Hard Drive Reliability in Data Centers Using Spark

By: Alex Morris, Nathan Kozakis, Sam Morris
Background Information

- Backblaze released data on their data center hard drives for the first 3 quarters of 2019
- By 2025 over 175 zettabytes of data expected
  - In 2018 there was 18 zettabytes of data
- Almost all data is stored in a data center
- Current plan is to build more data centers to meet demand
  - But how long can we keep adding new data centers?
  - Moore's Law
2019 Annualized Hard Drive Failure Rates

For drives models in service as of 12/31/2019

<table>
<thead>
<tr>
<th>MFG</th>
<th>Model</th>
<th>Drive Size</th>
<th>Drive Count</th>
<th>Drive Days</th>
<th>Failures</th>
<th>Annualized Failure Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGST</td>
<td>HSM5C4040ALE640</td>
<td>4TB</td>
<td>2,852</td>
<td>1,048,376</td>
<td>17</td>
<td>0.59%</td>
</tr>
<tr>
<td>HGST</td>
<td>HSM5C4040BLE640</td>
<td>4TB</td>
<td>12,746</td>
<td>4,674,986</td>
<td>57</td>
<td>0.45%</td>
</tr>
<tr>
<td>HGST</td>
<td>HUH728080ALE600</td>
<td>8TB</td>
<td>1,000</td>
<td>368,454</td>
<td>8</td>
<td>0.79%</td>
</tr>
<tr>
<td>HGST</td>
<td>HUH721212ALE600</td>
<td>12TB</td>
<td>1,560</td>
<td>327,080</td>
<td>5</td>
<td>0.56%</td>
</tr>
<tr>
<td>HGST</td>
<td>HUH721212ALN604</td>
<td>12TB</td>
<td>10,859</td>
<td>2,848,164</td>
<td>31</td>
<td>0.40%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST4000DM000</td>
<td>4TB</td>
<td>19,211</td>
<td>7,325,582</td>
<td>402</td>
<td>2.00%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST6000DX000</td>
<td>6TB</td>
<td>886</td>
<td>379,894</td>
<td>10</td>
<td>0.96%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST8000DM002</td>
<td>8TB</td>
<td>9,809</td>
<td>3,591,167</td>
<td>125</td>
<td>1.27%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST8000MN0055</td>
<td>8TB</td>
<td>14,447</td>
<td>5,242,891</td>
<td>225</td>
<td>1.57%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST10000NM0086</td>
<td>10TB</td>
<td>1,200</td>
<td>437,259</td>
<td>12</td>
<td>1.00%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST12000NM0007</td>
<td>12TB</td>
<td>37,004</td>
<td>12,721,076</td>
<td>1,156</td>
<td>3.32%</td>
</tr>
<tr>
<td>Seagate</td>
<td>ST12000NM0008</td>
<td>12TB</td>
<td>7,215</td>
<td>321,275</td>
<td>10</td>
<td>1.14%</td>
</tr>
<tr>
<td>Toshiba</td>
<td>MD04ABA400V</td>
<td>4TB</td>
<td>99</td>
<td>39,788</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Toshiba</td>
<td>MG07ACA14TA</td>
<td>14TB</td>
<td>3,619</td>
<td>564,829</td>
<td>10</td>
<td>0.65%</td>
</tr>
</tbody>
</table>

Totals 122,507 39,890,821 2,068 1.89%

2019 failure data as calculated by backblaze. Our data only went through September.
The Question: Which is the best hard drive for a datacenter?

- Is failure rate a good statistic?
- Which models last the longest?
  - ST6000DX00 - 6TB - 9 failures - 36351 hours
  - ST4000DM00 - 4TB - 283 failures - 31350 hours
  - ST500LM012 - 0.5TB - 21 failures - 29861 hours
- Is the longest lasting the best?
  - Smaller drives tend to last longer than larger drives
  - Larger drives take longer to replace lost data
  - Larger drives have better capacity density within the datacenter.
  - Does drive size affect data center performance?
Methodology

- Use statistical analysis methodologies to analyze the data set
  - ANOVA - Analysis of Variance
    - Use ANOVA to view within group and between group variance
    - Gives values such as mean square errors as well as sum of square error
    - F statistics can be used to show statistical significance in the variance
  - Failure rates
    - Use standard deviations to view a more normalized curve of failure rates
Methodology

- Spark
  - Utilize Spark to retrieve and process data from the 8gb data set
  - Run statistical analysis programs on Spark itself
  - Spark web application shows time it took to run
Methodology

- Measuring performance
  - Used two different Spark programs
    - One to gather data and write to a csv (Later used with a standalone python program)
    - One to gather data and run statistical analysis
  - Look at difference in run time on Spark versus the combined time to write to a csv and run on python program
    - Python program simply reads through the csv and runs the same ANOVA analysis
    - Running everything on Spark reduces I/O which significantly reduces run time
Performance - Spark Cluster

- 10 computers
- 2 workers per computer
- 8GB of data in HDFS
- Running pyspark

Finding all failures and writing them all to two csv files, one by model and one by serial number, took 8 minutes on this cluster.
Performance - ANOVA

Running ANOVA on the csv of failures (1500 line csv) took milliseconds, but that’s after 8.6 minutes building that csv with the spark cluster.

Gathering data and running ANOVA on the spark cluster without building the CSV and just printing the F statistic took 2.6 minutes.
Performance - current operating hours of working HDDs compared to failed

The spark cluster took 3 minutes to find the maximum operating hours per serial number, then find the average of those per model.

This data gave us an idea of which models might have better or worse statistics than the limited number of failures show. This was helpful when determining if the data from groups with fewer failed models was good.
HDD Capacity’s effect on average lifetime

Larger drives perform worse than smaller drives.

0.5TB hard drives do not appear to follow this trend.

There were fewer 6TB hdds so the trend holds within the error of that group.
ANOVA for failed drives

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>23</td>
<td>115739195491</td>
<td>5032138934</td>
<td>311</td>
</tr>
<tr>
<td>Error</td>
<td>1561</td>
<td>25185821062</td>
<td>16134414</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1584</td>
<td>140925016554</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

With an F statistic of 311, it’s more than clear that the differences between groups are more significant than the differences within groups.