ABSTRACT

ARCHIVAL DATA STORAGE SYSTEM

A cost-effective, durable and scalable archival data storage system is provided herein that allow customers to store, retrieve and delete archival data objects, among other operations. For data storage, in an embodiment, the system stores data in a transient data store and provides a data object identifier may be used by subsequent requests. For data retrieval, in an embodiment, the system creates a job corresponding to the data retrieval and provides a job identifier associated with the created job. Once the job is executed, data retrieved is provided in a transient data store to enable customer download. In various embodiments, jobs associated with storage, retrieval and deletion are scheduled and executed using various optimization techniques such as load balancing, batch processed and partitioning. Data is redundantly encoded and stored in self-describing storage entities increasing reliability while reducing storage costs. Data integrity is ensured by integrity checks along data paths.

Figure 1