PERMABIT
TECHNOLOGY CORPORATION

Albireo Software Development Kit (SDK) for Virtual Environment Backup

Albireo

Both server and desktop virtualization have compelling propositions for IT teams. The use of Virtual Machines (VMs) has caused organizations of all sizes to rethink their data protection strategies. Many companies have realized that having a server backup and disaster recovery plan built specifically for virtual environments is essential. New data protection software offerings have emerged to address these specific requirements, and one key element of a strong VM backup strategy is deduplication.

Since the backup needs of virtual environments involve backing up images of each VM, the application of deduplication technology lends itself perfectly because there is a high duplicate ratio. A dedupe rate of 35X or higher is very common, saving storage space and reducing the impact on backup cycles. Albireo provides extremely granular (4 KB chunk) deduplication at a very low resource cost, so when applied to VM backup operations Albireo delivers extremely high backup efficiency and the highest possible cost reduction while improving overall operations (including recovery cycles).

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 VM backup software 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.
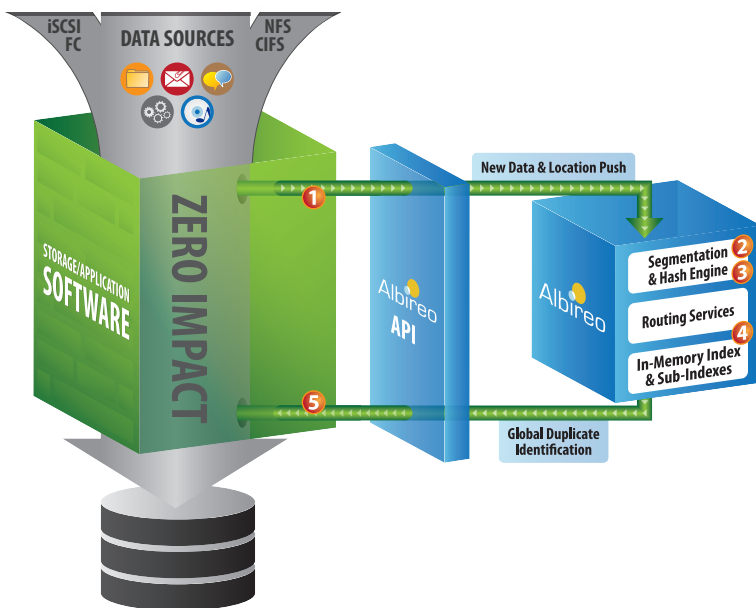
> *"The Albireo technology from Permabit will save an OEM 18-24 months getting to market… This stuff is so far ahead in its capabilities and performance I can't see why you would want to do it yourself, unless you already have it baked."*
>
> — *Steve Duplessie*
> *Founder & Sr. Analyst*
> *Enterprise Strategy Group*

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®

Business Benefits

- Lowers cost of backup storage with 35X or more data reduction
- Improves data recovery efficiency
- Reduces maintenance costs because less storage is needed

Technical Benefits

- Increases backup performance because data volume is markedly reduced
- Increases reliability with less data to manage
- Increases efficiency by lowering the effective cost per GB of data stored

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 *microseconds* across petabytes of data
Bandwidth Optimization (Source Deduplication)	Improves bandwidth efficiency for writes both within the Data Center network and into the Cloud
Content-Aware Segmentation	Optimizes efficiency based on data type
Mechanical Disk Optimization	Provides optimal performance for hard disk-based environments
Solid State Disk Optimization	Provides optimal performance for flash-based storage environments
Portable Source Code	Allows for ports of the SDK to custom embedded devices

PERMABIT
TECHNOLOGY CORPORATION