



YCBS-257 - Data at Scale

HBase Workshop - Keys

Exercise 1:

```
create 'students', 'data'
list
put 'students', 'id1', 'data:name', 'michel'
put 'students', 'id2', 'data:age', '45'
put 'students', 'id3', 'data:city', 'paris'
put 'students', 'id4', 'data:country', 'canada'

scan 'students'
put 'students', 'id4', 'data:name', 'jean'
put 'students', 'id4', 'data:age', '35'
put 'students', 'id4', 'data:class', 'data at scale'
put 'students', 'id4', 'data:phone', '123456789'
scan 'students'

put 'students', 'id5', 'data:name', 'khaled'

count 'students'

put 'students', 'id1', 'data:age', '56'
put 'students', 'id2', 'data:age', '36'

scan 'students', {COLUMNS => 'data:age'}
alter 'students', NAME => 'data', VERSIONS => 5
alter 'students', NAME => 'city', VERSIONS => 3

put 'students', 'id3', 'data:city', 'paris'
put 'students', 'id3', 'data:city', 'nyc'

scan 'students', VERSIONS => 5

put 'students', 'id3', 'data:city', 'roma'

scan 'students', VERSIONS => 5

put 'students', 'id3', 'data:city', 'london'
scan 'students', VERSIONS => 5
put 'students', 'id3', 'data:city', 'berlin'
scan 'students', VERSIONS => 5
put 'students', 'id3', 'data:city', 'montreal', 1538612416273
scan 'students', VERSIONS => 5

alter 'students', NAME => 'personal', VERSIONS => 3

describe 'students'
```



Exercise 2:

Pig & HBase

HBase

```
create 'stations','stations_infos'
list
count 'stations'
```

Pig

```
data = Load '/home/cloudera/Downloads/Bixi/Stations_2018.csv' using
PigStorage(',') as
(code:chararray,name:chararray,latitude:double,longitude:double);

stations = FILTER data BY (code != 'code');

STORE stations INTO 'hbase://stations' USING
org.apache.pig.backend.hadoop.hbase.HBaseStorage('stations_infos:code,
stations_infos:name, stations_infos:latitude,
stations_infos:longitude');
```

HBase

```
count 'stations'
```

Exercise 3:

Hive & HBase

Hive

```
create table Stations (
    code string,
    name string,
    latitude double,
    longitude double)
Row format delimited fields terminated by ',' stored as TextFile
TBLPROPERTIES ("skip.header.line.count"="1");

load data inpath '/hivelab1/Stations_2018.csv' into table stations;

CREATE TABLE stations_hbase (code STRING, name STRING, latitude
DOUBLE, longitude DOUBLE) STORED BY
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH
SERDEPROPERTIES ('hbase.columns.mapping' =
':key,data:name,data:latitude,data:longitude') TBLPROPERTIES
('hbase.table.name' = 'stations_hive');

show tables;

select count(*) from stations_hbase;

insert into stations_hbase select * from stations;
```



```
select count(*) from stations_hbase;
```

HBase

```
list
```

```
count 'stations_hive'
```