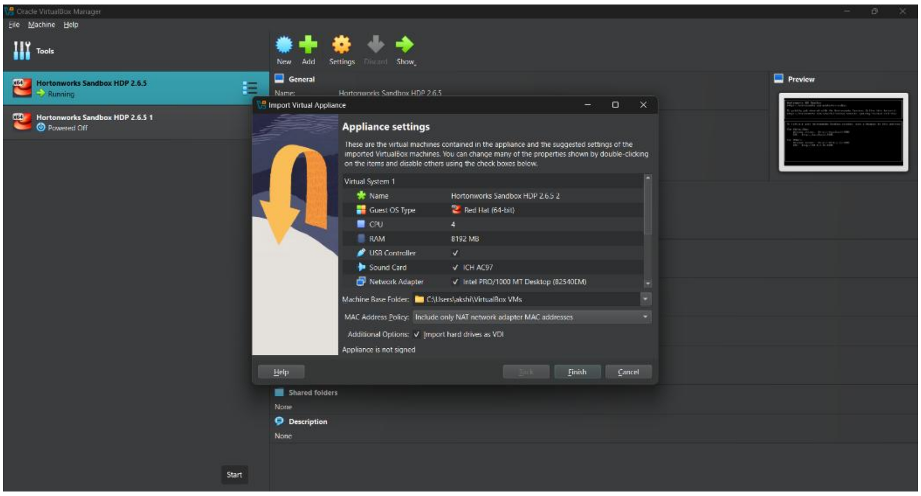
Practical-02

Aim:- Install and configure Hadoop using HDP Sandbox in a VirtualBox, and perform Hive queries on a dataset.

1. Download HDP Sandbox and import into virtualbox.


Oracle VirtualBox Manager
File Machine Help
Tools
New Add Settings Discard Show
Hortonworks Sandbox HDP 2.6.5
Running
Hortonworks Sandbox HDP 2.6.5 1
Powered Off
General
Name: Hortonworks Sandbox HDP 2.6.5
Preview
Shared folders
None
Description
None
Start

Import Virtual Appliance

Appliance settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Virtual System 1

Name	Hortonworks Sandbox HDP 2.6.5 2
Guest OS Type	Red Hat (64-bit)
CPU	4
RAM	8192 MB
USB Controller	✓
Sound Card	✓ ICH AC97
Network Adapter	✓ Intel PRO/1000 MT Desktop (82540EM)

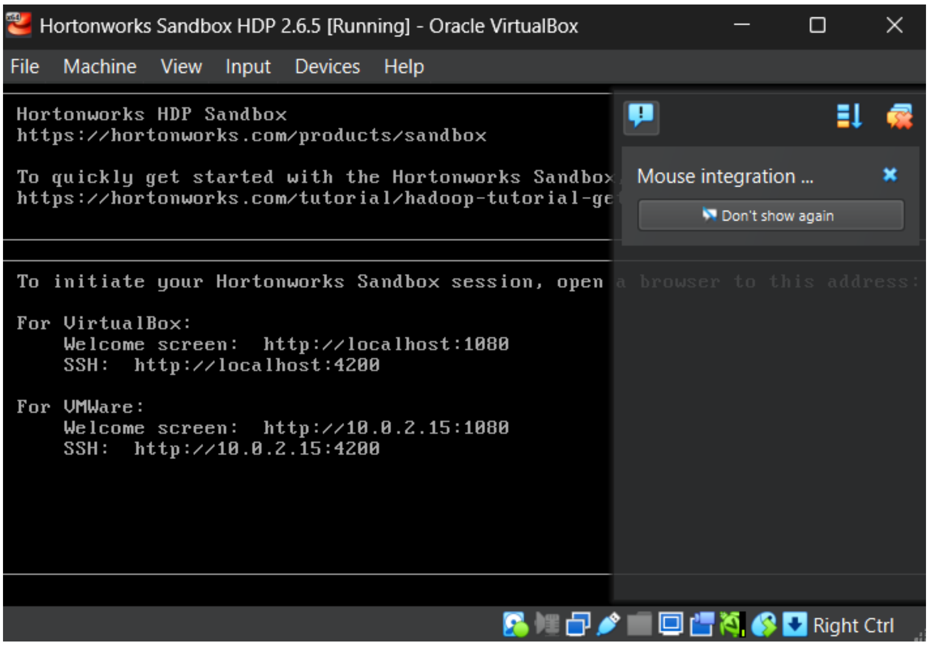
Machine Base Folder: C:\Users\akshi\VirtualBox VMs

MAC Address Policy: Include only NAT network adapter MAC addresses

Additional Options: ✓ Import hard drives as VDI

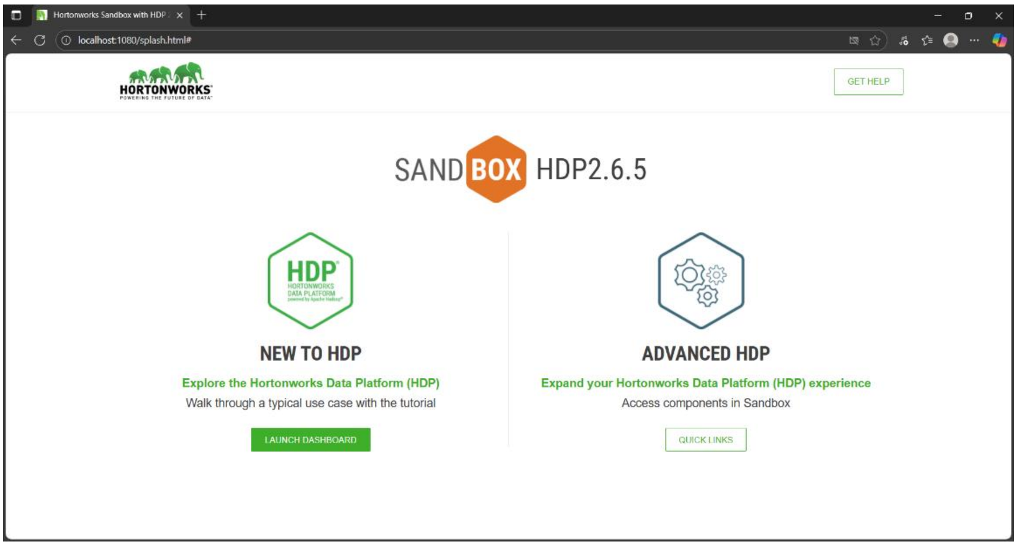
Appliance is not signed

2. Start the Sandbox VM , it will display the URL to access Ambari.


Hortonworks Sandbox HDP 2.6.5 [Running] - Oracle VirtualBox
File Machine View Input Devices Help
Mouse integration ...
Don't show again
Right Ctrl

```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox
https://hortonworks.com/tutorial/hadoop-tutorial-ge

To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
    Welcome screen:  http://localhost:1080
    SSH:  http://localhost:4200

For VMWare:
    Welcome screen:  http://10.0.2.15:1080
    SSH:  http://10.0.2.15:4200
```

3. The URL will open Ambari Dashboard , click on launch dashboard it will open a login page, login using:
   - Username : maria_dev
   - Password : maria_dev

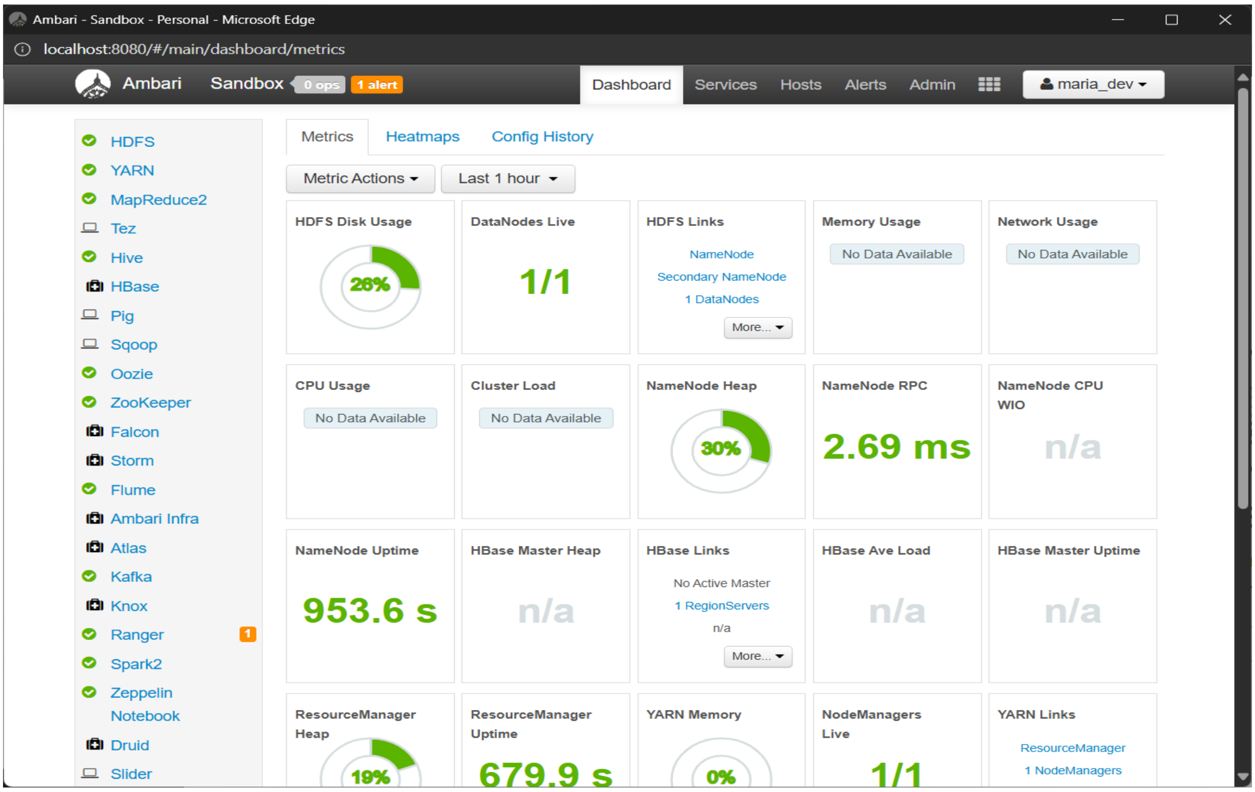


Explore the Hortonworks Data Platform (HDP)

Walk through a typical use case with the tutorial

Expand your Hortonworks Data Platform (HDP) experience

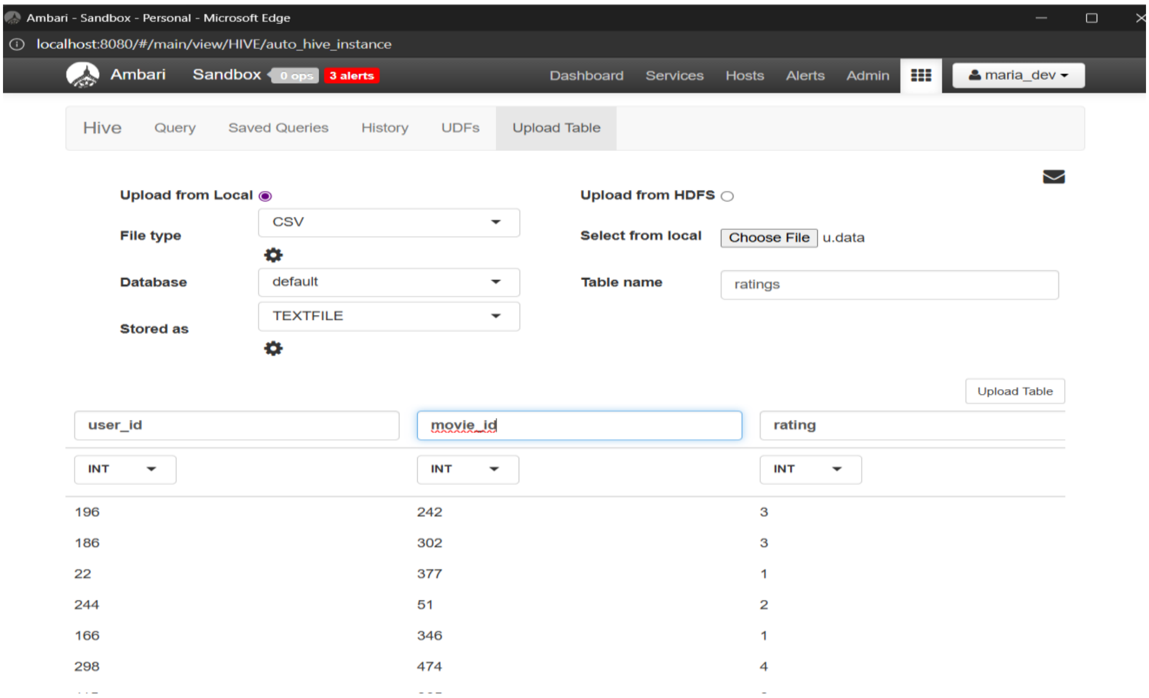
Access components in Sandbox



4. **Small Activity:**
   Download the dataset from grouplens
   link-https://grouplens.org/datasets/movielens/
   older dataset is provided on the website
   Hit the download button & download ml.100k.zip file
   Once you have downloaded extract the data

5. Go into the Ambari tool and from the menu go into the hive view, import the data from the local file there to import data open the hive view

After hive view click on the upload table option
Select the csv file type & set the file delimiter type to the 9 (i.e horizontal tab)
Choose the file from your local system)i.e u.data file
Rename the table name- ratings
column name-
user_id
movie_id
rating
rating_time
Hit the upload button

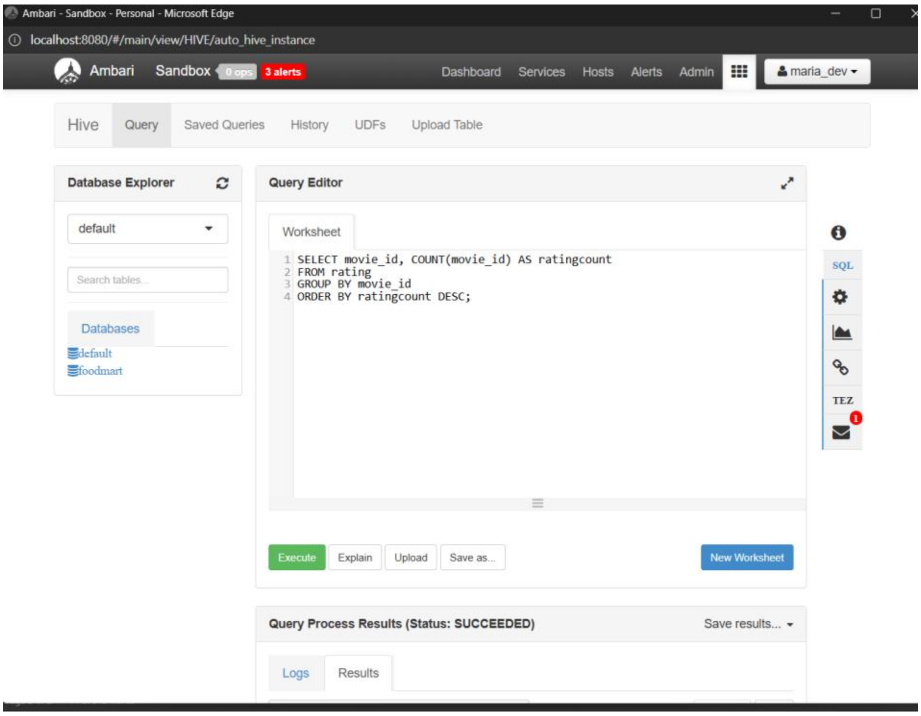


Upload from Local ●
File type: CSV
Database: default
Stored as: TEXTFILE
Upload from HDFS ○
Select from local: u.data
Table name: ratings

user_id	movie_id	rating
INT	INT	INT
196	242	3
186	302	3
22	377	1
244	51	2
166	346	1
298	474	4

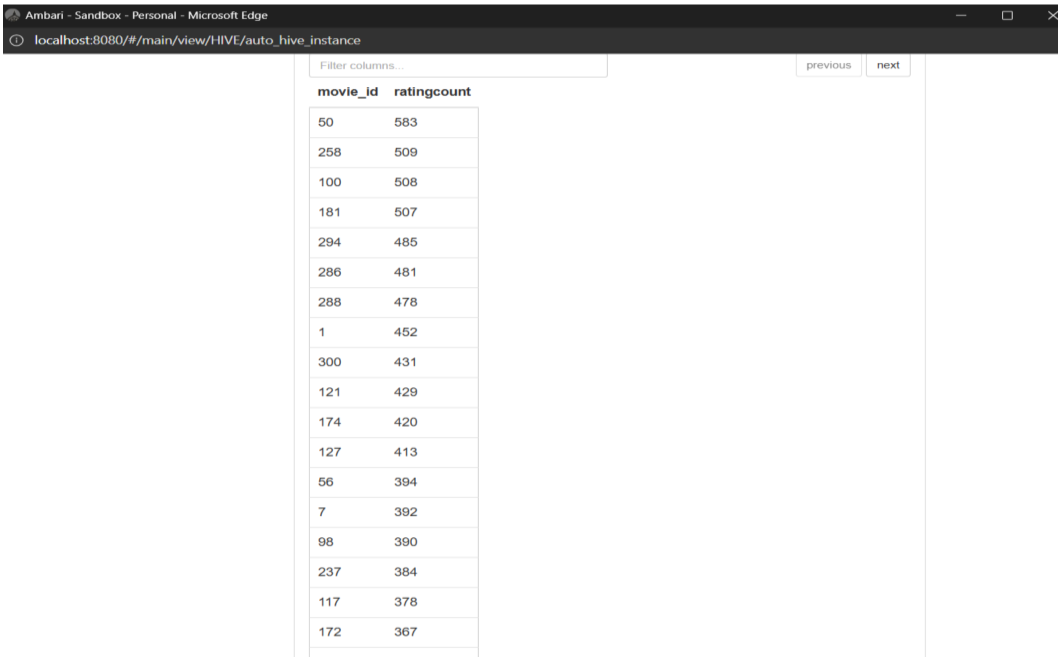
Same for the movie name table
Select the file type as above and set the file delimiter to 124
Rename the table name to movie
column name-
movie_id
name
After this hit the upload button

6. Write the SQL Query to perform operations
SQL Query1-
SELECT movie_id, count(movie_id), AS ratingcount
FROM ratings
GROUP BY movie_id
ORDER BY ratingCount
DESC
After writing the query execute it and see the results



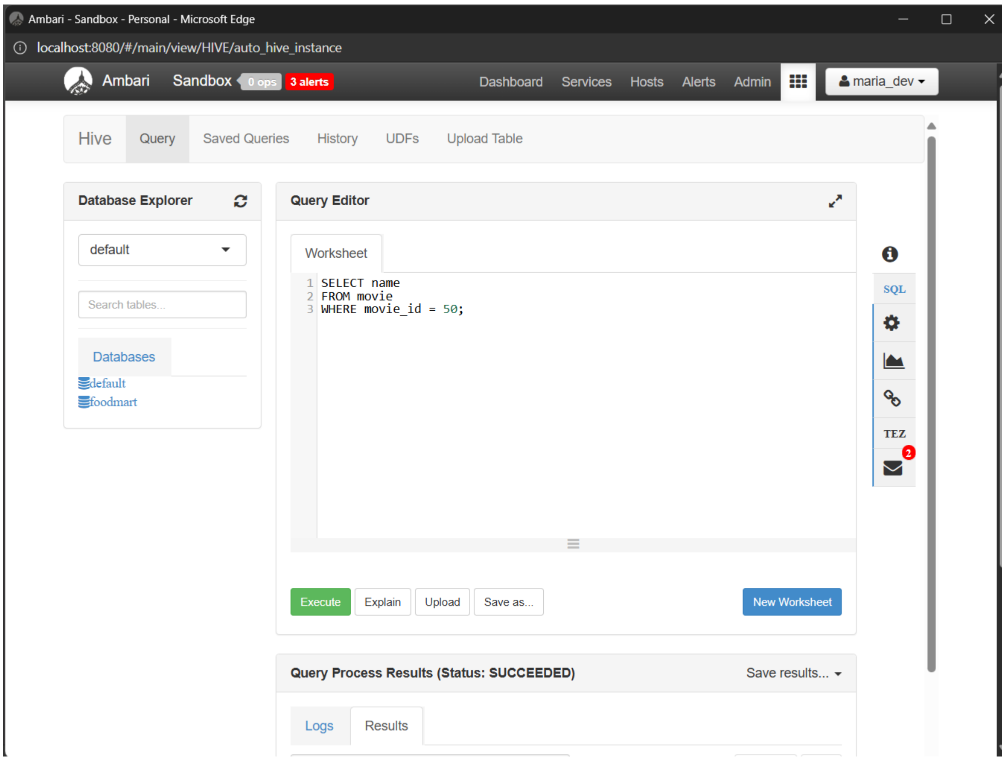
```
SELECT movie_id, COUNT(movie_id) AS ratingcount
FROM rating
GROUP BY movie_id
ORDER BY ratingcount DESC;
```

Query Process Results (Status: SUCCEEDED)

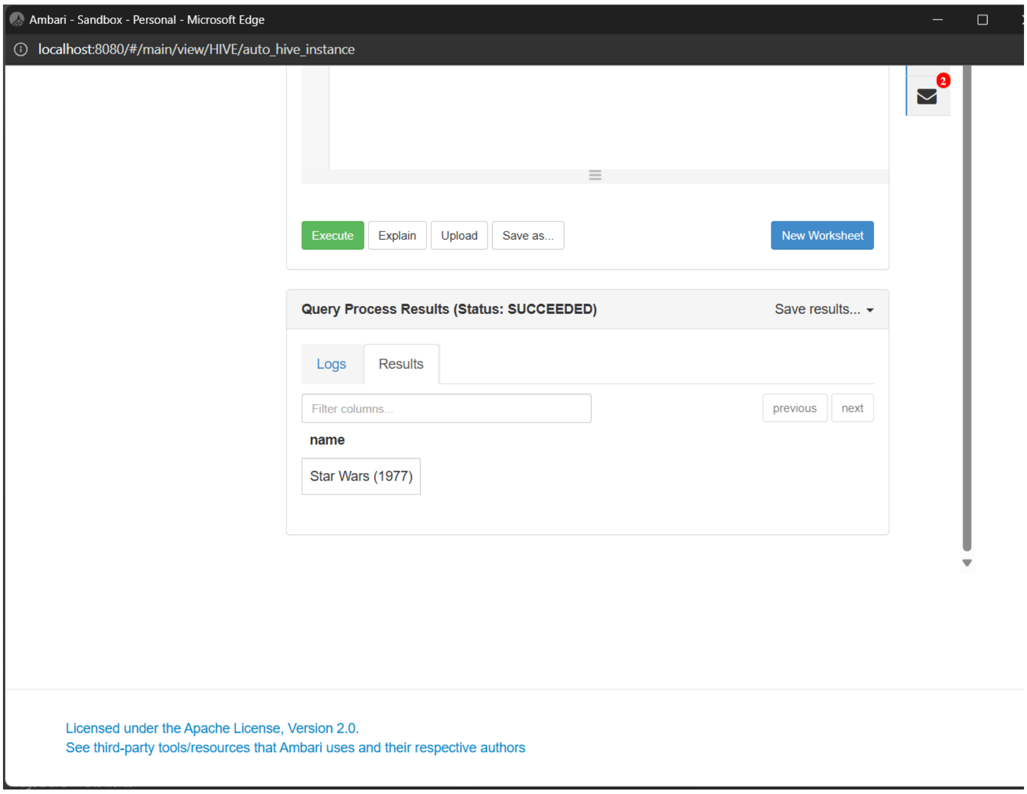


Ambari - Sandbox - Personal - Microsoft Edge
localhost:8080/#/main/view/HIVE/auto_hive_instance
Filter columns...
previous
next

movie_id	ratingcount
50	583
258	509
100	508
181	507
294	485
286	481
288	478
1	452
300	431
121	429
174	420
127	413
56	394
7	392
98	390
237	384
117	378
172	367



```
SELECT name
FROM movie
WHERE movie_id = 50;
```


Ambari - Sandbox - Personal - Microsoft Edge
localhost:8080/#/main/view/HIVE/auto_hive_instance
2
Execute
Explain
Upload
Save as...
New Worksheet
Query Process Results (Status: SUCCEEDED)
Save results...
Logs
Results
Filter columns...
previous
next

name
Star Wars (1977)

Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors