ceph days

CLYSO

Ceph Durability
How Safe Is My Data?

Dan van der Ster (CLYSO & Ceph Council)

Ceph Days Seattle - May 15, 2025

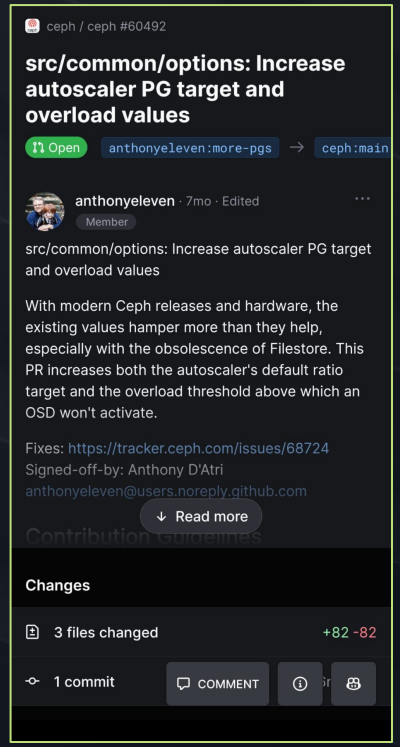
CL'SO

Outline

- **Why This Matters**
  - The motivation behind rethinking durability in modern Ceph clusters
- **Understanding the Concepts**
  - Durability vs. Availability – what they really mean
- **Measuring Durability in Practice**
  - Introducing a new **Ceph Reliability Calculator**

Motivation

- The shift to very large disks is forcing a re-evaluation of key Ceph constants (e.g. PR #60492)
- These changes could have unexpected consequences for Ceph's reliability
- We need a Ceph Reliability Calculator to guide our decisions!



If we double the number of PGs per OSD, does our risk of data loss double?

Motivation

- The shift to very large disks is forcing a re-evaluation of key Ceph constants (e.g. PR #60492)
- These changes could have unexpected consequences for Ceph's reliability
- We need a Ceph Reliability Calculator to guide our decisions!

If we double the number of PGs per OSD, does our risk of data loss double?

Previous Work

- In 2013, Mark Kampe created the original Ceph reliability model
- In 2014, Veronica Estrada Galinanes, Kyle Bader, and Loïc Dachary expanded on this in their Google Summer of Code final report
- These works developed the mathematical foundation for Ceph reliability, similar to traditional RAID calculators

- This talk takes a different approach — using **Monte Carlo** simulation to explore real-world durability risks

Background: Durability vs. Availability

- Terms like “reliability,” “accessible,” and “data protection” are often used loosely
- It’s helpful to separate two key concepts:

Durability

- Is my data safe?
  Can I read the exact bytes I wrote?
- Flip the question:
  - *How much data will I lose per year?*
  - *How long until the first byte is lost?*

Availability

- Is the system up when I need it?
  Can I read or write now?
- Flip the question:
  - *How much downtime should I expect per year?*

Durability vs. Availability – An Example

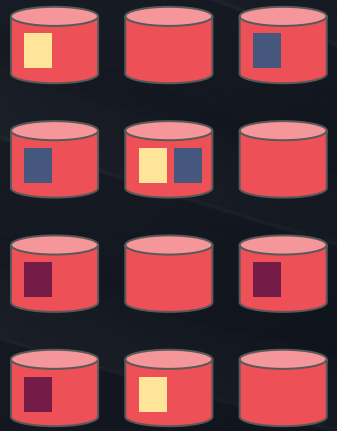
- **Amazon S3 SLA:**
  - "Designed to provide 99.999999999% durability and 99.99% availability of objects over a given year."
- **11-nines Durability:**
  - What does that mean in practice?

    ```
    10 PiB × (1 - 0.99999999999) = 100 KiB lost per year
    → ~1 KiB lost per 100 TiB, per year
    ```
- **4-nines Availability:**
  - How much downtime is that?

    ```
    (1 - 0.9999) × 1 year = 52.6 minutes of downtime per year
    ```

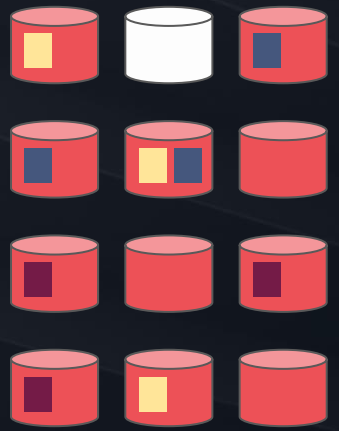
Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost



Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

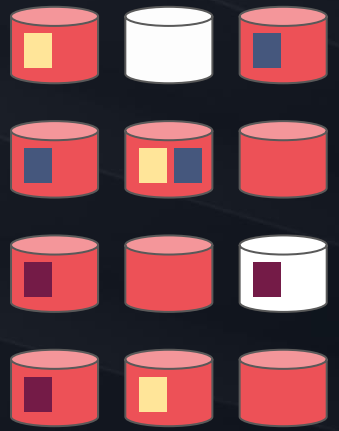


Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

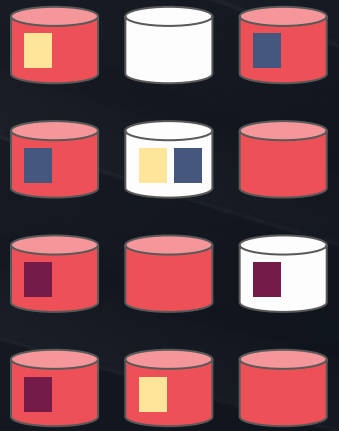


Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

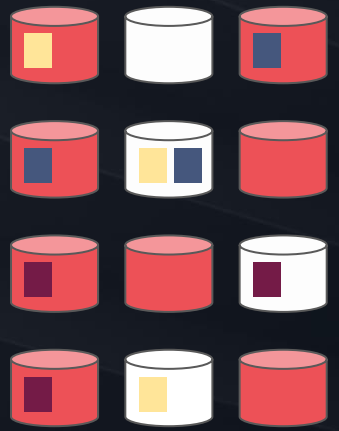


Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

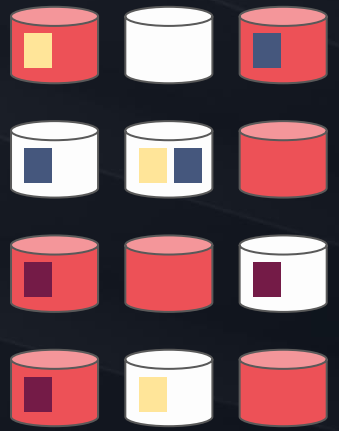


Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

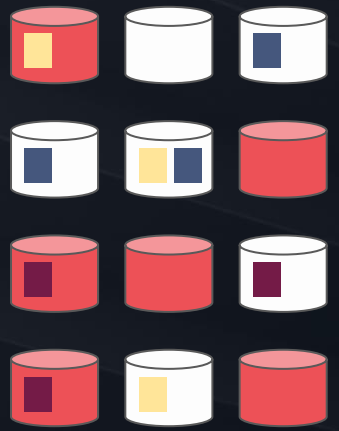


Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost



Ceph Reliability

- **Ceph has several innovations to improve durability and availability:**
  - **CRUSH** and **Placement Groups** store data intelligently to minimize the impact of disk, host, rack, ... failures on durability and availability.
  - And **Ceph is very flexible**!... highly configurable Replication and Erasure Coding profiles
- Ceph differentiates availability and durability:
  - Our standard profile is 3 replicas, min_size 2:
    - Data is available if 2 out of 3 copies online
    - Data is durable if at least 1 copy is not lost
  - Similar for erasure coding, e.g. k=4, m=2, min_size 5:
    - Data is available if 5 out of 6 shards are online
    - Data is durable if 4 out of 6 shards are not lost

How did CERN “discover” the Higg’s Boson?

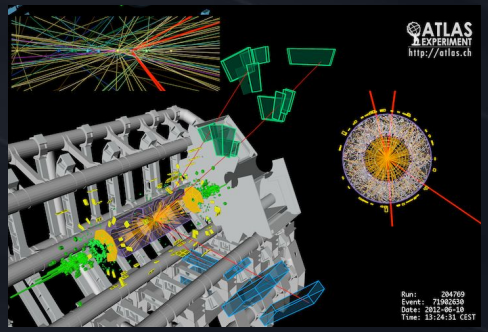
There are two types of particle physicists:

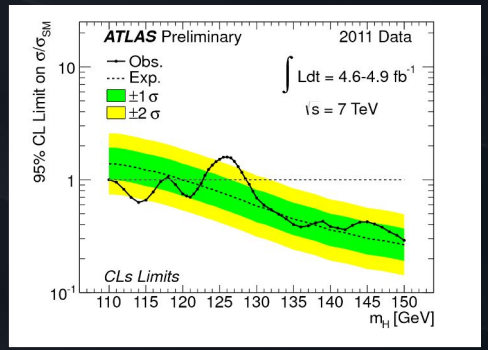
- **Theorists** write equations describing how we think the universe works
- **Experimentalists** build machines to test if reality agrees

The **Large Hadron Collider** was built to compare theory with observation – *difference? we discovered something!*

- Physicists run **Monte Carlo simulations** of particle collisions to predict what results should look like
  → These predictions form the "expected curve" (dashed line with green/yellow error bands)

BTW, you can trust your local particle physicist.
They don’t call something a “discovery” unless it’s **5σ away** from expectations – that’s five standard deviations!
It takes *an absurd amount* of simulated and real data to get there.





Calculating Ceph's Durability

- **What's the goal?**
  - Understand the impact on durability when increasing the default pg_num per OSD from **100 to 200**
- **How can we calculate this?**
  - Be a **Ceph Theorist**?
    → That requires complex probability math (see the earlier work...)
  - Or be a **Ceph Experimentalist**?
    → Let's **simulate it** using Monte Carlo methods ✅
- We don't need exact equations – we need realistic risk estimates.

Simulating Ceph

- **RADOS Recap**
  - Objects are placed into Placement Groups (PGs)
  - PGs use replication (e.g., 3×) or erasure coding (k+m) across OSDs
  - OSDs are organized into failure domains (disk, host, rack, …)
- **Failure Scenarios**
  - Single-disk failures, entire-host outages, correlated multi-host failures
  - Random timing: e.g., two disks fail minutes apart in different domains
- **When Does Data Loss Occur?**
  - **Replication**: if all N replicas of a PG's data fail together
  - **Erasure Coding**: if more than m shards of a k+m profile are lost
- *Final exam: Does that mean if I lose 3 OSDs at the same time, I will lose data?*

Simulating Ceph

Does that mean if I lose 3 OSDs at the same time, I will lose data?

Not necessarily! You only lose data if all 3 failed OSDs are part of the same PG.

- **Example:**
  - You have **100 OSDs**: `osd.0` … `osd.99`
  - Then `osd.3`, `osd.42`, and `osd.71` all fail
  - Do you lose data?
- **Answer:**
  - Only if at least one PG has `acting = [3, 42, 71]`
  - If no PG depends on all 3, **no data loss occurs**

Simulating Ceph

- Back in 2014, I built a simulator that connects to a real Ceph cluster and evaluates the impact of concurrent failures.
- **Simulation Loop (10,000+ iterations):**
  - Pick *N* random OSDs to “fail” simultaneously
  - Check if any PG becomes unavailable, record the result.
- **Output:**
  - $P_{unlucky}$ = num_failures_with_PG_loss / total_simulations
- If we multiply by the probability of *N*+ concurrent failures, we can estimate the overall durability of the cluster.

$$P_{dataloss} = P_{N\_concurrent} \times P_{unlucky}$$

Ceph Failure Simulator

- **Live Demo Time.**
- Note: $P_{N\_concurrent}$ is very difficult to estimate accurately

  So let's focus on **comparative** studies...

  - PGs per OSD: 100 vs 200
  - OSDs per cluster: 500 vs 1000
  - Replication or EC: 3x vs 4+2
  - Other EC profiles: 2+2 vs 10+2

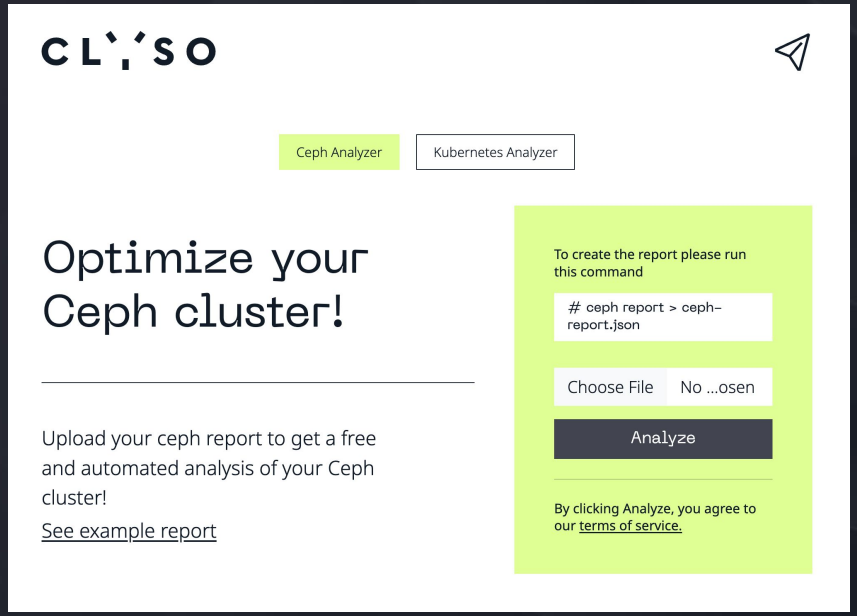
CL'SO

Next Steps

- **Add Availability Modeling**
  - Estimate expected downtime per year alongside durability
- **Reconnect to Live Clusters**
  - Enable simulations using real CRUSH maps and PG states
- **Share the Results**
  - Write a detailed blog post for ceph.io with examples and guidance

CLYSO

Free Stuff!



analyzer.clyso.com

ceph days

CLYSO

Thank you!

Dan van der Ster – dan.vanderster@clyso.com