Big Data Storage Technologies

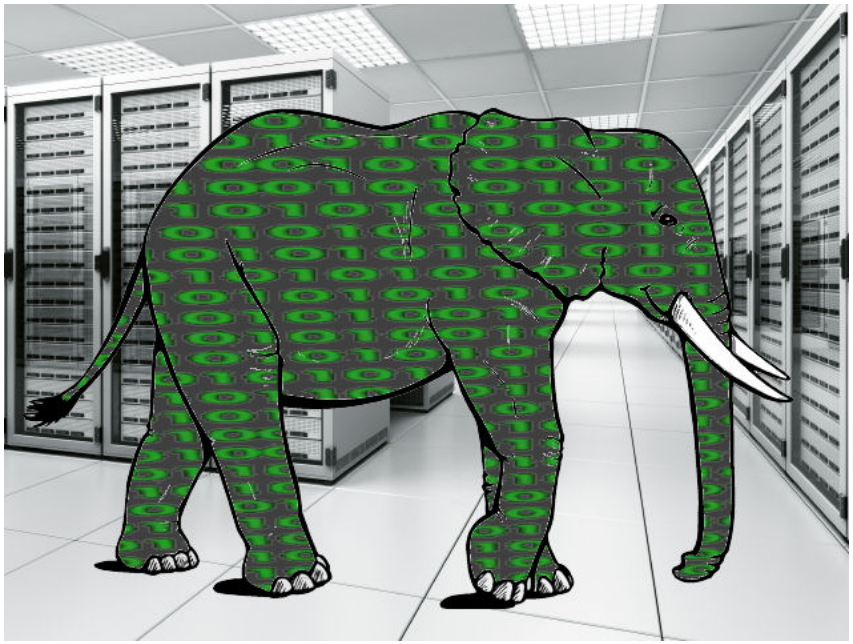
James Lee

The George Washington University

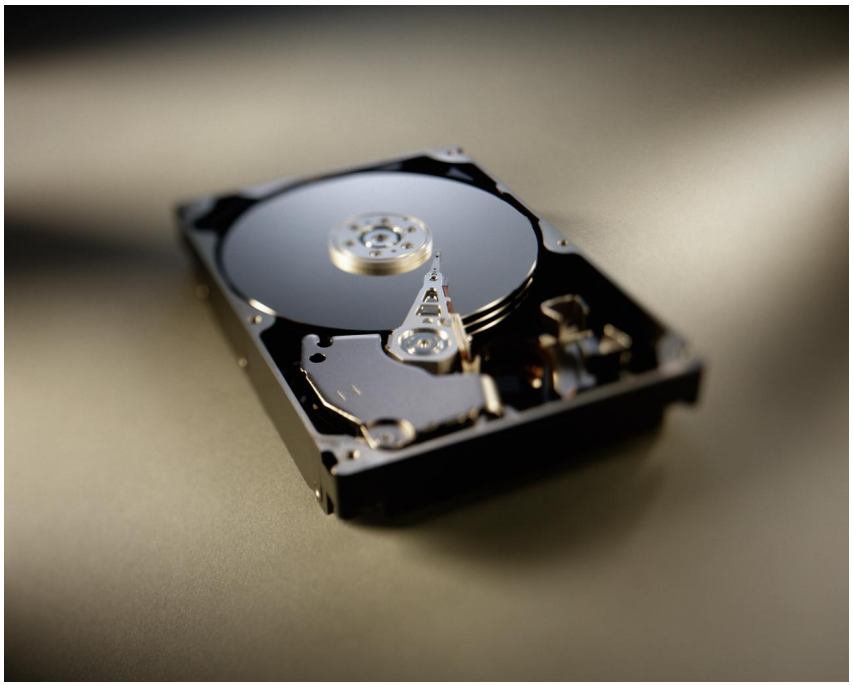
April 11, 2012

What is Big Data?

- When the size of the data grows to become as big of a problem to store and process as the problem you are trying to solve with the data.



Why are traditional filesystem insufficient?



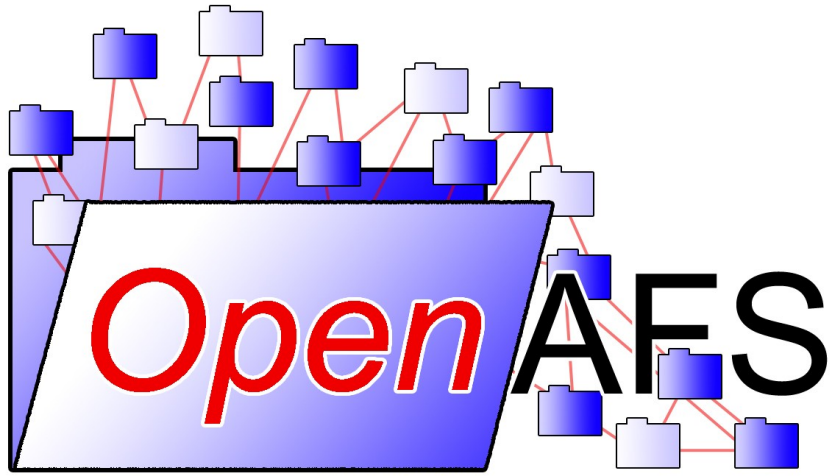
- Upper limit on filesystem size
- Limited redundancy
- Limited bandwidth

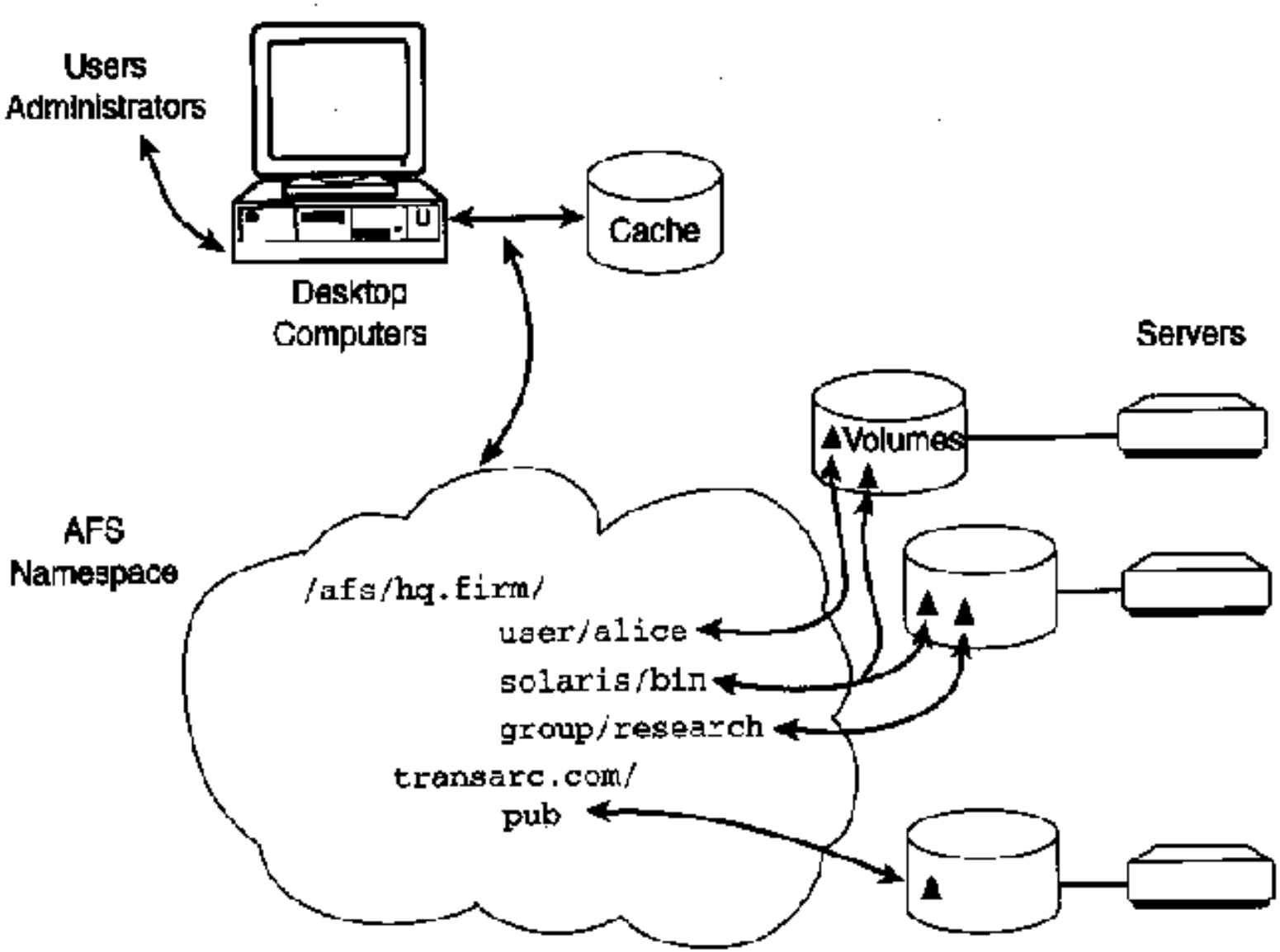
So what are the options for scaling out?

- Depends on business needs.
- Scale within a rack, within a datacenter, or across wide-area networks.
- Several different technologies available for achieving those goals.
- May have to make compromises in places.

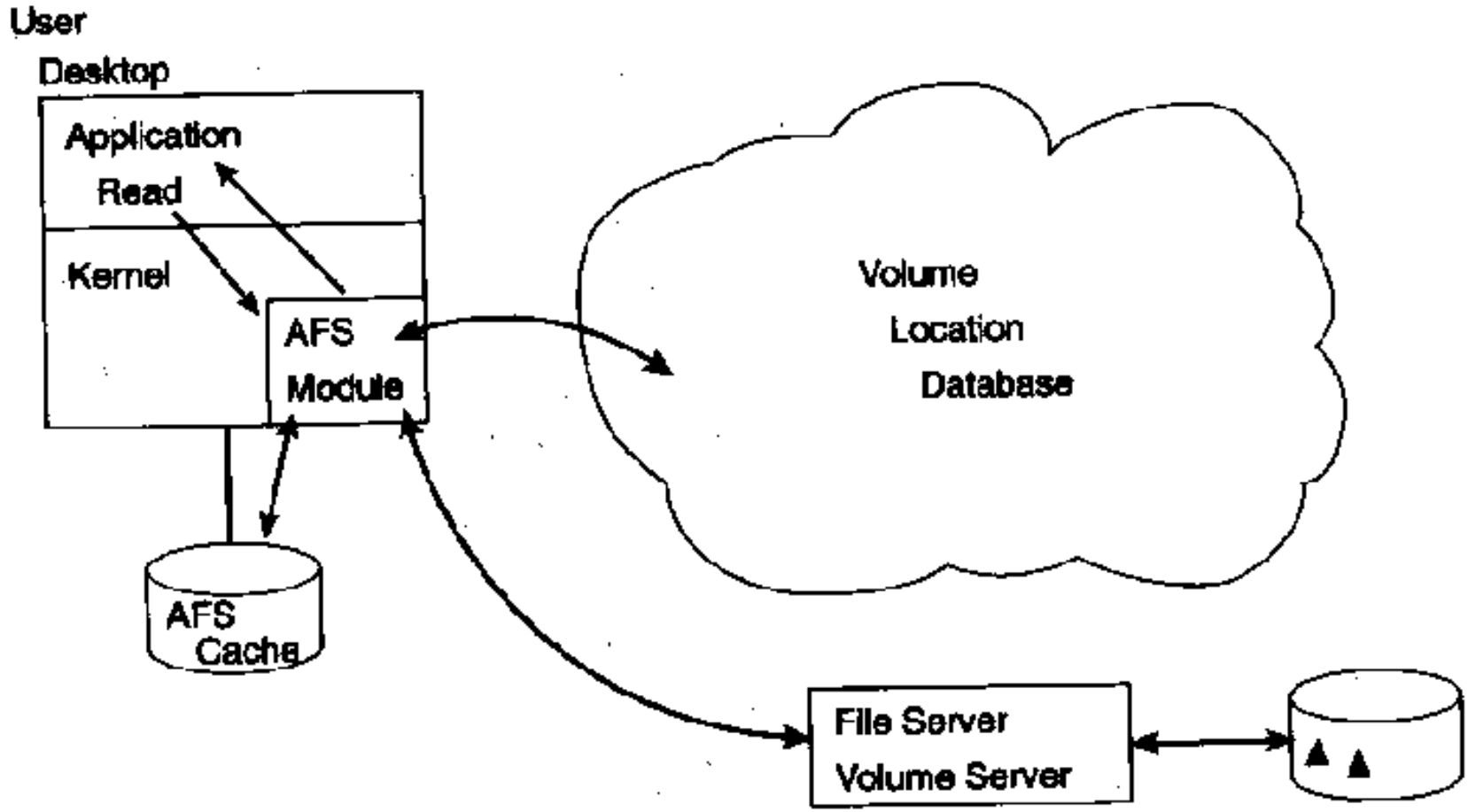
Andrew File System

- Distributed filesystem developed in 1980s.
- Used primarily by Universities.
- Has traditional filesystem semantics.
- Scales to hundreds of terabytes.



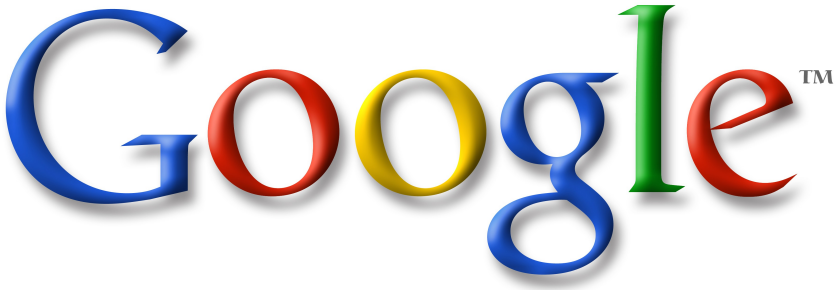


Source: http://caligari.dartmouth.edu/classes/afs/print_pages.shtml



Source: *http://caligari.dartmouth.edu/classes/afs/print_pages.shtml*

What does Google do?



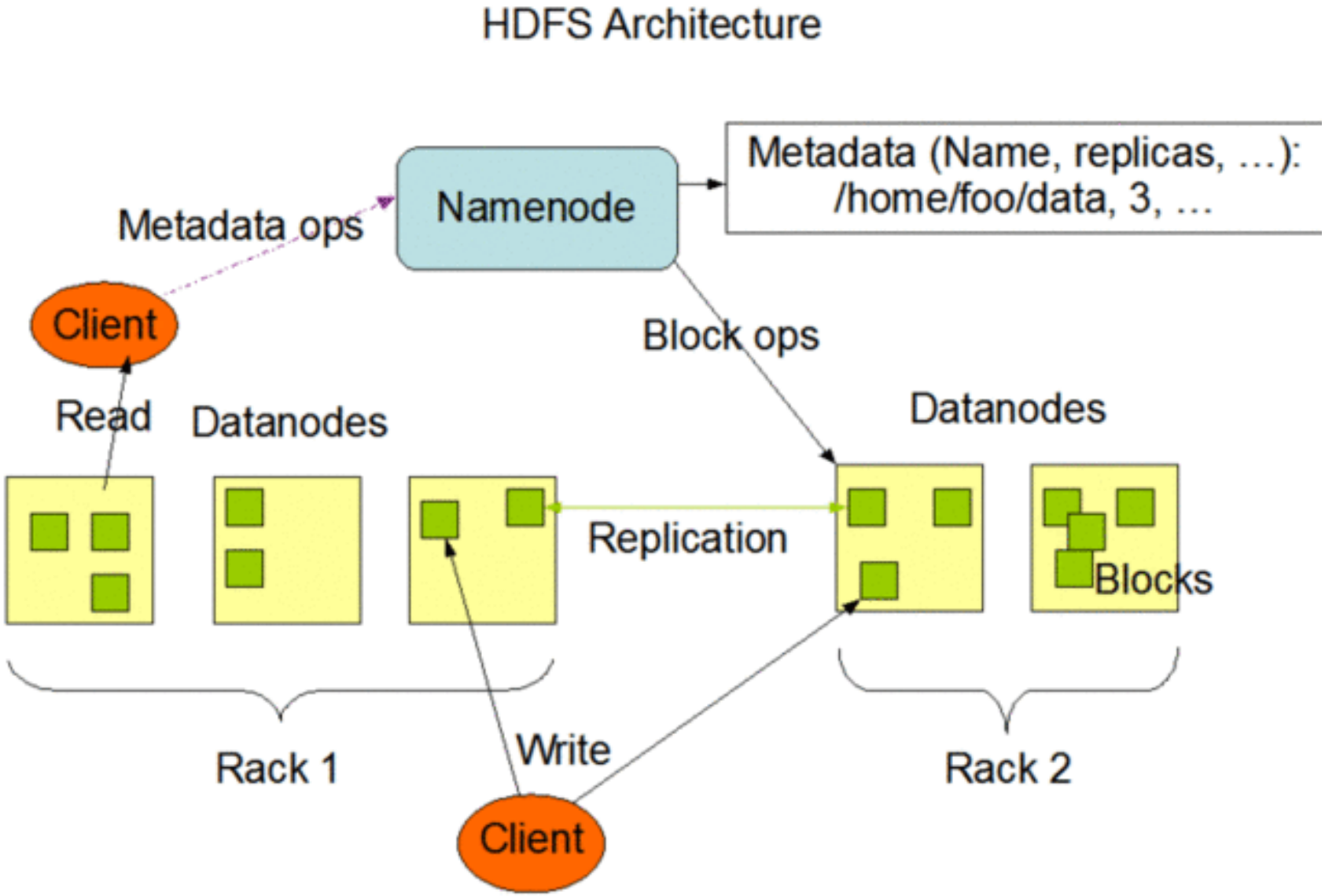
Look at Google's requirements:

- hundreds of millions of huge files
- have to be read very quickly
- writes less important
- have to be redundant, but not synchronous
- concurrent access to files should have low overhead

These ideas have been implemented in the Apache Hadoop project.

Hadoop

- Written in Java (no filesystem semantics)
- Stores files in large blocks (64 MB) that get lazily-replicated
- Rack-aware replication
- Master 'NameNode' tracks location of blocks
- Writes only optimized for appending data
- Scales to tens of thousands of nodes; $>$ 100 PB



Source: http://arst.ch/s9l

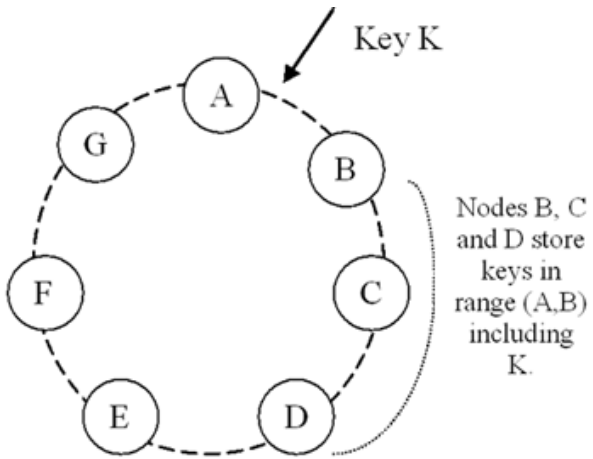
Amazon has very different requirements than a search engine:

- Willing to compromise on data consistency across system for HA
- Deal with more general-purpose data access
- Handle random access to smaller components

Amazon developed their own distributed FS called Dynamo.

Dynamo

- Decentralized, peer-to-peer architecture.
- System determines node to select by MD5 hash.
- Nodes always query neighbors for latest version.
- Implemented in Apache Cassandra project.



Source: http://arst.ch/s9l