Dig Data Management Assignment 3

信科四 郑元嘉 1800920541

Task1 - Hive

1. Create a table first.

   ```
   $ create table titanic(PassengerId int, Survived int, Pclass int, Name string,
   Sex string, Age float, SibSp int, Parch int, Ticket string, Fare float, Cabin
   string, Embarked string) ROW FORMAT SERDE
   'org.apache.hadoop.hive.serde2.OpenCSVSerde' location
   '/user/hive/warehouse/mydb.db/titanic';
   ```

2. Check out the creation.

   ```
   > DESCRIBE titanic;
   ```


3. Load csv file `TitanicData.csv` to Hive.

   ```
   $ load data local inpath '/media/jack/File/hadoopHw/hw3/Titanic/TitanicData.csv'
   overwrite into table titanic;
   ```

4. Check out the loading.

   ```
   > SELECT * FROM titanic
   ```

```
| 874                      | 0                 | 3             | Vander Cruyssen, Mr. Victor                              | male   | 47
   | 0               | 0         | 345765        | 9              | S                 |
| 875                      | 1                 | 2             | Abelson, Mrs. Samuel (Hannah Wizosky)                   | female | 28
   | 1               | 0         | P/PP 3381     | 24             | C                 |
| 876                      | 1                 | 3             | Najib, Miss. Adele Kiamie "Jane"                        | female | 15
   | 0               | 0         | 2667          | 7.225          | C                 |
| 877                      | 0                 | 3             | Gustafsson, Mr. Alfred Ossian                           | male   | 20
   | 0               | 0         | 7534          | 9.8458         | S                 |
| 878                      | 0                 | 3             | Petroff, Mr. Nedelio                                    | male   | 19
   | 0               | 0         | 349212        | 7.8958         | S                 |
| 879                      | 0                 | 3             | Laleff, Mr. Kristo                                      | male   |
   | 0               | 0         | 349217        | 7.8958         | S                 |
| 880                      | 1                 | 1             | Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)          | female | 56
   | 0               | 1         | 11767         | 83.1583    | C50    | C                 |
| 881                      | 1                 | 2             | Shelley, Mrs. William (Imanita Parrish Hall)           | female | 25
   | 0               | 1         | 230433        | 26             | S                 |
| 882                      | 0                 | 3             | Markun, Mr. Johann                                      | male   | 33
   | 0               | 0         | 349257        | 7.8958         | S                 |
| 883                      | 0                 | 3             | Dahlberg, Miss. Gerda Ulrika                            | female | 22
   | 0               | 0         | 7552          | 10.5167        | S                 |
| 884                      | 0                 | 2             | Banfield, Mr. Frederick James                          | male   | 28
   | 0               | 0         | C.A./SOTON 34068 | 10.5        | S                 |
| 885                      | 0                 | 3             | Sutehall, Mr. Henry Jr                                 | male   | 25
   | 0               | 0         | SOTON/OQ 392076 | 7.05        | S                 |
| 886                      | 0                 | 3             | Rice, Mrs. William (Margaret Norton)                   | female | 39
   | 0               | 5         | 382652        | 29.125         | Q                 |
| 887                      | 0                 | 2             | Montvila, Rev. Juozas                                  | male   | 27
   | 0               | 0         | 211536        | 13             | S                 |
| 888                      | 1                 | 1             | Graham, Miss. Margaret Edith                           | female | 19
   | 0               | 0         | 112053        | 30         | B42    | S                 |
| 889                      | 0                 | 3             | Johnston, Miss. Catherine Helen "Carrie"               | female |
   | 1               | 2         | W./C. 6607    | 23.45          | S                 |
| 890                      | 1                 | 1             | Behr, Mr. Karl Howell                                  | male   | 26
   | 0               | 0         | 111369        | 30         | C148   | C                 |
| 891                      | 0                 | 3             | Dooley, Mr. Patrick                                    | male   | 32
   | 0               | 0         | 370376        | 7.75           | Q                 |
+--------------------------+-------------------+---------------+--------------------------------------------------------+--------+--------
------+-----------------+---------+---------------+----------------+-------------------+

891 rows selected (0.179 seconds)
```

5. Query I: Find all data where the gender is female and she survived.

```
> SELECT * FROM titanic WHERE sex == 'female' AND survived == 1;
```

```
| titanic.passengerid | titanic.survived | titanic.pclass |                        titanic.name                        | titanic.sex  | tita
nic.age | titanic.sibsp | titanic.parch |    titanic.ticket   | titanic.fare |  titanic.cabin  | titanic.embarked |
+---------------------+------------------+----------------+------------------------------------------------------------+--------------+-----
---------+---------------+---------------+---------------------+--------------+-----------------+------------------+
| 2                   | 1                | 1              | Cumings, Mrs. John Bradley (Florence Briggs Thayer) | female      | 38
     | 1             | 0             | PC 17599            | 71.2833  | C85             | C                |
| 3                   | 1                | 3              | Heikkinen, Miss. Laina                              | female      | 26
     | 0             | 0             | STON/O2. 3101282    | 7.925    |                 | S                |
| 4                   | 1                | 1              | Futrelle, Mrs. Jacques Heath (Lily May Peel)        | female      | 35
     | 1             | 0             | 113803              | 53.1     | C123            | S                |
| 9                   | 1                | 3              | Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)   | female      | 27
     | 0             | 2             | 347742              | 11.1333  |                 | S                |
| 10                  | 1                | 2              | Nasser, Mrs. Nicholas (Adele Achem)                 | female      | 14
     | 1             | 0             | 237736              | 30.0708  |                 | C                |
| 11                  | 1                | 3              | Sandstrom, Miss. Marguerite Rut                     | female      | 4
     | 1             | 1             | PP 9549             | 16.7     | G6              | S                |
| 12                  | 1                | 1              | Bonnell, Miss. Elizabeth                            | female      | 58
     | 0             | 0             | 113783              | 26.55    | C103            | S                |
| 16                  | 1                | 2              | Hewlett, Mrs. (Mary D Kingcome)                     | female      | 55
     | 0             | 0             | 248706              | 16       |                 | S                |
| 20                  | 1                | 3              | Masselmani, Mrs. Fatima                             | female      |
     | 0             | 0             | 2649                | 7.225    |                 | C                |
| 23                  | 1                | 3              | McGowan, Miss. Anna "Annie"                         | female      | 15
     | 0             | 0             | 330923              | 8.0292   |                 | Q                |
| 26                  | 1                | 3              | Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson) | female
   | 38                | 1           | 5             | 347077              | 31.3875  |                 | S                |
| 29                  | 1                | 3              | O'Dwyer, Miss. Ellen "Nellie"                       | female      |
     | 0             | 0             | 330959              | 7.8792   |                 | Q                |
| 32                  | 1                | 1              | Spencer, Mrs. William Augustus (Marie Eugenie)      | female      |
     | 1             | 0             | PC 17569            | 146.5208 | B78             | C                |
| 33                  | 1                | 3              | Glynn, Miss. Mary Agatha                            | female      |
```

6. Query II: List the rearranged age labels and names.

```
1  > SELECT name,
2  > CASE WHEN age < 20 THEN 'young'
3  > WHEN 20 <= age AND age < 40 THEN 'adult'
4  > WHEN 40 <= age THEN 'senior'
5  > ELSE 'none'
6  > END AS age
7  > FROM titanic;
```

```
+----------------------------------------------------+---------+
|                       name                         |   age   |
+----------------------------------------------------+---------+
| Braund, Mr. Owen Harris                            | adult   |
| Cumings, Mrs. John Bradley (Florence Briggs Thayer)| adult   |
| Heikkinen, Miss. Laina                             | adult   |
| Futrelle, Mrs. Jacques Heath (Lily May Peel)       | adult   |
| Allen, Mr. William Henry                           | adult   |
| Moran, Mr. James                                   | none    |
| McCarthy, Mr. Timothy J                            | senior  |
| Palsson, Master. Gosta Leonard                     | young   |
| Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  | adult   |
| Nasser, Mrs. Nicholas (Adele Achem)                | young   |
| Sandstrom, Miss. Marguerite Rut                    | young   |
| Bonnell, Miss. Elizabeth                           | senior  |
| Saundercock, Mr. William Henry                     | adult   |
| Andersson, Mr. Anders Johan                        | adult   |
| Vestrom, Miss. Hulda Amanda Adolfina               | young   |
| Hewlett, Mrs. (Mary D Kingcome)                    | senior  |
| Rice, Master. Eugene                               | young   |
| Williams, Mr. Charles Eugene                       | none    |
| Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)| adult|
| Masselmani, Mrs. Fatima                            | none    |
| Fynney, Mr. Joseph J                               | adult   |
| Beesley, Mr. Lawrence                              | adult   |
| McGowan, Miss. Anna "Annie"                        | young   |
| Sloper, Mr. William Thompson                       | adult   |
| Palsson, Miss. Torborg Danira                      | young   |
| Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)| adult|
| Emir, Mr. Farred Chehab                            | none    |
| Fortune, Mr. Charles Alexander                     | young   |
| O'Dwyer, Miss. Ellen "Nellie"                      | none    |
| Todoroff, Mr. Lalio                                | none    |
| Uruchurtu, Don. Manuel E                           | senior  |
| Spencer, Mrs. William Augustus (Marie Eugenie)     | none    |
| Glynn, Miss. Mary Agatha                           | none    |
| Wheadon, Mr. Edward H                              | senior  |
| Meyer, Mr. Edgar Joseph                            | adult   |
| Holverson, Mr. Alexander Oskar                     | senior  |
| Mamee, Mr. Hanna                                   | none    |
```

Task2 - HBase

1. Create table named 'titanic' first.

   ```
   > create 'titanic', 'survived', 'pclass', 'name', 'sex', 'age', 'sibSp', 'parch',
   'ticket', 'fare', 'cabin', 'embarked'
   ```

   Note that we don't pass the attribute 'passengerId', for it will ba used as 'HBASE_ROW_KEY'
   later on.

2. Check out the creation.

```
hbase(main):010:0> list
TABLE
test
titanic
2 row(s) in 0.0160 seconds

=> ["test", "titanic"]
```

3. Load csv file `TitanicData.csv` to HBase.

   `$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.columns=HBASE_ROW_KEY,survived,pclass,name,sex,age,sibSp,parch,ticket,fare,cabin,embarked '-Dimporttsv.separator=,' titanic TitanicData.csv`

   Note that the first column name must be 'HBASE_ROW_KEY' which indicates the indexes (ids) for the row data.

4. Error: caused by wrong column names

   I was confronted by the error due to not completely identical columns the table holds and my loading command holds. The uppercase/lowercase requires the same as well.

```
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.doBatchOp(RSRpcServices.java:925)
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.doNonAtomicRegionMutation(RSRpcServices.java:844)
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.multi(RSRpcServices.java:2405)
        at org.apache.hadoop.hbase.protobuf.generated.ClientProtos$ClientService$2.callBlockingMethod(ClientProtos.
java:36621)
        at org.apache.hadoop.hbase.ipc.RpcServer.call(RpcServer.java:2380)
        at org.apache.hadoop.hbase.ipc.CallRunner.run(CallRunner.java:124)
        at org.apache.hadoop.hbase.ipc.RpcExecutor$Handler.run(RpcExecutor.java:297)
        at org.apache.hadoop.hbase.ipc.RpcExecutor$Handler.run(RpcExecutor.java:277)
: 891 times, servers with issues: jack-sy-series,16201,1556759654613
        at org.apache.hadoop.hbase.client.AsyncProcess$BatchErrors.makeException(AsyncProcess.java:297)
        at org.apache.hadoop.hbase.client.AsyncProcess$BatchErrors.access$2300(AsyncProcess.java:273)
        at org.apache.hadoop.hbase.client.AsyncProcess.waitForAllPreviousOpsAndReset(AsyncProcess.java:1906)
        at org.apache.hadoop.hbase.client.BufferedMutatorImpl.backgroundFlushCommits(BufferedMutatorImpl.java:250)
        at org.apache.hadoop.hbase.client.BufferedMutatorImpl.close(BufferedMutatorImpl.java:186)
        at org.apache.hadoop.hbase.mapreduce.TableOutputFormat$TableRecordWriter.close(TableOutputFormat.java:124)
        at org.apache.hadoop.mapred.MapTask$NewDirectOutputCollector.close(MapTask.java:670)
        at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:793)
        at org.apache.hadoop.mapred.MapTask.run(MapTask.java:341)
        at org.apache.hadoop.mapred.LocalJobRunner$Job$MapTaskRunnable.run(LocalJobRunner.java:243)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:748)
2019-05-02 09:57:39,848 INFO  [main] mapreduce.Job: Job job_local1423913549_0001 running in uber mode : false
2019-05-02 09:57:39,850 INFO  [main] mapreduce.Job:  map 0% reduce 0%
2019-05-02 09:57:39,852 INFO  [main] mapreduce.Job: Job job_local1423913549_0001 failed with state FAILED due to: N
A
2019-05-02 09:57:39,873 INFO  [main] mapreduce.Job: Counters: 9
        Map-Reduce Framework
                Map input records=891
                Map output records=891
                Input split bytes=112
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
        ImportTsv
                Bad Lines=0
        File Input Format Counters
                Bytes Read=61113
        File Output Format Counters
                Bytes Written=0
```

```
        at org.apache.hadoop.mapred.LocalJobRunner$Job.runTasks(LocalJobRunner.java:462)
        at org.apache.hadoop.mapred.LocalJobRunner$Job.run(LocalJobRunner.java:522)
Caused by: org.apache.hadoop.hbase.client.RetriesExhaustedWithDetailsException: Failed 891 actions: org.apache.hado
op.hbase.regionserver.NoSuchColumnFamilyException: Column family Age does not exist in region titanic,,155676178090
1.be9cf8b551f7a2736b8e94a6bcecd5ac. in table 'titanic', {NAME => 'age', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_M
EMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'N
ONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'cabin',
 BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING =>
 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536',
REPLICATION_SCOPE => '0'}, {NAME => 'embarked', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_D
ELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0
', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'fare', BLOOMFILTER => 'ROW', VE
RSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER'
, COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
, {NAME => 'name', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA
_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOC
KSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'parch', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY =>
'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN
_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'passengerId', BL
OOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'N
ONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REP
LICATION_SCOPE => '0'}, {NAME => 'pclass', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETE
D_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BL
OCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'sex', BLOOMFILTER => 'ROW', VERSIONS
 => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMP
RESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAM
E => 'sibSp', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOC
K_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE
 => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'survived', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'f
alse', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_V
ERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}, {NAME => 'ticket', BLOOMFILT
ER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', T
TL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATIO
N_SCOPE => '0'}
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.doBatchOp(RSRpcServices.java:925)
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.doNonAtomicRegionMutation(RSRpcServices.java:844)
        at org.apache.hadoop.hbase.regionserver.RSRpcServices.multi(RSRpcServices.java:2405)
        at org.apache.hadoop.hbase.protobuf.generated.ClientProtos$ClientService$2.callBlockingMethod(ClientProtos.
```

5. Screenshot for success in loading task.

```
2019-05-02 10:08:53,884 INFO  [LocalJobRunner Map Task Executor #0] zookeeper.ZooKeeper: Session: 0x16a724046180027
 closed
2019-05-02 10:08:53,884 INFO  [LocalJobRunner Map Task Executor #0-EventThread] zookeeper.ClientCnxn: EventThread s
hut down for session: 0x16a724046180027
2019-05-02 10:08:53,979 INFO  [LocalJobRunner Map Task Executor #0] mapred.LocalJobRunner:
2019-05-02 10:08:54,101 INFO  [LocalJobRunner Map Task Executor #0] client.ConnectionManager$HConnectionImplementat
ion: Closing zookeeper sessionid=0x16a724046180026
2019-05-02 10:08:54,102 INFO  [LocalJobRunner Map Task Executor #0] zookeeper.ZooKeeper: Session: 0x16a724046180026
 closed
2019-05-02 10:08:54,102 INFO  [LocalJobRunner Map Task Executor #0-EventThread] zookeeper.ClientCnxn: EventThread s
hut down for session: 0x16a724046180026
2019-05-02 10:08:54,102 INFO  [LocalJobRunner Map Task Executor #0] mapred.Task: Task:attempt_local444967832_0001_m
_000000_0 is done. And is in the process of committing
2019-05-02 10:08:54,107 INFO  [LocalJobRunner Map Task Executor #0] mapred.LocalJobRunner: map
2019-05-02 10:08:54,107 INFO  [LocalJobRunner Map Task Executor #0] mapred.Task: Task 'attempt_local444967832_0001_
m_000000_0' done.
2019-05-02 10:08:54,107 INFO  [LocalJobRunner Map Task Executor #0] mapred.LocalJobRunner: Finishing task: attempt_
local444967832_0001_m_000000_0
2019-05-02 10:08:54,107 INFO  [Thread-29] mapred.LocalJobRunner: map task executor complete.
2019-05-02 10:08:54,711 INFO  [main] mapreduce.Job: Job job_local444967832_0001 running in uber mode : false
2019-05-02 10:08:54,713 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2019-05-02 10:08:54,716 INFO  [main] mapreduce.Job: Job job_local444967832_0001 completed successfully
2019-05-02 10:08:54,725 INFO  [main] mapreduce.Job: Counters: 21
        File System Counters
                FILE: Number of bytes read=30063172
                FILE: Number of bytes written=30670617
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=61113
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=3
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Map-Reduce Framework
                Map input records=891
                Map output records=891
                Input split bytes=112
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=4
                Total committed heap usage (bytes)=251396096
        ImportTsv
                Bad Lines=0
        File Input Format Counters
                Bytes Read=61113
        File Output Format Counters
                Bytes Written=0
```

6. Check out the table 'titanic.'

```
> count 'titanic'
```

```
hbase(main):001:0> count 'titanic'
891 row(s) in 0.4100 seconds

=> 891
```

7. Query: Find the survials.

```
scan 'titanic', {COLUMN => 'survived', FILTER => "ValueFilter(=, 'binary:1')"}
```

```
hbase(main):102:0> scan 'titanic', {COLUMN => 'survived', FILTER => "ValueFilter(=, 'binary:1')"}
ROW                         COLUMN+CELL
 10                          column=survived:, timestamp=1556766665510, value=1
 107                         column=survived:, timestamp=1556766665510, value=1
 108                         column=survived:, timestamp=1556766665510, value=1
 11                          column=survived:, timestamp=1556766665510, value=1
 110                         column=survived:, timestamp=1556766665510, value=1
 12                          column=survived:, timestamp=1556766665510, value=1
 124                         column=survived:, timestamp=1556766665510, value=1
 126                         column=survived:, timestamp=1556766665510, value=1
 128                         column=survived:, timestamp=1556766665510, value=1
 129                         column=survived:, timestamp=1556766665510, value=1
 134                         column=survived:, timestamp=1556766665510, value=1
 137                         column=survived:, timestamp=1556766665510, value=1
 142                         column=survived:, timestamp=1556766665510, value=1
 143                         column=survived:, timestamp=1556766665510, value=1
 147                         column=survived:, timestamp=1556766665510, value=1
 152                         column=survived:, timestamp=1556766665510, value=1
 157                         column=survived:, timestamp=1556766665510, value=1
 16                          column=survived:, timestamp=1556766665510, value=1
 162                         column=survived:, timestamp=1556766665510, value=1
 166                         column=survived:, timestamp=1556766665510, value=1
 167                         column=survived:, timestamp=1556766665510, value=1
 173                         column=survived:, timestamp=1556766665510, value=1
 18                          column=survived:, timestamp=1556766665510, value=1
 184                         column=survived:, timestamp=1556766665510, value=1
 185                         column=survived:, timestamp=1556766665510, value=1
 187                         column=survived:, timestamp=1556766665510, value=1
 188                         column=survived:, timestamp=1556766665510, value=1
 191                         column=survived:, timestamp=1556766665510, value=1
 193                         column=survived:, timestamp=1556766665510, value=1
 194                         column=survived:, timestamp=1556766665510, value=1
 195                         column=survived:, timestamp=1556766665510, value=1
 196                         column=survived:, timestamp=1556766665510, value=1
 199                         column=survived:, timestamp=1556766665510, value=1
 2                           column=survived:, timestamp=1556766665510, value=1
 20                          column=survived:, timestamp=1556766665510, value=1
 205                         column=survived:, timestamp=1556766665510, value=1
```

8. (Extra) Drop the table.

If we attempt to drop, we need to 'disable' the table fisrt then drop it.

```
> disable 'titanic'
```

```
> drop 'titanic'
```

Task3 - PIG

1. We don't need to create one table or load the csv file from local (if it has already been loaded to HDFS). The only thing to extract the csv data is to use `LOAD` command and corresponding package `org.apache.pig.piggybank.storage.CSVExcelStorage()` in PIG shell.

```
> records = LOAD 'TitanicData.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage() AS (passengerIp:int,
```

```
survived:int, pclass:chararray, name:chararray, sex:chararray, age:float, sibSp:
chararray, parch:chararray, ticket: float, fare:float, cabin: chararray,
embarked: chararray);
```

2. Check out the loading by dumping `records`.

```
> dump records;
```

```
(851,0,3,Andersson, Master. Sigvard Harald Elias,male,4.0,4,2,347082.0,31.275,,S)
(852,0,3,Svensson, Mr. Johan,male,74.0,0,0,347060.0,7.775,,S)
(853,0,3,Boulos, Miss. Nourelain,female,9.0,1,1,2678.0,15.2458,,C)
(854,1,1,Lines, Miss. Mary Conover,female,16.0,0,1,,39.4,D28,S)
(855,0,2,Carter, Mrs. Ernest Courtenay (Lilian Hughes),female,44.0,1,0,244252.0,26.0,,S)
(856,1,3,Aks, Mrs. Sam (Leah Rosen),female,18.0,0,1,392091.0,9.35,,S)
(857,1,1,Wick, Mrs. George Dennick (Mary Hitchcock),female,45.0,1,1,36928.0,164.8667,,S)
(858,1,1,Daly, Mr. Peter Denis ,male,51.0,0,0,113055.0,26.55,E17,S)
(859,1,3,Baclini, Mrs. Solomon (Latifa Qurban),female,24.0,0,3,2666.0,19.2583,,C)
(860,0,3,Razi, Mr. Raihed,male,,0,0,2629.0,7.2292,,C)
(861,0,3,Hansen, Mr. Claus Peter,male,41.0,2,0,350026.0,14.1083,,S)
(862,0,2,Giles, Mr. Frederick Edward,male,21.0,1,0,28134.0,11.5,,S)
(863,1,1,Swift, Mrs. Frederick Joel (Margaret Welles Barron),female,48.0,0,0,17466.0,25.9292,D17,S)
(864,0,3,Sage, Miss. Dorothy Edith "Dolly",female,,8,2,,69.55,,S)
(865,0,2,Gill, Mr. John William,male,24.0,0,0,233866.0,13.0,,S)
(866,1,2,Bystrom, Mrs. (Karolina),female,42.0,0,0,236852.0,13.0,,S)
(867,1,2,Duran y More, Miss. Asuncion,female,27.0,1,0,,13.8583,,C)
(868,0,1,Roebling, Mr. Washington Augustus II,male,31.0,0,0,,50.4958,A24,S)
(869,0,3,van Melkebeke, Mr. Philemon,male,,0,0,345777.0,9.5,,S)
(870,1,3,Johnson, Master. Harold Theodor,male,4.0,1,1,347742.0,11.1333,,S)
(871,0,3,Balkic, Mr. Cerin,male,26.0,0,0,349248.0,7.8958,,S)
(872,1,1,Beckwith, Mrs. Richard Leonard (Sallie Monypeny),female,47.0,1,1,11751.0,52.5542,D35,S)
(873,0,1,Carlsson, Mr. Frans Olof,male,33.0,0,0,695.0,5.0,B51 B53 B55,S)
(874,0,3,Vander Cruyssen, Mr. Victor,male,47.0,0,0,345765.0,9.0,,S)
(875,1,2,Abelson, Mrs. Samuel (Hannah Wizosky),female,28.0,1,0,,24.0,,C)
(876,1,3,Najib, Miss. Adele Kiamie "Jane",female,15.0,0,0,2667.0,7.225,,C)
(877,0,3,Gustafsson, Mr. Alfred Ossian,male,20.0,0,0,7534.0,9.8458,,S)
(878,0,3,Petroff, Mr. Nedelio,male,19.0,0,0,349212.0,7.8958,,S)
(879,0,3,Laleff, Mr. Kristo,male,,0,0,349217.0,7.8958,,S)
(880,1,1,Potter, Mrs. Thomas Jr (Lily Alexenia Wilson),female,56.0,0,1,11767.0,83.1583,C50,C)
(881,1,2,Shelley, Mrs. William (Imanita Parrish Hall),female,25.0,0,1,230433.0,26.0,,S)
(882,0,3,Markun, Mr. Johann,male,33.0,0,0,349257.0,7.8958,,S)
(883,0,3,Dahlberg, Miss. Gerda Ulrika,female,22.0,0,0,7552.0,10.5167,,S)
(884,0,2,Banfield, Mr. Frederick James,male,28.0,0,0,,10.5,,S)
(885,0,3,Sutehall, Mr. Henry Jr,male,25.0,0,0,,7.05,,S)
(886,0,3,Rice, Mrs. William (Margaret Norton),female,39.0,0,5,382652.0,29.125,,Q)
(887,0,2,Montvila, Rev. Juozas,male,27.0,0,0,211536.0,13.0,,S)
(888,1,1,Graham, Miss. Margaret Edith,female,19.0,0,0,112053.0,30.0,B42,S)
(889,0,3,Johnston, Miss. Catherine Helen "Carrie",female,,1,2,,23.45,,S)
(890,1,1,Behr, Mr. Karl Howell,male,26.0,0,0,111369.0,30.0,C148,C)
(891,0,3,Dooley, Mr. Patrick,male,32.0,0,0,370376.0,7.75,,Q)
```

3. Query: List the names and genders of those who are male survivors.

```
> male_survivor = FOREACH (FILTER records BY survived == 1 AND sex == 'male')
GENERATE name, sex;
```

```
dump male_survivor;
(McGough, Mr. James Robert,male)
(Beane, Mr. Edward,male)
(Padro y Manent, Mr. Julian,male)
(Davies, Master. John Morgan Jr,male)
(Thayer, Mr. John Borland Jr,male)
(Leeni, Mr. Fahim ("Philip Zenni"),male)
(Jonsson, Mr. Carl,male)
(Harris, Mr. George,male)
(Flynn, Mr. John Irwin ("Irving"),male)
(Jussila, Mr. Eiriik,male)
(Frolicher-Stehli, Mr. Maxmillian,male)
(Duff Gordon, Sir. Cosmo Edmund ("Mr Morgan"),male)
(Homer, Mr. Harry ("Mr E Haven"),male)
(Daniel, Mr. Robert Williams,male)
(Kimball, Mr. Edwin Nelson Jr,male)
(Nakid, Mr. Sahid,male)
(Barkworth, Mr. Algernon Henry Wilson,male)
(Stahelin-Maeglin, Dr. Max,male)
(Foo, Mr. Choong,male)
(Harper, Mr. Henry Sleeper,male)
(Simonius-Blumer, Col. Oberst Alfons,male)
(Frauenthal, Dr. Henry William,male)
(Lindqvist, Mr. Eino William,male)
(Wilhelms, Mr. Charles,male)
(Cardeza, Mr. Thomas Drake Martinez,male)
(Hassab, Mr. Hammad,male)
(Dick, Mr. Albert Adrian,male)
(Lam, Mr. Ali,male)
(Silverthorne, Mr. Spencer Victor,male)
(Calderhead, Mr. Edward Pennington,male)
(Moubarek, Master. Halim Gonios ("William George"),male)
(Taylor, Mr. Elmer Zebley,male)
(Chambers, Mr. Norman Campbell,male)
(Lesurer, Mr. Gustave J,male)
(Hawksford, Mr. Walter James,male)
(Stranden, Mr. Juho,male)
(Moor, Master. Meier,male)
(Hamalainen, Master. Viljo,male)
(Barah, Mr. Hanna Assi,male)
(Dean, Master. Bertram Vere,male)
(Carter, Master. William Thornton II,male)
(Thomas, Master. Assad Alexander,male)
(Hedman, Mr. Oskar Arvid,male)
(Lulic, Mr. Nikola,male)
(Mallet, Master. Andre,male)
(McCormack, Mr. Thomas Joseph,male)
(Richards, Master. George Sibley,male)
(Chip, Mr. Chang,male)
(Marechal, Mr. Pierre,male)
(Daly, Mr. Peter Denis ,male)
(Johnson, Master. Harold Theodor,male)
(Behr, Mr. Karl Howell,male)
```

References

Hive 导入CSV文件 - 小海的专栏 - CSDN博客

HBase shell 命令介绍 - 纯洁的微笑 - 博客园

Hbase 上传CSV文件 - mega-victor - CSDN博客