

>create a dir in local file system FOR SQOOP IMPORT  
>IMPORT ALL TABLES BY USING --as-avrodatafile EXTENSION  
>now copy the .avsc file created by sqoop import back to HDFS  
>BY `hadoop fs -copyFromLocal *.avsc /user/cloudera/dataForAvscFiles`

## NOW MOVE TO HIVE

The data should be in hive/warehouse or else you will end up with errors

```
CREATE EXTERNAL TABLE orders
STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/retail_stage.db/orders'
TBLPROPERTIES
('avro.schema.url'='hdfs://quickstart.cloudera/user/cloudera/avscData/orders.avsc');
We need to extract the schema by using avro-tools
```

## LOAD DATA

```
load data inpath '/user/cloudera/candy/orders/part-m-00000.avro' into table
orders;
load data inpath '/user/cloudera/candy/orders/part-m-00001.avro' into table
orders;
```

```
CREATE EXTERNAL TABLE fit
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
STORED AS INPUTFORMAT
'org.apache.hadoop.hive.ql.io.avro.AvroContainerInputFormat'
OUTPUTFORMAT
'org.apache.hadoop.hive.ql.io.avro.AvroContainerOutputFormat'
LOCATION 'hdfs:///user/hive/warehouse/retail_stage.db/categories'
TBLPROPERTIES
('avro.schema.url'='hdfs://quickstart.cloudera/user/cloudera/avscData/categories.avsc');
```

```
CREATE EXTERNAL TABLE janma
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
```

STORED AS INPUTFORMAT

'org.apache.hadoop.hive.ql.io.avro.AvroContainerInputFormat'

OUTPUTFORMAT

'org.apache.hadoop.hive.ql.io.avro.AvroContainerOutputFormat'

LOCATION 'hdfs:///user/hive/warehouse/retail\_stage.db/departments'

TBLPROPERTIES ('avro.schema.literal' = '{

"type" : "record",

"name" : "departments",

"doc" : "Sqoop import of departments",

"fields" : [ {

"name" : "department\_id",

"type" : [ "null", "int" ],

"default" : null,

"columnName" : "department\_id",

"sqlType" : "4"

}, {

"name" : "department\_name",

"type" : [ "null", "string" ],

"default" : null,

"columnName" : "department\_name",

"sqlType" : "12"

} ],

"tableName" : "departments"

}');  
=====