DATACORE

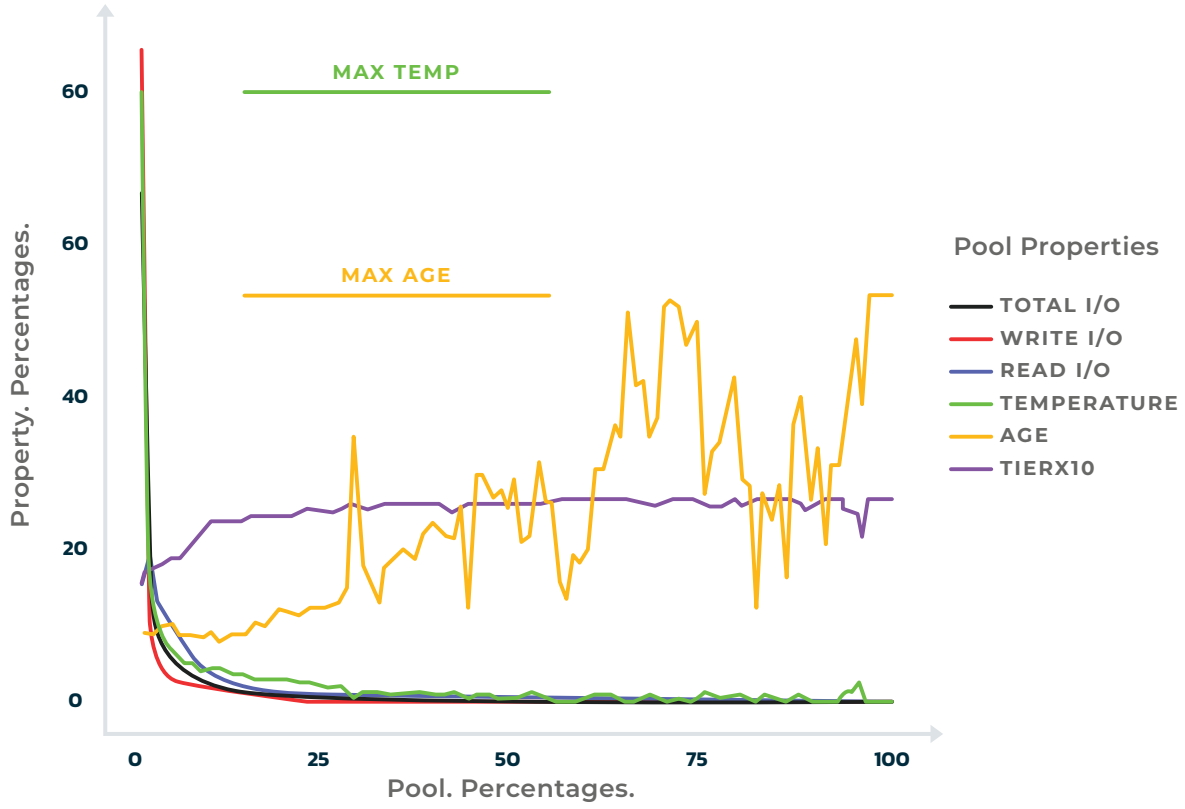
# How data temperature drives data placement decisions – and what to do about it

NEWS FLASH

**We're spending too much on dormant data. How do we know? Telemetry.**

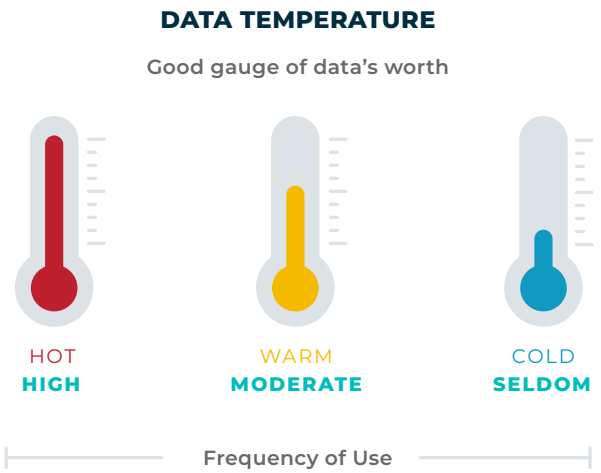
Have a peek at the traces below. It mirrors the characteristics of numerous organizations. The vertical axis reflects the temperature of the data. The most frequently accessed records appear hotter (higher in the chart) mostly at the left edge. The temperature slopes off quite quickly at about 20%. Translation – less than 20% of the data in this storage pool is actively being read, written or updated. The other 80% is mostly idle. Just chilling.

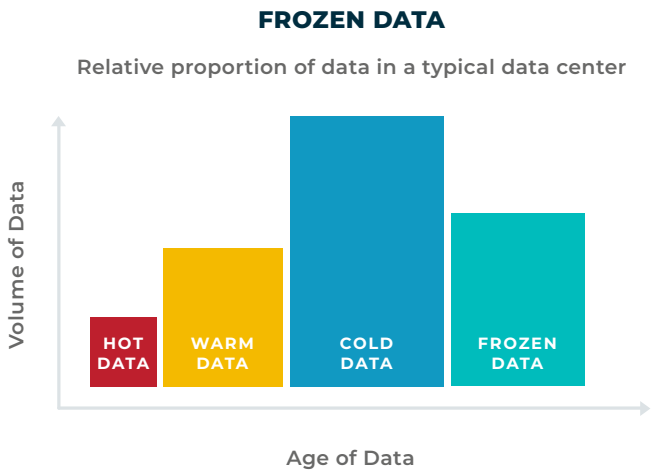
The squiggly orange graph tells us the age. Older data appears higher in the chart, with the maximum holding period occurring around the 40 line. Anything below that is younger data. This reveals that older data is seldom accessed – no surprise here. But the same is true for relatively recent arrivals.



The chart can be expressed more clearly using familiar representations. The thermometer depiction below works in Fahrenheit and Celsius alike. High frequency access = **Hot**; moderate = **Warm**; seldom touched = **Cold**.

Actually, archive data is sometimes referred to as Frozen. The relative proportion in each temperature range for a typical data center looks something like this:



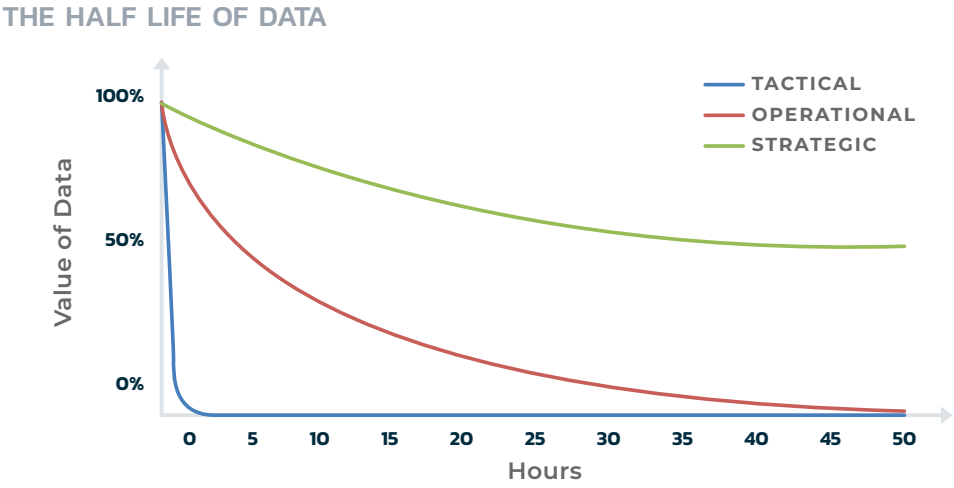




Truth is, the absolute amount of hot data remains relatively constant when the business is running at steady state. But the cooler data continues to accumulate. So the proportion of cold over hot shifts over time. Which implies that your capacity growth will be more pronounced on secondary storage.

## HOT NOT FOR LONG

Nucleus Research[1] referred to the dramatic falloff in access frequency as the half-life of data. Analogous to the rate of radioactive decay, except instead of years for uranium, we're talking hours, even minutes or seconds for data. Said differently, within a few ticks of the clock, data goes from being super popular, to no one cares.

Well they may care later on, but for all practical purposes no one touches it after a short frenzy of activity.



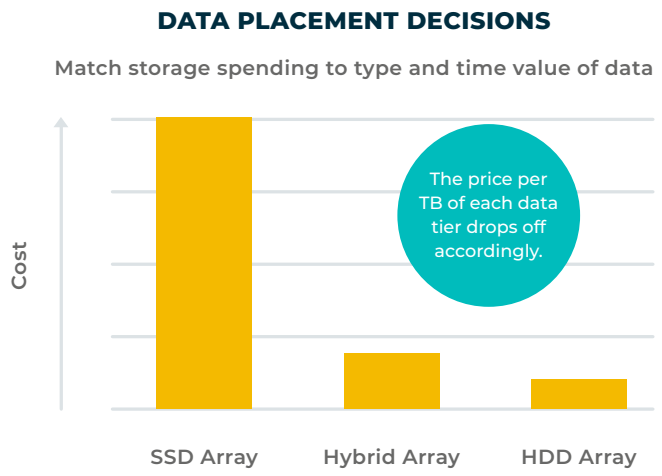
Yet, most of this relatively dormant data sits on expensive storage for a long time because no one has the time to move it elsewhere. Can you see where I'm going with this?

You spend oodles of money getting super-fast storage for the frenzy, then fill it up with idle information. That in turn slows down your next access, and forces you to buy more expensive capacity. Cool business model if you sell flash cards. Not so nice on a tight IT budget.

Of course, "the rate of decline in data value can be correlated to the tempo of its use in a company's decision making processes, which ranges from tactical to strategic to operational.[1]"

## TIERS OF STORAGE – A STEP IN THE RIGHT DIRECTION

Storage manufacturers offer specific models tailored for different data temperatures. The All Flash Arrays (AFAs) cater to the red-hot extreme, hybrid arrays target some compromise, with high capacity / low-cost HDD arrays covering the colder range. Similar categories are available on-premises and in the cloud. The price per TB of each tier drops off accordingly. Gartner estimates that there's about a 6X price difference between Tier 1 AFAs and Tier 2 hybrids, so it really pays off to evacuate data out of the premium storage as quickly as you can.




Source: Gartner Research, Storage Utilization Is Decreasing, Stop Wasting Money
Published: 25 May 2018 ID: G00347626

## MOVING DATA BETWEEN TIERS – NOT SO EASY

Now put on your operational hats and think through how hard it is to move data from your AFA to cheaper storage. Where would you start? How often would you have to do it? How long would it take you each time, and how disruptive would it be on you and the users?

Confronted with those obstacles it's no wonder that most people give up. You could make the rational argument that it's not worth it. That is, until you had to justify buying another pair of AFAs. Two because you wouldn't want to trust your most valuable data to a single point of failure.

Why bother with tiers at all? Why not stick with one type of storage and save yourself from having to move it? Because it gets really expensive – that's why!

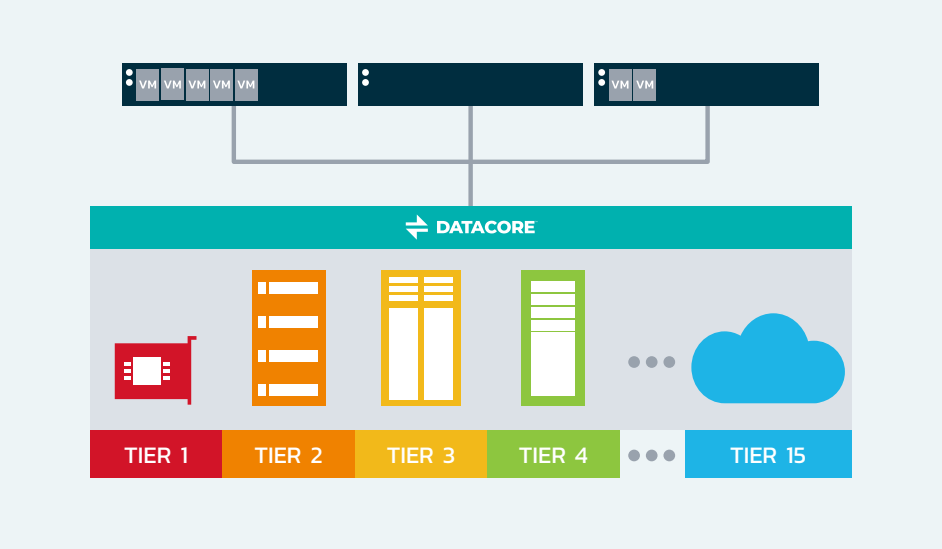
## AUTO-TIERING REQUIRED

Detecting data temperature gradients is really not something any human can be expected to do. Much less take responsibility for moving it to the right bucket. That's a problem machine learning (ML) can help us with.

Automated storage tiering techniques in hybrid arrays track access frequencies and migrate data in the background between SSDs and HDDs. This form of ML works fine as long as all of your data fits in one box. But buying a big enough box in anticipation of your future needs has its own financial challenges. And trusting one array with all of your jewels is equally irresponsible. What then?

## HANDS-FREE DATA PLACEMENT ACROSS BOX BOUNDARIES

Extending auto-tiering techniques across arrays and even into the cloud is the obvious next step. DataCore customers take good advantage of this capability every day, with artificial intelligence (AI) silently making room on pricey AFAs for critical hot data. It's one of the differentiating aspects in DataCore's software-defined storage services.




This advanced level of storage virtualization has the added benefit that models of storage chosen for each tier become interchangeable. Consider how this alone would strengthen your negotiating position during the next round of capacity expansion. If your current supplier tries to hold you hostage, you shop around for another one without fear of disruption.

## WHAT TO DO NEXT

Now that you have a better conceptual understanding for the dynamic migration of data from premium storage to cheaper housing and the savings this can bring year after year, take the next steps.

- Schedule a live demonstration of auto-tiering from a DataCore solution architect to see it in action.
- Watch this webinar and learn more from one of our experts on how data temperature influences data placement decisions.

### Choice Drives Down Costs

"The biggest benefit with DataCore is not being locked into a single vendor for storage. This gives us the flexibility to purchase what we need – high performance SSD or low cost archival storage at affordable prices.

— Ryan Tetzlaff, IT Manager, Open Systems International Inc

Source: Ryan Tetzlaff, IT Manager, Open Systems International Inc
Validated Published: Jan. 22, 2016 TVID: 66E-73A-AF7
DATACORE TechValidate

[1] Nucleus Research, "Measuring the Half Life of Data", 2012

0719

REQUEST A DEMO

### Discover the Ultimate Flexibility of DataCore Software

DataCore software-defined & hyperconverged storage solutions reduce costs, eliminate vendor lock-in, and deliver ultimate flexibility in how organizations manage, build and modernize their storage infrastructures.

**See why over 10,000 customers recognize DataCore Software as the most flexible software-defined storage platform and visit www.datacore.com.**

DATACORE