HBase

Column Oriented - Distributed Databases

NoSQL Support

HBase is a distributed database part of Hadoop Ecosystem. It utilizes HDFS to store data and metadata. Also temporary files pertaining to the framework are stored in a local configured directory.

It is important to understand the purpose of columnar storage. Know that the way arrays can be physically designed to store as row vs column major, similarly records pertaining to databases can be designed with physical storage. RDBMS stores data physically row after another row. But with respect to data analytics many times the complete column is focused and having all values within a single column physically stored next to each other shall help in performance w.r.to disk I/O.

Further requirement incurred development of a concept called ‘column family’. Instead of having just a single atomic column, a column family contains a set of related columns. All the data within the column family physically stored next to each other. Of course this helps during the analytics.

While working on HBase Data Model it is very important to consider column families correctly.

Know that HBase queries are automatically converted to MapReduce Task behind the scene to get you records.

HBase is considered an online database where other thand batch or sequential access to HDFS data is made possible.

To interact to HBase you can use:

Hbase shell

start-dfs.sh
start-yarn.sh

start-hbase.sh

hbase shell

create 'student', 'id', 'name'

put 'student','r1','id',1

put 'student','r1','name','James'

put 'student','r2','id',2

put 'student','r2','name','John'

put 'student','r3','id',3

put 'student','r3','name','Dave'

scan 'student', {LIMIT=>2}

disable 'student'

drop 'student'

stop-all.sh

Or Java Programming API

And to monitor via gui

http://localhost:16010/master-status

Important link:

http://hbase.apache.org

Documented By:

Jigar M. Pandya
https://www.linkedin.com/in/jigar-pandya

Document Last updated: July 31, 2020.