Deploying Filesystem Performance Metrics Through XSEDE Metrics on Demand (XDMoD)

Broday Walker

July 31, 2019

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

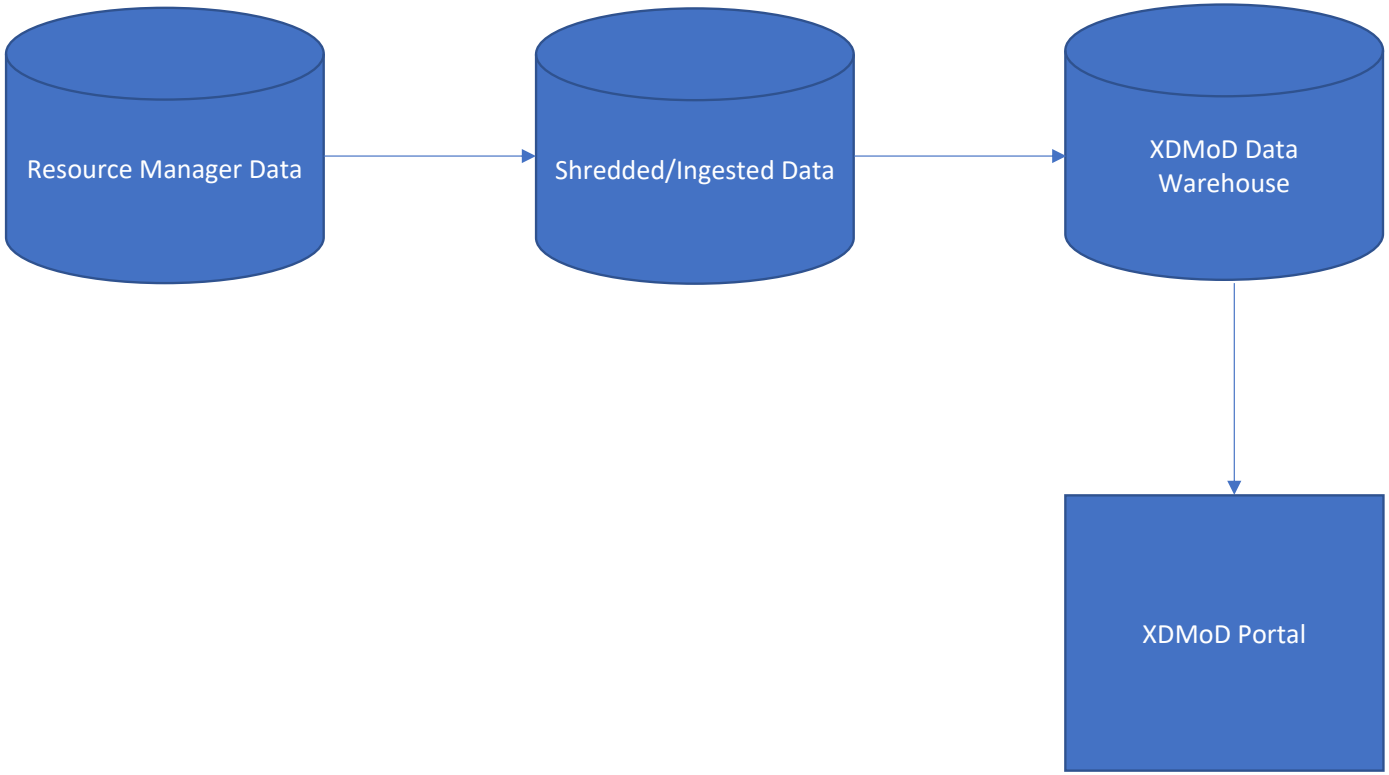
Motivation

- Upgrade XDMoD instance to version 8.1
- Upgrade SUPReMM to version 1.2
- Include new filesystem I/O metrics
- Generate a weekly report characterizing resource usage on Cheyenne
- Update and improve procedure for installing, upgrading, and configuring XDMoD and SUPReMM on internal wiki page and GitHub

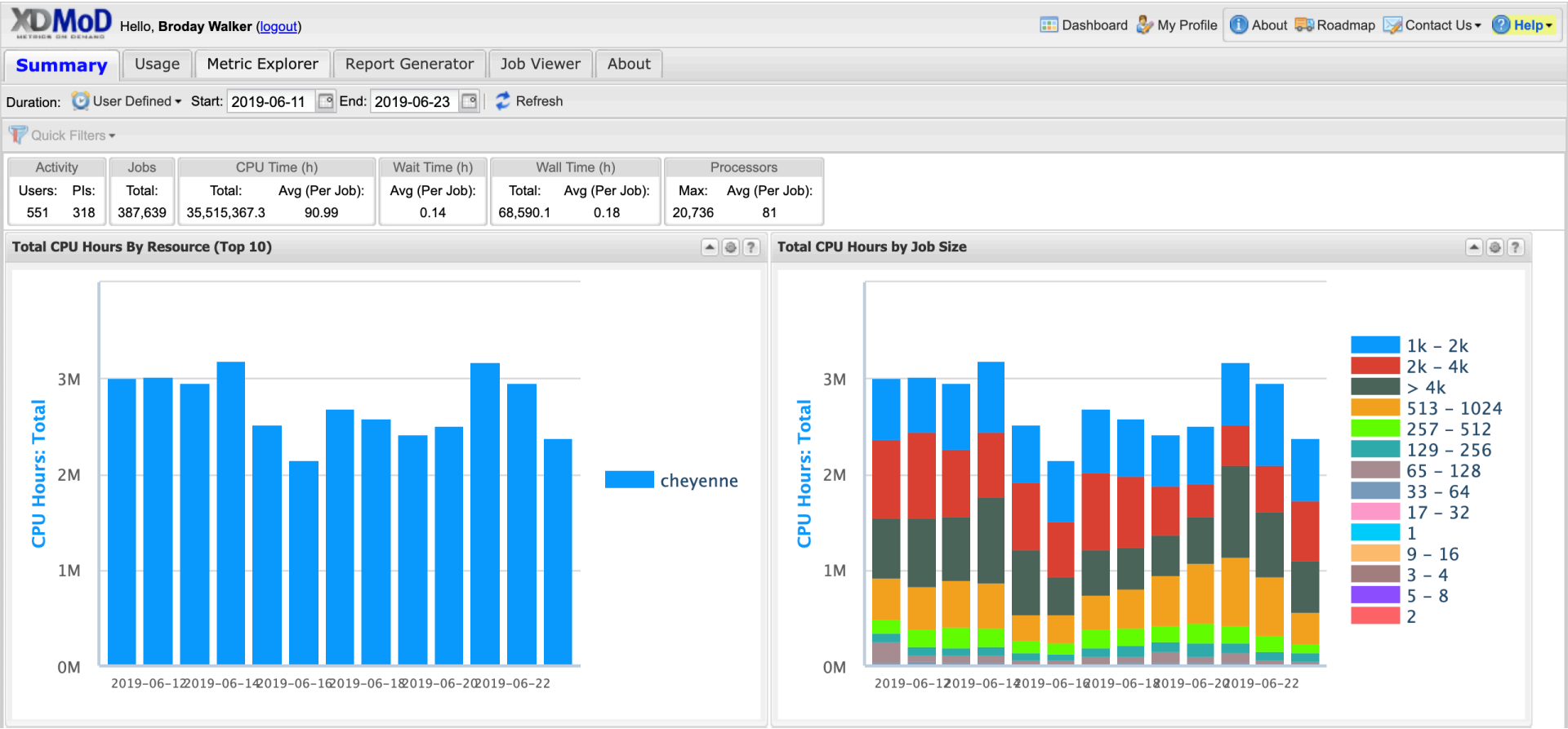
XDMoD

- An open source tool for HPC resource management
- Great for providing a general overview of resource usage
- Metrics include CPU hours consumed, average job wall time, and average wait time
- Metrics can be grouped by user or PI

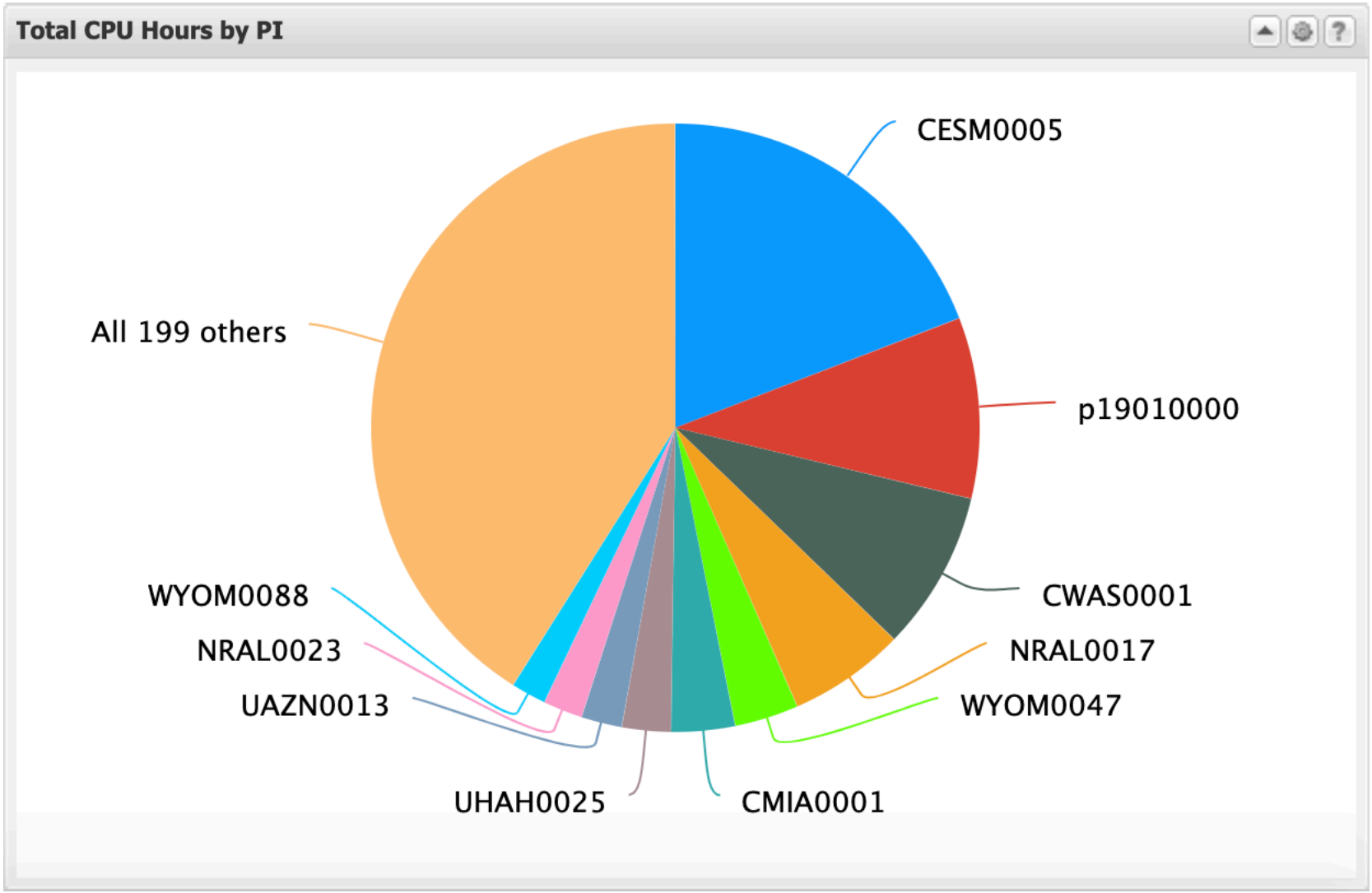
XDMoD Architecture



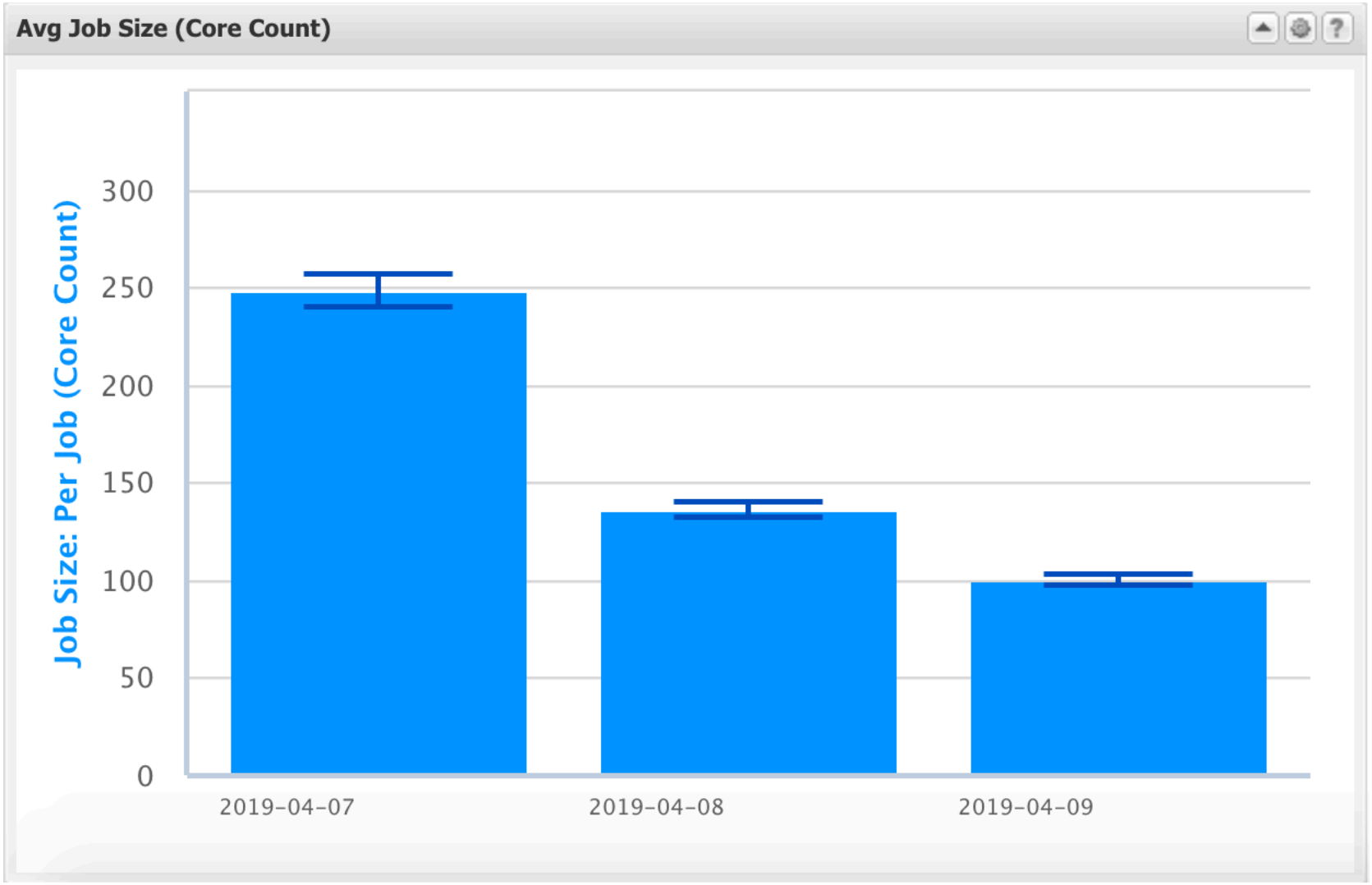
XDMoD Portal



XDMoD Portal



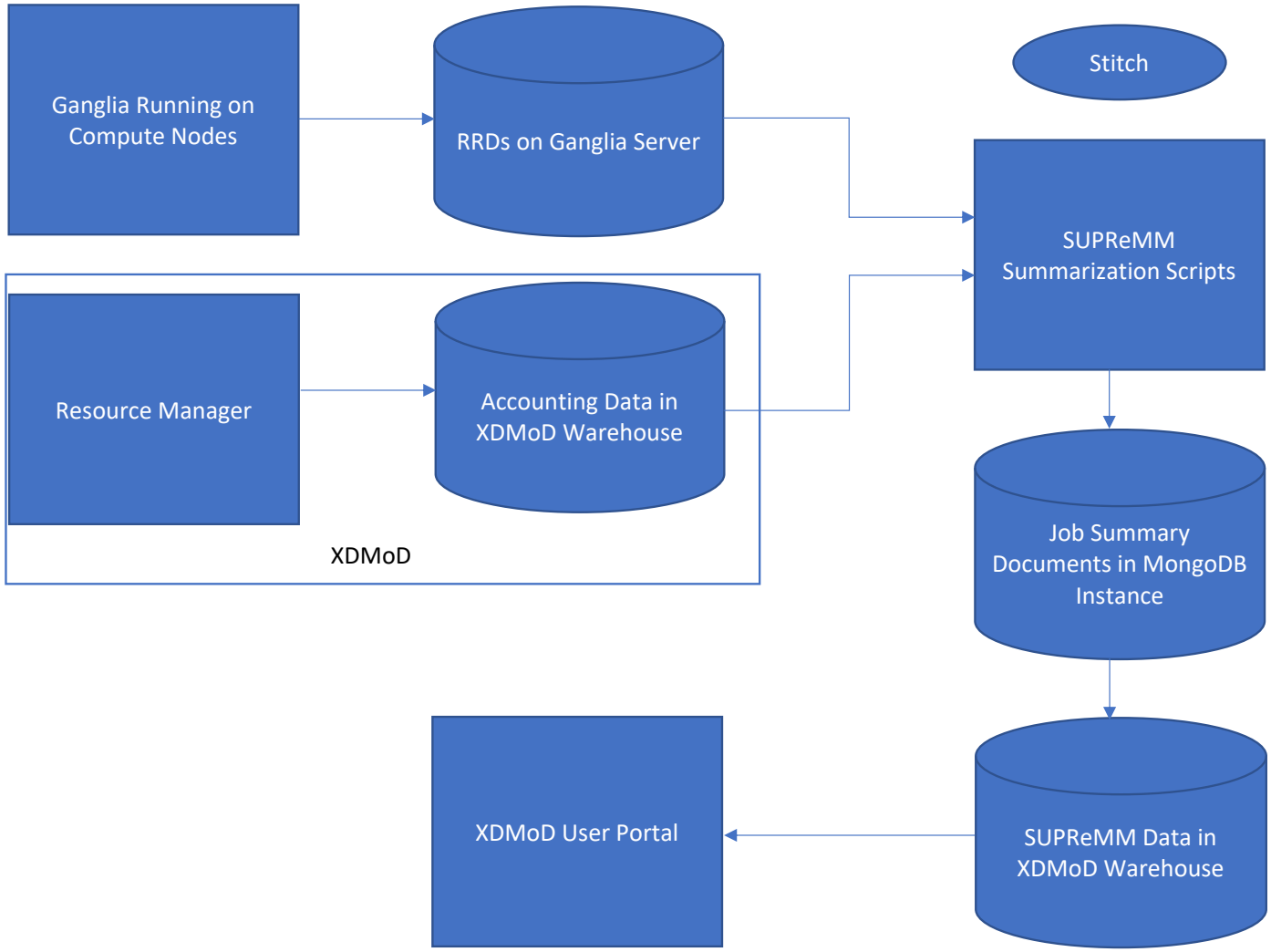
XDMoD Portal



SUPReMM

- An XDMoD module
- Provides job performance data
- Generates job-level data summarization from node-level data
- Raw node-level metrics: 70MB per node per day
- Job-level summary: 36KB per job
- XDMoD Data Warehouse storage: 2KB per job

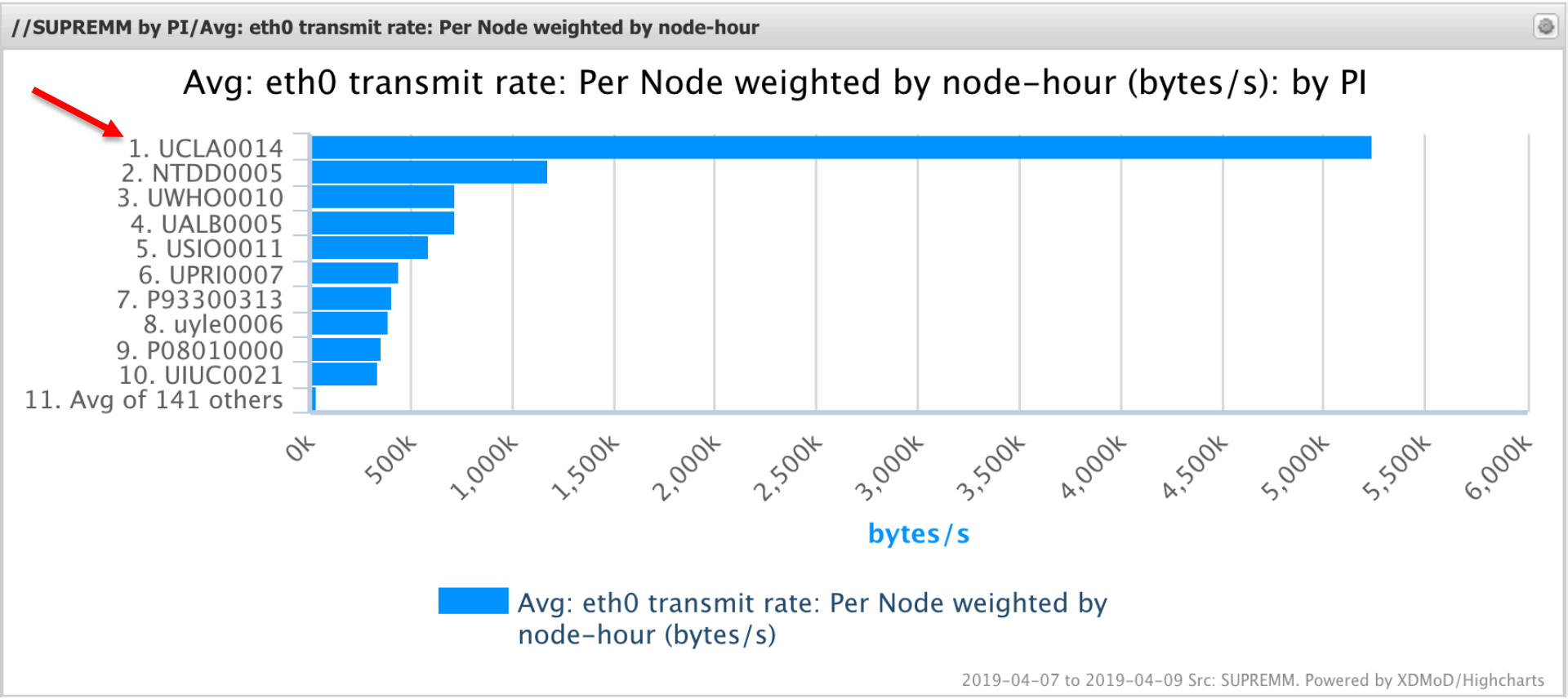
SUPReMM Architecture



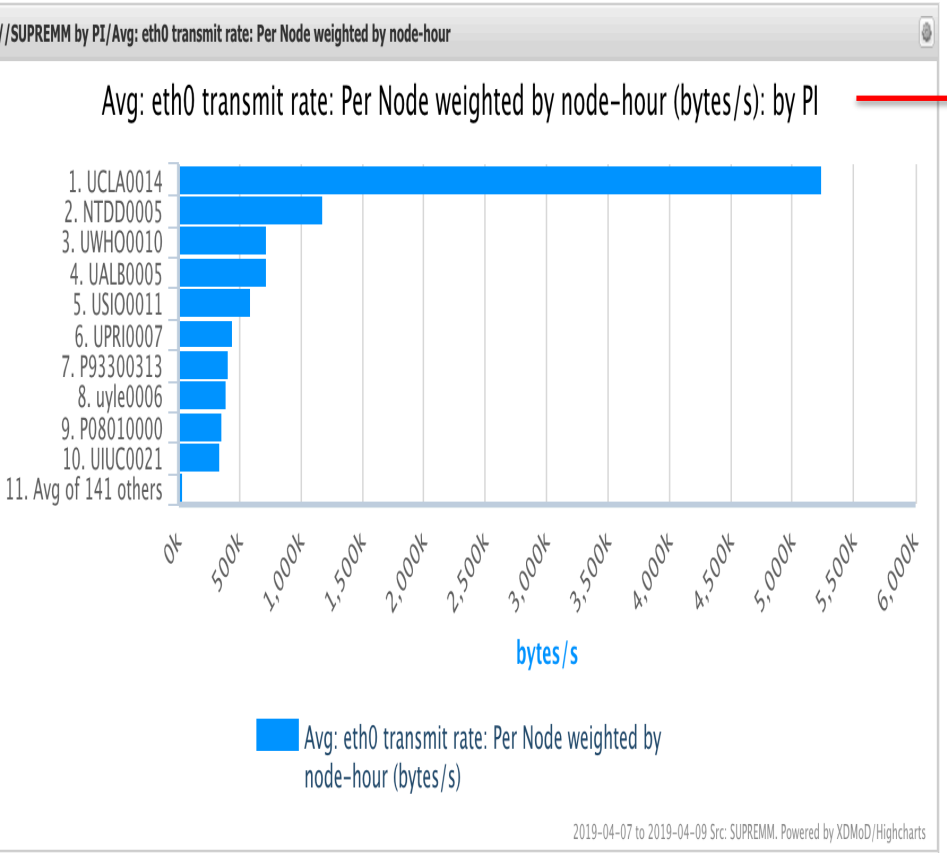
SUPReMM Stitch

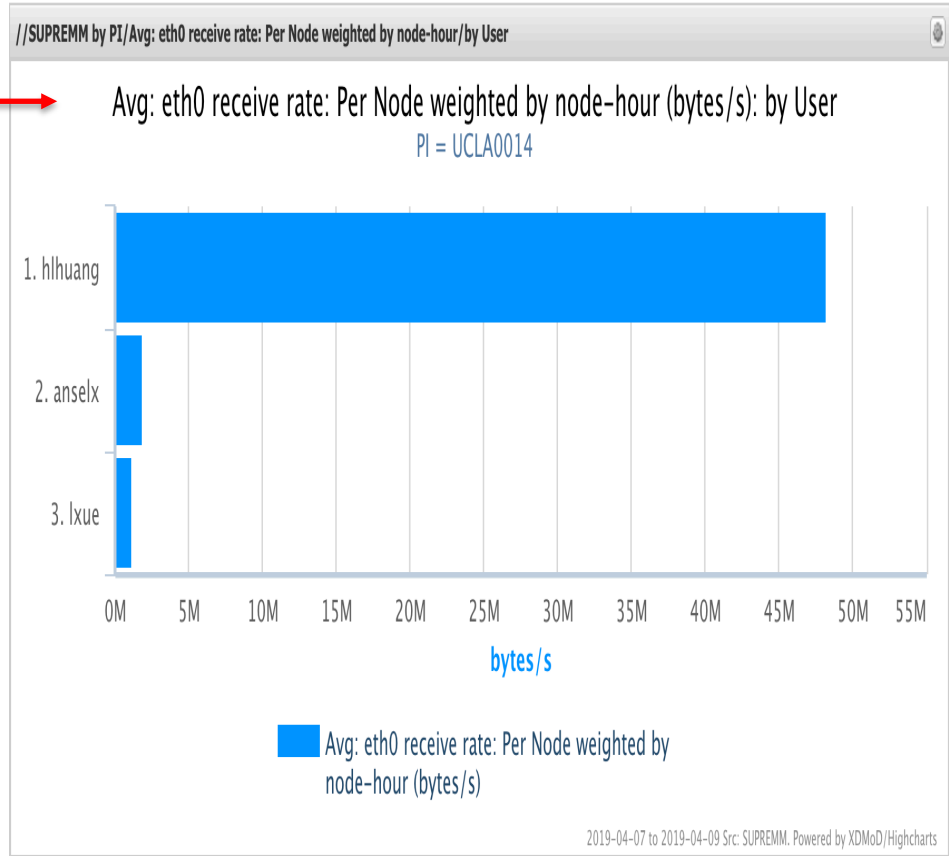
- Job performance data collected by Ganglia is stored as an RRD
- Using the accounting data from XDMoD, the node-level job performance data can be attributed to the appropriate job
- The result is a job performance document stored in a MongoDB instance

SUPReMM Drilldown



SUPReMM Drilldown





SUPReMM Job Analysis

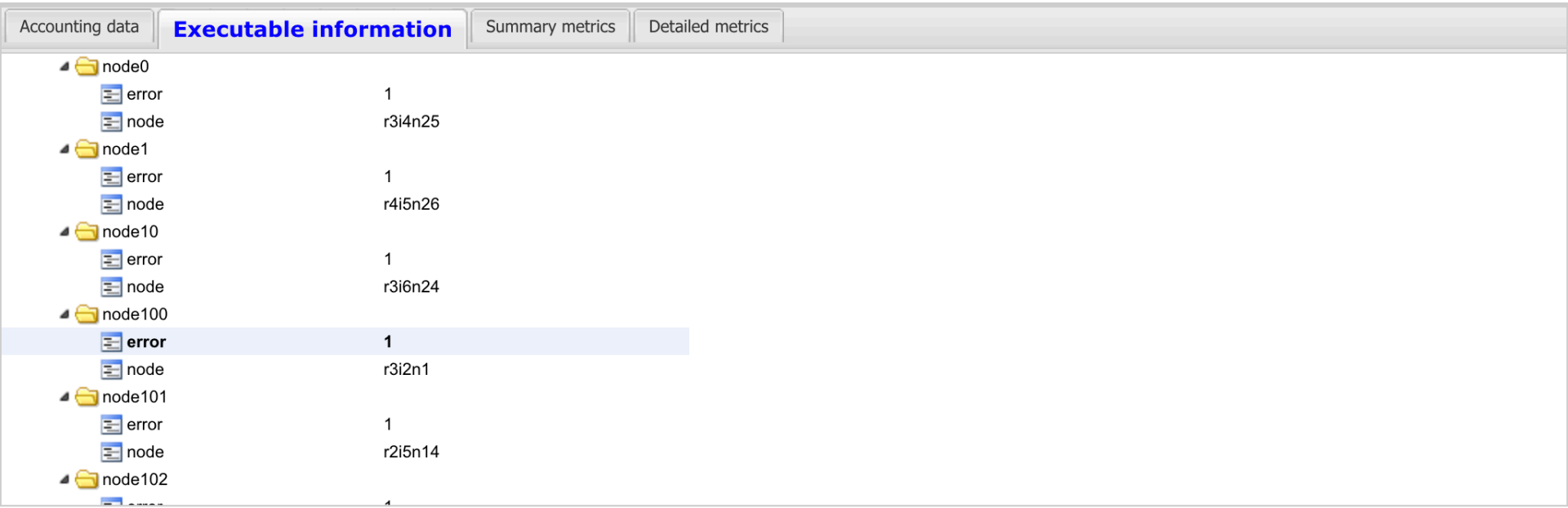
- Accounting Data and Executable Information
- Summary Metrics
- Detailed Metrics

Accounting Data

Accounting data | Executable information | Summary metrics | Detailed metrics

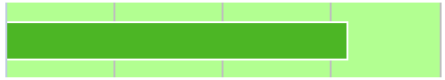
Key	Value
Working directory	NA
Category: Requested resource (3 Items)	
Requested Nodes	141
Requested Wall Time	12 hours 0.0 minute
Queue	economy
Category: Timing (6 Items)	
Wait Time	11 seconds
Wall Time	8 hours 45.7 minutes
Eligible Time	2019-04-09T09:57:39 UTC
End Time	2019-04-09T18:43:31 UTC
Start Time	2019-04-09T09:57:49 UTC
Submit Time	2019-04-09T09:57:38 UTC

Executable Information



SUPReMM Job Analysis

cheyenne-4954705

CPU User: 0.463	Homogeneity: N/A	CPU User Balance: 0.79	Memory Headroom: N/A
	Metric Missing: Not Available On The		Metric Missing: Unknown Reason

Accounting data | Executable information | **Summary metrics** | Detailed metrics

Key	Value
CPU Idle	53.04 %
CPU System	0.7033 %
CPU User	46.26 %
CPU User cov	0.0279
Node CPU idle	0.53
Category: Memory Statistics (4 Items)	
Memory Used	1.947 Tibytes
Total memory used	2.391 Tibytes
Memory Used Cov	0.0447
Total memory used cov	0.0373
Category: Network I/O Statistics (4 Items)	
Net Eth0 Rx	33.41 Gibytes
Net Eth0 Tx	396.9 Mibytes
Net Eth0 Rx Cov	0.0835
Net Eth0 Tx Cov	2.52

Detailed Metrics

Accounting data | Executable information | Summary metrics | **Detailed metrics**

Device	Average	Count	Standard Dev.	Median	Skew	Minimum	Maximum
cpu							
load1							
max	33.702334675476	137	0.8371062749722	33.918525695801	-3.9825466835109	27.117050170898	34.898342132568
maxpercore	0.46808798160383	137	0.011626476041281	0.4710906346639	-3.9825466835109	0.37662569681803	0.48469919628567
mean	32.425619848252	137	0.97378588851259	32.734918961158	-3.0722633090131	25.40191929157	33.432080562298
meanpercore	0.45035583122573	137	0.013524804007119	0.45465165223831	-3.0722633090131	0.35280443460513	0.46433445225414
network							
total							
in-bytes	35.87 Gbyte	137	2996289459.054	36.07 Gbyte	-11.185803353596	1.413 Gbyte	40.97 Gbyte
out-bytes	416.1 Mbyte	137	1048685729.1394	211.4 Mbyte	9.0623570606524	1.807 Mbyte	11.57 Gbyte
nodememory							
free	45.31 Gibyte	136	678664.02775055	45.29 Gibyte	1.9409255534116	44.05 Gibyte	48.79 Gibyte
maxfree	45.54 Gibyte	136	779391.86662608	45.48 Gibyte	0.85993356692252	44.15 Gibyte	48.81 Gibyte
maxused	17.59 Gibyte	136	717558.86376226	17.65 Gibyte	-2.2494212582031	13.94 Gibyte	18.68 Gibyte
maxused_minus_cache	17399851.198529	136	1696823.7764307	18229413.5	-0.5424019088541	13901418	21490732
physmem	62.68 Gibyte	136					
used	17.37 Gibyte	136	678664.02775055	17.39 Gibyte	-1.9409255534116	13.90 Gibyte	18.63 Gibyte
used_minus_cache	14828463.865102	136	662742.69648094	14842982.903846	-1.5800790636976	11429435.346154	16153626.115385
cores	72	137					

XDMoD/SUPReMM

- Easy to see drops in utilization by project ID, user, season
- Diagnose poorly performing jobs using detailed job performance data
- Better understanding of how users run jobs on Cheyenne

Future Work

- Improve RRD retrieval time
- Application Kernels – a proactive quality of service module

Accomplishments

- Upgraded XDMoD
- Delivered XDMoD/SUPReMM service
- Included new metrics (bytes_in, bytes_out, pkts_in, pkts_out)
- Fixed double quote bug in project ID
- Submitted poster to SC19
- Full update of internal wiki page

References

1. https://open.xdmod.org/8.1/
2. https://supremm.xdmod.org/8.1/supremm-architecture.html
3. https://wiki.ucar.edu/display/csg/test2+-+XDMoD+with+SUPReMM+Set+Up+Procedure
4. http://ganglia.sourceforge.net/

Acknowledgements

- University at Buffalo
- SSG
- CSG
- EIO
- SIParCS

Questions