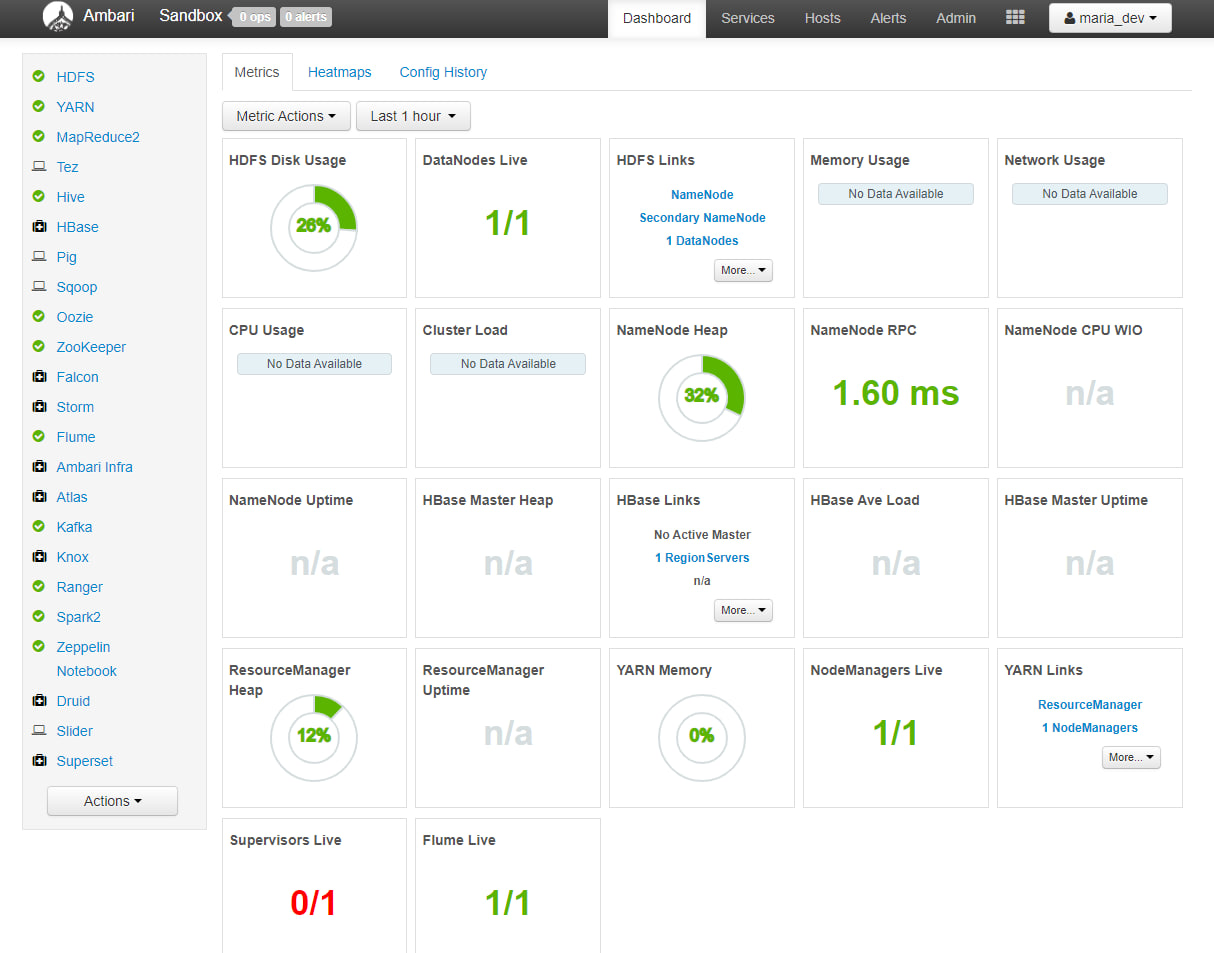
|  |  |
| --- | --- |
| Track: M24-DS-01 | **Week 1** |
| Name: Anastassiya Luzinsan | **Big Data course** |
| Email: luzinsan@mail.ru | **January 2025** |

**Report**

**Task 2**

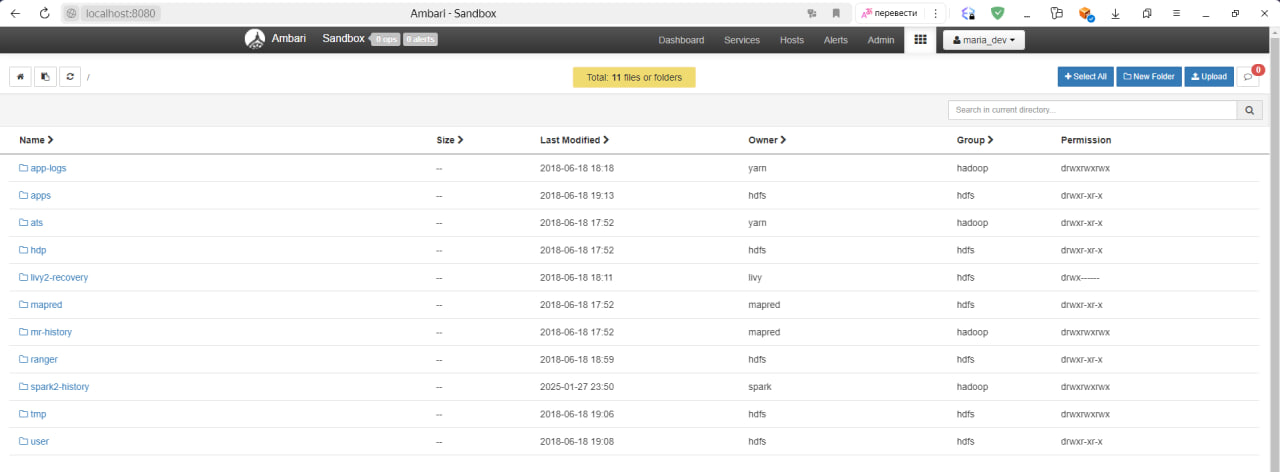
Install HDP on your machine. Take a full screenshot of Ambari dashboard and put it in the report.



|  |
| --- |
| **Your comments (optional)** |
| Firsts I was trying to do it via Arch (Archcraft), but I got “Failed to get D-Bus connection: No such file or directory” to “docker exec -t sandbox-hdp sh -c ‘rm -rf /var/run/postgresql/\*; systemctl restart postgresql-9.6.service;’’” Then I install alongside with existing systems the Ubuntu 24.04 and got same error. Reinstalling on the Ubuntu 20.04 (which don’t support docker-desktop) didn’t have that error. But… port wasn’t available and got “504 bad gateway time-out” even after 20 minutes of waiting. (UPD: Also it was created the docker group (‘sudo groupadd docker’) and added the user to this docker group (‘sudo usermod -aG docker $USER’) and activated (‘newgrp docker’). All these steps are for running docker commands without requiring root (following post-install instruction: https://docs.docker.com/engine/install/linux-postinstall/)) After all, I run Windows 11, install hypervisor - Oracle Virtual Box, import the appliance HDP\_2.6.5\_virtualbox\_180626.ova (HDP 2.6.5 ver.) with standard settings and all works!  Then I accessed the splash webpage of the cluster via localhost:1080, clicked to Quick Links, transferred to Ambari login page (localhost:8080), inputted the login/password as maria\_dev/maria\_dev and was waiting a couple of minutes to load all services. Result (Ambari dashboard) is above. |

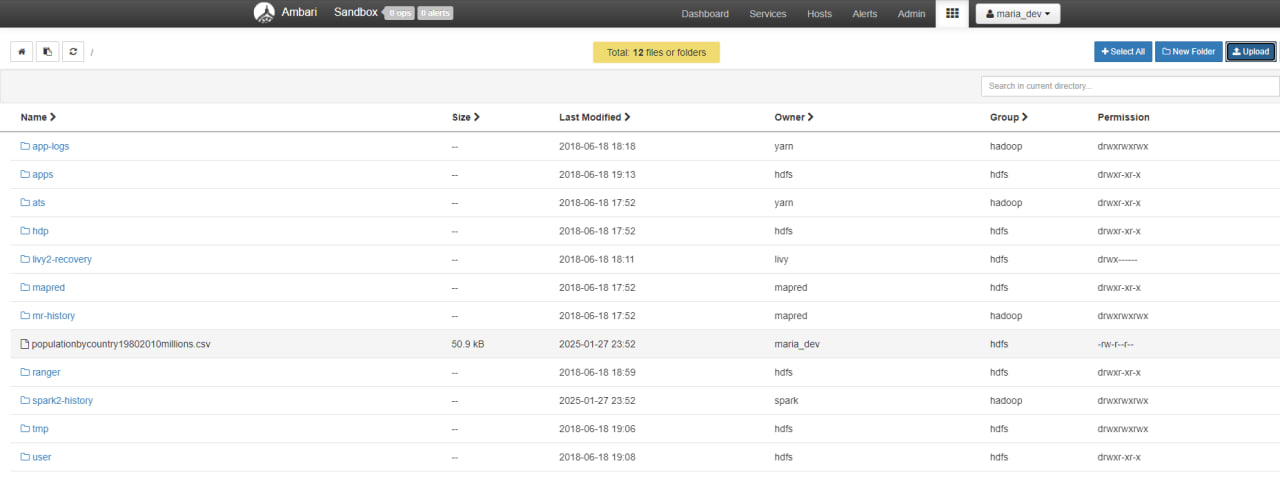
**Task 3**

Access HDFS. Screenshot the content of the root folder in HDFS and attach it to the report.



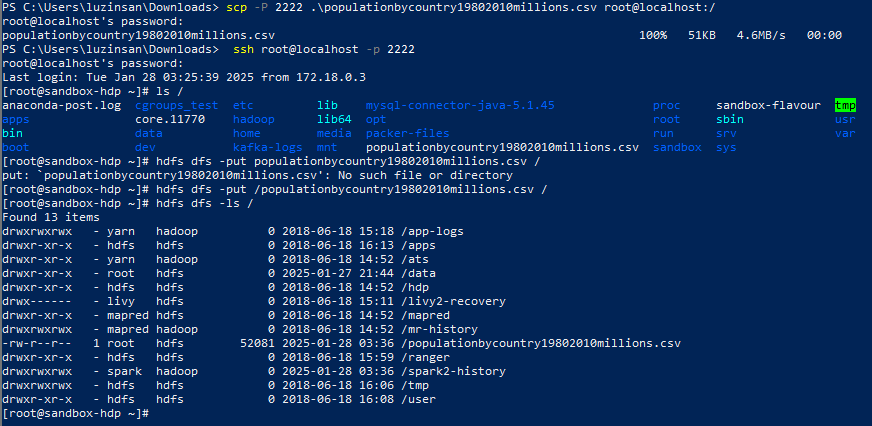
**Task 4**

Download the data file from here. Add it to HDFS root folder (/) via Ambari Files View and take a screenshot of the content of the HDFS root folder and attach it to the report. Delete the file from HDFS.

****

**Task 5**

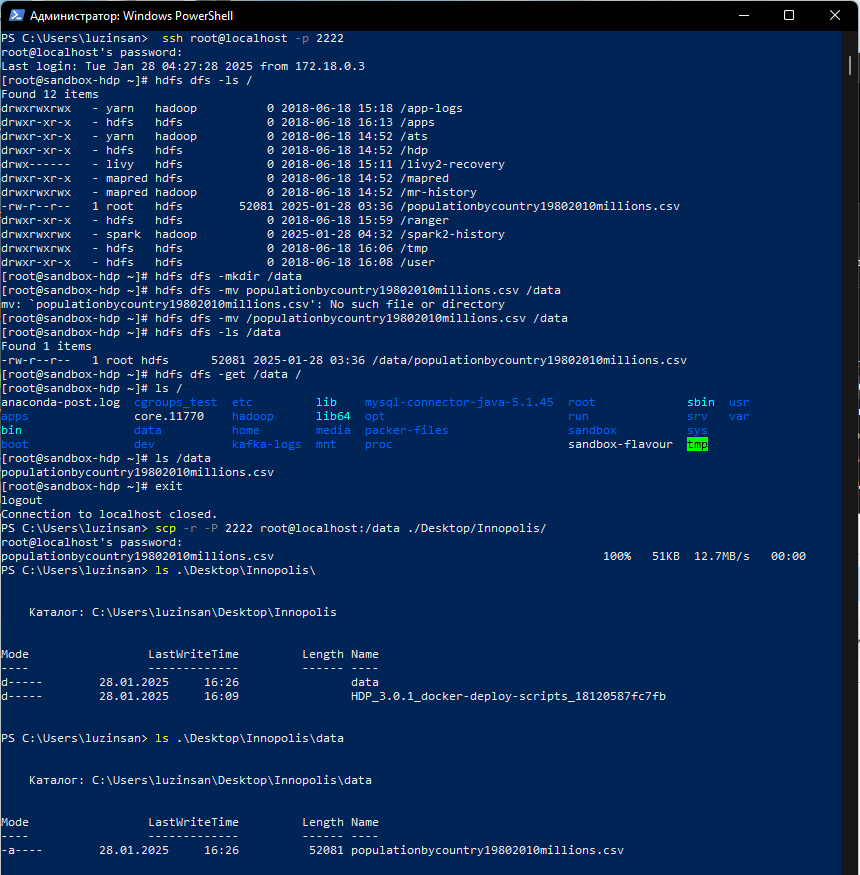
Send the same data file from the local machine (where you boot up) to HDFS root folder (/) via scp or docker cp. Take a screenshot of the command and the output and add it to the report.



|  |
| --- |
| **Your comments (optional)** |
| First, I sent data from the local file system in the host machine (Windows) to local system of the cluster node with port 2222 by using command scp  Second, I moved data from the local file system of the *cluster node* to HDFS by *hdfs dfs* with *-put* option and show the results. |

**Task 6**

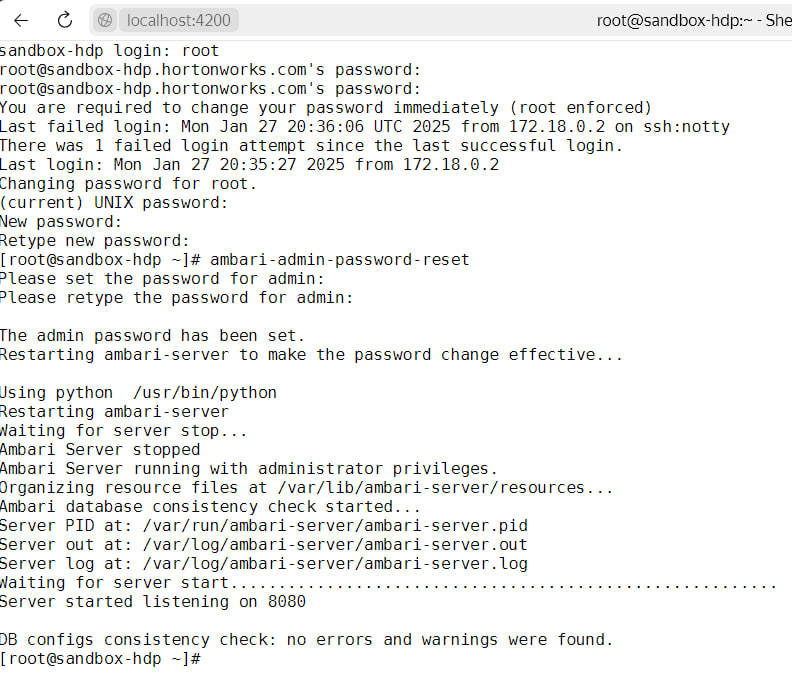
Access the shell of the cluster node via port 2222 in the terminal or 4200 in the browser and create a folder data in HDFS root folder (/). Move the file to the folder data. Send a copy of the folder to the local machine. Take screenshots of the commands and the output and add them to the report.

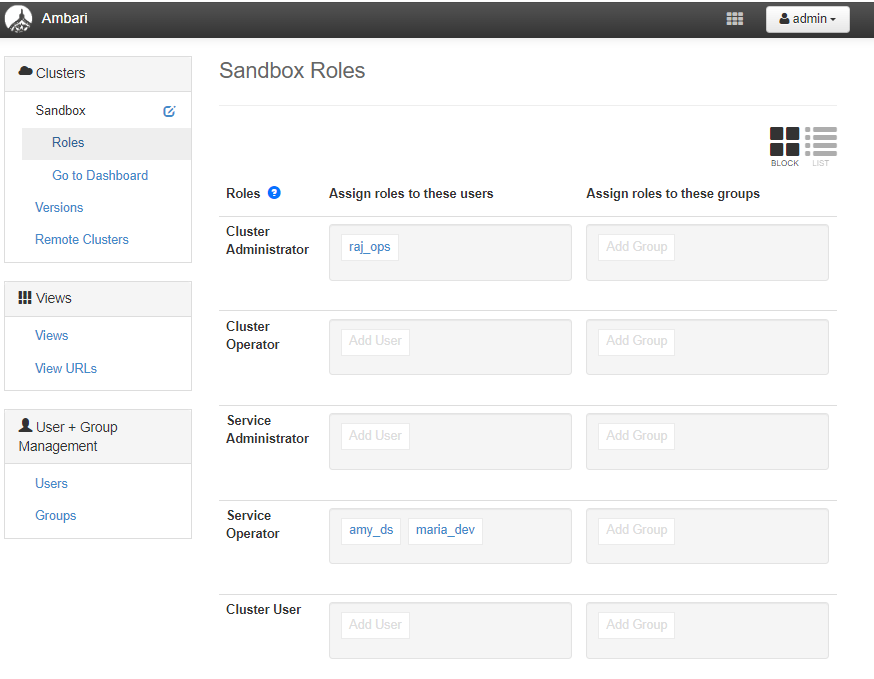


|  |
| --- |
| **Your comments (optional)** |
| After accessing the shell of the cluster I created a directory data in the root via ‘hdfs dfs -mkdir /data’. Then, I moved earlier loaded in the hdfs populationbycountry19802010millions.csv file via ‘hdfs dfs -mv’ command. Then I get the /data directory from hdfs to the root of the cluster node with command ‘hdfs dfs -het /data /’. Further I closed the connection and get the file from cluster node to local machine using the secure scp utility with -r (recursive option). In the end I’ve got the above result. |

**Task 7**

Reset the password of Ambari-Admin and access to Ambari as an admin. Display the roles of the users in the cluster. Screenshot and add it to the report.





**Task 8**

Create a new user with your name and give it cluster manager (admin) role. Access it, screenshot and add to the report. You can check the roles of HDP from here.

