Job Accounting Data Analysis for the Quanah Cluster

CS5352 Course Project, Spring 2021

[This project accepts one student only.]

**Contact**

Jie Li ([jie.li@ttu.edu](mailto:jie.li@ttu.edu))

**Description**

Job accounting information of a High-Performance Computing platform records very detailed metrics of each submitted job, including job submission time, start time, end time, exit status, cpu time usage and integral memory usage etc. These metrics are invaluable for understanding user behaviors and evaluating the HPC platform performance. In this project, we analyze two months of the job accounting data collected from the previous Quanah cluster, which contains 139,261 jobs from 233 unique users. You are expected to analyze the resource usage (e.g., number of cores, cpu time, memory usage etc.) of each job and of each user. The following deliverables will be used for the evaluation:

1. Understand each metric in the job accounting information.
2. Analyze the job accounting data and plot diagrams to visualize the aforementioned resource usage in job-level and user-level.
3. Discuss findings from the diagrams (bonus).

**Requirements**

* Experience of Linux
* Experience of Python programming
* Experience of matplotlib library (or other plotting libraries)

**References**:

1. Univa Grid Engine accounting file format <https://docs.hpc.qmul.ac.uk/using/man/man5/accounting.5.txt>
2. A Workload Analysis of NSF's Innovative HPC Resources Using XDMoD <https://arxiv.org/pdf/1801.04306>
3. Comprehensive job level resource usage measurement and analysis for XSEDE HPC systems <https://dl-acm-org.lib-e2.lib.ttu.edu/doi/pdf/10.1145/2484762.2484781>
4. Experience with Using the Parallel Workloads Archive <https://doi-org.lib-e2.lib.ttu.edu/10.1016/j.jpdc.2014.06.013>