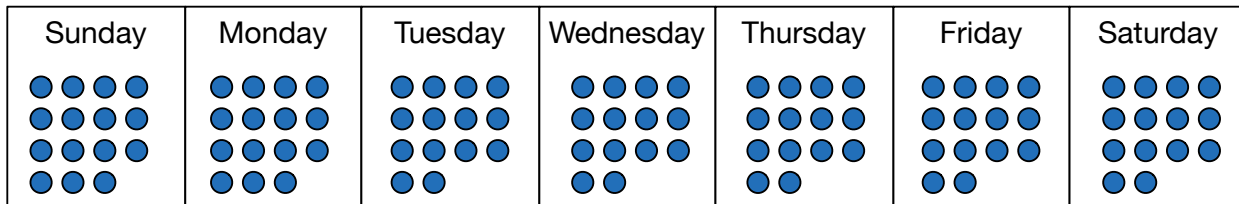


How To Compute Failure Rates with the Backblaze Hard Drive Stats

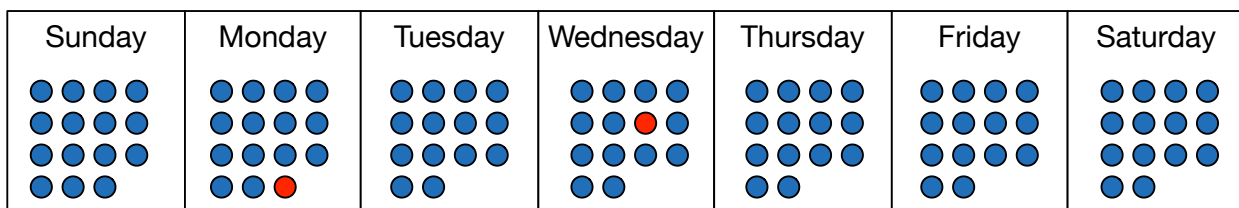
A failure rate is the fraction of drives that have failed over a given time span. Let's start by calculating a daily failure rate, which will tell us what fraction of drives fail each day. We'll start by counting "drive days" and "failures".

To count drive days, we'll take a look every day and see how many drives are running. As an example, we'll take a look at a week in the life of a (small) data center which starts out with 15 running hard drives:



Each of the blue dots represents a drive running on a given day. On Sunday and Monday, there are 15 drives running. Then one goes away, and from Tuesday through Saturday there are 14 drives each day. Adding them up we get $15 + 15 + 14 + 14 + 14 + 14 + 14 = 100$. That's 100 drive days.

Now, let's look at drive failures. One drive failed on Monday and was not replaced. Then one died on Wednesday and was promptly replaced. The red dots indicate the drive failures:



So we have 2 drive failures in 100 drive days of operation. To get the daily failure rate, you simply divide. 2 divided by 100 is 0.02, or 2%. The daily failure rate is 2%.

The annual failure rate is the daily failure rate multiplied by 365. If we had a full year made of weeks like the one above, the annual failure rate would be 730%.

Annual failure rates **can** be higher than 100%. Let's think this through. Say we keep 100 drives running in our data center at all times, replacing drives immediately when they fail. At a daily failure rate of 2%, that means 2 drives fail each day, and after a year 730 drives will have died. We can have an annual failure rate above 100% if drives last

How To Compute Failure Rates with the Backblaze Hard Drive Stats

less than a year on average.

Computing failure rates from the data that Backblaze has released is a matter of counting drive days and counting failures. Each row in each daily drive stats file is one drive day. Each failure is marked with a "1" in the failure column. Once a drive has failed, it is removed from subsequent daily drive stats files.

To get the daily failure rate of drives in the Backblaze data center, you can take the number of failures counted in a given group of daily stats files, and divide by the number of rows in the same group of daily stats files.