<https://www.backblaze.com/b2/hard-drive-test-data.html>

**Hard Drive to Buy: Drive Definitions & Specs**

<https://www.backblaze.com/blog/what-hard-drive-should-i-buy/>

**Hard drives last for X time**

<https://www.backblaze.com/blog/how-long-do-disk-drives-last/>

**How to calculate data durability**

<https://www.backblaze.com/blog/cloud-storage-durability/>

**Helium filled hard drives**

<https://www.backblaze.com/blog/helium-filled-hard-drive-failure-rates/>

#### **In 2007 we began Backblaze to backup our friends' computers online. We had a mission to make backup software elegant and provide complete peace of mind. Today we have expanded to protect business and offer our B2 Cloud Storage to developers and IT.**

## **Personal Backup**

Never lose a file again. Backup an unlimited amount of data. Unlimited bandwidth. Unlimited file size. You can download a free restore or have a USB hard drive or USB key Fedex to you. Backblaze is just $5/month.

## **Business Backup**

The same award winning backup service is available to business for the price of $50/year/computer. Unlimited Data. No throttling. Backup all your companies laptop and desktop computers to the cloud.

## **B2 Cloud Storage**

Developers and IT people can now build their own cloud storage solutions with Backblaze. Using our storage pods you can build applications, host content or build your very own backup service on B2. Backblaze offers a huge discount over other cloud providers at $0.005/GB/Month for storage.

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.
* **2015-2017 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.
* **2018 (Q1) SMART Stats** – 100 columns of data, that are the Raw and Normalized values for 50 different SMART stats as reported by the given drive. Each value is the number reported by the drive.
* **2018 (Q2) SMART Stats** – 104 columns of data, that are the Raw and Normalized values for 52 different SMART stats as reported by the given drive. Each value is the number reported by the drive.
* **2018 (Q4) SMART Stats** – 124 columns of data, that are the Raw and Normalized values for 62 different SMART stats as reported by the given drive. Each value is the number reported by the drive.

### Helpful Hints and Caveats

#### **The Q4 2018 versus Q2 2018 Schema**

For Q4 2018 we began tracking 10 additional SMART attributes, meaning there are 20 additional fields in the Q4 2018 schema. The new SMART attributes being collected for Q4 2018 are the raw and normalized values for: smart\_16, smart\_17, smart\_168, smart\_170, smart\_173, smart\_174, smart\_218, smart\_231, smart\_232, and smart\_233. 

#### **The Q2 2018 versus Q1 2018 Schema**

For Q2 2018 we began tracking 2 additional SMART attributes, meaning there are 4 additional fields in the Q2 2018 schema. The new SMART attributes being collected for Q2 2018 are the raw and normalized values for: smart\_23 and smart\_24. 

#### **The Q1 2018 versus the 2015-2017 Schema**

For Q1 2018 we began tracking 5 additional SMART attributes, meaning there are 10 additional fields in the Q1 2018 schema. The new SMART attributes being collected for Q1 2018 are the raw and normalized values for: smart\_177, smart\_179, smart\_181, smart\_182, and smart\_235. 

#### **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. 

#### **Missing Data**

In Q1 2017, the SMART stats for some of the hard drives during the period of January 28 thru 31, 2017 were not recorded. On February 1, 2017 complete reporting resumed. While this had no effect on the how we use the data, it may, depending on how you use the data, effect your efforts.

### 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.