

CLAIMS:

1. A method, comprising:

performing, by one or more computers:

storing columnar data of a database table in a plurality of physical data blocks in a distributed data storage system on behalf of one or more clients, wherein the distributed data storage system comprises a cluster of one or more nodes, each of which comprises one or more disks on which physical data blocks are stored, and wherein each of the plurality of physical data blocks is associated with a respective unique identifier;

storing a copy of each of the plurality of physical data blocks in a remote key-value durable backup storage system, wherein for each of the plurality of physical data blocks, the respective unique identifier serves as a key to access the data block in the remote key-value durable backup storage system;

detecting a failure in the distributed data storage system affecting two or more physical data blocks of the plurality of physical data blocks in which the columnar data was stored;

in response to said detecting, automatically restoring the two or more physical data blocks, wherein said restoring comprises:

analyzing a plurality of previously executed queries to determine a pattern of access to the columnar data stored in the distributed data storage system;

determining a priority order in which to restore the two or more physical data blocks based, at least in part, on the pattern of access, the priority order indicating a relative likelihood of subsequent access to each of the two or more physical data blocks;

retrieving a copy of one of the two or more physical data blocks having the highest priority from the remote key-value durable backup storage system, wherein said retrieving comprises using the respective unique identifier associated with the one of the two or

more physical data blocks as a key to access the copy of the one of the two or more physical data blocks in the key-value durable backup storage system;
writing a primary copy of the retrieved copy of the physical data block to a given disk on a given node in the distributed data storage system;
and
initiating replication of the retrieved copy of the physical data block on one or more disks in the distributed data storage system other than the given disk.

2. The method of claim 1, wherein said determining the priority order in which to restore the two or more physical data blocks is further based, at least in part, on a respective count value associated with each of the two or more physical data blocks indicating a number of recent accesses targeting the physical data block.

3. The method of claim 1, wherein said restoring further comprises updating an entry in a superblock data structure that stores information about the physical data blocks stored on the given node to indicate that the one of the two or more physical data blocks has been restored.

4. The method of claim 1, further comprising:
retrieving a copy of each of the two or more physical data blocks other than the one of the two or more physical data blocks from the remote key-value durable backup storage system in the determined priority order;
writing a primary copy of each of the retrieved copies of the two or more physical data blocks other than the one of the two or more physical data blocks to a disk in the distributed data storage system; and
initiating replication of each of the retrieved copies of the two or more physical data blocks other than the one of the two or more physical data blocks on one or more disks in the distributed data storage system other than the disk on which its primary copy was written.

5. A method, comprising:

performing, by one or more computers:

maintaining data in a plurality of physical data blocks of a data storage system on behalf of one or more clients, wherein each physical data block is associated with a unique identifier;

beginning an operation to restore data that was stored in two or more physical data blocks of the plurality of physical data blocks in the data storage system from copies of the two or more physical data blocks stored in a key-value storage system that is distinct from the data storage system;

determining a highest priority data block to be restored based, at least in part, on a pattern of access to the data stored in the data storage system, the pattern of access determined through an analysis of a plurality of previously executed queries of the data; and

restoring a copy of the highest priority data block to the data storage system from the key-value storage system prior to restoring copies of the two or more physical data blocks other than the highest priority data block.

6. The method of claim 5,

wherein the method further comprises detecting a failure in the data storage system affecting the two or more physical data blocks; and

wherein said beginning the operation to restore the data is performed in response to said detecting.

7. The method of claim 5,

wherein the method further comprises receiving a request from a client to restore the two or more physical data blocks; and

wherein said beginning the operation to restore the data is performed in response to said receiving.

8. The method of claim 5,

wherein the data storage system stores columnar data of one or more database tables in the plurality of physical data blocks on behalf of one or more clients; and
wherein said determining the highest priority data block to be restored comprises determining which of the one or more database tables was most recently loaded into the data storage system.

9. The method of claim 5,
wherein the data storage system stores columnar data of one or more database tables in the plurality of physical data blocks on behalf of one or more clients; and
wherein said determining the highest priority data block to be restored comprises determining which of the one or more database tables or columns thereof were most frequently queried in a recent time period.

10. The method of claim 5,
wherein the data storage system stores columnar data of one or more database tables in the plurality of physical data blocks on behalf of one or more clients; and
wherein said determining the highest priority data block to be restored comprises identifying the physical data block that stores the data for a same table or column thereof as a copy of the physical data block that was recently retrieved from the key-value storage system in response to receiving a query request targeting the data in the physical data block that was no longer available in the data storage system.

11. The method of claim 5, wherein said restoring the copy of the highest priority data block comprises:

retrieving a copy of the highest priority data block from the remote key-value storage system, wherein said retrieving comprises using the unique identifier associated with the highest priority data block as a key to access the copy of the highest priority data block in the key-value storage system;
writing a primary copy of the retrieved copy of the highest priority data block to a storage device in the data storage system.

12. The method of claim 11, further comprising:
prior to said writing:
receiving a query that targets the data maintained in the highest priority data block; and
satisfying the query, wherein satisfying the query comprises returning at least a portion of the data in the retrieved copy of the highest priority data block.
13. The method of claim 11, wherein said restoring the copy of the highest priority data block further comprises:
initiating replication of the highest priority data block on one or more storage devices in the data storage system other than the storage device on which the primary copy was written.
14. The method of claim 5,
wherein said restoring the copy of the highest priority data block comprises restoring the data stored in the highest priority data block from the key-value storage system to the data storage system while accepting and servicing queries directed to the data maintained on behalf of the one or more clients.
15. The method of claim 5, wherein the pattern of access is a historical pattern of access exhibited by the plurality of previously executed queries to target the plurality of physical data blocks.
16. The method of claim 5,
wherein said restoring is performed by a background process; and
wherein the method further comprises, while the background process performs said restoring:
a foreground process receiving one or more queries directed to the data maintained on behalf of the one or more clients; and
the foreground process servicing the one or more queries.

17. The method of claim 5,
wherein the method further comprises receiving a query that targets the data previously maintained in the two or more physical data blocks that are no longer available in the data storage system; and
wherein said beginning the operation to restore the data is performed in response to said receiving.
18. A computing system, comprising:
one or more computing nodes, each of which comprises at least one processor and a memory, wherein the one or more computing nodes are configured to collectively implement a database service; and
an interface to a remote key-value storage system;
wherein the database service is configured to:
store data on behalf of one or more subscribers to the database service in a plurality of physical data blocks on one or more storage devices, wherein each of the plurality of physical data blocks is associated with a unique identifier;
back up the data stored on behalf of the one or more subscribers, wherein to back up the data, the database service is configured to send to the remote key-value storage system, via the interface, a copy of each of the plurality of physical data blocks for storage in the remote key-value storage system, and wherein the unique identifiers associated with each of the plurality of physical data blocks are employable as access keys to access copies of the plurality of physical data blocks in the remote key-value storage system;
detect a condition or event that triggers a restore operation to restore two or more physical data blocks of the plurality of physical data blocks;
determine an order in which to restore the two or more physical data blocks, the order based, at least in part, on an access pattern exhibited by a plurality of previously executed queries to access the data; and
restore the two or more physical data blocks in the determined order.

19. The computing system of claim 18, wherein to determine the order in which to restore the two or more physical data blocks, the database service is configured to determine how recently each of the two or more physical data blocks was a target of a query.

20. The computing system of claim 18, wherein to determine the order in which to restore the two or more physical data blocks, the database service is configured to determine how recently each of the two or more physical data blocks was written to.

21. The computing system of claim 18, wherein to determine the order in which to restore the two or more physical data blocks, the database service is configured to determine how recently each of the two or more physical data blocks was backed up.

22. A non-transitory computer-readable storage medium storing program instructions that when executed on one or more computers cause the one or more computers to perform:

- maintaining data in a plurality of physical data blocks of a data storage system on behalf of one or more clients, wherein each physical data block is associated with a unique identifier;
- during a pre-determined time period:
 - receiving a query targeting data in a given one of the plurality of physical data blocks;
 - servicing the query, wherein servicing the query comprises accessing the given one of the plurality of physical data blocks; and
 - incrementing a counter associated with the given one of the plurality of physical data blocks, wherein a value of the counter indicates a number of times the given one of the plurality of physical data blocks was accessed within the pre-determined time period;
- beginning an operation to restore two or more of the plurality of physical data blocks including the given one of the plurality of physical data blocks;
- determining an order in which to restore the two or more of the plurality of physical data blocks based, at least in part, on the value of the counter associated with the given

one of the plurality of physical data blocks and the value of a respective counter associated with each of the two or more of the plurality of physical data blocks other than the given one of the plurality of physical data blocks.

23. The non-transitory computer-readable storage medium of claim 22, wherein said determining the order in which to restore the two or more of the plurality of physical data blocks is further dependent on the value of another respective counter associated with each of the plurality of physical data blocks, wherein the value of the other respective counter associated with each of the plurality of physical data blocks indicates the number of times that the physical data block was accessed within a previous pre-determined time period.

24. The non-transitory computer-readable storage medium of claim 23, wherein said determining the order in which to restore the two or more of the plurality of physical data blocks comprises computing a value representing a logical or mathematical combination of the values of the respective counter and the other respective counter associated with each of the two or more of the plurality of physical data blocks.