CHULALONGKORN
BUSINESS SCHOOL
FLAGSHIP FOR LIFE

# Apache HBase

APACHE
HBASE

Thaveewat Khanan

2603615 Organization Data Management

# Understanding HBase

## An open source, non-relational, distributed database

HBase is an open source, non-relational, distributed database modeled after Google's BigTable and is written in Java. It is developed as part of Apache Software Foundation's Apache Hadoop project and runs on top of HDFS providing BigTable-like capabilities for Hadoop. That is, it provides a fault-tolerant way of storing large quantities of sparse data.

# HBase Features

- Hadoop database modelled after Google's Bigtable
- Column oriented data store, known as Hadoop Database
- Support random realtime CRUD operations (unlike HDFS)
- No SQL Database
- Opensource, written in Java
- Run on a cluster of commodity hardware

Hive.apache.org

# When to use HBase ?

- When you need high volume data to be stored
- Un-structured data
- Sparse data
- Column-oriented data
- Versioned data (same data template, captured at various time, time-elapse data)
- When you need high scalability

Hive.apache.org

# Which one to use ?

**HDFS**

Only append dataset (no random write)

Read the whole dataset (no random read)

**HBase**

Need random write and/or read

Has thousands of operation per second on TB+ of data

**RDBMS**

Data fits on one big node

Need full transaction support

Need real-time query capabilities

# HBase Components

**Region**

Row of table are stores

**Region Server**

Hosts the tables

**Master**

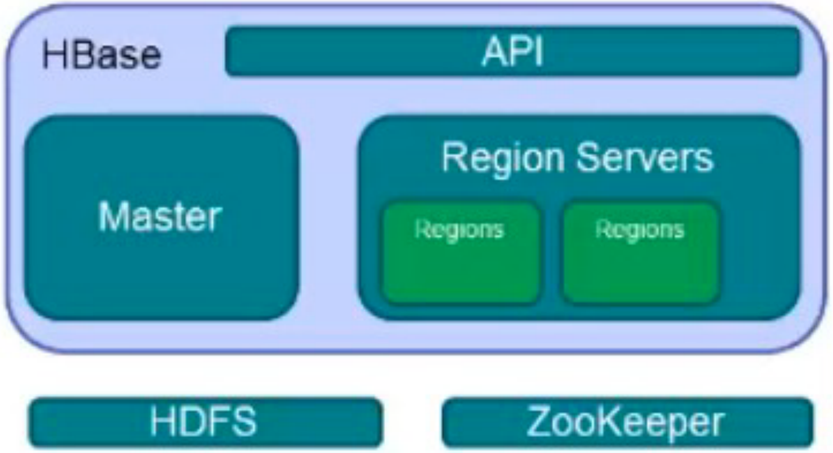
Coordinating the Region Servers

**ZooKeeper**

**HDFS**

**API**

The Java Client API



Hive.apache.org

# HBase Shell Commands

List all tables in hbase.

```
hbase> list
```

Create table; pass table name

```
hbase> create 'ns1:t1', {NAME => 'f1', VERSIONS => 5}
hbase> create 't1', {NAME => 'f1'}, {NAME => 'f2'}, {NAME => 'f3'}
```

Put a cell ‘value’ at specified table/row/column

```
hbase> put ‘t1’, ‘r1’, ‘c1’, ‘value’, ts1
```

Get row or cell contents

```
hbase> get ‘t1’, ‘r1’
```

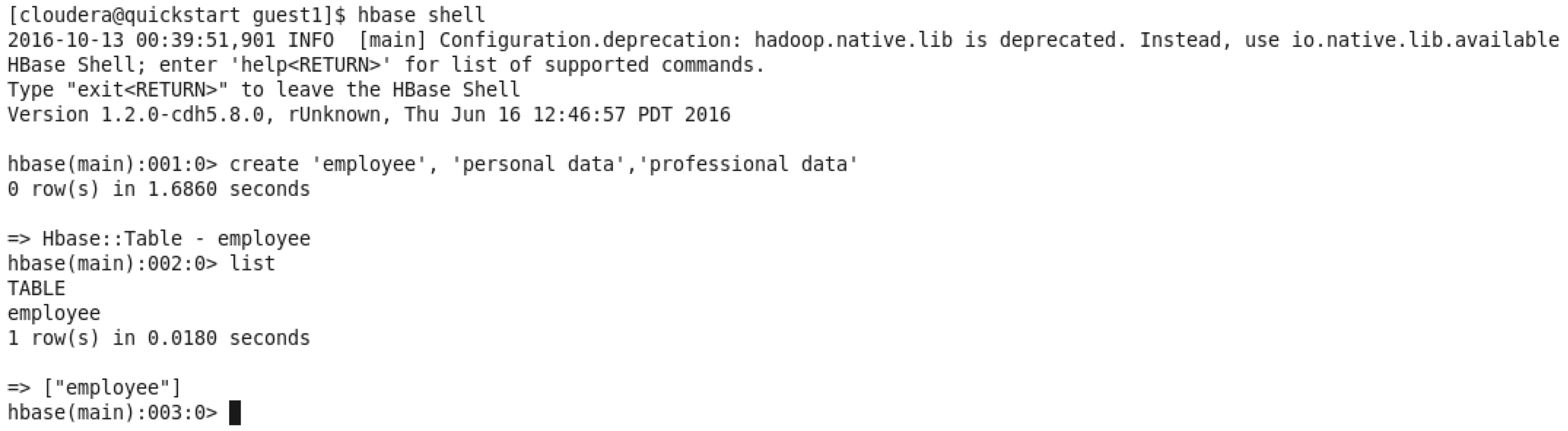
https://learnhbase.wordpress.com/2013/03/02/hbase-shell-commands/

# Running HBase

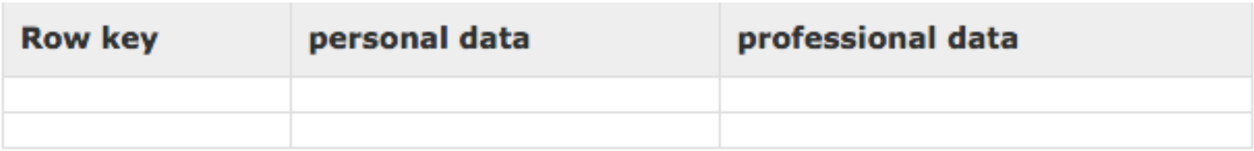
**$ hbase shell**

**hbase(main):001:0> create 'employee', 'personal data', 'professional data'**

**hbase(main):002:0> list**

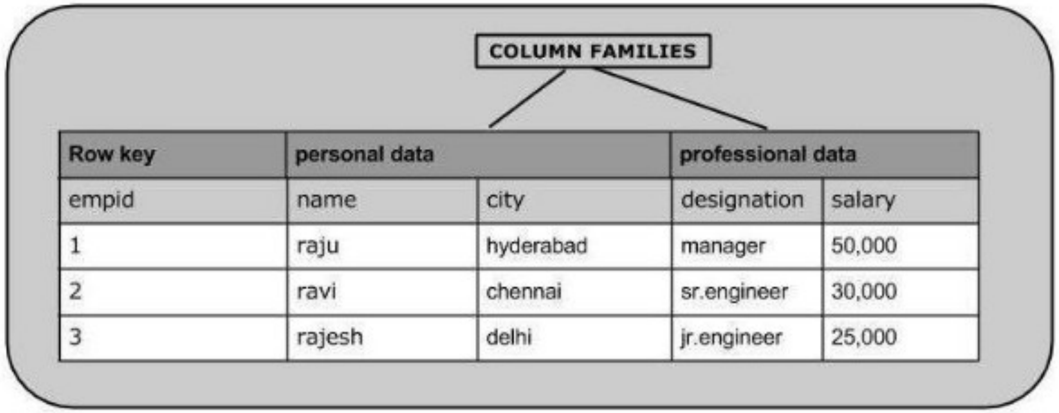


```
[cloudera@quickstart guest1]$ hbase shell
2016-10-13 00:39:51,901 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.8.0, rUnknown, Thu Jun 16 12:46:57 PDT 2016

hbase(main):001:0> create 'employee', 'personal data','professional data'
0 row(s) in 1.6860 seconds

=> Hbase::Table - employee
hbase(main):002:0> list
TABLE
employee
1 row(s) in 0.0180 seconds

=> ["employee"]
hbase(main):003:0> █
```



| Row key | personal data | professional data |
|---|---|---|
| | | |
| | | |

# Create Data


COLUMN FAMILIES

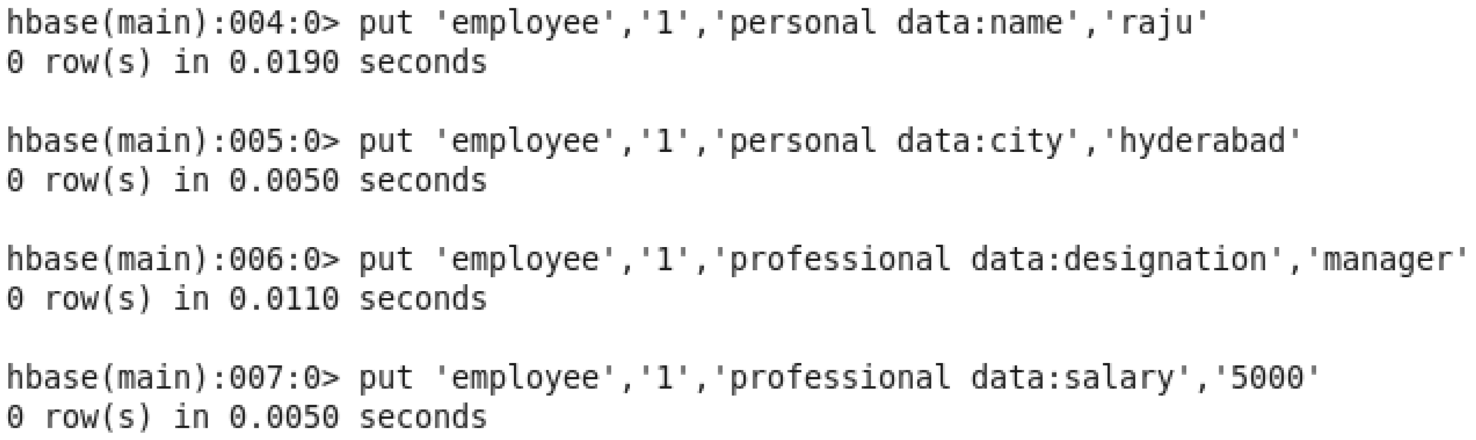
| Row key | personal data | | professional data | |
|---|---|---|---|---|
| empid | name | city | designation | salary |
| 1 | raju | hyderabad | manager | 50,000 |
| 2 | ravi | chennai | sr.engineer | 30,000 |
| 3 | rajesh | delhi | jr.engineer | 25,000 |

**hbase(main):004:0> put 'employee','1','personal data:name','raju'**

**hbase(main):005:0> put 'employee','1','personal data:city','hyderabad'**

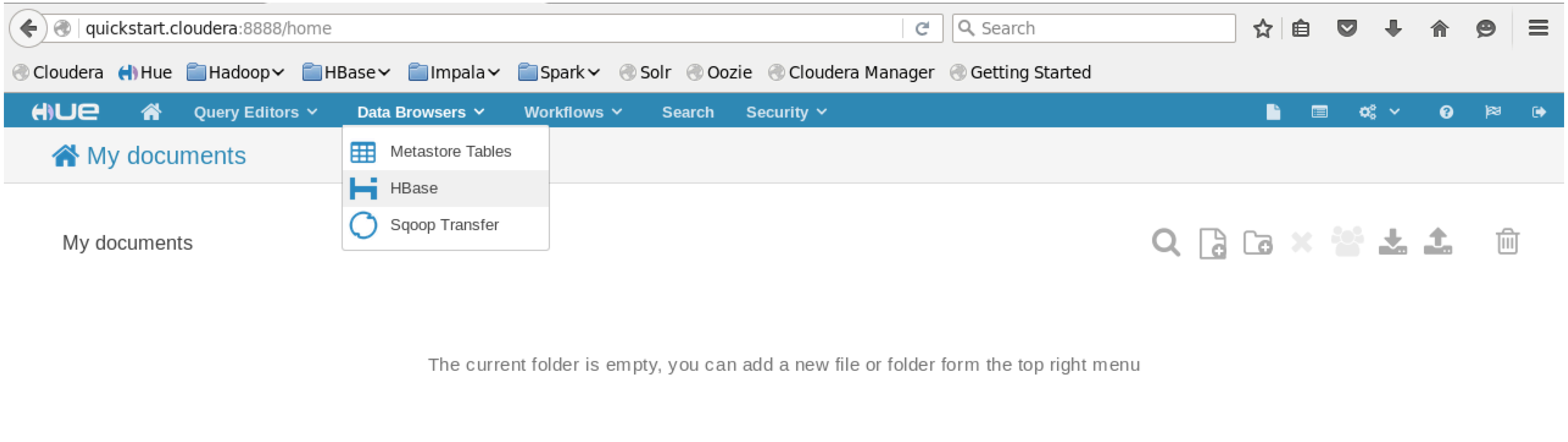
**hbase(main):006:0> put 'employee','1','professional data:designation','manager'**

**hbase(main):007:0> put 'employee','1','professional data:salary','5000'**



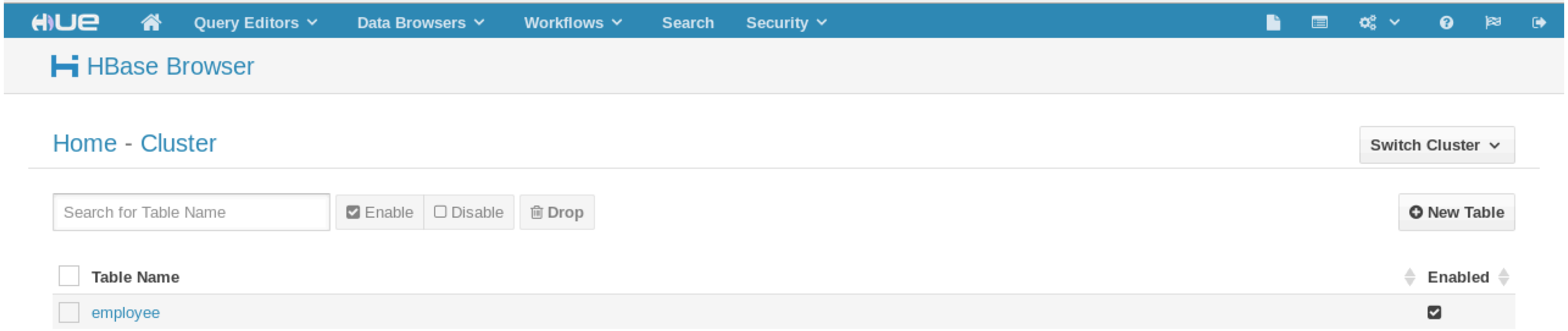
```
hbase(main):004:0> put 'employee','1','personal data:name','raju'
0 row(s) in 0.0190 seconds

hbase(main):005:0> put 'employee','1','personal data:city','hyderabad'
0 row(s) in 0.0050 seconds

hbase(main):006:0> put 'employee','1','professional data:designation','manager'
0 row(s) in 0.0110 seconds

hbase(main):007:0> put 'employee','1','professional data:salary','5000'
0 row(s) in 0.0050 seconds
```

# Running HBase Browser



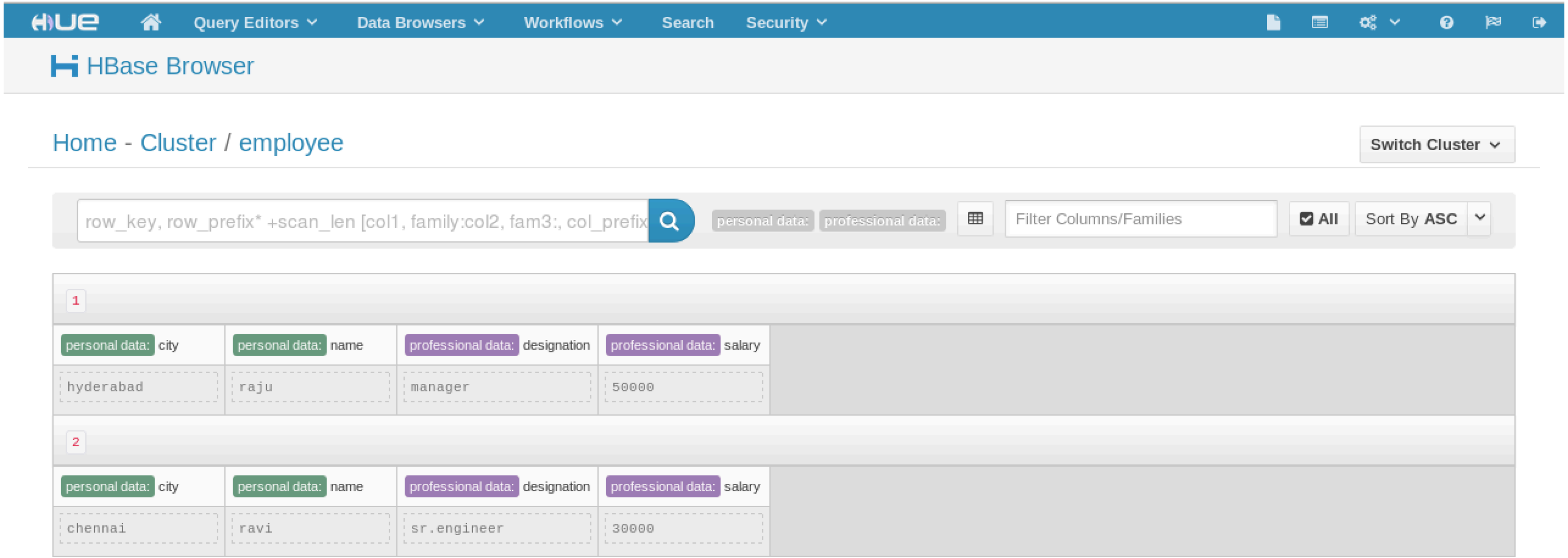
My documents

The current folder is empty, you can add a new file or folder form the top right menu



| □ | Table Name | Enabled |
|---|---|---|
| □ | employee | ☑ |

# Viewing Employee Table



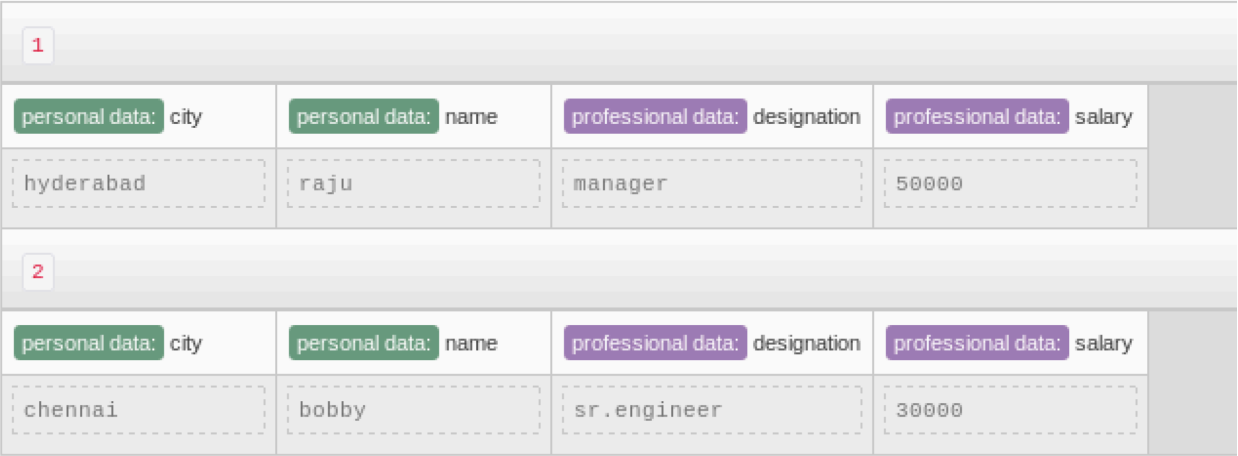
1

| personal data: city | personal data: name | professional data: designation | professional data: salary |
|---|---|---|---|
| hyderabad | raju | manager | 50000 |

2

| personal data: city | personal data: name | professional data: designation | professional data: salary |
|---|---|---|---|
| chennai | ravi | sr.engineer | 30000 |

**hbase(main):025:0> put 'employee','2','personal data:name','bobby'**



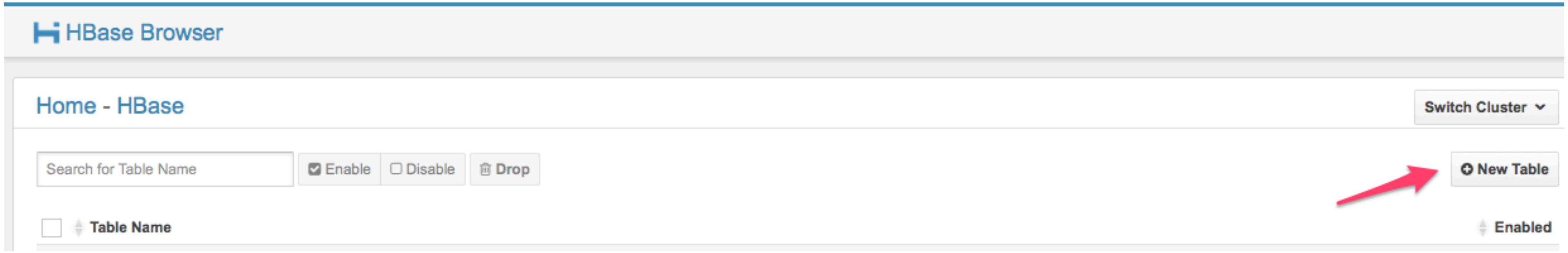
1

| personal data: city | personal data: name | professional data: designation | professional data: salary |
|---|---|---|---|
| hyderabad | raju | manager | 50000 |

2

| personal data: city | personal data: name | professional data: designation | professional data: salary |
|---|---|---|---|
| chennai | bobby | sr.engineer | 30000 |

# Create a table in HBase



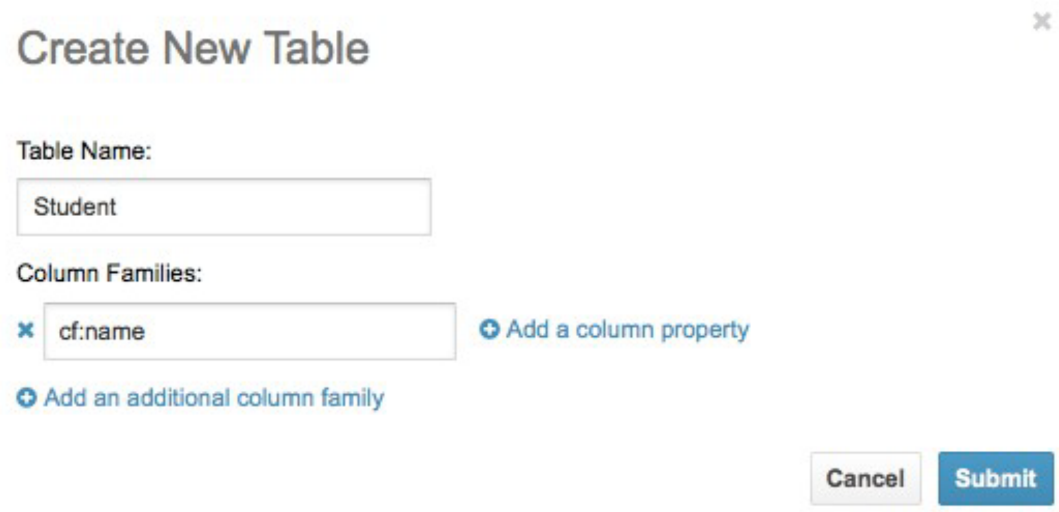
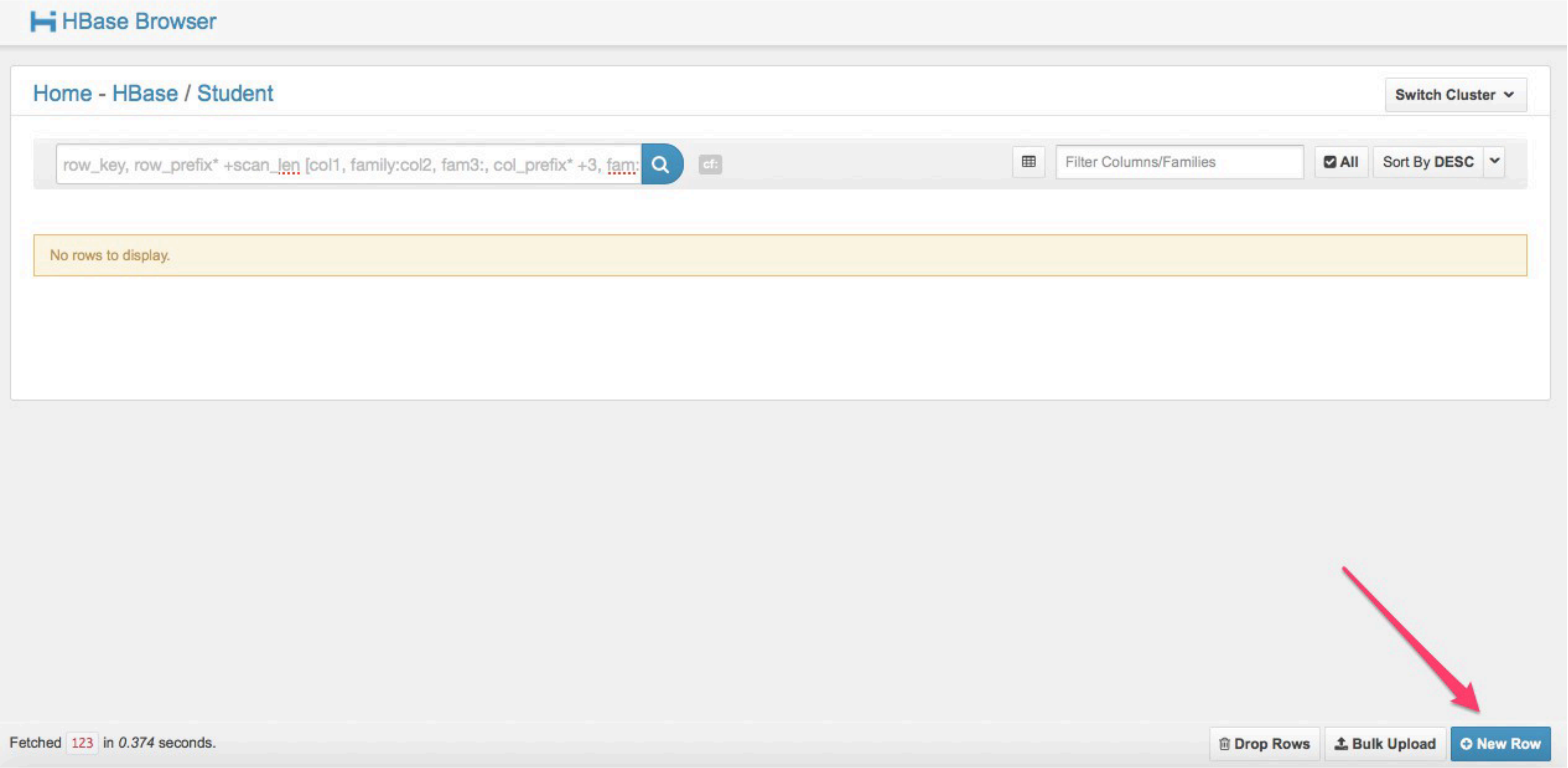


Table Name: Student

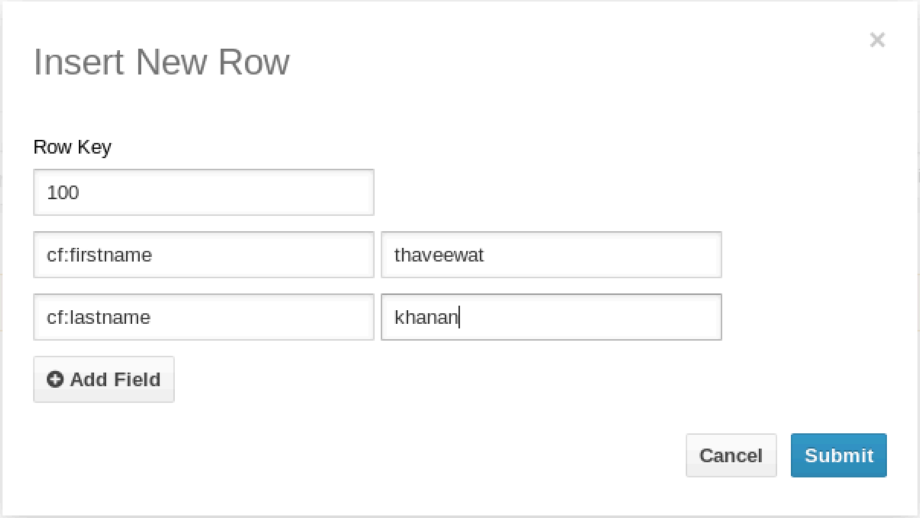
Column Families: cf:name



No rows to display.

Fetched 123 in *0.374* seconds.

# Add field into a new row

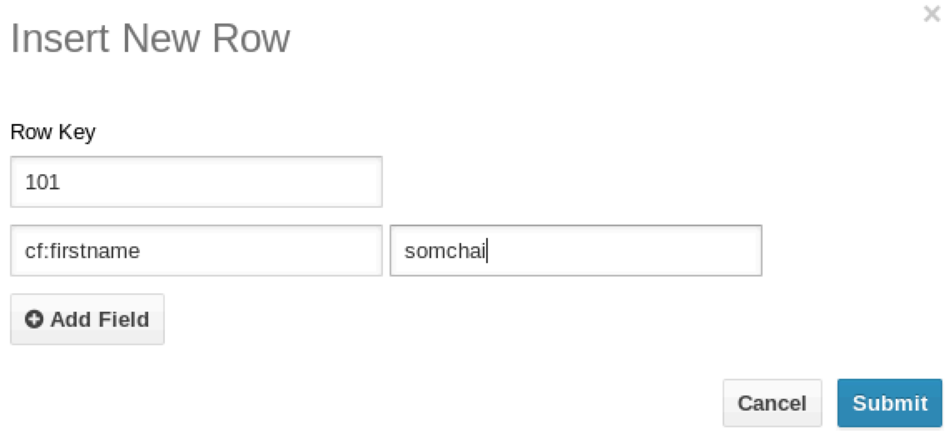


Insert New Row

Row Key: 100

cf:firstname: thaveewat

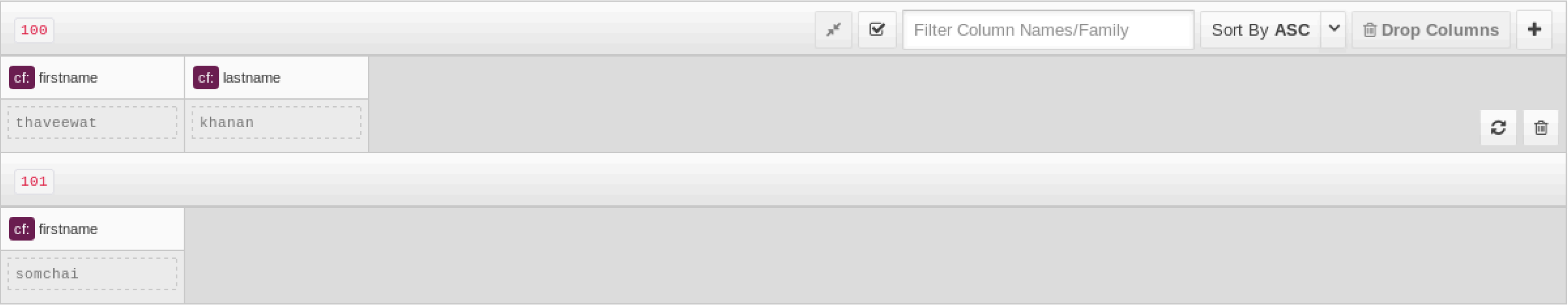
cf:lastname: khanan



Insert New Row

Row Key: 101

cf:firstname: somchai



100

| cf: firstname | cf: lastname |
|---|---|
| thaveewat | khanan |

101

| cf: firstname |
|---|
| somchai |