A Study of SSD Reliability in Large Scale Enterprise Storage Deployments

FAST 2020

Stathis Maneas, Kaveh Mahdaviani, Tim Emami, Bianca Schroeder

Seminar in Computer Architecture
Presented by Marco Imboden
15.12.2022
Executive Summary

**Problem:** Reliability of SSDs in enterprise storage systems has not been tested so far

**Goal:** Examine the reliability of SSDs and the reasons for failures in enterprise storage systems.

**Challenges:**
- Find sufficient data sets
- Finding reasonable failure categories and other investigation criteria

**Key Results:**
- One third of all failures are preventive (based on predictions)
- Regularly updating the firmware can prevent from data loss
Outline

Paper presentation
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

Analysis
5. Strengths
6. Weaknesses
7. Takeaways
8. Discussion
Most important points when storing data:

• Access time

• Price

• Reliability
Introduction – HDD and SSD

Changing Focus

• Faster
• Less power consumption
• smaller
Background – Storage Types

Distributed Storage systems

• Faster for remote Access
• Same time data access possible
• More fault tolerant

Enterprise Storage systems

• Faster for local Access
• Same time data access may cause problems
• Higher data integrity
Outline

Paper presentation
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

Analysis
5. Strengths
6. Weaknesses
7. Takeaways
8. Discussion
Related Work

Several studies based on data centers from Facebook, Google, Microsoft and Alibaba

- No failure distinction
  - No indication of the severity
- Distributed data centers
Related Work - Facebook

- Published 2015
- First large scale study of SSD in the Field
- Focusing on how many memory errors manifest
Related Work - Microsoft

- Published 2016
- Analysis of what, when and why characteristics of SSD failures
- 500’000 SSDs analyzed
Related Work - Google

- Published 2016
- Large-scale field study
- different flash technologies (MLC, eMLC, SLC)
Related Work - Alibaba

- Published 2019
- Analyze both device errors and system failures
- Particularly, focusing on failures that are Reported As “SSD-Related” (RASR)
Goal of this Study

• Focusing on large enterprise storage systems
• Focusing on the reason of failures (reason of drive replacements)
  • Scope of the underlying failure
  • Data lost or just prediction
Outline

**Paper presentation**

1. Introduction and Background
2. Related Work

3. Methodology

4. Results

**Analysis**

5. Strengths
6. Weaknesses

7. Takeaways
8. Discussion
Active IQ bundles
- Contains no information about the stored data
- **Logfile** with information about the Drive (failures, PE cycles, device role, etc)
- Send to NetApp *every week*
Methodology – Overview

• No creation of data

• Telemetry data of **1.4 million SSDs** (NetApp) in real use analysed

• Separate data set contains information about **each drive failure** and replacement

• **Different drives families** were investigated by the study
# Drive Characteristics

<table>
<thead>
<tr>
<th>Manufact./Model</th>
<th>Cap. (GB)</th>
<th>Flash Tech.</th>
<th>Lithography</th>
<th>PE Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>I - A</td>
<td>200</td>
<td>eMLC</td>
<td>2</td>
<td>K</td>
</tr>
<tr>
<td>I - B</td>
<td>400</td>
<td>eMLC</td>
<td>1</td>
<td>K</td>
</tr>
<tr>
<td>I - C</td>
<td>400</td>
<td>eMLC</td>
<td>1</td>
<td>K</td>
</tr>
<tr>
<td>II - A</td>
<td>3800</td>
<td>3D-TLC</td>
<td>V3</td>
<td>10K</td>
</tr>
<tr>
<td>II - B</td>
<td>3800</td>
<td>3D-TLC</td>
<td>V2</td>
<td>10K</td>
</tr>
<tr>
<td>II - C</td>
<td>8000</td>
<td>3D-TLC</td>
<td>V3</td>
<td>10K</td>
</tr>
<tr>
<td>II - D</td>
<td>960</td>
<td>3D-TLC</td>
<td>V3</td>
<td>10K</td>
</tr>
<tr>
<td>II - E</td>
<td>400</td>
<td>3D-TLC</td>
<td>V2</td>
<td>10K</td>
</tr>
<tr>
<td>II - F</td>
<td>400</td>
<td>3D-TLC</td>
<td>V2</td>
<td>10K</td>
</tr>
</tbody>
</table>

**Diagram:**
- **SLC:** 1 bit/cell
- **MLC:** 2 bits/cell
- **TLC:** 3 bits/cell

**Legend:**
- **P/E Cycle:**
  - K: 10K cycles
  - V2: 20K cycles
  - V3: 30K cycles

**Flash Technology:**
- **eMLC:** Enhanced Multi-Level Cell
- **3D-TLC:** 3D Tri-Level Cell
Methodology – Health Metrics

• Percentage of **Spare Blocks** Consumed (Spare blocks = 2.5% of capacity)
  • “Substitute” storage space

• Number of **bad sectors**
  • Every unrecoverable error counts as bad sector

• Annual Replacement Rate (**ARR**)
  • Number of device failures divided by number of devices
Methodology – Health Metrics

• Percentage of Spare Blocks Consumed (2.5% of capacity)
  • “Substitute” storage space

white = normal drive storage
Gray = additional drive storage (spare blocks)
Methodology – Health Metrics

- Number of bad sectors
  - Sector that cannot be read (physical damage or software related problem)
    → Can lead to use a spare block
- Annual Replacement Rate (ARR)
  - \( \frac{\text{Number of device failures}}{\text{Number of devices}} \)
Outline

**Paper presentation**
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

**Analysis**
5. Strengths
6. Weaknesses
7. Takeaways
8. Discussion
Results – Reasons for Replacements

<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Pct.</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCSI Error</td>
<td>32.78%</td>
<td>Full</td>
</tr>
<tr>
<td></td>
<td>Lost Writes</td>
<td>13.54%</td>
<td>Partial</td>
</tr>
<tr>
<td></td>
<td>Aborted Commands</td>
<td>13.56%</td>
<td>Partial</td>
</tr>
<tr>
<td></td>
<td>Disk Ownership I/O Errors</td>
<td>3.27%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Command Timeouts</td>
<td>1.81%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Predictive Failures</td>
<td>12.78%</td>
<td>Zero</td>
</tr>
<tr>
<td></td>
<td>Threshold Exceeded</td>
<td>12.73%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recommended Failures</td>
<td>8.93%</td>
<td></td>
</tr>
</tbody>
</table>

Finding 1

One third of replacements are associated with one of the most severe reason types, but on the other hand, one third of drive replacements are preventative.
Factors Impacting Device Failures

- Usage and Age
- Flash and Drive Type
- Capacity
- Lithography
- Firmware Version
- .. and others
Usage and Age – PE Cycles

• Most of the devices have **not used more than 1%** of the PE cycles (very little used)
• Lost writes error (cat. B) decreases rapidly in later stages
• For RL *Used > 50%* predictive failures increase
• **Expected Behaviour:** Bathtub curve known from reliability engineering
  • Short initial period of high failure rates
  • Followed by constant failure rate
  • Increasing failures by the end of life
Usage and Age – Actual Behaviour

Finding 2
We observe a very **drawn-out period of infant mortality** and see failure rates 2-3X larger than later in life.
Flash and Drive Types

• Types: SLC (Single Level Cell), MLC (Multi Level Cell), 3D-TLC (Triple Level Cell)
• Consumer class: cMLC (just one family)
• Enterprise class: eMLC

Finding 3

Overall, the highest replacement rates are associated with 3D-TLC SSDs. No single flash type has noticeably higher replacement rates than the other flash types studied in this work.
• Different capacities from **100GB to 15.3TB** analyzed

• Number of bad sectors increases with higher capacities:

<table>
<thead>
<tr>
<th>Capacity</th>
<th>400 GB</th>
<th>800 GB</th>
<th>1600 GB</th>
<th>3800 GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad sectors</td>
<td>1.9</td>
<td>5.6</td>
<td>6.4</td>
<td>14.97</td>
</tr>
</tbody>
</table>

- x2
- x2
- x2
- x3
- x1.1
- x2.3
Capacity – ARR

• Differences begin for capacity > 1600 GB

• Rate of predictive failures (cat. D)

Finding 4

Drives with very large capacities not only see a higher replacement rate overall, but also see more severe failures and fewer of the (more benign) predictive failures.
Lithography

- eMLC: ARR is twice as high with 1xnm as with 2xnm
- For 3D-TLC the trend is reversed
- Failures cat.A are not related to lithography

Finding 5

*Higher density drives do not always see higher replacement rates.* In fact, we observe that, although higher density eMLC drives have higher replacement rates, this trend is reversed for 3D-TLC.
Firmware Version

- Same Firmware Version over all data snapshots in 70% of drives
- Earliest Versions much less reliable

Finding 6

*Earlier firmware versions can be correlated with significantly higher replacement rates*, emphasizing the importance of firmware updates.
Finding 7
SSDs with a non-empty defect list have a higher chance of getting replaced, not only due to predictive failures, but also due to other replacement reasons as well.

Finding 8
SSDs that make greater use of their over-provisioned space are quite likely to be replaced in the future.
Correlation between Drive Failures

• Random drive **failure probability** in a RAID: 0.0504%

• Drive failure **probability following within a week** on a previous drive replacement: 9.39%

• RAID reconstruction imposes **an additional load to other drives**

• Shared **environmental conditions**: overheating, power surge etc.
Finding 9

While large RAID groups have a larger number of drive replacements, we find no evidence that the rate of multiple failures per group is correlated with RAID group size.

The reason seems to be that the likelihood of a follow-up failure after a first failure is not correlated with RAID group size.
Executive Summary

Problem: Reliability of SSDs in enterprise storage systems has not been tested so far

Goal: Examine the reliability of SSDs and the reasons for failures in enterprise storage systems.

Challenges:
- Find sufficient data sets
- Finding reasonable failure categories and other investigation criteria

Key Results:
- One third of all failures are preventive (based on predictions)
- Regularly updating the firmware can prevent from data loss
Outline

**Paper presentation**
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

**Analysis**
5. Strengths
6. Weaknesses
7. Takeaways
8. Discussion
Strengths

• First large scale enterprise storage systems
• Large amount of data analyzed (1.4 million SSDs)
  • Different drive families
  • Different storage size
• Long time data range (2.5 years)
• Use of existing data
• Comparison of data is very detailed (different cases)
Outline

**Paper presentation**
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

**Analysis**
5. Strengths
6. Weaknesses
7. Takeaways
8. Discussion
Weaknesses

• **Failure Type** is *missing* in 40% of the collected metadata
  • These failures were *proportionally spread over all categories*

• In the chapter “Age and Usage” the used metric was *time (months).* It would be better to measure the “Age and Usage” in **PE Cylces**

• No comparison to similar research about HDD
Outline

**Paper presentation**
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

**Analysis**
5. Strengths
6. Weaknesses

7. Takeaways
8. Discussion
Takeaways

• **Firmware** Version should be **always updated**!
• Using **already available data** can lead to new insights
• Some failures can be **predicted** with software
• **Data backup** or a **RAID** is important, even if the drives are very reliable today
Outline

Paper presentation
1. Introduction and Background
2. Related Work
3. Methodology
4. Results

Analysis
5. Strengths
6. Weaknesses
7. Takeaways

8. Discussion
What kind of data storage would you prefer for a **start-up**, centralized enterprise data storage or distributed data storage?

- Which one is **safer**?
  - Distributed: Not all data at the same location
  - Centralized: All data at the same location

- **Access** time
How can we use this study to improve reliability or other things?
How can failures be reduced?

• Improve **failure prediction**:
  • Fitting failure prediction to specific failure category
  • With higher capacity, the failure prediction decreases

• others
Discussion - III

For a centralized enterprise storage system, which flash type would you choose?

- **the more bits per cell, the less PE cycles possible (2D)**
- 3D TLC has great advantages:
  - **More PE cycles** than 2D TLC
  - **Higher densities at lower cost**
  - **Lower Power Consumption**
Acknowledgement

Thanks to my mentors

Mohammad Sadrosadati
Rakesh Nadig
Haocong Luo
Thank you for Listening!