# When Should I Implement an Active Archive?

By Floyd Christofferson, Chairman of the Board of the Active Archive Alliance

There are a pair of recurring and contradictory themes I frequently hear from storage IT managers when it comes to managing their data and storage infrastructure. On the one hand, they'll say that an archive is viewed a low priority project in their company. When IT budgets are under pressure, prioritizing resources to manage 'old' data seems like a difficult choice.

But then in the next breath they will talk about how their existing storage capacity is filling up or approaching end of life. Some are in a situation where there are multiple user communities in their organization with differing performance requirements, which only makes this problem worse. Often these conflicting use cases lead to the creation of storage silos that increase management burden and inhibit collaboration. In other cases, it means that different stakeholders are faced with a compromise that is less than optimal for each.

These urgent, short-term needs are often perceived as a much higher budget priority than building an archive. The squeaky wheel is going to get the most attention, particularly when IT budgets are under pressure to remain flat, even as data volumes are increasing. IT managers complain that this leaves them with a Hobson's choice where they must set aside long-term data management needs for short term storage infrastructure reality. In many organizations this often means simply deleting data, or exporting it into an unmanaged or offline collection.

What is missing in this conversation about "archive" is the word "active." That is, in an analog world, an archive is where information is put to rest. It is put on the shelf and there it stays, gathering dust. But in a digital world, all information should be online and accessible to satisfy both the immediate and long-term needs. This is particularly important as more and more companies are seeking to extract added value from their legacy data, whether by monetizing it through repurposing to other uses or by gaining business intelligence from Big Data analytics.

Given that the data being stored typically lives longer than the storage systems that house it today, the long-term problem is not only accommodating the growth of the data, but also figuring out when and how to migrate to next generation platforms. At its core, this is a data management problem, not a storage problem.

There is no shortage of studies from storage vendors, universities and even the National Science Foundation that show how primary storage infrastructures contain a high proportion of infrequently accessed, or cool/cold data. But because the data needs to be online and available, the choice defaults back to the squeaky wheel of short-term priorities. So the most expensive storage is filling up with data that has often not been touched for months or years. As a result, power, cooling, and infrastructure management costs grow.  Data utilization analysis says that infrastructures are solving for the highest performance use case, but often for data that is rarely touched.

An active archive strategy can bridge these conflicting priorities. By enabling hot and cold data to always be online and accessible even across different storage technologies and different performance requirements, IT managers do not need to decide whether to prioritize expensive high performance infrastructure over long-term retention needs. Both are solved with the same strategy. Moreover, by enabling data to spread out across a fabric of multiple storage technologies and price points, the high-performance portion can be reduced significantly without adding complexity, effecting SLAs or limiting access. A flexible active archive fabric can be scaled up or down in multiple ways to accommodate these various priorities with one common environment.

There are numerous choices for how to do this, but the common denominator is that all data is available all the time, whether cold or hot. Different performance requirements can be accommodated in a shared open fabric. Future technologies or use cases become options to plug into this open architecture rather than looming roadblocks. Vendor or technology lock-in can be minimized and the focus then can be on what to do with the data, not how to manage more isolated pockets of data stranded in silos. The active archive strategy helps IT managers future-proof their infrastructure.

The good news is that this is not a one-size fits all proposition. The Active Archive Alliance includes vendors with a broad range of options for implementing such a strategy so that the storage architecture can be tuned to fit the workflow and the data, not the other way around. These options include everything from flash to disk to tape, as well as object storage and conventional POSIX file systems. In many cases all of the above are virtualized into the same fabric, with individual tiers optimized for the particular workflow of the customer.

Thus, the decision is not about deciding when the right time is to build an archive to house old data. The decision is about how to manage all the data with an integrated strategy to accommodate present and future needs with the most cost effective choices.

In a talk to Harvard Business School students, noted Harvard psychology professor Ellen Langer addressed how companies often approach technology choices: "Mindlessness is the application of yesterday's business solutions to today's problems. Mindfulness is attunement to today's demands to avoid tomorrow's difficulties…."

An active archive strategy to data management enables companies to evolve their storage choices, to bridge existing realities with future requirements, and to do it all in a way that reduces the cost and complexity. It is about managing the data, not the storage.