**BÁO CÁO LAB1 LƯU TRỮ XỬ LÝ DỮ LIỆU LỚN : HDFS**

**Nhóm: squad game**

**Thành viên: Lại Ngọc Thăng Long 20183581**

**Nguyễn Đình Dũng 20183506**

**Nguyễn Thành Long 20183586**

**Nguyễn Khương Duy 20183513**

**BÀI LÀM:**

1. **Trên toàn bộ máy**

# 1st Step: Cài đặt SSH, PDSH

Cài đặt thành công SSH, PDSH

Text

Description automatically generated

# 2nd Step: Cài đặt Java 8

Cài thành công Java

Text

Description automatically generated

# 3rd Step: Tải Hadoop

# 4th Step: Tạo 1 user hadoopuser mới trên các máy

Text

Description automatically generated

Setup network

https://www.youtube.com/watch?v=ADmgVT4ovak

Cấu hình các node

A screenshot of a computer

Description automatically generated with medium confidence

1. **Trên máy hadoop-master**

# 1st Step: Dùng SSH để kết nối từ hadoop-master tới hadoop-slave1, hadoop-slave2

# 2nd Step: Config các file

Cấu hình 2 bản sao:

Text

Description automatically generated

# 3rd Step:

* Gửi các file đã config sang các máy slave

# 4th Step: Format HDFS file system

* Format HDFS file system:

source /etc/environment  
hdfs namenode –format

https://lh6.googleusercontent.com/1Cclfmu4FShPnJwZ5m6lRFk3SFi25SU0cA15ClVkxn_BFrifZCx4DsHbz9cq8DI-FY9PgDtMFj9ly1LBmWXEknQy-bLH544RK8yH1VO0UsBKNYPWOxyPWqgV5kwMK289QLaRoatN

# 5th Step: Start HDFS

* Start HDFS

start-dfs.sh

Text

Description automatically generated

* Trên máy master, kiểm tra:

A screenshot of a computer

Description automatically generated with medium confidence

* Trên máy slave1 và slave2

Text

Description automatically generated

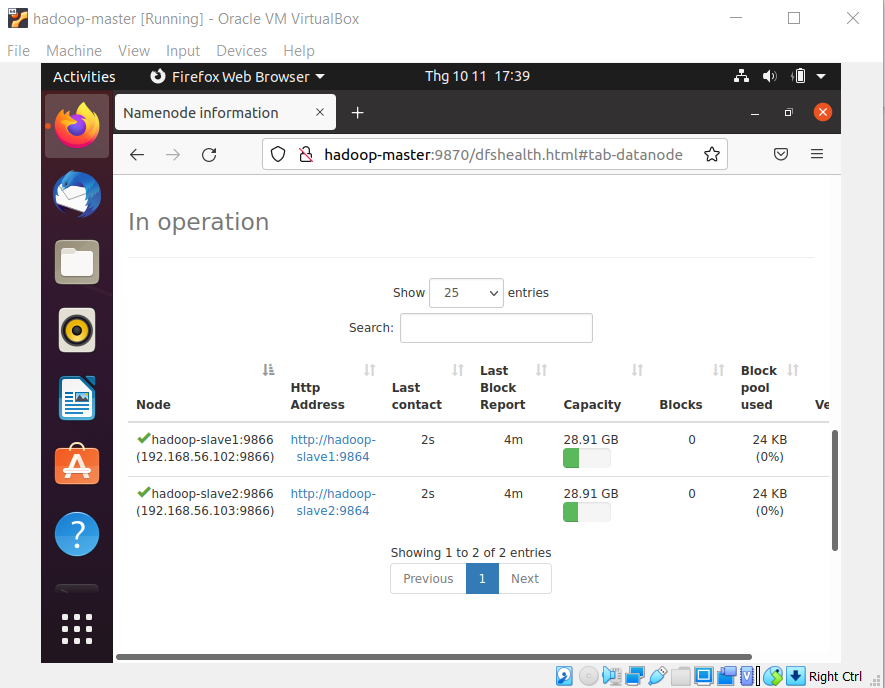
Text

Description automatically generated

Check trên web:

Graphical user interface, text

Description automatically generated



1. **Đẩy dữ liệu lên cụm HDFS**

Đẩy 1GB dữ liệu lưu trữ trên HDFS, replication=2

Graphical user interface

Description automatically generated