

Predicting job idle time before execution using machine learning

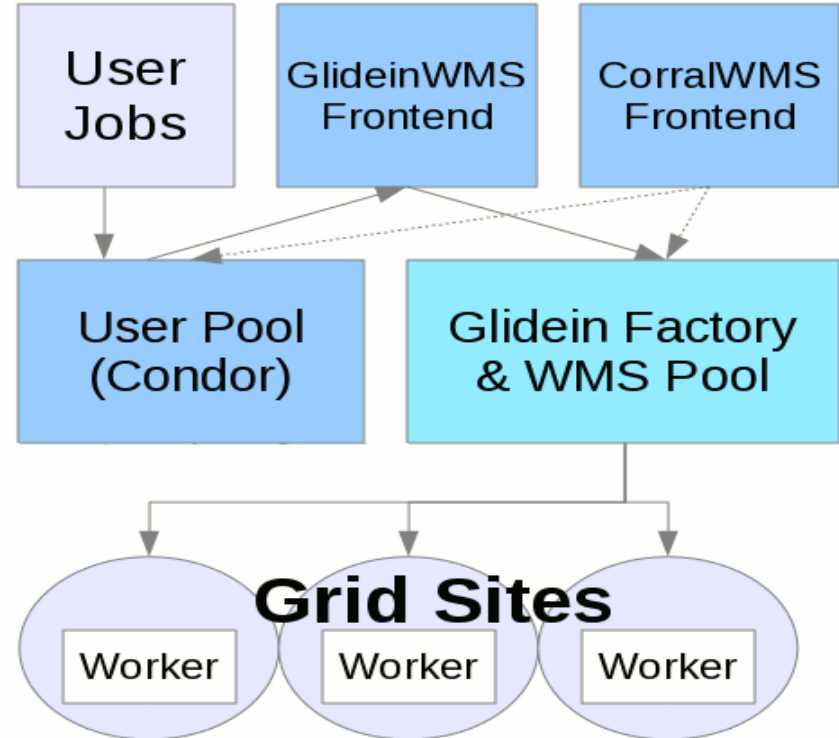
Mentee: Francisco Laris
Mentor: Bruno Coimbra

Abstract

Abstract: When jobs are submitted through HTCondor they are first submitted and added to a queue before beginning execution. This idle time can vary wildly from job to job resulting in some jobs start running later than expected. Machine learning coupled with data from Landscape on the jobs status and submission time, should give us means to create a model for predicting which jobs take longer to get out of the queue and possibly why.

Overview

1. Introduction
2. HTCondor
3. glideinWMS
4. Machine Learning
5. Limits



Work to do

- Job data is stored mainly in Landscape. The data must be extracted and formatted in such a way for it to be usable by our model.
- The data must be cleaned, pre processed and important variables extracted.
- Once this is done the training for our model can begin. The model will have to be adjusted iteratively for a better performance.

Method

- The current plan is to have 2 models one mainly for prediction and one for analysis.
- Due to the nature of neural networks it will only be used as a predictive tool, since it acts as a black box.
- The other model will be using regression to find what variables affect the idle time of jobs most.