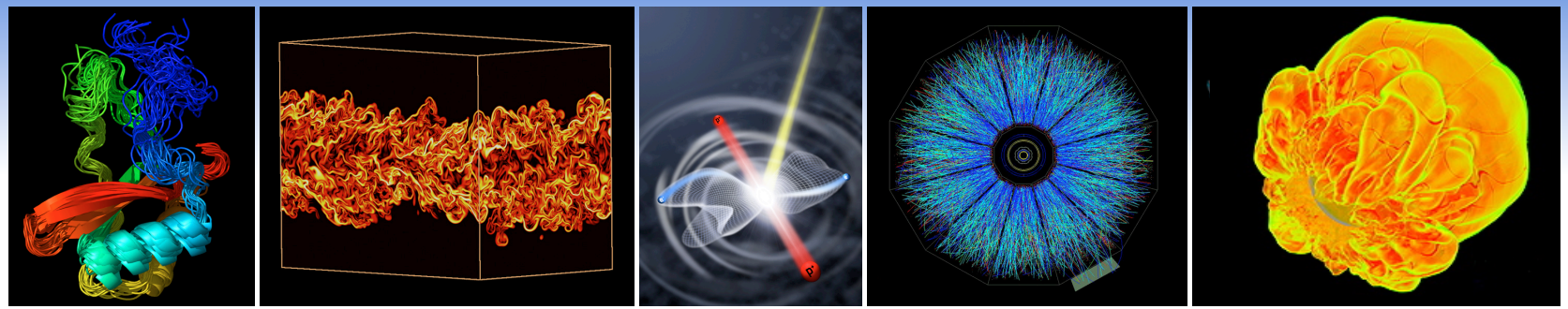


# Monitoring Lustre I/O on the Franklin Cray XT4


**Andrew Uselton**

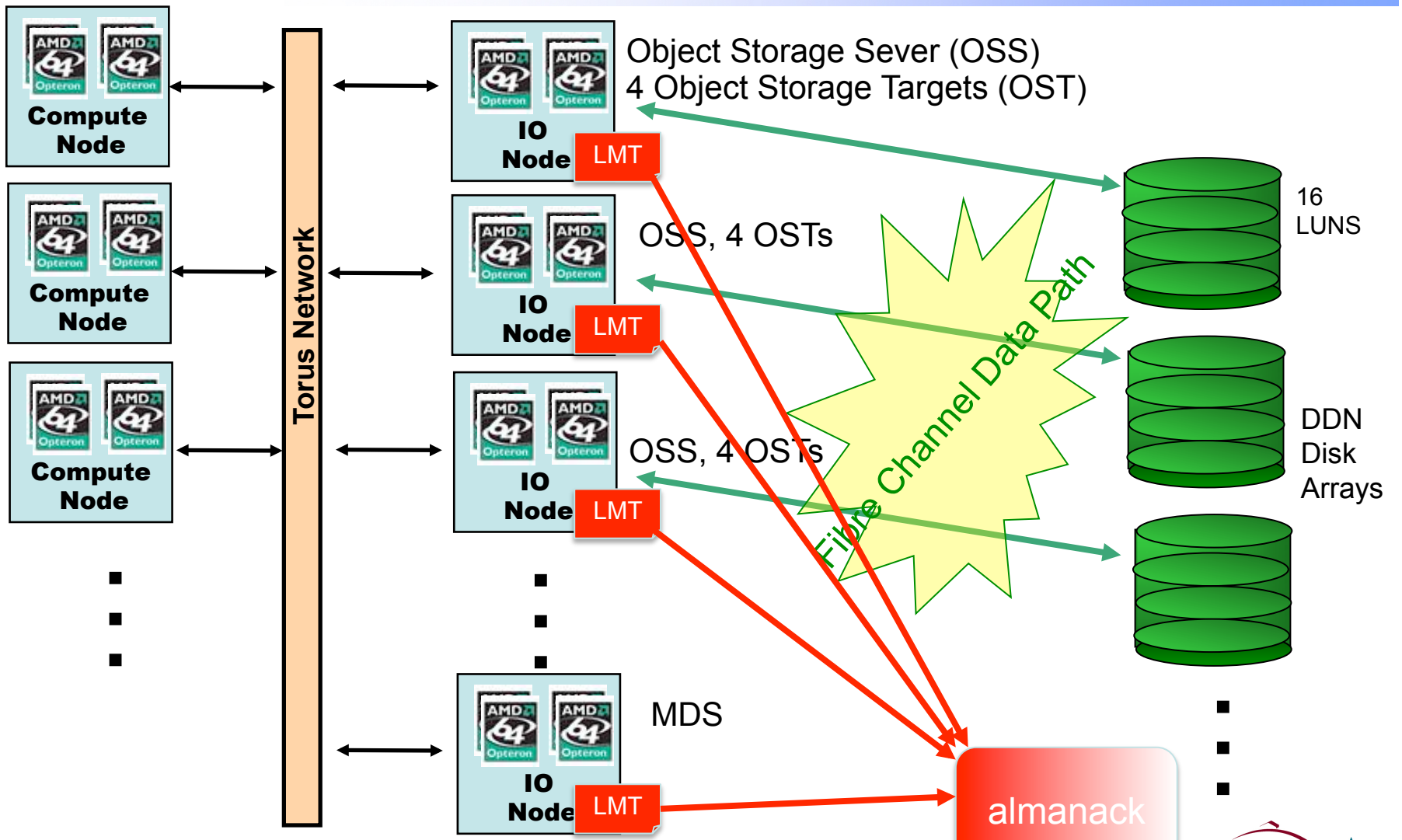
**NERSC**

**acuselton@lbl.gov**

# The Uses of Monitoring

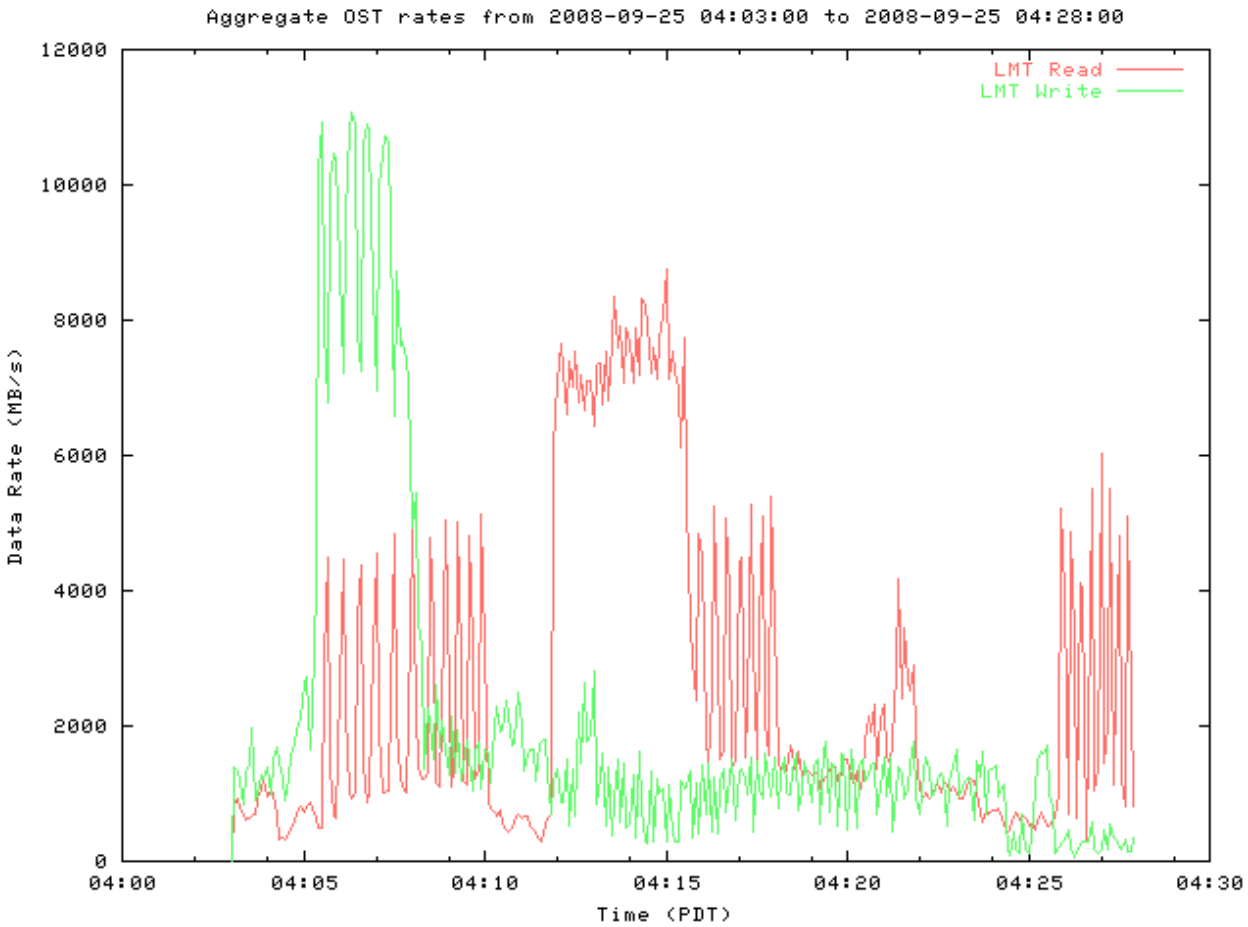
- After-the-fact incident investigation
- Checking the health of the system
- Noticing and exploring anomalous behavior
- Documenting use statistics
- System analysis and tuning

# The Franklin I/O Pipeline

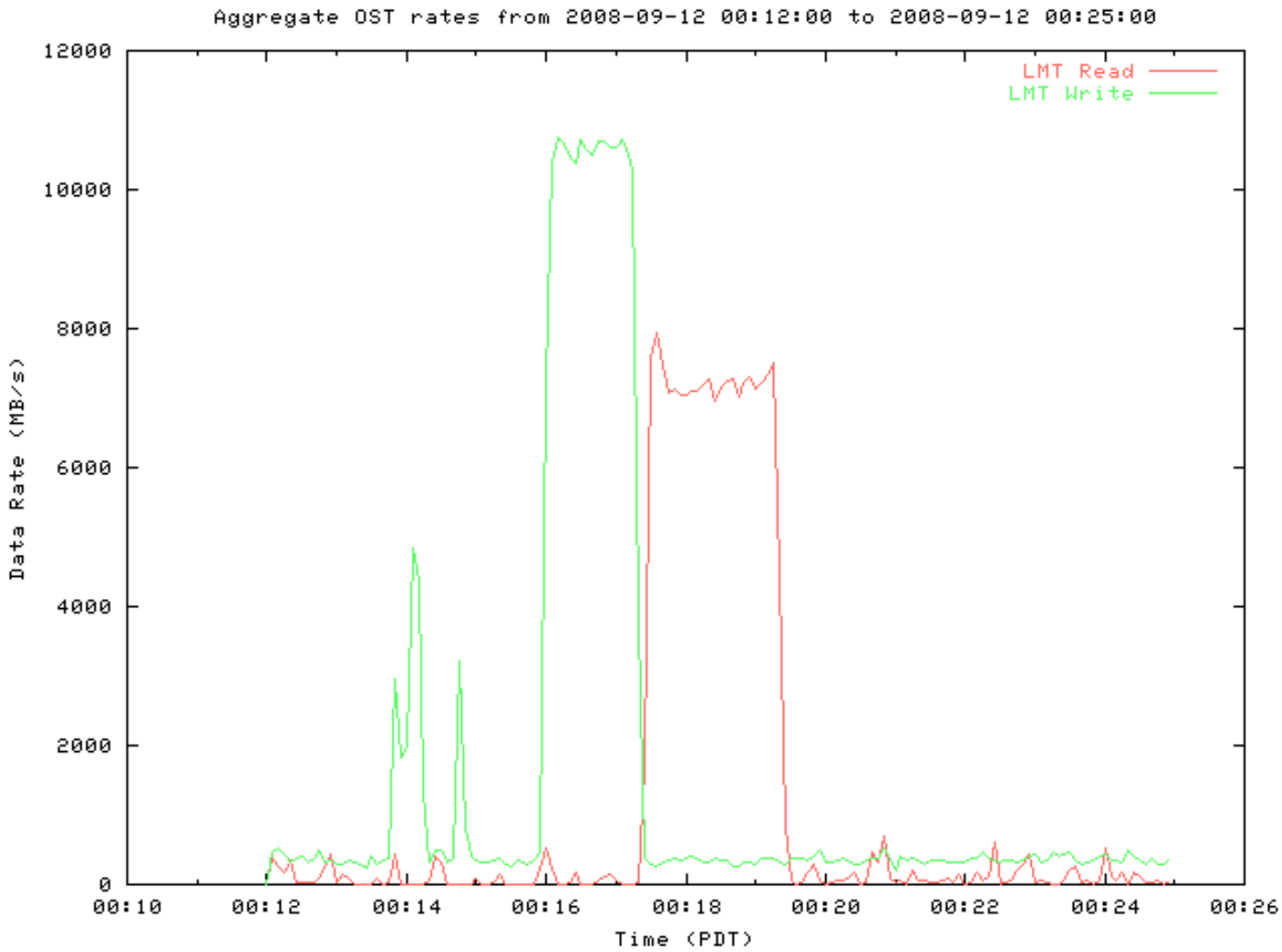



Monitoring I/O on Franklin

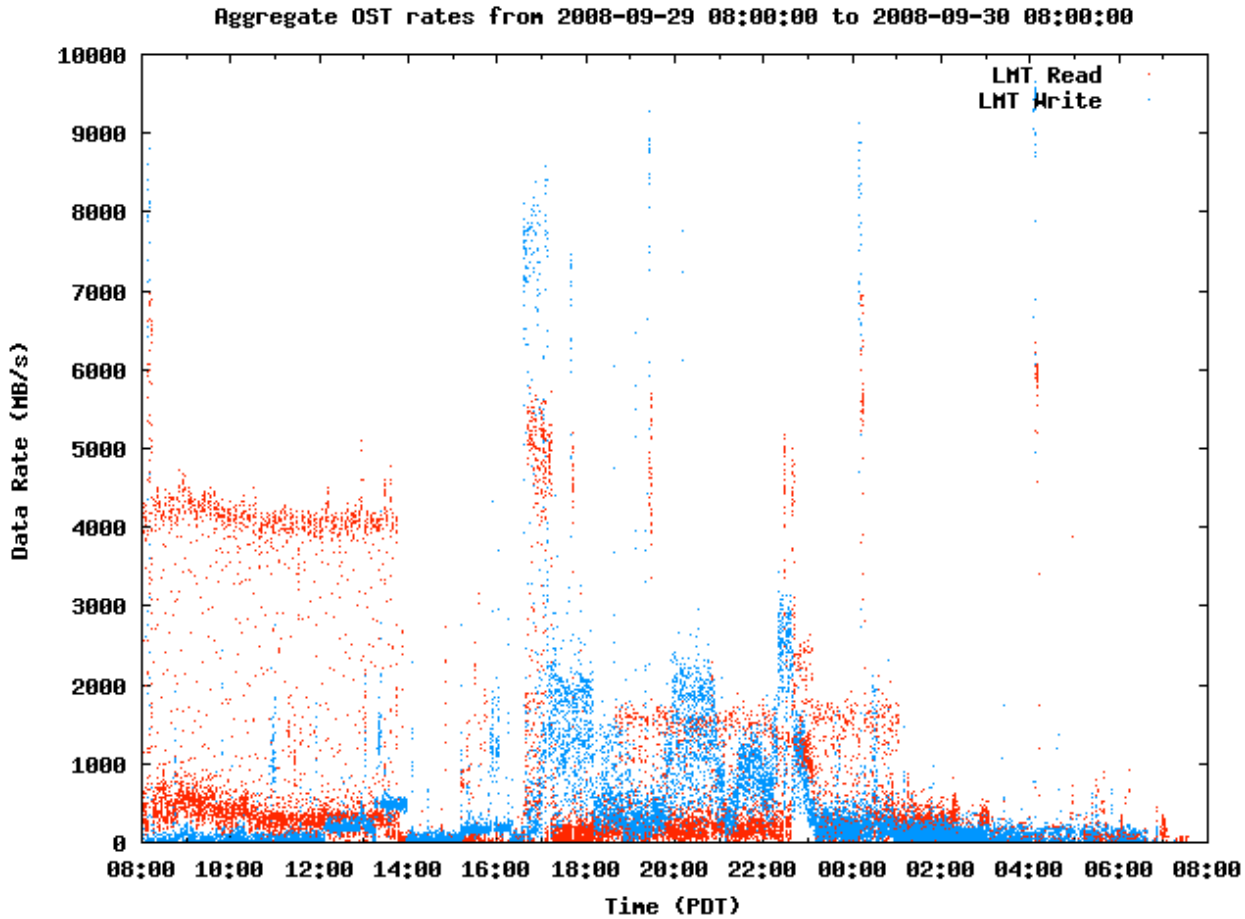
# The View from LMT



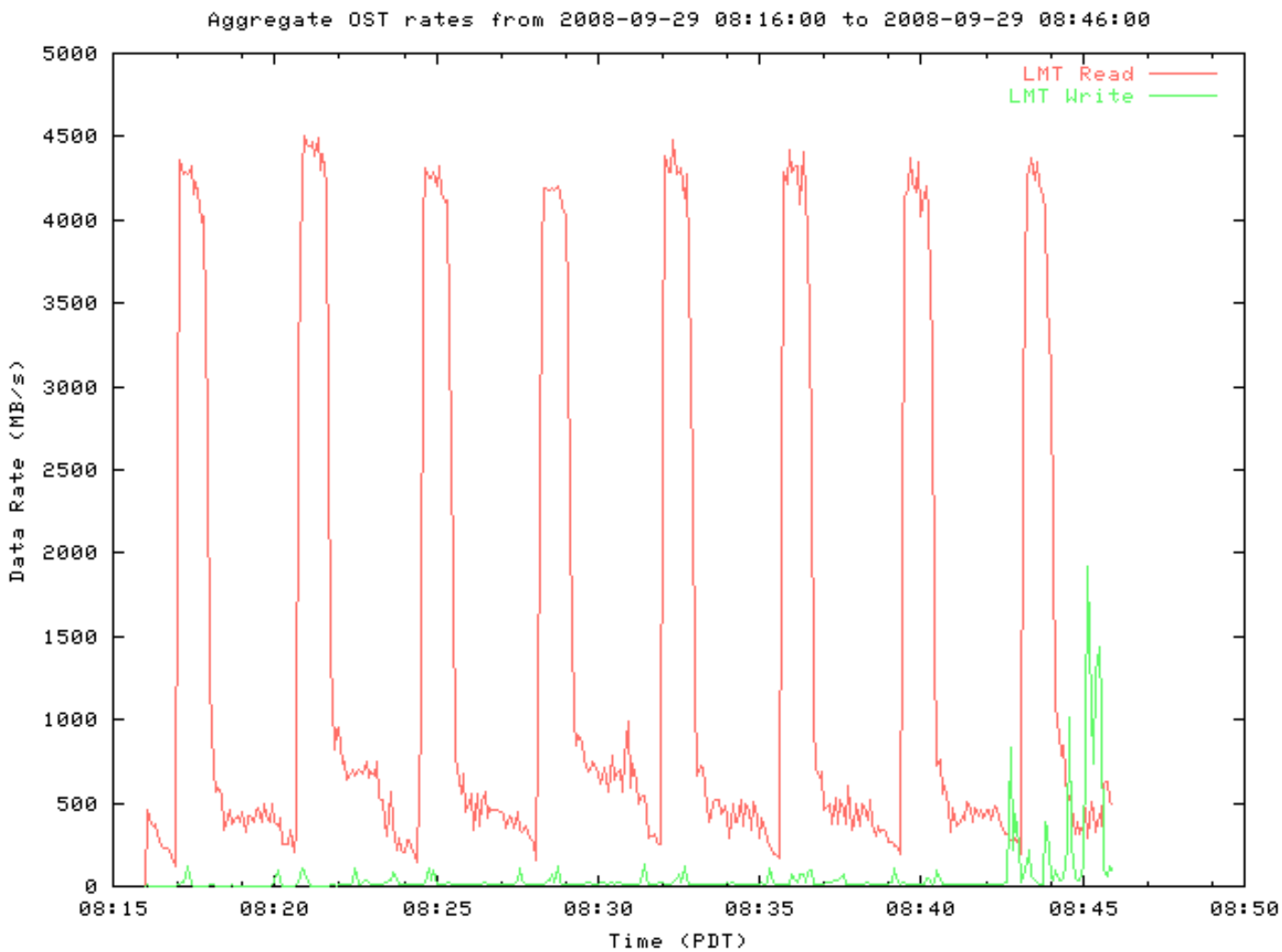
# A “Clean” IOR Run



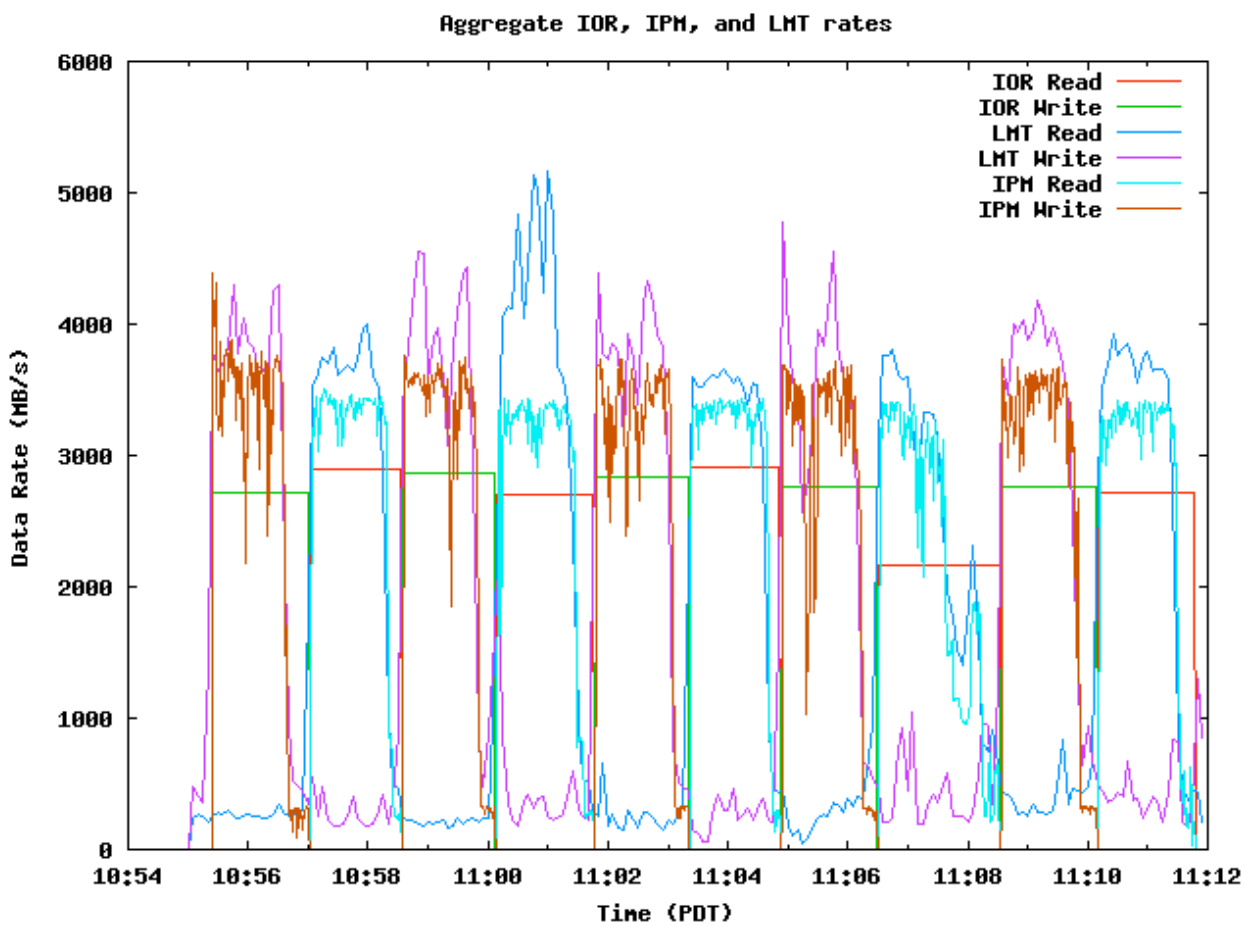
# 24 hours of I/O



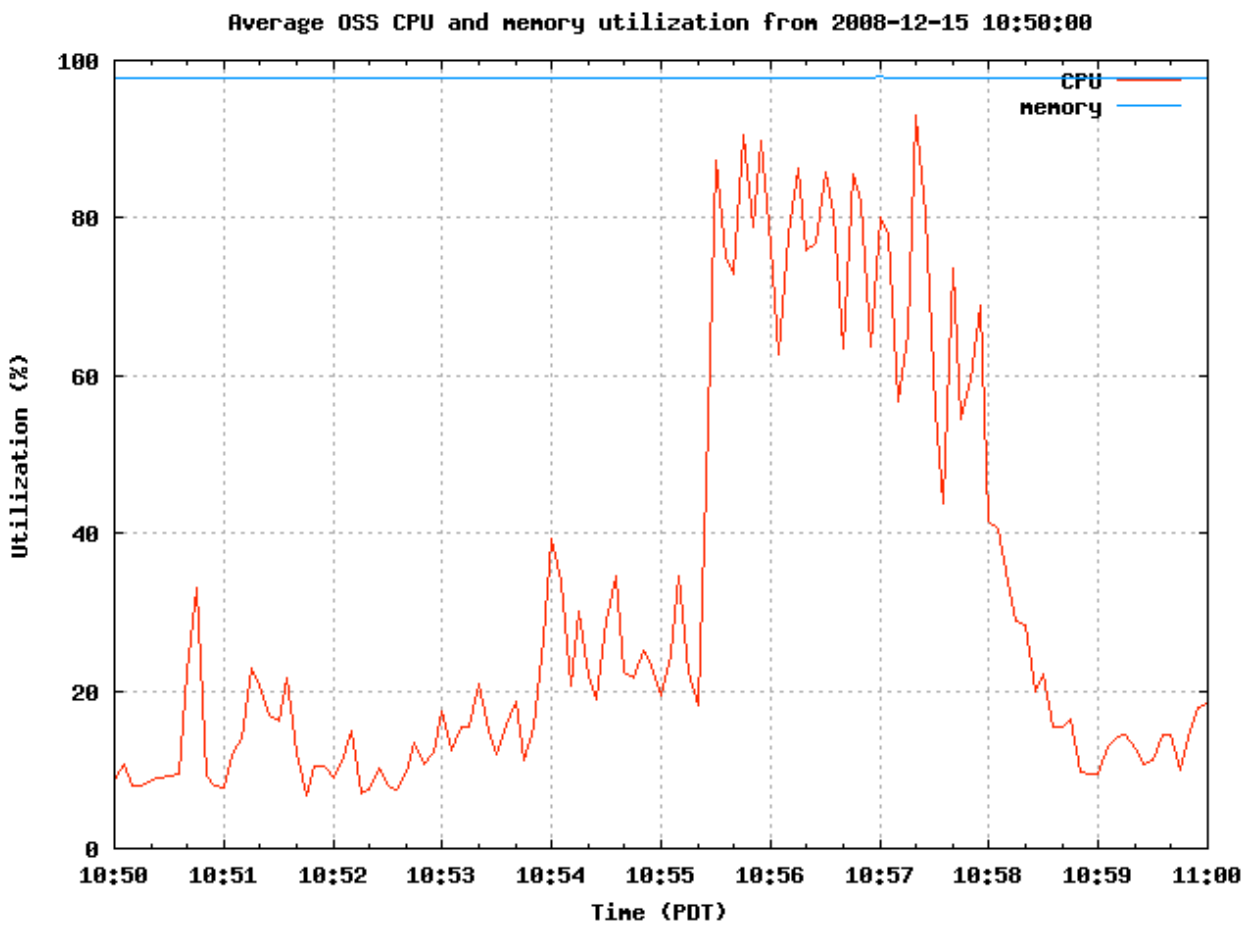
# Zooming in on the Reads



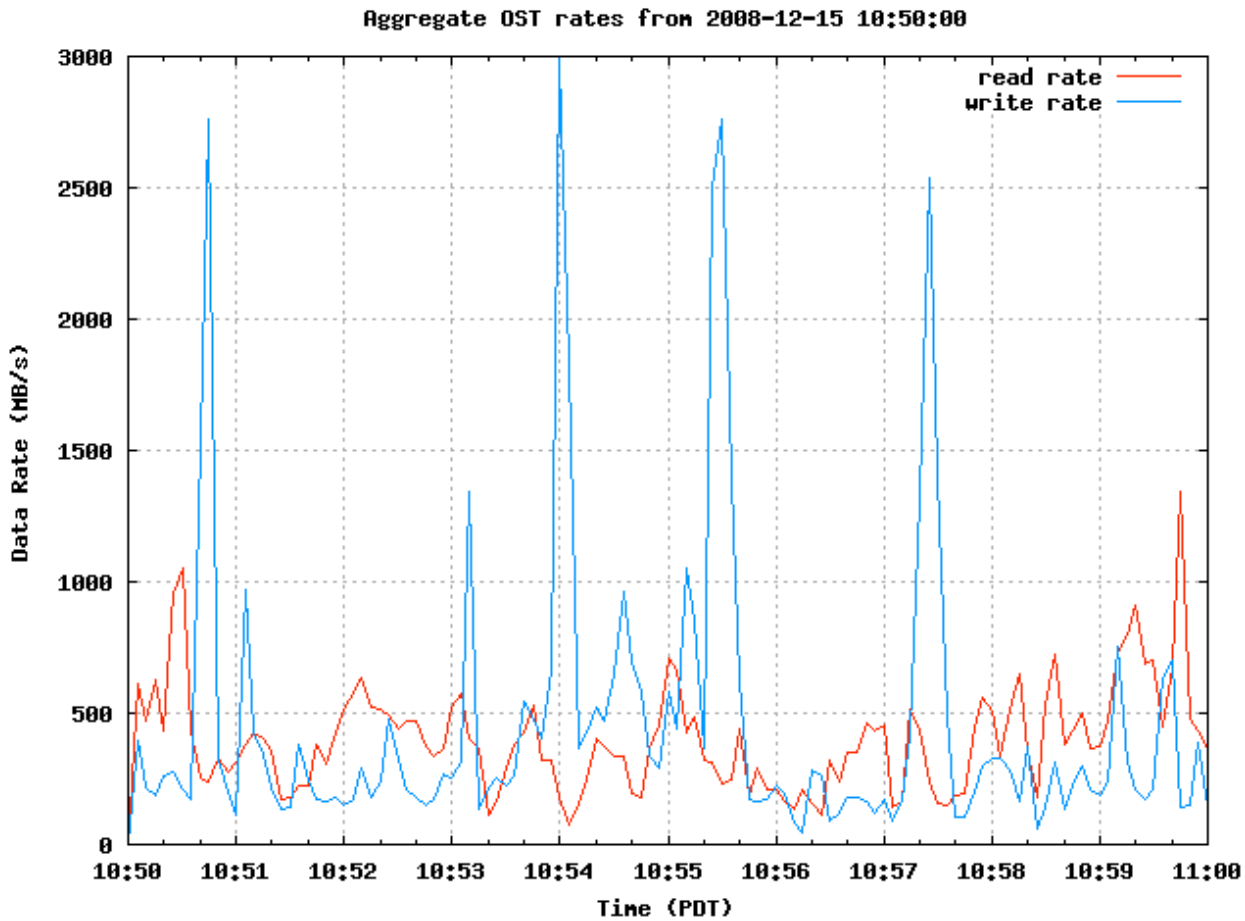
# LMT and IPM Illuminate and IOR



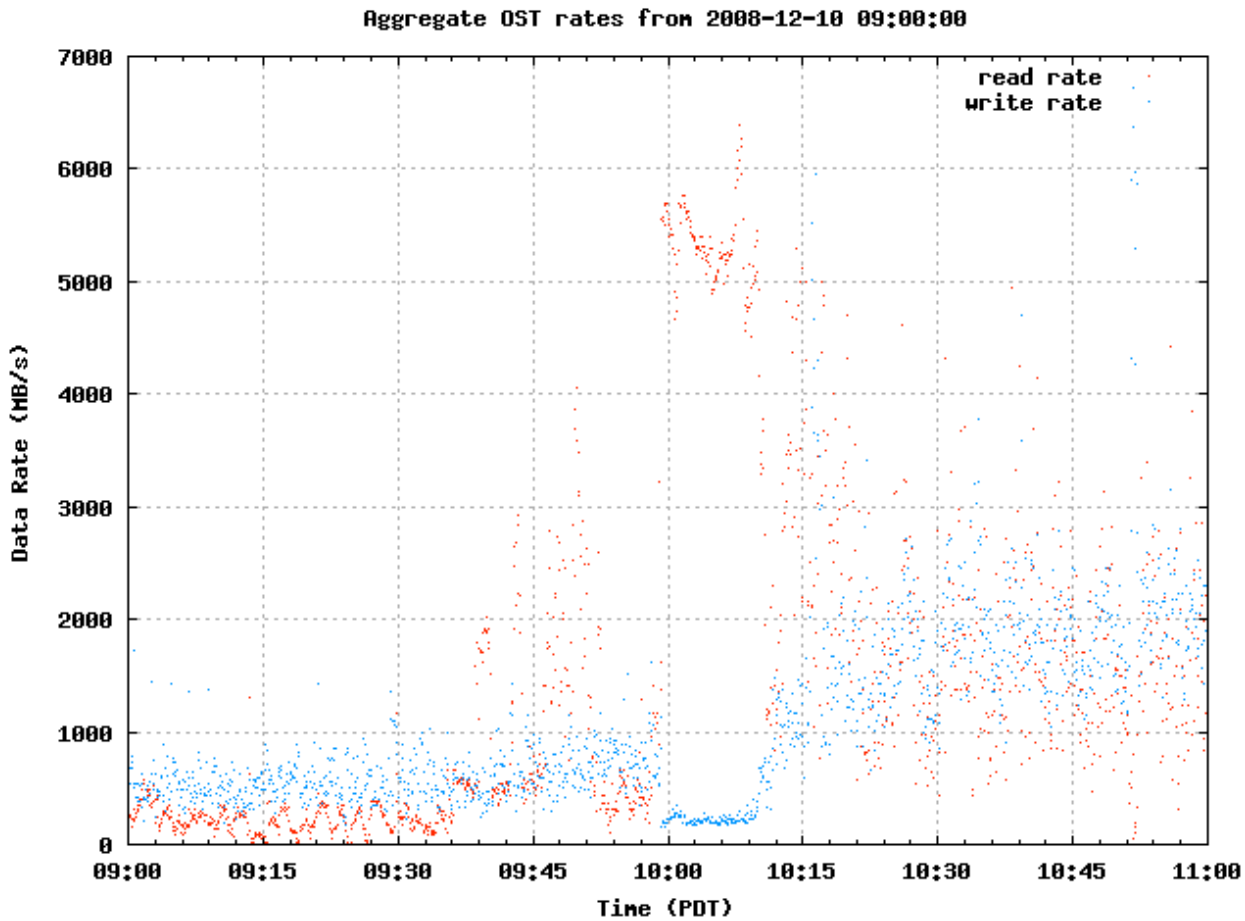
# Memory and CPU Utilization



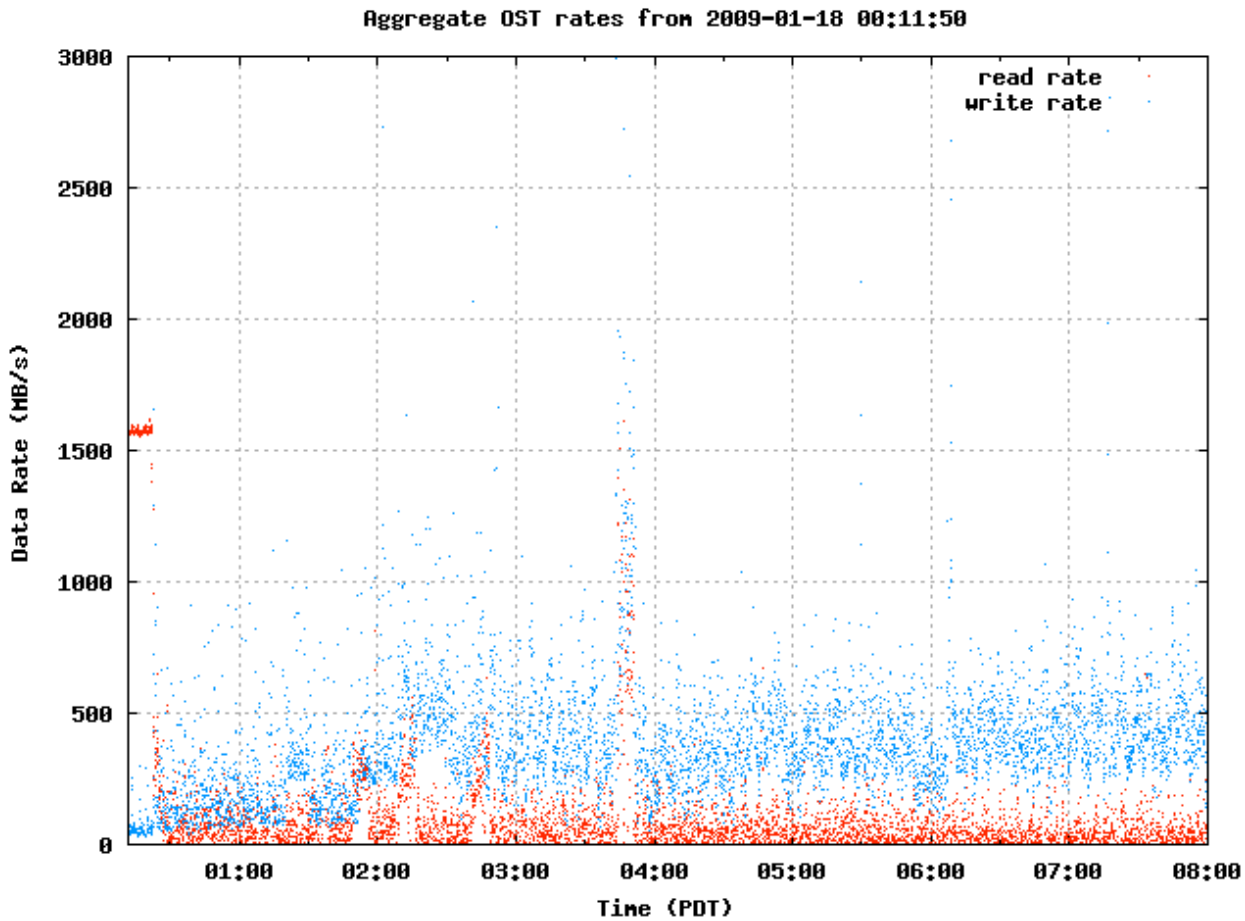
# The Corresponding Data Rates



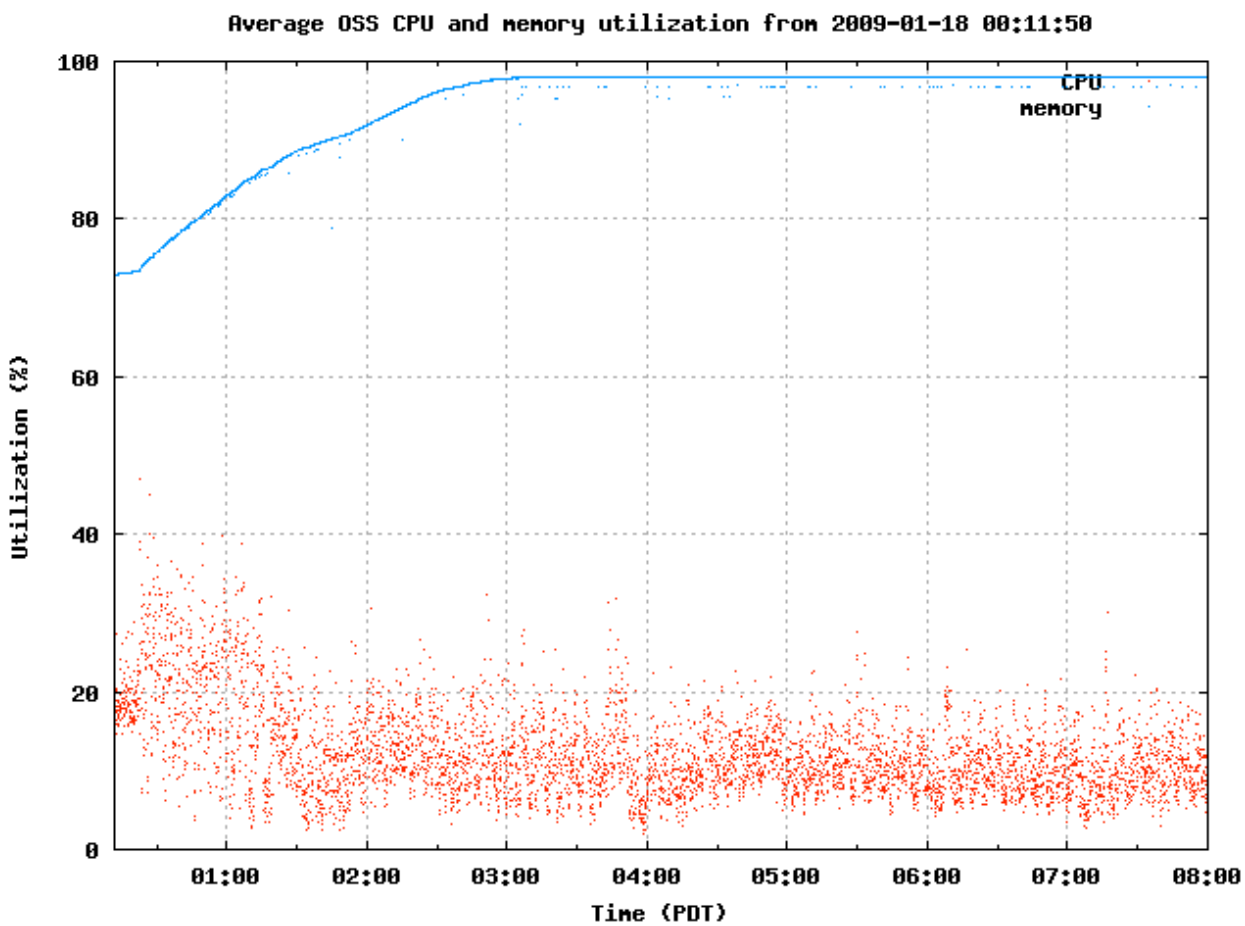
# Write Suppression



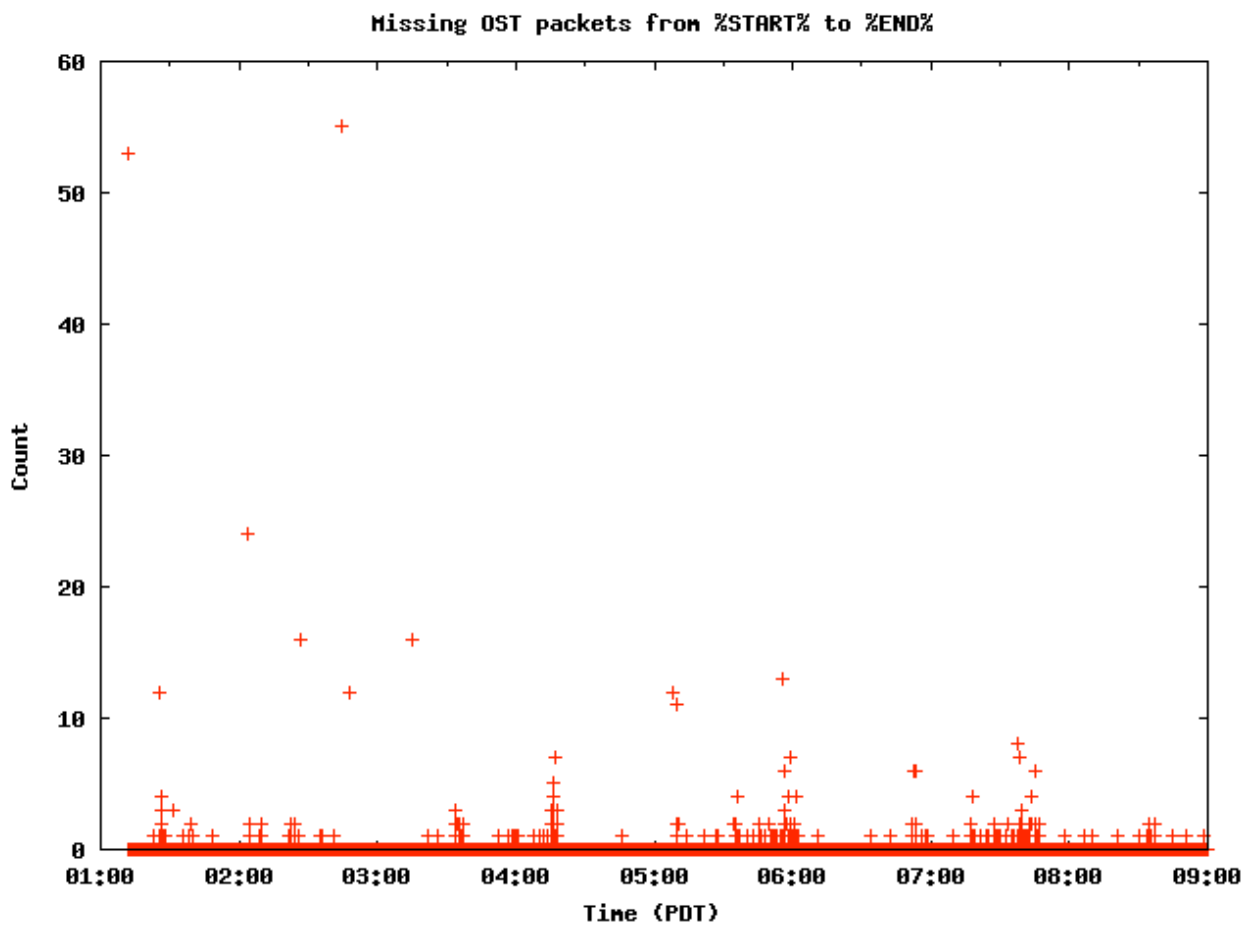
# A Recent Day


Aggregate OST rates from 2009-01-18 00:11:50
read rate
write rate
Data Rate (MB/s)
3000
2500
2000
1500
1000
500
0
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
Time (PDT)

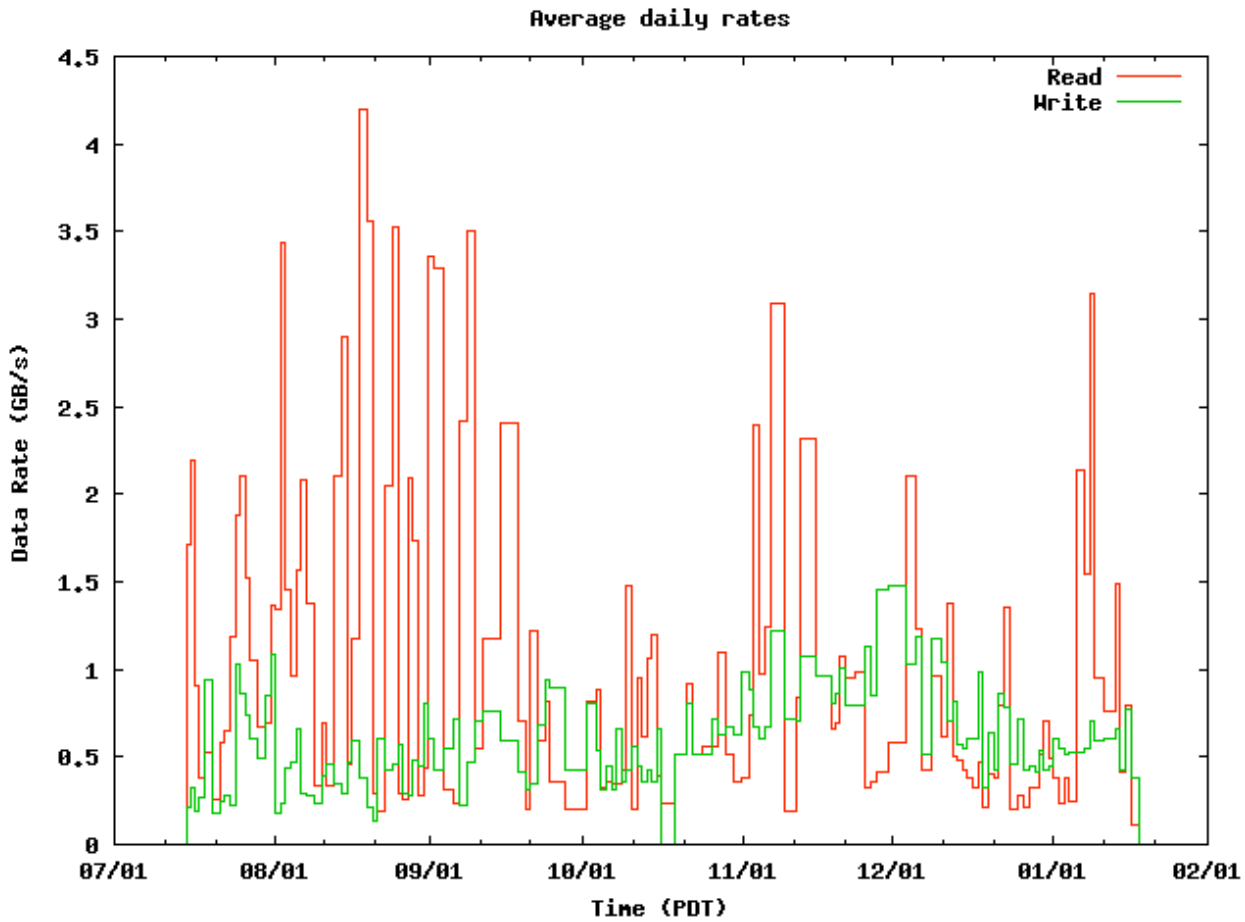
# The Corresponding CPU and Memory Utilization



# Incomplete Data



# History


Average daily rates
Read
Write
Data Rate (GB/s)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
07/01
08/01
09/01
10/01
11/01
12/01
01/01
02/01
Time (PDT)

# Useful Links

- **Franklin LMT web interface (may change)**

  http://www.nersc.gov/nusers/systems/franklin/lmt/

- **Franklin benchmarking page**

  https://www.nersc.gov/nusers/systems/franklin/monitor.php

- **LMT source code**

  http://sourceforge.net/projects/lmt/

- **Cerebro source code**

  http://sourceforge.net/projects/cerebro/

- **Additional slides**

# The LMT Dashboard

LWatch-lustre

File Configure

nid04116_mds | nid04140_mds | nid00032_mds

nid04116_mds 2008-09-30 14:57:10.0

| %CPU | %KB | %Inodes |
|---|---|---|
| 0.70 | 0.56 | 6.08 |

| Operation | Samples | Sample /Sec | Avg Valu |
|---|---|---|---|
| mds_close | 0 | 0.00 | 0.00 |
| mds_connect | 0 | 0.00 | 0.00 |
| mds_disconnect | 0 | 0.00 | 0.00 |
| mds_done_writing | 0 | 0.00 | 0.00 |
| mds_getattr | 0 | 0.00 | 0.00 |
| mds_getattr_lock | 0 | 0.00 | 0.00 |
| mds_getstatus | 0 | 0.00 | 0.00 |
| mds_getxattr | 0 | 0.00 | 0.00 |
| mds_pin | 0 | 0.00 | 0.00 |
| mds_quotacheck | 0 | 0.00 | 0.00 |
| mds_quotactl | 0 | 0.00 | 0.00 |
| mds_readpage | 0 | 0.00 | 0.00 |
| mds_reint | 70 | 14.00 | 286 |
| mds_set_info | 0 | 0.00 | 0.00 |
| mds_setxattr | 0 | 0.00 | 0.00 |
| mds_statfs | 0 | 0.00 | 0.00 |
| mds_sync | 0 | 0.00 | 0.00 |
| mds_unpin | 0 | 0.00 | 0.00 |
| obd_ping | 0 | 0.00 | 0.00 |

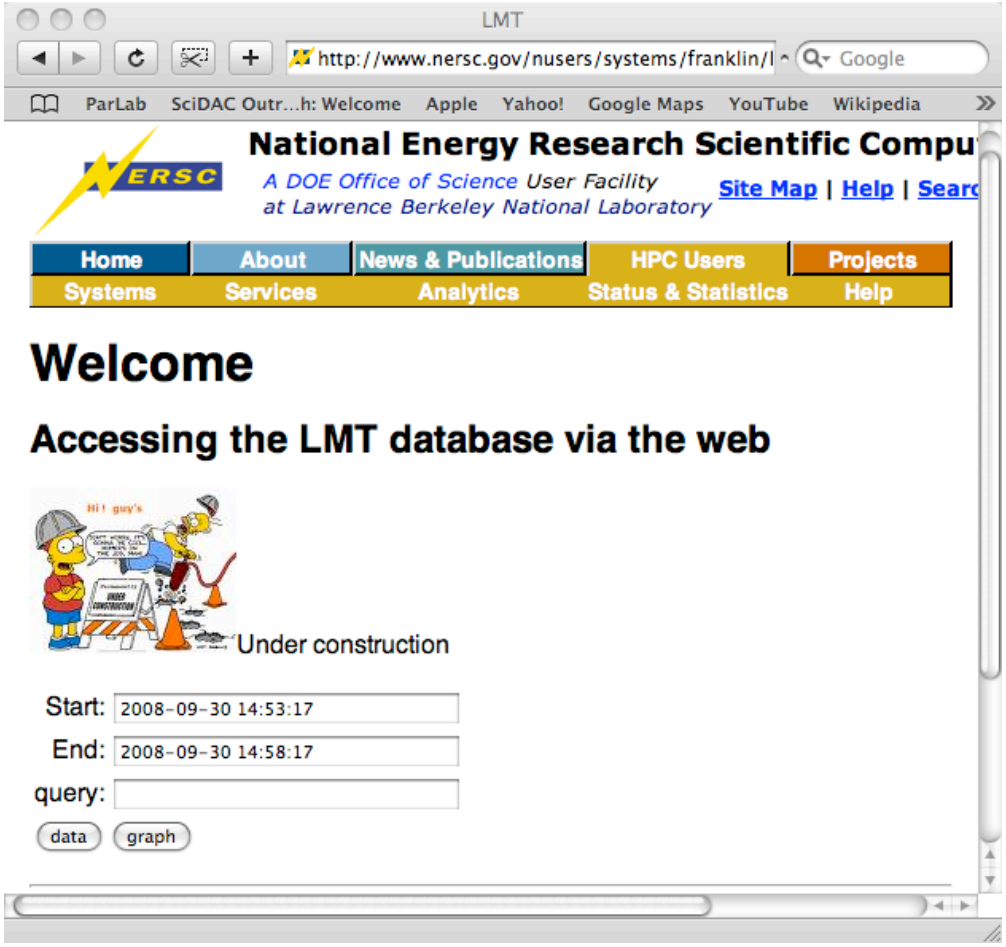
OST 2008-09-30 14:57:10.0

| Ost Name | Read Rate | Write Rate | %CF |
|---|---|---|---|
| ost16 | 0.10 | 2.00 | **** |
| ost8 | 0.40 | 0.31 | **** |
| ost18 | 0.10 | 0.80 | **** |
| ost5 | 0.23 | 0.21 | **** |
| ost15 | 0.15 | 1.40 | **** |
| ost7 | 0.35 | 0.95 | **** |
| ost17 | 0.50 | 2.53 | **** |
| ost9 | 0.25 | 1.72 | **** |
| ost19 | 0.40 | 0.80 | **** |
| ost46 | 1.70 | 109.44 | **** |
| ost56 | 0.58 | 0.00 | **** |
| ost48 | 1.05 | 0.40 | **** |
| ost58 | 0.50 | 0.40 | **** |
| ost45 | 2.06 | 126.70 | **** |
| ost55 | 0.20 | 0.88 | **** |
| ost47 | 1.09 | 81.44 | **** |
| ost57 | 0.20 | 0.60 | **** |
| ost49 | 0.30 | 0.41 | **** |
| ost59 | 0.05 | 0.48 | **** |
| AGGREGATE | 126.49 | 701.26 | **** |
| MAXIMUM | 26.25 | 126.70 | **** |
| MINIMUM | 0.00 | 0.00 | **** |
| AVERAGE | 1.58 | 8.77 | 0.0 |

OSS 2008-09-30 14:57:10.0

| Oss Name | Read Rate | Write Rate |
|---|---|---|
| nid00024 | 1.25 | 0.72 |
| nid00027 | 0.40 | 3.70 |
| nid00032 | 27.10 | 2.66 |
| nid00035 | 1.31 | 4.90 |
| nid00040 | 27.92 | 105.3 |
| nid00043 | 2.35 | 4.43 |
| nid04124 | 25.45 | 2.87 |
| nid04127 | 0.55 | 2.25 |
| nid04132 | 19.35 | 0.79 |
| nid04135 | 1.91 | 4.12 |
| nid00504 | 3.77 | 166.3 |
| nid00507 | 1.20 | 2.78 |
| nid00512 | 2.75 | 57.03 |
| nid00515 | 1.50 | 1.65 |
| nid04596 | 2.87 | 183.8 |
| nid04599 | 0.65 | 4.94 |
| nid04604 | 2.49 | 138.6 |
| nid04607 | 1.15 | 5.51 |
| nid04612 | 1.35 | 3.98 |
| nid04615 | 0.80 | 2.78 |
| AGGREGATE | 126.14 | 699.3 |
| MAXIMUM | 27.92 | 183.8 |
| MINIMUM | 0.40 | 0.72 |
| AVERAGE | 6.31 | 34.97 |

# Detail from the Dashboard

| | | |
|---|---|---|
| ost57 | 0.20 | 0.60 |
| ost49 | 0.30 | 0.41 |
| ost59 | 0.05 | 0.48 |
| AGGREGATE | 126.49 | 701.26 |
| MAXIMUM | 26.25 | 126.70 |
| MINIMUM | 0.00 | 0.00 |
| AVERAGE | 1.58 | 8.77 |

# Accessing LMT Data on the Web