

Introduction to NoSql HBase

The important of Databases :

* Requirements of a Database

- 1) Structured: Rows & columns.
- 2) Random access: Update one row at a time.
- 3) Low latency: Very fast read/write/update operations.
- 4) ACID compliant: Ensure data integrity.

* What are ACID properties?

1) Atomicity:

Transactions on a database should be all-or-nothing.

E.g. Transferring Money

"Both withdrawal and deposit should occur or none at all".

2> Consistency:

Database updates should not violate any constraints.

E.g. Enrolling Students

"Every student should have a unique student id".

3> Isolation:

Concurrent operations on the database should appear as though they were applied in some sequence.

4> Durability:

Safety of Data.

In case of power loss, crashes, errors.

unfortunately, Hadoop makes a very poor database.

* Limitations of Hadoop

1) Unstructured data

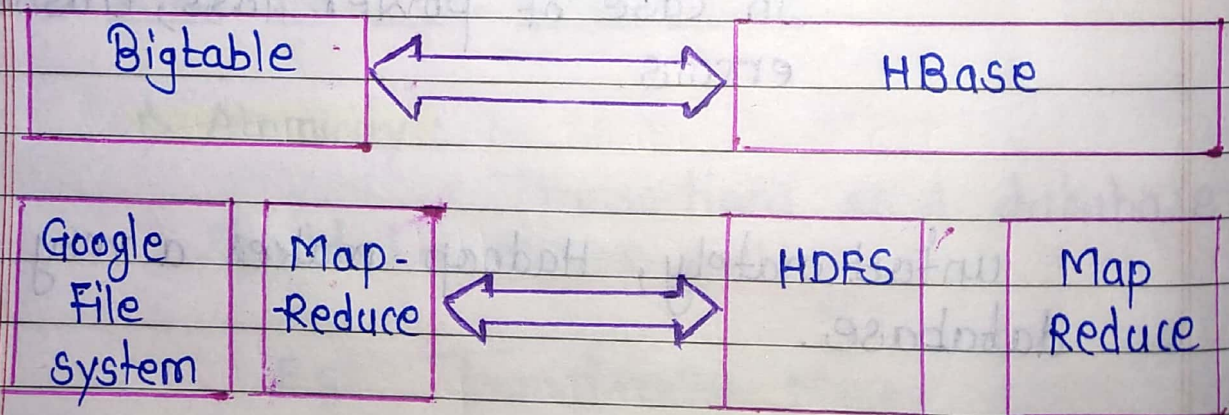
2) No random access

3) High latency

4) Not ACID compliant

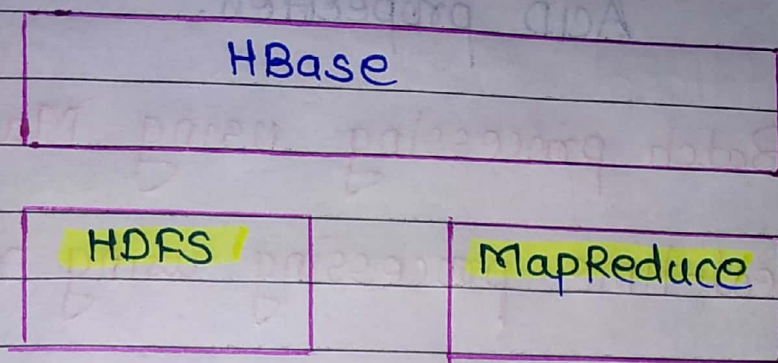
Google published a paper on Bigtable a distributed storage system for structured data.

* How did google solve this for search?



Page No. _____
Date _____

HBase is a distributed database management system which runs on top of Hadoop.



HBase: HBase is Nosql database.

1) Distributed : Stores data in HDFS.

2) Scalable : Capacity directly proportional to number of nodes in the cluster.

3) Fault tolerant: piggybacks on Hadoop.

4) Structured: A loose data structured.

5) Low latency: Real-time access using row based indices called row keys.

6) Random access: Row keys allow access updates to one record.

7) Somewhat ACID compliant:

Some transactions will have ACID properties.

Batch processing using MapReduce.

Real-time processing using row keys

* properties of HBase:

1) Columnar store.

2) Denormalized storage.

3) Only CRUD operations.

4) ACID at the row level.

Advantage of a columnar store:

① Sparse tables:

No wastage of space when storing sparse data.

② Dynamic attributes:

dynamically update attributes without changing storage structure.

Denoormalized Storage:

Read a single record to get all details about an employee in one read operations.

CRUD operations:

Only a limited set of operations are allowed in HBase.

- 1) Create
- 2) Read
- 3) Update
- 4) Delete

* all details need to be self contained in one row.

ACID at the row level:

* updates to a single row are atomic

All columns in a row are updated or none are.

* updates to multiple rows are not atomic.

Even if the updates is on the same column in multiple rows.

Traditional RDBMS vs HBase

RDBMS

- 1) Data arranged in rows & columns
- 2) Supports Sql.
- 3) Complex queries such as grouping, aggregates, joins etc.
- 4) Normalized storage to minimize redundancy & optimize space
- 5) ACID compliant

HBase

- 1) Data arranged in a Column-wise manner.
- 2) Support NoSql database.
- 3) only basic operations such as create, read, update, delete.
- 4) Denormalized storage to minimize disk seeks.
- 5) ACID compliant at the row level.