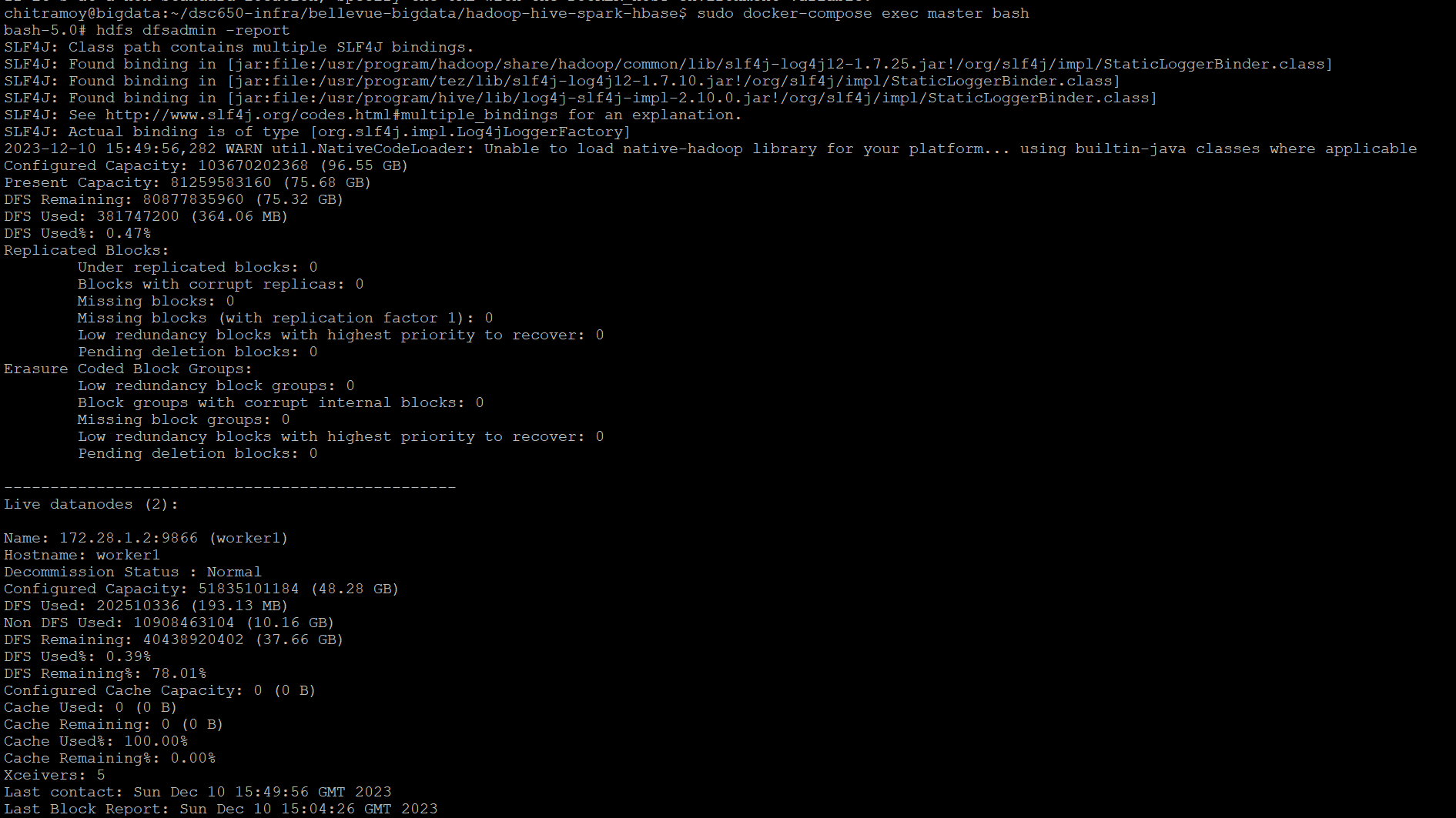
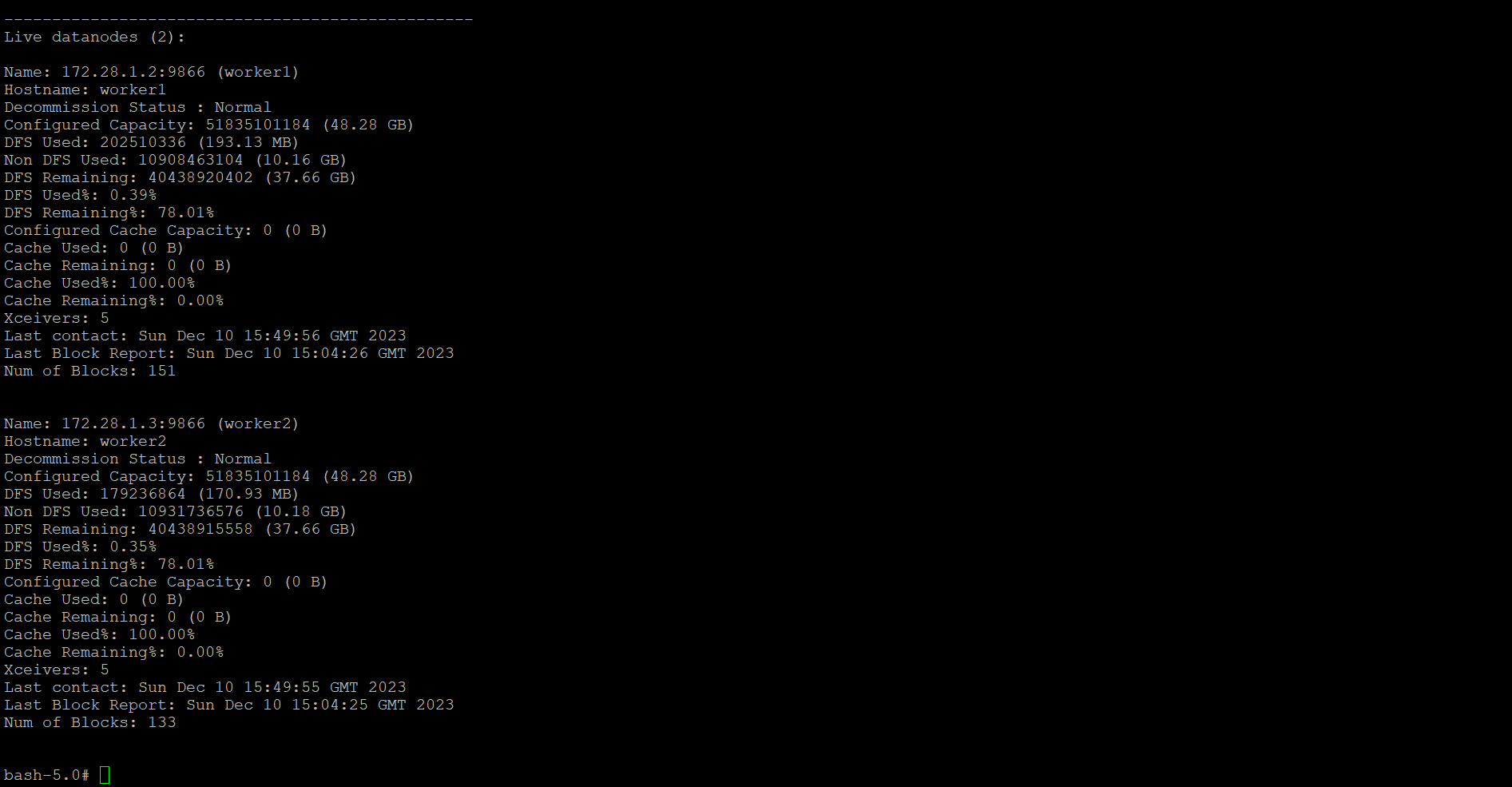
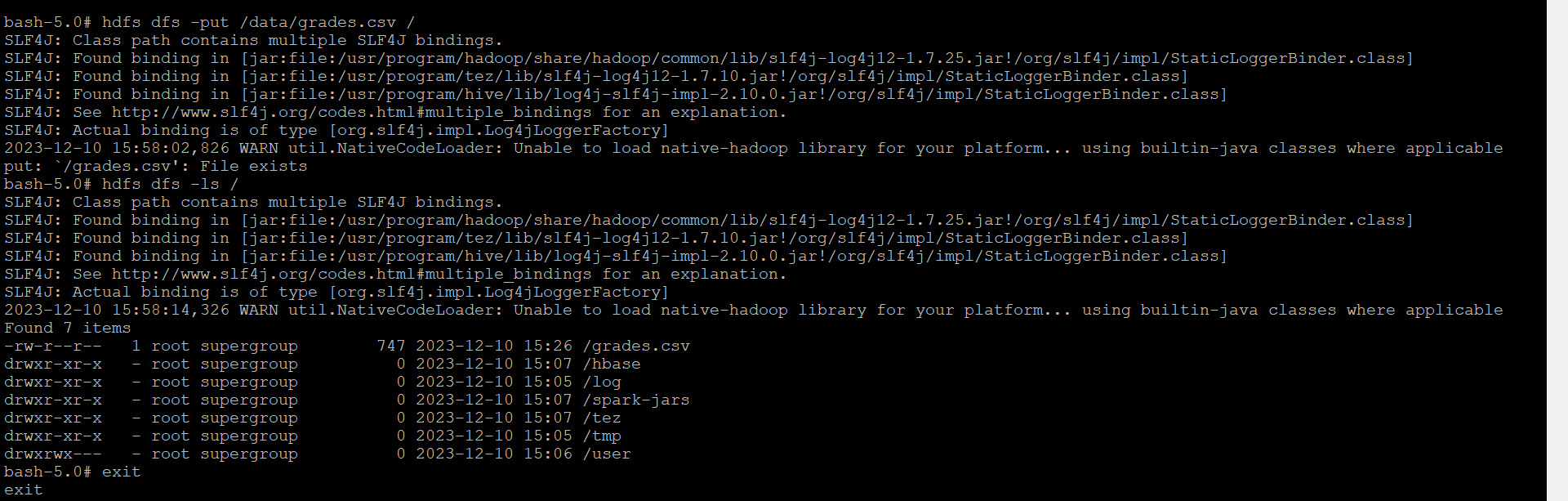
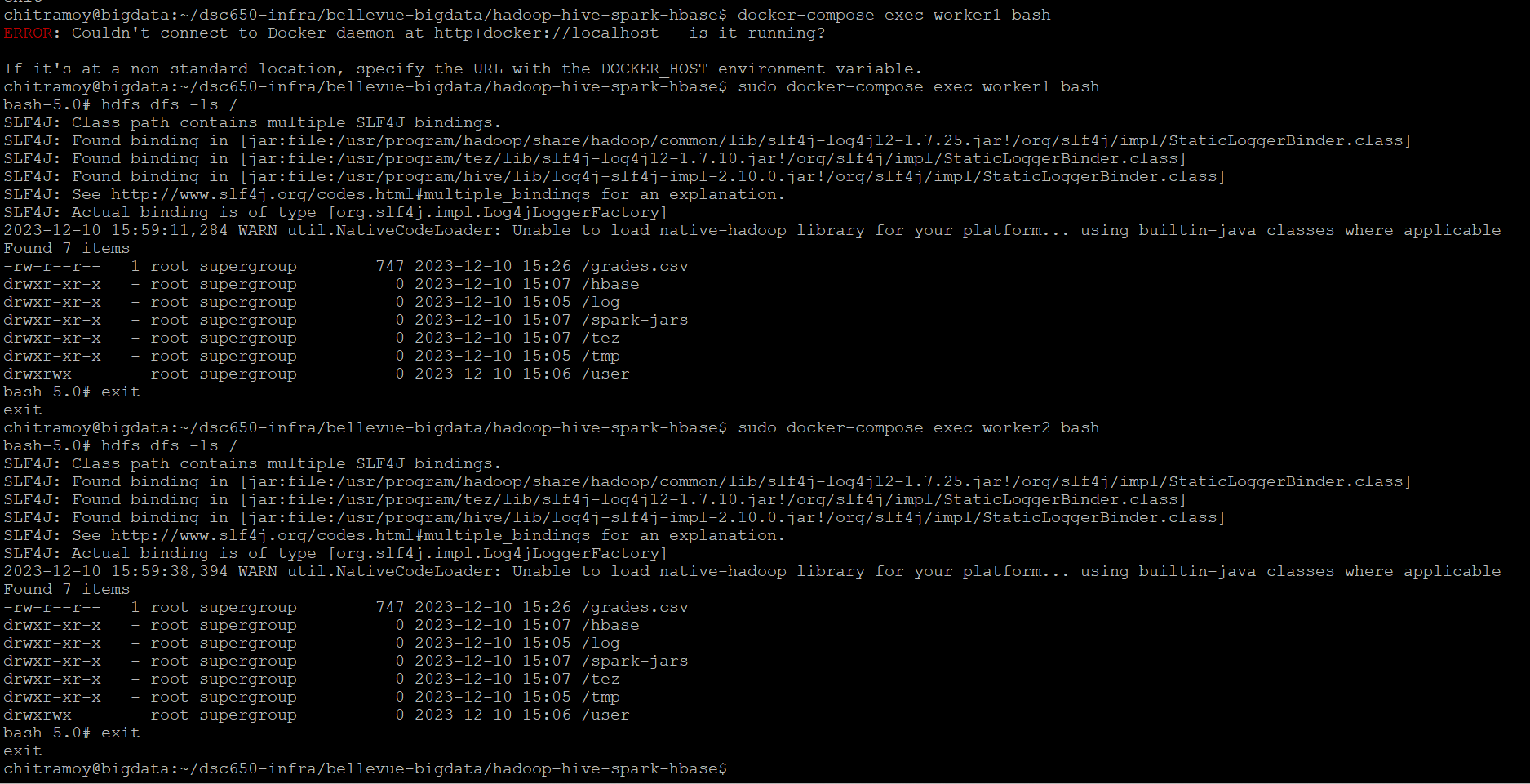
**hdfs dfsadmin –report Screenshot :**

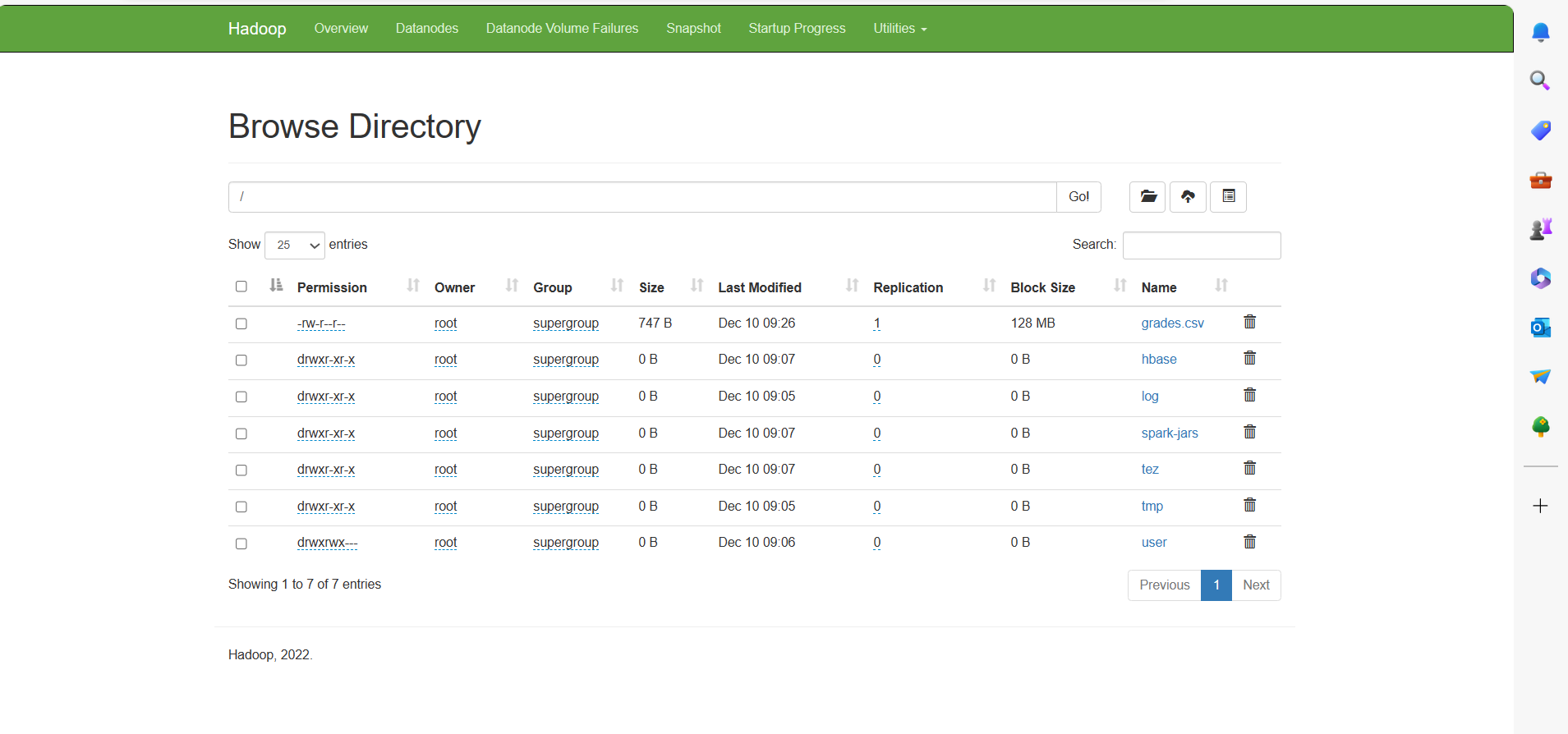




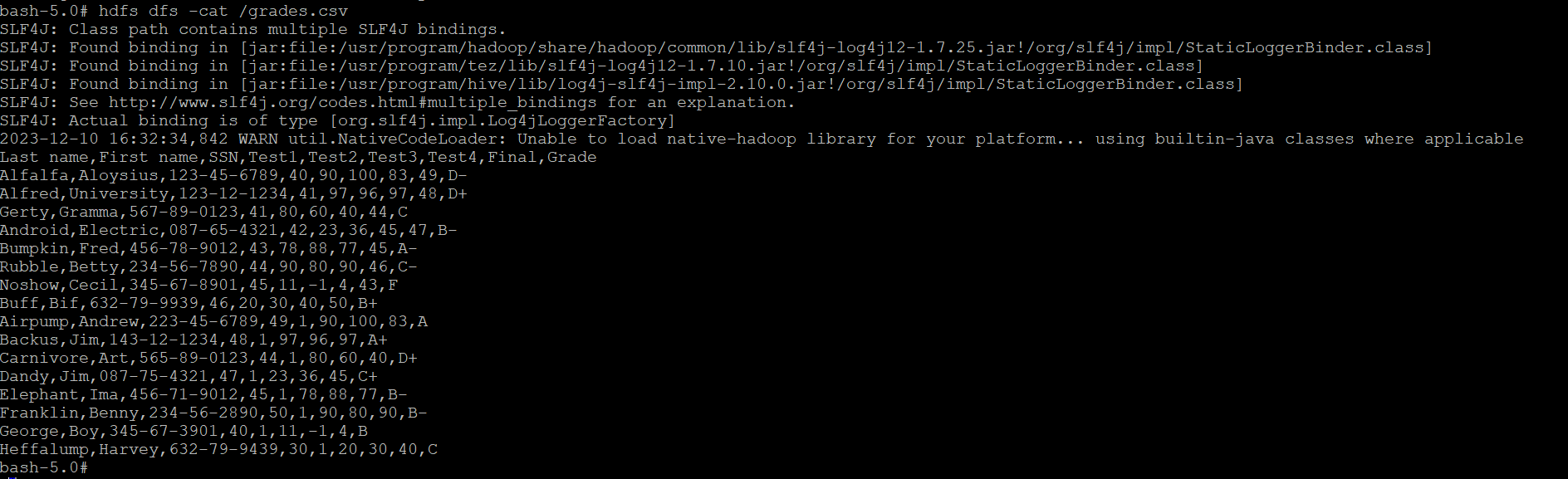
**Screenshot of uploading grades.csv file :**

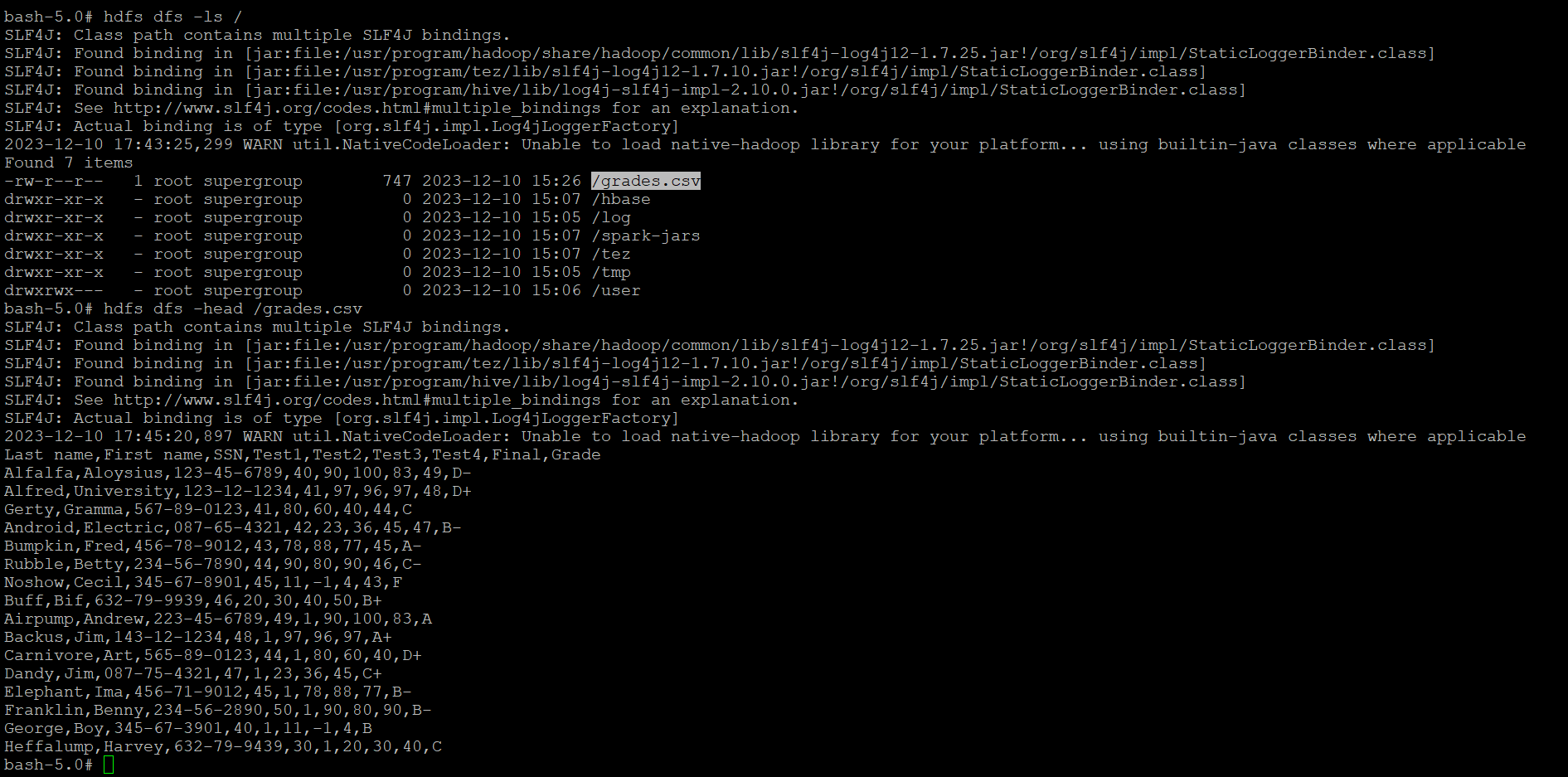


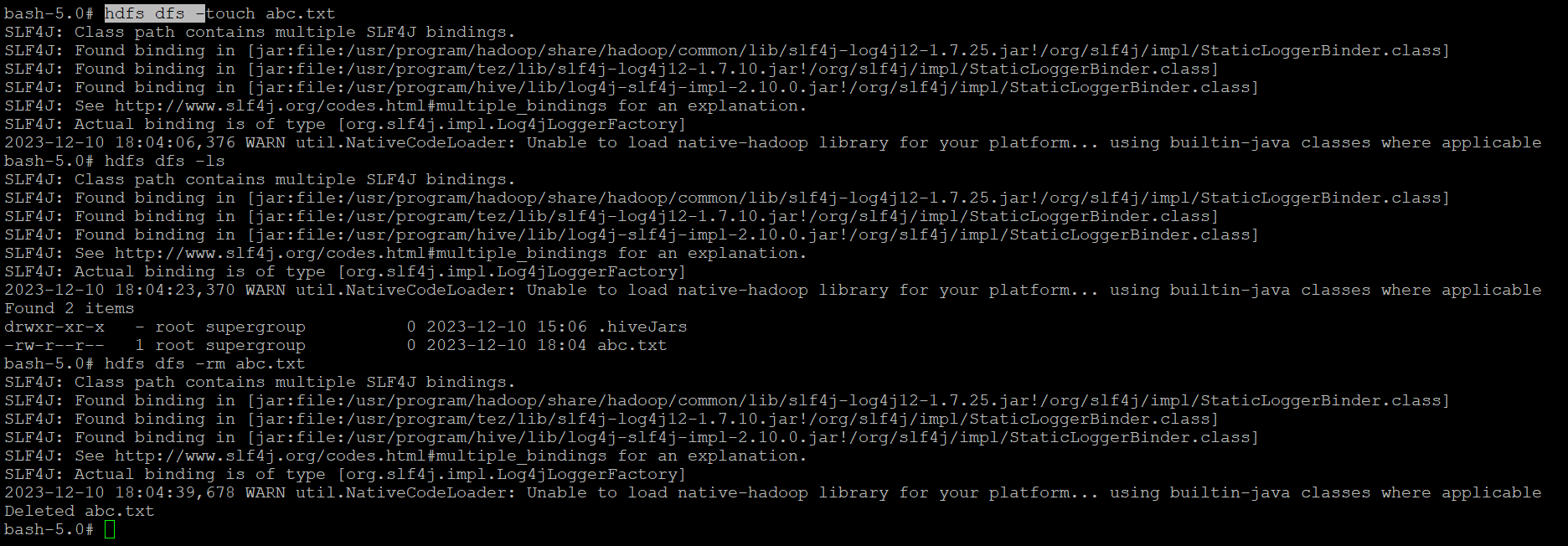




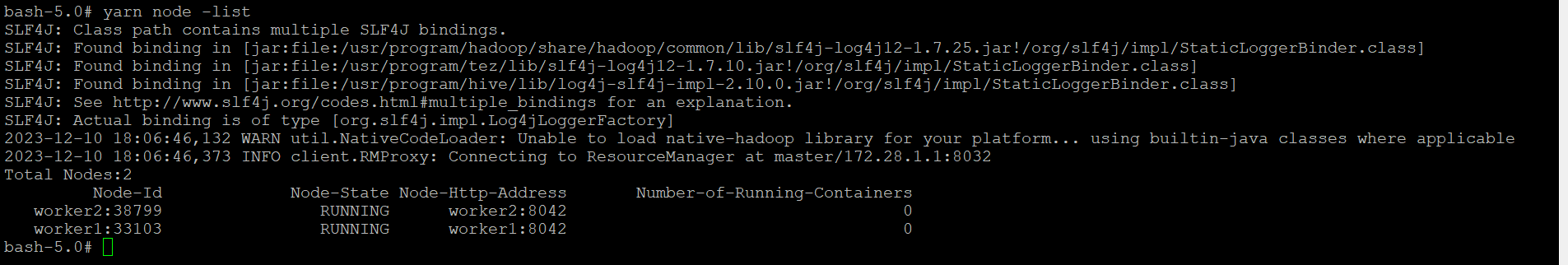
**Screenshots of the three chosen HDFS command outputs :**



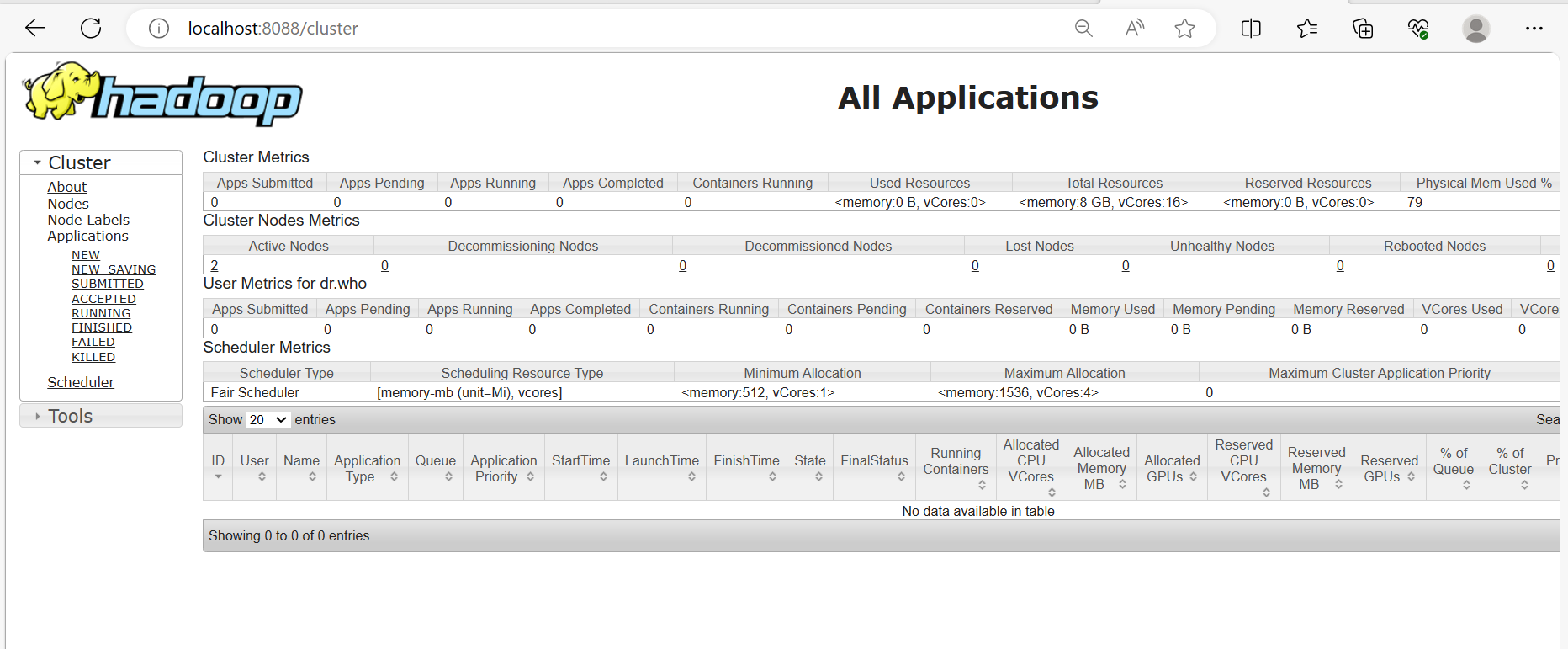


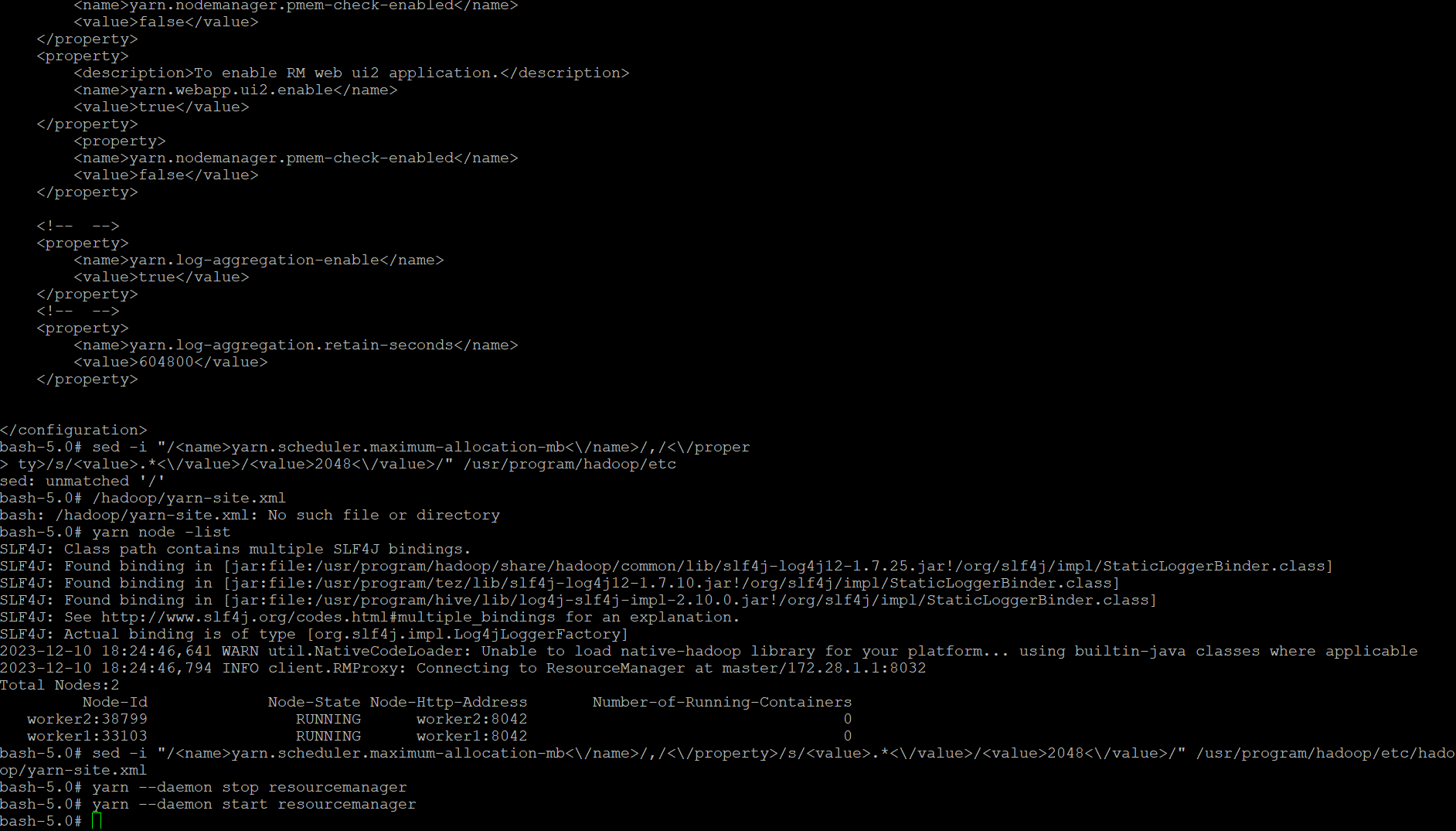


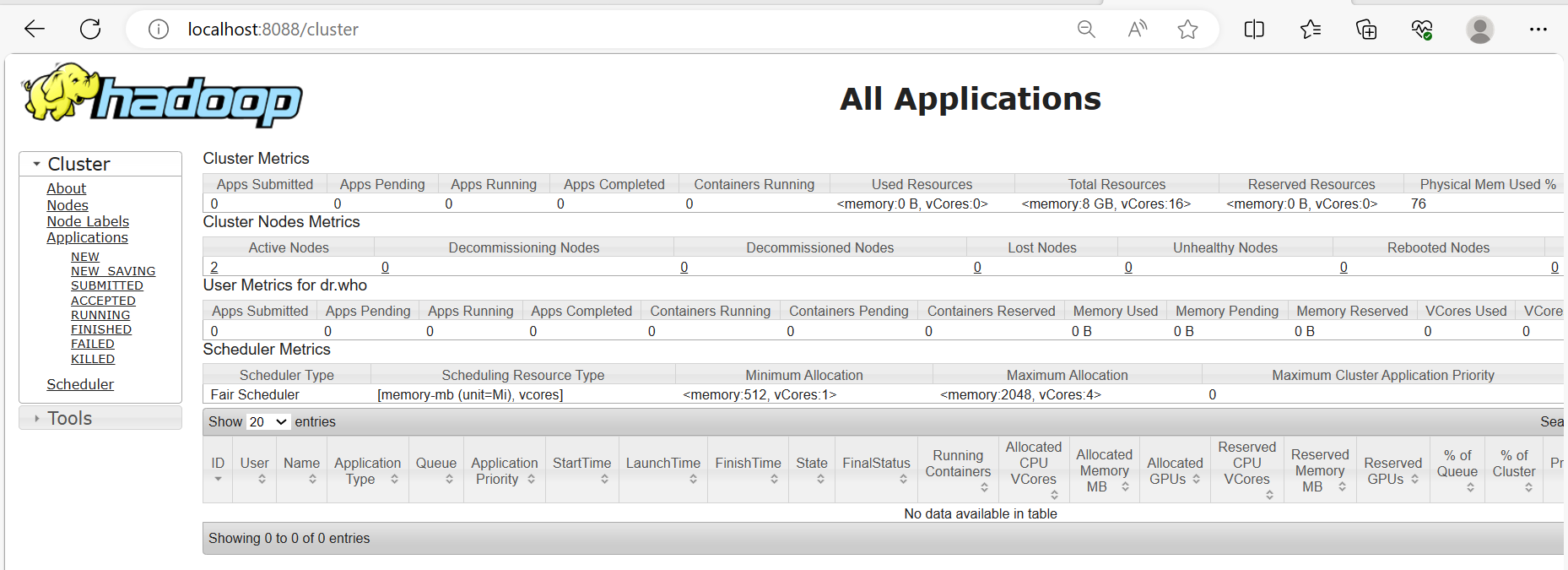
**Screenshot of the results YARN :**



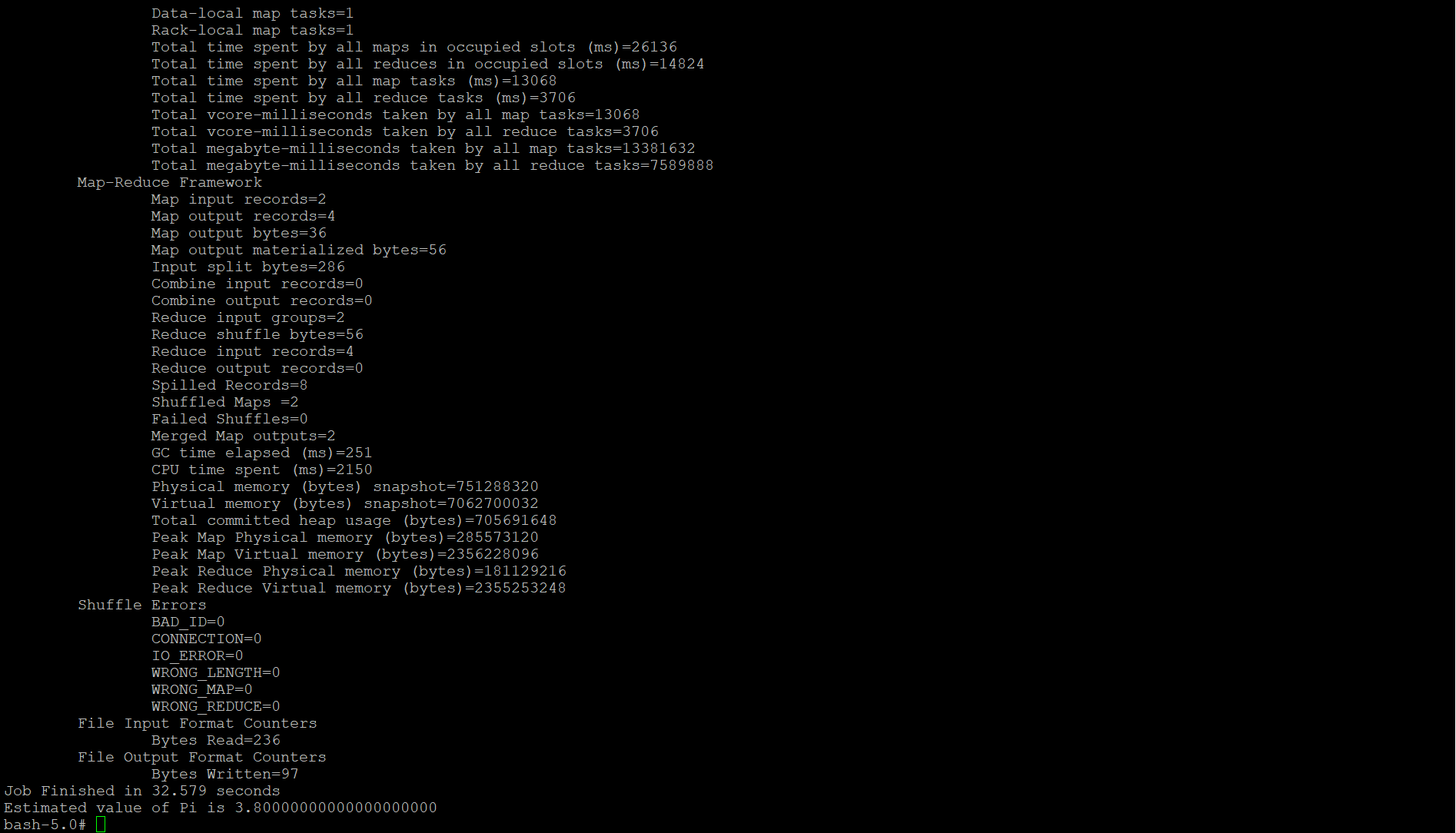
**Screenshot from the YARN UI showing the updated maximum memory (2048 MB) :**







**Summary of the result and its significance :**



**Analysis :**

The number of maps and the number of samples per map are parameters that affect the accuracy and efficiency of the Monte Carlo simulation used to approximate the value of Pi.

**Number of Maps (2 in the example):**

* The number of maps determines how many parallel tasks Hadoop will use to distribute the work.
* In the Pi example, each map performs a certain number of random samples using the Monte Carlo method to estimate the value of Pi independently.
* A higher number of maps can potentially lead to better parallelism and faster computation, especially when dealing with large datasets. However, it also depends on the size of your dataset and the nature of your computation.

**Number of Samples per Map (10 in the example):**

* Each map task generates a certain number of random samples to contribute to the Pi approximation.
* A higher number of samples per map generally leads to a more accurate Pi approximation, as it increases the precision of the Monte Carlo simulation.
* However, increasing the number of samples per map also increases the computational load on each map task, potentially affecting performance and parallelism.

In summary, the values **2** and **10** in the command represent the number of maps and the number of samples per map, respectively. Adjusting these values allows to balance the trade-off between accuracy and computational efficiency based on your specific requirements, available resources, and the characteristics of your data.

Optimal values for these parameters can depend on the size of your dataset, the computing resources available in Hadoop cluster, and the desired level of precision in the Pi approximation. Below are the pi values based on the change of parameters.

|  |  |  |  |
| --- | --- | --- | --- |
| **Iteration** | **Number of Maps** | **Number of Samples per MAP** | **pi value** |
| 1 | 2 | 10 | 3.80 |
| 2 | 3 | 10 | 3.60 |
| 3 | 4 | 10 | 3.40 |
| 4 | 5 | 10 | 3.28 |
| 5 | 6 | 10 | 3.33 |
| 6 | 2 | 30 | 3.33 |
| 7 | 2 | 45 | 3.24 |