

# Luminous upgrade plans

- Built new packages for XrootD (server + plugin) and GridFTP (plugin) against the release of Luminous we intend to run
  - Some differences in package names, otherwise went smoothly
- Upgraded test environment
  - Mixing Kraken and Luminous OSDs and MONs works fine during the upgrade
  - The first RGW to be brought up as Luminous got stuck in a crash-restart loop trying to reshard a large bucket index
    - Had to turn off the dynamic resharding feature
    - Running 2 Luminous + 1 Kraken RGW seems to be fine
    - Have not tested dynamic resharding yet
  - GridFTP and XrootD tested and confirmed to be interoperable as well
  - EC backfill bug still present
    - Luminous behaves slightly better: when receiving corrupted data from a shard it will sometimes ignore the shard and go to the others, this is the desired behaviour in that situation, not clear what the condition is that causes this to not be the case all the time
  - BlueStore is the new default, we're not ready to move Echo onto BlueStore yet but I didn't find an option to change the default in ceph.conf (maybe switch passed to ceph-deploy when deploying new OSDs?)
  - Maximum object size has been drastically reduced to 128MB in Luminous (can cause problem if not using striping and not expecting it)
- Planning to do a rolling upgrade
  - Do Monitors (and Managers) first, then OSDs and GWs last
  - No downtime expected but a period of hand-holding everything through a restart of ceph-osd daemons and peering everything again is expected
  - Plan is to roll the Spectre/Meltdown patches into this as well, avoid having to restart twice

# Luminous current experiences

- We have a new instance of Ceph running Luminous since the end of summer
  - This is going to evolve into the final, production version of the SCD Cloud's backing storage
  - It consists of 3 Monitors (HyperV), 11 hosts, 131 OSDs
  - It hosts 1 pool (2048 PGs) and has a capacity of 476TB
  - It was deployed as a clean Luminous install, with Luminous defaults, including BlueStore
  - It's being tested with OpenStack and has been stable and performant according to user reports
- The #1 concern right now is the opaqueness of BlueStore, we have not yet got around to getting FUSE set up properly so we can inspect what's going on at the disk layer.

# Spectre & Meltdown

- We have kernel and microcode patches available
  - Already applied to one gateway (under observation)
  - Plan to do outward facing gateways this week
- No idea what the actual performance impact could be
  - Some community reports are scary (up to 30% performance hits reported)
  - If problematic, mitigations can be disabled
  - Current plan is to patch along with Luminous upgrade
    - We'll be rolling out the patch to a monitor and a couple of SNs first
    - Can change if we decide to separate them

# OSD LevelDB inflation

- Shortly after CMS started using Echo we noticed the LevelDB sizes of some OSDs jumped up
- The expected size across the cluster would fall in the 400-800MB range, some of these are now 25GB.
- Looking into which OSDs have the largest LevelDBs, the common thread was one PG in the CMS pool. The PG in question appears completely ordinary.
- There is a growth trend both in the number of OSDs with LevelDBs well above average and the size of the LevelDB itself.
- Compaction happens normally but doesn't trim the size down. Ceph developers have speculated this may be due to problematic states on the PG not clearing and suggest upgrading to Luminous and reporting back whether or not that helps.