

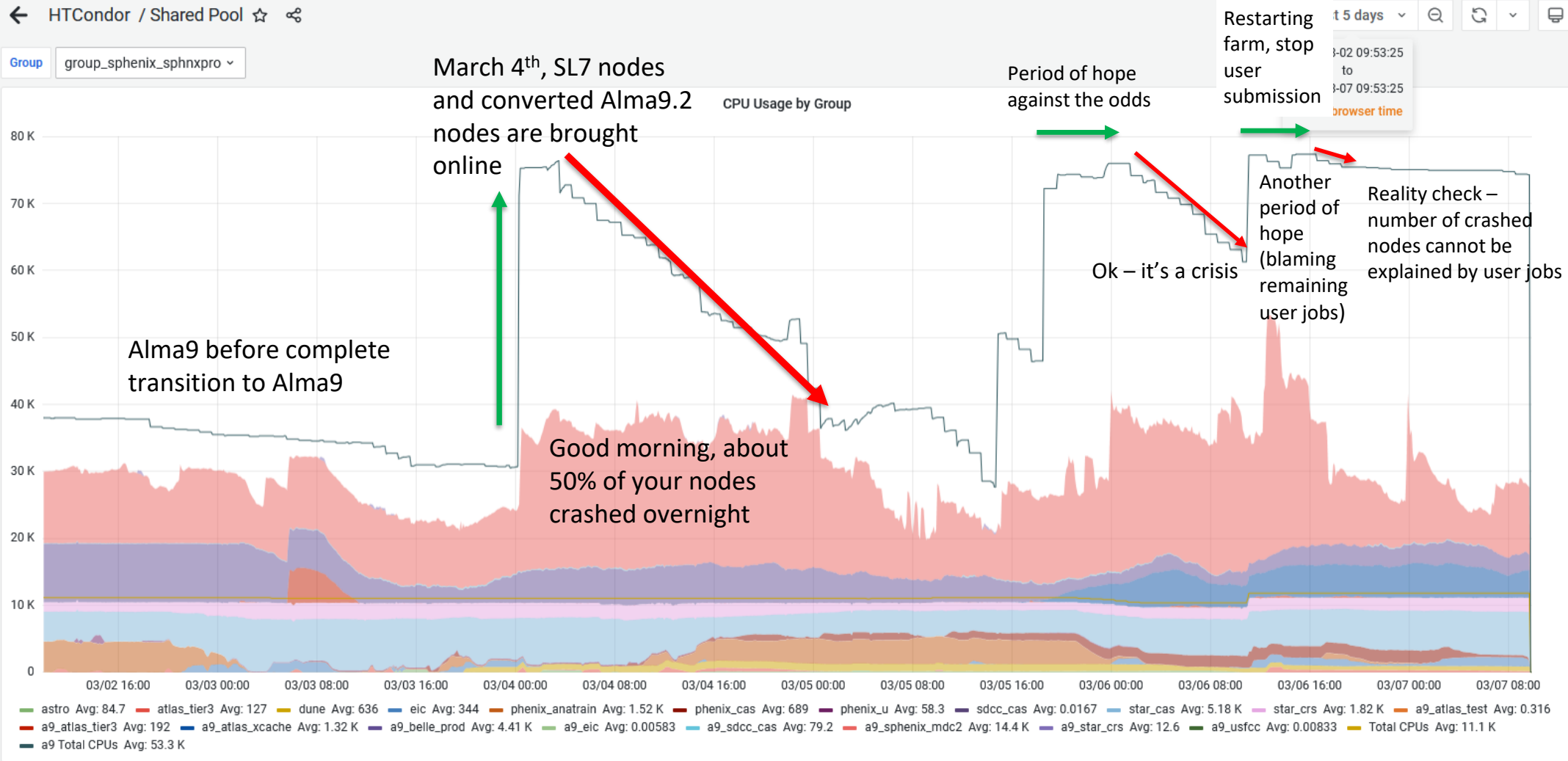
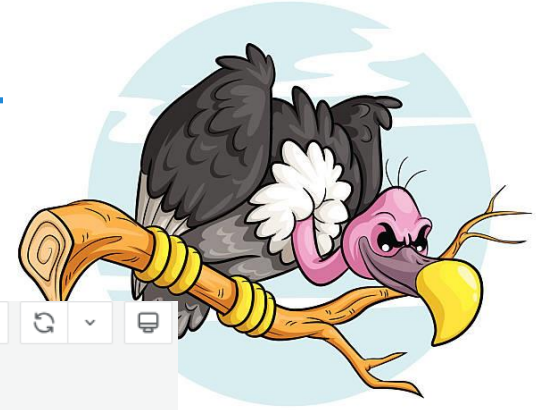
# sPHENIX

NPPS Group Meeting 03/19/2025

- 5 days till cooldown
- Blue Beam March 28<sup>th</sup>/29<sup>th</sup>
- First Collisions April 1<sup>st</sup>
- Given uncertainty, go to physics data taking asap – we have working points for all detectors (remedying the effects of those MVTX splash events)
- Computing
  - 163 new nodes (120 logical cores each) coming online this week
  - Another ~300 nodes by beginning of April
  - Move sPHENIX to it's own condor pool (which will be then the largest condor pool in the world)
  - First RHIC experiment to move to Alma9 (with SL7 still supported)
  - Lustre Reads of 350GB/s seen when building events

# What's going on in our farm?

- It's not condor – don't shoot the messenger



# Costin to the rescue, the art of kernel debugging



Before throwing in the towel – the kernel writes out a crash dump which can be analyzed

```
[63382.277076] BUG: kernel NULL pointer dereference, address: 0000000000000004
[63382.277081] #PF: supervisor write access in kernel mode [63382.277083] #PF: error_code(0x0002) - not-present page
[63382.277084] PGD 34575d067 P4D 34575d067 PUD 149645067 PMD 0
[63382.277088] Oops: 0002 [#1] PREEMPT SMP NOPTI
[63382.277091] CPU: 5 PID: 2688981 Comm: python3 kdump: loaded Tainted: G OE ----- 5.14.0503.19.1.el9_5.x86_64#1
```

Python???

ps finds the pid:

```
> 2688981 2688081 5 ffff9279d9150000 RU 0.0 171568 19188 python3
```

vm pid finds the process and all libraries it uses:

```
PID: 2688981 TASK: ffff9279d9150000 CPU: 5 COMMAND: "python3" MM PGD RSS TOTAL_VM ffff9273cc5d4100 ffff92897e15c000 19188k 171568k VMA START END FLAGS FILE
ffff927480e29ec0 400000 401000 8000071 /cvmfs/sphenix.sdcc.bnl.gov/gcc-12.1.0/opt/sphenix/core/stow/Python-3.10.8/bin/python3.10 ffff927480e29a40 401000 402000 8000075
/cvmfs/sphenix.sdcc.bnl.gov/gcc-12.1.0/opt/sphenix/core/stow/Python-3.10.8/bin/python3.10
```

```
...
#7 [ffffb2976019f210] ll_lock_cancel_bits at ffffffff1ef7c7a [lustre]
#8 [ffffb2976019f270] ll_md_blocking_ast at ffffffff1ef8333 [lustre]
```

Okay – we tried the SL7 stock python but it was a bug in Lustre

Bug details

<https://jira.whamcloud.com/projects/LU/issues/LU-18085?filter=allissues>

...  
Looks like above patch set for b2\_15 unfortunately did not make it into 2.15.6.

No rpm with patch:  
Costin pulled patch from git, handled the merging conflicts and rebuild lustre with this patch  
Patched nodes have been stable

# Hpss Readback

## ERADAT

### FILE STAGING REAL-TIME STATUS

sphnxrdat													
Tape Info	Tape ID	Files	Avg size	Status	Files staged	GB Staged	Avg MB/s	Files failed	Last staged	Mount Time	Drv Addr	Drv Type	Waiting
sPhenix_C	<a href="#">P91300</a>	2 / 954	18,096,476,160	Reading	45	850.15	334.57		3-03 14:47:30	3-03 14:05:15 ( 00:42:29 )	<a href="#">000788F0BB</a>	IBM LTO9	
sPhenix_D	<a href="#">P92300</a>	2 / 880	21,474,824,192	Reading	48	894.45	348.92		3-03 14:47:35	3-03 14:05:15 ( 00:42:29 )	<a href="#">000788F24B</a>	IBM LTO9	
TOTAL:		2 Tapes	4 Files				Avg 341.75 MB/s/dr						

- Hpss readback test shows great performance
- Staging all files from a raw data tape (we have 2 silos, C and D)
- LTO9 specs is 400MB/sec, we get around 350MB/sec
  - The 20GB filesize really helps
- With gaps (fetching some 100 files), still around 250MB/sec
- Writing with the new movers: sustained 10GB/sec → 20GB/sec
  - Daq rates north of 30GB/sec