



IT Knowledge • Business Results

Intelligent Tiered Storage

Focus on Compellent In-the-Box Tiering

By Tony Asaro
Senior Analyst

March 2006

Intelligent Tiered Storage

Not all data is created equal. While this is true, it is easier to store everything the same way, on the same class of storage system with the same protection levels. However, the cost associated with this approach can be substantial. The task of classifying data is somewhat daunting and requires time that most companies do not have. However, there are methods and technologies that can enable companies to implement intelligent tiered storage without requiring upfront classification of data. This report discusses “in-the-box tiering”, which provides policy-driven data movement within a storage system with a specific focus on Compellent.

In-the-box tiering is a combination of two things, a storage system that has different classes of hard disk drives (e.g. FC and SATA), and the intelligence to move data between these different storage classes. The intelligent data movement should be an online non-disruptive process that is transparent to the users and applications. There is a great deal of messaging around in-the-box tiering of storage with various vendors providing this functionality. Compellent in-the-box tiering is unique since they are the only storage system that ESG is aware of that can perform data movement at a block-level and not just at a volume level. This block-level movement becomes essential to the fundamental value proposition of intelligent tiered storage, which we will explore in more detail in this report.

Multiple Storage Tiers

A storage system that supports different drive types is valuable in of itself. Companies can use a single storage system for different types of applications and data. For example, a company may have a database application that is highly transaction-oriented and they need the fastest disk drives they can get to meet its performance needs. In this case they can use 15,000 RPM FC drives. Additionally, they have an email application that also requires fairly rapid performance but not to the degree of their database application. In this case, they can use 10,000 RPM FC drives. The bulk of their data however is unstructured content including Word documents, Excel spreadsheets, PDF files, etc. This data does not require a great deal of performance but represents 70 percent of their storage capacity requirements. For this data they can use 7,200 RPM SATA drives that are 250 GB in size.

In addition to creating tiers through the use of different drive technologies, different protection levels can be set at each tier. System administrators may want to set RAID levels and snapshot policies at a higher level on Tier One to minimize any risk of data loss. Lower tiers can have lower levels of protection creating a balance between protection and pricing based on the importance of the data to the business.

There are a number of uses and configurations that can be used to provide the right storage class to meet the business need. Not only can multiple tiers of storage lower cost but it can enable customers to do more with their storage that they could not previously cost justify. ESG worked with one customer that implemented an in-the-box tiered storage environment. They were able to meet the performance and protection levels for their Tier One applications. Since only 20 percent of their data was stored on Tier One they were able to cost justify implementing RAID 10 and a snapshot policy of copies on an hourly basis. They were able to add additional applications stored on SATA drives that they could not cost justify using just one tier of FC disks. And they implemented disk-to-disk (D2D) backup using SATA drives, which helped them improve backup performance and reliability.

Policy Driven Migration

One of the objections of implementing an intelligent tiered storage environment is the challenges of classifying data. Taking the time and resource to actually classify data can be a Herculean task for companies. Additionally, the importance of data will change rapidly over a short period of time. Therefore classification can become obsolete quickly.

Another approach to implementing intelligent storage systems is to monitor the frequency of data accessed within the storage system. In other words, if data is not being accessed by a user or application over a long period of time, then it can be assumed that the value of the data has diminished. Based on user defined

policies, the infrequently accessed data can be moved to another tier of storage within the storage system. If at some point, the data is being accessed again the data can be moved back to the higher tier of storage.

In theory this approach is extremely compelling. Companies do not have to go through an arduous data classification process and data can be demoted and promoted as needed. However, the chink in the armor is that most SAN-based storage systems monitor frequency of access at the volume level. This means that if any user or application accesses any data within that volume, it will be classified as tier one. The flaw with this approach is that there may be data within a volume that is constantly being accessed, while other data is never being used. Companies could have 10 percent of data within a volume being accessed and the other 90 percent could be just sitting there.

A more useful approach is to monitor and move data at the block-level. SAN storage systems do not store data in human-readable file formats but rather in machine language. The storage system stores blocks of data within its volumes. By monitoring which blocks are being accessed the process of classification is done at a very granular level. If a block has not been accessed for days, weeks or months, it can be moved to a lower tier of storage. This level of granularity is much more efficient than volume level tiering.

Compellent Data Progression provides both volume level and block-level data movement enabling unique in-the-box tiering that is a policy-driven, online and transparent process.

The Value of Intelligent Tiered Storage

The following are various ways that the Compellent Intelligent Tiered Storage provides value to companies:

Reduce Storage Expenditures

- Reduced initial storage system acquisition costs
- Acquire more capacity for less or the same cost
- Cost justify adding more applications onto the SAN
- Lower the cost of data protection by storing snapshots on less expensive disk drives

Increase Operational Efficiencies

- Improved backup performance and data protection.
- Improve application performance by optimizing Tier One storage through storing only active data on it
- Reduce or eliminate the time and resource required to perform data classification
- Improve protection levels on Tier One storage and keep other tiers at good enough protection levels

Intelligent tiered storage provides value to any company regardless of size or industry. The tiers involve different drive types and protection levels. The intelligence includes data movement and policies to know when and where to move the data. And none of this should impact users, applications or productivity. Compellent provides different drive types, protection levels and its unique Data Progression functionality that moves data at the volume or block level. ESG believes that intelligent tiered storage can significantly reduce cost and enable companies to better leverage their storage investment.

Storage Tiers

In-the-box tiering requires a storage system to have different drive types that it supports. At the very least there should be a high performance tier and a high capacity tier. The Compellent Storage Center storage system provides multiple drive type support including the following:

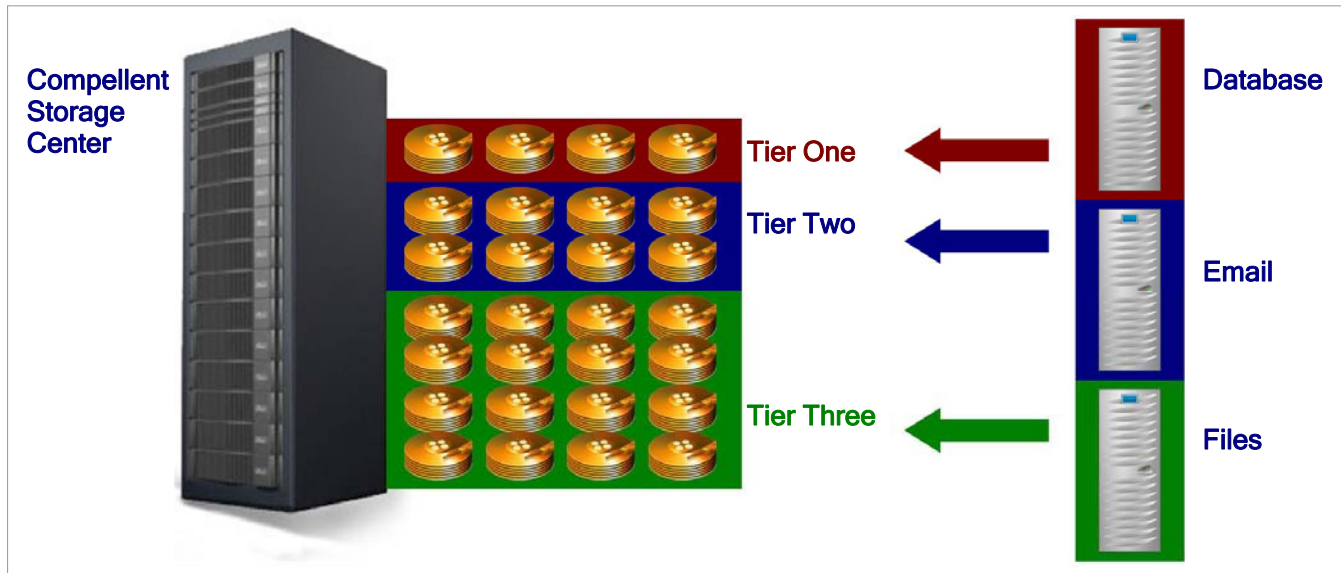
- FC 15,000 RPM 146 GB
- FC 10,000 RPM 146 GB
- FC 10,000 RPM 300 GB
- SATA 7,200 RPM 300 GB
- Mid-Tier FC 7,200 RPM 500 GB

The cost difference between different drive types can be substantial. The cost of storage capacity has multiple variables, differs by storage vendor and product, and cost of capacity is being driven down on an on-going basis by approximately 20 percent annually. The difference in pricing between FC capacity and SATA capacity can be as high as 4 to 1. [For illustrative purposes, this section uses FC drive cost is \$4.00/GB and typical SATA drive cost is \$1.00/GB]

Increasing Capacity

Another way to look at it using tiered storage is leveraging the difference in pricing to buy more storage and still stay within budget. If the budget allowed for \$25,000 of raw capacity then using the above pricing, the company could either acquire 6.25 TB of FC disk or 3 TB of FC disks and 13 TB of SATA disks by using a Tiered Storage approach. That equates to over 2.5 times the amount of capacity for the same price. This much additional capacity can enable more projects and provide greater protection levels.

Figure One: Compellent Intelligent Tiered Storage



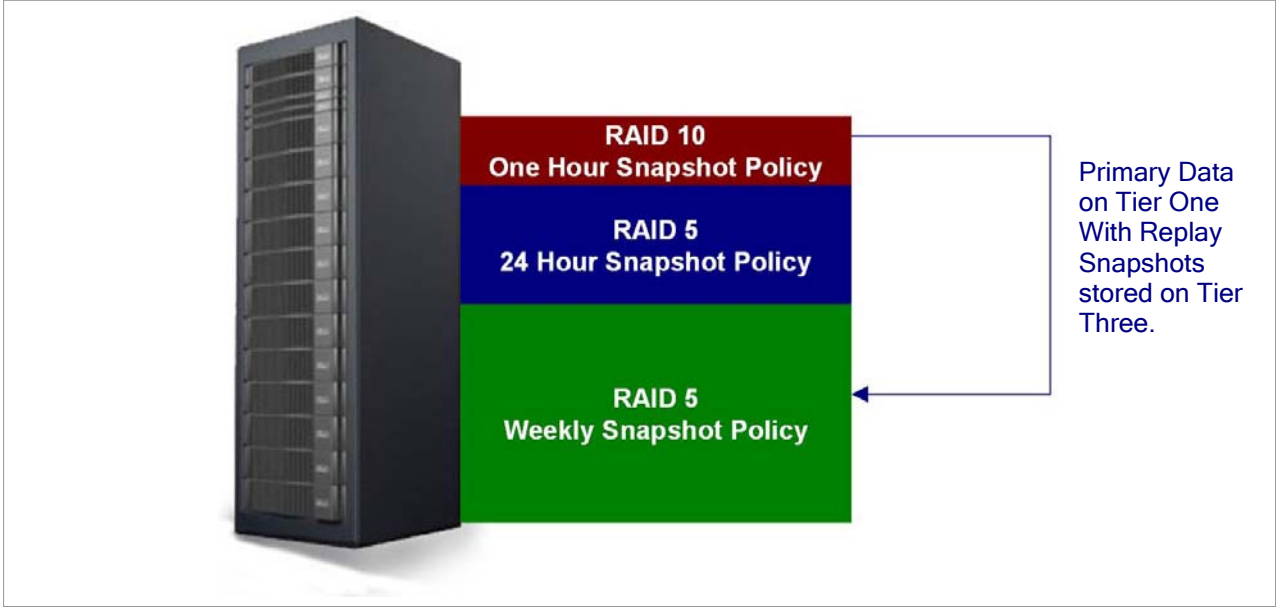
Protection Levels

System administrators can implement varying protection levels on different tiers. It is somewhat intuitive to increase protection levels on Tier One storage. For example, Tier One storage can use RAID 10, which provides a mirrored copy of every write for a high level of protection and it stripes data across all of the drives in the RAID group for optimal performance. Additionally, Tier One storage can have hourly snapshot policies for

greater levels of protection and a relatively granular recovery point objective (RPO). The cost would not be prohibitive using a tiered storage strategy.

Tier Two storage can have RAID 5 using a parity drive versus mirroring to protect against drive failures. A snapshot policy of once a day can be created for a 24 hour RPO. Tier Three can also use RAID 5 but instead have no snapshot policy or one that executes infrequently, such as once a week or month.

Figure Two: Tiered Protection Levels



Additionally, instead of using the Tier One storage to actually store the snapshot copies, Compellent allows primary data running on FC disks to store its snapshot copies onto lower cost SATA drives. Tier One storage is the highest cost/performance tier and should be used just for primary data and not for snapshot copies. The industry average for the amount of capacity that snapshot copies consume is approximately 20 percent of the actual primary data¹. For example 2 TB of primary data would yield 400 GB of snapshot capacity. More snapshots can be created providing greater levels of protection. The positive impact on the business is realized when a file is recovered that otherwise would have been lost because more snapshot copies were made. By enabling more snapshots there will be more recovery points. That one very important file that was deleted or corrupted may have been lost by making fewer snapshot copies.

¹ ESG hasn't done a formal study on this but instead bases this assumption on discussions with a number customers and storage vendors.

Intelligent Data Movement

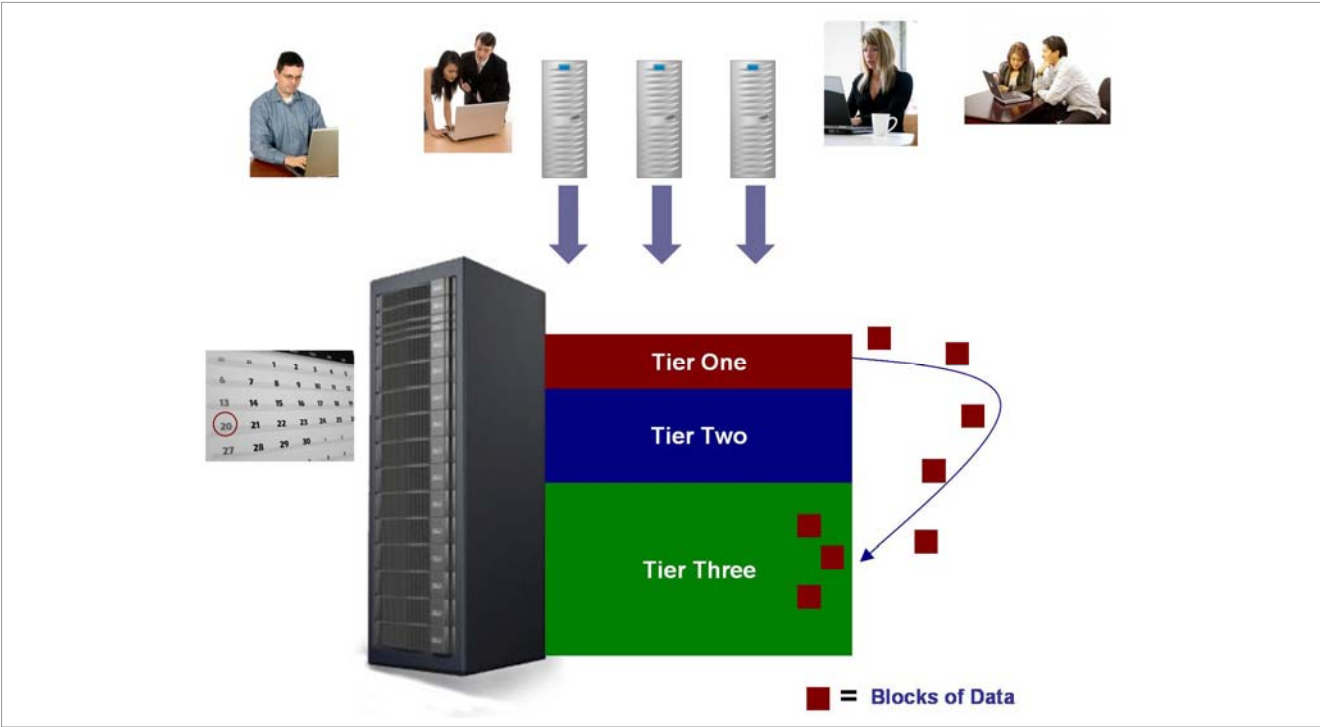
The ability to move data between tiers of storage can have a significant impact on reducing costs. However, in order to make this functionality useful and practical, some or all of the following attributes are required:

- Online and transparent data movement
- Movement between different drive types, sizes and protection levels
- Policy-driven data movement including by frequency of access
- The ability to demote and promote to different tiers of storage based on the needs of the application
- Volume and/or block-level data movement between tiers

Online and Transparent Movement

The ability to move data between tiers should not negatively impact users or applications. The process should be done online with the storage system providing the intelligence to keep track of where data is and making it available as needed. Servers and storage systems should not have to be re-booted.

Figure Three: Online and Transparent Data Movement



Movement between Different Drive Types, Sizes and Protection Levels

The storage system should be able to move data between different drive types and RAID levels seamlessly. In the case of Compellent, they can move data between different drives types, sizes and RAID settings. System administrators can move data from 15,000 RPM FC drives supporting 146 GB of capacity to 7,200 RPM SATA drives supporting 250 GB of data. The FC drives can be set with RAID 10 and the SATA drives with RAID 5. Additionally, Replay snapshot policies will match the tiers. For example, Tier One could have snapshots taken every hour while Tier Three has a weekly snapshot schedule.

Data Movement between Tiers

The notion of automating data movement is still a bit controversial. System administrators may not want their storage system to make decisions on moving data. They would prefer to move this data manually. However, the advantage of using an automated process is that you don't have to classify your data or do in-depth analyses to determine which data gets moved.

Typically data is classified at the point of creation. However, the value of this data changes rapidly over time. The process of re-classifying and then moving data to different tiers is such a cumbersome process that companies typically leave dormant files on Tier One storage indefinitely. Once the data is moved...it usually stays that way for a long time because it is very time consuming to move it again.

One useful metric is measuring the frequency of access, or how often data is being used. If data has not been accessed in weeks, months or even years then the storage system can move data to another tier. Additionally, policies can be made to move data back to a higher tier of storage once it is being accessed again. This can be useful with projects that are cyclical in nature. An end of month process could be moved between tiers storing the data on Tier One during peak times and Tier Three the rest of the time.

Compellent Data Progression provides policy-driven data movement between tiers enabling the demotion and promotion of data. Data Progression is completely and automatically optimized by the Compellent Storage Center storage system based on actual usage patterns. All the administrator has to do is decide what kind of automatic tiered storage they want to "allow" for each volume (FC 10K RAID 10, FC RAID 5, SATA RAID 5, for instance). The Compellent Data Progression technology does the rest. The system analyzes block-level usage and demotes the least used blocks over the last 12-days (default setting). The default for the promotion of data from lower tiers up to higher ones is four days. If data is being accessed over a four day period it gets moved back up to the higher tier. This is a smart strategy since the access of infrequently used data may have anomalous activity but then go dormant again after a short period of time.

Compellent can move data by frequency of access at the volume or block-level. In the case of a project that is active at the end of every month, it may be more efficient to move the data associated with that project at the volume level.

Compellent Data Progression also has the unique ability to move data at the block-level. This approach can be extremely valuable since some data within a volume may never be accessed while other data is always accessed. If frequency of access is based on the level of activity within a volume, even if only one piece of data is being accessed within that volume, it will be deemed as active. For example, if 10 percent of the data within a volume is active and the other 90 percent is inactive, the volume level approach is extremely inefficient. By monitoring and moving data at the block-level, Compellent is able to move the 90 percent of data that is dormant to a lower tier and keep the other 10 percent of data that is active on Tier One.

Compellent Data Progression data movement is unique and valuable in its ability to move data at the volume or the block-level providing compelling in-the-box tiering capability. In-the-box tiering ultimately optimizes performance, better capacity management, reduces cost, provides the ability to implement greater protection levels and enables more applications to be stored on the storage network.

The Value of Intelligent Tiered Storage

The business implications of intelligent tiered storage include economic value, and the creation of levels of efficiency when managing data storage. This section explores the value proposition of intelligent tiered storage in general and with specific focus on the Compellent approach, which includes support of FC and SATA drives and volume and block-level data movement between tiers.

Reduce Storage Expenditures

Reduced Initial Storage System Acquisition Costs. When acquiring a storage system that only supports high-end disk options (e.g. FC drives), the overall cost will be greater than a mixed solution. By combining different drive types companies can get the best of both worlds balancing price and performance without sacrificing anything.

Acquire More Capacity for Less or The Same Cost. Using lower cost, higher density drives can provide more capacity for less or the same cost as just using more expensive disk types. More capacity can enable companies to support more projects and provide greater levels of protection. In the example described on page 3, for the same cost a company can acquire over 2.5 times the capacity with a tiered storage solution versus solely using expensive disk drives.

Cost Justify Adding More Applications Onto The SAN. Adding more capacity also enables adding more applications to the SAN. In many cases, companies will only put the highest priority applications onto SAN-based storage due to the high cost of these systems. However, by adding lower cost storage to the storage system it becomes economically justified.

Lower The Cost of Data Protection by Storing Snapshots on Less Expensive Disk Drives. The goal of intelligent tiered storage is to keep the most important primary data on Tier One. Based on this objective, the notion of keeping copies of primary data on Tier One seems counterintuitive. Compellent can keep snapshot copies of primary data on lower tiers of storage.

Increase Operational Efficiencies

Improved Backup Performance and Data Protection. Keeping only active data on Tier One storage improves backup performance, which results in better protection of data. Reducing the backup window also reduces risk. If a backup takes 16 hours to complete, a file lost during that period of time may not be recoverable. Having less data to back up shrinks the backup window and therefore reduces the risk of data loss, which results in better protection of data. Additionally, both recovery point objectives (RPO) and recovery time objectives (RTO) are improved because there is less primary data to backup and recover.

Improve Application Performance by Optimizing Tier One Storage Through Storing Only Active Data On It. One of the main benefits for using high performance (and expensive) FC drives is for the performance. However, having dormant data stored on FC volumes actually hampers performance. Moving unused data to lower tiers will optimize performance where it is needed (and costs) most.

Reduce or Eliminate the Time and Resource Required to Perform Data Classification. The process of classifying data can be time consuming and resource intensive. The Herculean task of classifying data is often what prevents customers from actually implementing intelligent tiered storage. Using policy-driven data movement software within a storage system essentially eliminates the need to do up front data classification. Significant economic benefits can be gained by using this method without having to do the up front classification.

Improve Protection Levels on Tier One Storage and Keep Other Tiers at Good Enough Protection Levels. By keeping snapshot copies on lower cost storage, system administrators can make more snapshot copies of primary data. This will improve RPO and recoverability of good data. Another way that protection levels can be increased is by using RAID 10 instead of RAID 5 on Tier One. Without intelligent tiered storage the use of RAID 10 may be cost prohibitive. However, with the economics savings derived with intelligent tiered storage, RAID 10 may be a practical and cost efficient choice for Tier One.

ESG's View

The Compellent Data Progression capability provides transparent and online data migration. Like other solutions in the market, Compellent can move data at the volume level. Uniquely, the Compellent Data Progression software can also move data at the block-level. ESG believes that the ability to move data at such a granular level provides superior value when implementing an intelligent tiered storage environment.

Compellent is an extremely competitive solution that companies evaluating SAN storage systems should put on their shortlist. It is one of the most versatile, feature-rich SAN solutions and provides real value to customers. ESG believes that Compellent has crafted an excellent SAN storage system providing value on multiple levels. There are a number of innovative aspects of the Compellent Storage Center solution including its clustered architecture, FC and iSCSI support, thin provisioning (Dynamic Capacity), comprehensive data management software for snapshot copies and remote replication, multiple drive types including FC and SATA disks, and unique policy-driven data movement capability for intelligent, cost effective in-the-box tiering.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.