**Báo cáo Lưu Trữ Và Xử Lý Dữ Liệu Lớn**

**Lab2:MapReduce**

**Nhóm**:LoveData

**Thành viên**:

Phương Trung Kiên 20183776

Lê Nguyễn Thành Chung 20183696

Đào Trọng Hiếu 20183739

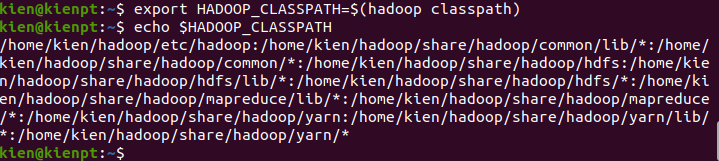
HIM PHON 20170283

**1.Mô tả kết quả cài đặt:**

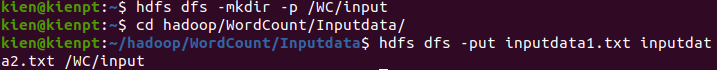
-Sử dụng 3 máy đã cài HDFS ở lab1 có cấu hình file Yarn và Mapreduce

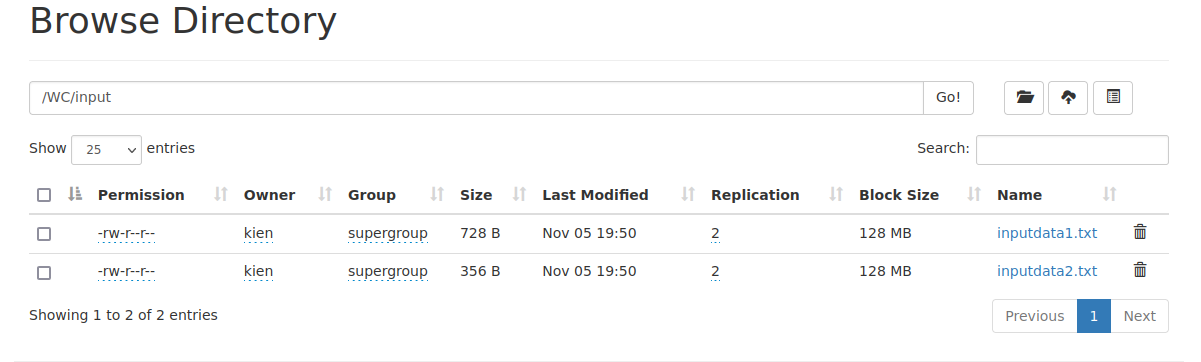
-Các bược cấu hình để chạy Mapreduce:

+Tạo classpath:



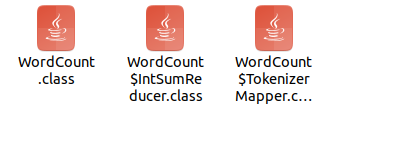
+Tạo 2 file text đẩy lên HDFS:



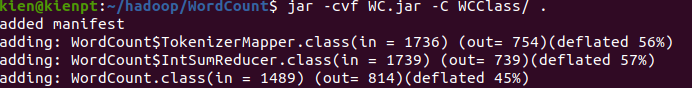


+Biên dịch WordCount.java thành các Class:





+Export 1 file jar từ các Class:

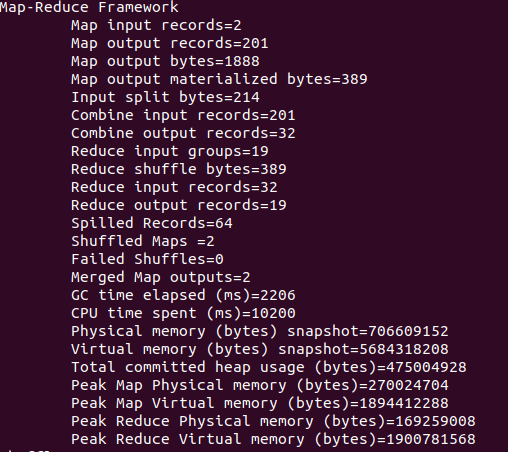


+Cuối cùng chạy Job:với 2 đường dẫn input và output tạo trên HDFS:

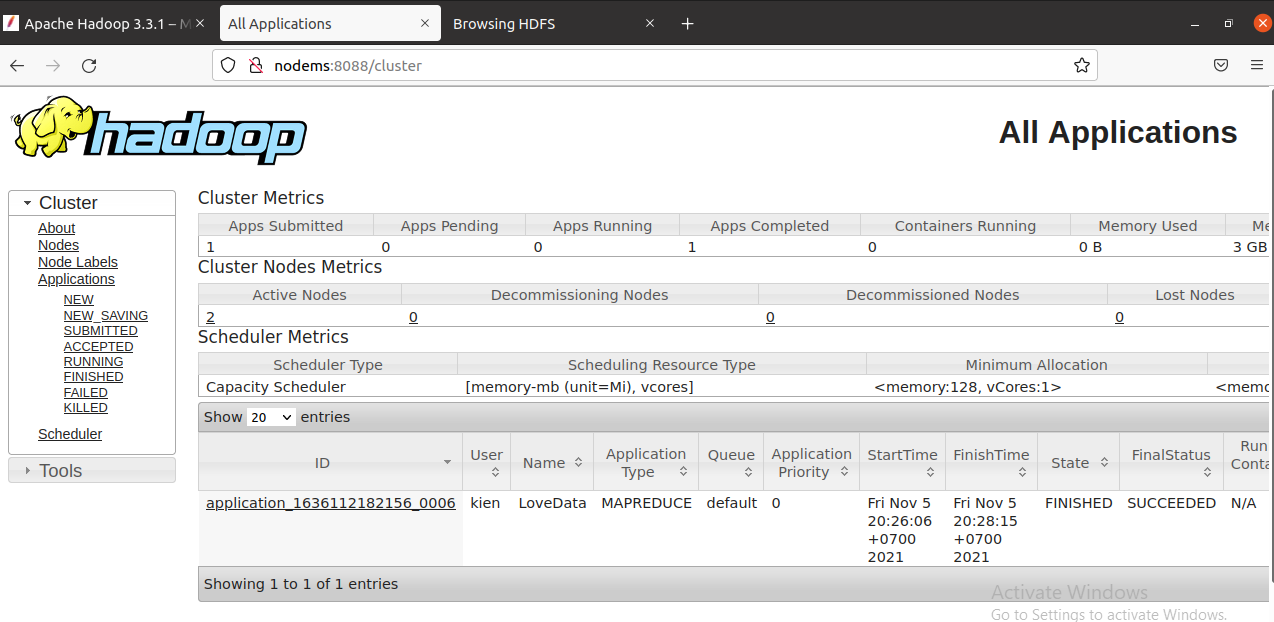


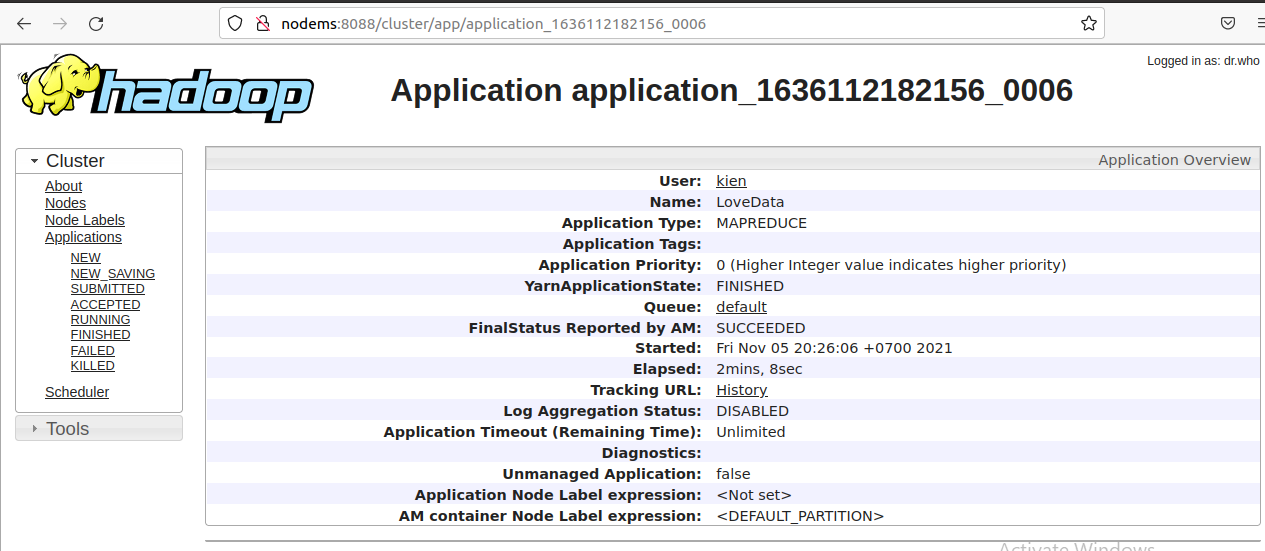
**2.Kết quả thu được:**

**-**Chương trình chạy với 2 mappers:

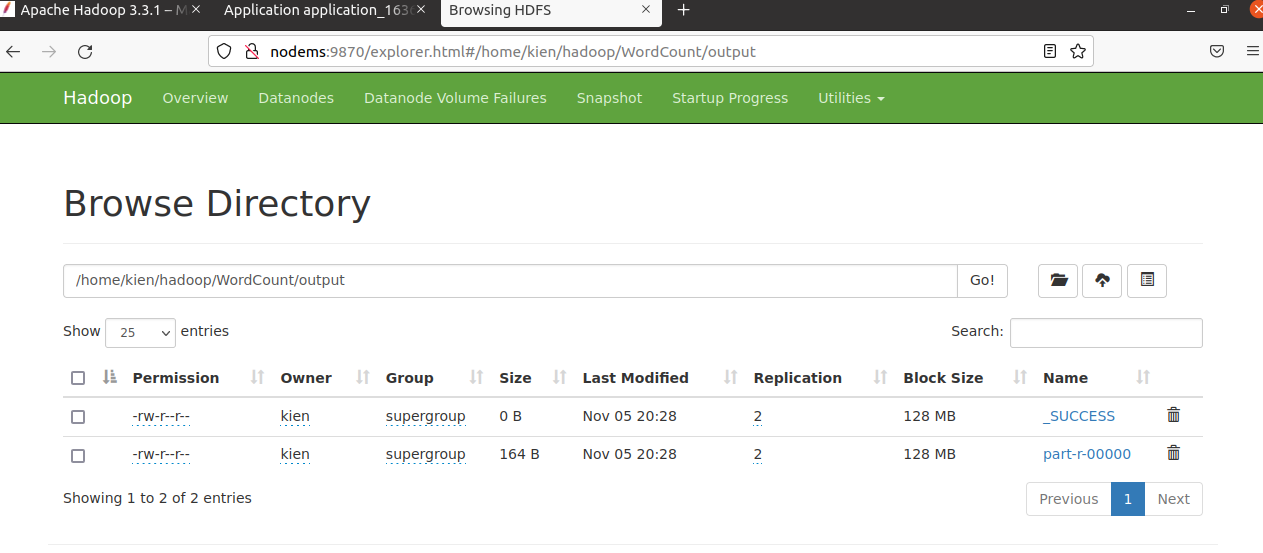


+job đã hiển thị được trên Cluster:





+Sauk khi thực hiện thành công ta được 2 file trong output:



+Kết quả thu được sau khi xử lý được hiển thỉ trong file part-r-00000:

