Discovering Energy Resource Usage Patterns on Scientific Clusters

Matthew Bae[1], Wucherl Yoo[2] (advisor), Alex Sim[2] (advisor), John Wu[2] (advisor)

[1]Harvey Mudd College, [2]Lawrence Berkeley National Laboratory

Background and Motivation

- Motivated by observations that discovering resource usage patterns can be conducted by monitoring performance from scientific clusters
- Energy efficiency is a concern currently addressed by dynamic power management, frequency scaling, etc.
- Simple Linux Utility for Resource Management (SLURM) is a widely used job scheduler on many supercomputers.
- Characteristics of jobs from SLURM logs, especially energy consumption patterns, can be read on Cori, NERSC's Cray XC40 supercomputer.
- Challenges in understanding energy usage patterns:
  - Energy consumption is nonlinearly dependent on multiple variables, requiring nonlinear metrics
  - Large dataset of energy and system resources includes noisy data

Data and Analysis Design

- At time of analysis, Cori had 1630 compute nodes, each with 32 cores.
- Data points from job steps in SLURM logs for May 2016
- 5951 jobs and 57210 job steps in logs
- Logs include elapsed time, page faults, resident set size, average CPU frequency, energy consumption, and more.
- Preprocessing utilizing Python and Apache Spark
- Interactive plot filters based on thresholds for given variables with different colors
- Grouping and annotations to further understand resource usage patterns

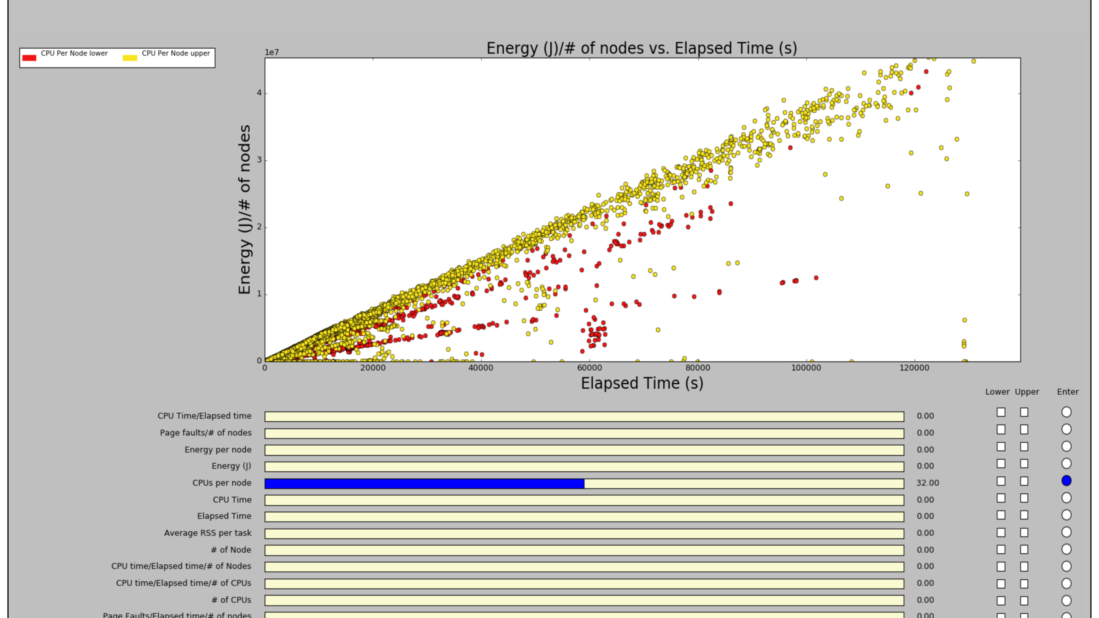

Figure 1: The interactive tool in use. Sliders at the bottom of different variables allow users to partition points with different colors based on a given threshold.

Research Problem

To identify patterns of HPC jobs that consume different amounts of electricity on NERSC supercomputers in the context of system resources.

Results

- Figure 2 shows jobs that use 80 nodes from 1 user.
  - 3 types of energy usage signatures
  - Cluster with 32 cores/node has largest different in CPU time and Elapsed time, implying there is more time spent doing I/O.
- Lines in figure 3 formed by different core/node values on left-hand graph and differences in elapsed time minus CPU Time in right-hand graph.
- Linear regression in figure 4 approximates growth in power consumption and shows a baseline for power consumption (regression only for x-values less than 32)

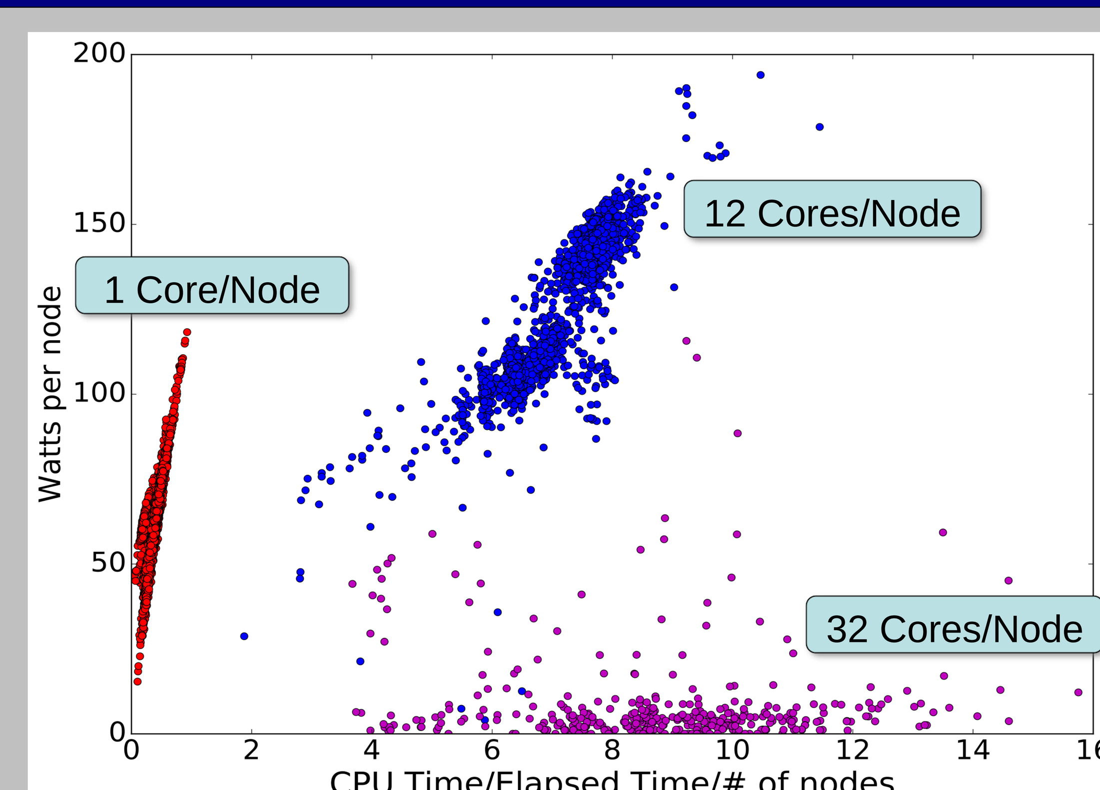

Figure 2: Colorings of groupings based on cores/# of nodes for values of 1, 12, and 32 from left to right respectively. Memory and page faults are largest in purple cluster.

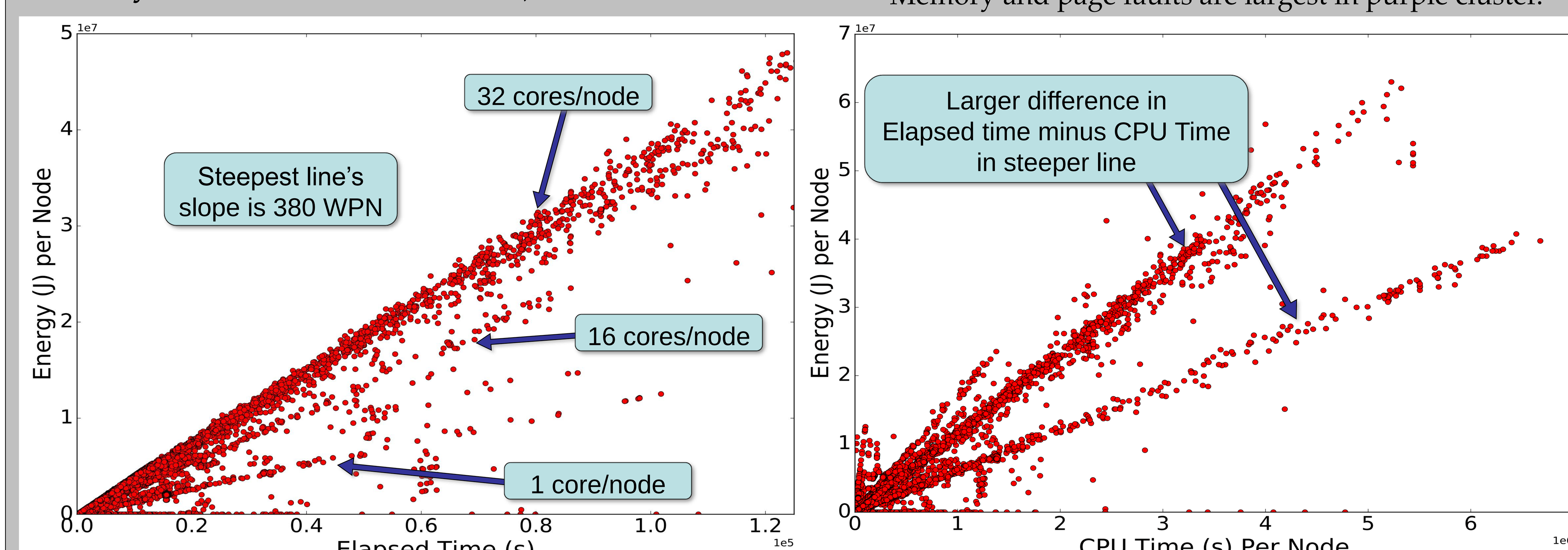

Figure 3: Plots of Energy per node vs. Elapsed time and CPU Time per node.

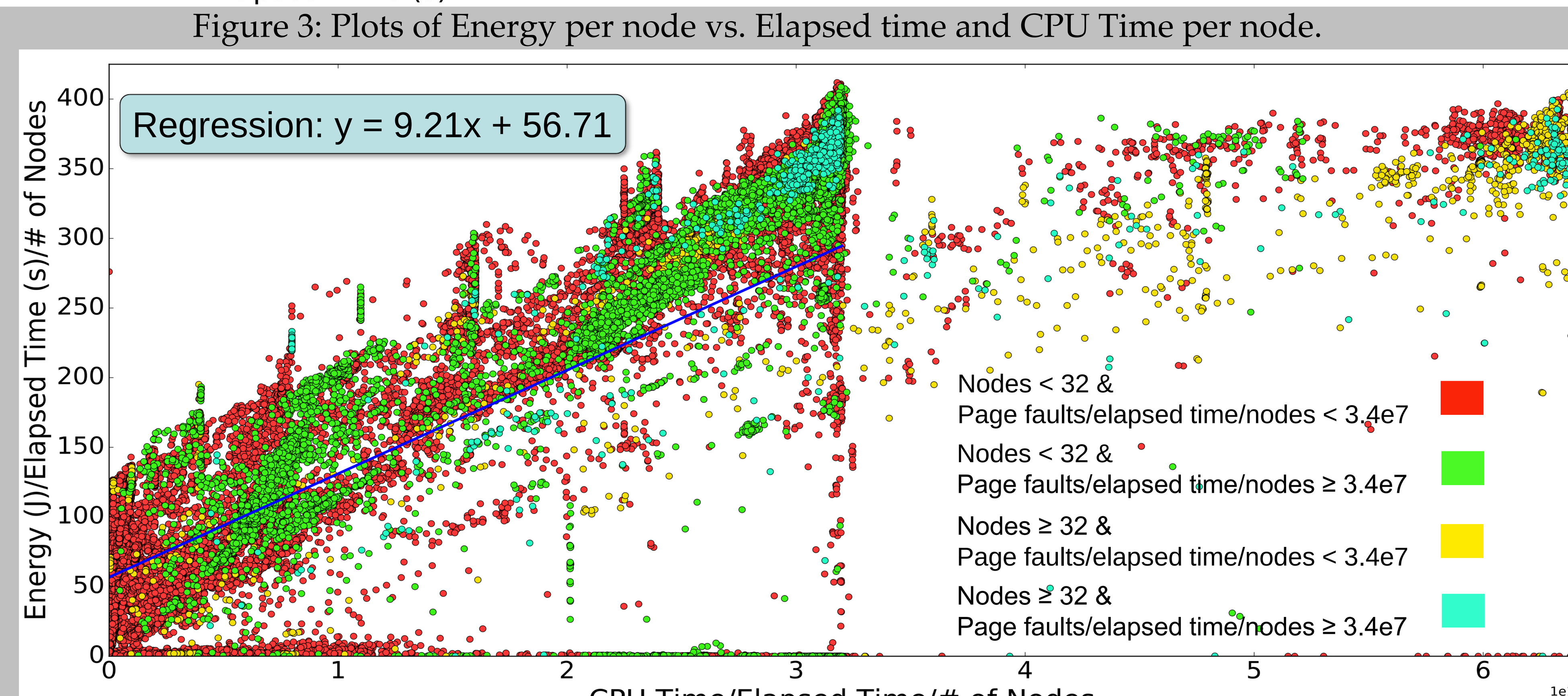

Figure 4: An interactive plot showing Watts per node vs. CPU Load. The interactive plot allow users to partition data with thresholds and manually grouping the points on the plot.

Tables

Cluster	Elapsed time (s)	AveCPU time (s)	Energy (J)/Node
Red	8.79e0	4.65e-2	5.47e2
Blue	2.44e1	1.37e1	3.22e3
Purple	1.23e3	3.35e2	6.74e3

Cluster	AveRSS/Node (b)	AvePages/Node	AveCPUFreq
Red	5.16e4	4.56e4	1.42e9
Blue	6.11e3	1.41e6	2.05e9
Purple	1.14e6	2.26e6	5.84e8

Figure 5: Table of averages for each cluster in figure 2. AveRSS and AvePages are averages per core.

Conclusions

- Monitoring energy performance in relation to other resources can lead to discovering energy usage patterns.
- Energy consumption patterns arise based on different variables such as CPU load and CPU utilization.
- Differences in WPN shows potential in energy savings.
- With the interactive plotting tool, one can observe that assigning the proper number of CPUs per node is important for energy consumption.

Future Work

- Analyzing spread of WPN values and understand causes of low WPN values
- Optimizing the number of CPUs per node depending on the resource usage patterns of job executions
- Develop suggestions to allow users and those maintaining the supercomputers to conserve energy

Acknowledgements

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists under the Science Undergraduate Laboratory Internship program. This work was also supported by the Office of Advanced Scientific Computing Research, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Contact Information: Matthew Bae - mbae@hmc.edu