Bigtable: A Distributed Storage System for Structured Data

Chang, Fay, Jeffrey Dean, Sanjay Ghemawat, Wilson Hsieh, Deborah Wallach, Mike Burrows, Tushar Chandra, and Andrew Fikes. "Bigtable: A Distributed Storage System for Structured Data." 2006.

Krzyzanowski, Paul. "Bigtable: A NoSQL massively parallel table." *Rutgers Computer Science*. N.p., 18 Sept 2013. Web.

Jonathan Pistilli

11/25/2013

Bigtable

- "A distributed storage system for managing structured data that is designed to scale to a very large size."

- Provides Google with significant advantages by having their own data model and allowing access to other Google systems that Bigtable uses as building blocks
  - Google File System and SSTable

Bigtable Implementation

- Bigtable is a large map that is indexed by a row key, column key, and a timestamp

- Building Blocks
    - Built on Google File System, used to store log and data files
    - Google SSTable internally stores Bigtable data
    - Chubby, a lock service that manages leases for resources and stores configuration information

- A client library is linked with user's code, there is a master server to monitor and coordinate activity for tablets, and large amounts of tablet servers (tablets manage sets of tables)

- Uses Chubby (distributed lock service that manages leases for resources and stores configuration information) to ensure there is only one active master, discover tablet servers, store Bigtable schema information and store access control lists

Analysis

- This system seems to be very efficient for Google systems

- Bigtable makes efficient use of other Google systems like Google File System and SStable

- The user interface is odd and less understandable than a relational database

Pros and Cons

- Pros
  - Works very well with other Google systems and is an excellent storage system for the uses that it was created for
  - Efficiently uses a master server to assign and control other servers, distributing work load evenly while allowing for mass amounts of data and information

- Cons
  - Does not allow for joins and there is no SQL
  - Does not support ACID transactions across row keys

Real-World Applications

- Google Analytics
    - Bigtable organizes website information and compresses size

- Google Earth
    - Stores the satellite imagery

- Personalized Search
    - Tracks users search history and allows users to search based of previous searches