Cross Reference Monitoring of Supercomputers and Support Infrastructure

Tracking data Issues in Cheyenne HPC

Steven Rivera

SIParCS Intern, University of Puerto Rico

JULY 28, 2020

Mission

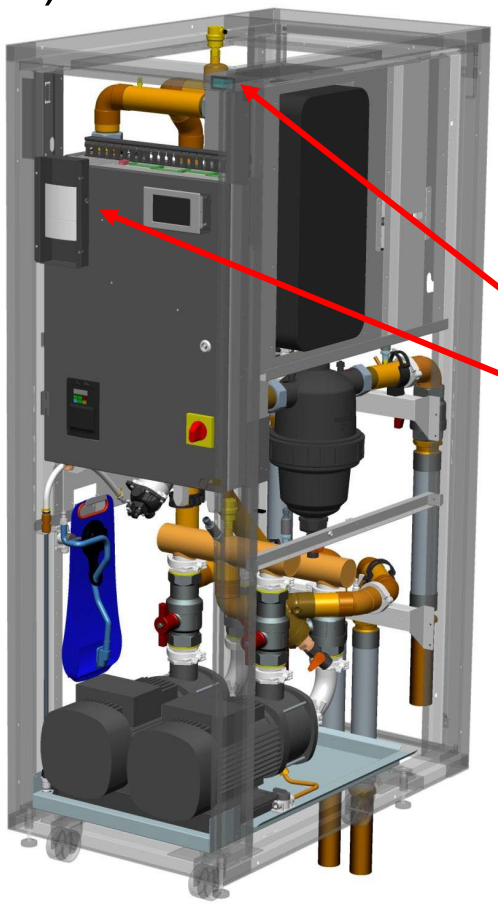
- Researching inside data of the Supercomputer Infrastructure.
- Finding discrepancies of hardware equipment failure.
- Differentiating between one anomaly and another.
- Finding solutions to improve internal processes.

Cheyenne

- Computing performance of 5.34 PFLOPS.
- 4032 nodes
  - A total of 145,152 cores.
  - There are 33 sensors in each node.
  - Giving us a total of 133,056 sensors in computes nodes .
- There are total of 7 valves and 7 Cooling Distribution Unit (CDU).
  - The CDUs provide 75 °F water to cool the compute nodes processors and other hardware.

CDUs Infrastructure

1)

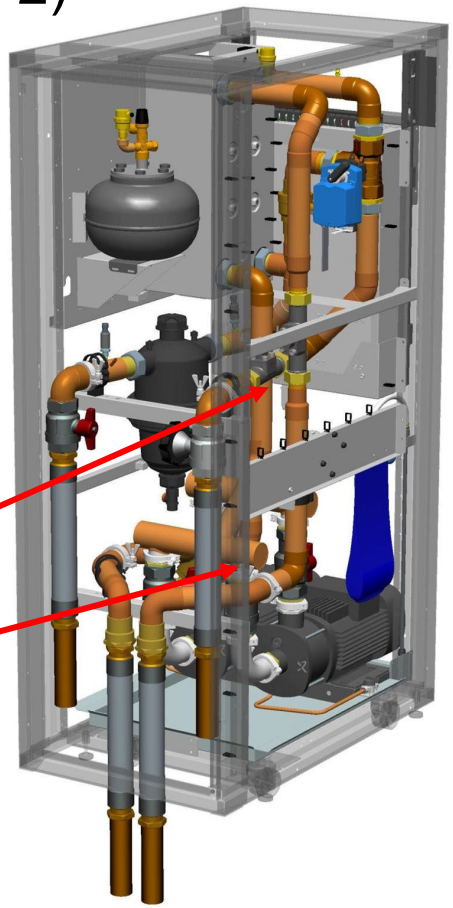


Cooling Distribution Unit (CDU)

1) Front
- Temperature display.
- Room temperature sensor.

2) Back
- Flow meters
- Level Sensor

2)



Physical Infrastructure

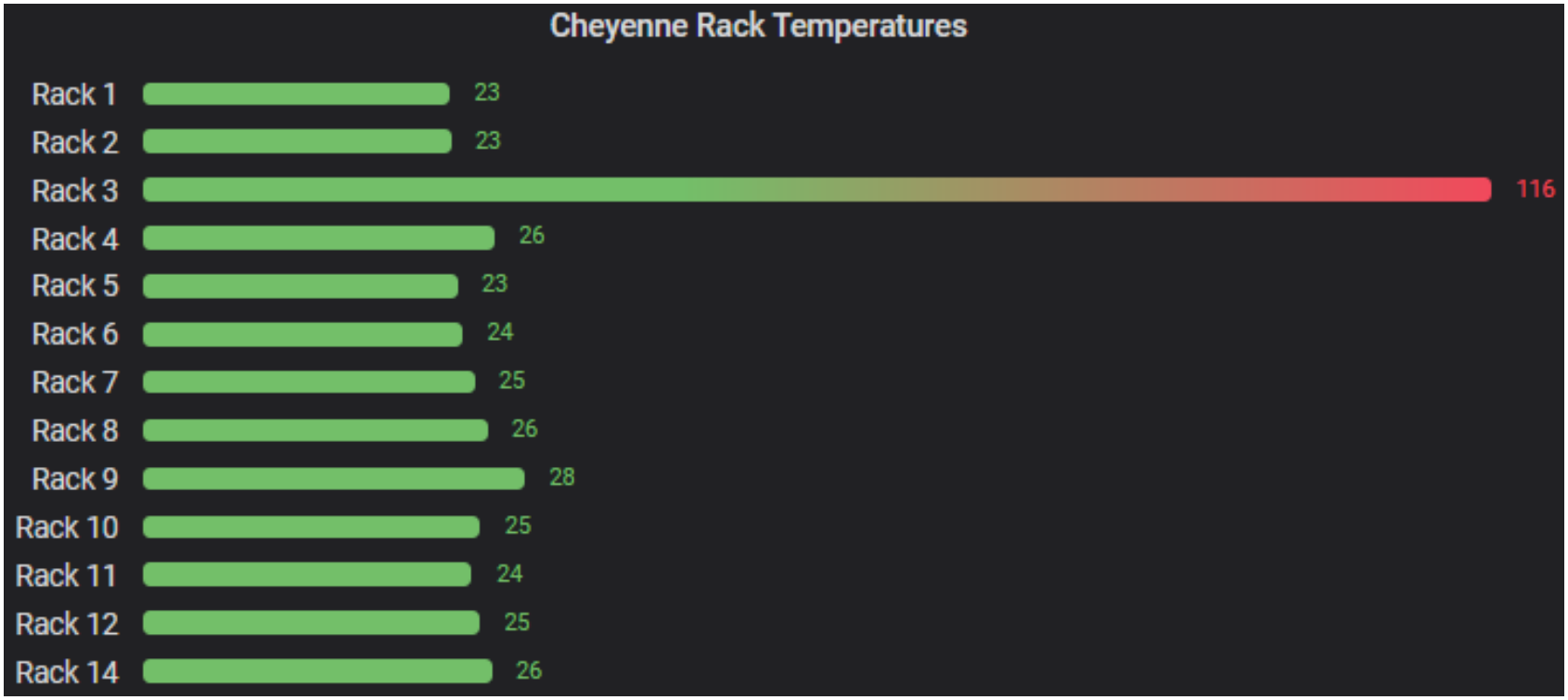
- The power usage through all the system is around
  - 1.3MW
  - 1.75MW (max.)
- NWSC evaporates around 25000-35000 gallons/day at 1.5MW.
  - 40% more efficient than most data centers.
- Racks runs with 96KPa of water pressure.
- Each compute Rack uses:
  - 78.2kVA
  - 79.8kVA (max.)
- A water flow of 350 l/m loops around each CDUs.

Software Tools

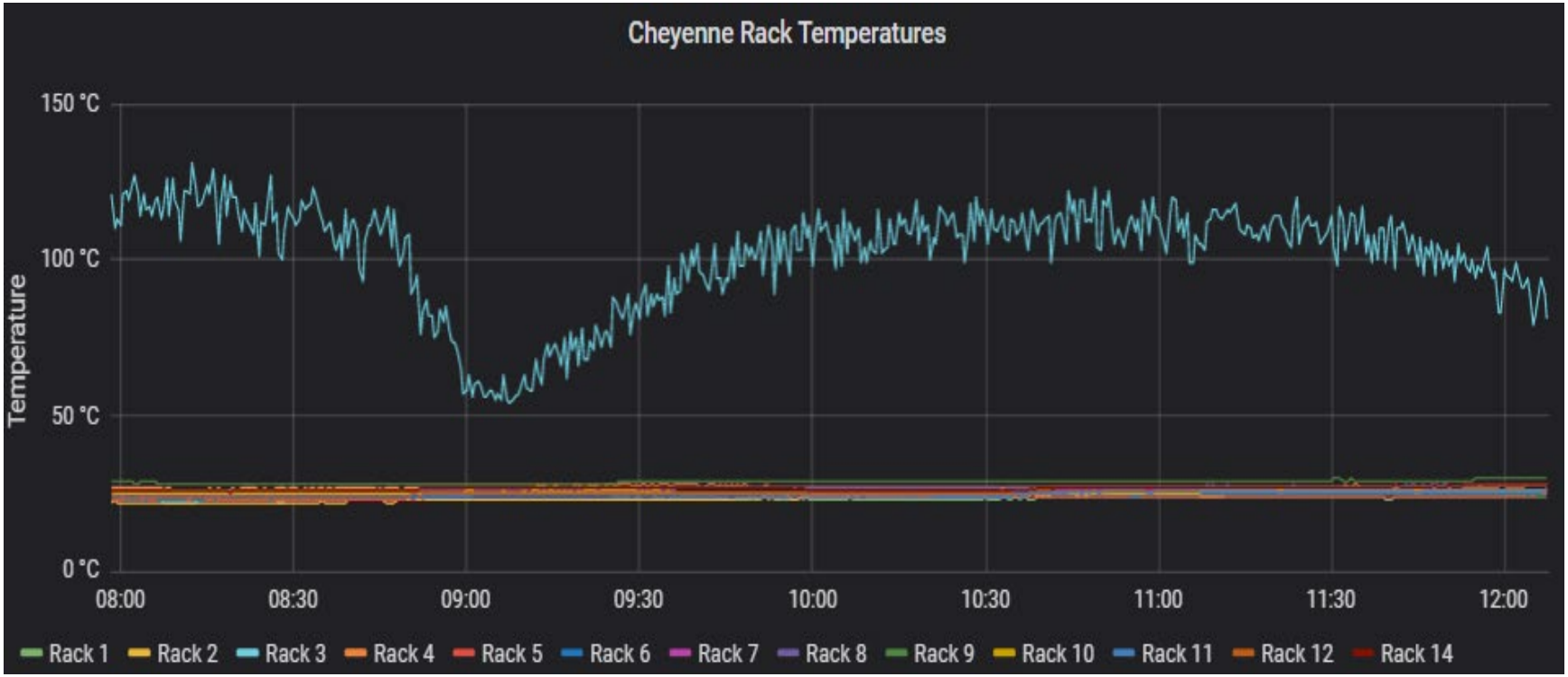
- Grafana
- InfluxDB
- Python
- Johnson Control
  - MetaSys

Process/Monitoring

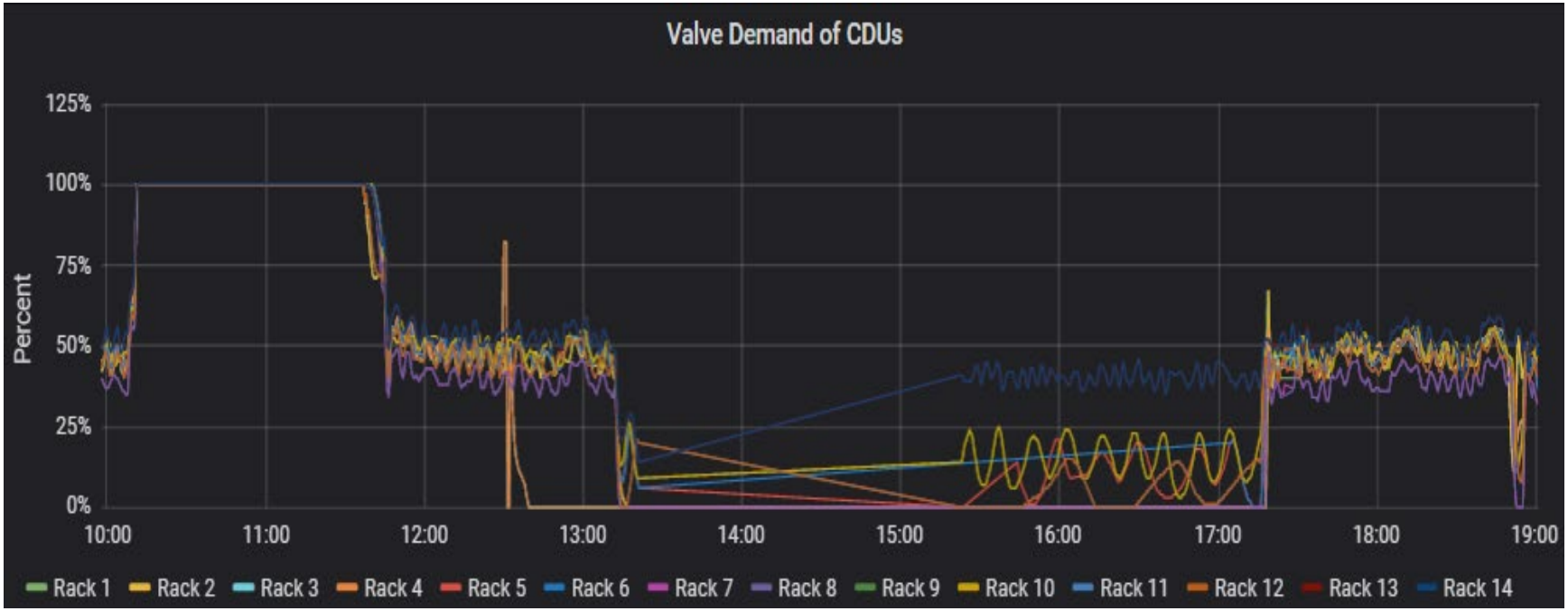
- Analyzing what data was shown from Cheyenne.
- Understanding the different limits.
- Finding anomalies on specific devices.
  - Different types of queries: data query language.
- Devices names in the queries:
  - Showing a general behavior.
- Scaling any difference with other similar behavior such as:
  - Increases and reductions of load.
- With large number of Racks and nodes it was presented:
  - Unwanted behavior.
  - Large significance load.



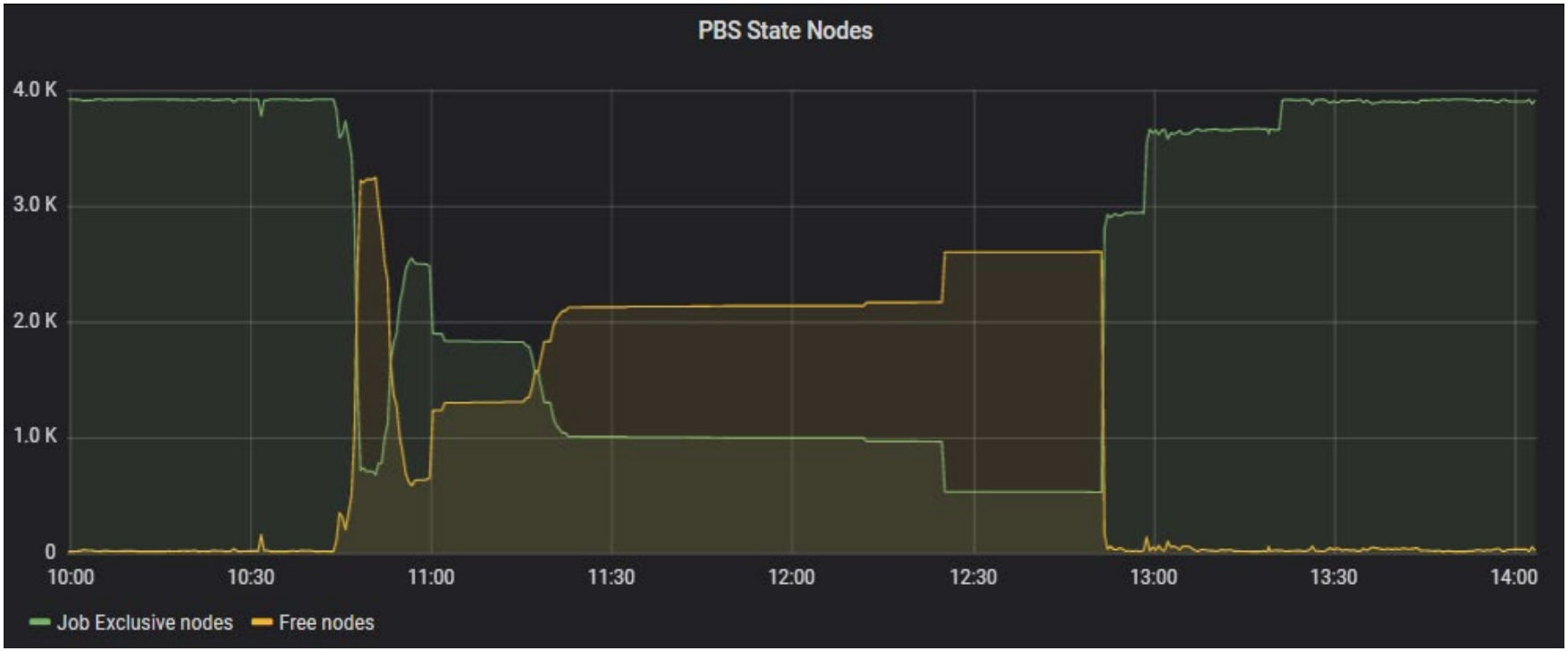
- A similar flow is expected between all Racks in °C.
- Suspicion in a device value higher than 26 °C.
  - In this case, Rack 3.
- Any operation big enough to be noticable can be real anomalies.



- Possible causes for the behaviour in Rack 3:
  - Bad water flow
  - Power outage
  - Defective sensor



- A rise of 50% additional usage in every Rack.
  - 100% is a critical state
- Random downgrade of their function, first Rack 4 followed by Rack 8.
- In 4hrs there was a percentage usage below 23%, except Rack 13.



- Increase over 2k of free nodes system.
- Decrease of Job Exclusive nodes 1k and below.
- Opposite process can freeze, damage or delete job submissions.

- Power output from the power su pply 0 to power supply 8.
- Approximate reduction of 0.5MW.
- Increased occurred with 0 .3MW less.

Power Supplies	Minimum	Maximum	Average
• Supply 0	0.382 MW	1.376 MW	1 MW
• Supply 1	0.383 MW	1.332 MW	0.996 MW
• Supply 2	0.383 MW	1.328 MW	1 MW
• Supply 3	0.381 MW	1.368 MW	0.996 MW
• Supply 4	0.382 MW	1.348 MW	0.999 MW
• Supply 5	0.400 MW	1.328 MW	0.999 MW
• Supply 6	0.384 MW	1.328 MW	0.998 MW
• Supply 7	0.379 MW	1.346 MW	0.997 MW
• Supply 8	0.382 MW	1.330 MW	0.998 MW

- Minimum operation was around 0.400 MW.
- Average operation is 1MW.
- Maximum operation is 1.3 MW.

Results

- An anomaly was found in one of the 14 Racks:
  - Real case for temperature of liquid in Rack 3.
- Appearances of high Fluctuation were shown:
  - With a downscale of over 20% of usage in Racks.
- System discrepancies can lead to:
  - Decrease in node availability
  - Temperature increases
  - Loss of running jobs
  - Performance irregularities
- System maintenance can downscale severe temperatures, lower flow of water and reduce future expenses.
- System functions are most of the time stable in temperatures.

Future Improvements

- Live monitoring versus (current) historical analysis
- Automatic alerting via email or web interface when known discrepancies occur
- A way to log historical system behaviours with a database backend

Acknowledgements

Big Thanks to:

- The SIParCS program for the internship
- My mentors Michael Kercher, Jonathan Roberts and helpers Ben Matthews, Jared Baker for their support.
- NCAR and UCAR
- National Science Foundation