

Bigtable: A Distributed Storage System for Structured Data

Google Inc.

A Comparison of Approaches to Large- Scale Data Analysis

SIGMOND'09, Copyright 2009

Nicole Rae
May 8, 2015

1 - Google's Bigtable

- A distributed storage system providing clients with a simple data model enabling them to manage structured data with dynamic control over layout, format, and whether to serve data out of memory or from disk.
- By April of 2005, Google's Bigtable had achieved goals including wide applicability, scalability, high performance, and high availability.
- A sparse and persistent multi-dimensional map, indexed by a row key, whose row range is the unit of distribution and load balancing (tablet), a column key, which is grouped within a set called column family, and a 64-bit timestamp.
- Bigtable uses the Google File System to store and log data files, the Google *SSTable* file format to internally store data, and a distributed lock service known as Chubby.

Implementation

- Several tables store data that is served to users, while others are used to store data for batch processing; all tables ranging in size, cell size, percentage of data served from memory, and complexity of the table schema.
- Google Analytics
 - Analysis of traffic patterns on websites by providing webmasters with aggregate statistics such as visitors per day, page views and site-tracking reports.
 - Uses the two tables *raw click table*, storing the website's name and creation time of the session, and *summary table*, containing predefined summaries which are generated from the *raw click table* by periodically scheduled MapReduce jobs.
- Google Earth
 - Provides users with high-resolution satellite imagery and the ability to navigate the world's surface
 - Utilizes MapReduce over Bigtable to structure data with each row in the imagery table corresponding to a single geographic segment, along with Google File System to serve tens of thousands of queries per second per datacenter with low latency.
- Personalized Search
 - An optional service that records user queries and clicks across Google properties such as web search, images and news.
 - Each user is assigned a row named by a unique userid and receive profiles generated using a MapReduce over Bigtable.

Analysis

- Being a Google system, Bigtable has the capability to scale across hundreds or thousands of servers that collectively can store petabytes of data.
- The use of tablet servers allows for a company to dynamically add or remove those servers from a cluster to accommodate changes in workloads.
- Bigtable branches away from the traditional approach and focuses mainly on simplicity and robustness.
- The fact that this system is used to run one of the craziest Google products, Google Earth, is enough to sell me.

2 - MapReduce vs. DBMS

- Both systems run on a “shared nothing” collection of computers.
- A relational database management system requires a structured schema and SQL programming, whereas Hadoop’s MapReduce allows flexibility in regards to schema and programming model.
- In regards to indexing, MapReduce allows for creation of original indexes, where in DBMS a traditional hash or b-tree index is supported.
- For data distribution, DBMS uses a parallel query optimizer to seek balance of computational workloads, whereas MapReduce must perform these tasks manually.
- Fault tolerance for MapReduce is significantly more sophisticated because if one task fails, that one specific task can be restarted; In DBMS, if one task fails, the entire query must be restarted.

Implementation

- System Installation, Configuration & Tuning
 - Parallel DBMSs are found to be much more challenging than Hadoop's MR to install and configure properly.
 - In regards to tuning, in DBMSs it is done prior to query execution, whereas in MR occasional tuning is required for the system as a whole, but potentially also each individual task.
- Data Loading
 - In DBMS, the data was loaded on each node sequentially, therefore as the amount of data increases, so does the load time.
 - In MR, each node is copying all data files from the local disk and distributing replicas to other nodes in the cluster, making load time much smaller.
- Execution
 - DBMS performs full table scans on data with the use of a complete query plan.
 - Within MR, the immense use of control messages exist to search for patterns within data; the Reduce function takes the output of the first function and compresses data into single files.

Analysis

- Due to the start up cost advantage of Hadoop's MR, the rapid popularity it has gained did not surprise me.
- Although SQL is a powerful tool, many people find it to be difficult to use due to it's complexity.
- Even though the setup and installation process for MR was deemed easier, the relative performance that was provided by the parallel database systems was impressive.

Bigtable vs. Large-Scale Data Analysis

- The first paper I read focused on Google's distributed storage system, where in the second one, Hadoop's systems and the traditional DBMS were discussed.
- In both Bigtable and Hadoop, the MapReduce function is implemented – however, there are two different versions (Google's said to be faster).
- On a minimal level, all three systems will get the job done without much of a difference being noticed by the user. However large scale, I believe Bigtable's implementation is the most "distributed".

Michael Stonbraker – “10 Year Test of Time”

- One Size Fits None!
- The traditional use of row stores are soon to be obsolete within all markets listed below:
 - Data Warehouses
 - NoSQL Markets
 - Complex Analytics
 - Streaming Markets
 - Graph Analytics
- When gathering influential ideas on new implementations of a database management system, keep in mind the following:
 - Non-Volatile RAM
 - Big main memory
 - Processor diversity
 - Higher speed networks
 - Low Level Virtual Machine
 - Vectorization
- Strive to be a data scientist – it's a great time for it!

Advantages & Disadvantages

- Advantages of Bigtable
 - Supports access control at column family level
 - Index building is extremely powerful
 - High scalability and availability
 - With no assembly required, it is time efficient
 - NoSQL - Uses the programming model of MapReduce
- Disadvantages of Bigtable
 - No joins or consistency constraint checks
 - Technically only offered within Google, no open source copies.
 - No consistent guarantees for any updates aside from single row transactions.
 - Unpredictable costs