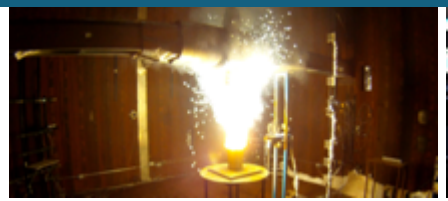
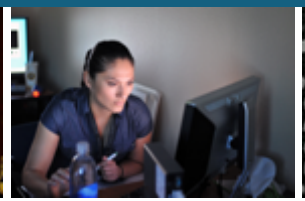
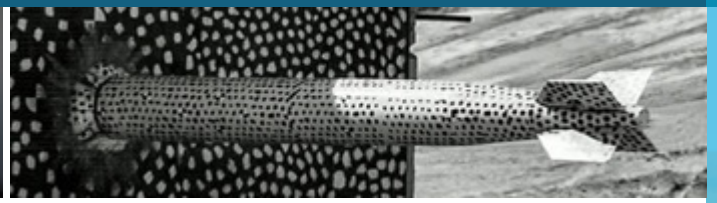
SAND2020-12691C

Sandia National Laboratories

AI/ML for HPC Operations







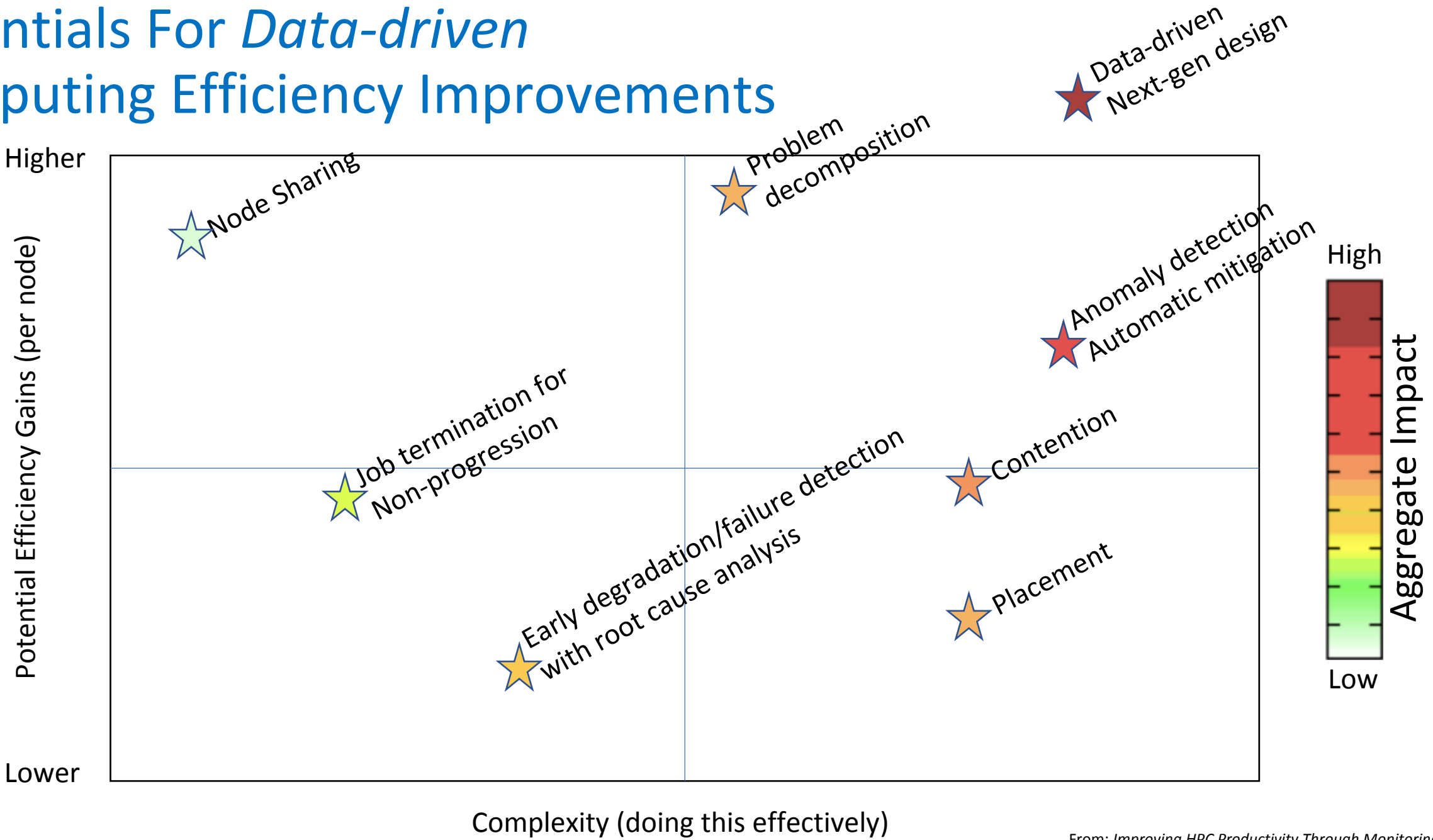
Current Issues in Computational Methods Roundtable
ANS Virtual Winter Meeting 2020

A. Gentile, HPC Development

Representing SNL's HPC Operational Analytics Team, Open Grid Computing, UIUC, and many other collaborators

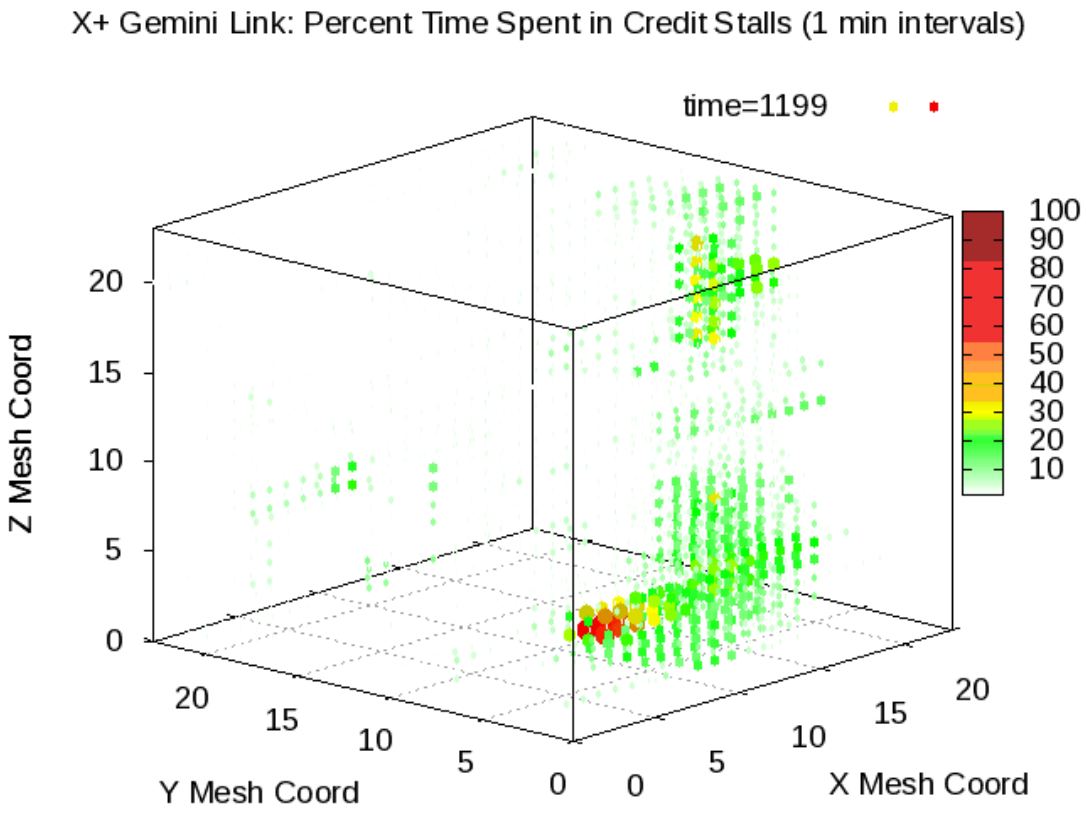
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Potentials For *Data-driven* Computing Efficiency Improvements



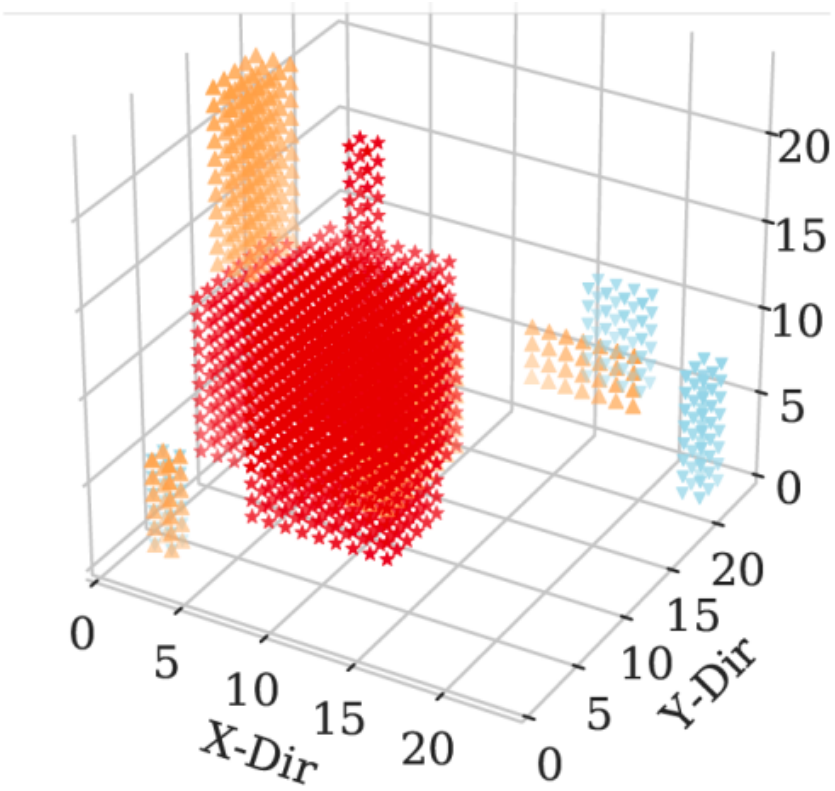
From: *Improving HPC Productivity Through Monitoring, Analysis, and Feedback SNL PESP 2019*

Characterizing Network Congestion



Plays at 10 real minutes per second

Automated Characterization of "Congestion Clouds" In Cray Gemini Networks

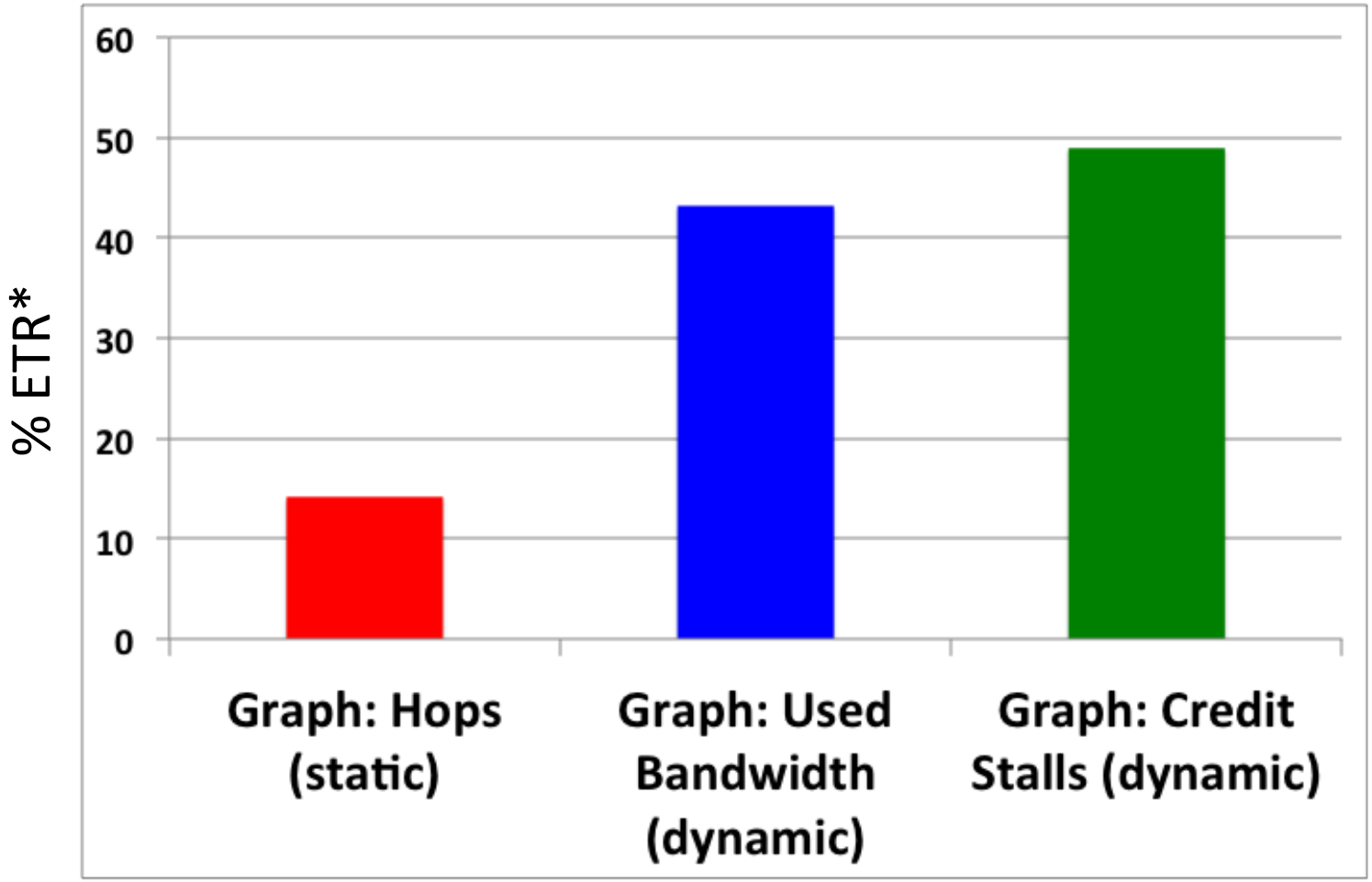


Long duration congestion clouds (direction independent) (NCSA Blue Waters)

Analysis from NCSA's Blue Waters, *Lightweight Distributed Metric Service: A Scalable Infrastructure for Continuous Monitoring of Large Scale Computing Systems and Applications SC14*

From *Measuring Congestion in High Performance Data Center Interconnects NDSI20*

System Assessment Drives Application Response



Remapping based on **dynamic congestion assessment and application communication patterns** recovered ~50% of the time otherwise lost to heavy congestion.

Graph edges weighted by network:

- hops required
- bandwidth used
- credit stalls

*Execution **Time Recovered**
(higher is better)

From *Demonstrating Improved Application Performance Using Dynamic Monitoring and Task Mapping HPCMASPA2014*

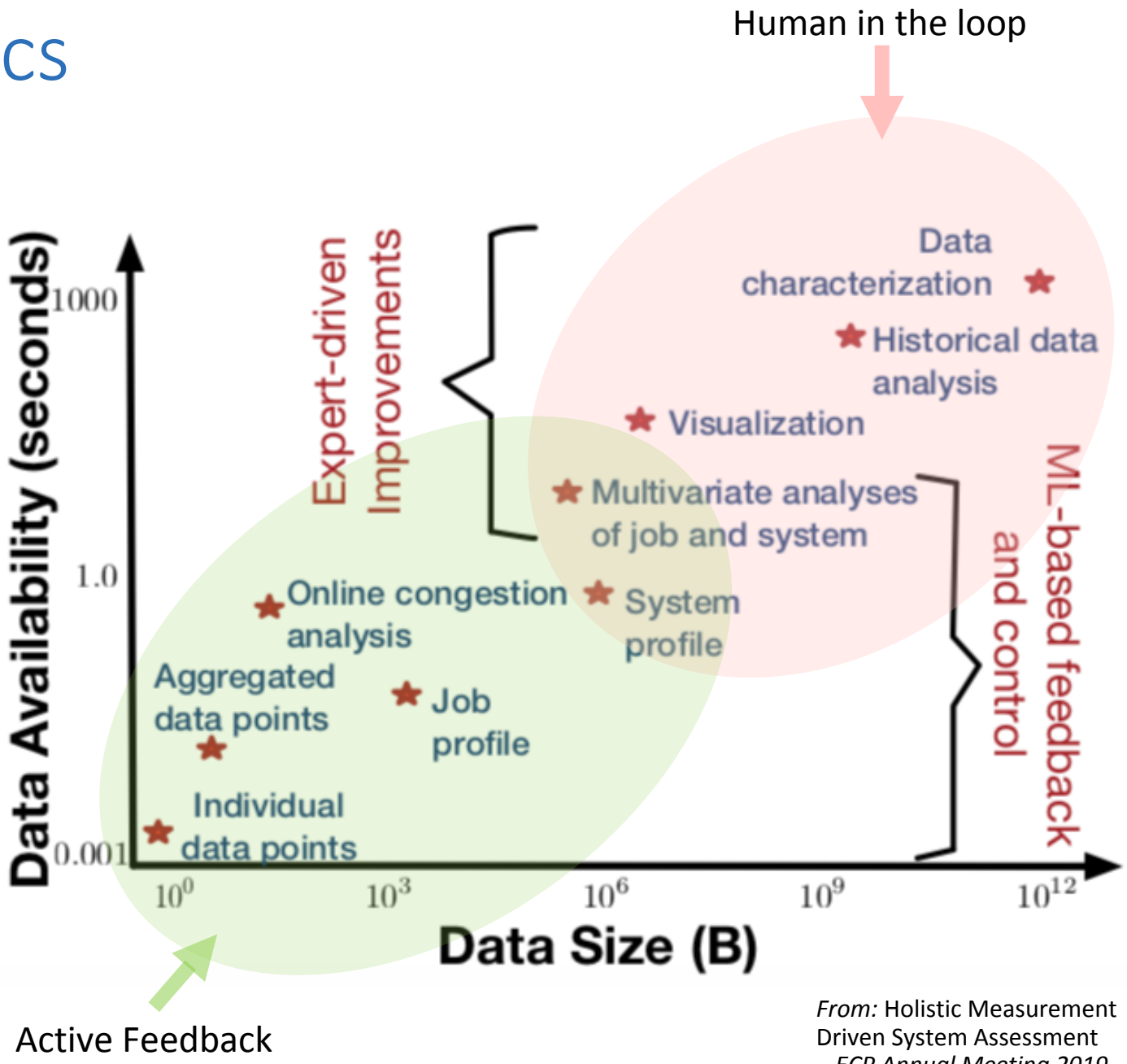
System Assessment Drives System Software Response

- ML-based characterizations of applications as bandwidth-intensive or latency-sensitive
- Dynamic detection of congestion and resource-targeted injection throttling to ensure overall workload throughput

From Application-aware Congestion Mitigation for High-Performance Computing Systems in submission

Actionable Data Analytics

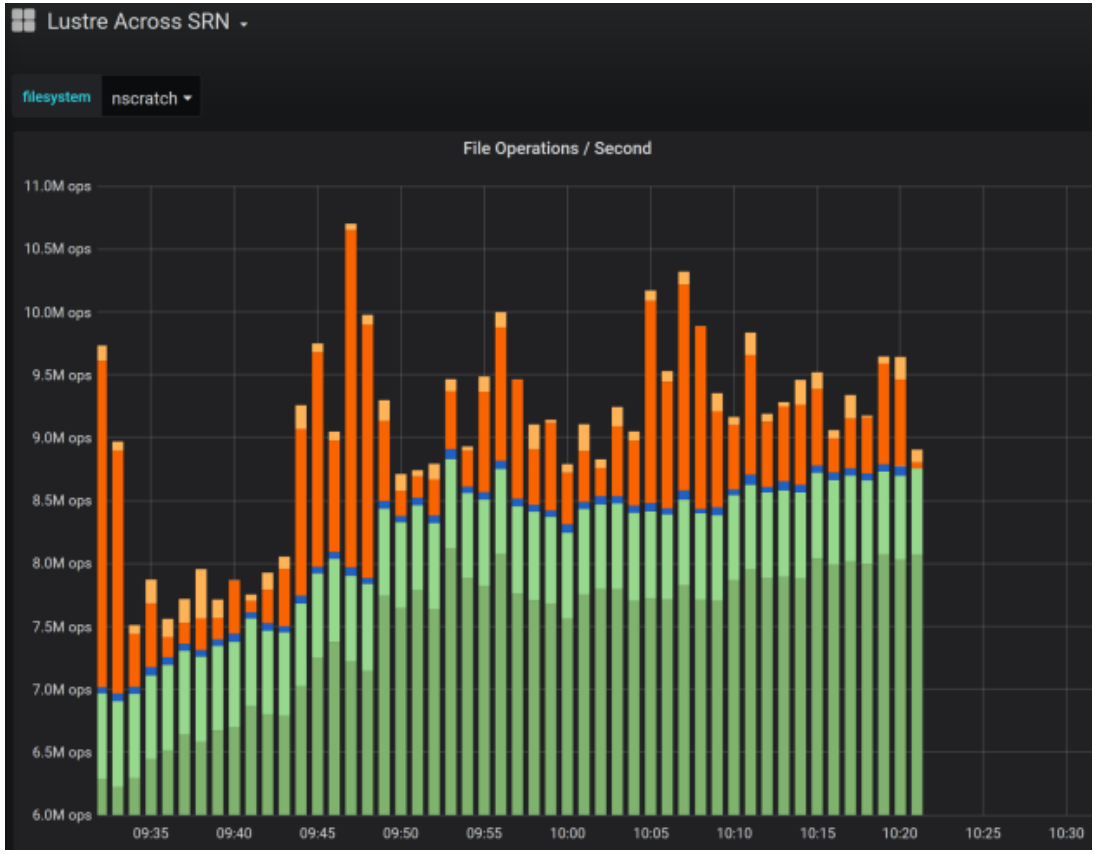
- Data features:
  - High volume (~10s of TB/day)
  - High dimension (100s to 1000s of discreet variables) data
  - **Getting** data is *not* a challenge!
- Machine Learning (ML) focus areas:
  - Fundamental techniques for time-series, rare events
  - Physics-constrained ML -> Architectural constraints
  - Validated and explainable ML for automated response

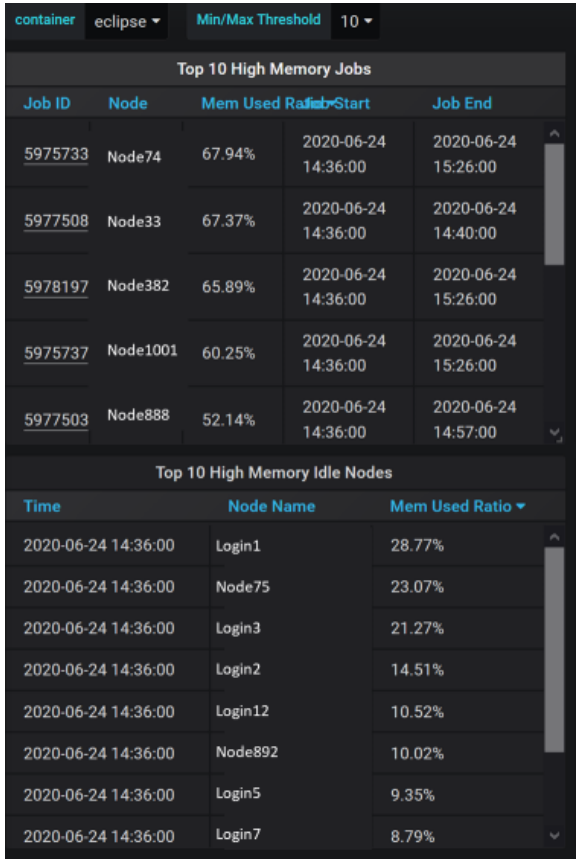


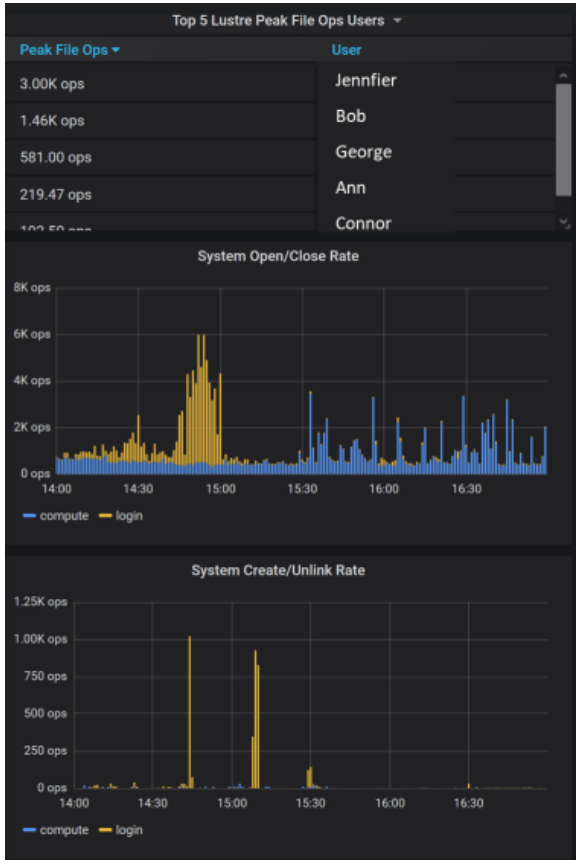
From: Holistic Measurement Driven System Assessment *– ECP Annual Meeting 2019*

Visualization Portal for Admins and Users

- System Status e.g., center-wide filesystem performance, high memory usage jobs on a system
- Goal: Extend to application progress/performance overlaid on system performance measures







Partially From: *HPC System Data Pipeline to Enable Meaningful Insights through Analytic-Driven Visualizations HPCMASPA2020*

Applications Can Facilitate Data-driven Computing Efficiency Improvements!

- Can we get at submission time which runs should be comparable?
  - Build normal profiles and detect abnormal progress
  - Use learned resource-utilization information for co-scheduling decisions
- What can you expose in a run-time data stream to indicate performance and throughput? To assess performance sensitivity?
  - Performance counters? Call stack? Science variables?
- What would be meaningful representations of performance-impacting conditions?
  - Full network topology would be too complex
- Can your application be responsive to feedback (e.g., rebalancing, migrating)?
- Long-running dialog with applications team to understand and diagnose issues (e.g., “slow nodes”).