

ABSTRACT

STREAMING RESTORE OF A DATABASE FROM A BACKUP SYSTEM

A distributed data warehouse system may maintain data blocks on behalf of clients, and may store primary and secondary copies of each data block on different disks or nodes in a cluster. The warehouse system may back up data blocks in a remote key-value backup storage system. A restore operation may retrieve data blocks from backup storage using their unique identifiers as keys (while incoming queries are serviced) in response to a failure or a query targeting data that was lost or corrupted. The order in which data blocks are restored may be dependent on the relative likelihood that they will be accessed in the near future (e.g., based on how recently or frequently they were accessed, written, or backed up; the values of one or more access counters associated with each data block; or how recently a database table containing data in each data block was loaded).

[Figure 13]