

Block-level Tiering vs File-level Tiering

What's the difference?

Many IT Leaders Confuse Storage with Data Management

As data grows, storage costs are escalating. It is easy to think the solution is more efficient storage. But the real cause of storage costs is poor data management. Over 70% of data is cold and has not been accessed in months, yet it sits on expensive storage and consumes the same backup resources as hot data. As a result, storage costs are rising, backups are slow, recovery is unreliable, and the sheer bulk of this data makes it difficult to leverage new options like Flash and Cloud. Sounds familiar?

Managing Cold Data Can Save Millions

Since the bulk of data is cold, finding and tiering cold data can save millions, since it offloads cold data from expensive storage and backups. Tiering has been a solution for years, but has disrupted end users and applications. Tiering needs to be frictionless with no disruption to users and applications. Even after data is moved, it needs to be accessed by users and applications exactly the same way as before the move.

Did you know that the way tiering is done can significantly change your actual savings and affect your options to access the cold data? Cold data can be tiered at the block level or at the file level, and there are many significant differences between these two approaches.

Block-Level Tiering is a Storage-Centric Solution

Block-level tiering was first introduced as a technique within a storage array to make the storage box more efficient by leveraging a mix of technologies such as more expensive SAS disks as well as cheaper SATA disks. As the name implies, block-level tiering moves blocks between the various tiers to increase performance while reducing costs. Hot blocks and metadata are typically kept in the higher, faster, and more expensive storage tiers while cold blocks are migrated to lower, less expensive tiers.

Storage vendors are now using block-level tiering to move data out of the file server and into an object or cloud tier. All file access must be done through the original file server. The moved blocks cannot be directly accessed from their new location, such as the cloud, because they are meaningless without all the other data blocks and the file context and attributes (the file's metadata).

Many operations like 3rd party backup software that operate at a file level require that the blocks of the file be brought back (re-hydrated) before they can be backed up. This defeats the purpose of tiering to an external storage, such as the cloud, since no space is saved on the original file server.

If the file server is to be end-of-life'd, all blocks tiered to the cloud from it must first be rehydrated, then migrated to the new file server. Given that there will not likely be enough space on the original file server to re-hydrate all of the blocks moved over its lifetime, the only way to end-of-life the file server without losing data will be to iteratively rehydrate a small number of files at a time, migrate them to the new file server, and repeat. Once migrated, the rehydrated blocks will need to be tiered to the cloud from the new file server. This creates unnecessary egress costs. One way to resolve this would be to reserve the space left by all cold blocks tiered to the cloud, but this defeats the very purpose of tiering cold blocks.

All of these limitations dramatically impact the actual savings you can receive from block-level tiering to just some storage efficiency. You do not get the full benefit of eliminating cold data from storage, backups, migrations, and you get locked into the storage device.

File-Level Tiering is a Data Management Solution

File-tiering is a more advanced technology and is standards-based. File-level tiering means the file along with all its metadata moves to the new tier. Whether you have NTFS extended attributes or Posix ACLS you need the ability to move the file and all of its associated metadata with high fidelity and rehydrate it back into its exact original form if needed. Moving just the file is not enough. Many applications rely on attributes of the file to operate. The file system imposes access control through basic and extended attributes. File-level tiering maintains full file fidelity and preserves all the attributes and metadata along with the file at each tier.

The savings is more significant since the entire file is preserved. You should be able to access it directly from the target storage and yet be able to return it to the source storage exactly as it was before. Also, by transparently moving the entire file, all other operations such as 3rd party backup applications and migrations can be done without rehydrating the data – thus maximizing the full savings.

Summary of Differences:

	Block-level Tiering	File-level Tiering	Impact
Approach	Proprietary solution that moves cold blocks of a file without the metadata	Standards-based solution that moves the entire file with metadata	Block-based tiering creates proprietary lock-in, file-based tiering does not
Backup Savings	Most 3rd party backup software will require data to be rehydrated, eliminating space savings	3rd party backup solutions can access without rehydration, so backup savings are maximized	Rehydration erases potential backup savings of block-based tiering
Migrations, End-of-life of File Server	All data needs to be brought back to the original tier before it can be migrated to another file server – since there may not be enough space, this is slow, iterative, laborious	Shrinks migration footprint and costs since cold data does not need to be rehydrated	File-based tiering speeds up migrations and shrinks footprint, block-based tiering does not
Space Savings	Limited as space needs to be reserved for rehydration	Maximizes space savings since access does not require rehydration	File-based tiering typically gives 3x more savings than block-based
Fast Access to Moved Data	One argument for block-based tiering is that recalls can be fast because just the blocks that are accessed can be retrieved	File-based tiering can also do intelligent reads and bring back the needed blocks first	File-based tiering can provide the same speed of access without the limitations
Data Access from Lower Tiers	Blocks tiered outside the file server (e.g. the cloud) cannot be directly accessed from the cloud/lower tiers	Preserves native access to the file with all its security intact at each tier	File-based tiering enables direct access of the file from the target storage device (e.g. cloud) and from the source
Lock-in	Creates both device lock-in and vendor lock-in	No lock-in to the storage devices or data management vendor	File-based tiering done right maximizes savings and flexibility

Easier for Vendors to Build Block-Level

Although file and block level tiering both sound the same at the high level, the differences between the two are

significant. Block-level tiering is a storage efficiency technique that is primarily focused on lowering storage cost. File-level tiering is storage agnostic and provides full, secure access to your files from anywhere, and it should be everyone's data management solution. A block-level tiering solution is technically much easier to build, so many data management solutions are starting to leverage this technology. Marketing messages are blurring the line between the two, so it's imperative to understand each vendor's approach.

In addition to the technology differences, cost savings from file-based tiering is typically 3x more than block-based tiering solutions because you do not need to reserve as much space for rehydration during backups, recalls, and migrations. The backup footprint is reduced in file-level tiering because the entire file has been tiered off and does not need to be backed up. While in block-based tiering, the backup typically rehydrates and backs up the full file, so there are no backup savings.

Ask Your Vendor: Block-Level or File-Level Tiering Approach?

Block-based tiering is typically used by storage vendors. Storage tiering, aka pools solutions, use block-based tiering. Only the operating system of the NAS storage knows exactly what blocks were moved, so you can only access the file through the original source. If you decide to end-of-life the device, you must re-hydrate all of the archived data. Given that there will likely not be enough space on the device, this can be a painful, slow, iterative approach. Also, most NAS tiering solutions are limited – for instance, NetApp FabricPools is not available on their disk-based storage, and it can only tier data after the SSD is at least 50% full and less than 63 days old. Similarly, EMC CloudPools needs a policy per cluster, and its licensing costs can be prohibitive if you tier to non-EMC platforms.

Secondary storage vendors also starting to transparently tier data to their device which is moving the storage concerns from Tier 1 to Tier 2 storage. You are now tied to that secondary storage vendor and lose the same flexibility on secondary storage based on need, costs and the direction of your company's infrastructure initiatives as you had on Tier 1 storage. Ultimately, data management is not something that should be left to storage devices. You should be able to freely move from one storage device to another.

Komprise is a data management software solution that tiers and archives data at the file-level and fully preserves file fidelity and standards-based access to your data at each tier. Komprise enables you to freely move data across different vendor storage and clouds without lock-in to either the storage or to Komprise. The solution is analytics-driven, so you can choose what you move, when, and how.

Manage Data, Not Storage — File-Level Tiering Maximizes Savings without Lock-In

The first response to the exponential data growth should not be to call your storage vendor. As discussed, file-level tiering fully preserves file access at each tier by keeping the metadata and file attributes along with the file no matter where it lives, even on object storage and cloud. This maximizes space savings by eliminating the need to rehydrate data for common operations such as data access, backups, and migration which significantly shrinks your storage footprint. File-level tiering provides up to 3x greater savings than block-level tiering because it not only reduces cold files on the primary tier, it also shrinks backup footprint without rehydration, and it shrinks DR footprint without rehydration.

File-level tiering, when done correctly, is a storage agnostic, data management solution that puts you in control of your data without lock-in to either the storage vendor or the data management solution itself. Since block-level tiering is a storage-centric approach that creates device and vendor lock-in, reach out to Komprise to provide compelling data literacy and become proactive in your data management.

To understand how Komprise leverages file-level tiering with its patented Transparent Move Technology (TMT™), go to komprise.com/product



Komprise, Inc.
1901 S. Bascom Ave. Suite 400
Campbell, CA 95008
United States

For more information:
Call: 1-888-995-0290
Email: info@komprise.com
Visit: komprise.com

For media requests email
marketing@komprise.com
©2019 Komprise, Inc. All rights reserved.