

Exascale-ready operations, improved performance, and balanced system design via comprehensive system monitoring and analysis

National Center for Supercomputing Applications and University of Illinois: Bill Kramer (PI)

Greg Bauer, Brett Bode, Mike Showerman, Jeremy Enos, Aaron Saxton, Saurabh Jha, Zbigniew Kalbarczyk, Ravi Iyer

Sandia National Laboratories: James Brandt, Ann Gentile

Institutional Collaborators: Boston University, New Mexico State University, University of Central Florida



Answering Key Operational Questions

Is my performance variation due to system conditions or code changes?

User

Capability: Combined reasoning based on changes in application characteristics (e.g., MPI message size) and system characteristics (e.g., network congestion) to diagnose cause(s) of performance variation and predict application runtime.

Figures of merit on Cray Aries interconnect system

Job ID	App ID	Node ID	Start Time	End Time	HPages	NW Congest	Mens Score	Anomalies	PAPI Perf	App Perf	Comm Perf
10399	CoMD	nid000[53,59]	2018-09-21 00:28:40	2018-09-21 00:30:40	0.0	1.57	0.042	0x1	0.0	0.59	0.0
10398	iutesh	nid000[60,61]	2018-09-21 00:19:26	2018-09-21 00:22:05	0.0	0.21	0.041	0x0	0.29	1.55	0.0
10397	CoMD	nid000[53,59]	2018-09-21 00:19:19	2018-09-21 00:21:28	0.0	0.06	0.042	0x0	0.1	0.59	0.0
10396	CoMD	nid000[21,22]	2018-09-21 00:19:16	2018-09-21 00:21:25	0.0	2.75	0.053	0x0	0.0	0.61	0.0

Different Cray Aries with competing application, different Aries, same Aries (self-contention)

Capability: Diagnose network issues using extracted congestion regions (e.g. Cray Gemini interconnect)

System Managers & Users

Capability: Diagnose file-system issues using periodic latency measurements of file system operations (e.g. write)

Architects & Acquisitions Teams

Capability: Inform design decisions for future systems

Insight: Assessing dynamic range of memory needs for facilities' workflows

Architects, System Managers & Support Staff

Capability: Dynamic task re-mapping based on congestion measurements

Capability: Dynamic task re-mapping based on congestion measurements

Insight: Identifying issues with network routing and congestion mitigation procedures

Capability: Topology-aware scheduling (TAS) based on congestion measurements

Impact of TAS on system performance

Exascale-ready Features

Data-driven and Machine Learning (ML) based mechanisms for data collection, analysis, and automated runtime feedback and control

Scalable APIs that enable flexible runtime interactivity (e.g., streaming analytics)

Platform independent open-source solutions

Key Strengths

Lightweight, low overhead mechanisms enable high fidelity (e.g. sub-second) synchronized, whole system numeric and event data capture with negligible impact on application runtime

Discovery and analysis of performance and resilience related phenomena via integrated system logs and numeric data

Low latency feedback of analysis results to system software, applications, and system managers

High resolution extraction and classification of phenomena with respect to locality, severity, and temporal extent

Significance and Impact

- Enable **highly efficient HPC system usage** and inform future system improvements to produce more science

- Transform **real-time data into actionable intelligence** at runtime

Scalable System Design and Architecture

Data Collection: Processor Interconnect, Storage, Event Logs, Blade and Cabinet Controller Sensor Data, Job Application Resource and Performance Data, LDMS, Human in the Loop Analysis Feedback, Machine in the Loop Analysis Feedback.

Analysis: HIGH PERFORMANCE/PROJECT DRIVEN ANALYSIS AND PREDICTION TOOLS (Log and Event Analysis, Event Prediction and Correlation, ML/AI, Signal Analysis), Database and Active Data Repository, Notification System, Web Server, BUILDING AND FACILITIES (Power System, Facility Systems, Environmental Systems).

Actionable Intelligence: Email, Text and Event Notifications, Admin Web MOTO, Log Browsing, Data Mini, Web Portals API, Dashboards, Analysis Research, System Visualization.

Feedback and Control: Feedback and Control loop between Analysis and Actionable Intelligence.

Expert-driven Improvements: Data characterization, Historical data analysis, Visualization, Multivariate analyses of job and system, Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.

ML-based feedback: ML-based feedback loop between Data Collection and Analysis.

Architecture supports big data analysis, provenance, and curation. Currently in production on Blue Waters:

- Real-time collection and processing for 28,000 nodes and all sub-systems

- * Process 20 B datums per day

- * Active storage of 11 trillion datums

- * 8,100 active unique log and data event types

- * 50,000 different event types seen to date

- * Up to 1,470M log and data events/day (30M avg)

Data Availability (seconds) vs **Data Size (B)**

Expert-driven Improvements: Online congestion analysis, Job profile, Aggregated data points, Individual data points.