HIVE - LABS

**Step 1.** Let’s do some hbase and hive integration

create table studentname (studentid string, name string, email string) stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' with SERDEPROPERTIES ("hbase.columns.mapping" = "post:name,post:email");

Connect to HBase Shell; hbase(main):001:0> describe "studentname" Also in hive, describe the table and see.

Can you think about a use of this setup?

**Step 2.** Let’s create a complex table which is more likely be the case when you will create a real table for you project.

Create table mycomplex\_studentname (names array<int>,

properties map<string, int>,

others struct<choice:string, Choice\_id:INT>);

Let’s select data from this table;

Select names[0], properties[‘height’], others.choice\_id from mycomplex;

Do you see a map reduce job is running?

**Step 3.** Let’s create a view using same syntax in standard sql;

Create view test\_studentname as select \* from your table;

**Step 4.** How can you view all the UDF functions available?

To view all the available functions, please type “show functions” in hive prompt.

**Step 5.** How can you view what one UDF functions do?

To understand what a function can do, you can type “desc function year;”

**Step 6.** Create a simple Hive table and drop respective hdfs directory to delete all the data.

What happened?

**Step 7.** Running hive commands in unix prompt:

$ hive –e ‘select \* from YOURTABLE’

$hive –e ‘show tables’

**Step 8.** Installation and configuration:

Go to $HIVE\_HOME and get familiar with hive-site.xml or other configuration files.

**Step 9.** Let’s create a UDF Function:

export CLASSPATH=$CLASSPATH:$HIVE\_HOME/lib/hive-exec-0.7.1-cdh3u2.jar

export CLASSPATH=$CLASSPATH:$HADOOP\_HOME/hadoop-core.jar

Create a java File with the following content;

+++++++++++++++++++++++++++++++++++++

import org.apache.hadoop.io.Text;

import org.apache.hadoop.hive.ql.exec.UDF; public final class MonthExtract extends UDF {

public Text evaluate(final Text col) { String str = col.toString();

if (str==null) return null;

return new Text(str.substring(6,2));

}

}

+++++++++++++++++++++++++++++++++++++

Compile this file as java MonthExtract.java Create a jar file out of this file as

jar cvf monthextract.jar MonthExtract.class Now connect to Hive and do the following:

add jar /home/um/monthextract.jar

Create a table with one column as date and store the following value : 2011-10-10;

Next, write a select query to select the data from your table and apply the udf function that you just created.

Hive > select monthextract(columnname) from Yourtable;