Data Integrity

Data that becomes corrupted through any number of environmental or operational problems can wreak havoc on your workflows and workloads. Your storage solution must always ensure that data read is exactly what was initially written. Here are some factors to consider when exploring data integrity.

1 Checksum process

Checksums are algorithmically-derived small-sized blocks of data (also called a *hash*) which are calculated from a larger block of data. By comparing checksums—say, when data is written to disk and subsequently when read from disk—a storage solution can determine whether data has full integrity or has been corrupted.

2 Checksum variations

Because the checksum process depends on algorithms, many different variations exist in the market. Different checksum algorithms include Fletcher, SHA-256, and MD5. Different methods require different computational needs, which is the reason we implement Fletcher 4 which is lightweight on computational resources while still being highly accurate in detecting errors.

3 Error correction

Detecting data integrity errors—such as corrupted data that has succumbed to bit rot or mismatched data due to read, write, or transmission problems—is really only half the story. You have to ask, how well does your storage solution recover corrupted data once detected? Our solutions can instigate automated error correction by reconstituting data to match calculated checksums.

4 Checksum redundancy

A single checksum certainly has value. But who checks the checksums? Well, we happen to do that. Our solutions create a tree structure in which checksums themselves are checksummed (known as a merkle tree) allowing for very rapid validation of data integrity across large quantities of data.

5 Checksums and data security

While a checksum operation can detect data integrity errors, it probably shouldn't be your only method of data-centric security. The reason is that bad actors can actually modify data and, with enough time, recalculate checksums to match. So, adopt a broad security posture that includes checksums, snapshots, and backup & recovery but isn't exclusive of other forms of cybersecurity.

open DRIVES

“Data integrity is more than just ensuring that the data being read is exactly what was written. Data integrity is the basis for hugely important initiatives such as compliance, security, and operational efficiency.”

James Divito—
Director of Solutions @ OpenDrives

open DRIVES