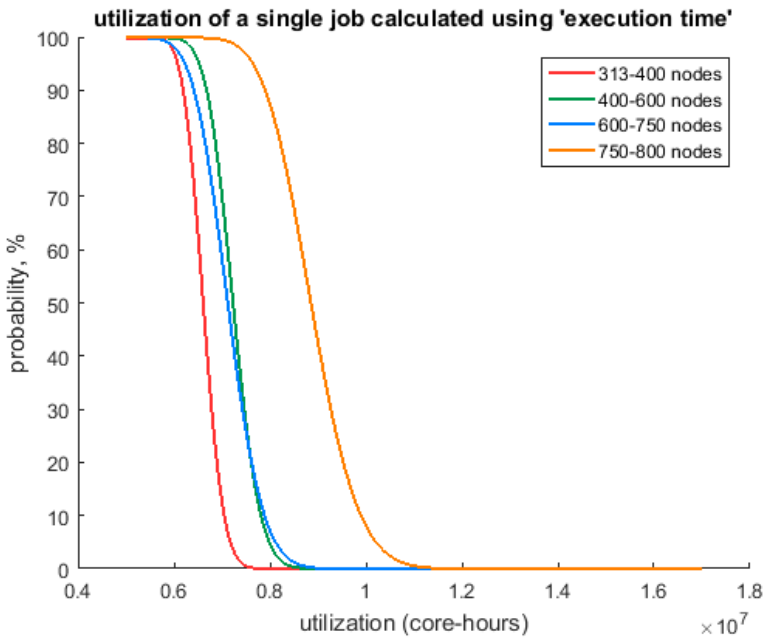


data collected for 6 months (from aug'17 to jan'18)

HEP110 (ALCC) log data analysis

Probability distribution of utilization of the resource during 3 months





Plans / next steps

- Analysis of the extended log data (for 1 year)
  - New logs were obtained from Sarp O.
- Update the new module in the simulator
  - Module “schedule” should be optimized
- Conduct real tests with HEP113 (ALCC) project
  - Use discovered execution strategies