



BigTable: Introduction

Let's explore Bigtable and its use cases.

We'll cover the following ^

- Goal
- What is BigTable?
- Background
- BigTable use cases

Goal#

Design a distributed and scalable system that can store a huge amount of structured data. The data will be indexed by a row key where each row can have an unbounded number of columns.

What is BigTable?#

BigTable is a **distributed** and **massively scalable** wide-column store. It is designed to store huge sets of structured data. As its name suggests, BigTable provides storage for very big tables (often in the terabyte range).

In terms of the CAP theorem, BigTable is a CP system, i.e., **it has strictly consistent reads and writes**. BigTable can be used as an input source or output destination for MapReduce (https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html) jobs.

Background#



BigTable was developed at Google and has been in use since 2005 in dozens of Google services. Because of the large scale of its services, Google could not use commercial databases. Also, the cost of using an external solution would have been too high. That is why Google chose to build an in-house solution. BigTable is a highly available and high-performing database that powers multiple applications across Google — where each application has different needs in terms of the size of data to be stored and latency with which results are expected.

Though BigTable is not open-source itself, its paper was crucial in inspiring powerful open-source databases like Cassandra (<https://cassandra.apache.org/>) (which borrows BigTable's data model), HBase (<https://hbase.apache.org/>) (a distributed non-relational database) and Hypertable (<https://hypertable.org/>).

BigTable use cases#

Google built BigTable to store large amounts of data and perform thousands of queries per second on that data. Examples of BigTable data are billions of URLs with many versions per page, petabytes of Google Earth data, and billions of users' search data.

BigTable is suitable to store large datasets that are greater than one TB where each row is less than 10MB. Since BigTable does not provide ACID (atomicity, consistency, isolation, durability) properties or transaction support, Online Transaction Processing (OLTP) (https://en.wikipedia.org/wiki/Online_transaction_processing) applications with transaction processes should not use BigTable. For BigTable, data should be structured in the form of key-value pairs or rows-columns. Non-structured data like images or movies should not be stored in BigTable.

Here are the examples of data that Google stores in BigTable:



- URL and its related data, e.g., PageRank, page contents, crawl metadata (e.g., when a page was crawled, what was the response code, etc.), links, anchors (links pointing to a page). There are billions of URLs with many versions of a page.
- Per-user data, e.g., preference settings, recent queries/search results. Google has hundreds of millions of users.

BigTable can be used to store the following types of data:

1. Time series data: As the data is naturally ordered
2. Internet of Things (IoT) data: Constant streams of writes
3. Financial Data: Often represented as time-series data

← Back

Mock Interview: HDFS

Next →

BigTable Data Model



Mark as Completed



Report an Issue