Towards Application Agnostic HPC Profiling

Hari Teja Jajula[1], Dhruva Kulkarni[2], Brian Austin[2], P.V. Bangalore[1]

[1] The University of Alabama , [2] Berkeley Lab (NERSC)

ABSTRACT

Modern HPC systems generate large amounts of GPU and network telemetry, typically used for system health monitoring. At NERSC, we are developing a Performance API that generates a Job Report Card from this telemetry, providing an overview of performance characteristics. Using DCGM counters, we report GPU memory, compute, and power usage, and present preliminary investigations of job-level network activity. This application-agnostic approach, which does not require traditional profiling tools, helps identify resource utilization imbalances, detect anomalies such as memory leaks, and assess overall performance for users without additional effort.

BACKGROUND

- NERSC collects and stores operational telemetry at **petabyte scale** and at **sampling rates (~0.1–1 Hz**), which requires the use of several advanced tools (e.g., LDMS [4]) and mechanisms, labeled OMNI.
- This data is useful from several perspectives, including assessing system performance, monitoring workflow and application performance, debugging application performance regressions, and enabling insights across the system-application boundary.
- NERSC is developing an API to provide curated access to performance and power data at the workflow/job level, intended to be released to NERSC users in the near future.

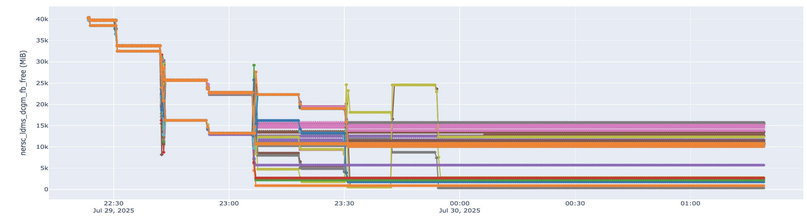
JOB REPORT CARD

From raw LDMS/DCGM/NIC counters, we generate a **job-level report card** that summarizes key performance characteristics and can plot time-series metric data.

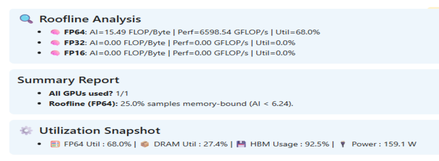
Implemented features:

- **Roofline Heatmap** – shows arithmetic intensity vs achieved performance; validated with MixBench to confirm compute vs memory-bound regimes.
- **GPU Utilization** – reports average usage, highlights imbalance across GPUs.
- **Activity Breakdown** – percentage split across GPU compute, network communication, and I/O.
- Helps detect imbalance across computation vs communication vs I/O

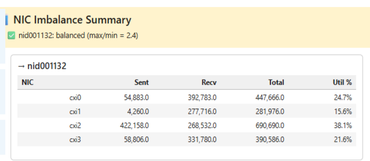
EXAMPLE PLOTS & REPORT CARD



GPU free memory over time



GPU utilization summary

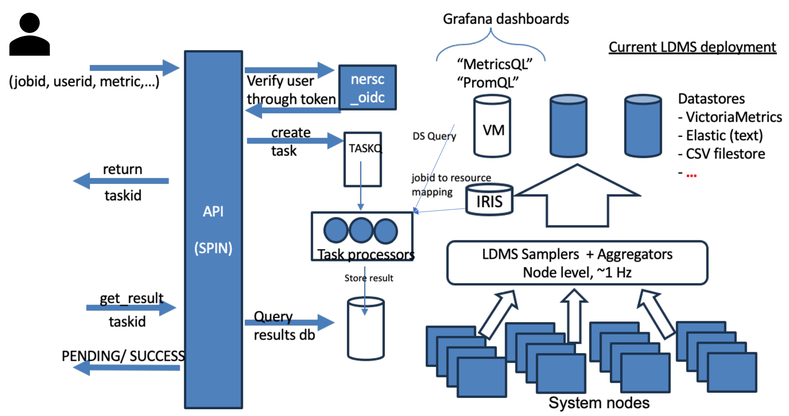


NIC utilization summary

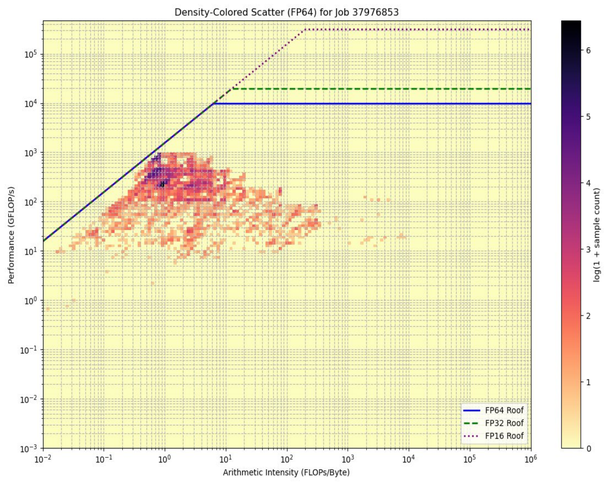
NIC COUNTER DATA

- we identified the mapping of traffic classes to communication types: TC 0–1 correspond to MPI traffic, TC 2–3 to I/O traffic, and TC 6 to TCP traffic.,
- Our NIC analysis further highlighted the need to collect traffic-class packet size distribution counters

ARCHITECTURE



ROOFLINE HEATMAP



- $AI = \frac{FP64\ util}{dram\ UTIL} \times \frac{Peak\ Perf}{Peak\ Bandwidth}$
- $Performance = FP64\ util \times Peak\ Perf$

FUTURE WORK

- Overlap metrics (CPU–GPU, comm–compute),
- load balancing factors, and anomaly detection (e.g., memory leaks).

CONCLUSION

- System telemetry provides immediate performance insights without requiring traditional profilers.
- They help identify underutilization, imbalance, or congestion, and can evolve into automated alerts (e.g., idle GPUs, hangs, memory leaks).
- Ultimately, correlating job-level insights with system-level telemetry enables holistic, system-wide performance understanding.

REFERENCES

1) Austin, B., Kulkarni, D., Cook, B., Williams, S., & Wright, N. J. (2024). ***System-wide roofline profiling**: A case study on NERSC's Perlmutter supercomputer.* In *Proceedings of the 15th International Workshop on Performance Modeling, Benchmarking and Simulation of High-Performance Computing Systems (PMBS '24)*, SC Workshops (pp. 1398-1404). IEEE.
2) **NVIDIA Corporation.** (2024). *NVIDIA Data Center GPU Manager (DCGM) User Guide*. Retrieved from https://docs.nvidia.com/datacenter/dcgm/latest/user-guide
3) **Hewlett Packard Enterprise Development LP.** (2024). *HPE Cassini Performance Counters User Guide*. Retrieved from https://cpe.ext.hpe.com/docs/latest/getting_started/HPE-Cassini-Performance-Counters.html
4) **Sandia National Laboratories.** (n.d.). *Lightweight Distributed Metric Service (LDMS) Documentation*. Retrieved from https://www.sandia.gov/sandia-computing/high-performance-computing/lightweight-distributed-metric-service-ldms/
5) Elias Konstantinidis, Yiannis Cotronis, "A quantitative roofline model for GPU kernel performance estimation using micro-benchmarks and hardware metric profiling", Journal of Parallel and Distributed Computing, Volume 107, September 2017, Pages 3756,ISSN0743-7315,https://doi.org/10.1016/j.jpdc.2017.04.002.
URL: http://www.sciencedirect.com/science/article/pii/S0743731517301247

ACKNOWLEDGEMENTS

I thank Dr. Purushotham Bangalore (University of Alabama) for academic guidance, and Dr. Sam Wellborn for discussions on iperf scripts for validation of NIC counters. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility (project m888-2025).