OAK RIDGE National Laboratory

A Comprehensive Informative Metric for Analyzing HPC System Status using the LogSCAN Platform

Yawei Hui
Byung Hoon Park
Christian Engelmann

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Outline

- HPC logging at OLCF (Oak Ridge Leadership Computing Facility)
- LogSCAN - Log processing by Spark and Cassandra-based Analytics
- SIE - System Information Entropy
- Application

Titan at OLCF

Facts about Titan at OLCF:

- A Cray machine with 27 petaflops (peak performance)
- 18,688 AMD Opteron 16-core CPUs
- 18,688 NVidia Tesla K20X GPUs
- 693.5 TiB total memory
- 400 cabinets (of 25 by 8 physical floor layout)
- 3 chasses per cabinet; 8 slots per chassis; 4 nodes per slot;

System Logs of Titan

Titan logs are recorded in many categories:

- Application
- Console
- Consumer
- Netwatch
- Etc.

Console Log

Time Stamp | Source ID | Record Type | Log Message

```
2017-02-20T00:04:22.994827-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:22.994840-05:00 c6-7c2s7n1 Error: krca_add_evqueue out of mempool (evt)
2017-02-20T00:04:24.443771-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:28.976757-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:28.976834-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:30.464027-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:35.044149-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:35.044211-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:36.474531-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:40.984901-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:40.984958-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:42.495131-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:46.966215-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:46.966270-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:48.525635-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:52.985832-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:52.985890-05:00 c6-7c2s7n1 Error: krca_add_evqueue out of mempool (evt)
2017-02-20T00:04:54.466204-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:59.256389-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:04:59.256445-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:00.476375-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:01.969841-05:00 c1-1c0s4n3 [Hardware Error]: CPU 8: Machine Check Exception: 0 Bank 4: dc4a4090001c017b
2017-02-20T00:05:01.969893-05:00 c1-1c0s4n3 [Hardware Error]: TSC 0 ADDR 401a4a580 MISC c01a000001000000
2017-02-20T00:05:01.969911-05:00 c1-1c0s4n3 [Hardware Error]: PROCESSOR 2:600f12 TIME 1487567101 SOCKET 0 APIC 8 microcode 0
2017-02-20T00:05:03.139137-05:00 c3-3c1s0n3 [Hardware Error]: CPU 0: Machine Check Exception: 0 Bank 4: dc284540011c010b
2017-02-20T00:05:03.139194-05:00 c3-3c1s0n3 [Hardware Error]: TSC 0 ADDR c84 MISC c01a000001000000
2017-02-20T00:05:03.139203-05:00 c3-3c1s0n3 [Hardware Error]: PROCESSOR 2:600f12 TIME 1487567101 SOCKET 0 APIC 0 microcode 0
2017-02-20T00:05:03.294305-05:00 c13-3c0s7n1 [Hardware Error]: CPU 8: Machine Check Exception: 0 Bank 4: dc1dc0e0001f010b
2017-02-20T00:05:03.294329-05:00 c13-3c0s7n1 [Hardware Error]: TSC 0 ADDR be00 MISC c01a000001000000
2017-02-20T00:05:03.294337-05:00 c13-3c0s7n1 [Hardware Error]: PROCESSOR 2:600f12 TIME 1487567101 SOCKET 0 APIC 8 microcode 0
2017-02-20T00:05:05.066755-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:05.066807-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:06.507334-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:08.387147-05:00 c7-2c2s0n0 LNetError: 26041:0:(o2iblnd_cb.c:1110:kiblnd_init_rdma()) RDMA has too many fragments
2017-02-20T00:05:08.387202-05:00 c7-2c2s0n0 LNetError: 26041:0:(o2iblnd_cb.c:1110:kiblnd_init_rdma()) Skipped 1 previous similar m
2017-02-20T00:05:08.387211-05:00 c7-2c2s0n0 LNetError: 26041:0:(o2iblnd_cb.c:432:kiblnd_handle_rx()) Can't setup rdma for PUT to l
2017-02-20T00:05:08.387240-05:00 c7-2c2s0n0 LNetError: 26041:0:(o2iblnd_cb.c:432:kiblnd_handle_rx()) Skipped 1 previous similar me
2017-02-20T00:05:11.167921-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:11.167985-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:12.507938-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:17.097821-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:17.097894-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:18.517881-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:23.118926-05:00 c6-7c2s7n1 NVRM: os_schedule: Attempted to yield the CPU while in atomic or interrupt context
2017-02-20T00:05:23.118983-05:00 c6-7c2s7n1 INFO: rcu_sched_state detected stall on CPU 4 (t=15000 jiffies)
```

Parsing Logs

Counts	Percentage	ID	Description
10	0	1	DVS Confusion
2,998,492	2	2	NVRM Xid
5,671,348	4	3	Machine Check Exception (MCE)
1,229	0	4	NVRM DBE
49	0	5	Unknown GPU Error (UGE)
302,969	0.2	6	Graphics Engine Error (GEE)
5,732	0	7	Kernel Panic
782,337	0.5	8	Out of Memory (OOM)
16,938,194	11.6	9	HWERR
1,215,780	0.8	10	Seg. Fault
43,268,141	29.5	11	Lustre
31,498,746	21.5	12	LNet
992,997	0.7	13	LNet Error
42,809,426	29.2	22	Lustre Error
146,485,450	100		

Table 1: The non-zero event types and their occurrences in Titan's logs between January 2015 to March 2018.

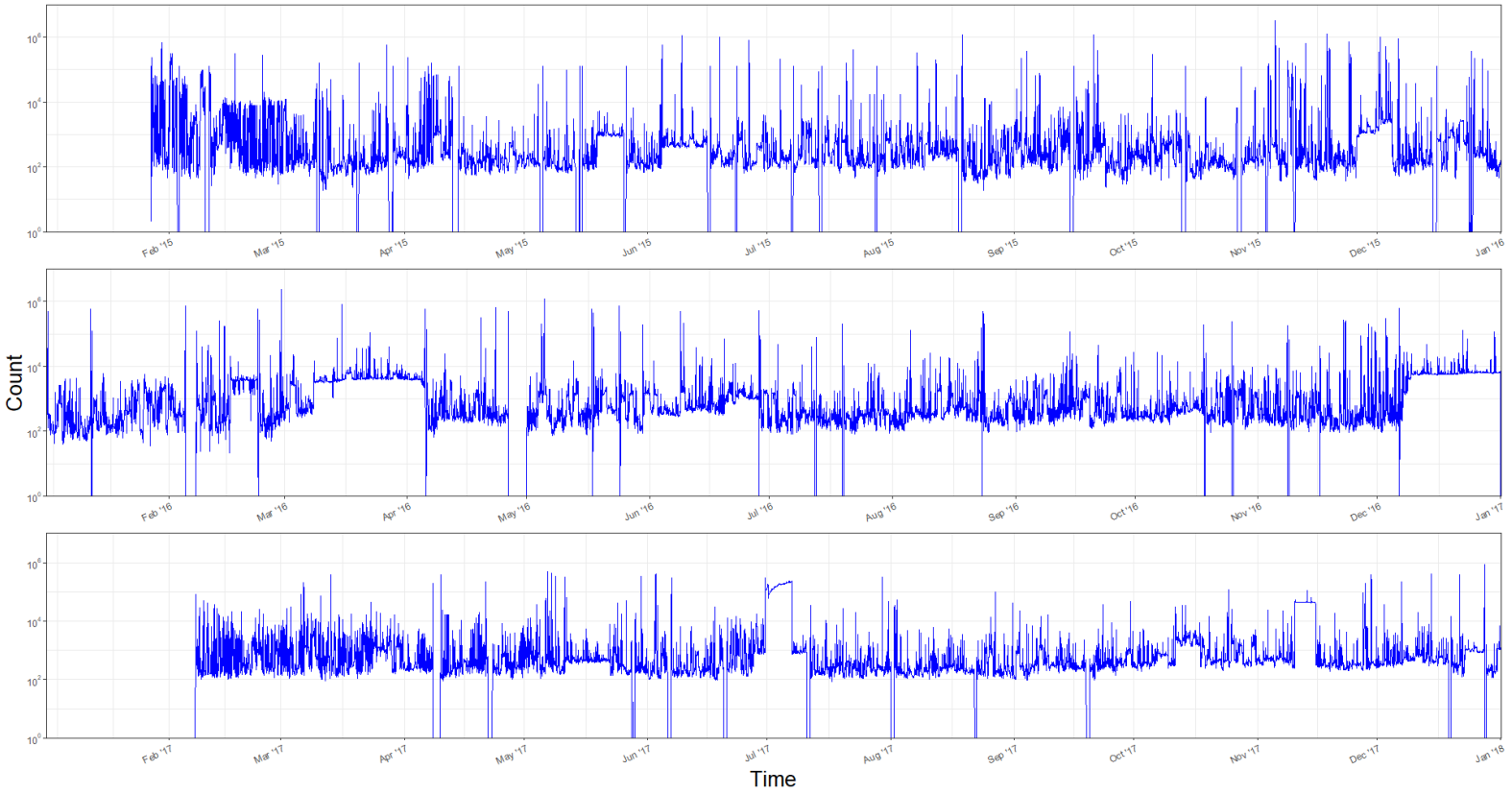


Figure 1: The total event counts accumulated in calendar year 2015, 2016 and 2017. The resolution in time is by hour, which means the count number plotted at any given time is the total counts recorded on all Titan nodes within the past hour.

Challenge in Event Analysis

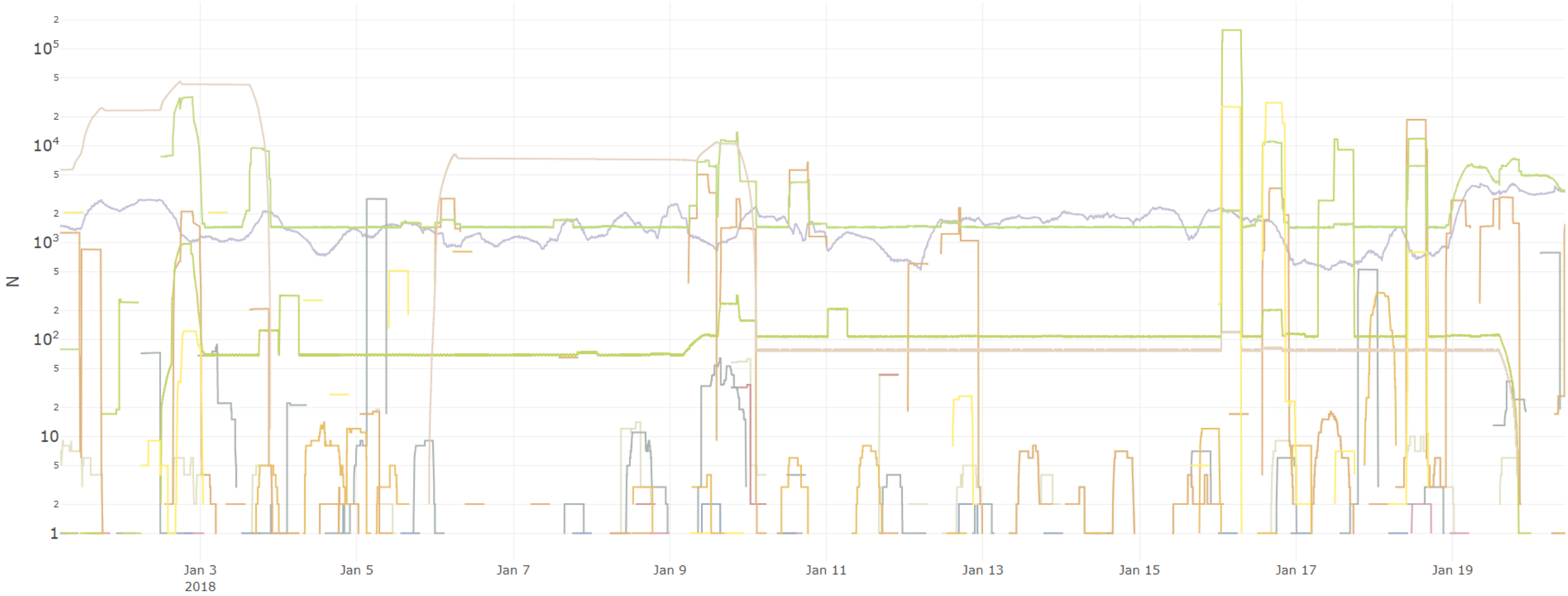


Figure 2: Event counts for different event types (in different colors) during [Jan. 01, 2018 to Jan. 20, 2018].

Log processing by Spark and Cassandra-based Analytics (LogSCAN)

Oak Ridge Leadership Computing Facility (OLCF)

Publish

kafka

Topic

Topic

Topic

Subscribe

cassandra

Spark

MESOS

Parsers

Spark

MESOS

jupyter

Scala

Java

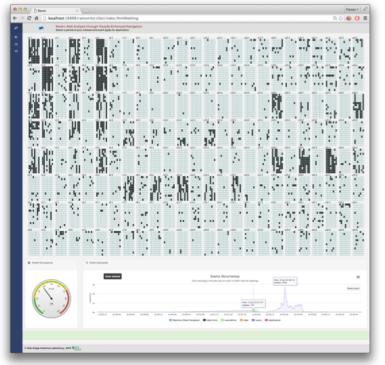
Tornado

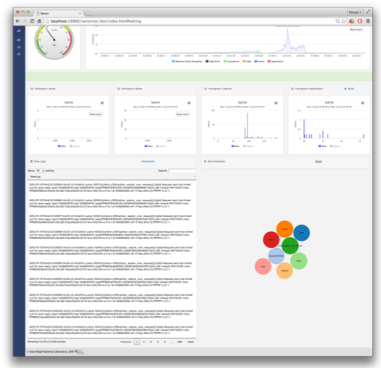
Shiny

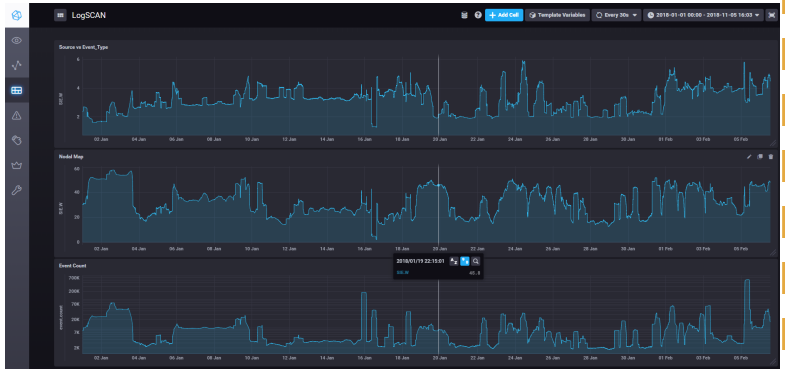
influxdb

chronograf

Kapacitor







Compute and Data Environment for Science (CADES)

OAK RIDGE National Laboratory

LogSCAN – Data Transformation

Time	Event Type	Source
2015-01-28 05:00:42	3	c10-4c2s6n0
2015-01-28 05:52:18	22	c7-2c1s6n0
2016-04-05 13:48:16	12	c20-0c1s4n3
2016-08-24 01:00:35	11	c2-1c0s1n2
2015-12-16 03:19:36	9	c3-5c0s4n2
2017-12-07 10:05:01	11	c18-7c0s2n1

Table 2: The event table contains 146.5 million events logged on all Titan nodes during the period of January 2015 to March 2018.

$Y \longrightarrow$

$X \downarrow$

Source	Machine Check Exception	Out of Memory	Seg. Fault	Lustre
c0-2c2s1n2	1	0	0	0
c10-4c0s0n0	0	0	0	4
c10-4c0s3n0	0	1	0	3
c10-4c1s6n1	0	0	3	1

Table 3: An example of the time-windowed event table. For an illustrative purpose, only 4 rows are shown out of a total of 400.

Source: c0-2c2s1n2
CabinetRow: 0
CabinetColumn: 2
Chassis: 2
Slot: 1
Node: 2

Nodal Map

$$X = 12\times Cabinet_{Row} + 4\times Chassis + Node$$
$$Y = 8\times Cabinet_{Column} + Slot$$

Figure 3: The nodal layouts for every event type and the combined total for one-hour window prior to ”2015-02-09 01:49:54”. Each layout has dimensions of [300, 64] in pixel and each pixel represents a unique Titan node with its coordinates [X, Y].

System Information Entropy (SIE)

3D **General Form of Data Table**

	Feature 1	Feature 2	…	…	Feature N
Record 1					
Record 2					
…					
…					
…					
…					
Record M					

Variance Distribution of Principal Components

2D

$$\xi_i = \frac{\sigma_i}{\sum_1^k \sigma_i}$$

1D **Shannon Entropy**

$$H = -\sum_1^k \xi_i log_b(\xi_i)$$

Entropy: in a general ”b-ary” form

2D **Principal Components**

$$\text{SVD} \Longrightarrow \sigma_i$$

σ_i: *i*-th variance out of k eigenvalues of the SVD decomposition

1D **System Information Entropy (SIE)**

$$W(t) = b^{H(t)}$$

b: the logarithmic base used in calculating H. In our analysis, b = 10.

Application – overall SIE

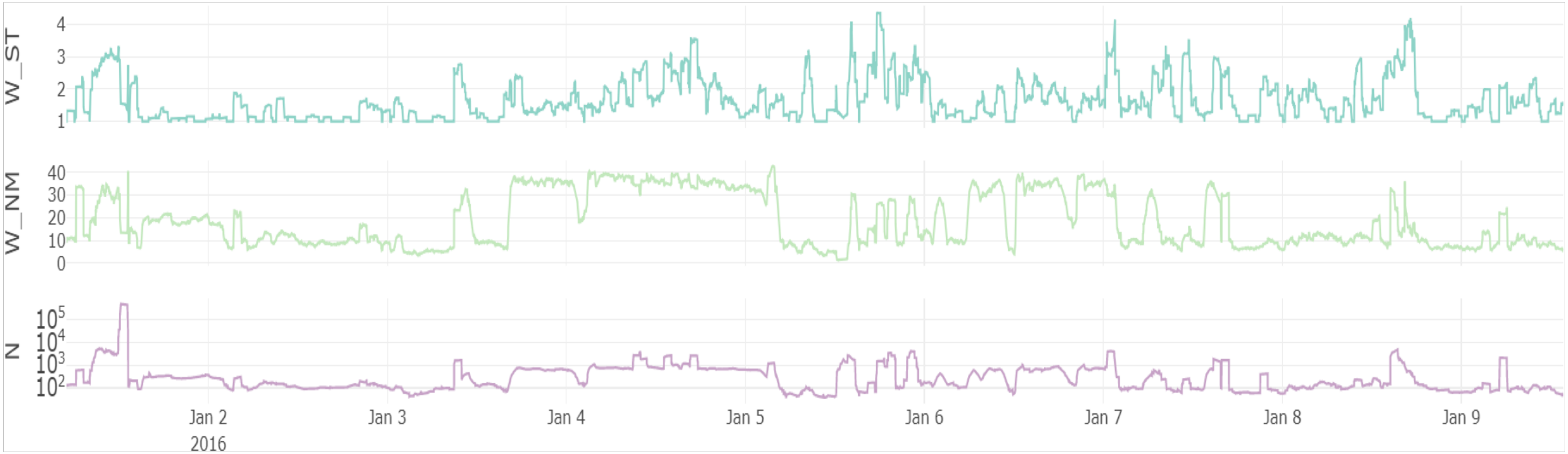


Figure 4: The top panel represents the SIE for the “Source Type” layout, while the middle panel represents the SIE for the “Nodal Map” layout. The bottom panel shows the total event counts. All three plots share the same time resolution of one minute for an hour-window and use distinctive color codes, i.e., green for “Source Type”, blue-green for “Nodal Map” and purple for total counts.

SIE – Source vs. Type

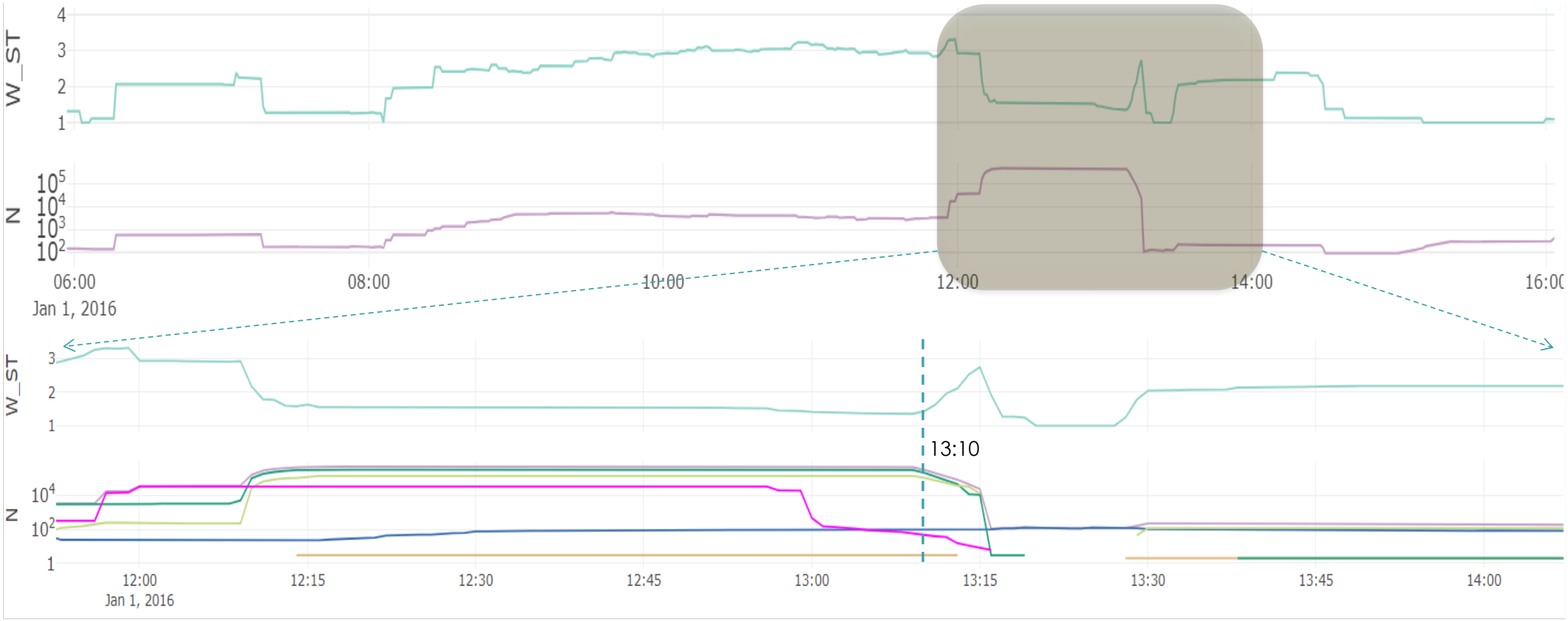


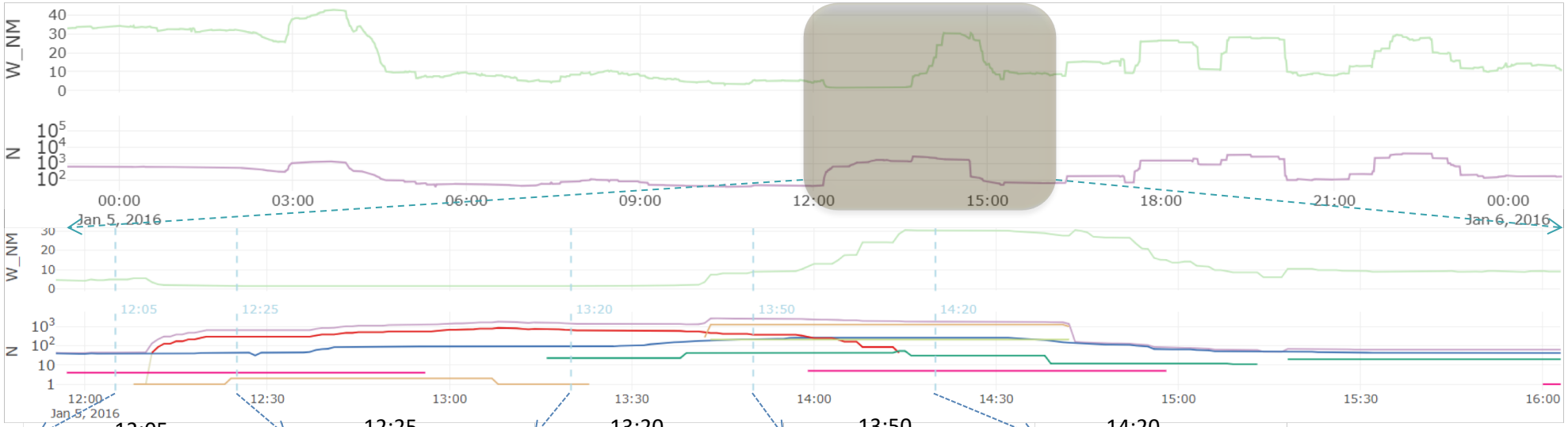
Figure 5: A close-up for SIEs of the layout “Source Type”.

The color codes for event types are:

Lustre LustreError HWERR

Lnet Machine Check Exception

SIE – Nodal Map



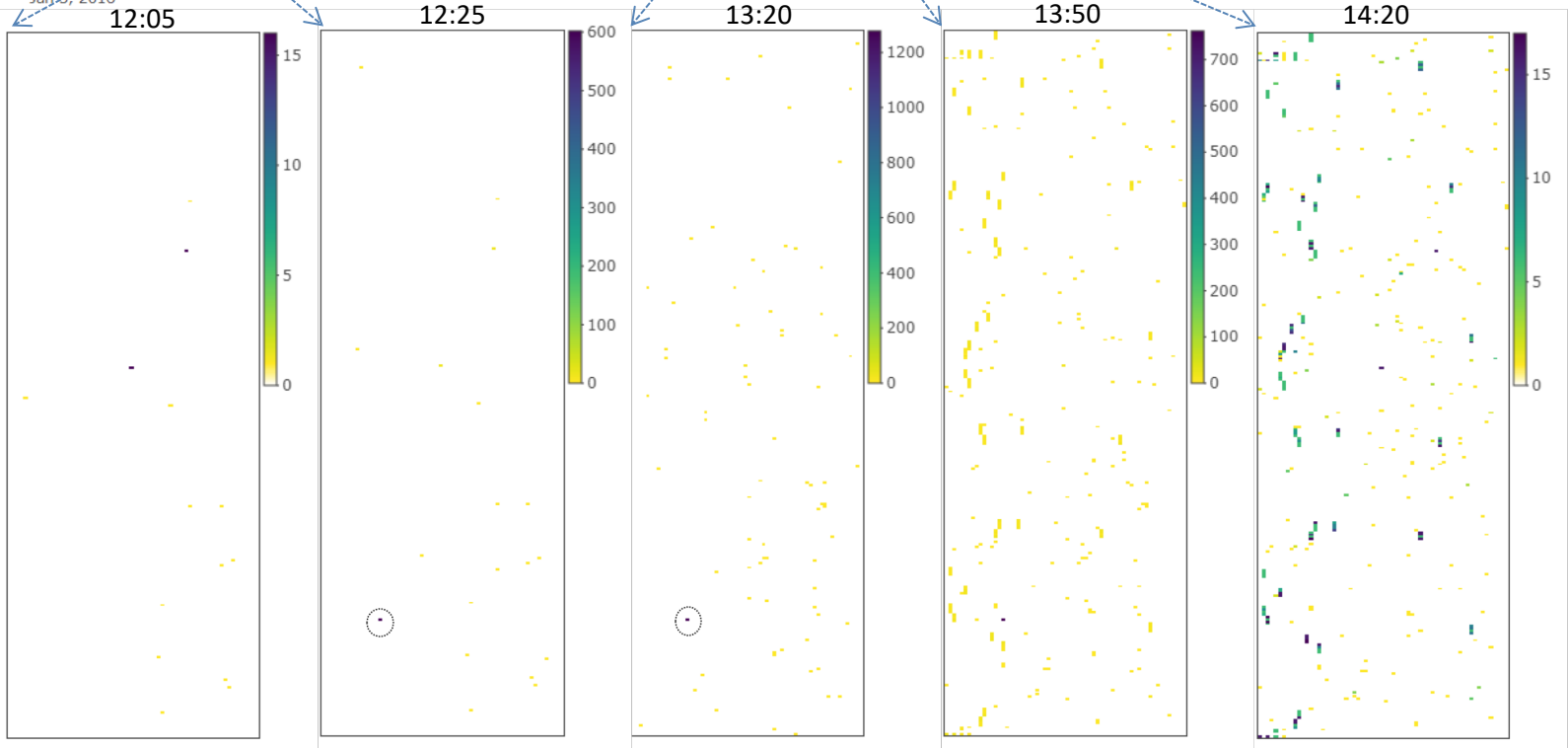


Figure 6: A close-up for SIE of "Nodal Map" layout. The color codes used are:

Machine Check Exception
Seg Fault
HWERR
NVRM Xid
Graphics Engine Error
Lustre
LNet

Conclusion

- A general-purpose metric – SIE
- A concise time series – cost efficient in both computation and visualization
- An indicator of system status – comprehensive and sensitive
- A valuable input for further analysis

Acknowledgement

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, program manager Lucy Nowell, under contract number DE-AC05-00OR22725. This work was supported by the Compute and Data Environment for Science (CADES) facility and the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is managed by UT Battelle, LLC for the U.S. DOE (under the contract No. DE-AC05-00OR22725).

Question?