SSD Failures in Datacenters: What? When? And Why?

Iyswarya Narayanan, Di Wang, Myeongjae Jeon, Bikash Sharma, Laura Caulfield, Anand Sivasubramaniam, Ben Cutler, Jie Liu, Badriddine Khessib, Kushagra Vaid

The 9th ACM Systems And Storage Conference (SYSTOR 2016)
Why SSD Reliability?

Data reliability

01001100 01001101 11010010 01000000
10011100 10111111 10101111 11000101

SSDs’ popularity

Datacenter decision support

Limited field data

46.5% annual growth*

*Source: IDC, Dec 2015
Why SSD Reliability?

Data reliability

SSDs’ popularity

46.5% annual growth*

Datacenter support

Limited field data

Large scale Field data

01001100 01001101 11010010 01000000 10011100 10111111 10101111 11000101

*Source: IDC, Dec 2015
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.
SSD Failures

- Flash failures
  - Media wear-out
  - Data Retention
  - Program disturb
  - Erase disturb

- FTL Mechanisms
  - Wear levelling
  - Error detection
  - Error correction
  - Flash correct and refresh, etc.
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

Fail-stop failures
SSD Reliability

- AFR=0.61
- AFR=0.73

SSD Model

1-A

1-B

1-C

1-D

2-A

Annualized Failure Rate %

Consumer

Enterprise

SSD Model
SSD Reliability

Annualized Failure Rate %

- AFR=0.61
- AFR=0.73

SSD Model

1-A
1-B
1-C
1-D
2-A

SSD
SSD Reliability

![SSD Reliability Chart]

Annualized Failure Rate %

SSD Model

AFR=0.61
AFR=0.73

5 large datacenters
SSD Reliability

SSD Model

Annualized Failure Rate %

1-A 1-B 1-C 1-D 2-A

AFR=0.61  AFR=0.73

4 major workloads
SSD Reliability

- AFR=0.61
- AFR=0.73

6 different rack SKUs
Various factors in production environment could affect SSD failure trends very differently from lab test conditions.

Can we understand SSD failures in the presence of various factors?
Understanding SSD Failures – An analogy

Reactive

Proactive
What are the symptoms?

Fever

Unexpected weight loss

Low blood pressure

Reallocated sectors

Data errors

Program and erase failure

SATA downshift
Insufficiency of symptom only diagnosis

Symptoms seen only in 62% of failed devices
What are the factors?

- Lifestyle
- Environmental agents
- Genetics
- Workload
- Production environment
- Design decisions
- SSD
Device level correlating factors

- Average write rate of a device
- Average read rate of a device
- Total read and/or write usage
- Write Amplification
- Read Write Ratio

![Graph showing AFR % vs. Avg. host writes per day with increasing failure trend at higher write rates.]

More results in the paper
Server level correlating factors

- SSD space utilization
- Disk space utilization
- Memory utilization
- Processor utilization

More results in the paper
Datacenter factors

Same model
different behavior

More results in the paper
Understanding SSD Failures – An analogy

MULTI FEATURE ANALYSIS

Symptoms

Factors

SSD

Symptoms

Factors
Understanding SSD Failures – An analogy

Symptoms

Factors

Random forest based binary classification
Permutation feature ranking
Understanding *What*?

What are the important factors?  
is their order of importance?  
are the important combinations?
Understanding *What?*

**SYMPTOMS**

- DataErrors
- ReallocSectors
- TotalNANDWrites
- HostWrites
- TotalReads+Writes
- AvgMemory
- AvgSSDSpace
- UsagePerDay
- TotalReads
- ReadsPerDay

**Feature Importance**

![Bar chart showing feature importance](chart.png)
Understanding *What?*

![Bar Chart](Image)

- **DEVICE**
  - WORKLOAD
  - **Feature Importance**
    - DataErrors
    - ReallocSectors
    - TotalNANDWrites
    - HostWrites
    - TotalReads+Writes
    - AvgMemory
    - AvgSSDSpace
    - UsagePerDay
    - TotalReads
    - ReadsPerDay

---

28
Understanding What?

Feature Importance

0 0.2 0.4 0.6 0.8 1

DataErrors
ReallocSectors
TotalNANDWrites
HostWrites
TotalReads+Writes
AvgMemory
AvgSSDSpace
UsagePerDay
TotalReads
ReadsPerDay

SERVER WORKLOAD
Understanding *What*?

Combinations of top 8 important features

**Frequent Combinations**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Errors &lt;=1 &amp; Reallocated Sectors &lt;=5</td>
<td>H</td>
</tr>
<tr>
<td>Data Errors &lt;=1 &amp; WAF &lt;=1</td>
<td>H</td>
</tr>
<tr>
<td>Media Wear-out = 100 &amp; WAF &lt;= 1</td>
<td>H</td>
</tr>
<tr>
<td>Avg. SSD space &gt;= 10</td>
<td>F</td>
</tr>
</tbody>
</table>
Understanding *What*?

Combinations of top 8 important features

### Frequent Combinations

<table>
<thead>
<tr>
<th>Condition</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Errors &lt;=1 &amp; Reallocated Sectors&lt;=5</td>
<td>H</td>
</tr>
<tr>
<td>Data Errors&lt;=1&amp; WAF&lt;=1</td>
<td>H</td>
</tr>
<tr>
<td>Media Wear-out=100 &amp; WAF&lt;=1</td>
<td>H</td>
</tr>
<tr>
<td>Avg. SSD space &gt;=10</td>
<td>F</td>
</tr>
</tbody>
</table>
Understanding *What?*

Combinations of top 8 important features

**Frequent Combinations**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Errors &lt;=1 &amp; Reallocated Sectors &lt;=5</td>
<td>H</td>
</tr>
<tr>
<td>Data Errors &lt;=1 &amp; WAF &lt;= 1</td>
<td>H</td>
</tr>
<tr>
<td>Media Wear-out = 100 &amp; WAF &lt;= 1</td>
<td>H</td>
</tr>
<tr>
<td>Avg. SSD space &gt;= 10</td>
<td>F</td>
</tr>
</tbody>
</table>
Understanding When?

What is the duration between detection and failure? Signatures characterize SSD survivability?
Understanding *When*?

![Graph showing CDF(x) and Time To Fail (months)]

- 50% of failures occur after > 4 months.
- Sufficient time to intervene.
Understanding *When?*

Early failures (<1 month): Rules include symptoms and their thresholds

50% of failures

Late failures: Rules contain only workload factors

Time To Fail (months)

CDF(x)
Understanding SSD Failures – An analogy

Symptoms

Factors

Observation based causal estimate
Probabilistic causal models and Pearl’s do-calculus
Understanding *Why*?

**What** factors impact SSD reliability? is their magnitude of impact?
Understanding Why?

SSD model and symptoms have direct impact

Workload impacts failures through media wearout
Concluding Remarks

• SSD Failures in the field
• Factors -> Symptoms -> Failures
• Important Symptoms: Data Errors and Reallocated Sectors
  • High intensity and rapid progression fails early
• Important factors: NAND Writes, Total Reads and Writes, etc.
• Direct impact: SSD Model and Symptoms
• Indirect impact: Workload through wear-out
• Future direction: prediction and control