Performance Analysis of Backup Storage Systems

Modern backup systems must balance deduplication ratio against ingest throughput. Inline deduplication indexes chunks during backup, while post-process deduplication performs indexing as a background task after initial storage.

Sparse indexing techniques reduce memory requirements by sampling the chunk index. Locality-based approaches exploit the observation that similar backups share chunk sequences, using bloom filters to identify candidate segments.

Restore performance depends heavily on chunk fragmentation. As data ages and is partially overwritten, chunks become scattered across storage containers. Forward referencing and container packing help mitigate fragmentation effects.