Bigtable overview

Cloud Bigtable is a sparsely populated table that can scale to billions of rows and thousands of columns, enabling you to store terabytes or even petabytes of data. A single value in each row is indexed; this value is known as the row key. Bigtable is ideal for storing very large amounts of single-keyed data with very low latency. It supports high read and write throughput at low latency, and it is an ideal data source for MapReduce operations.

Bigtable is exposed to applications through multiple client libraries, including a supported extension to the [Apache HBase library for Java](). As a result, it integrates with the existing Apache ecosystem of open-source Big Data software.

Bigtable's powerful back-end servers offer several key advantages over a self-managed HBase installation:

- **Incredible scalability.** Bigtable scales in direct proportion to the number of machines in your cluster. A self-managed HBase installation has a design bottleneck that limits the performance after a certain threshold is reached. Bigtable does not have this bottleneck, so you can scale your cluster up to handle more reads and writes.
- **Simple administration.** Bigtable handles upgrades and restarts transparently, and it automatically maintains high [data durability](). To replicate your data, simply add a second cluster to your instance, and replication starts automatically. No more managing replicas or regions; just design your table schemas, and Bigtable will handle the rest for you.
- **Cluster resizing without downtime.** You can increase the size of a Bigtable cluster for a few hours to handle a large load, then reduce the cluster's size again—all without any downtime. After you change a cluster's size, it typically takes just a few minutes under load for Bigtable to balance performance across all of the nodes in your cluster.

What it's good for

Bigtable is ideal for applications that need very high throughput and scalability for key/value data, where each value is typically no larger than 10 MB. Bigtable also excels as a storage engine for batch MapReduce operations, stream processing/analytics, and machine-learning applications.

You can use Bigtable to store and query all of the following types of data:

- **Time-series data,** such as CPU and memory usage over time for multiple servers.
- **Marketing data,** such as purchase histories and customer preferences.
- **Financial data,** such as transaction histories, stock prices, and currency exchange rates.
- **Internet of Things data,** such as usage reports from energy meters and home appliances.
- **Graph data,** such as information about how users are connected to one another.

Bigtable storage model

	Column family 1		Column family 2		
	Column 1	Column 2	Column 1	Column 2	
Row key 1	☐	☐		☐	t1 / t2 / t3
Row key 2	☐	☐	☐		