Introduction to Hortonworks Sandbox

What is Hortonworks?

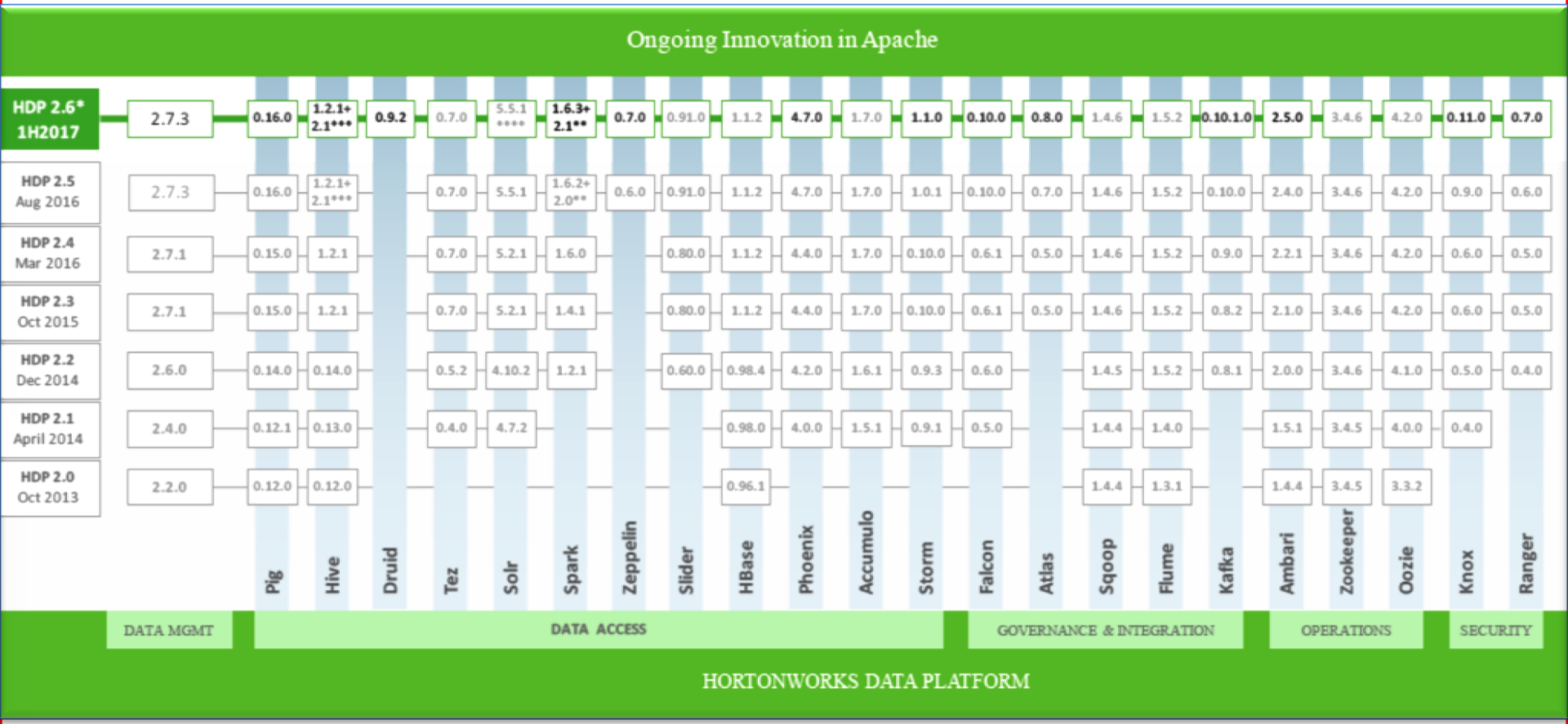
- Hortonworks is a business computer software company focusing on the development and support of Apache Hadoop.
- Hortonworks' product named Hortonworks Data Platform (HDP).
- HDP includes Apache Hadoop and is used for storing, processing, and analyzing large volumes of data from many sources and formats.
- Provides world-class support for Enterprise Hadoop from development to production.
- Is an open and single platform for any data and any workload.

Why Hortonworks Data Platform (HDP)?

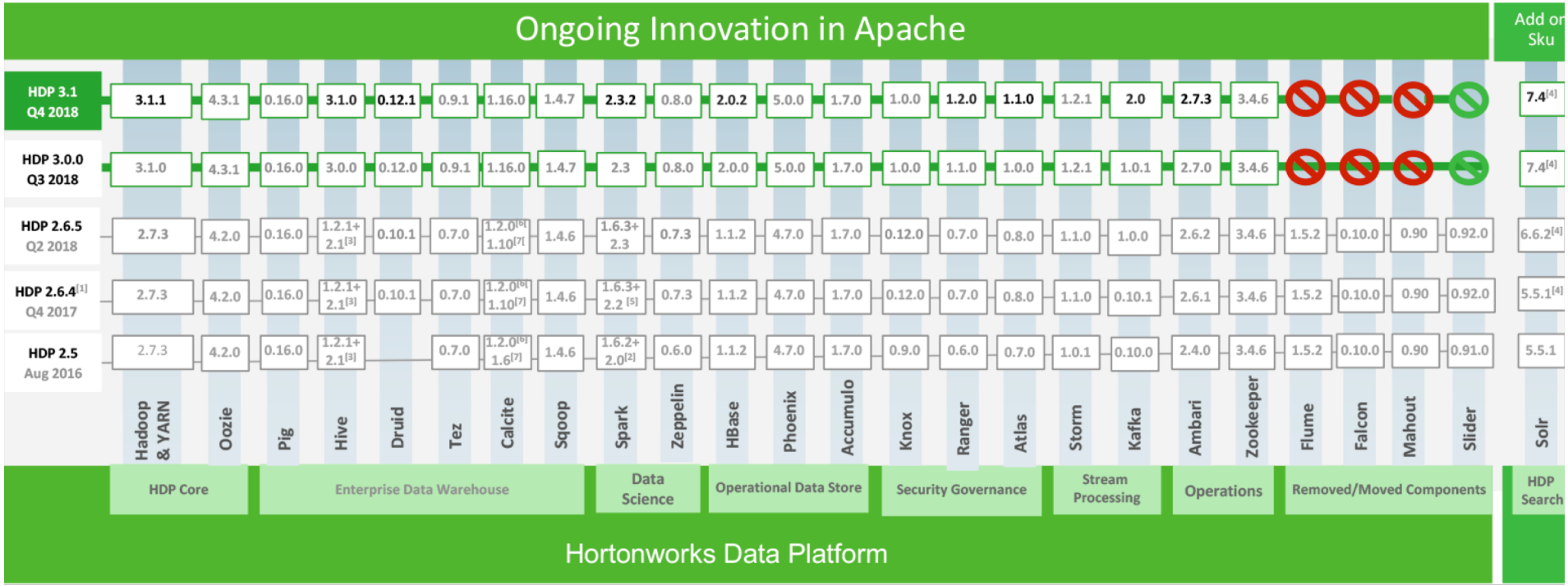
- Is complete, offering not just data processing and management, but the enterprise capabilities to match the demands of an enterprise spanning security, governance and operations. All of this is delivered in 100% open source.
- Is wholly integrated with existing data center investments and can be deployed in almost all computing environments.
- The platform includes various Apache Hadoop projects including the Hadoop Distributed File System (HDFS), MapReduce, Pig, Hive, HBase, Hue, Oozie, Ambari, Storm, Spark, Zookeeper and more.

HDP 2.x

The HDP (2.x) distribution increasingly incorporates many innovations in the Hadoop ecosystem



HDP 3.x

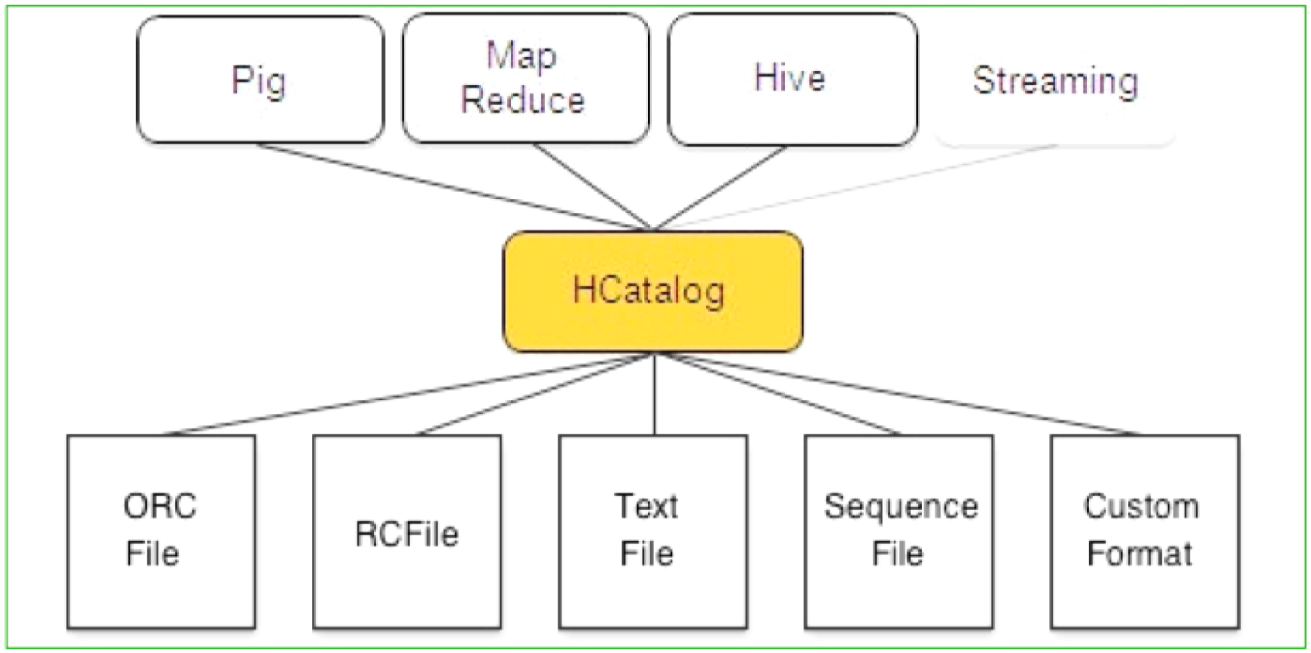

Ongoing Innovation in Apache
Add or Sku
HDP 3.1 Q4 2018
3.1.1 4.3.1 0.16.0 3.1.0 0.12.1 0.9.1 1.16.0 1.4.7 2.3.2 0.8.0 2.0.2 5.0.0 1.7.0 1.0.0 1.2.0 1.1.0 1.2.1 2.0 2.7.3 3.4.6 7.4[4]
HDP 3.0.0 Q3 2018
3.1.0 4.3.1 0.16.0 3.0.0 0.12.0 0.9.1 1.16.0 1.4.7 2.3 0.8.0 2.0.0 5.0.0 1.7.0 1.0.0 1.1.0 1.0.0 1.2.1 1.0.1 2.7.0 3.4.6 7.4[4]
HDP 2.6.5 Q2 2018
2.7.3 4.2.0 0.16.0 1.2.1+ 2.1[3] 0.10.1 0.7.0 1.2.0[6] 1.10[7] 1.4.6 1.6.3+ 2.3 0.7.3 1.1.2 4.7.0 1.7.0 0.12.0 0.7.0 0.8.0 1.1.0 1.0.0 2.6.2 3.4.6 1.5.2 0.10.0 0.90 0.92.0 6.6.2[4]
HDP 2.6.4[1] Q4 2017
2.7.3 4.2.0 0.16.0 1.2.1+ 2.1[3] 0.10.1 0.7.0 1.2.0[6] 1.10[7] 1.4.6 1.6.3+ 2.2[5] 0.7.3 1.1.2 4.7.0 1.7.0 0.12.0 0.7.0 0.8.0 1.1.0 0.10.1 2.6.1 3.4.6 1.5.2 0.10.0 0.90 0.92.0 5.5.1[4]
HDP 2.5 Aug 2016
2.7.3 4.2.0 0.16.0 1.2.1+ 2.1[3] 0.7.0 1.2.0[6] 1.6[7] 1.4.6 1.6.2+ 2.0[2] 0.6.0 1.1.2 4.7.0 1.7.0 0.9.0 0.6.0 0.7.0 1.0.1 0.10.0 2.4.0 3.4.6 1.5.2 0.10.0 0.90 0.91.0 5.5.1
Hadoop & YARN
Oozie
Pig
Hive
Druid
Tez
Calcite
Sqoop
Spark
Zeppelin
HBase
Phoenix
Accumulo
Knox
Ranger
Atlas
Storm
Kafka
Ambari
Zookeeper
Flume
Falcon
Mahout
Slider
Solr
HDP Core
Enterprise Data Warehouse
Data Science
Operational Data Store
Security Governance
Stream Processing
Operations
Removed/Moved Components
HDP Search
Hortonworks Data Platform

What is Hortonworks Sandbox?

- Hortonworks Sandbox is a personal, portable Hadoop environment that provides an easy and effective way to learn Enterprise Hadoop on a single-node cluster in a virtual machine.
- It comes with a pre-configured image.

HCatalog

- HCatalog is a table and storage management layer for Hadoop.
- It enables users with different data processing tools (Pig, MapReduce) to more easily read and write data.
- HCatalog's table abstraction presents users with a relational view of data in HDFS and ensures that users need not worry about where or in what format their data is stored.
- By default, HCatalog supports RCFile, CSV, JSON, and SequenceFile, and ORC file formats.

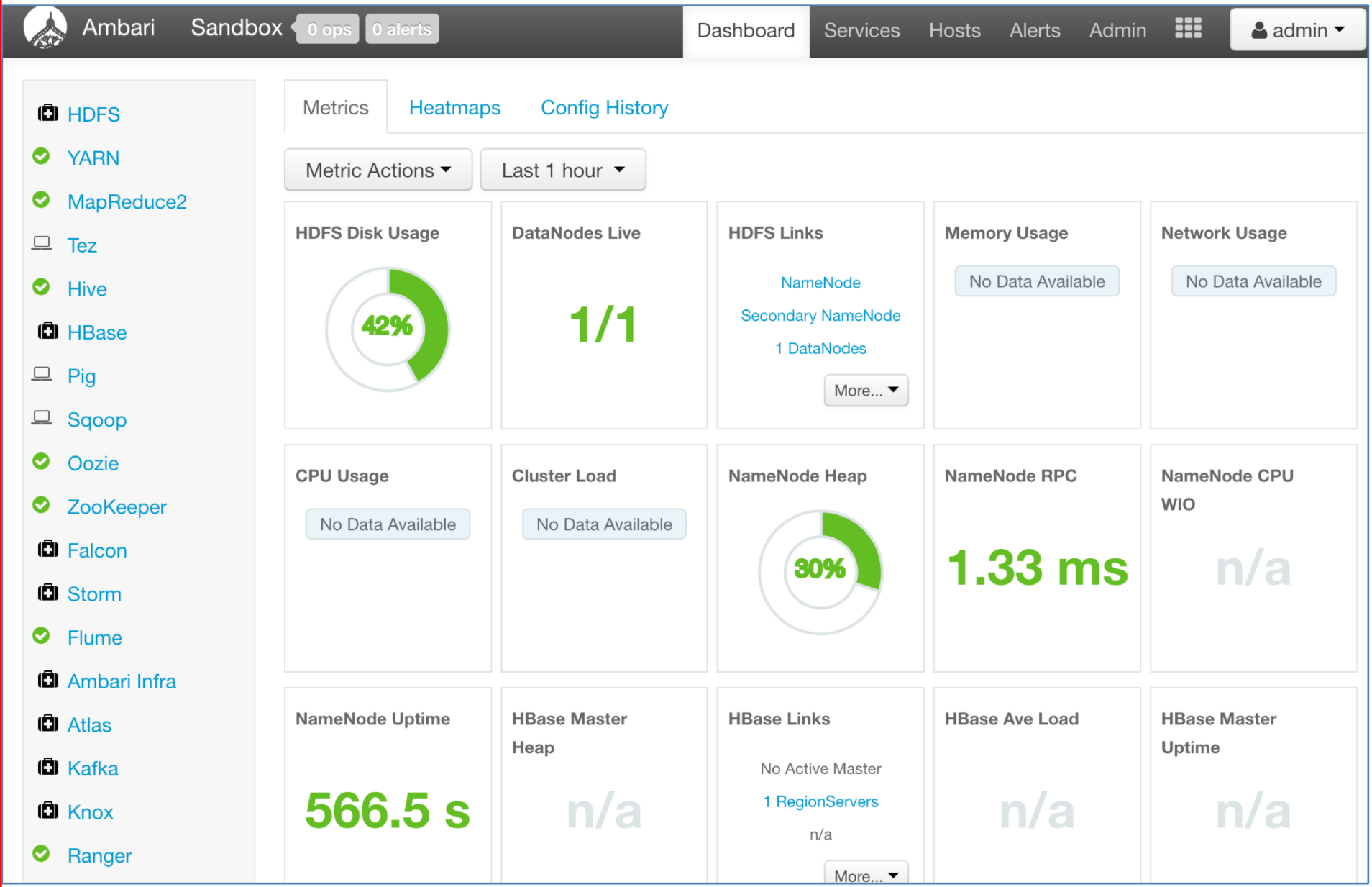


Ambari

A completely open framework for provisioning, managing and monitoring Apache Hadoop clusters

- Offers an intuitive collection of tools and APIs that mask the complexity of Hadoop, simplifying the operation of clusters no matter the size of the Hadoop cluster.
- Features:
  - Wizard-driven interface
  - API-driven installations
  - Granular service control
  - Configuration change history
  - Extensible framework
  - Customizable user interface
  - User views
  - File Browser for accessing HDFS.
  - Metastore Browser for accessing Hive metadata and HCatalog.
  - Hive Editor for developing and running Hive queries.
  - Pig Editor for submitting Pig scripts.
  - …

Ambari


Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
admin
HDFS
YARN
MapReduce2
Tez
Hive
HBase
Pig
Sqoop
Oozie
ZooKeeper
Falcon
Storm
Flume
Ambari Infra
Atlas
Kafka
Knox
Ranger
Metrics
Heatmaps
Config History
Metric Actions
Last 1 hour
HDFS Disk Usage
42%
DataNodes Live
1/1
HDFS Links
NameNode
Secondary NameNode
1 DataNodes
More...
Memory Usage
No Data Available
Network Usage
No Data Available
CPU Usage
No Data Available
Cluster Load
No Data Available
NameNode Heap
30%
NameNode RPC
1.33 ms
NameNode CPU WIO
n/a
NameNode Uptime
566.5 s
HBase Master Heap
n/a
HBase Links
No Active Master
1 RegionServers
n/a
More...
HBase Ave Load
n/a
HBase Master Uptime
n/a

Access the Linux Shell Web Client

- Start HDP Sandbox in VM and keep it running
- Go to http://127.0.0.1:4200/ to access the shell web client (Or use ssh client via ssh root@127.0.0.1 -p 2222)
- Log in:
  - Username: root
  - Password: hadoop (very 1st time, change to a new password) or the updated password you set

Access Hortonworks Sandbox via Ambari

- Keep the Hortonworks Sandbox running in VM
- Go to http://127.0.0.1:8080/
- You may use the following users and passwords (that came with HDP) to log in

Username	Password	Roles
raj_ops	raj_ops	For infrastructure build and R&D activities
maria_dev	maria_dev	For preparing and getting insight from data
holger_gov	holger_gov	For the management of data elements
amy_ds	amy_ds	For exploratory data analysis, cleanup and transformation

Reset Ambari the admin Password

- Run this command at the prompt (127.0.0.1:4200)

  ambari-admin-password-reset

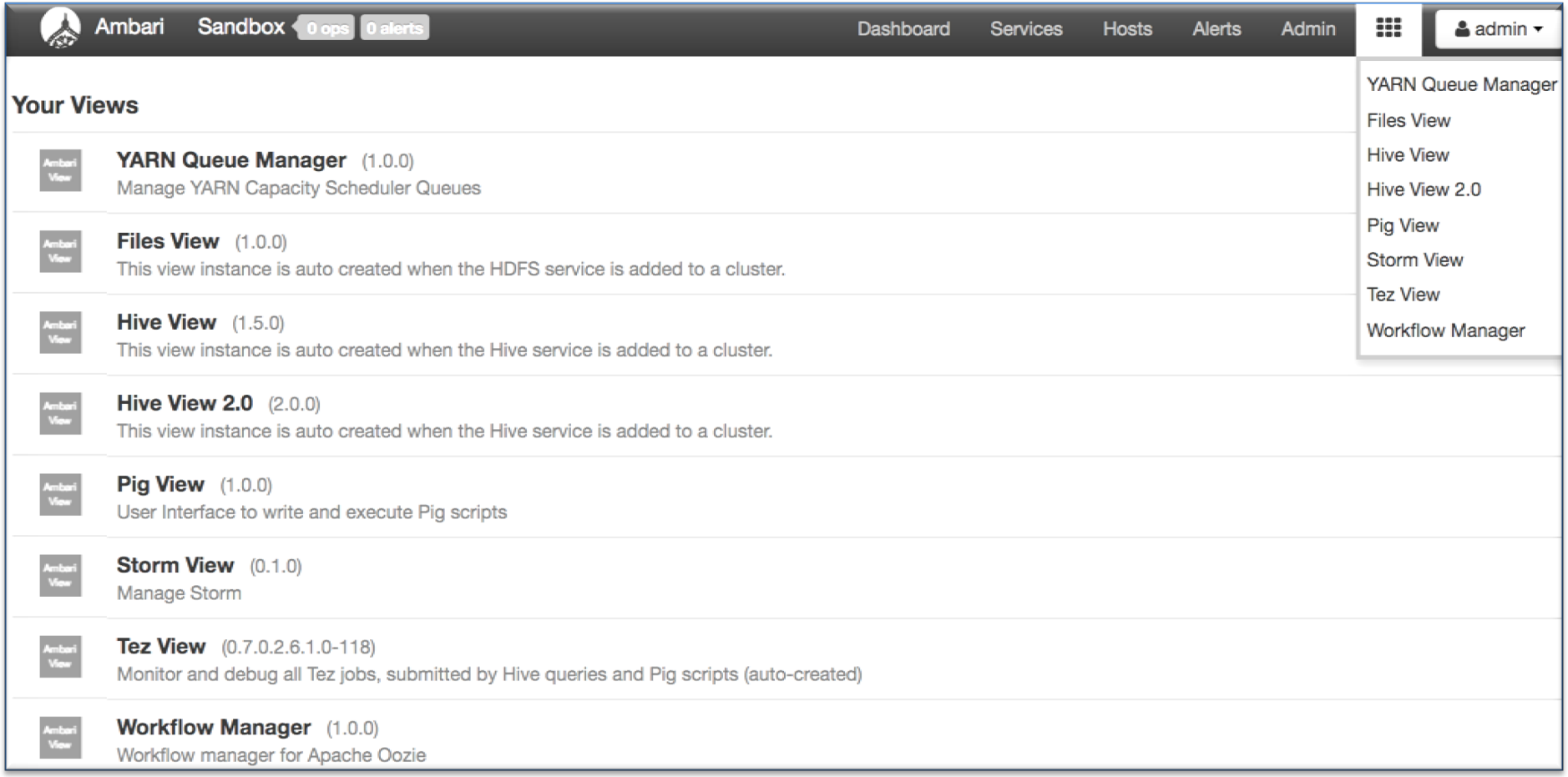
- Give **your new password** for Ambari.

  It may take some time to complete. If your Ambari doesn't restart automatically, restart ambari service with command:

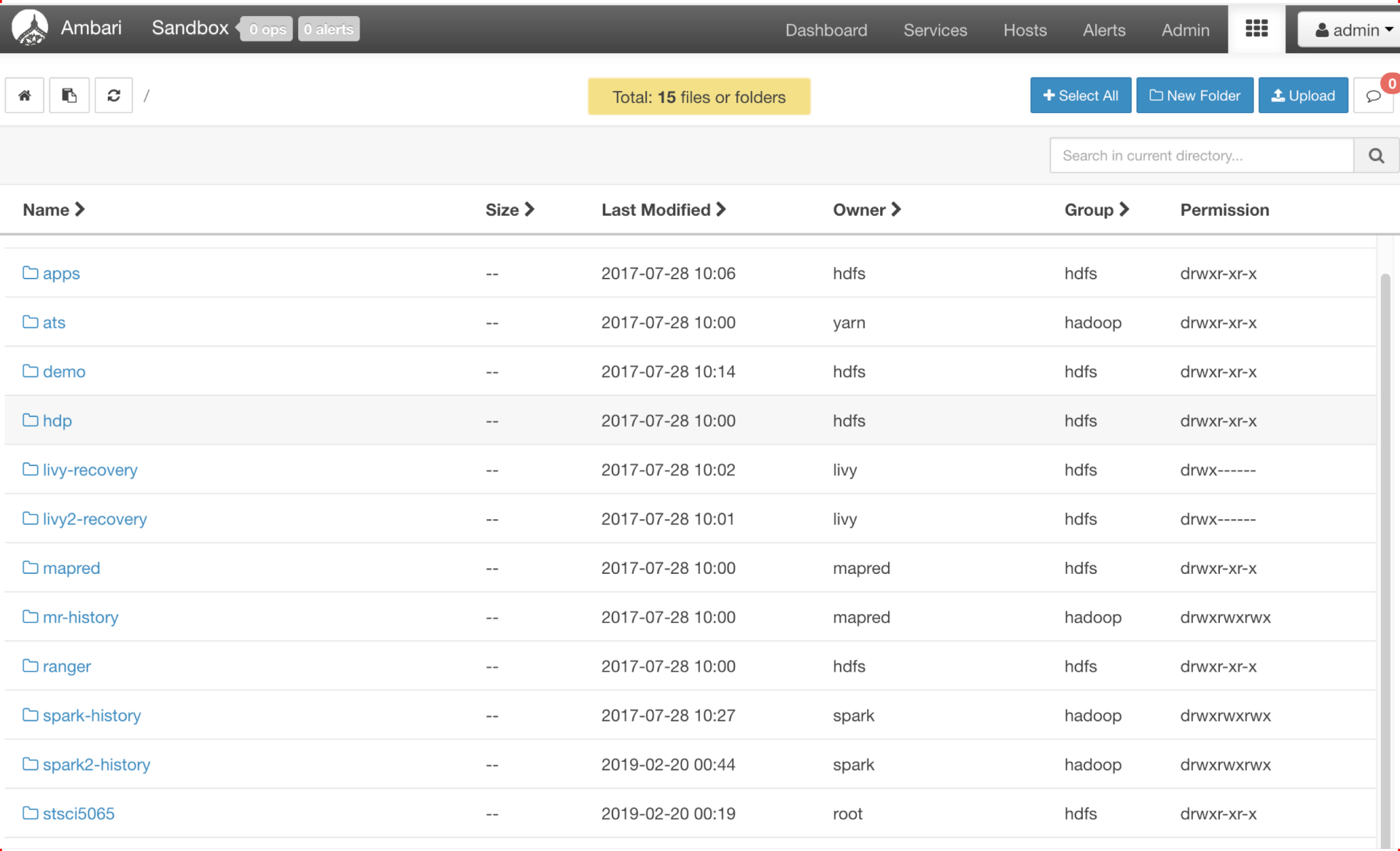
  ambari-agent restart

- Next time you can access Ambari with
  - Username: admin
  - Password: **your new password**

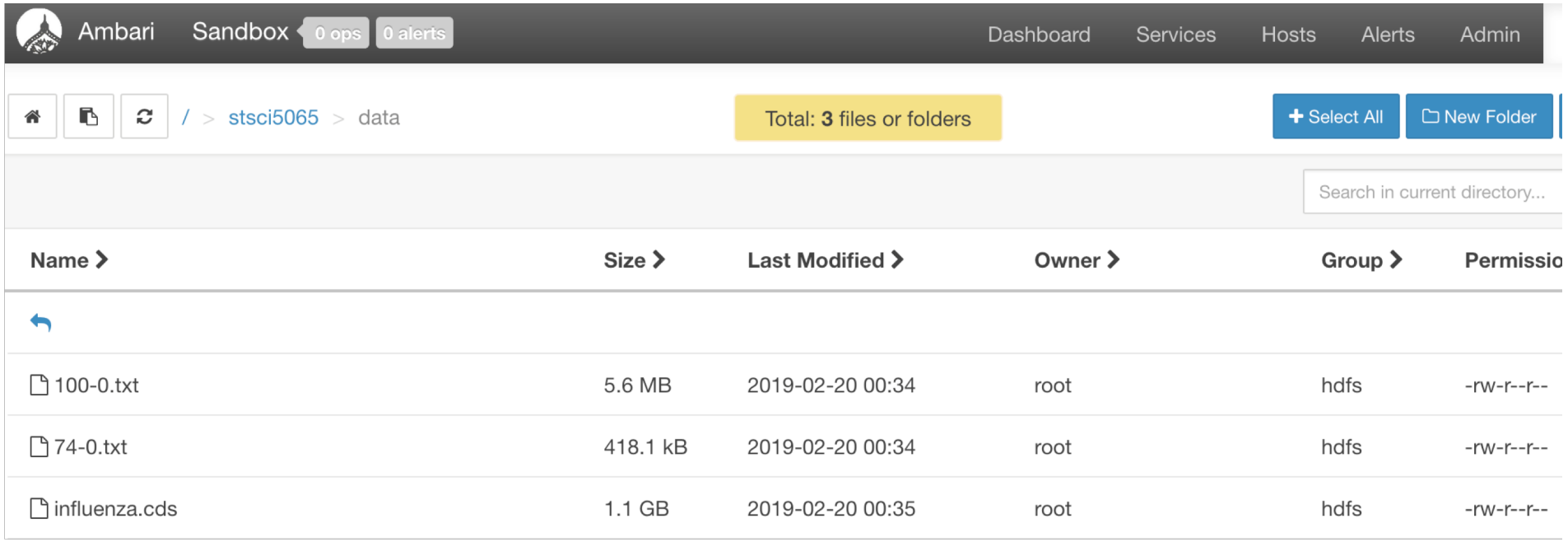
Using Ambari: the User Views



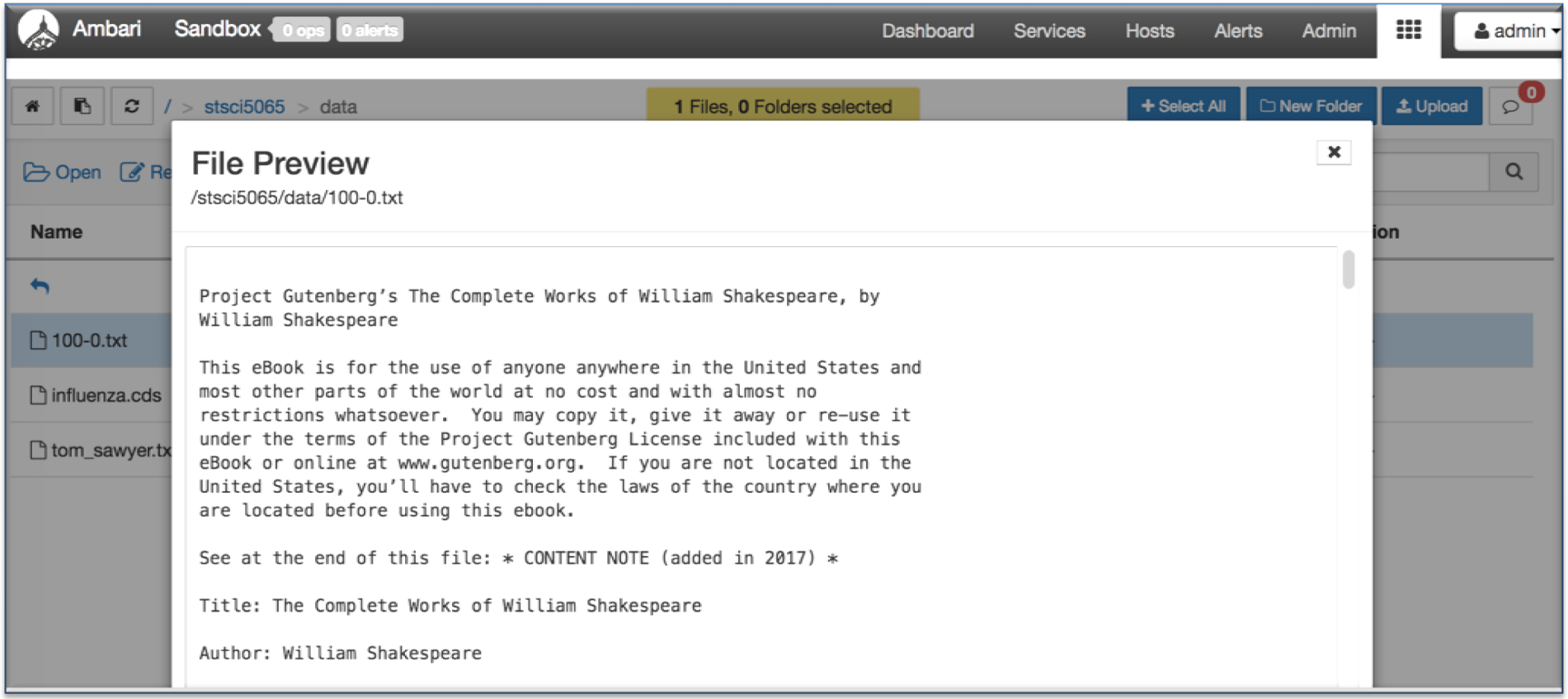
Ambari User Views: Files View



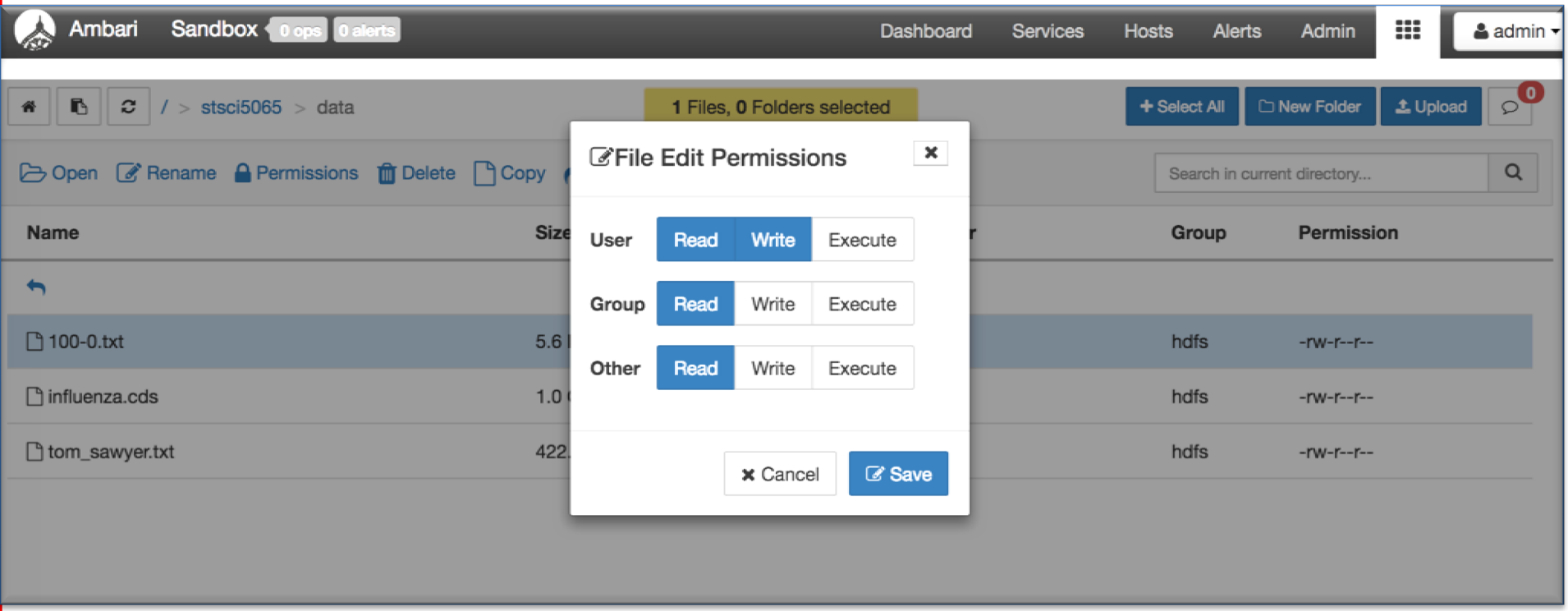
Ambari User Views: Files View, Browse Files



Ambari User Views:
Files View, Click Open to Preview a File

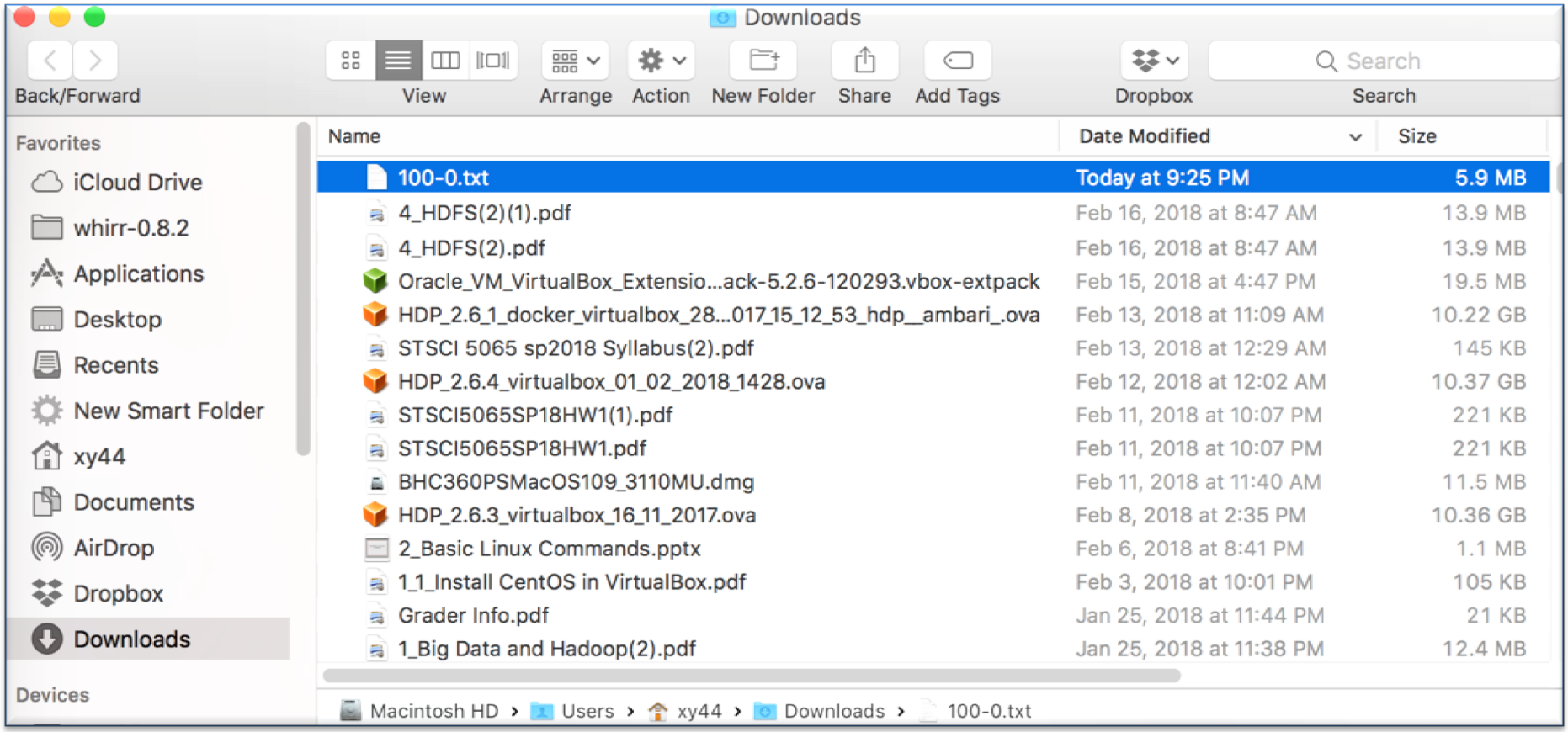


Ambari User Views: Files View, Click Permissions to Change Permissions of a File

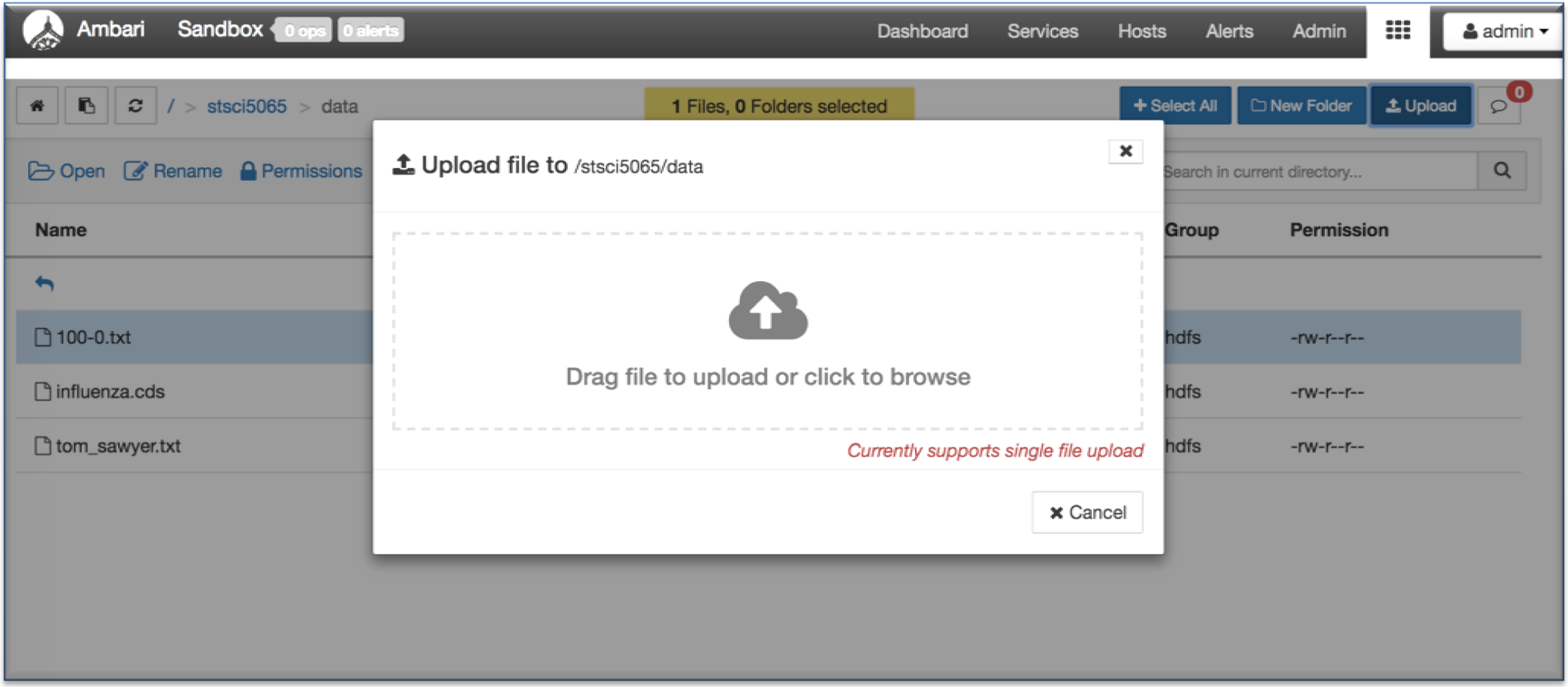


Ambari User Views:

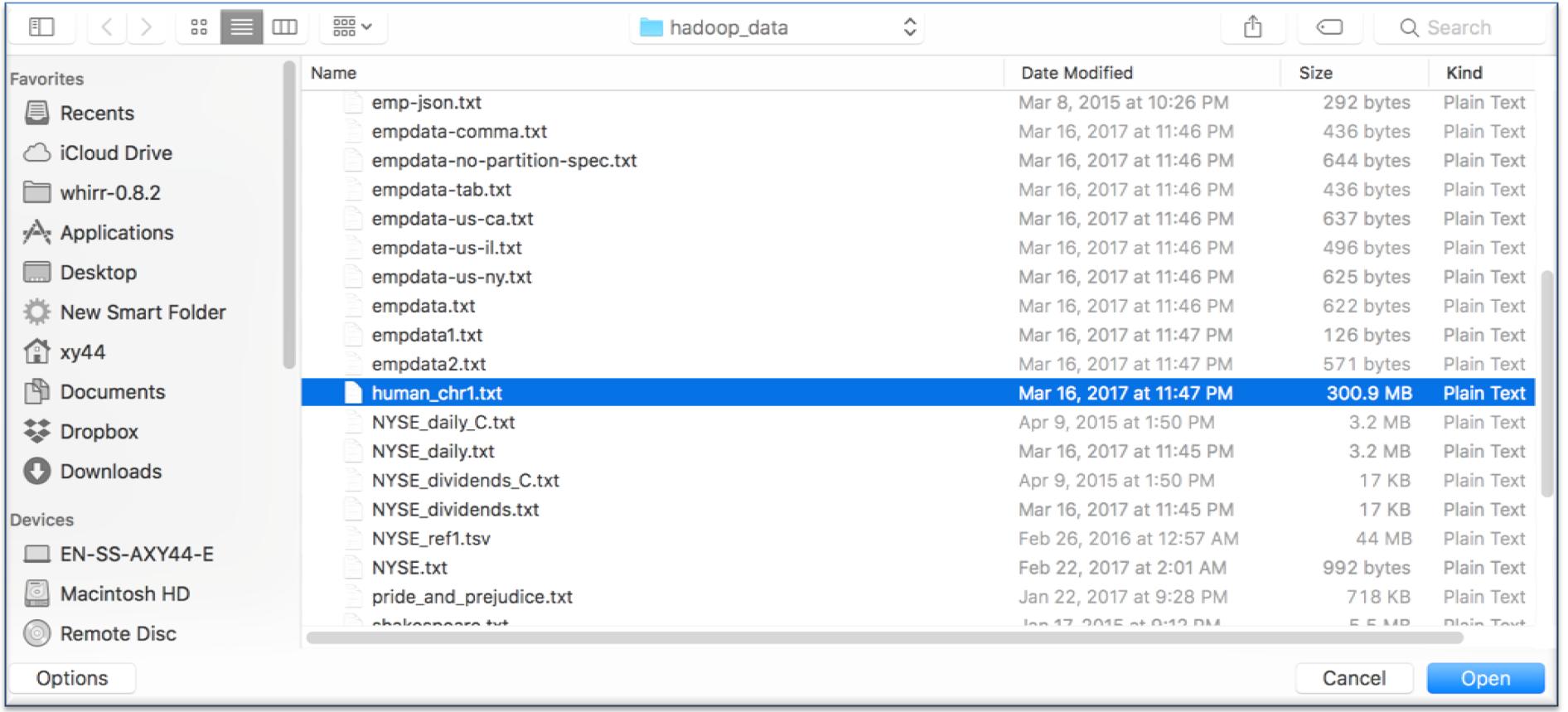
Files View, Click Download to Download a File to Your Local OS (Mac or Windows)



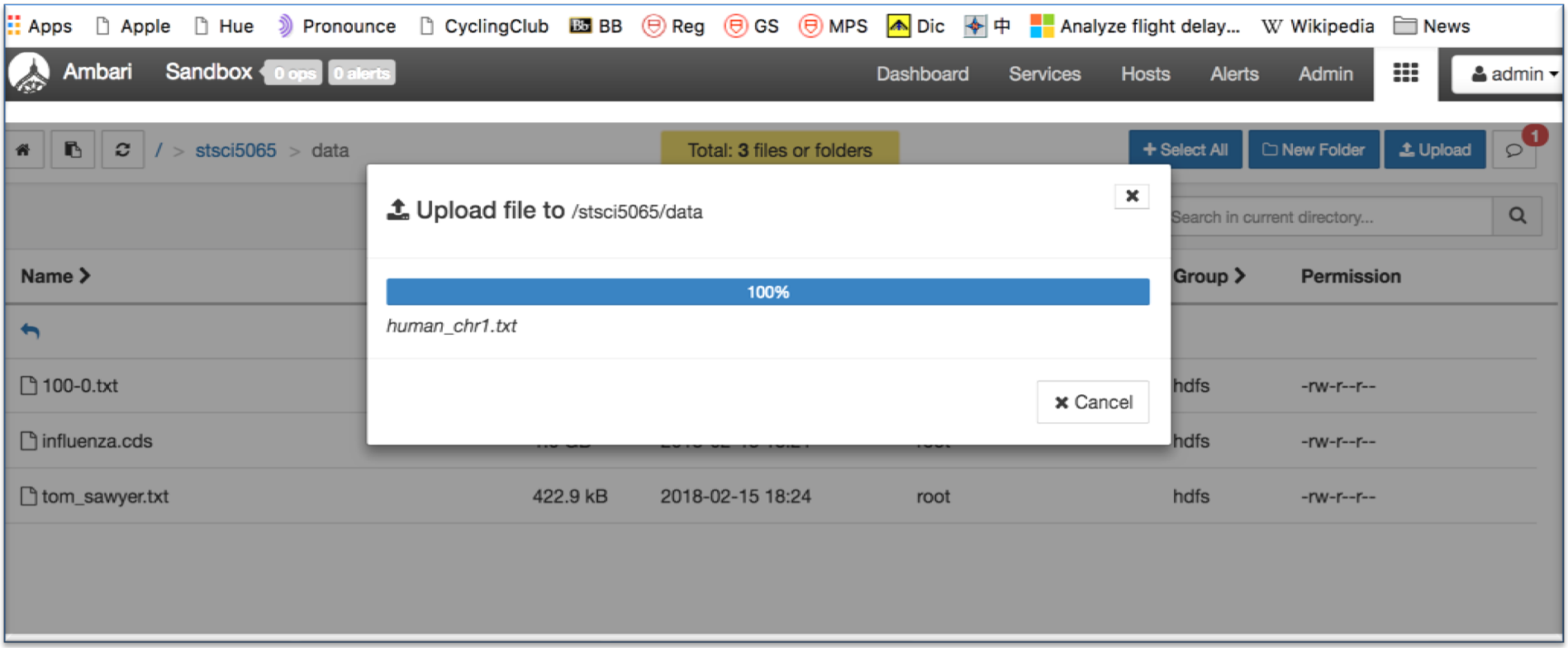
Ambari User Views: Files View, Click Upload to Upload a File from Your Local OS (Mac or Windows) (1)



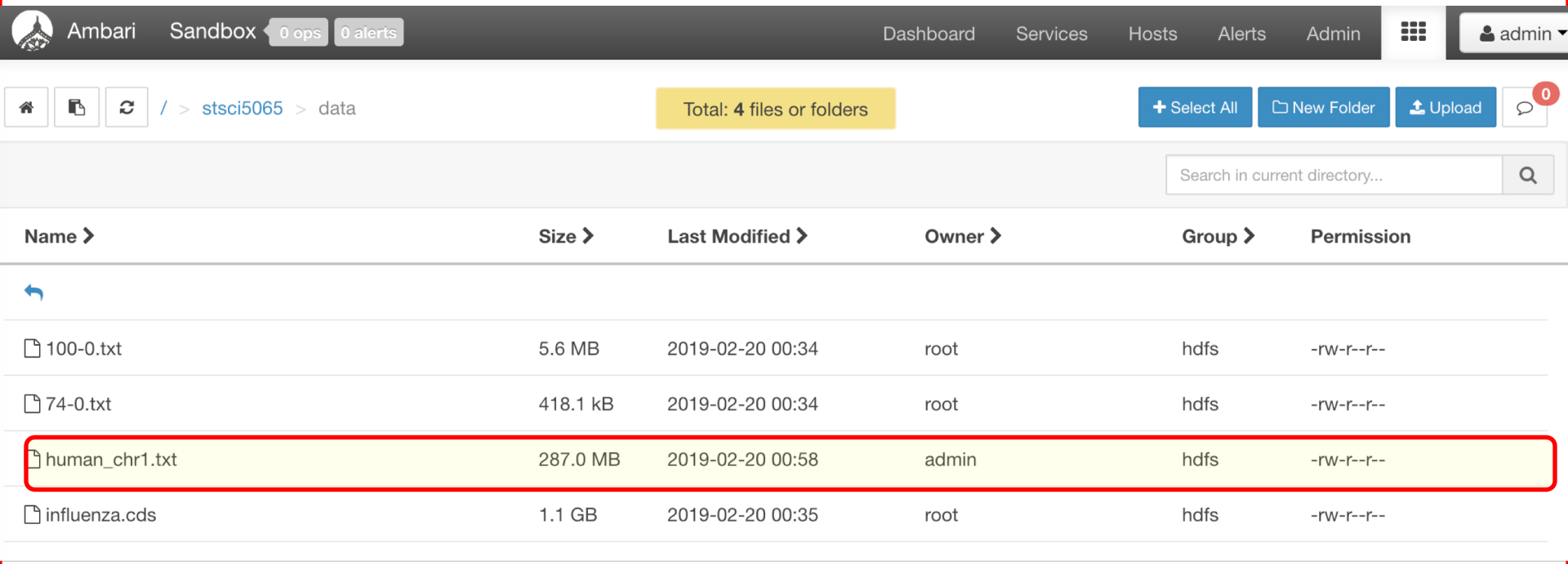
Ambari User Views: Files View, Click Upload to Upload a File from Your Local OS (Mac or Windows) (2)



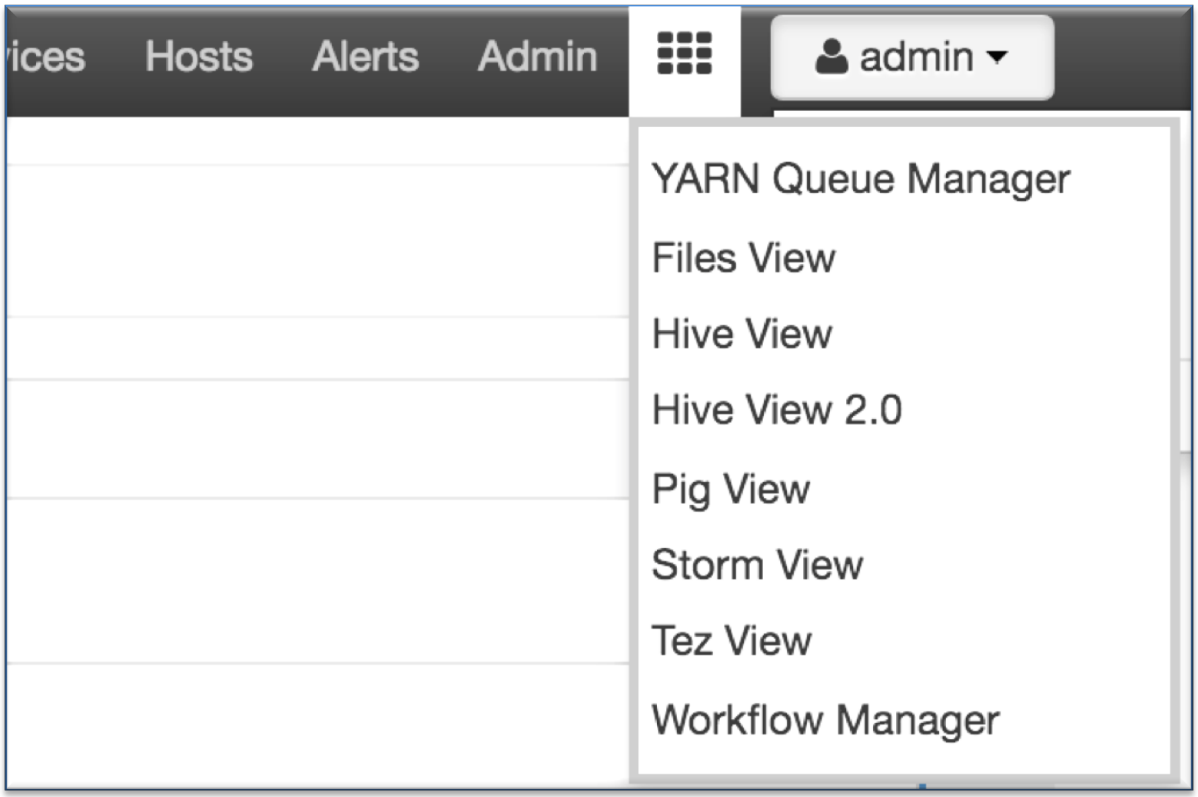
Ambari User Views: Files View, Click Upload to Upload a File from Your Local OS (Mac or Windows) (3)



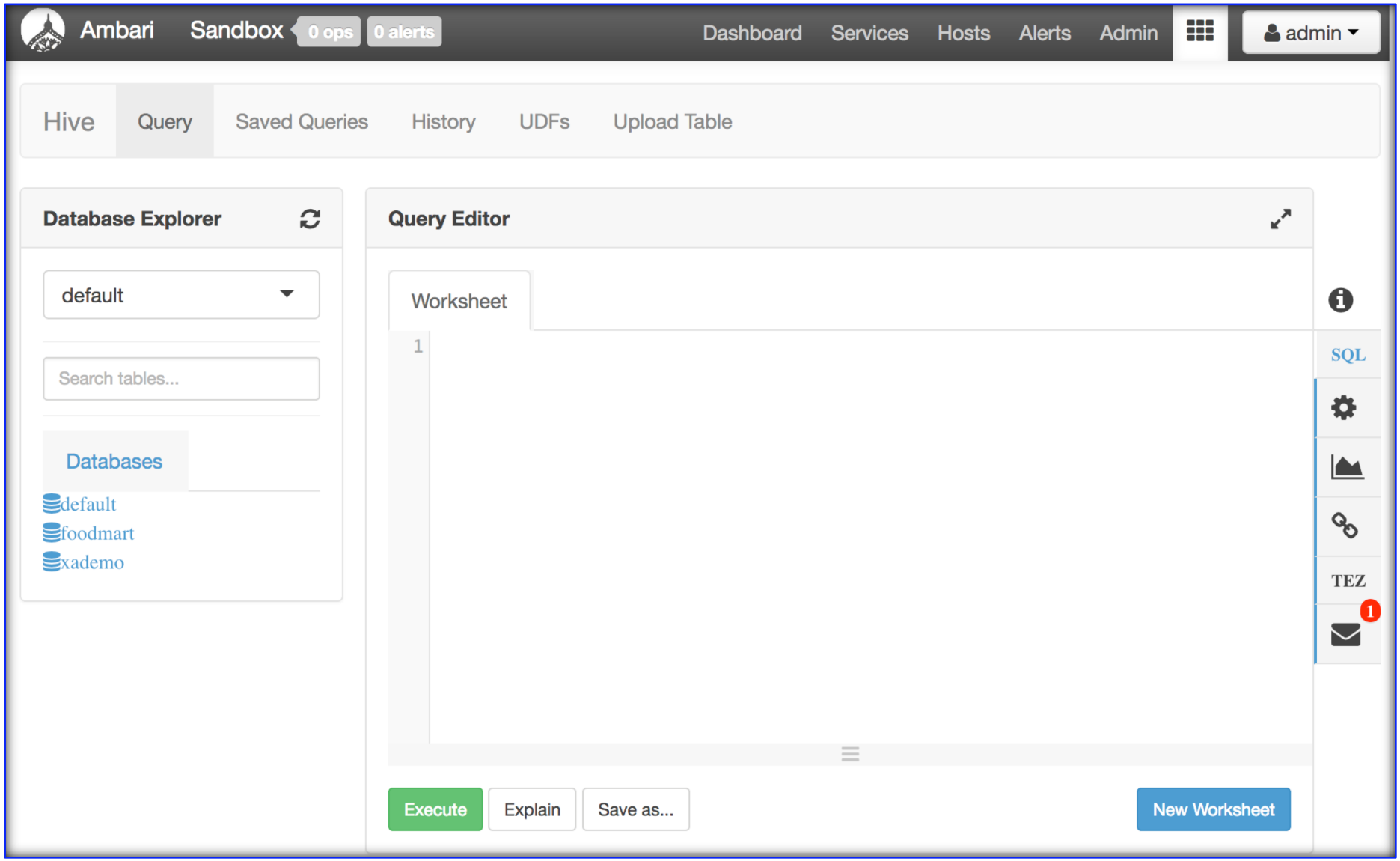
Ambari User Views: Files View, Click Upload to Upload a File from Your Local OS (Mac or Windows) (4)



Ambari User Views: Hive View
There are Two Hive Views

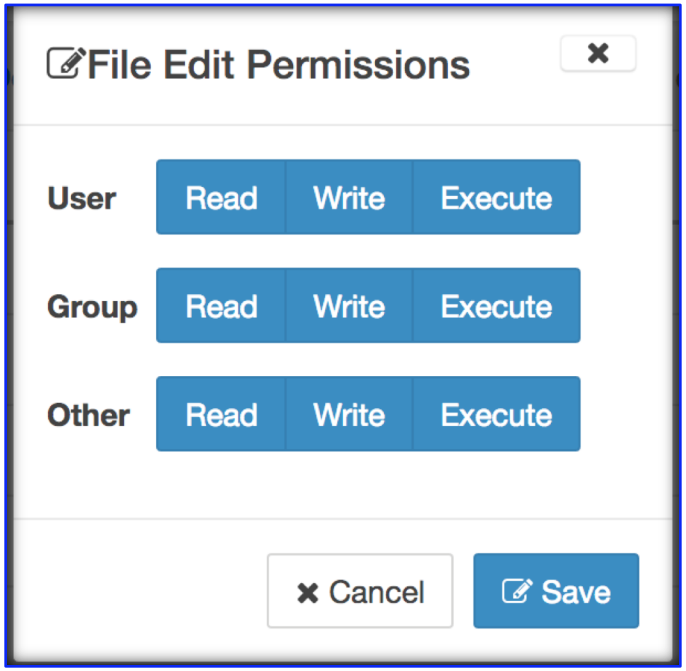


Ambari User Views: Hive View

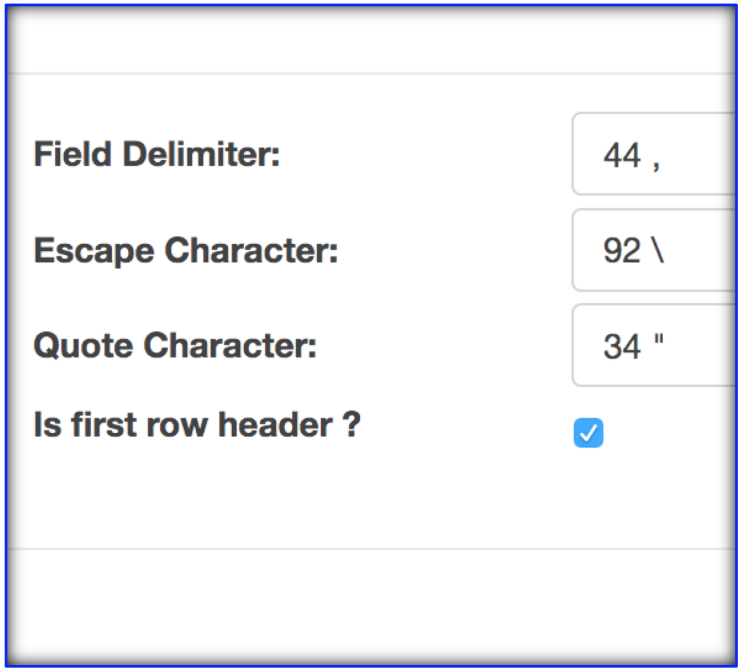


Ambari User Views: Hive View
Prepare to Upload/Create a Table

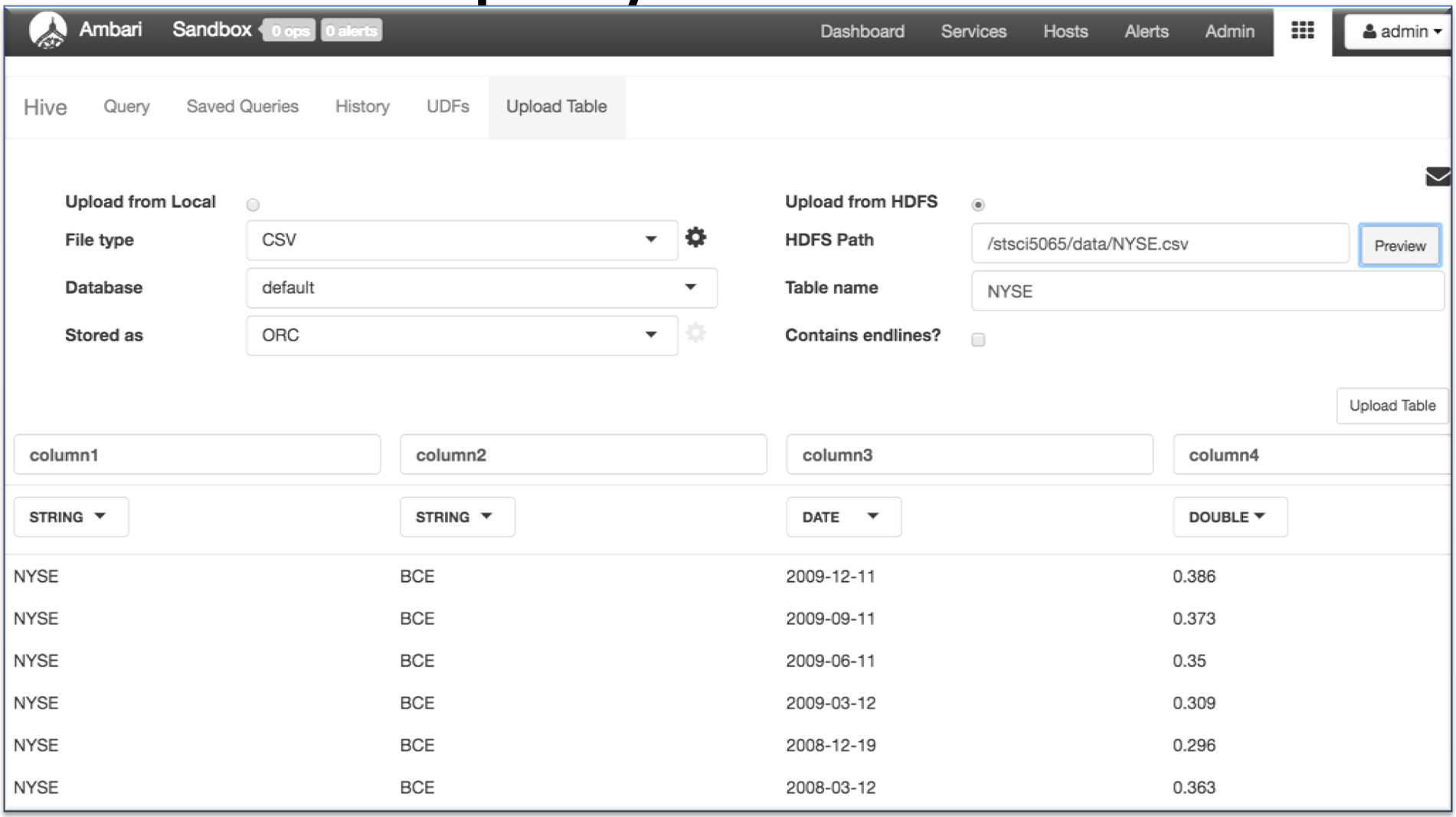
- Change file permissions to 777 (often) in Files View



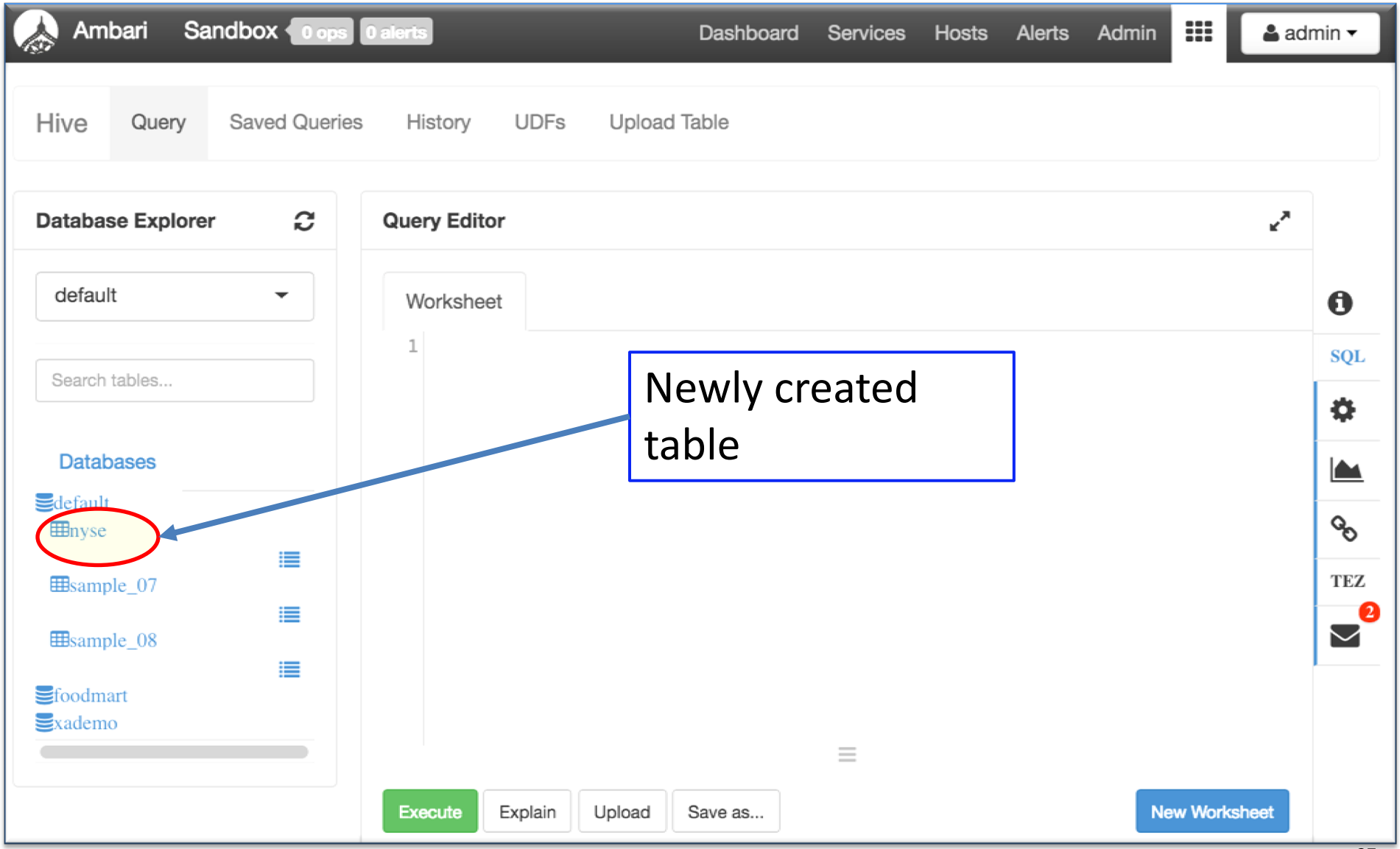
- Set right delimiter and import row header if any



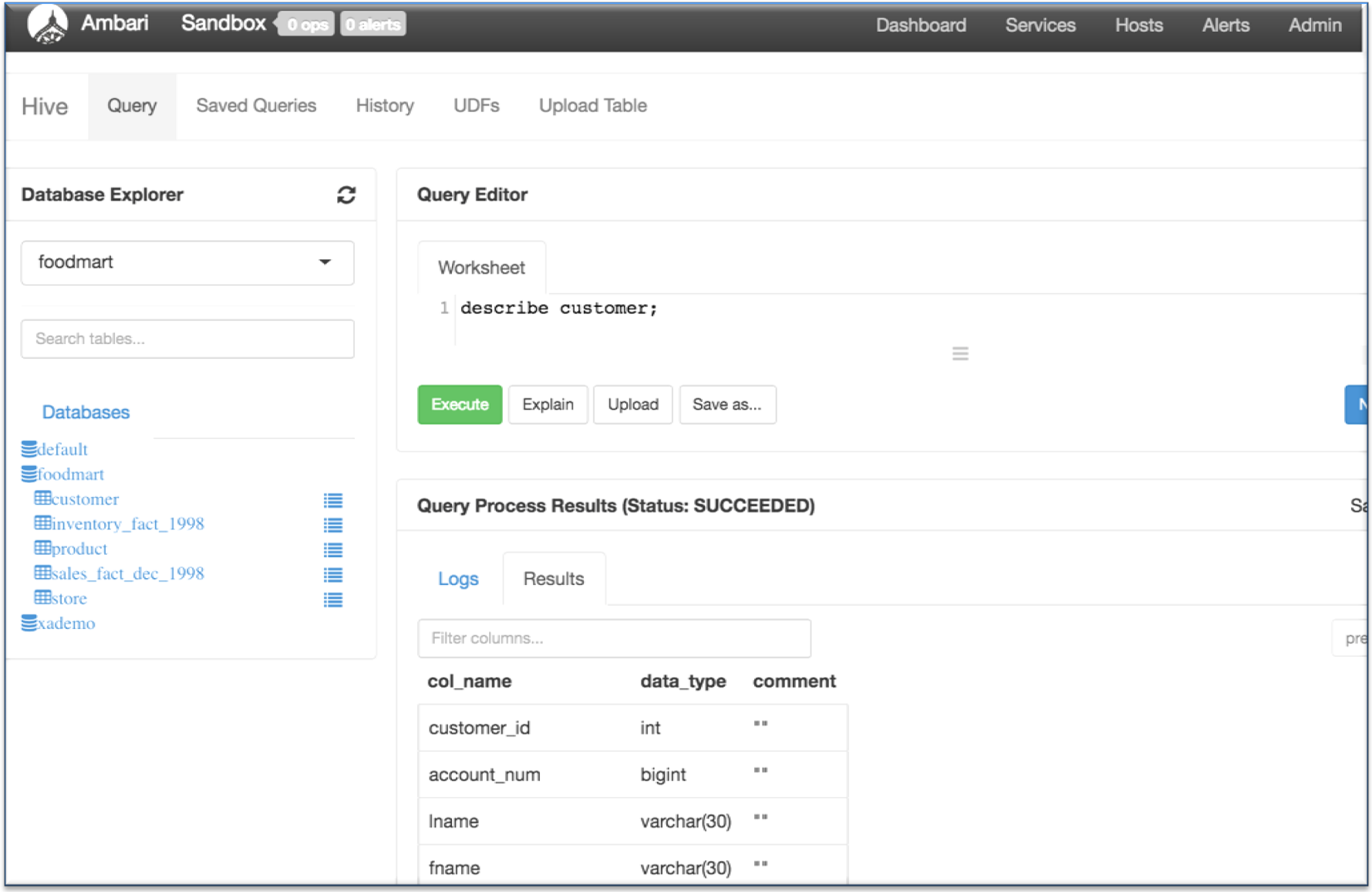
Ambari User Views: Hive View Upload/Create a Table



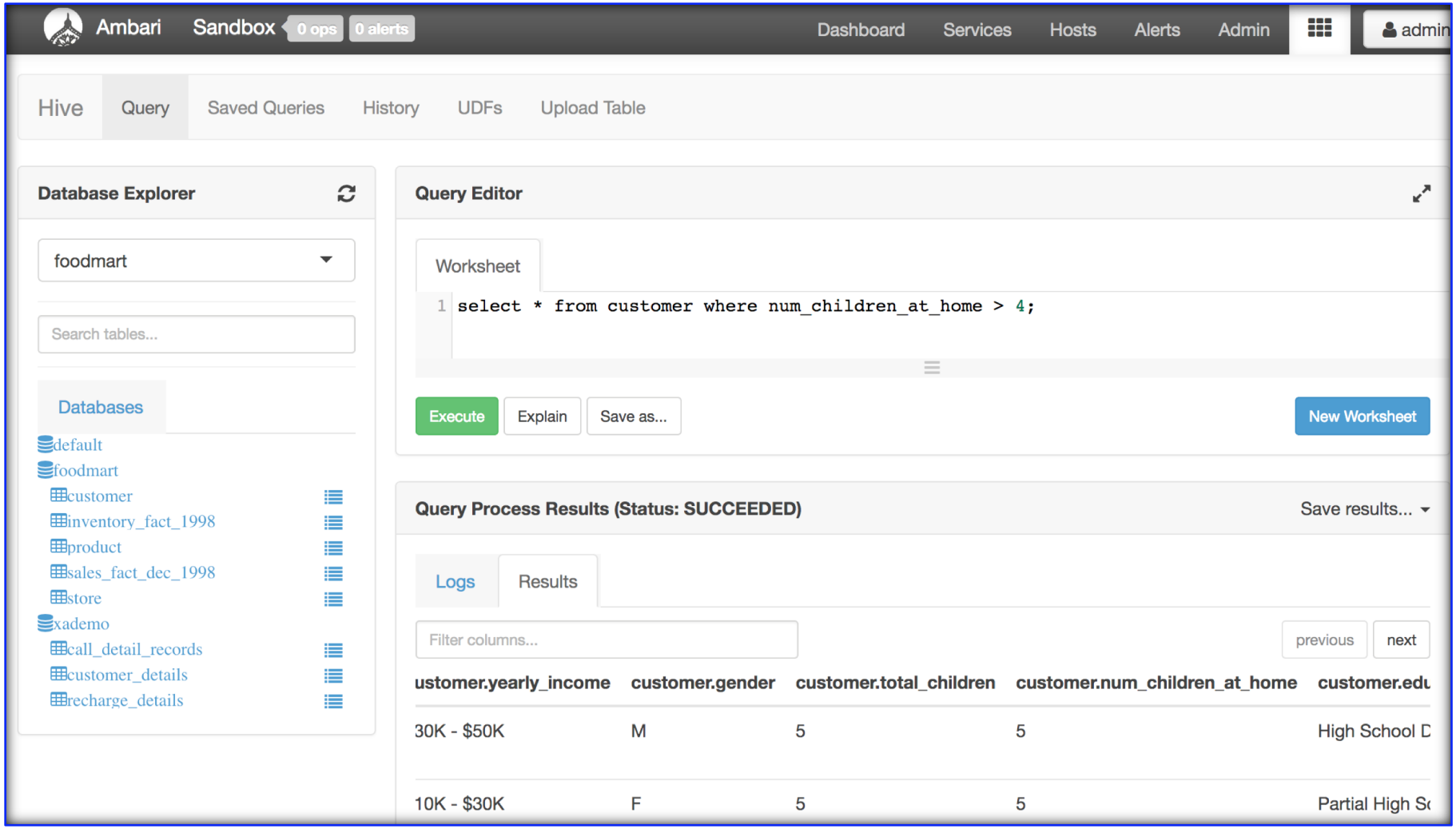
Ambari User Views: Hive View Upload/Create a Table



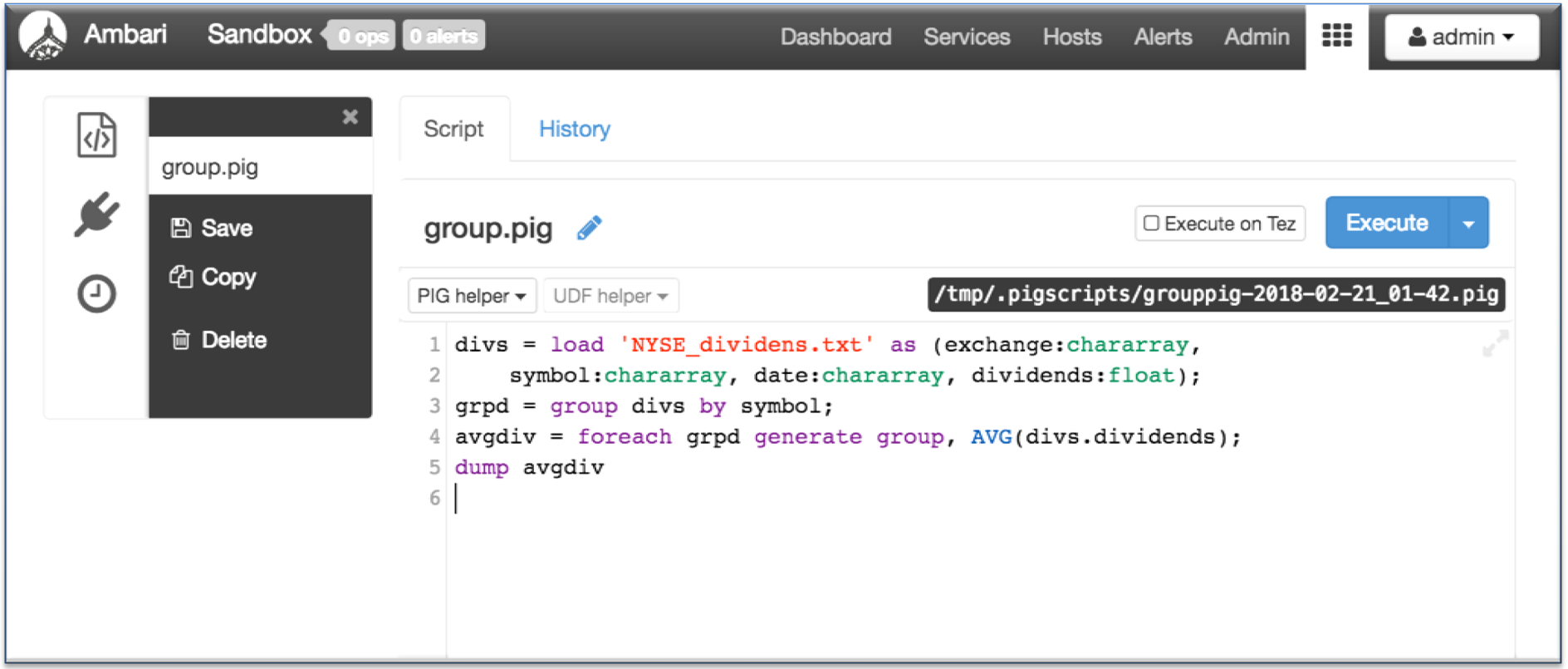
Ambari User Views: Hive View
Get a Description of a Table



Ambari User Views: Hive View
An SQL Query



Ambari User Views: Pig View



Ambari User Views: Pig View