View Reviews

Paper ID
204

Paper Title
Wasted Cycles and Waiting Games Analysis of HPC Resource Usage Using Production Cluster Data in FRESCO

Track Name
Applied Data Science Track

Reviewer #1

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.
In this paper, the authors perform an analysis on predicting wait time for HPC jobs which are in a queue. The authors divide the wait times in different bins such as 10 mins, 10min-1hr, 1hr-6hr and more than 6hr in the training data and train model and then try to predict wait time based on a few features of the job, user metadata, and the state of the then system. The authors also perform an important analysis on wastage of resources in terms of CPUs being idle or memory unused during the job execution.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.
1. The analysis was done on a reasonably large dataset which includes the data on the resource utilization details against each of the job ids.

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel).
1. Although the authors mention about different kind of clusters in Section 3.1, while doing the final analysis they only considered the job in Anvil cluster. It would be good to see the analysis for other clusters as well.
2. In the paper some plots were given in Figure 2 but that figure was never referred in the write up. However, explanations of the figures were given to a certain extent in the write up.
3. Going back to analysis of RQ1, it would have been great if there was an after-analysis based on the predicted wait time to check how much efficiently jobs could have been

scheduled.
4. Some more details about the prediction model would be a good addition to the write up.

4. [Final rating] Paper rating
Weak Reject

5. Justification of final rating
The analysis seems to be relevant but it would be good to see the generalization of the results across multiple clusters. The entire write up could be improved a bit for better readability.

Reviewer #2

Questions

1. [Summary] In 3-5 sentences, describe the key ideas, experiments, and their significance.
The paper analyzes production HPC operations using the FRESCO dataset that links 20.9 million job records to fine-grained telemetry across three clusters, and studies two questions: predicting queue wait times and quantifying resource waste. For RQ1, the paper recasts wait time as a four-way classification problem and, on a temporal holdout from Anvil, achieve a macro F1 of 0.92 with XGBoost using only submission-time features, with SHAP showing that pending jobs, free cores, and queue choice matter most. For RQ2, it defines CPU, memory, and a composite waste score, and in a 10 percent random job sample report an average composite waste near 48.5 percent, with 69.3 percent of jobs exceeding 50 percent waste and waste rising with job size. The results are positioned to inform user-facing wait-time guidance and administrative policy such as waste-aware scheduling.

2. [Strengths] What are the strengths of the paper? Clearly explain why these aspects of the paper are valuable.
1. The work uses a public, multi-institutional dataset that links job accounting and time-series telemetry, enabling analyses that many prior traces could not support. The study spans three clusters with heterogeneous architectures and includes a clear data schema that ties requested to used resources.
2. Framing wait time as four actionable categories produces strong performance on a temporal holdout from a production system (macro F1 0.92). Features are available at submission time, and SHAP analysis identifies load and queue choice as dominant drivers, which is easy for users and admins to reason about.
3. The paper reports high memory underuse, an average composite waste near 48 to 50 percent, and a monotonic rise in waste with job size, plus exit-code and duration effects.
4. The methodology, feature set, and sampling are described clearly, and an artifacts link is provided. The paper translates findings into concrete suggestions for users and

administrators, including waste-based policies and targeted education.
5. The paper connects to strands on queue prediction and utilization telemetry, and explains how FRESCO closes a long-standing gap by linking accounting and node-level metrics at job granularity.

3. [Weaknesses] What are the weaknesses of the paper? Clearly explain why these aspects of the paper are weak. Please make the comments very concrete based on facts (e.g. list relevant citations if you feel the ideas are not novel).

1. Although the dataset is very large, the waste study uses a 10 percent simple random sample from a frame built by scanning about every twentieth hourly shard, yielding 567 jobs. Given the strong claims about prevalence and costs, a larger sample or stratified sampling by cluster, job size, and duration would increase confidence.
2. Memory waste assumes fixed memory-per-core values per system and gives equal weight to CPU and memory in the composite score. Sensitivity analysis on these choices and a justification for equal weighting would be helpful, especially since memory dominates the waste signal.
3. The prediction study is limited to Anvil from January to June 2023. It would be stronger to report cross-cluster or cross-period validation, especially under policy changes. The class distribution is highly imbalanced (about 88 percent under 10 minutes), and while macro F1 is high, precision for some minority bins is lower, which could affect user trust for those cases.

4. [Final rating] Paper rating

Weak Accept

5. Justification of final rating

This paper makes a strong contribution by analyzing HPC operations through the unique FRESCO dataset, which links scheduler accounting with fine-grained telemetry at job level across three large clusters. This linkage enables two important analyses: wait-time prediction and resource waste quantification. The results are both practically relevant and well supported. The wait-time model achieves high predictive accuracy using only submission-time features, which is actionable for users and administrators. The waste study shows that a majority of jobs substantially underutilize allocated resources, with memory being the dominant factor, and highlights patterns such as higher waste in large jobs. These findings have immediate operational implications, including user guidance, administrative policy adjustments, and design of waste-aware scheduling.

The work is clearly written, well grounded in prior literature, and careful in methodology, including use of SHAP for interpretability and statistical tests for waste patterns. Importantly, it provides a public artifact, increasing reproducibility and impact.

The main limitations are the relatively small sample size for waste quantification, the assumptions in defining composite waste, and the restriction of prediction experiments to a

single cluster and time window. These choices limit generalizability, but the authors acknowledge them and suggest reasonable directions for future work.

Overall, despite some scope constraints, the paper delivers valuable insights into HPC inefficiencies with both scientific rigor and practical relevance. It meets the CODS criteria for novelty, clarity, and impact, and is recommended for acceptance.