How to create example tables in HBase

Posted by Hue Team (http://gethue.com/author/hue/) on August 13, 2013 in HBase (http://gethue.com/category/hbase/), Tutorial (http://gethue.com/category/tutorial/), Video (http://gethue.com/category/video/) 5 Comments

Hue (http://gethue.tumblr.com/post/55581863077/hue-2-5-and-its-hbase-app-is-out) brings another new app for making Apache Hadoop easier to use: HBase Browser. Apache HBase (http://hbase.apache.org/) is the main keyvalue datastore for Hadoop. This post is the first episode describing the new user experience brought by the app. We will start by describing how to create some sample tables with various HBase schemas.

To help getting started with HBase, Hue now comes directly with theses examples! Just got to Hue on the /about/ (http://127.0.0.1:8888/about/) page and on Step 2 click on HBase and Hue will install them for you:

Screenshot from 2014-04-09 08:29:00

(http://gethue.com/wp-content/uploads/2013/08/Screenshot-from-2014-04-09-082900.png)

If you want to see the HBase Browser demo, jump episode 2 (http://gethue.com/the-web-ui-for-hbase-hbase-browser/)!

Tutorial

When building the new HBase Browser, we wanted to test the app against various HBase tables. It happened to be difficult to find some ready to play with schema and data to load. Hence, base on the most common uses cases, we created our our own HBase schemas and decided to share them in order to help anybody wanting to start with HBase.

This how-to describes how to create a very simple table that counts the daily number of votes for certain candidates and get you warmed-up. Then part 2 focuses on creating a HBase table with lot of columns and part 3 about inserting and visualizing binary data.

Setup

The HBase Browser application is tailored for quickly browsing huge tables and accessing any content. You can also create new tables, add data, modify existing cells and filter data with the autocompleting search bar.

The first step is to install HBase in your Hadoop cluster. We recommend to use the CDH packages (http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/CDH4-Installation-Guide/cdh4ig_topic_20_2.html). HBase Browser requires the Thrift 1 service (http://www.cloudera.com/content/cloudera-content/cloudera-docs/CDH4/latest/CDH4-Installation-Guide/cdh4ig_topic_20_5.html#topic_20_5_4_unique_1) to be started.

Then, grab the app from a special tarball release (http://gethue.com/downloads/releases/hbase/hue-hbase-2.5.0.tgz) of Hue or get the latest and slickest version from the nightly 'hue' package (http://nightly.cloudera.com/cdh4/). CDH 4.4 (target date early September) will bring a stable v1. After the installation, if HBase master is not running on the same host as Hue, have the app pointing to it by updating the hue.ini (https://github.com/cloudera/hue/blob/master/desktop/conf.dist/hue.ini#L505) and restarting Hue.

Then go to http://127.0.0.1:8888/hbase/ (http://127.0.0.1:8888/hbase/) to check that all is setup correctly! We show in the video how to create a table and add some columns in just a few clicks. In the next steps, we are showing how to create and populate a real life example table.

The sample data and scripts are published on github (https://github.com/romainr/hadoop-tutorials-examples/tree/master/hbase-tables). In a terminal, use git (http://git-scm.com/) to retrieve the repository:

```
cd /tmp
git clone https://github.com/romainr/hadoop-tutorials-examples.git
(https://github.com/romainr/hadoop-tutorials-examples.git)
cd hbase-tables
```

Analytics table

The goals of this data is to show the search and smart layout of HBase Browser.

This table contains more than 1000 columns of text. The idea is to have counters for 3 Web domains of 3 countries for each hour of the day. The data is then aggregated by day and for all the countries.

hour	1-24									day	1-365						total			
1				2				...	24	1				2	...	365	Fran	Italy	US	total
Fran	Italy	US	total	Fran	Italy	US	total	...	...	Fran	Italy	US	total							
10	5	100	115	...						100	50	1000	1150	...		...	...	...	...	...

Schema of the table

How to create the HBase table and insert some data:

1. Generate column names and data with create_schemas.py (https://raw.github.com/romainr/hadoop-tutorials-examples/master/hbase-tables/create_schemas.py). Run it with ./create_schemas.py
2. Upload the date data /tmp/hbase-analytics.tsv to HDFS with File Browser
3. In HBase Browser create a 'analytics' table with 3 column families 'hour', 'day', 'total'
4. Load the data into the analytics table with the HBase bulk import command (https://raw.github.com/romainr/hadoop-tutorials-examples/master/hbase-tables/load_data.sh).

It will trigger a MapReduce job and display the progress (https://github.com/romainr/hadoop-tutorials-examples/blob/master/hbase-tables/load_data.log) of the import.

That's it! Go open the analytics table in HBase Browser (http://127.0.0.1:8888/hbase/#Cluster/analytics)!

Binary table

This second tables focus on big data cells, various formats, demonstrating the preview and editing of data within HBase Browser.

We are re-using the app API for inserting into HBase some cells of various content types, e.g. text, json, pictures, binary…

1. First create a table ‘events’ with a column family ‘doc’.
2. Then cd in the root of Hue
3. cd /usr/share/hue
4. /opt/cloudera/parcels/CDH-4.X/share/hue (if using parcels)

And start the Hue shell build/env/bin/hue shell and type the content of locad_binary.py: (https://github.com/romainr/hadoop-tutorials-examples/blob/master/hbase-tables/load_binary.py)

Load the HBase API and insert some text data:

```
from hbase.api import HbaseApi

HbaseApi().putRow('Cluster', 'events', 'hue-20130801', {'doc:txt': 'Hue is
awesome!'})
HbaseApi().putRow('Cluster', 'events', 'hue-20130801', {'doc:json': '{"use
r": "hue", "coolness": "extra"}'})
HbaseApi().putRow('Cluster', 'events', 'hue-20130802', {'doc:version': 'I
like HBase'})
HbaseApi().putRow('Cluster', 'events', 'hue-20130802', {'doc:version': 'I
LOVE HBase'})
```

Then insert a picture, and HTML page and a PDF:

```
root='/tmp/hadoop-tutorials-examples'

HbaseApi().putRow('Cluster', 'events', 'hue-20130801', {'doc:img': open(ro
ot + '/hbase-tables/data/hue-logo.png', "rb").read()})
HbaseApi().putRow('Cluster', 'events', 'hue-20130801', {'doc:html': open(r
oot + '/hbase-tables/data/gethue.com.html', "rb").read()})
HbaseApi().putRow('Cluster', 'events', 'hue-20130801', {'doc:pdf': open(ro
ot + '/hbase-tables/data/gethue.pdf', "rb").read()})
```

Notice that the column names do not matter for the type detection. The go look at the events (http://127.0.0.1:8888/hbase/#Cluster/events) table and play around!

Conclusion

These two schemas and data enable the user to easily get started with HBase. This first version of HBase Browser brings a new way to quickly explore and search for some rows and columns. New versions will support bulk loads and upload in order to completely free the user from the command

line.

The new HBase Browser app will be demo-ed on these two tables in the upcoming blog posts, so stay tuned!

Share this article:

5 Comments

Isadora 2 years ago

Reply (http://gethue.com/hadoop-tutorial-how-to-create-example-tables-in-hbase/?replytocom=52831#respond)

When I do step “2 – Upload the date data /tmp/hbase-analytics.tsv to HDFS with File Browser” it gives me an error “Api Error: Failed setting up proxy interface org.apache.hadoop.hbase.ipc.HRegionInterface to localhost.localdomain/127.0.0.1:60020 after attempts=1”

Hue Team 2 years ago

Reply (http://gethue.com/hadoop-tutorial-how-to-create-example-tables-in-hbase/?replytocom=52859#respond)

60020 is the port of a HBase Region server, are you sure that it is up?

Sabanam Lakhey 12 months ago

Reply (http://gethue.com/hadoop-tutorial-how-to-create-example-tables-in-hbase/?replytocom=66969#respond)

Can we upload tiff files in the binary table? I tried uploading tiff files but while viewing the file I only get the binary text.

Hue Team 12 months ago

Reply (http://gethue.com/hadoop-tutorial-how-to-create-example-tables-in-hbase/?replytocom=66970#respond)

The live preview of files is just for these types: ‘image/png’, ‘image/gif’, ‘image/jpg’ and ‘application/pdf’

Leave a reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

POST COMMENT

Languages

Categories

ADLS (1) (http://gethue.com/category/adls/) ›

MapReduce (2) (http://gethue.com/category/mapreduce/) ›

Metastore (11) (http://gethue.com/category/metastore/) ›

Metastore @fr (1) (http://gethue.com/category/metastore-fr/) ›

Non classifié(e) (http://gethue.com/category/non-classifiee/) ›
(7)

Notebook (6) (http://gethue.com/category/notebook/) ›

Oozie (30) (http://gethue.com/category/oozie/) ›

Pig (9) (http://gethue.com/category/pig/) ›

Pig @fr (1) (http://gethue.com/category/pig-fr/) ›

Presentation (26) (http://gethue.com/category/presentation/) ›

Programming (23) (http://gethue.com/category/programming/) ›

Release (22) (http://gethue.com/category/release/) ›

Release @fr (3) (http://gethue.com/category/release-fr/) ›

SDK (26) (http://gethue.com/category/sdk/) ›

Search (25) (http://gethue.com/category/search/) ›

Search @fr (1) (http://gethue.com/category/search-fr/) ›

Search @it (1) (http://gethue.com/category/search-it/) ›

Security (22) (http://gethue.com/category/security-2/) ›

Spark (19) (http://gethue.com/category/spark/) ›

Spark @fr (1) (http://gethue.com/category/spark-fr/) ›

Spark @it (1) (http://gethue.com/category/spark-it/) ›

SQL (20) (http://gethue.com/category/sql/) ›

Sqoop (5) (http://gethue.com/category/sqoop/) ›

Team (16) (http://gethue.com/category/team/) ›

Tutorial (78) (http://gethue.com/category/tutorial/) ›

Tutorial @fr (6) (http://gethue.com/category/tutorial-fr/) ›

Tutorial 2 (8) (http://gethue.com/category/season2/) ›

Archives

Archives Select Month

Latest Tweets (http://www.twitter.com/gethue)

- Community: Using Hue to interact with Apache Kylin in your cluster or on AWS https://t.co/xm4fXIsdyf (https://t.co/xm4fXIsdyf) https://t.co/0ggbNbgu0J
- Browsing Impala Query Execution within the SQL Editor #hue (https://twitter.com/search?q=%23hue&src=hash) #impala (https://twitter.com/search?q=%23impala&src=hash) https://t.co/AKcN3tM0HI (https://t.co/AKcN3tM0HI) https://t.co/0EqEkCDqIH
- Browsing ADLS data, querying it with SQL and exporting the results back in Hue 4.2 https://t.co/Bd1SdlTewV (https://t.co/Bd1SdlTewV)… https://t.co/R7CWJWldsw (https://t.co/R7CWJWldsw)
- Hue 4.1 is released! Stability, Hue 4 UI, SQL and Solr improvements! https://t.co/oAozrpHhpU (https://t.co/oAozrpHhpU) #gethue (https://twitter.com/search?q=%23gethue&src=hash) #analytics (https://twitter.com/search?q=%23analytics&src=hash) https://t.co/2T560l3XrM
- The story about the new Hue 4 user interface: https://t.co/jJtgtnaZ70 (https://t.co/jJtgtnaZ70) https://t.co/EMN19VagnL