

# Albireo Software Development Kit (SDK) for Data Protection Appliances



*A key enabling technology for disk-based data protection appliances is deduplication. The effective capacity of a deduplication-enabled system can expand significantly by not storing the same information multiple times. A solution that achieves a 20X data deduplication ratio can store 100 TB of data in just 5 TB of physical capacity. When deployed within inexpensive disk systems, deduplication moves the cost of disk even closer to that of tape offerings.*

Albireo delivers advances in data deduplication technology that allow vendors to deliver faster and more scalable deduplication than ever before, without requiring a substantial increase in hardware resources. Albireo High Performance Indexing enables scale-out well beyond the multiple petabyte level. Its memory-based indexing and collision-free hashing, delivered in microseconds, enables the backup performance levels needed for today's petabyte-scale data stores. In addition, Albireo can be used to provide source-based deduplication capabilities that enhance performance and reduce network bandwidth requirements. With Albireo, data can be backed up in real time, while delivering immediate in-line space savings and addressing performance requirements. Applying Albireo data deduplication in data protection appliances delivers more resource-efficient backup that scales to meet the needs of today's enterprise data protection stores.

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 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.

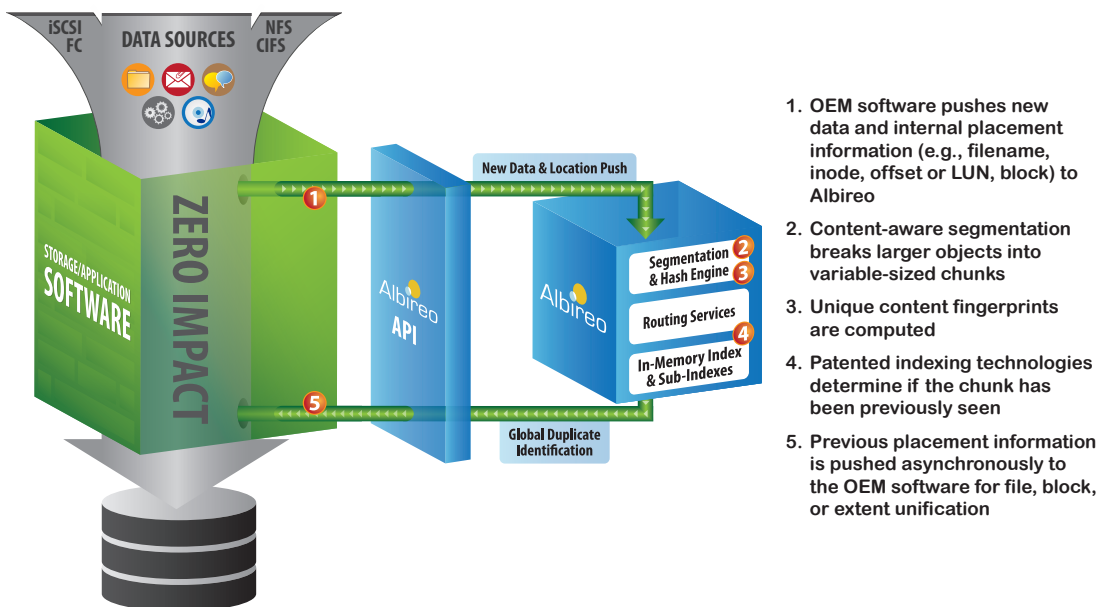


Figure 1: Albireo Process Flow

*"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

## Business Benefits

- Reduces deduplication resource requirements
- Lowers customers' cost of data protection cycles
- Reduces the manpower and resource load of backup
- Delivers real-time efficiency

## Technical Benefits

- Delivers enterprise-class inline performance
- Offers resource-efficient, massively scalable indexing
- Provides bandwidth optimization without the traditionally associated overhead
- Reduces deduplication memory requirements by over 90%
- Increases efficiency by reducing the effective cost per GB stored

## 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®.

## 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.

## 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
High Performance Index Engine	Albireo handles heavy workloads and can identify duplicate data in <i>microseconds</i> across petabytes of data
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)
Mechanical Disk Optimization	Provides optimal performance for hard disk-based environments
Content-Aware Segmentation	Optimizes efficiency based on data type
Bandwidth Optimization (Source Deduplication)	Improves bandwidth efficiency for writes both within the Data Center network and into the Cloud
Portable Source Code	Allows for ports of the SDK to custom embedded devices

## Find out More

To learn more about Permabit Albireo technology, or to license our products, visit our website at [www.permabit.com](http://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.



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

[info@permabit.com](mailto:info@permabit.com)  
[www.permabit.com](http://www.permabit.com)