

HBase

HBase is an open source, multidimensional, distributed, scalable and a NoSQL database written in Java. HBase runs on top of HDFS (Hadoop Distributed File System) and provides BigTable like capabilities to Hadoop. It is designed to provide a fault tolerant way of storing large collection of sparse data sets.

It is a schema-less database which contains keys and values. Each key, points to a value which is an array of bytes, can be a string, BLOB, XML, etc.

Features of HBase

- HBase is linearly scalable.
- It has automatic failure support.
- It provides consistent read and writes.
- It integrates with Hadoop, both as a source and a destination.
- It has easy java API for client.
- It provides data replication across clusters.

Properties of HBase:

1. Columnar Storage
2. Denormalization Storage
3. Only CRUD operations
4. ACID properties follow at only row level

Where to Use HBase

- Apache HBase is used to have random, real-time read/write access to Big Data.
- It hosts very large tables on top of clusters of commodity hardware.
- Apache HBase is a non-relational database modeled after Google's Bigtable. Bigtable acts up on Google File System, likewise Apache HBase works on top of Hadoop and HDFS.

HBase Architecture and its Important Components

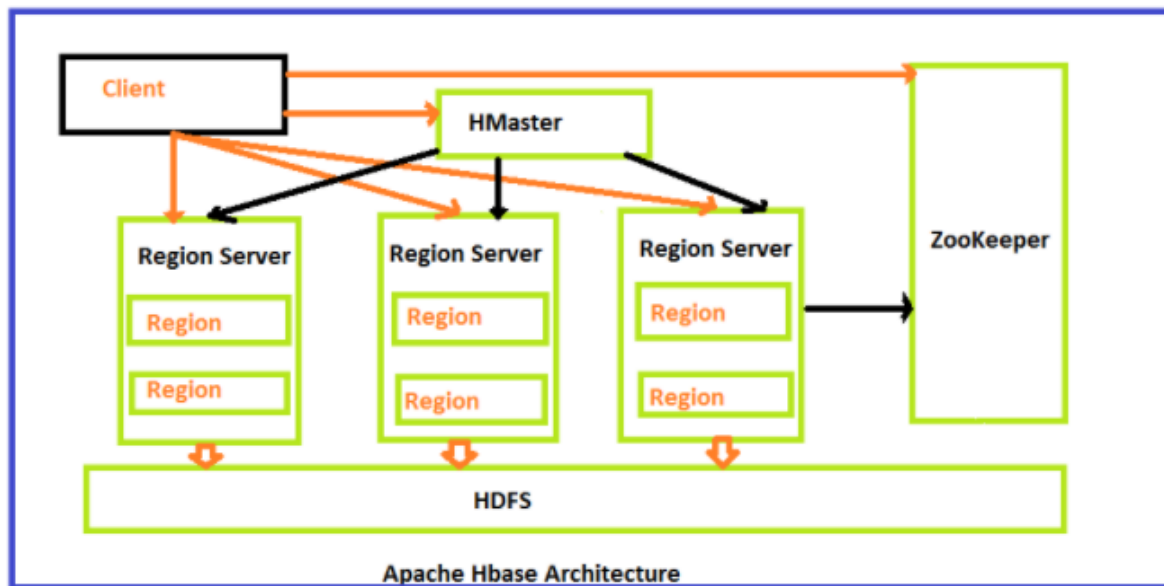


Fig: HBase Architecture

HBase architecture consists mainly of four components

- HMaster
- HRegionserver
- HRegions
- Zookeeper
- HDFS

HMaster:

Role of HMaster in HBase:

- Plays a vital role in terms of performance and maintaining nodes in the cluster.
- HMaster provides admin performance and distributes services to different region servers.
- HMaster assigns regions to region servers.
- HMaster has the features like controlling load balancing and failover to handle the load over nodes present in the cluster.
- When a client wants to change any schema and to change any Metadata operations, HMaster takes responsibility for these operations.

HBase Regions:

Role of HBase Regions in HBase:

- It contains distribution of tables and are comprised of Column families.
- It contains multiple stores, one for each column family.
- It consists of mainly two components, which are Memstore and Hfile.

HBase Regions Servers:

Role of HBase Regions Servers in HBase:

- Hosting and managing regions
- Splitting regions automatically
- Handling read and writes requests
- Communicating with the client directly

ZooKeeper:

Role of ZooKeeper in HBase:

- Maintains Configuration information
- Provides distributed synchronization
- Client Communication establishment with region servers
- Provides ephemeral nodes for which represent different region servers
- Master servers usability of ephemeral nodes for discovering available servers in the cluster
- To track server failure and network partitions

HDFS:

Role of HDFS in HBase:

- It provides a distributed environment for the storage.
- It is a file system designed in a way to run on commodity hardware.
- It stores each file in multiple blocks and to maintain fault tolerance, the blocks are replicated across a Hadoop cluster.

CRUD operations using Hbase

Creating a Table

We can create a table using the create command.

Syntax: `create '<table name>', '<column family>'`

Example:

```
2.7.1 :003 > create 'census', 'personal', 'professional'
Created table census
Took 1.3996 seconds
=> Hbase::Table - census
2.7.1 :004 > list
TABLE
census
1 row(s)
Took 0.0281 seconds
=> ["census"]
```

Inserting Data

We can insert data in a table using the put command.

Syntax: `put '<table name>', 'row1', '<colfamily:colname>', '<value>'`

Example:

```
2.7.1 :018 > put 'census', 2, 'professional:education', 'Post Graduate'
Took 0.0122 seconds
2.7.1 :019 > scan 'census'
ROW COLUMN+CELL
1 column=personal:gender, timestamp=2021-05-15T01:36:29.295, value=Male
1 column=personal:marital_status, timestamp=2021-05-15T01:36:07.713, value=Unmarried
1 column=personal:name, timestamp=2021-05-15T01:35:21.895, value=Shivam
1 column=professional:education, timestamp=2021-05-15T01:40:58.776, value=High School
1 column=professional:employed, timestamp=2021-05-15T01:40:15.205, value=Yes
2 column=personal:gender, timestamp=2021-05-15T01:43:22.452, value=Female
2 column=personal:marital_status, timestamp=2021-05-15T01:42:59.151, value=Married
2 column=personal:name, timestamp=2021-05-15T01:42:36.123, value=Tanvi
2 column=professional:education, timestamp=2021-05-15T01:45:01.538, value=Post Graduate
2 column=professional:employed, timestamp=2021-05-15T01:43:36.967, value=Yes
```

Reading Data

We can read data from a table using the get command.

Syntax: `get '<table name>', 'row1'`

Example:

```
2.7.1 :022 > get 'census', 2
COLUMN                                CELL
personal:gender                       timestamp=2021-05-15T01:43:22.452, value=Female
personal:marital_status               timestamp=2021-05-15T01:42:59.151, value=Married
personal:name                         timestamp=2021-05-15T01:42:36.123, value=Tanvi
professional:education                timestamp=2021-05-15T01:45:01.538, value=Post Graduate
professional:employed                 timestamp=2021-05-15T01:43:36.967, value=Yes
1 row(s)
```

Deleting Data From Table

We can read data from a table using the command.

Syntax: `delete '<table name>', '<row>', '<column name >'`

Example:

```
2.7.1 :036 > delete 'census', 1, 'personal:marital_status'
Took 0.0228 seconds
2.7.1 :037 > scan 'census'
ROW                                COLUMN+CELL
1                                column=personal:gender, timestamp=2021-05-15T01:36:29.295, value=Male
1                                column=personal:name, timestamp=2021-05-15T01:35:21.895, value=Shivam
1                                column=professional:education, timestamp=2021-05-15T01:40:58.776, value=High School
1                                column=professional:employed, timestamp=2021-05-15T01:40:15.205, value=Yes
2                                column=personal:gender, timestamp=2021-05-15T01:43:22.452, value=Female
2                                column=personal:marital_status, timestamp=2021-05-15T01:42:59.151, value=Married
2                                column=personal:name, timestamp=2021-05-15T01:42:36.123, value=Tanvi
2                                column=professional:education, timestamp=2021-05-15T01:45:01.538, value=Post Graduate
2                                column=professional:employed, timestamp=2021-05-15T01:43:36.967, value=Yes
```

Deleting Table

We can delete a table using the drop command before deleting the table we have to disable it first.

Syntax: `drop '<Table_Name>'`

```
2.7.1 :038 > disable 'census'
Took 1.3278 seconds
2.7.1 :039 > drop 'census'
Took 1.3901 seconds
2.7.1 :040 > list
TABLE
0 row(s)
Took 0.0231 seconds
=> []
```

CRUD operations using Hbase Java API

Creating a Table

Create Table Code:

```
hbase/pom.xml *GettingStarted.java
1 package hbase;
2
3 import java.io.*;
4
5
6
7
8
9
10
11 public class GettingStarted {
12     public static void main(String[] args) throws IOException {
13
14         Configuration conf = HBaseConfiguration.create();
15         Connection connection = ConnectionFactory.createConnection(conf);
16         try {
17             Admin admin = connection.getAdmin();
18             HTableDescriptor tableName = new HTableDescriptor(TableName.valueOf("census"));
19             tableName.addFamily(new HColumnDescriptor("personal"));
20             tableName.addFamily(new HColumnDescriptor("professional"));
21
22             if (!admin.tableExists(tableName.getTableName())) {
23                 System.out.println("Creating the Census Table");
24                 admin.createTable(tableName);
25                 System.out.println("Done");
26             }
27             else {
28                 System.out.println("Table Already Exists");
29             }
30         }
31         finally {
32             connection.close();
33         }
34     }
35 }
36
37
```

Output:

```
2.7.1 :002 > list
TABLE
census
1 row(s)
Took 0.0451 seconds
=> ["census"]
```

Inserting Data

Insert Data Code:

```
hbase/pom.xml *GettingStarted.java
1 package hbase;
2
3 import java.io.*;
10
11 public class GettingStarted {
12     private static byte[] PERSONAL_CF = Bytes.toBytes("personal");
13     private static byte[] PROFESSIONAL_CF = Bytes.toBytes("professional");
14     private static byte[] NAME_COLUMN = Bytes.toBytes("name");
15     private static byte[] GENDER_COLUMN = Bytes.toBytes("gender");
16     private static byte[] MARITAL_STATUS_COLUMN = Bytes.toBytes("marital_status");
17     private static byte[] EMPLOYED_COLUMN = Bytes.toBytes("employed");
18     private static byte[] FIELD_COLUMN = Bytes.toBytes("field");
19
20     public static void main(String[] args) throws IOException {
21         Configuration conf = HBaseConfiguration.create();
22         Connection connection = ConnectionFactory.createConnection(conf);
23         Table table = null;
24         try {
25             table = connection.getTable(TableName.valueOf("census"));
26             Put put1 = new Put(Bytes.toBytes("1"));
27             put1.addColumn(PERSONAL_CF, NAME_COLUMN, Bytes.toBytes("Mike Jon"));
28             put1.addColumn(PERSONAL_CF, GENDER_COLUMN, Bytes.toBytes("Male"));
29             put1.addColumn(PERSONAL_CF, MARITAL_STATUS_COLUMN, Bytes.toBytes("Married"));
30             put1.addColumn(PROFESSIONAL_CF, EMPLOYED_COLUMN, Bytes.toBytes("Yes"));
31             put1.addColumn(PROFESSIONAL_CF, FIELD_COLUMN, Bytes.toBytes("Construction"));
32             table.put(put1);
33             Put put2 = new Put(Bytes.toBytes("2"));
34             put2.addColumn(PERSONAL_CF, NAME_COLUMN, Bytes.toBytes("Dolcy"));
35             put2.addColumn(PERSONAL_CF, GENDER_COLUMN, Bytes.toBytes("Female"));
36             put2.addColumn(PROFESSIONAL_CF, EMPLOYED_COLUMN, Bytes.toBytes("Yes"));
37             Put put3 = new Put(Bytes.toBytes("3"));
38             put3.addColumn(PERSONAL_CF, NAME_COLUMN, Bytes.toBytes("David"));
39             put3.addColumn(PERSONAL_CF, GENDER_COLUMN, Bytes.toBytes("Male"));
40             put3.addColumn(PERSONAL_CF, MARITAL_STATUS_COLUMN, Bytes.toBytes("Unmarried"));
41             List<Put> putList = new ArrayList<Put>();
42             putList.add(put1);
43             putList.add(put2);
44             putList.add(put3);
45             table.put(putList);
46             System.out.println("Inserted row for Mike Jon, Dolcy, David");
47         } finally {
48             connection.close();
49             if (table != null) {
50                 table.close();
51             }
52         }
53     }
54 }
```

Output:

```
2.7.1 :007 > scan 'census'
ROW COLUMN+CELL
1 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Male
1 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Married
1 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Mike Jon
1 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
1 column=professional:field, timestamp=2021-05-15T19:04:42.214, value=Construction
2 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Female
2 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Dolcy
2 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
3 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Male
3 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Unmarried
3 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=David
3 row(s)
```

Reading Data

Read Data from Table Code:

```
hbase/pom.xml GettingStarted.java
1 package hbase;
2
3 import java.io.*;
4
10
11 public class GettingStarted {
12     private static byte[] PERSONAL_CF = Bytes.toBytes("personal");
13     private static byte[] PROFESSIONAL_CF = Bytes.toBytes("professional");
14     private static byte[] NAME_COLUMN = Bytes.toBytes("name");
15     private static byte[] FIELD_COLUMN = Bytes.toBytes("field");
16
17     public static void main(String[] args) throws IOException {
18         Configuration conf = HBaseConfiguration.create();
19         Connection connection = ConnectionFactory.createConnection(conf);
20         Table table = null;
21         try {
22             table = connection.getTable(TableName.valueOf("census"));
23             Get get = new Get(Bytes.toBytes("1"));
24             get.addColumn(PERSONAL_CF, NAME_COLUMN);
25             get.addColumn(PROFESSIONAL_CF, FIELD_COLUMN);
26             Result result = table.get(get);
27             byte[] nameValue = result.getValue(PERSONAL_CF, NAME_COLUMN);
28             System.out.println("Name: " + Bytes.toString(nameValue));
29             byte[] fieldValue = result.getValue(PROFESSIONAL_CF, FIELD_COLUMN);
30             System.out.println("Field: " + Bytes.toString(fieldValue));
31             List<Get> getList = new ArrayList<Get>();
32             Get get1 = new Get(Bytes.toBytes("2"));
33             get1.addColumn(PERSONAL_CF, NAME_COLUMN);
34             Get get2 = new Get(Bytes.toBytes("3"));
35             get1.addColumn(PERSONAL_CF, NAME_COLUMN);
36             getList.add(get1);
37             getList.add(get2);
38             Result[] results = table.get(getList);
39             for (Result res : results) {
40                 nameValue = res.getValue(PERSONAL_CF, NAME_COLUMN);
41                 System.out.println("Name: " + Bytes.toString(nameValue));
42             }
43         }
44         finally {
45             connection.close();
46             if (table != null) {
47                 table.close();
48             }
49         }
50     }
}
```

Output:

```
Problems Javadoc Declaration Console
<terminated> GettingStarted [Java Application] /usr/lib/jvm/java-11-openjdk-amd64/bin/java (15-May-2
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.hbase.util.UnsafeAvailChecker (file:/home/piyushpp/.m2/repo
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.hbase.util.UnsafeAvailChecker
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Name: Mike Jon
Field: Construction
Name: Dolcy
Name: David
```


Deleting Data from Table

Code to Delete Data from Table:

```
hbase/pom.xml *GettingStarted.java
1 package hbase;
2
3 import java.io.*;
10
11 public class GettingStarted {
12
13     private static byte[] PERSONAL_CF = Bytes.toBytes("personal");
14     private static byte[] PROFESSIONAL_CF = Bytes.toBytes("professional");
15     private static byte[] GENDER_COLUMN = Bytes.toBytes("gender");
16     private static byte[] FIELD_COLUMN = Bytes.toBytes("field");
17
18     public static void main(String[] args) throws IOException {
19         Configuration conf = HBaseConfiguration.create();
20         Connection connection = ConnectionFactory.createConnection(conf);
21         Table table = null;
22         try {
23             table = connection.getTable(TableName.valueOf("census"));
24
25             Delete delete = new Delete(Bytes.toBytes("1"));
26
27             delete.addColumn(PERSONAL_CF, GENDER_COLUMN);
28             delete.addColumn(PROFESSIONAL_CF, FIELD_COLUMN);
29
30             table.delete(delete);
31             System.out.println("Deletion Complete!");
32         }
33         finally {
34             connection.close();
35             if (table != null) {
36                 table.close();
37             }
38         }
39     }
40 }
41
```

Output:

```
2.7.1 :008 > scan 'census'
ROW
1 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Male
1 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Married
1 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Mike Jon
1 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
1 column=professional:field, timestamp=2021-05-15T19:04:42.214, value=Construction
2 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Female
2 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Dolcy
2 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
3 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Male
3 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Unmarried
3 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=David
3 row(s)
Took 0.0316 seconds
2.7.1 :009 > scan 'census'
ROW
1 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Married
1 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Mike Jon
1 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
2 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Female
2 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=Dolcy
2 column=professional:employed, timestamp=2021-05-15T19:04:42.214, value=Yes
3 column=personal:gender, timestamp=2021-05-15T19:04:42.214, value=Male
3 column=personal:marital_status, timestamp=2021-05-15T19:04:42.214, value=Unmarried
3 column=personal:name, timestamp=2021-05-15T19:04:42.214, value=David
3 row(s)
```

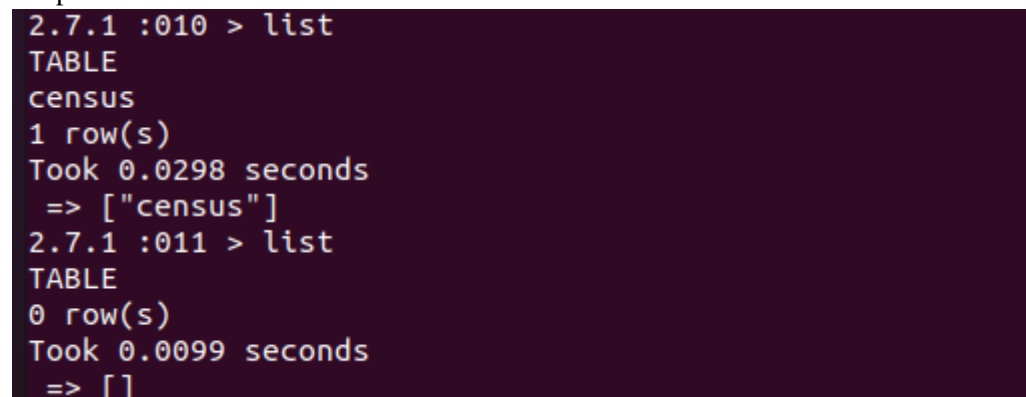
Deleting Table

Code to delete table:



```
1 package hbase;
2
3 import java.io.*;
4
5
6
7
8
9
10
11 public class GettingStarted {
12
13     public static void main(String[] args) throws IOException {
14         Configuration conf = HBaseConfiguration.create();
15         Connection connection = ConnectionFactory.createConnection(conf);
16         try {
17             TableName tableName = TableName.valueOf("census");
18             Admin admin = connection.getAdmin();
19
20             if(admin.tableExists(tableName)) {
21                 System.out.println("Deleting Table ...");
22                 admin.disableTable(tableName);
23                 admin.deleteTable(tableName);
24                 System.out.println("Done!");
25             }
26             else {
27                 System.out.println("Table does not exists.");
28             }
29         }
30         finally {
31             connection.close();
32         }
33     }
34 }
35
```

Output:



```
2.7.1 :010 > list
TABLE
census
1 row(s)
Took 0.0298 seconds
=> ["census"]
2.7.1 :011 > list
TABLE
0 row(s)
Took 0.0099 seconds
=> []
```

Filter rows based on condition

```
hbase/pom.xml *GettingStarted.java
1 package hbase;
2
3 import java.io.*;
4
18
19 public class GettingStarted {
20
21 private static void printResults(ResultScanner scanResult) {
22     System.out.println();
23     System.out.println("Results: ");
24
25     for(Result res : scanResult) {
26         for(Cell cell : res.listCells()) {
27             String row = new String(CellUtil.cloneRow(cell));
28             String family = new String(CellUtil.cloneFamily(cell));
29             String column = new String(CellUtil.cloneQualifier(cell));
30             String value = new String(CellUtil.cloneValue(cell));
31
32             System.out.println(row + " " + family + " " + column + " " + value);
33         }
34     }
35 }
36
37 public static void main(String[] args) throws IOException {
38     Configuration conf = HBaseConfiguration.create();
39     Connection connection = ConnectionFactory.createConnection(conf);
40     Table table = null;
41     ResultScanner scanResult = null;
42     try {
43         table = connection.getTable(TableName.valueOf("census"));
44
45         SingleColumnValueFilter filter = new SingleColumnValueFilter(
46             Bytes.toBytes("personal"), Bytes.toBytes("gender"), CompareFilter.CompareOp.EQUAL, new BinaryComparator(Bytes.toBytes("f"));
47         filter.setFilterIfMissing(true);
48         Scan userScan = new Scan();
49         userScan.setFilter(filter);
50
51         scanResult = table.getScanner(userScan);
52         printResults(scanResult);
53         filter = new SingleColumnValueFilter(Bytes.toBytes("personal"),
54             Bytes.toBytes("name"), CompareFilter.CompareOp.EQUAL, new SubstringComparator("Jones"));
55         userScan.setFilter(filter);
56         scanResult = table.getScanner(userScan);
57         printResults(scanResult);
58     }
59     finally {
60         connection.close();
61         if (table != null) {
62             table.close();
63         }
64         if (scanResult != null) {
65             scanResult.close();
66         }
67     }
68 }
69 }
```

Output:

```
Problems @ Javadoc Declaration Console
<terminated> GettingStarted [Java Application] /usr/lib/jvm/java-11-openjdk-amd64/bin/java (16-May-2021, 1:57)
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release

Results:
1 personal gender male
1 personal marital_status Unmarried
1 personal name Mike Jones
1 professional education_level high school
1 professional employed yes
1 professional field construction
2 personal gender male
2 personal marital_status divorced
2 personal name Ben

Results:
1 personal gender male
1 personal marital_status Unmarried
1 personal name Mike Jones
1 professional education_level high school
1 professional employed yes
1 professional field construction
```

MapReduce with Hbase

MapReduce is a process which is designed to solve the problem of processing in excess of terabytes of data in a scalable way. MapReduce follows a divide-and-conquer approach by splitting the data located on a distributed filesystem so that the servers available can access these chunks of data and process them as fast as they can.

Classes that are involved in the Map reduce:

- **InputFormat:** First it splits the input data, and then it returns a RecordReader instance that defines the classes of the key and value objects, and provides a next() method that is used to iterate over each input record.
- **Mapper:** In this step, each record read using the RecordReader is processed using the map() method.
- **Reducer:** The Reducer stage and class hierarchy is very similar to the Mapper stage. This time we get the output of a Mapper class and process it after the data has been shuffled and sorted.
- **OutputFormat:** Its job is to persist the data in various locations. There are specific implementations that allow output to files, or to HBase tables in the case of the TableOutputFormat class.
- **TableMapReduceUtil:** The MapReduce support comes with the TableMapReduceUtil class that helps in setting up MapReduce jobs over HBase.

Mapper Code

```
hbase/pom.xml  GettingStarted.java  MaritalStatusMapper.java  MaritalStatusReducer.java
1 package mapreduce;
2
3 import java.io.IOException;
4
5 import org.apache.hadoop.hbase.client.Result;
6 import org.apache.hadoop.hbase.io.ImmutableBytesWritable;
7
8 import org.apache.hadoop.io.IntWritable;
9 import org.apache.hadoop.hbase.mapreduce.TableMapper;
10
11 public class MaritalStatusMapper extends TableMapper<ImmutableBytesWritable, IntWritable>{
12
13     public static final byte[] CF = "personal".getBytes();
14     public static final byte[] MARITAL_STATUS = "marital_status".getBytes();
15
16     private final IntWritable ONE = new IntWritable(1);
17     private ImmutableBytesWritable key = new ImmutableBytesWritable();
18
19     public void map(ImmutableBytesWritable row, Result value, Context context) throws IOException, InterruptedException{
20         key.set(value.getValue(CF, MARITAL_STATUS));
21         context.write(key, ONE);
22     }
23 }
24
25
```

Reducer Code

```
hbase/pom.xml GettingStarted.java MaritalStatusMapper.java MaritalStatusReducer.java Main.java
1 package mapreduce;
2
3 import java.io.IOException;
4
14 public class MaritalStatusReducer extends TableReducer<ImmutableBytesWritable, IntWritable, ImmutableBytesWritable>{
15
16     public static final byte[] CF = "marital_status".getBytes();
17     public static final byte[] COUNT = "count".getBytes();
18
19     public void reduce(ImmutableBytesWritable key, Iterable<IntWritable> value, Context context) throws IOException, InterruptedException{
20         Integer sum = 0;
21         for(IntWritable val : value) {
22             sum += val.get();
23         }
24         Put put = new Put(key.get());
25         put.addColumn(CF, COUNT, Bytes.toBytes(sum.toString()));
26         context.write(key, put);
27     }
28 }
29
30 }
31
```

Driver Code

```
hbase/pom.xml GettingStarted.java MaritalStatusMapper.java MaritalStatusReducer.java Main.java
1 package mapreduce;
2
3 import org.apache.hadoop.hbase.HBaseConfiguration;
4
10
11 public class Main {
12     public static void main(String[] args) throws Exception{
13         Configuration conf = HBaseConfiguration.create();
14         Job job = Job.getInstance(conf);
15         String sourceTable = "census";
16         String targetTable = "summary";
17
18         Scan scan = new Scan();
19         scan.setCaching(500);
20         scan.setCacheBlocks(false);
21
22         TableMapReduceUtil.addDependencyJars(job);
23         TableMapReduceUtil.initTableMapperJob(sourceTable, scan,
24             MaritalStatusMapper.class, ImmutableBytesWritable.class, IntWritable.class, job);
25         TableMapReduceUtil.initTableReducerJob(targetTable, MaritalStatusReducer.class, job);
26         job.setNumReduceTasks(1);
27         job.waitForCompletion(true);
28     }
29 }
30
```

Output:

```
2.7.1 :011 > scan 'summary'
ROW                                COLUMN+CELL
divorced                          column=marital_status:count, timestamp=2021-05-16T14:11:47.365, value=2
married                           column=marital_status:count, timestamp=2021-05-16T14:12:20.344, value=1
unmarried                         column=marital_status:count, timestamp=2021-05-16T14:12:42.705, value=1
3 row(s)
```