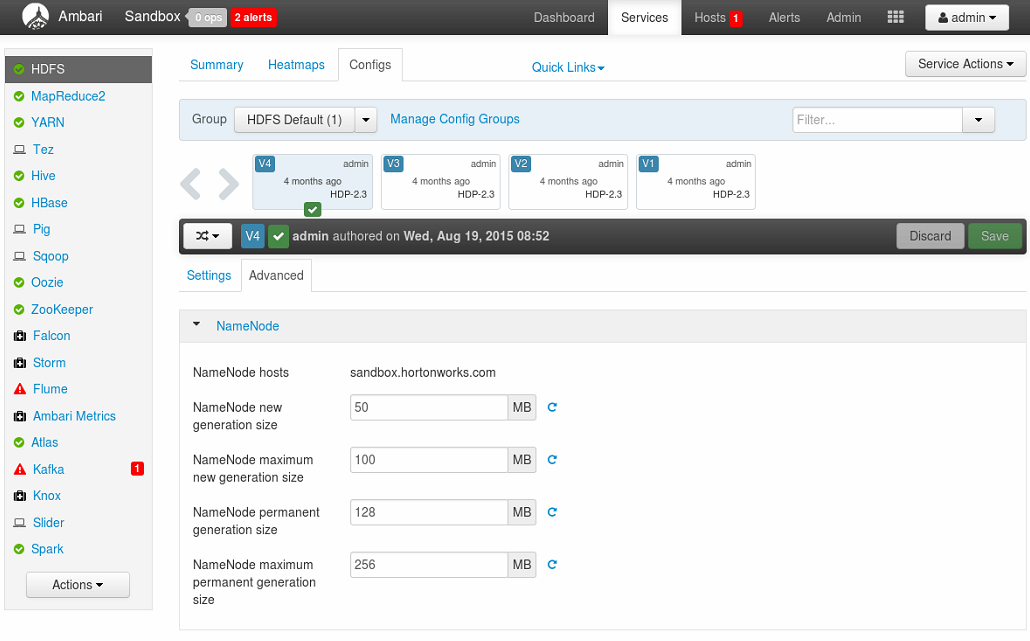
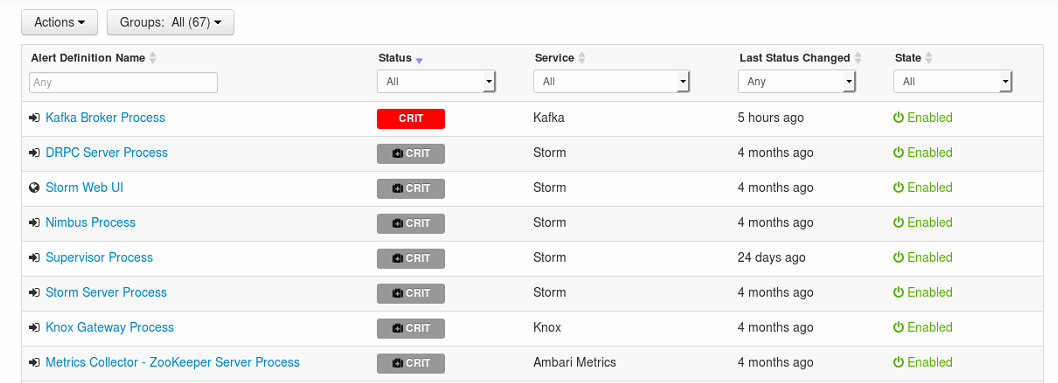
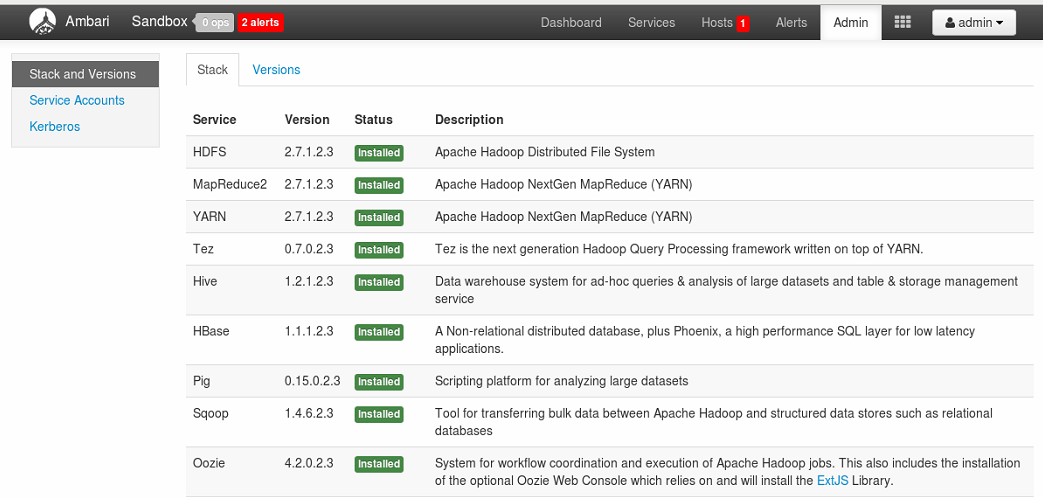
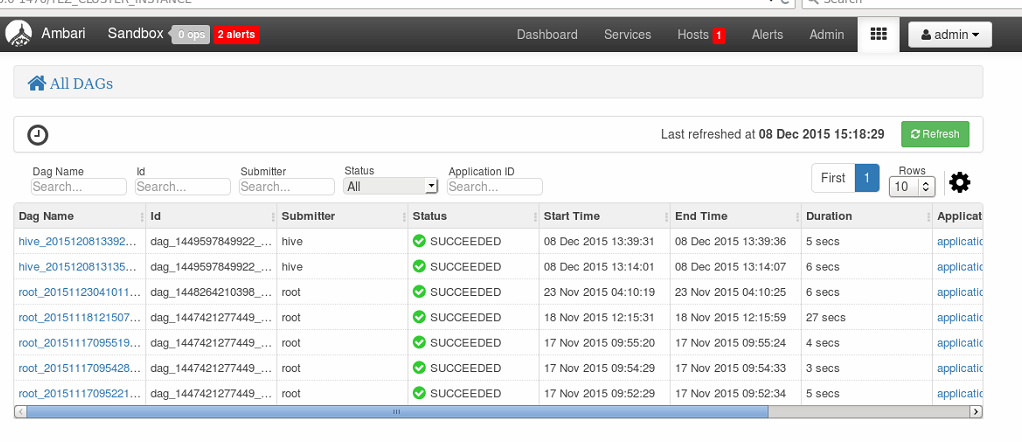
**Ambari**  
  
A completely open framework for provisioning, managing and monitoring Apache Hadoop clusters.It can be used for monitoring of the all services states (left menu). It's possible to get some information about different settings of a particular service. Besides, the settings above can be adjusted in a heartbreak.

  
Besides, Ambari gives a possibility to get into details about different alerts.   
  
  
  
  
In addition, Ambari shows detailed explanation of each of its services. Current version is also provided in this description.

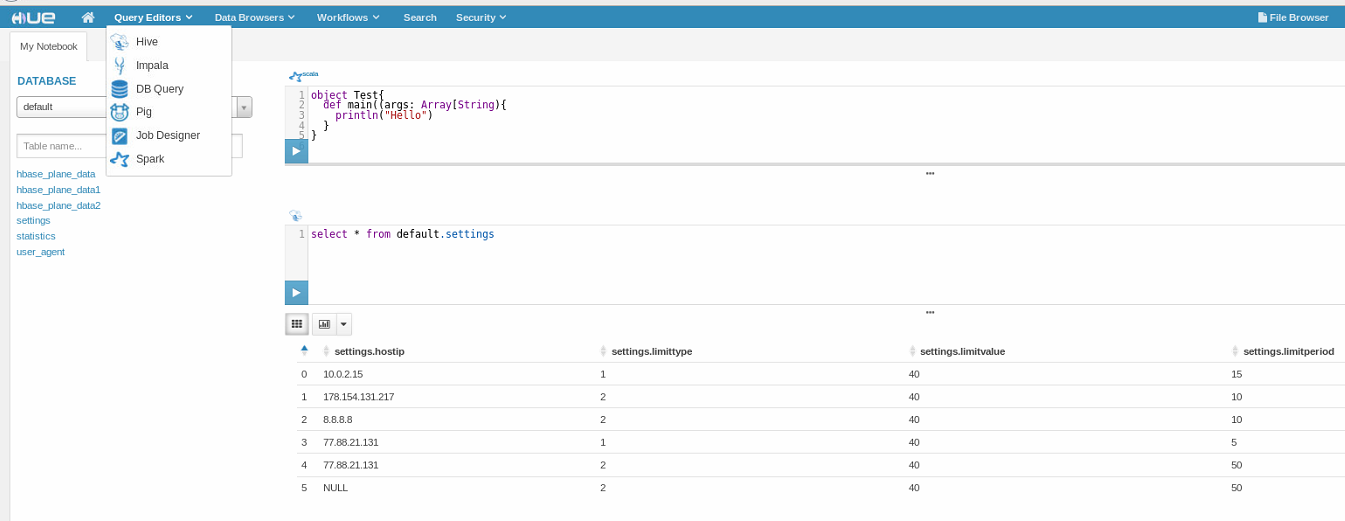


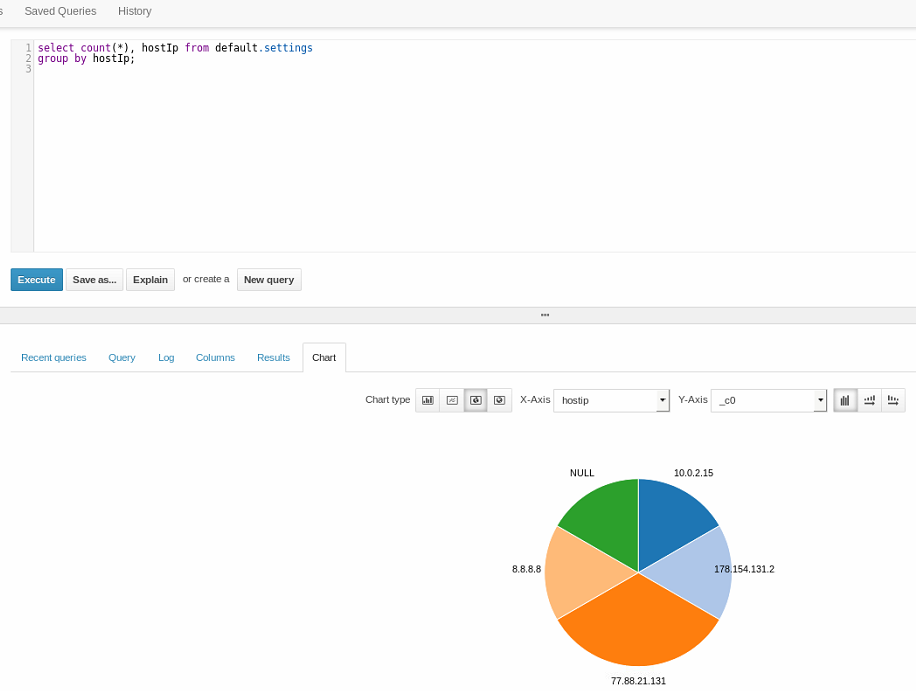
Finally, it's also possible to find some information about Tez engine: list of jobs, their statuses, their submitters and etc



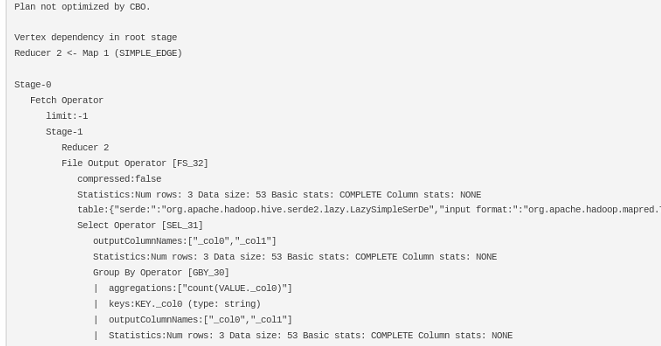
**Hue**

Hue provides a rich interface of choice for data analysts. It's possible to get data from different sources. It possible to get data, using Spark and its integration with Hive.

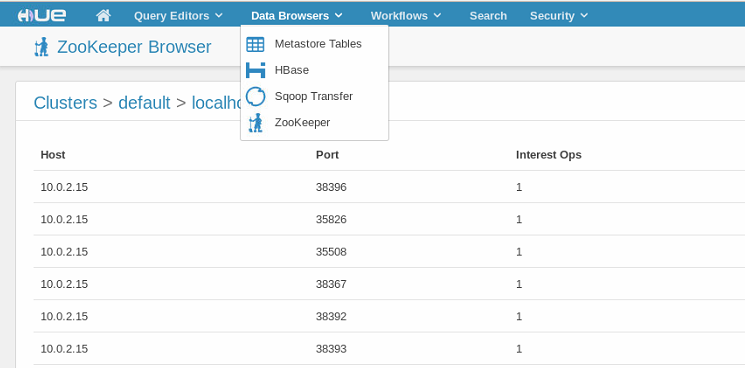


It's also possible to create a basic visualization of the query result  
  


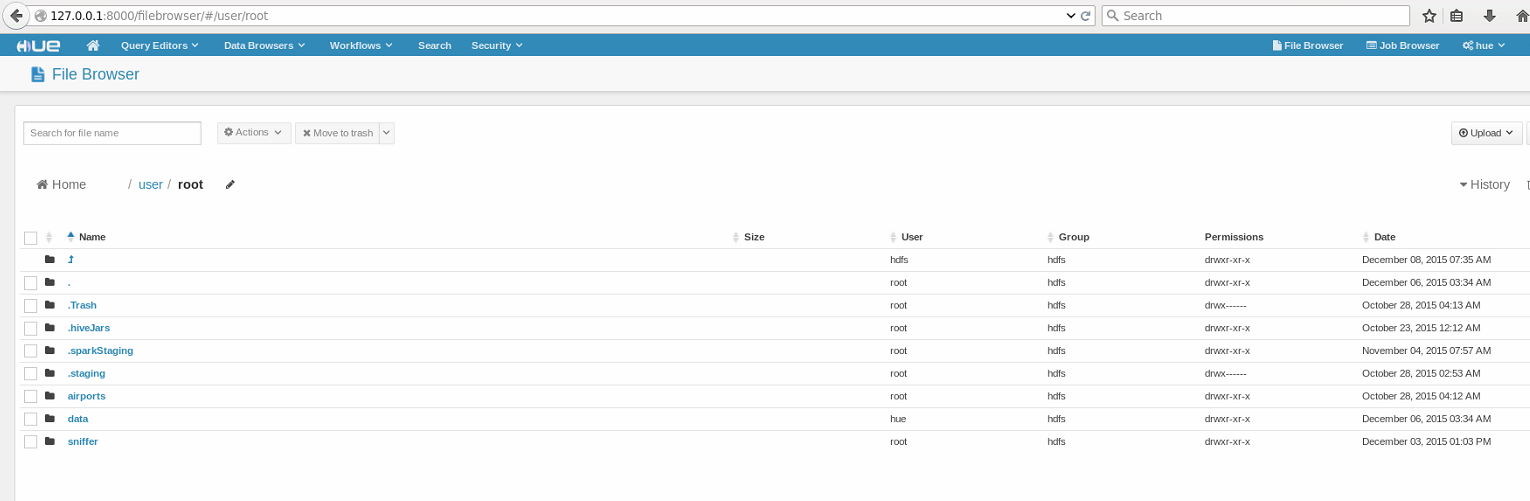
Besides, Hue provides a good opportunity to analyze a query. It's possible to get some information about the execution plan



In addition, Hue is connected with many HDP services, such as Zookeeper.



Finally, it's possible to get some information about files, stored on HDFS or in the local file system

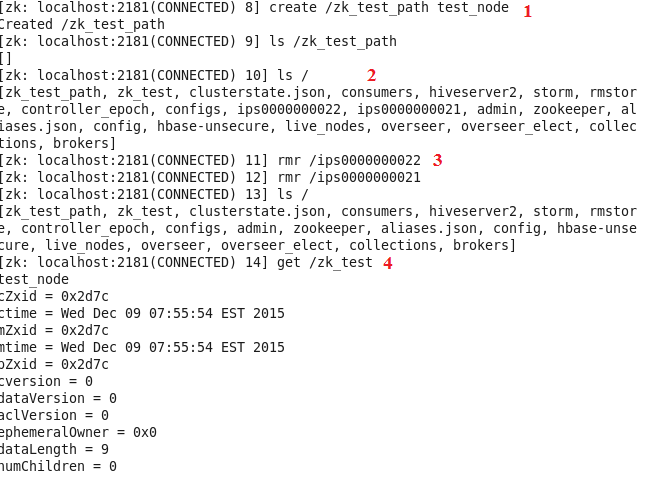


**Zookeeper**

ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.

It’s possible to work with this web-server from either cmd or using Zookeeper Rest.

There are a typical workflow of working with Zookeeper, using cmd:



1. A new znode was created.
2. It’s possible to list all existing nodes
3. It’s possible to delete the unused ones
4. It’s possible to get information about a particular node, which was created in particular directory.

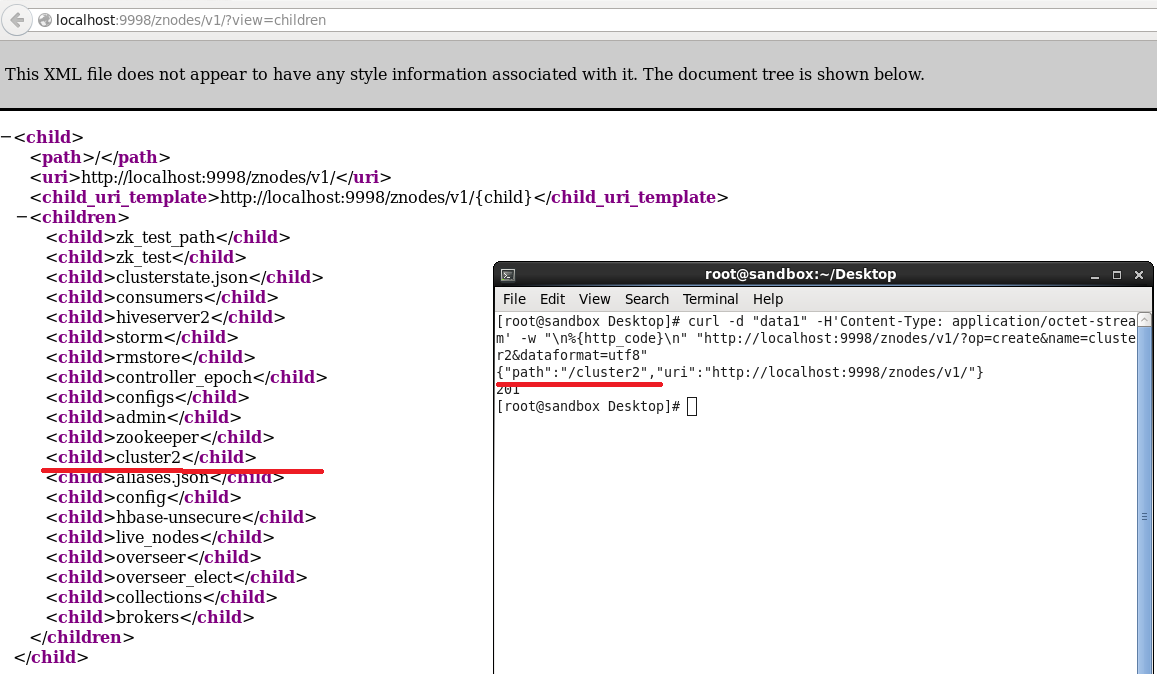
Besides, there is a rest-client for more comfortable work with Zookeeper:  
 <https://github.com/apache/zookeeper/tree/trunk/src/contrib/rest>

After running the Zookeeper rest service, the list of all znodes will be available by

curl <http://localhost:9998/znodes/v1/>



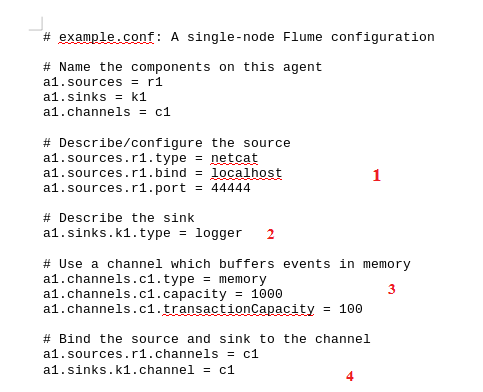
It’s also possible to manage znodes:



**Flume**

Apache Flume is a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store.

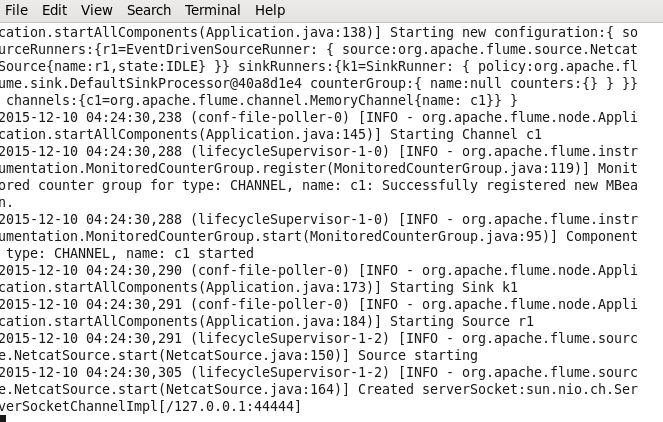
The main file for integration different sources through Flume is called \*.conf.



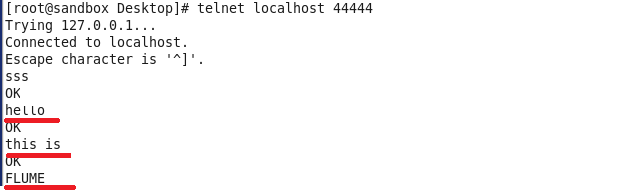
1. There are the properties of the source. (input)
2. There are the properties of the sink. (output)
3. There are some properties of the channel, such as capacity and type of the channel
4. There are properties, related to the integration between source (input) and sink (output) through channel.

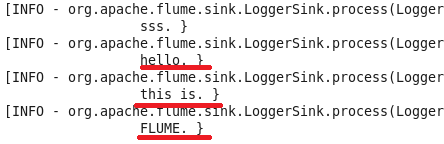
Flume can be started with the following command:

bin/flume-ng agent --conf conf --conf-file example.conf --name a1 Dflume.root.logger=INFO,console



There are the result of sending messages:





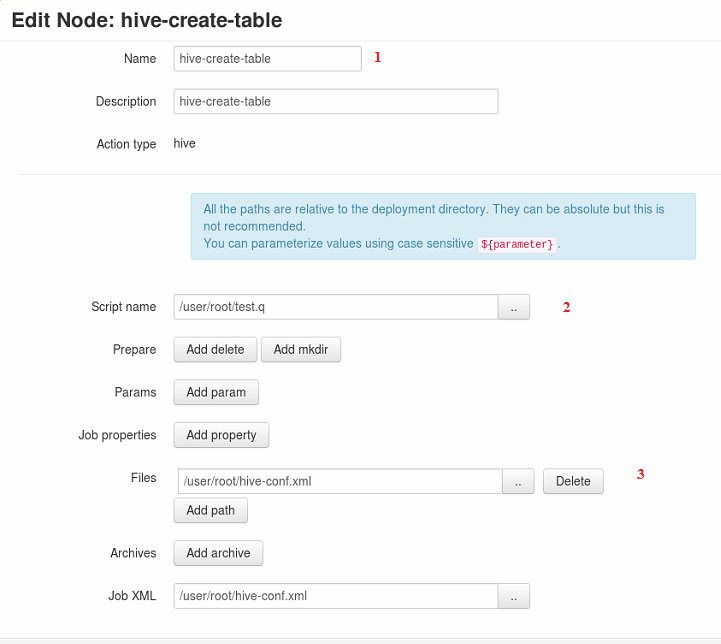
**Oozie**

Oozie is a workflow scheduler system to manage Apache Hadoop jobs.

It’s possible to work with Oozie, using Hue. It’s also possible to schedule different jobs (Hive, Sqoop, Map-Reduce etc).

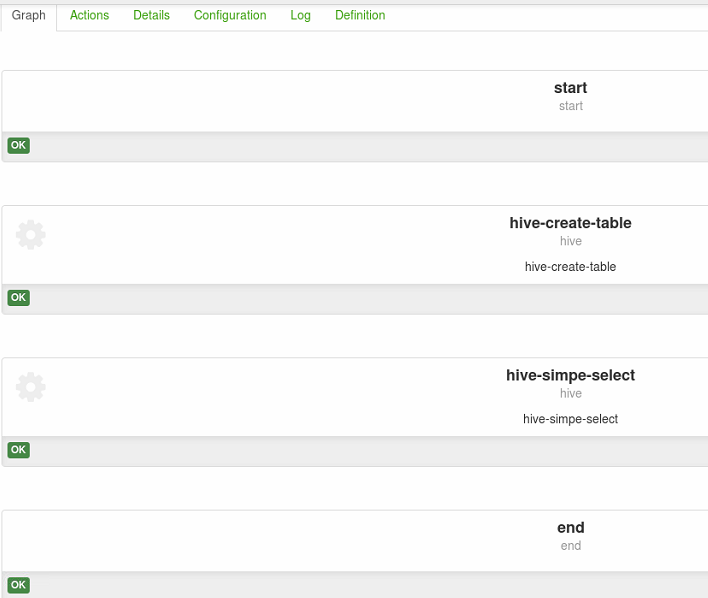
There is a small example of scheduling 2 Hive jobs. The first one creates a table and inserts all rows from an existing one to the new one. The second one selects all inserted rows.

This screenshot shows how the first jobs can be created:

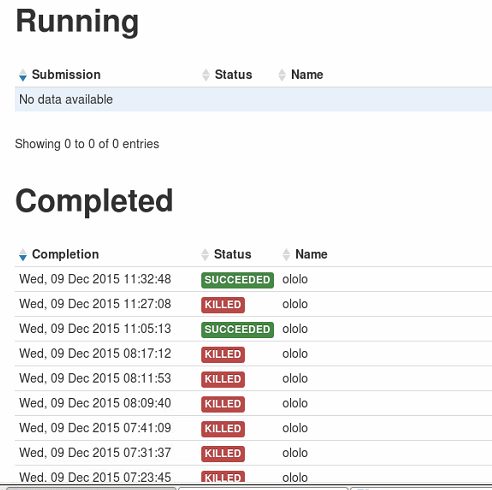


1. The name of the job
2. The path to the Hive script
3. Some basic hive configs

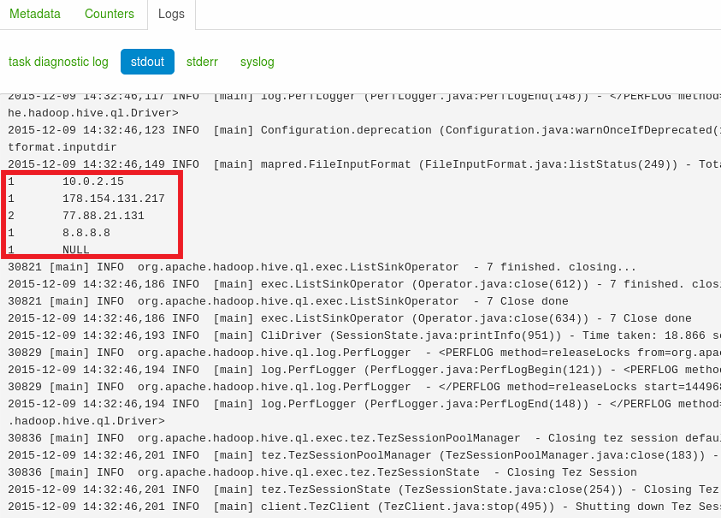
This panel is used to create a workflow of the jobs. It’s also possible to edit the jobs and to get some information about the current state and to read logs.



It’s also possible to get some information about the status of submitted workflows



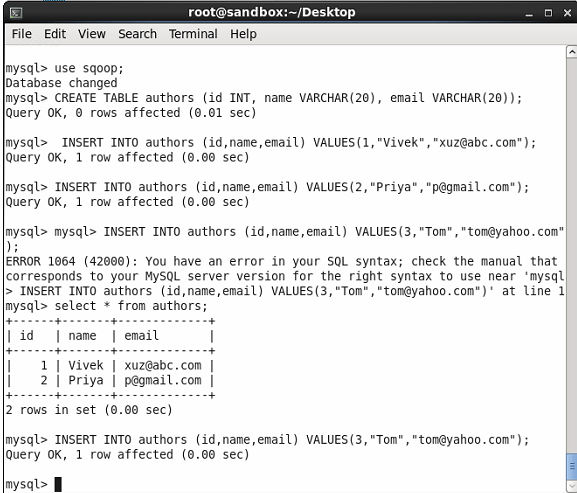
This is the result of the workflow from the log file:



**Sqoop**

Apache Sqoop(TM) is a tool designed for efficiently transferring bulk data between [Apache Hadoop](http://hadoop.apache.org/) and structured datastores such as relational databases.

To run a basic example, I’ve created a table in MySql and inserted some rows



The following command is used to import data from MySql :

sqoop import --connect jdbc:mysql://localhost/sqoop --username root --password root --table authors --m 1

C:\Users\Vera_Sidarovich\Desktop\screenshots\sqoop\sqoop_notes.png

After all, all data is places on HDFS. It’s possible to see them by the command

hadoop fs -cat /authors/part-m-\*

or through Hue:

