BIGTABLE

Robert Mitola

BIGTABLE: A DISTRIBUTED STORAGE SYSTEM FOR STRUCTURED DATA

- PowerPoint by **Robert Mitola**.
- Article by Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber.
- For information contact: {fay,jeff,sanjay,wilsonh,kerr,m3b,tushar,fikes,gruber} @google.com
- Google, Inc.
- November 25, 2013.

MAIN IDEA

- Create a solution for managing structured data which must be highly scalable to petabytes across thousands of servers, and work with programs with large data sizes and latency requirements.
- Provide clients with a data model that gives control over data layout and formation, and allow clients to specify locality properties of the data represented in storage.
- A Bigtable is a “sparse, distributed, persistent multidimensional sorted map indexed by a row key, column key, and timestamp.”
  - (row:string, column:string, time:int64) → string
- Data is indexed using row and column names that can be uninterpreted, arbitrary strings. Every read and write of data under a single row key is atomic.
- Row ranges, known as tablets, can be used by the client to get a good locality for data accesses.
- Column families, column keys of usually the same data type grouped together, form the basic unit of access control.
- Uses Google File System to store log and data files and Chubby as a distributed lock service.

IMPLEMENTATION

- Bigtable consists of three major components.
  - Library linked to every client.
  - One master server which assigns tablets to tablet servers, detects addition or removal of tablet servers, balances tablet server load, performs garbage collection of Google File System files, and handles schema changes.
  - Many tablet servers, which can be added or removed from a cluster for changes in workloads.
- The Chubby file is put into a three-level hierarchy table similar to a B+ tree. This stores tablet location information.
- Each tablet is assigned to one tablet server at a time while the master server keeps track of the set of live tablet servers and their current assignment of tablets.
  - Master assigned unassigned tablets to an available tablet server.
- When a master server is started, it discovers its tablet assignments by interacting with the Chubby and scanning the METADATA table.
- The persistent state of a tablet is stored in the Google File System.
  - Tablet servers read a tablet's metadata from the METADATA table.

ANALYSIS

- The Bigtable distributed storage system is an effective way to manage a vast amount of structured data.
- Provides a high amount of scalability, which is key in all of the applications Bigtable is used for.
- Allowing users the ability to specify the locality properties of the data represented is good design for clients.
- Storage of data as uninterpreted strings saves memory.
- Implementation of Chubby and the three level storage hierarchy is beneficial for speed and efficiency.
- The way in which the master server handles changes in the tablet tables is effective.

ADVANTAGES & DISADVANTAGES

- ADVANTAGES
  - Master server is lightly loaded since clients do not rely on master for tablet location information.
  - “Minor compaction” of recently added data into older updates shrinks memory usage of tablet servers.
  - As number of tablet servers increase, so does aggregate throughput.
- DISADVANTAGES
  - Compression of SSTable blocks separately causes loss of space.
  - Read operations on SSTables not in memory causes many disk accesses to be performed.
  - Noticeable decrease in per-server throughput as the number of tablet servers increases.

REAL WORLD USE CASES

- Used by more than 60 Google products.
- Google Analytics
  - Webmasters use this to analyze traffic patterns in their websites.
  - Program is attached to web pages via insertable JavaScript.
- Google Earth
  - Gives users the ability to access satellite photos of the earth on an interactive globe.
  - Raw photos are stored in one table, which contains 70 terabytes of data.
  - Uses one table for servicing the indexing of data stored in Google File System.
- Personalized Search
  - Service that records user queries and clicks across a variety of Google properties.
  - Stores each user's data and actions in Bigtable.