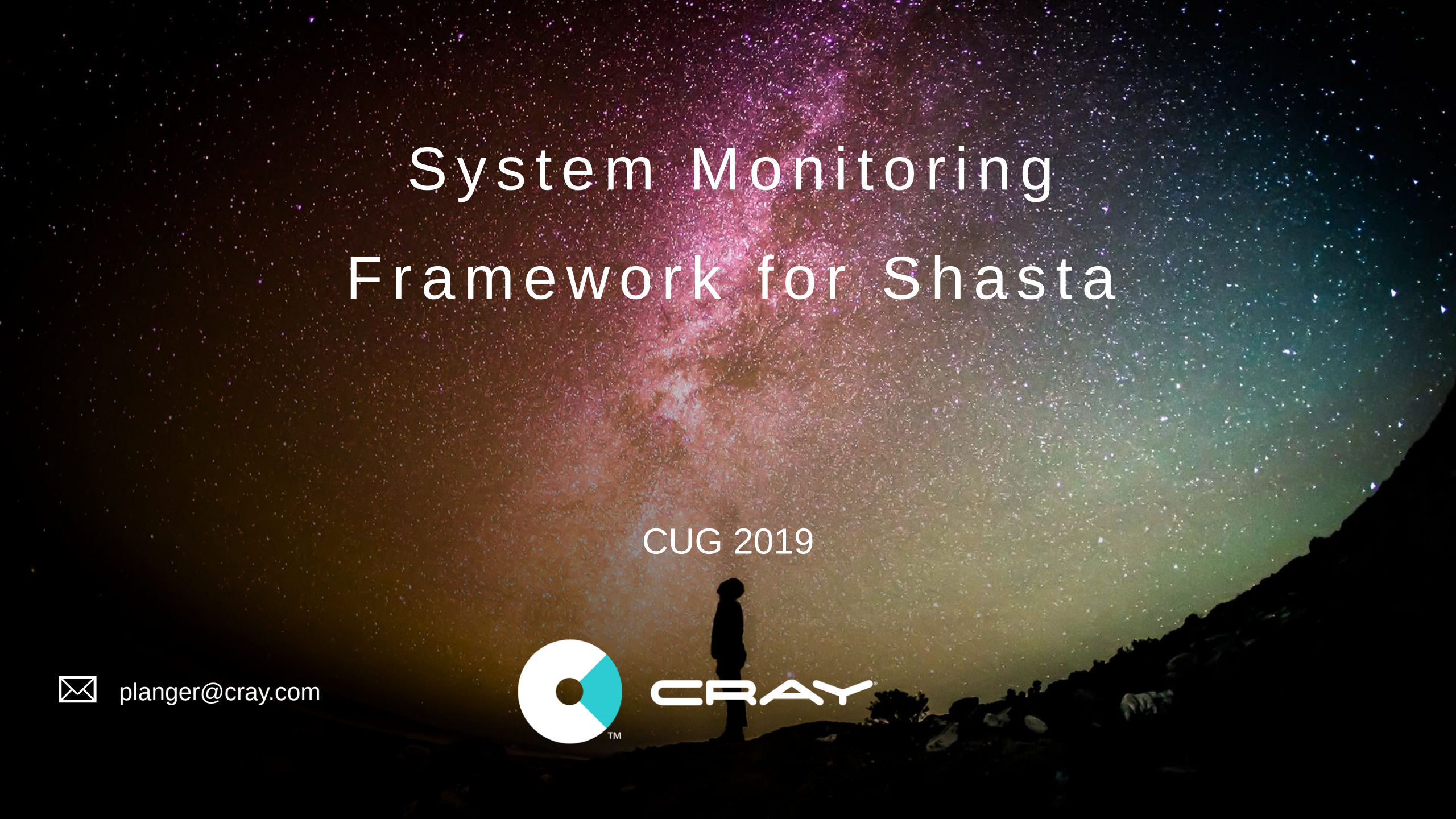
System Monitoring Framework for Shasta

CUG 2019



planger@cray.com

CRAY™

TOPICS

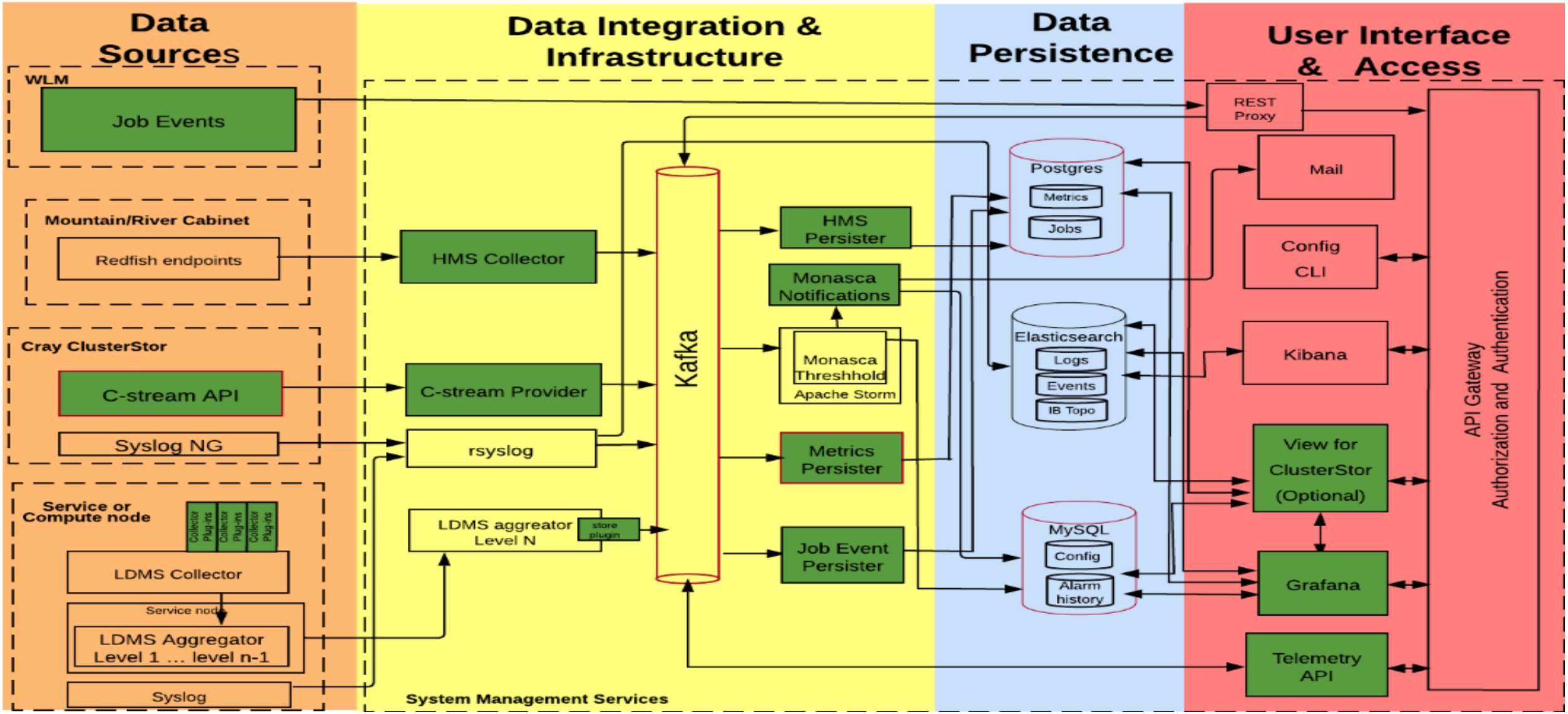
- Overview of the system monitoring framework
- Subsystems contributing metrics
- Correlating data with visualization tools
- Summary
- Q&A


Overview

SYSTEM MONITORING FRAMEWORK

- What is the System Monitoring Framework?
  - A tightly integrated framework for collecting and persisting metrics and logs
  - Consolidates telemetry data from multiple subsystems
    - Switch fabric
    - Network
    - Job Management
    - Storage
    - Power
    - User Applications
    - Compute
  - Integrated alarm and notification framework with threshold engine
  - Standard visualization tools for graphing metrics and searching logs
  - RESTful API for integration into customers monitoring solutions
  - Integrated with the diagnosability and serviceability solutions

ARCHITECTURE AND DATA SOURCES

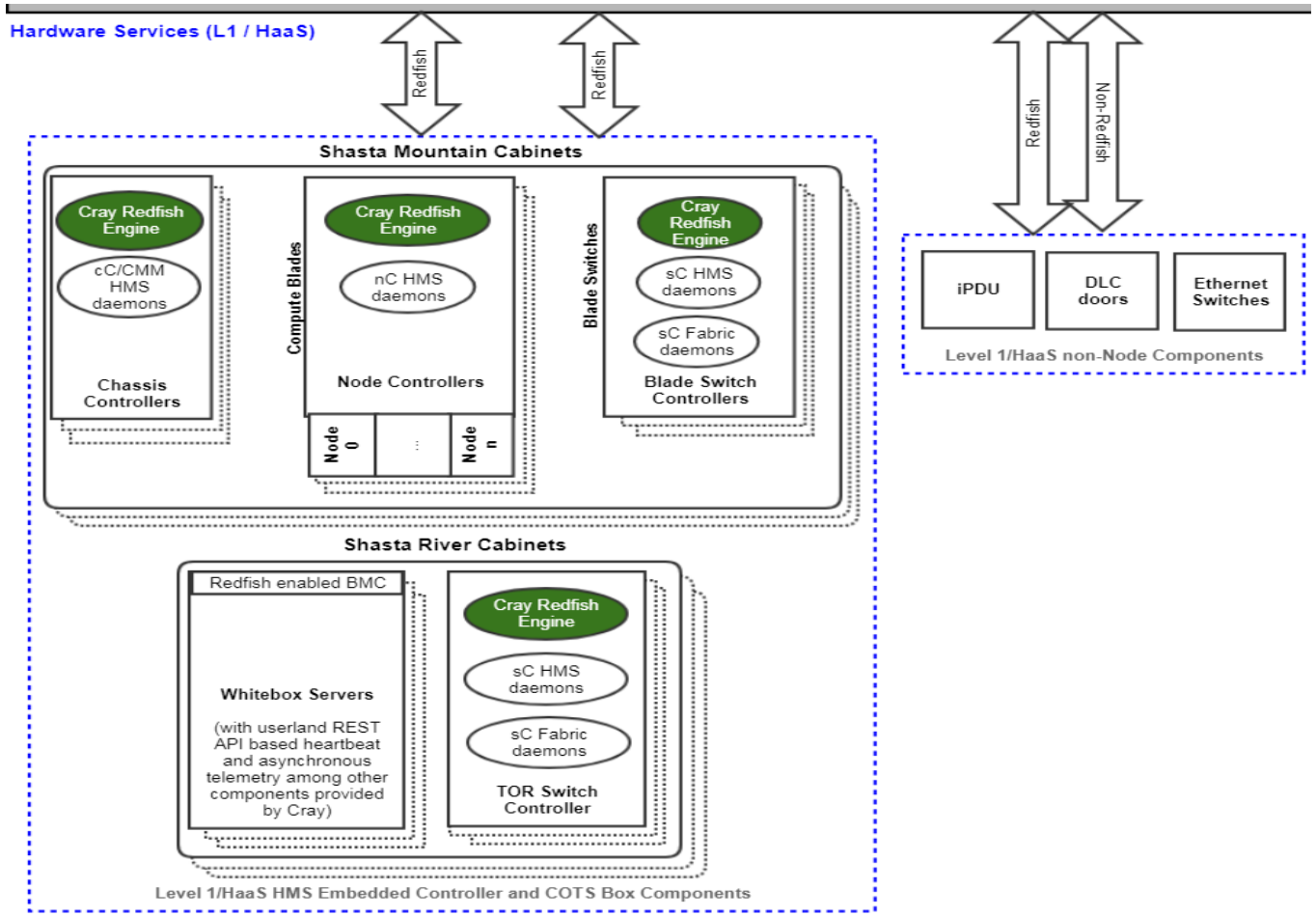



Data Sources

SUBSYSTEMS CONTRIBUTING METRICS

- Shasta hardware
- ClusterStor storage
- Compute nodes
- Network and fabric
- Logs

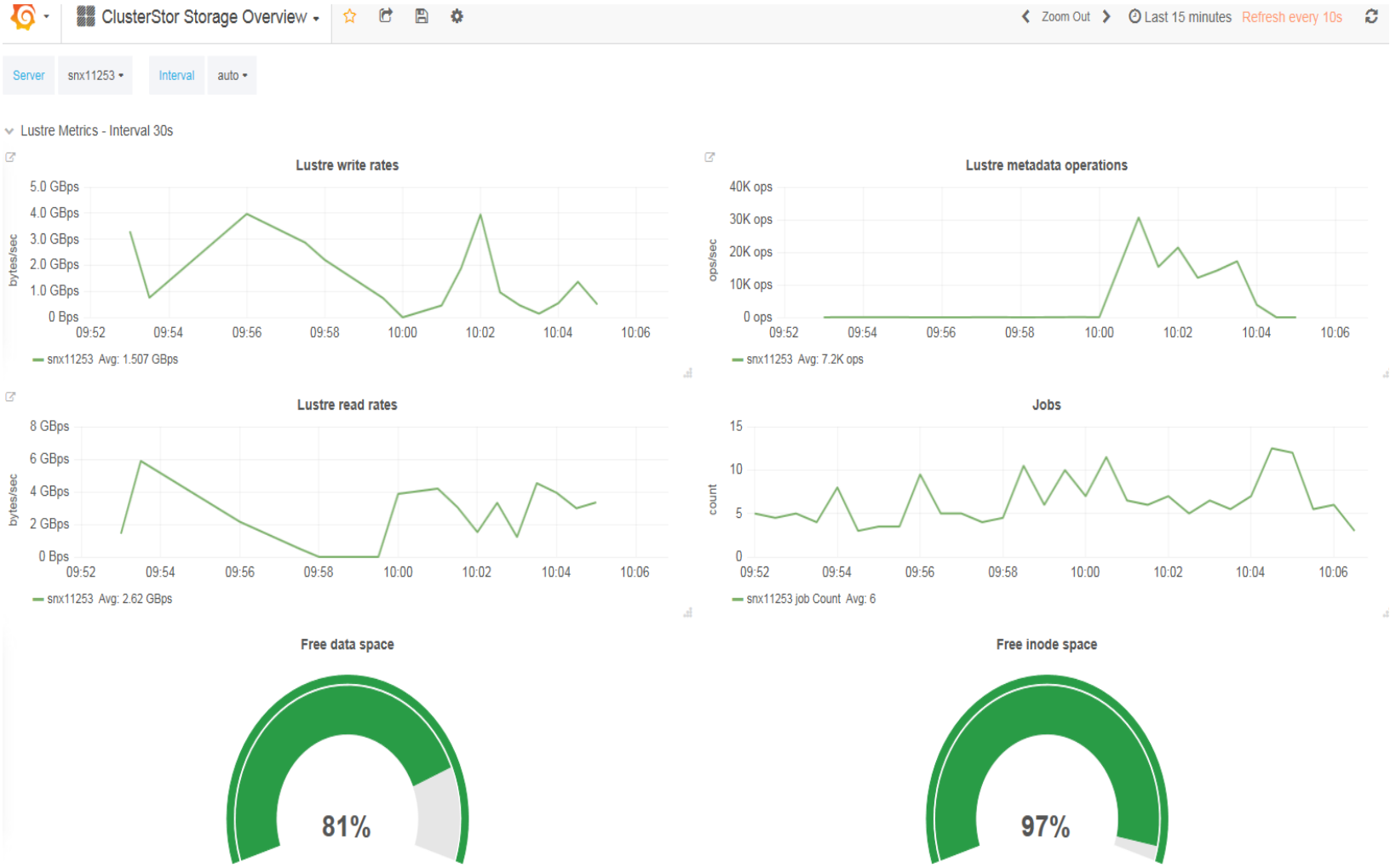
HARDWARE MANAGEMENT METRICS



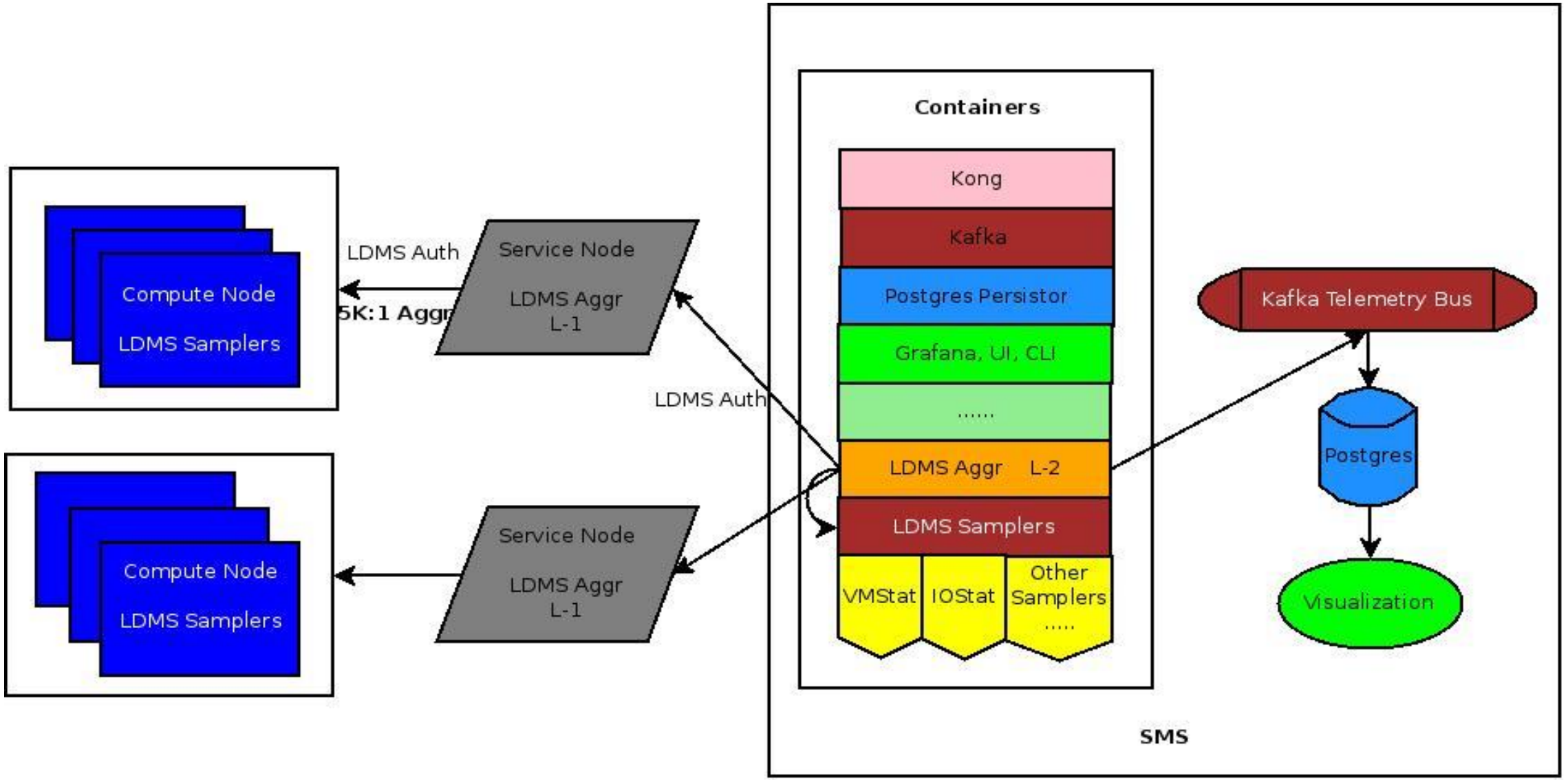
- Collect metrics from
  - Chassis controllers
  - Node controllers
  - Blade switch controllers
  - PDUs
  - TOR switches
- Collected using industry standard redfish API

CLUSTERSTOR STORAGE METRICS

- Metrics collected
  - Lustre performance
    - Metadata, OST I/O read/write
  - Lustre jobstats
  - Logs and events
- Collection rate : 15 to 30 seconds
- Calculated into delta rates and persisted
- Enables trend analysis

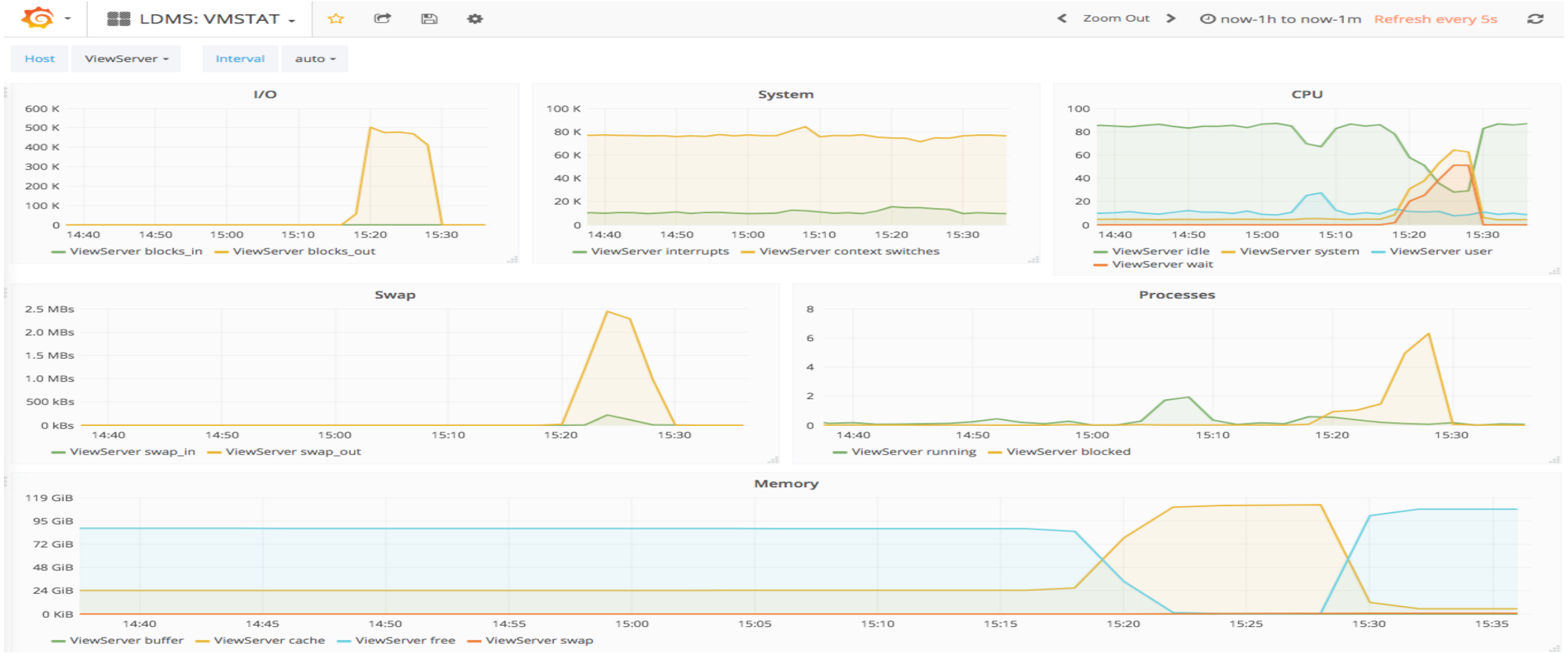


COMPUTE NODE METRICS VIA LDMS



COMPUTE NODE METRICS

- Six main categories: I/O, System, CPU, Swap, Processes & Memory
- Total of 13 metrics sampled at 10 second interval

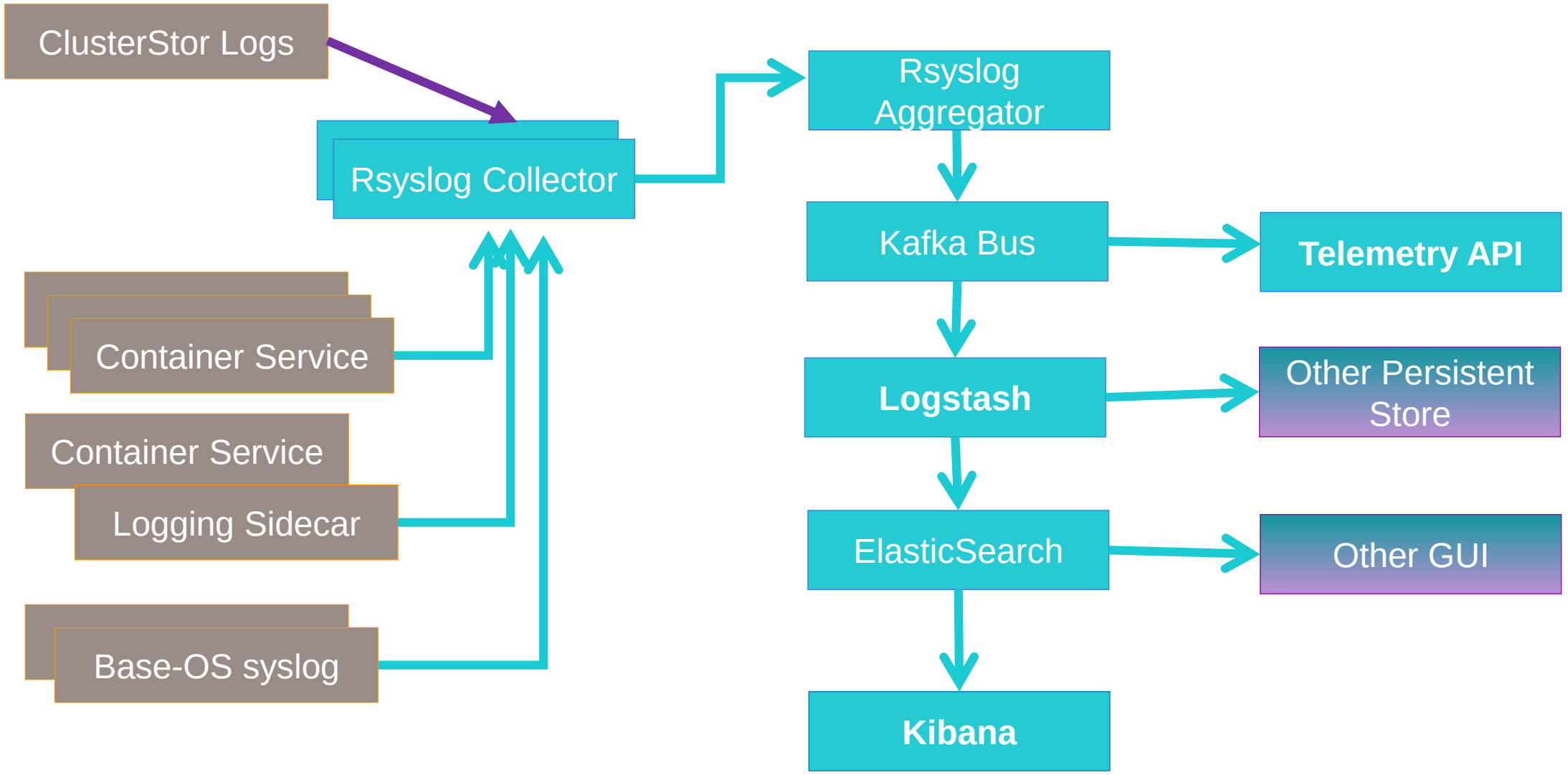


NETWORK/FABRIC METRICS

- Metrics are collected to enable monitoring and diagnosis of performance and congestion of the fabric
- These metrics will include:
  - Critical asynchronous link events and port state changes
    - e.g. used for diagnosis of link/cable issues
  - Running state data based on a configured set of standard SNMP MIBs
    - RFCs 1213, 2819, 2863, 3635, 4188, 4293
    - Data periodically posted, period is configurable
    - Types of bandwidth and congestion metrics collected
      - Packets/bytes in/out
      - Unicast/Multicast/Broadcast
      - Drops/errors
      - Pause Frames in/out
        - e.g. excessive transmit pause frames used to identify error at endpoint device
- All telemetry data includes locality of metric
  - Provides ability for focused query/heat map generation on specific area of the fabric

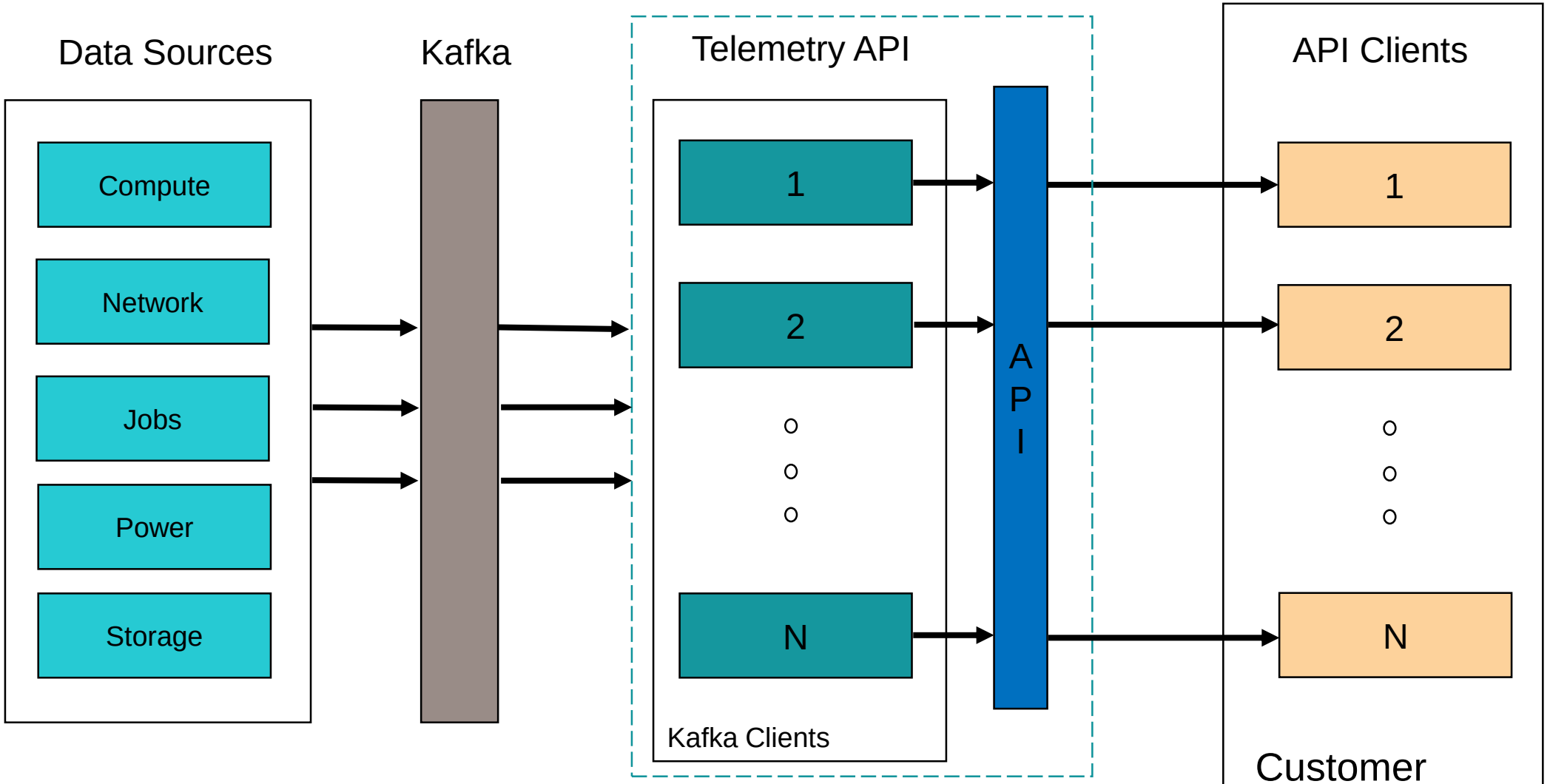

Log Aggregation

LOG AGGREGATION



Integration with 3rd Party Monitoring System

TELEMETRY API



Shasta Monitoring Framework

USE CASE DEMO

SUMMARY

- The System Monitoring Framework aggregates metrics into a single framework
- Telemetry collected includes :
  - Shasta hardware
  - Storage lustre and job metrics
  - Logs
  - VMStats from compute nodes
  - Network and fabric metrics
- Tools are provided to enable trend analysis, searching, and correlating of data
- A REST API is provided to allow streaming of telemetry off the kafka bus into customer monitoring solutions

SAFE HARBOR STATEMENT

This presentation may contain forward-looking statements that are based on our current expectations. Forward looking statements may include statements about our financial guidance and expected operating results, our opportunities and future potential, our product development and new product introduction plans, our ability to expand and penetrate our addressable markets and other statements that are not historical facts.

These statements are only predictions and actual results may materially vary from those projected. Please refer to Cray's documents filed with the SEC from time to time concerning factors that could affect the Company and these forward-looking statements.

THANK YOU

QUESTIONS?