Welcome to the Backblaze

Hard Drive Data and Stats

Since 2013, Backblaze has published statistics and insights based on the hard drives in our data center. You'll find links to those reports below. We also publish the data underlying these reports, so that anyone can reproduce them. You'll find an overview of this data and the download links further down this page.

Hard Drive Reliability Statistics

Hard Drive Stats for Q2 2016	Q3 2016
One Billion Drive Hours and Counting: Q1 2016 Hard Drive Stats	Q2 2016
Hard Drive Reliability Review for 2015	Q1 2016
What Can 49,056 Hard Drives Tell Us? Hard Drive Reliability Stats for Q3 2015	Q4 2015
Hard Drive Reliability Stats for Q2 2015	Q3 2015
Hard Drive Reliability Stats for Q1 2015	Q2 2015
What is the Best Hard Drive?	Q1 2015
Hard Drive Reliability Update	Q3 2014
What Hard Drive Should I Buy?	Q1 2014
How Long Do Disk Drives Last?	Q4 2013

Hard Drive Reviews and Analysis

Enterprise Drives: Fact or Fiction?	December 2013
Hard Drive Temperature – Does it Matter?	May 2014

Hard Drive SMART Stats	November 2014
Our 6 TB Hard Drive Face-Off	December 2014
Reliability Data Set For 41,000 Hard Drives Now Open Source	February 2015
A Look at Backblaze's Toshiba Hard Drives	May 2015
HGST 8TB Drives – Helium Makes Them Fly	June 2015

Overview of the Hard Drive Data

Each day in the Backblaze data center, we take a snapshot of each operational hard drive. This snapshot includes basic drive information along with the S.M.A.R.T. statistics reported by that drive. The daily snapshot of one drive is one record or row of data. All of the drive snapshots for a given day are collected into a file consisting of a row for each active hard drive. The format of this file is a "csv" (Comma Separated Values) file. Each day this file is named in the format YYYY-MM-DD.csv, for example, 2013-04-10.csv.

The first row of the each file contains the column names, the remaining rows are the actual data. The columns are as follows:

- **Date** – The date of the file in yyyy-mm-dd format.
- **Serial Number** – The manufacturer-assigned serial number of the drive.
- **Model** – The manufacturer-assigned model number of the drive.
- **Capacity** – The drive capacity in bytes.
- **Failure** – Contains a “0” if the drive is OK. Contains a “1” if this is the last day the drive was operational before failing.
- **2013-2014 SMART Stats** – 80 columns of data, that are the Raw and Normalized values for 40 different SMART stats as reported by the given drive. Each value is the number reported by the drive.
- **2015 SMART Stats** – 90 columns of data, that are the Raw and Normalized values for 45 different SMART stats as reported by the given drive. Each value is the number reported by the drive.

Helpful Hints and Caveats

The 2015 versus the 2013-14 Schema

For 2015 we began tracking 5 additional SMART attributes, meaning there are 10 additional fields in the 2015 schema. The new SMART attributes being collected for 2015 are the raw and normalized values for: smart_22, smart_220, smart_222, smart_224, and smart_226.

Blank Fields

The daily snapshots record the SMART stats information reported by the drive. Since most drives do not report values for all SMART stats, there are blank fields in every record. Also, different drives may report different stats based on their model and/or manufacturer.

Inconsistent Fields

Reported stats for the same SMART stat can vary in meaning based on the drive manufacturer and the drive model. Make sure you are comparing apples-to-apples as drive manufacturers don't generally disclose what their specific numbers mean.

Out-of-Bounds Values

The values in the files are the values reported by the drives. Sometimes, those values are out of whack. For example, in a few cases the RAW value of SMART 9 (Drive life in hours) reported a value that would make a drive 10+ years old, which was not possible. In other words, it's a good idea to have bounds checks when you process the data.

The Number of Drives Will Change

When a drive fails, the "Failure" field is set to "1" on the day it fails. The next day, the drive is removed from the list and is no longer counted, reducing the overall number of drives. On the other hand, new drives are added on a regular basis increasing the overall number of drives. In other words, count the number of drives each day.

Computing Drive Days

Each day a drive is listed in a daily snapshot file it counts as one drive day. For example, if there are 35,000 drives listed in a daily snapshot file that equals 35,000 drive days. In the docs.zip file you can download below, you'll find a PDF file named "computing_failure_rates.pdf" which describes how we compute drive days, drive years, and drive failures rates.

Drive Age

As noted, the RAW value of SMART 9 is the number of hours a drive has been in service up to that point. To determine the drive's age in days, you divide the reported number by 24.

How You Can Use the Data

You can download and use this data for free for your own purpose, all we ask is three things 1) you cite Backblaze as the source if you use the data, 2) you accept that you are solely responsible for how you use the data, and 3) you do not sell this data to anyone, it is free.

The Raw Hard Drive Test Data

Hopefully the information above has provided you with the information you need to access and use the hard drive data we have collected. Here is the data:

2013 Docs and Data

2013_docs.zip - 77.2 KB ZIP file, 111 KB on disk, 5 files.

2013_data.zip - 81.2 MB ZIP file, 775.2 MB on disk, 266 files.

2014 Docs and Data

2014_docs.zip – 77.5 KB ZIP file, 115 KB on disk, 5 files.

2014_data.zip – 587.3 MB ZIP file, 3.02 GB on disk, 365 files.

2015 Docs and Data

2015_docs.zip – 77.6 KB ZIP file, 115 KB on disk, 5 files.

2015_data.zip – 842.9 MB ZIP file, 4.5 GB on disk, 365 files.

2016 – Docs and Data

Q1_2016_docs.zip – 78 KB ZIP file, 106 KB on disk, 5 files

Q1_2016_data.zip – 270.3 MB ZIP file, 1.42 GB on disk, 92 files.

Q2_2016_docs.zip – 78 KB ZIP file, 106 KB on disk, 5 files

Q2_2016_data.zip – 292.4 MB ZIP file, 1.55 GB on disk, 91 files.