PERMABIT TECHNOLOGY CORPORATION

Data Sheet

Albireo Data Optimization Software

Deduplication for Today's OEM Storage Solutions

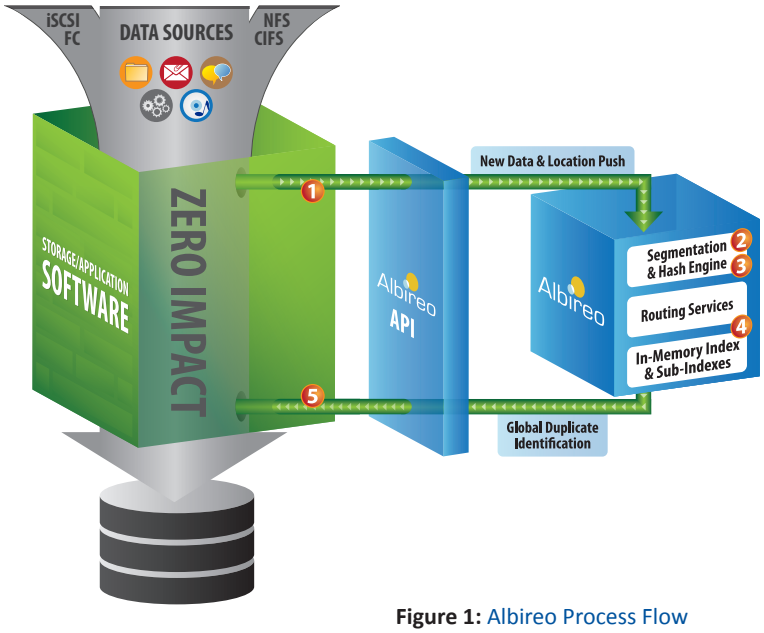
Albireo

With the amount of electronic data expected to exceed 35 zettabytes by 2020,[1] the conundrum facing every business is this: how can we to afford to store, analyze, manage and house this data? The increased storage costs in particular have driven IT organizations to explore data optimization technology as a more efficient means of housing information. It's no surprise that data efficiency is the number one priority of storage professionals in IT today.[2] Customer demand in turn drives storage manufacturers to offer data optimization capabilities.

Of the available data optimization technologies, data deduplication offers 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 does not negatively impact 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.

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 current or next-generation storage and platform architectures. Albireo deduplication is seamlessly deployed in primary, archive, and backup storage across the data center and the cloud. With Albireo, OEMs leverage their existing R&D investments while accelerating time to market for must-have, industry-leading data optimization capabilities.

An initial deduplication implementation for primary, archive, cloud, or backup-oriented storage can 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.



1. OEM software pushes new data and internal placement information (e.g., filename, inode, offset or LUN, block) to Albireo
2. Content-aware segmentation breaks larger objects into variable-sized chunks
3. Unique content fingerprints are computed
4. Patented indexing technologies determine if the chunk has been previously seen
5. Previous placement information is pushed asynchronously to the OEM software for file, block, or extent unification

Figure 1: Albireo Process Flow

Deduplication Efficiency

The Albireo architecture supports both fixed and variable block deduplication. Data can be intelligently segmented into chunks of variable length based on analysis of its content. Albireo uses content "scanners" to identify and optimize deduplication of objects within specific compound data formats (e.g. Microsoft® Office® documents, ZIP, PDF, tar). An API is available for vendors to implement their own application-specific scanners for further savings. Independent validation has demonstrated that Albireo offers 7X data reduction for general purpose environments, while data reduction rates of over 35X are seen for virtualization images generated by products such as VMware®.

"The Albireo technology from Permabit will save an OEM 18-24 months getting to market... 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

Enables Best-in-Class Deduplication for Solutions Across the OEM Portfolio:

- For Primary, Archive, and Backup
- In the Data Center and the Cloud

Addresses Key Requirements for Deduplication

- No Performance Impact
- Petabyte Scalability
- No Feature Set Impact
- No Data Integrity Impact

Albireo Performance Highlights:

- Identifies duplicate data in under 10 microseconds
- 11 GB/sec ingest performance (single processor core with 64 KB chunk size and hardware-accelerated hashing)
- 400 GB/sec ingest performance (32-node grid with 64 KB chunk size, multi-core processors, and hardware-accelerated hashing)

High Performance Index

The Albireo index utilizes advanced techniques to identify duplicates across massive storage pools using memory-resident information over 99% of the time, avoiding costly disk access and eliminating the largest single bottleneck in storage deduplication. Index operations average less than 10 *microseconds* — orders of magnitude faster than other deduplication solutions. This enables sustainable ingestion rates of 11 GB per second with a single processor core (64 KB chunks and hardware-accelerated hashing) and scales out linearly by leveraging Albireo grid technology.

Flexible Deployment Options

Albireo is integrated by the OEM as an inline, post-process, or parallel solution, depending on which is best suited to address the customer requirements for the vendor's architecture. Source-based deduplication deployment (where data is deduplicated at the source for bandwidth optimization) is supported along with the more traditional target-based approach.

As can be seen in Figure 2, storage applications that can benefit from real-time data optimization may deploy Albireo inline while those that wish to perform optimization as a scheduled event can integrate Albireo as a post-process operation. The Albireo parallel processing model offers the best of both worlds. Albireo always operates outside the data read path (shown in yellow) and does not write or alter data written to disk, regardless of which option is used. The OEM also retains full control of the data write path, managing data integrity, and maximizing performance.

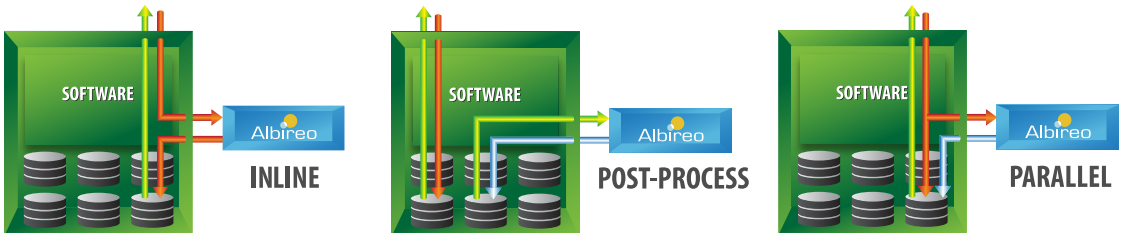


Figure 2: Inline, Post-Process, and Parallel Integrations

Complete SDK for Easy Implementation

Albireo is delivered as a Software Development Kit (SDK). The SDK contains the Albireo software library, full API documentation, code samples, and application notes for integration with primary, archive, and backup storage solutions. Permabit provides technical assistance to ensure rapid and seamless integration. OEM customers typically implement Albireo deduplication (and begin performance optimization and testing) in under 5 days.

Features and Benefits

Albireo Feature	Albireo Benefit
Proven deduplication architecture	OEMs can confidently integrate Albireo as a mature solution to deliver data optimization quickly and reliably to meet customer demand
High Performance Index Engine	Albireo handles heavy workloads and can identify duplicate data in *microseconds* across petabytes of data
Content-Aware Segmentation	Optimizes efficiency based on data type
Block, File, and Stream oriented APIs	Albireo leverages existing storage architectures, protecting OEMs' R&D investments
Inline, Parallel and Post-Process Integration	Can be flexibly deployed based on OEM storage architecture and design preferences
Target and Source APIs	Provides efficiency across the Data Center and the Cloud
Grid (Scale-out) Index	Offers linear performance scale-out as additional nodes are added. With the latest release, a 32-node grid with multi-core processors delivers ingest performance that exceeds 400 GB/sec (64 KB chunk size, hardware-accelerated hashing)

1 Extracting Value from Chaos, IDC, June 2011
2 Storage Infrastructure Spending Trends, ESG, Jan 2011

Find out More

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

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.

PERMABIT TECHNOLOGY CORPORATION

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

info@permabit.com
www.permabit.com