Acronis Software-defined Storage: Fast, Safe and Efficient

Thanks to Blockchain and Intel® Intelligent Storage Acceleration Library

Every piece of data is required to be stored somewhere. We all know about hard drives, optical discs, tapes and other means of storage. But nowadays cloud storage and software-defined storage are two modern methods of storing data that are useful, especially when we talk about large volumes of it. In this paper we will concentrate on Software-Defined Storage (SDS), a market Acronis entered in 2016.

SDS is software that separates common storage features from the hardware storing the data. It uses high-volume server hardware and provides flexibility and scalability without disruption, bringing local and cloud storage into a single pane of glass for its users.

Gartner predicted that by 2019, 70 percent of existing storage array solutions will be available as a "software only" version. The research firm also predicted that by 2020, 70 percent to 80 percent of unstructured data will be stored in less-expensive storage hardware, managed by SDS systems[1].

Innovative Acronis Archive 3 backup storage format

Beginning with the release of Backup 12 Advanced (which has since been replaced by Acronis Backup 12.5), Acronis introduced a new default archive format – TIBX. This is a modern format for hybrid (block and file level) environments, **which supports up to a billion files, 100K slices and 50 TB archive size**. This format was developed with reliability, speed and cost-effectiveness in mind. Avoiding corruption in case of power failure; it supports asynchronous data access, fast browsing and paging; and features built-in block-level deduplication for any type of data, adaptive compression, and other smart capabilities.

This single-format applies for all backup types: disk, file, app, mobile, Office 365 and so on. Archive 3 supports metadata inside the archive that has been checksummed and verified. With the TIBX format you can validate the recoverability and accessibility of encrypted archives without entering the password. It also supports re-usage of free

[1] Gartner Innovation Insight: **Separating Hype From Hope for Software-Defined Storage authored by Dave Russell and Arun Chandrasekaran published on October 2014.**

blocks for all types of backup by marking blocks in file as "free", thus OS can allocate them for other files and saving disk space during cleanup.

The format allows high-speed hybrid backup: any volumes that can be snapshotted are backed up using block-level method, other volumes by file-level method, and all backed up volumes are kept in single backup. This asynchronous pipeline delivers 500MB/sec backup speed with support for up to 8 files to be recovered in parallel. Archive 3 supports adaptive compression with ZSTD/LZ4 as well as encryption that includes keys in addition to passwords. Changing the password doesn't require re-encryption.

State-of-the-Art Storage from Acronis

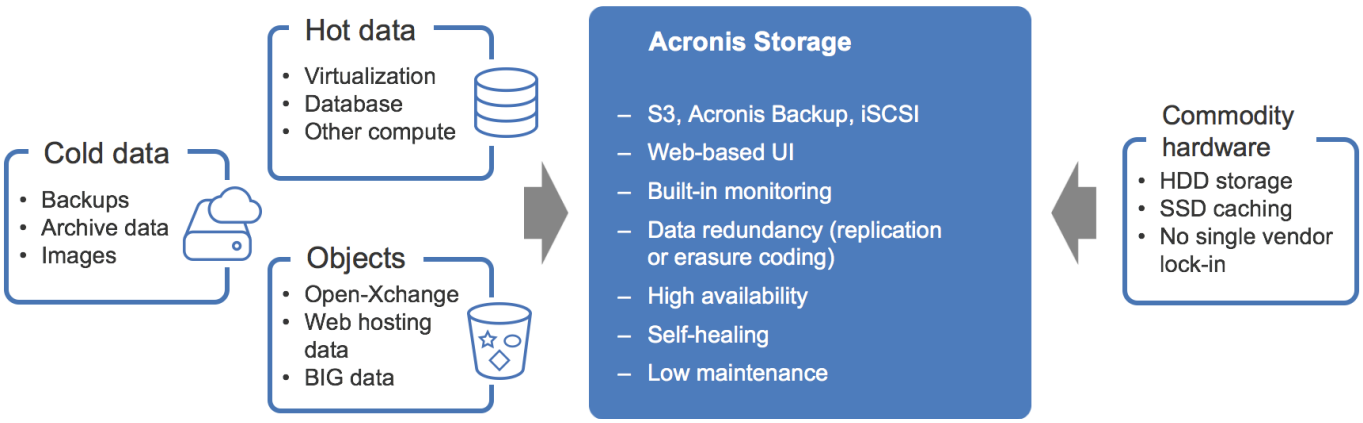
Acronis Storage is an SDS solution that allows service providers and end-use customers to quickly and easily transform heterogeneous hardware into protected, enterprise-grade, scalable storage to improve your total cost of ownership. When Acronis customers use Storage, backup data will be stored in Archive 3 format.

With features such as SSD caching, automatic load balancing, and parallel replication, Acronis Storage unites file, block and object-based storage in a single software-defined, scalable solution to cover the needs of a modern business. Acronis Storage is designed to provide a single-store identity across two or more data centers. Object data is actively replicated to all data centers with asynchronous replication. Single namespace allows access to data across all data centers. This protects against the failure of a single data center, while providing the highest levels of data protection and availability. Acronis Storage provides transparent encryption of all data stored using Advanced Encryption Standard (AES) with 256-bit key.

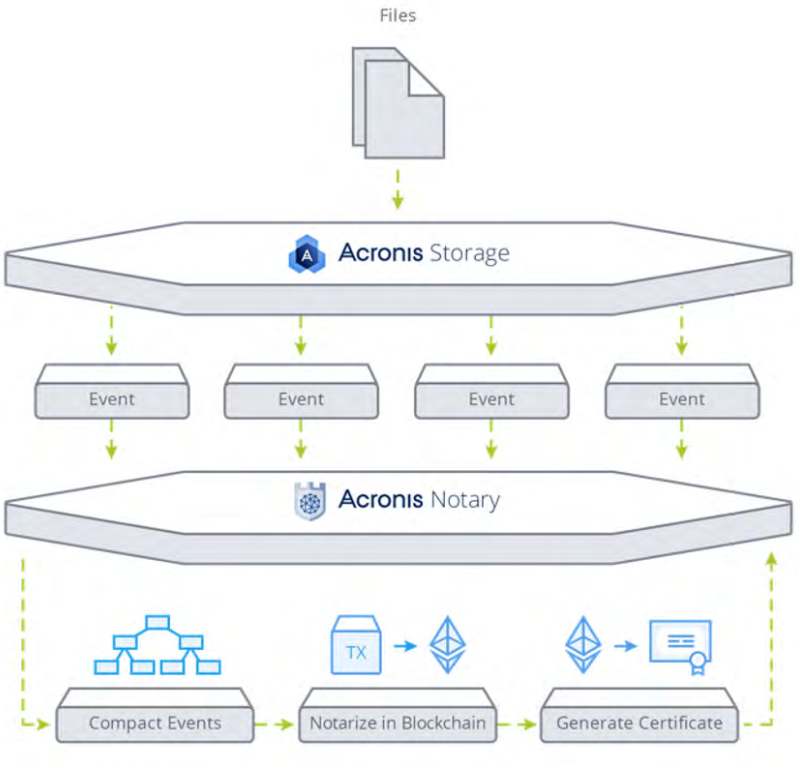
One important part of Acronis Storage is CloudRAID, which manages overhead and compute time, and solves data integrity as well as data rebuild and recovery problems. Another unique feature is Acronis Notary – blockchain-based technology that provides data immutability and authenticity verification. Let's look atthese in more detail.

CloudRAID and Notary to safeguard your data

Acronis CloudRAID uses Reed-Solomon erasure coding technology to reconstruct both protected data from data chunks and parity information stored elsewhere in the storage cluster. Erasure coding technology splits object data into multiple chunks that are distributed across a large number of disks and nodes, providing configurable levels of resiliency and fault tolerance. If a failure occurs, the data can be regenerated from the remaining erasure-coded chunks. Erasure coding comes with a significant advantage: it can be used to provide data redundancy that's equivalent to triple replication (e.g. it can withstand the loss



Acronis Storage optimized for commodity hardware usage.



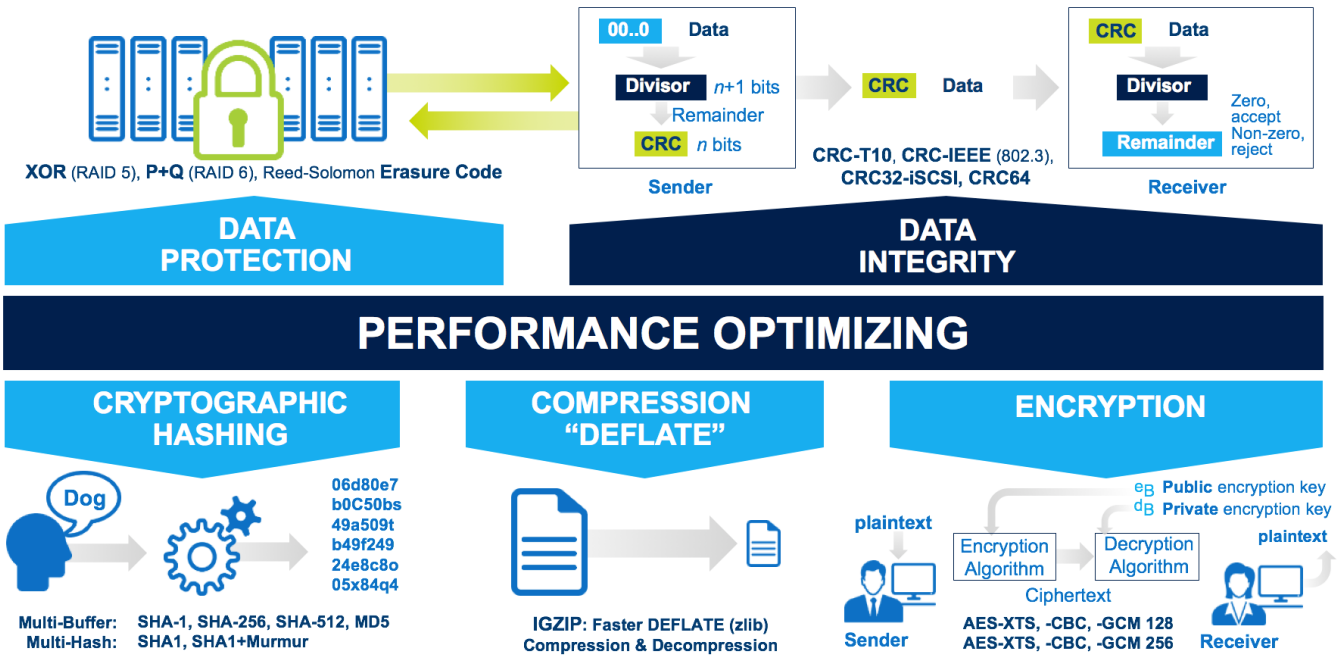
Acronis Notary uses Ethereum ledger to store certificate hashes.

of several copies of that data). By intelligently placing the data, it consumes just half of the raw disk capacity required for triple replication. The erasure coding algorithms powering CloudRAID are provided by Intel® Intelligent Storage Acceleration Library, which we will talk about in detail below. It delivers fast, flexible, efficient redundancy, protects against node failure, and offers a customizable protection level and disk overhead. These erasure code functions implement a general Reed-Solomon type encoding for blocks of data to deliver robust data protection. When there are only two replicas in Storage, RAID6 is used, and if more, Reed-Solomon encoding is used.

Acronis is a company pioneering blockchain for data protection. In fact, the release of Acronis Storage in 2016 was the first time a Software-Defined Storage (SDS) integrated blockchain into data protection. Acronis Notary with blockchain generates a time-stamped fingerprint of protected data and stores it in Ethereum, a public blockchain-based distributed ledger. The blockchain is distributed between hundreds of unrelated parties and encrypted. Every record in the blockchain is immutable and independently verifiable. By comparing two fingerprints (256-bit SHA-2 hashes) of the same data, Acronis Storage can verify the authenticity and integrity of stored data.

Intel helps to overcome storage challenges

Apart from its own innovative technologies, Acronis uses the Intel® Intelligent Storage Acceleration Library (Intel® ISA-L) in both the Archive 3 format and Acronis Storage products. ISA-L is an algorithmic library containing the core storage, cryptography, and compression algorithms that are optimized for applications where throughput and latency are the most critical factors. It can be used to increase performance for data integrity, security/encryption, data protection, and compression algorithms, which is what the Acronis engineers did. ISA-L is also designed for software-defined applications, providing an API which delivers the optimal binary implementation for past, present, and future Intel processors without requiring recompilation. From a developer's perspective, using ISA-L eliminates any concerns of the underlying processor architecture, delivering best-case performance regardless of CPU generation, virtualization environment, or even operating system.



ISA-L delivers optimal algorithms for data protection, integrity, compression, encryption and hashing on Intel CPUs.

ISA-L uses the latest processor capabilities, including the Intel® Advanced Vector Extensions (Intel® AVX-512) set of instructions, to accelerate performance for workloads and usages such as cryptography and data compression. Intel® AVX-512 has been available on the latest Intel® Xeon Phi™ processors and coprocessors and recently became available on Intel® Xeon® Scalable processors.

Acronis leverages many of the algorithms from ISA-L in the Archive 3 storage format, including using the CRC64 checksum calculations, SHA-256 multi-buffer and Rolling Hash for deduplication. Cyclic Redundancy Check (CRC) functions are used to detect accidental changes to raw data during transmission. CRC64 extends the familiar CRC capabilities to provide data integrity on objects up to 18 exabytes in size, which ISA-L delivers with no performance penalty relative to CRC32. However, the other algorithms used by Acronis (e.g.multi-buffer SHA-256 and 64-bit rolling hash) may require more explanation.

- **Multi-buffer hashing** functions provide cryptographic hash functions that use the unique capabilities of Intel CPUs. Intel® ISA-L supports MD5, SHA-1, SHA-256, and SHA-512 and Acronis Archive 3 uses SHA-256 calculations. These multi-buffer functions are used to increase the performance of the secure hash algorithms on a single processor core by operating on multiple jobs at once. By buffering jobs, the algorithm can exploit the instruction-level parallelism inherent in modern Intel cores to an extent not possible in a serial implementation. Acronis saw large reductions in CPU utilization by using ISA-L instead of other implementations, allowing the creation of more robust data services with fewer CPU resources.

- **A rolling hash,** on the other hand, is a hash function where the input is hashed in a window that moves through the input. In practice, this means rolling hashes provide the ability to calculate the hash values without re-hashing the whole string. A few hash functions allow a rolling hash to be computed very quickly—the new hash value is rapidly calculated given only the old hash value, the old value removed from the window, and the new value added to the window—similar to the way a moving average function can be computed much more quickly than other low-pass filters.

In Acronis Storage, CRC64 is used to provide very strong data protection and data integrity services, as are Reed-Solomon Erasure Codes and RAID6 algorithms from ISA-L. Additionally, the cryptographic cipher functions of ISA-L are used to deliver AES-128/256 cryptography. Erasure Code functions allow breaking up of objects into smaller fragments stored in different places, and regenerating the data from any combination of smaller numbers of those fragments.

The usage of ISA-L gave more convenience in development to Acronis engineers as well as a performance boost on Intel CPUs, especially ones which support AVX-512 instructions set (see a Table 1 for comparison). Acronis was able to take high-performance algorithms and deploy them immediately, instead of laboriously creating and tuning implementations in-house. All this Intel CPU improvements for the past 3-5 years allow storage vendors to avoid any overhead associated with data verification, deduplication and encryption, just enabling the functionality always by default.

This also means that storage and backup software uses ~25-70% less load of a single core in high loaded scenarios and releases this power for user workloads like Virtual Machines.

Beyond the performance improvements offered by ISA-L, Acronis was able to focus development on the features and data services, rather than chasing performance optimizations.

Combination of best in-house and open-source technologies

For storage formats and products, it is vital to: 1) be efficient and fast; 2) data integrity; and 3) provide security via encryption, compression, deduplication, and the other advanced functionalities mentioned above. Acronis Storage and Acronis Archive 3 deliver all of these thanks to internal, innovative – technologies, and the use of industry-leading service libraries like Intel® Intelligent Storage Acceleration Library. For Acronis customers, this means lower costs with heterogeneous hardware, improved IT productivity, reduction of disk overhead, and data protection and availability. Acronis products also deliver unique data protection where file integrity is authenticated using the blockchain Ethereum ledger. That means you can be sure that data is immutable, or exactly the same as it was when placed into Acronis Storage.

Operation	Sofware only	Previous Acronis Implementation	Impelementation with ISA-L
CRC32/CRC32 SSE4	2000 MB/sec	12500 MB/sec	12500 MB/sec
CRC64	2000 MB/sec	-	9200 MB/sec
RAID6 erasure codes	1GB/sec	14.6 GB/sec	21 GB/sec
Reed Solomon Codes	1.5 GB/sec	5.7 GB/sec	13 GB/sec (AVX-512)

Table 1. ISA-L significantly improved Acronis products performance.

Tested on system running on Intel(R) Xeon(R) Platinum 8168 CPU @ 2.70GHz

For additional information, please visit **www.acronis.com**