

# CEDA™ Data Deduplication

Product Brief



The CEDA Cryptographic Data Deduplication engine delivers high throughput inline compression eliminating coarse-grain redundant data. The sustained 25 Gbps per deduplication engine can be added to existing application servers and storage systems as well as an inline standalone appliance.



Accelium CEDA Data Deduplication engine



Accelium CEDA Data Deduplication appliance with upto 4 CEDA engines

**Data deduplication** is a specialized data compression technique for eliminating coarse-grained redundant data. This technique can be used to improve storage and network utilization by reducing the amount of data that needs to be stored or sent. In contrast to fine-grain data compression which reduces the storage size of frequently used words, data deduplication examines chunks of the document to detect if any change has been made and only stores or sends the changed blocks. For example, if only 1 block of a 100-block document is changed, then the compression ratio is 100:1. At this ratio, a change to a 100MB file would only require 1MB.

**Off-line vs In-line Deduplication.** Many advanced file systems use off-line (or post-processing) data deduplication to save disk space. This can result in greatly increased storage capacity for files that do not change frequently.

In-line deduplication is more advanced as it provides real-time analysis of data prior to storage or transmission. This can dramatically improve a wide-area network's bandwidth utilization while also improving responsiveness as more of the bandwidth is available. Real-time change-detection can be implemented with in-line deduplication and can be performed on a document or block-level. Additionally, in-line deduplication can be used to perform disk-based deduplication without any file system modification.

## Deduplication Quality and Trustworthiness.

How are duplicates detected? This is a critical question as a poor detection of similarity can result in data loss. Deduplication generates a key-value for each block of data and if two blocks have the same key-value, they are flagged as identical. If the key-value generator is not robust, then two different blocks can have the same key-value and the second block will not be stored, resulting in data loss.

*Cryptographic data deduplication* uses a cryptographic hashing algorithm to generate key-values. These algorithms are extensively tested to ensure that even a single bit change in a block results in a different key-value. Digital signatures and electronic commerce rely on cryptographic hash algorithms, such as the NIST certified SHA-1 and SHA-2 algorithms. Various seemingly-secure algorithms have been proposed and put into practice but later shown to have collisions (i.e. different blocks generate the same key-value.) Additionally, cryptographic hash algorithms are designed to protect against purposeful deceit. So it is computationally impossible to determine which bits of a block can be simultaneously changed to generate the same key-value result.

## 25 Gb/s Cryptographic Data Deduplication Accelerator.

Achieving in-line (i.e. real-time) performance of cryptographic deduplication is computationally intensive in software and can consume a server's processor capacity. Delivered on a DRC Accelium™ PCIe add-in card, the Cryptographic Deduplication Accelerator achieves both cryptographic reliability and real-time performance. Built upon the *Secure Hash Algorithm 2 (SHA-2)* cryptography standard, the accelerator ensures robustness and reliability. As a PCIe add-in card, the algorithm cannot be modified or harmed by viruses and will never overload the processor. SHA-2 performance is maintained at 25Gb/s. Key-value comparison performance is comparable but is dependent on the number of concurrent blocks in use.

## CEDA

Concurrent EDA (CEDA) is an application partner of DRC Computer Corp (DRC). Based in Pittsburg, PA CEDA is a recognized expert in developing FPGA-based complex applications.

## Made in America

DRC is a US based company with all staff US citizens, and all its engineering and manufacturing conducted in the US.

**DRC COMPUTER CORP**  
3375 Scott Blvd, Suite 206  
Santa Clara, CA 95054

**PHONE**  
+1. 408.562.0000

**WEB**  
drccomputer.com