



**Permabit Albireo
Data Optimization Software**

High Performance Deduplication Software Development Kit

January 2012



Permabit Technology Corporation

Ten Canal Park
Cambridge, MA 02141 USA

Phone: 617.252.9600

FAX: 617.252.9977

info@permabit.com

www.permabit.com

Contents

<i>Executive Summary</i>	3
<i>Market Dynamics</i>	3
<i>The Analysts Agree</i>	4
<i>The Emerging Market for Universal Storage Data Optimization</i>	4
<i>Embedded OEM Data Optimization Software</i>	4
<i>Albireo Data Optimization Software</i>	5
<i>Complete SDK for Easy Development</i>	5
<i>Albireo Architecture</i>	6
<i>Primary Storage</i>	6
<i>Archive/Cloud Storage</i>	7
<i>Backup Storage</i>	7
<i>Universal Storage for Primary, Archive and Backup</i>	7
<i>Block API</i>	7
<i>File API</i>	7
<i>Stream API</i>	7
<i>Flexible Deployment Options</i>	7
<i>Inline</i>	8
<i>Post-Process</i>	8
<i>Parallel</i>	8
<i>High Performance Index</i>	8
<i>Deduplication Throughput</i>	9
<i>Deduplication Latency</i>	9
<i>Deduplication Savings</i>	9
<i>Content Segmentation</i>	10
<i>Albireo Scale-out</i>	10
<i>Why Permabit?</i>	11
<i>Innovation</i>	11
<i>Focus</i>	11
<i>Expertise</i>	11
<i>Conclusion</i>	11
<i>About Permabit</i>	11
<i>Find Out More</i>	12

“The Albireo technology from Permabit will save an OEM 18-24 months getting to market, if they can do it at all. This stuff is so far ahead in its capabilities and performance I can’t see why you would want to do it yourself, unless you already have it baked.”

— Steve Duplessie
 Founder & Sr. Analyst
 Enterprise Strategy Group

Executive Summary

Rampant data growth is the IT industry's single most important issue, affecting budgets, operating costs, floor space and of course capital expenditure through the amount of data created and its associated cost. According to IDC, the amount of electronic data created is expected to reach 35 zettabytes by 2020.¹ The conundrum facing every business is how to afford to store, analyze, manage and house this data. The staggering growth of data and the associated costs of storing it have driven IT organizations to explore data optimization technology as a more efficient means of housing information. In fact another study, this time by ESG, shows data efficiency is the number one priority of storage professionals in IT today.² As a direct result of this customer demand, storage manufacturers have begun to offer data optimization capabilities. Of the available data optimization technologies, data deduplication is the one with the greatest potential to deliver substantial and recurring impact on the cost and manageability of data growth.

The challenge for many of these manufacturers is to develop deduplication technology that can be used without negative impact to storage performance and that can be leveraged across all of their storage platforms (many added via separate acquisitions) while meeting market timing demands and not derailing other high priority R&D projects. Several have made investments in deduplication technology specifically to address problems associated with backup, only to find the resulting solutions are not suitable for primary storage workflows.

Permabit Albireo is the industry's first purpose-built OEM data deduplication software designed to meet the needs of hardware, software, and service providers who wish to expand their existing solutions without negatively impacting existing differentiating capabilities or reducing overall performance. Albireo delivers deduplication at the sub-file level and can be flexibly integrated into existing or next generation storage and platform architectures with the ability to be deployed as an inline, post-process, or parallel solution.

Albireo's flexible architecture provides deduplication which can be seamlessly deployed across **primary, archive and backup** storage. This is done by offering specialized interfaces for block, file, and stream-oriented data formats that broadly address the needs of these storage tiers. Such broad coverage enables OEMs to deliver deduplication in a universal software stack capable of delivering performance and scalability across the widest range of use cases in both the data center and the cloud.

Market Dynamics

Unbridled storage growth is forcing all organizations to re-think their storage strategies in light of flat IT budgets. As a result, the \$25+ billion storage market is facing a sea change that promises to reset the competitive landscape. Technologies that substantially reduce overall storage costs stand to become "must have" requirements. OEMs that provide the highest storage efficiency are poised for the greatest success with top line and bottom line impacts.

Two examples of data optimization techniques are compression and data deduplication. Compression is a common and readily available technology that has been widely used to identify and eliminate recurring patterns within data objects. In contrast, data deduplication identifies and eliminates duplicates across multiple data objects. To illustrate, compression can be used to take 2 MB of data and reduce it to 1 MB, but deduplication can take all the copies of that 2 MB (found across the storage system), and store it as a single copy, consuming just 2 MB of space. In addition, that deduplicated file can then also be compressed so that it only consumes 1 MB. As can be seen from this example, compression is good but deduplication has a much broader impact. The combination of the two technologies provides the highest benefit in terms of storage savings.

Analysts believe that the most substantial opportunities for deduplication lie in end-to-end universal storage. Common data storage building blocks provide a shared basis for the primary, archive and backup storage of the future both in the data center and the cloud. By optimizing data at its origin, maximum storage efficiency is realized. The highly redundant nature of certain data types (including popular productivity files and virtual server images) enables storage optimization to reduce an

¹ IDC Digital Universe Study, sponsored by EMC, IDC, May 2010

² 2011 Storage Infrastructure Spending Trends, ESG, Jan 2011

organization's total storage capacity demands by 5 to 35 times, depending on the data mix. By reducing storage capacity, organizations will also enjoy related cost reductions in areas including storage licensing, storage management, power, cooling, and data center floor space.

The Analysts Agree

Here is some of what the industry is saying about the demand for deduplication today:

"A sea change is about to occur in primary storage, driven by the obvious mismatch between demand for capacity and the ability to supply that demand at a reasonable cost. There are only two variables here: since the price of storage isn't declining sufficiently fast, the only other option is for the amount of data that actually gets stored to decrease. Doing this while still serving demand for up-front capacity means that the necessary sea change for storage is (once again) spelled D-E-D-U-P-L-I-C-A-T-I-O-N."

— Steve Duplessie, *How Economics Alter the Storage Landscape*, ESG, May 2011

"Interest among Wikibon users for primary storage optimization is exploding (and)... primary optimization technologies should be embedded, have minimal application performance impacts and apply to both file and block storage."

— Dave Vellante, Wikibon, December 2010

"60% of respondents are either in the process of deduplicating or have plans to deduplicate their primary, backup or archive data in the coming year."

— Deduplication: An End-User Benchmark on Adoption and Value Metrics, IDC, Feb 2010

The Emerging Market for Universal Storage Data Optimization

Today's primary storage data optimization is a very different use case from that of disk backup. The substantial differences are in I/O performance and data integrity requirements. Some of today's primary deduplication systems use post-processing and impose severe capacity limitations (for example, some only scale to 16 TB) to limit the negative impact of deduplication on IOPS during production hours. In another example, today's most popular backup deduplication appliances work strictly inline and mask their very high I/O latency with good overall throughput, a reasonable trade-off for backup but not primary storage. Up until now, no single deduplication solution has been able to address both the IOPS requirements of block-oriented primary storage and the stream-oriented throughput requirements of backup in a single package.

Embedded OEM Data Optimization Software

Market dynamics point to the need for OEMs to develop comprehensive, sub-file-level deduplication capabilities that integrate with their existing single-tier storage solutions, while providing a viable roadmap to tomorrow's universal storage solutions. The overarching requirement for high performance data optimization is to provide petabyte-level scalability without incurring a performance penalty. All differentiating features of the storage platform must remain intact, with no compromises in functionality, data ingestion or data access performance.

The key requirements for data optimization involve the following areas:

- **Performance** — Data optimization must be extremely efficient and maintain a level of performance that does not impede overall storage performance on read and write operations. Storage vendors have made billion dollar R&D investments to optimize their storage performance as a means of differentiating their offering.
- **Scalability** — Less than 10 years ago, only a handful of IT organizations had a petabyte of data. Today, thousands of large organizations have requirements for more than a petabyte. Data optimization solutions must be able to scale to address the needs of these customers at petabyte and, in the future, exabyte capacities.
- **Feature Set Compatibility** — Data optimization software must operate in conjunction with existing storage software and not interfere with or impede existing features. Storage vendors have invested millions into storage features that are vital to the operation and value of their respective storage solutions.

To be viable, data optimization solutions for storage must not impact OEM application performance, feature set, or data integrity.

- **Data Integrity** — Data optimization technology must not interfere with the storage application software in a way that increases data safety risk. The OEM’s storage software must maintain control over writing the data to disk and the data optimization software cannot modify the data format in any way. This has the benefit of eliminating the need for complex data reassembly processes (commonly called “rehydration”), and protects data against possible corruption.

Today’s “add-on” primary data optimization methods are deployed as appliances and are unable to meet the four requirements described above. These alternative approaches insert a “bump in the wire” that undermines process flow and performance, limits scale, and inserts incremental complexity into the data storage process. In addition, these strategies create an immediate risk of data loss should they become inoperable.

High performance data optimization deployments will yield immediate overall performance and cost benefits, and will positively impact operational costs and resource consumption throughout its lifecycle across storage tiers.

Albireo Data Optimization Software

Permabit Albireo is the industry’s first and only embedded data optimization software for OEMs. At the core of Albireo is an advanced duplicate advisory service that allows storage vendors to deliver high speed data deduplication across primary, archive and backup storage tiers, both in the data center and the cloud. Albireo technology has broad applicability across differing workloads and content types. It is based on over a decade of development and a rich patent portfolio, also delivered today in the Permabit Enterprise Archive and Cloud Storage appliance offerings.

Albireo is designed to deal with the diverse block, file and stream oriented data formats of today’s special-purpose storage while laying the groundwork for the universal storage solutions of the future. It integrates at any point (inline, post-process, or parallel) and at a sub-file level. The software incorporates source-based capabilities, enabling deduplication to optimize primary storage without having to rehydrate data for backup, replication or migration purposes.

Albireo is the perfect solution for hardware, software, and service providers who wish to leverage their existing research and development investments, increase margins and accelerate time-to-market with leading-edge data optimization.

Albireo is the only solution to support universal storage, presenting the OEM with the opportunity to develop a single base platform for primary, archive and backup tiers of storage deployed both in the data center and the cloud, offering customers the greatest optimization value while leveraging shared manufacturing. Albireo is unique among deduplication technologies in that it is expressly designed to address the block, file and stream-oriented deduplication models required by emerging storage architectures.

Hash-based deduplication, as supported by Albireo, allows for data encryption with deduplication. This is a requirement for the cloud, where the data is stored in multiple remote locations, outside the control of the user.

Complete SDK for Easy Development

Albireo is delivered as a Software Development Kit (SDK). The SDK contains the Albireo data optimization software library, a comprehensive integration guide, full API documentation, and code samples. The Permabit development team provides extensive technical assistance to ensure rapid and seamless integration. Typically, OEM customers have their Albireo-enabled deduplication implemented and operational in just a few days so they can begin performance optimization and testing.

Albireo is the perfect solution for storage OEMs who wish to:

- *Leverage existing R & D investments*
- *Increase margins*
- *Accelerate time-to-market with next-generation data optimization*

Albireo Architecture

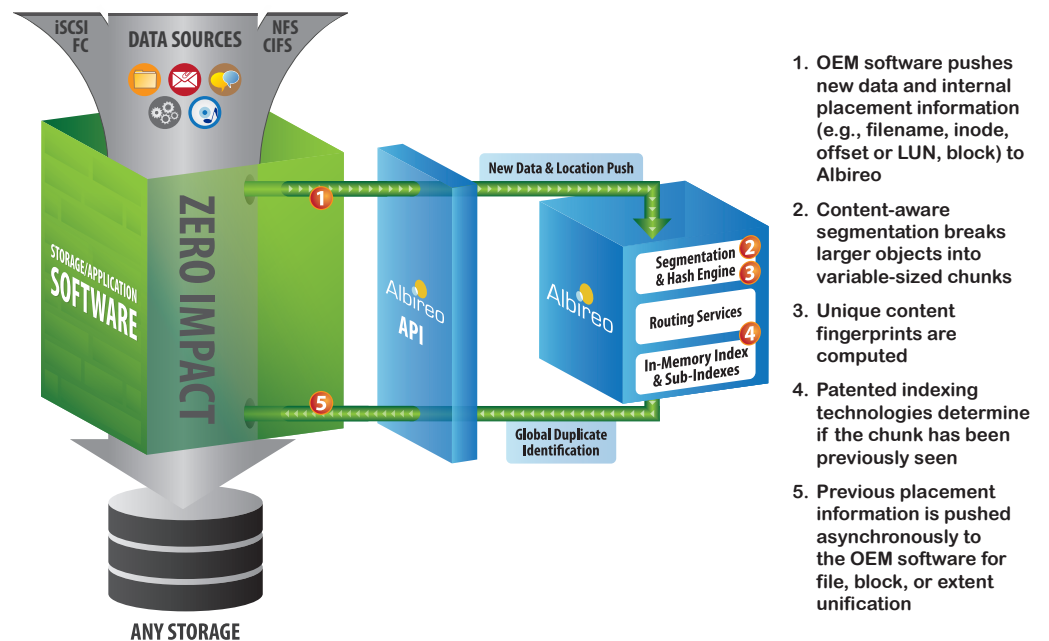
The single greatest challenge when implementing a deduplication system is in rapidly identifying duplicate information across a storage pool that can contain hundreds of billions of items. To achieve acceptable levels of performance the system must, for each new piece of data, quickly determine if that piece is identical to any previously stored piece of data. If a match is found, the storage system can then internally reference the existing item to avoid storing the same information a second time. On average, Albireo requires less than 10 microseconds to determine if a given piece of data is a duplicate. For flexibility within many existing architectures, Albireo can be implemented as an inline, post-process, or parallel operation.

An initial deduplication implementation for primary, archive, cloud or backup oriented storage can generally be completed in a matter of days using the Albireo SDK. As shown in Figure 1, Albireo operates outside the data path of the storage application software as a duplicate advisory service. This ensures that data integrity is never at risk and that there is zero performance impact. The advantage of the Albireo architecture is that it operates outside of the normal data storage flow and thus avoids any performance penalty. There is no risk to data integrity or of data format lock-in because Albireo does not modify or control the data — all reads and writes are managed by the storage software, and so data can always be read even if Albireo is disabled. Albireo is not involved in the reassembly of data because the OEM storage software handles all data retrieval as it normally would. That means no data lock-in, no performance penalty, and no reduction in data safety.

Albireo supports deduplication for primary, archive and backup storage solutions both in the data center and the cloud.

Figure 1: ▶

The Albireo process flow



Primary Storage

When implemented in primary storage, all stored data is managed by the existing storage software. In a parallel implementation, for example, once data is received by the storage system it is also delivered, with its corresponding location metadata (e.g. file name, offset, block, LUN), to Albireo. The incoming data objects are chunked into variable (or fixed) sized segments and a unique hash fingerprint (SHA-256 or MurmurHash3) value is calculated. These unique content fingerprints are compared to the existing stored data using patented high-speed indexing technologies that are capable of identifying duplicates orders of magnitude faster than similar technologies used in other storage products.

If the data chunk is a duplicate of previously stored information, deduplication advice is asynchronously pushed to the storage application software for file, block, or extent unification. If the data is unique, then no action is required. If the data is a duplicate, then the storage software updates its storage tables (e.g., inode block data structure for UNIX systems) to merge the duplicates and free the now unused storage space.

Archive/Cloud Storage

Integration with archive or cloud storage is similar to primary. The storage of existing data may be handled in a different way, but that is opaque to Albireo. When sending data to the archive or cloud system deduplication can be performed in the source application, removing the need to transmit data over the network that is already stored on the target system and thus vastly accelerating performance.

Backup Storage

Simply writing backups over a NAS protocol, such as NFS or CIFS, to a storage system integrated with Albireo will provide deduplication of the backup images. Albireo can also be integrated into a VTL solution, for compatibility with tape-oriented backup media agents, or into a traditional client-server backup application to provide source-level deduplication and additionally minimize network traffic associated with backups. Some backup container formats can benefit from using special chunking (content segmentation) of the backup stream for maximum efficiency.

Universal Storage for Primary, Archive and Backup

Albireo is capable of enabling deduplication in a single universal storage system serving primary, archive and backup applications. Three variations of the Albireo API are available to address these specific use-cases, they may be used stand-alone or to provide multiple tiers on the same platform. The APIs are:

Block API

The Block API supports deduplication at the granularity of fixed chunks of data. This is the preferred interface for storage systems implementing deduplication for high performance primary storage that deliver high IOPS with minimal latency. Block-level deduplication is particularly suited for block-oriented storage systems that have limited potential for file-level content awareness.

File API

The File API supports deduplication at the granularity of files (objects). This is the preferred interface for storage systems implementing deduplication for today's file-oriented archive storage repositories, whether in the data center or the cloud. To handle deduplication of chunks with arbitrary alignment, for example PowerPoint files which share graphics in common, Albireo supports content-awareness in the File API.

Stream API

The Stream API supports deduplication at the granularity of streams of data. This is the preferred interface for storage systems handling deduplication for backup devices, whether in the data center or the cloud. Content awareness is used to segment backup streams, which are typically in a specialized archive container format such as tar, into variable length chunks to ensure optimal alignment for deduplication.

Flexible Deployment Options

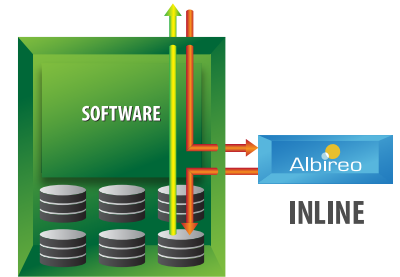
The Albireo architecture is flexible and can be integrated as an inline, post-process, or parallel solution by the OEM. Depending on the architecture and performance requirements of the existing software, an OEM can deploy Albireo as is best suited for its environment. For example, storage applications that must maintain the highest levels of performance can integrate Albireo as a post-process operation. Storage applications that benefit from real-time data optimization can deploy Albireo inline. The unique new parallel processing model offers the best of both worlds.

The three integration options for Albireo are shown in Figure 2. In each, it is important to note that Albireo always operates outside of the data read path (shown in yellow) and does not alter data written to disk. This avoids any reassembly performance penalty and protects against data corruption and lock-in. The OEM always controls the data write path, enabling them to manage data integrity and maximize performance.

Albireo integrates at any point (inline, post-process, or parallel) at a sub-file level, enabling deduplication to optimize primary storage and downstream replication processes.

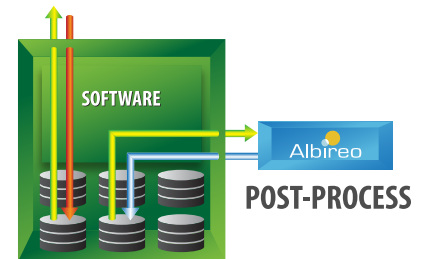
Inline

Albireo can operate 100% inline and process incoming segments in real-time. Inline processing is a good fit for applications that benefit from immediate data optimization and for which a slight latency can be masked by parallelism and write caching. Performing inline, Albireo intercepts the write path (shown in red) for only microseconds to determine if data is a duplicate, and then passes back advice to the storage software. Remember, Albireo never operates in the read path (shown in yellow) so data access is never impaired.



Post-Process

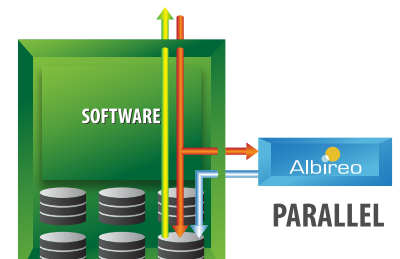
Albireo can also operate as a post-processing function, if desired by the OEM. Performing post-process, the storage stack reads data back from disk either on demand or on a fixed schedule, with Albireo identifying duplicates and providing update information to the storage application.



The Albireo mode of integration – inline, post-process, or parallel – is selected by the storage vendor. Having flexible integration options allows Albireo to perform optimally across all standard storage infrastructures, workloads, and content types.

Parallel

Parallel processing has the advantage of no latency in the write data flow. Duplicates are identified and updates for deduplication are managed in parallel to the write operation. Performing in parallel, Albireo receives a copy of the data as it is written to disk, so there is no delay in write. If a duplicate chunk is identified, the update (shown in blue) is asynchronously pushed to the storage software, where it can frequently be applied while data is still in the write cache. Parallel processing delivers no latency in write and real-time data optimization, which is perfect for a wide range of applications.



High Performance Index

Albireo uses a hash function (SHA-256 or MurmurHash3) to identify duplicate segments of data. Searching the resulting content fingerprints, which can quickly number in the billions (or trillions), is the bottleneck for all data deduplication solutions. As the number of fingerprints increase the size of the associated hash table data structure increases, very quickly exceeding available memory and spilling over to disk, which results in large processing delays. This bottleneck is the Achilles heel for all deduplication systems and it limits both scalability and performance.

The Albireo index uniquely addresses this challenge with patented Delta Indexing, Sparse Indexing, and Least Recently Used (LRU) discard technologies. Delta Indexing is a hybrid memory and disk resource technology that requires just 4 bytes of RAM per hash fingerprint. Assuming a 64 KB chunk size, the Albireo index, with 1 GB of RAM, can deduplicate 16 TB of unique data. Deduplication capacity is further increased by utilizing Sparse Indexing and LRU algorithms that optimize memory based on time. As a result, the Albireo index can deduplicate an amazing 10.2 PB of data with just 16 GB of RAM (Table 1).

Chunk Size	256 MB RAM	1 GB RAM	4 GB RAM	8 GB RAM	16 GB RAM
4 KB Chunk	10 TB	40 TB	160 TB	320 TB	640 TB
64 KB Chunk	170 TB	640 TB	2.6 PB	5.1 PB	10.2 PB

Figure 2:

Albireo can be adapted to inline, parallel, and post-process implementations

Table 1:

Albireo Index deduplication capacity based on chunk size and available RAM

The Albireo index can identify duplicate data in a matter of microseconds — orders of magnitude faster than other deduplication solutions.

The breakthroughs in the Albireo High Performance Index Engine result in an unparalleled combination of performance with extreme scalability. Even for the largest petabyte-scale storage environments, the Albireo index provides data deduplication with no performance penalty to the storage system.

There are three additional aspects of performance that should be considered when looking to implement storage deduplication: deduplication throughput, deduplication latency, and deduplication savings. Albireo achieves unparalleled performance in each of these key metrics.

Deduplication Throughput

Albireo deduplication throughput was measured in a test environment at 11 GB/sec on a single-core processor when operating with a 64 KB chunk size and hardware-based hashing. While hardware hashing provides the best performance, SHA-256 hash performance in software can still achieve speeds beyond 525 MB/sec on a single modern multi-core processor, even when used with more granular 4 KB chunks. Performance is sustainable as data is added to the system and scales linearly with additional cores to gigabytes per second. In a 32-node grid environment Albireo performance was measured at 400 GB/sec, with 64 KB chunk size and hardware-based hashing. Albireo deduplication throughput is over five hundred times greater than other deduplication technologies, which can only achieve rates in the 50-100 MB/sec.

Table 2: ▶

Albireo deduplication throughput performance with hardware hashing

Chunk Size	Single Processor, 1 core	32-Node Grid
4 KB	700 MB/sec	26 GB/sec
64 KB	11 GB/sec	400 GB/sec

Deduplication Latency

The Albireo Index addresses perhaps the most challenging aspect of data optimization — the amount of time it takes the deduplication engine to identify if an entry is a duplicate or not. Because of its highly efficient memory use, the High Performance Index Engine is able to identify duplicates in memory more than 99.95% of the time, eliminating the largest bottleneck in deduplication solutions — having to read from disk. Index lookups average less than 10 *microseconds* — orders of magnitude faster than other deduplication solutions. This enables sustainable ingestion rates of 11 GB/sec for a single-core processor (64 KB chunk size and hardware-based hashing) and scales linearly across multiple cluster nodes with Albireo grid technology.

Deduplication Savings

The final key area of performance is the deduplication savings ratio. Deduplication ratios are highly dependent on the data being processed. Albireo has been tested on a wide range of popular data types including common office productivity files, Microsoft Exchange data and VMware system images (Table 3). Albireo achieved the best results with VMware images with a deduplication rate as high as 99%. Excellent results were also achieved with the Exchange data and office files. Albireo reduced Exchange data by 86% and office files by 33%. Across the board, Albireo deduplication delivers massive cost savings.

Table 3: ▶

Permabit Optimization Results

Sample Data	Dedupe Rate 4 KB Chunks	Dedupe Rate 64 KB Chunks
User Directories, Fixed Chunk	2.8 : 1	2.7 : 1
User Directories, Variable Chunk	3.9 : 1	3.8 : 1
Tar Backups, Fixed Chunk with LZ77 compression	25.1 : 1	14 : 1
VMware Images, Fixed Chunk	36.3 : 1	26.4 : 1

Content Segmentation

As mentioned above, both the File API and the Stream API support content-aware deduplication.

With content-aware deduplication, data is analyzed intelligently based on its content type, resulting in more efficient variable sized chunks. For variable deduplication, Albireo utilizes specific content “scanners” to identify and optimize deduplication of objects within specific files (for example, Microsoft® Office® documents, ZIP, PDF, and Hyper-V files), which results in improved deduplication efficiency, saving more storage space and its associated costs. For common file sets, content awareness can yield efficiencies of 20:1. For backup streams, the benefits can exceed 30:1.

Albireo’s content-aware, variable block data deduplication delivers the same exceptional performance as fixed-block use, and delivers improved storage efficiency for a variety of use cases.

Albireo Scale-out

Albireo grid technology enables OEMs to increase performance and scalability far beyond what is possible with traditional data optimization solutions. Albireo grid technology incorporates patented indexing and memory resource utilization software with Grid Server to provide performance that scales out linearly, even when dealing with multi-petabyte capacities, without the performance drop-off common with the other data deduplication solutions on the market today. An Albireo grid consists of multiple Albireo clients, embedded in each storage ingest point, connected to multiple servers running the Albireo index over the network (Figure 3).

In an Albireo grid deployment scenario, each Albireo client performs its own content-level segmentation and hashing, then routes the resulting content fingerprints to the appropriate index on an Albireo server. The index returns deduplication advice to the client and stores the deduplication metadata if the new data is unique. The existing storage software is responsible for data storage, while the Albireo index on the servers stores content fingerprints.

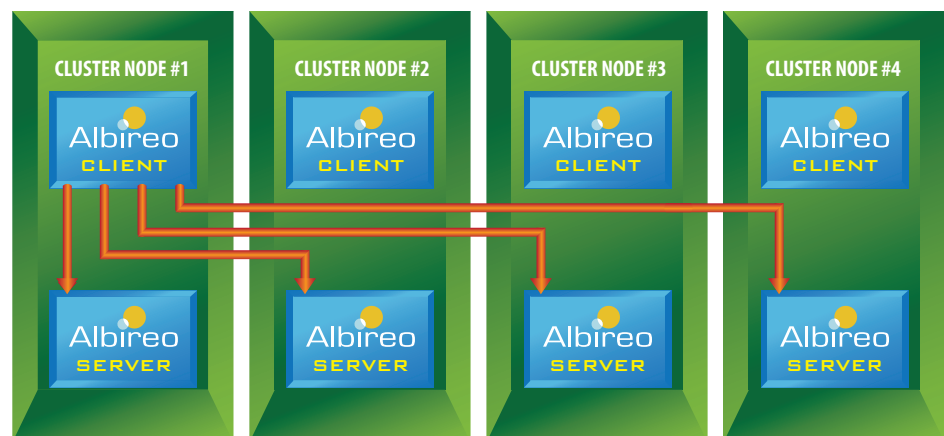
The index table is automatically distributed across the Albireo grid. Each server stores a balanced portion of the Albireo index. When servers are added to or removed from the Albireo grid, the index is re-balanced. Additionally, a single grid supports multiple namespaces; separate independent storage domains (e.g., file systems, logical volumes, or systems) can share a single Albireo index. Both unified storage pools and large quantities of independent storage volumes benefit from the performance and scalability of the Albireo grid technology.

Albireo grid technology provides:

- Massive indexes for unlimited scalability
- Distributed workloads for maximum performance
- Deduplication results in just microseconds

Figure 3: ▶

Albireo clients communicate with all servers to access the distributed index



Why Permabit?

Innovation

Only Permabit Albireo enables OEMs to quickly and reliably deliver high performance scalable deduplication with no performance impact into existing primary, archive/cloud and backup solutions. Albireo is an embedded OEM solution that flexibly integrates within the constraints of existing storage architectures and leverages existing significant R&D investments.

Focus

Permabit is an expert in the development of highly scalable, next-generation storage solutions that deploy full inline data deduplication. Universal data optimization promises to change the storage landscape forever. By offering the industry's first embedded OEM data optimization solution, Permabit is enabling storage OEMs to compete effectively with breakthrough technology. Emerging storage vendors can capitalize on this major market shift by introducing new storage solutions that take market share away from incumbents. Leading storage vendors can leverage Albireo to further solidify their market position.

Expertise

The Permabit track record in storage expertise and innovation is without peer for a company of its age and size. Permabit has a total of 37 patents filed and 28 patents granted, all in the storage-related field. Its MIT-educated engineers have earned multiple awards for product innovation. Since 2000, Permabit has worked to develop the latest storage technology to address the challenges of highly scalable storage. The Albireo software is the first time that Permabit has made its core intellectual property for data optimization and highly scalable storage available as an OEM offering to other manufacturers. The Albireo architecture is a proven technology that has been implemented in production environments as a core technology in the Permabit Enterprise Archive and Cloud Storage solutions.

Conclusion

Data centers are dealing with explosive data growth and flat budgets. As a result, IT organizations are making storage purchase decisions based on storage efficiency and total storage costs versus simply buying "cheap capacity." Storage vendors who will grow and flourish in today's business environment must adapt their existing storage solutions and/or introduce new solutions that offer greater storage efficiency and reduced operating cost. Emerging storage vendors can introduce new storage solutions based on data optimization and enter previously impenetrable markets.

Permabit is a pioneer in the next-generation data optimization market, providing the industry's only embedded data optimization software for primary, archive, and backup storage solutions from the data center to the cloud. Permabit Albireo enables OEMs to deliver next-generation universal storage, leverage their existing R&D investments, increase margin, and accelerate time-to-market.

Permabit Albireo software enables best-in-class:

- Block Storage OEM solutions
- File Storage OEM solutions
- Unified Storage OEM solutions

About Permabit

Permabit is a recognized leader in data efficiency technology. We enable OEMs to leverage their R&D investment, increase margin, accelerate time to market and achieve competitive advantage. Permabit Albireo software massively improves performance and efficiency of data creation, transmission and storage. Solutions built with Albireo are being delivered by leading hardware, software and service providers.

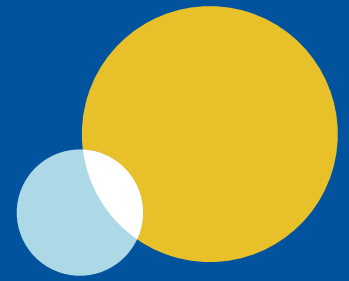
Only Albireo enables OEMs to quickly and reliably integrate deduplication into their existing primary, archive and backup storage solutions for the data center and the cloud.

"I don't see any other real alternatives for OEMs to be able to quickly get to market with lightening fast dedupe capabilities for primary, secondary, or really anydary storage. Albireo rocks."

*Steve Duplessie
Founder & Sr. Analyst
Enterprise Strategy Group
June 2011*

Find Out More

To learn more about the Permabit Albireo technology, or to license our products, visit our website at www.permabit.com or call us directly at 617.252.9600.



Albireo (al-BEER-ee-oh) appears to the naked eye to be a single star but can be resolved with a telescope into a double star, consisting of a brighter yellow star and a fainter blue star.



**Ten Canal Park
Cambridge, MA 02141
Phone: 617.252.9600
FAX: 617.252.9977**