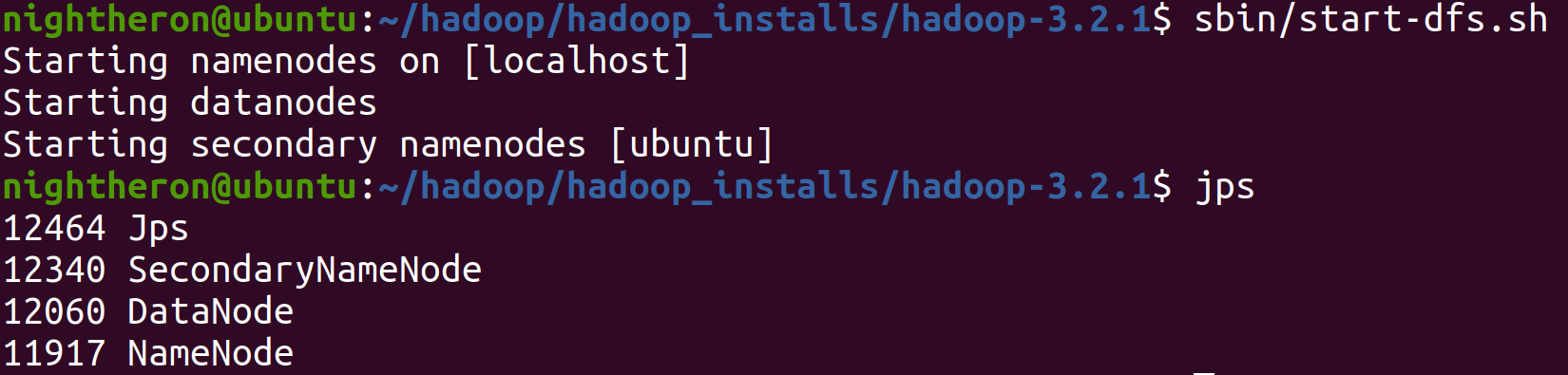
**实验 1：Hadoop 系统安装、开源大数据系统实验工具链使用**

陈德丹 221220159

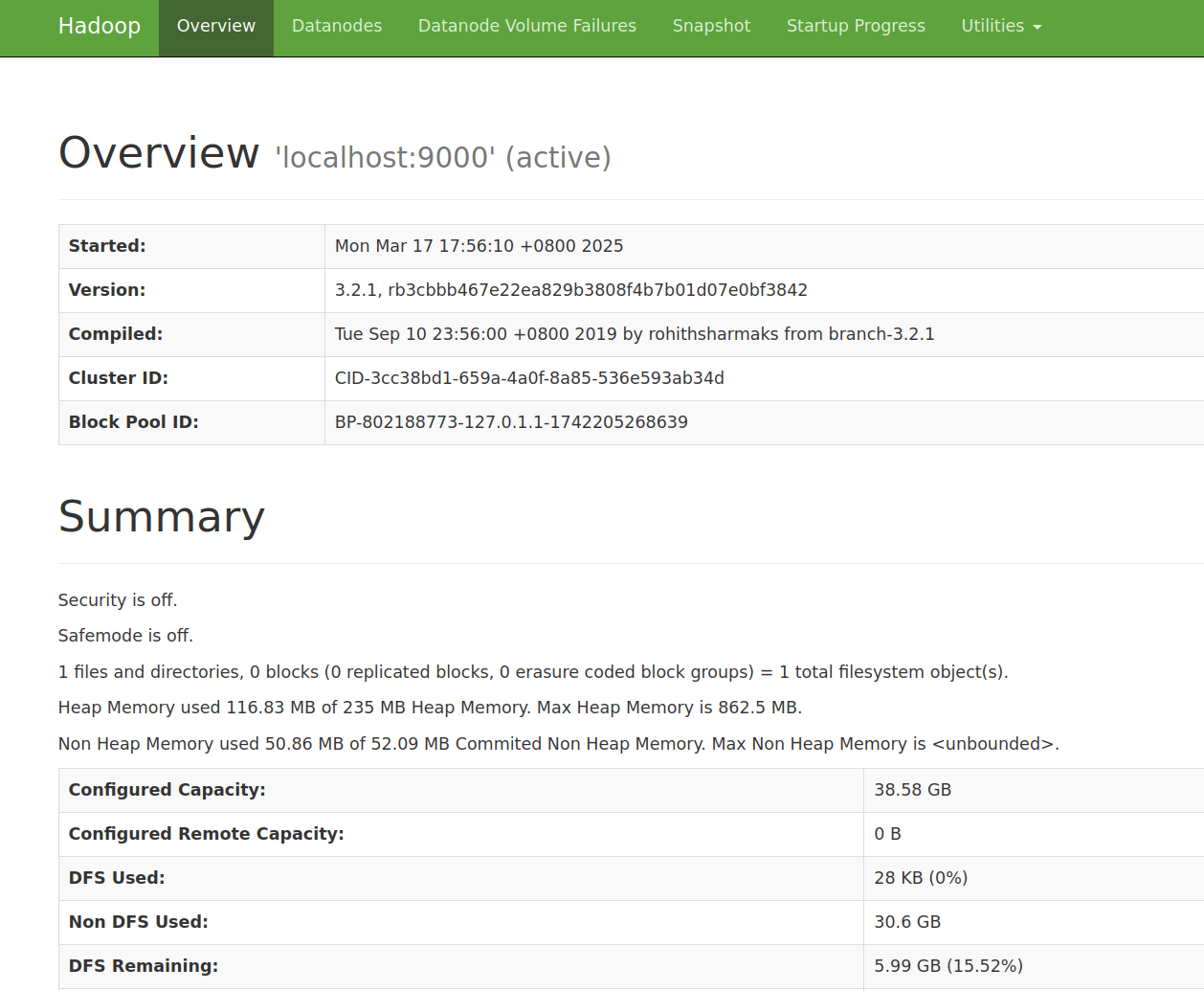
**任务1：Hadoop伪分布式环境安装与配置，HDFS基本操作**

⚪按照教程安装并配置hadoop和java

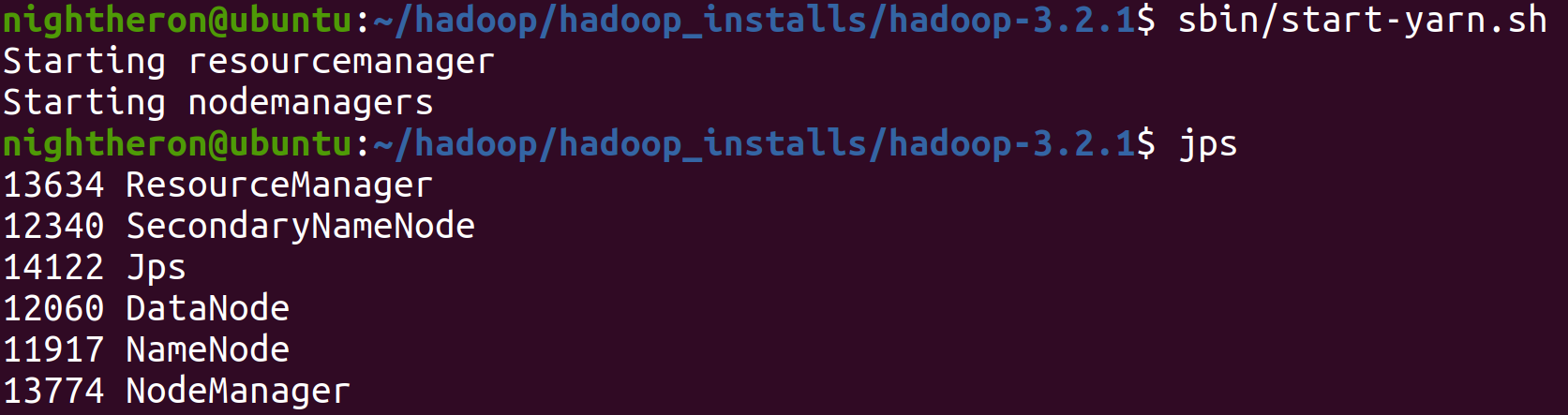
⚪启动NameNode daemon和DataNode daemon



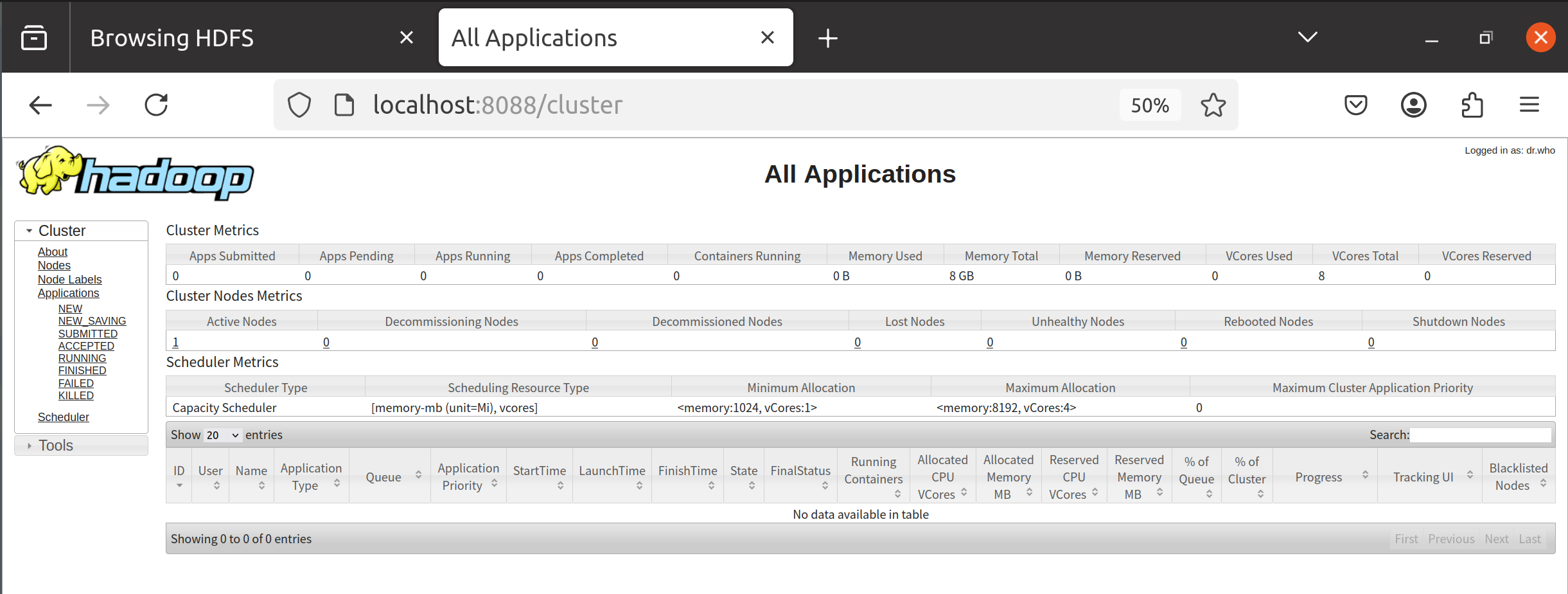
⚪浏览HDFS NameNode的Web接口



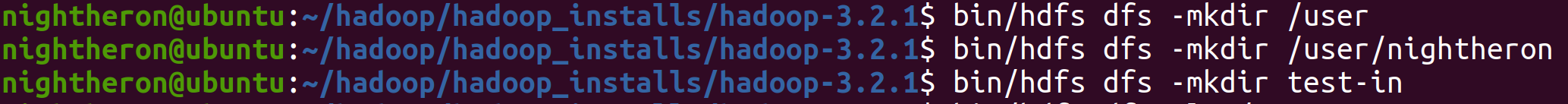
⚪启动ResourceManage daemon和NodeManage daemon



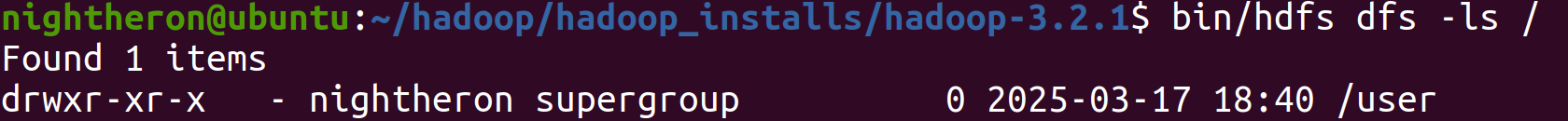
⚪浏览Resource Manager的Web接口



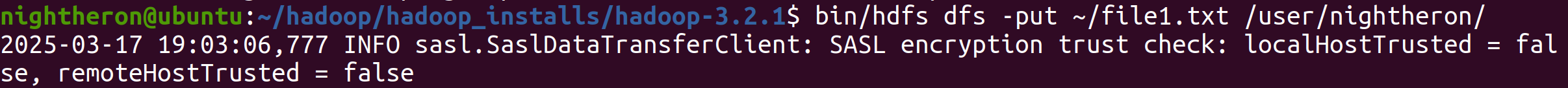
⚪创建HDFS目录



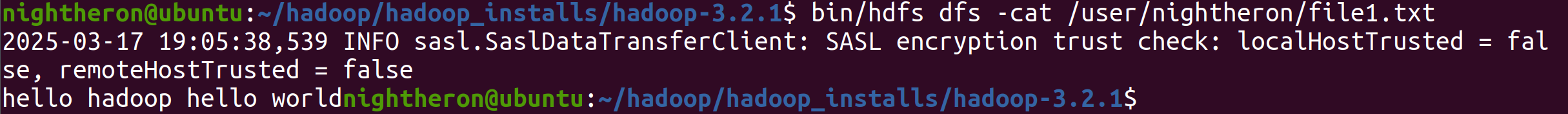
⚪查看HDFS根目录



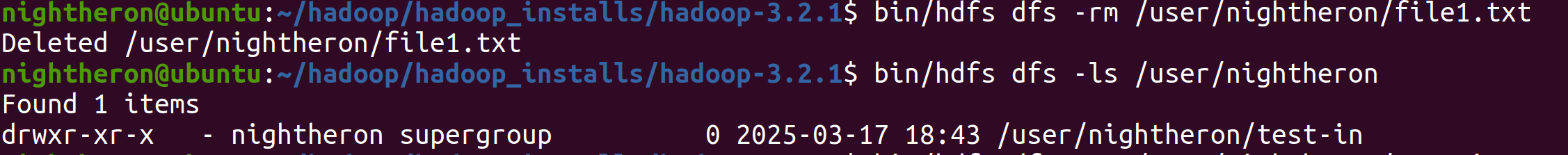
⚪向HDFS上传一个文件 file1.txt: hello hadoop hello world



⚪显示上传的文件的内容



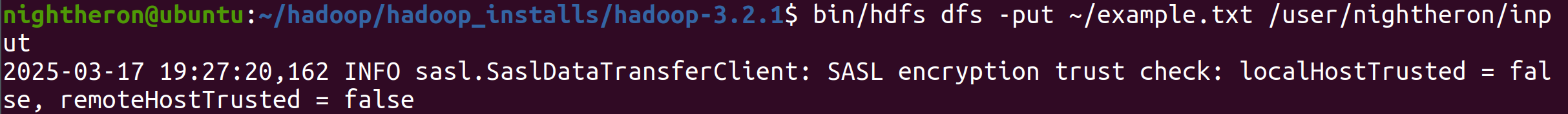
⚪删除上传的文件



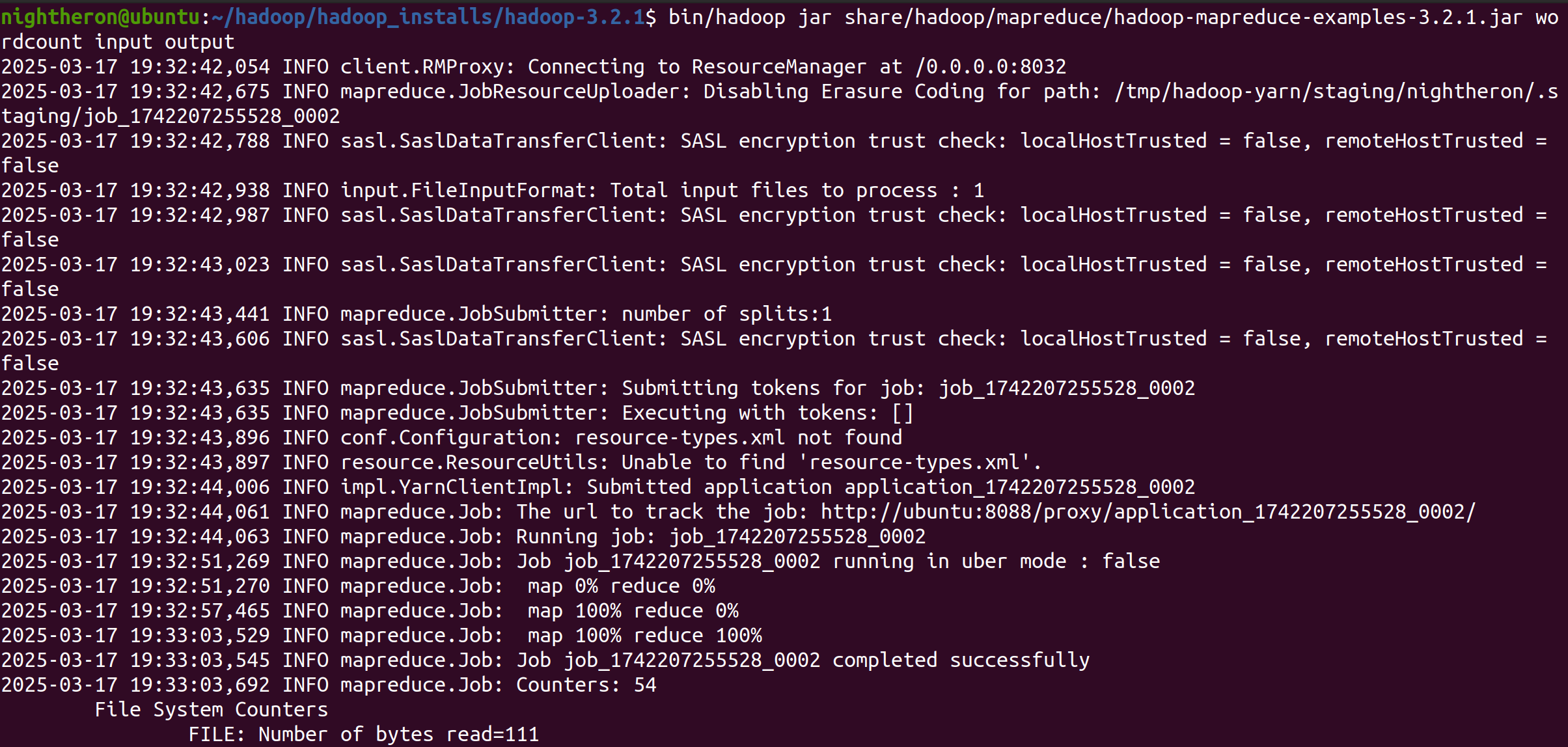
可以看到已成功删除

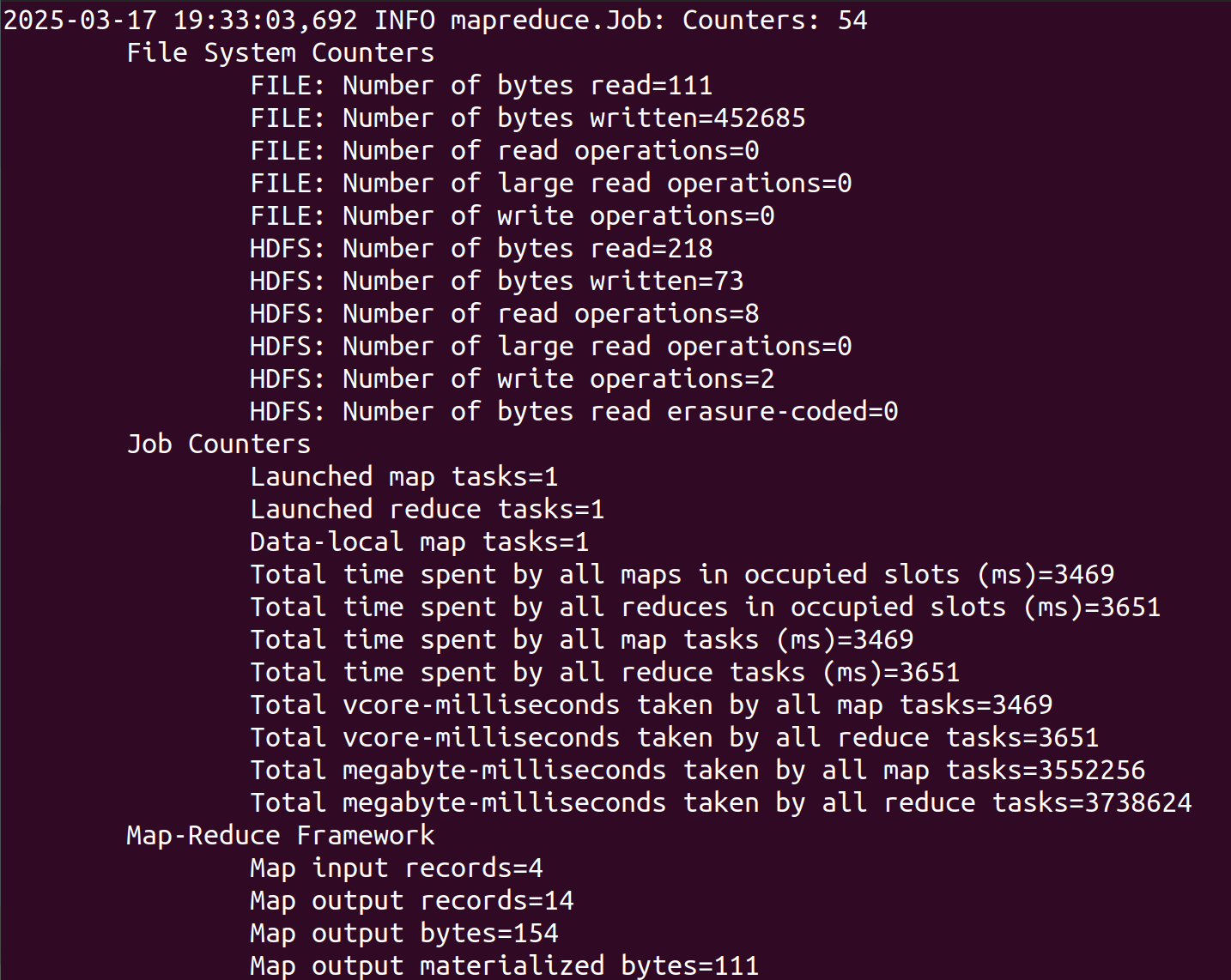
**任务2：执行Hadoop官方示例程序wordcount**

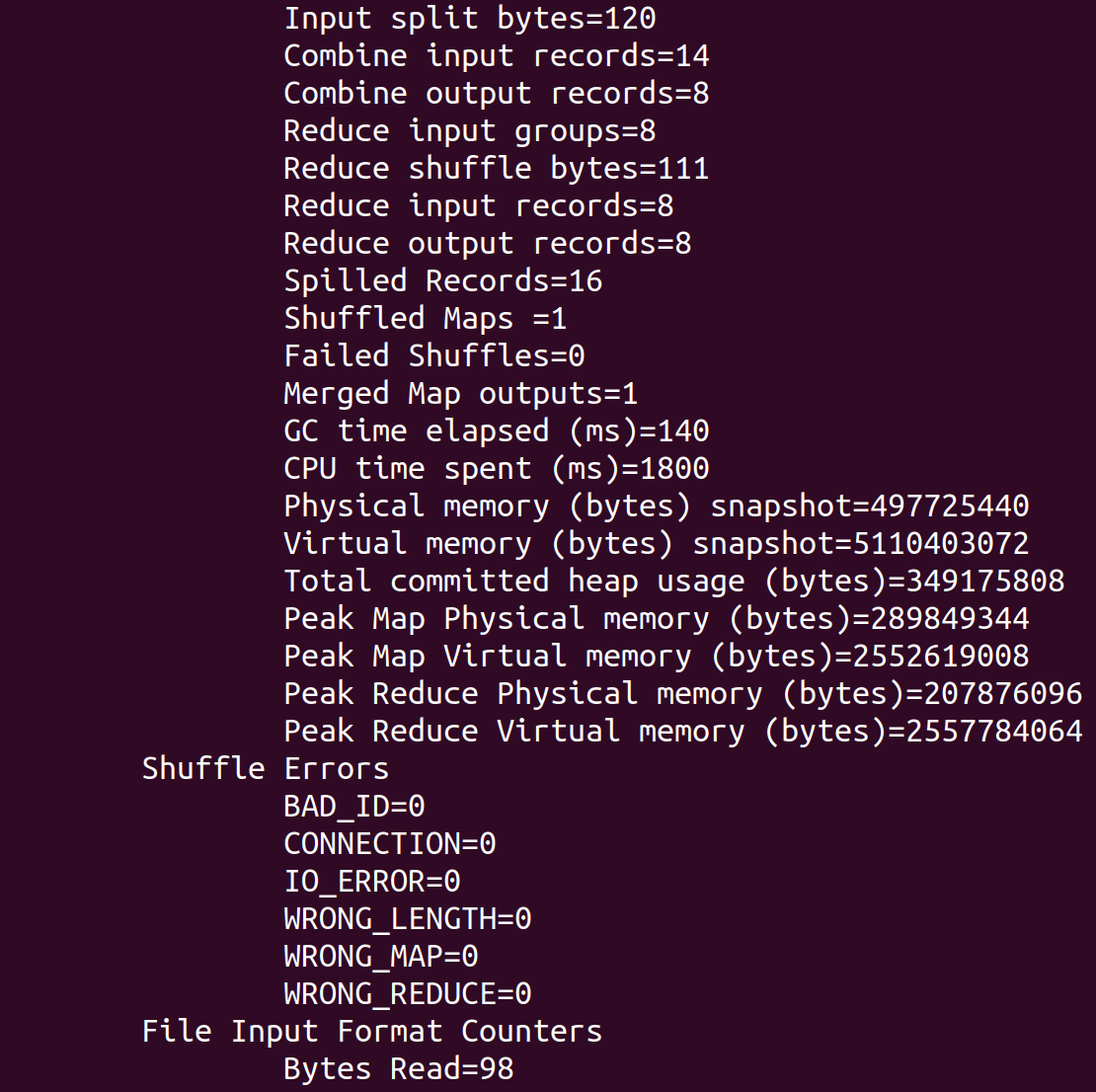
⚪上传example.txt到HDFS

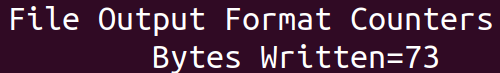


⚪执行wordcount

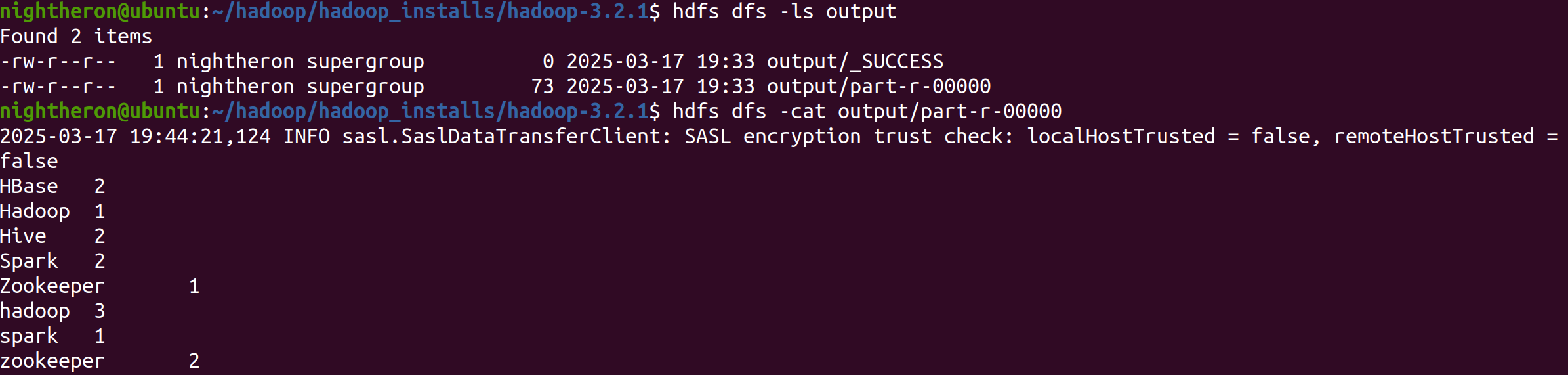






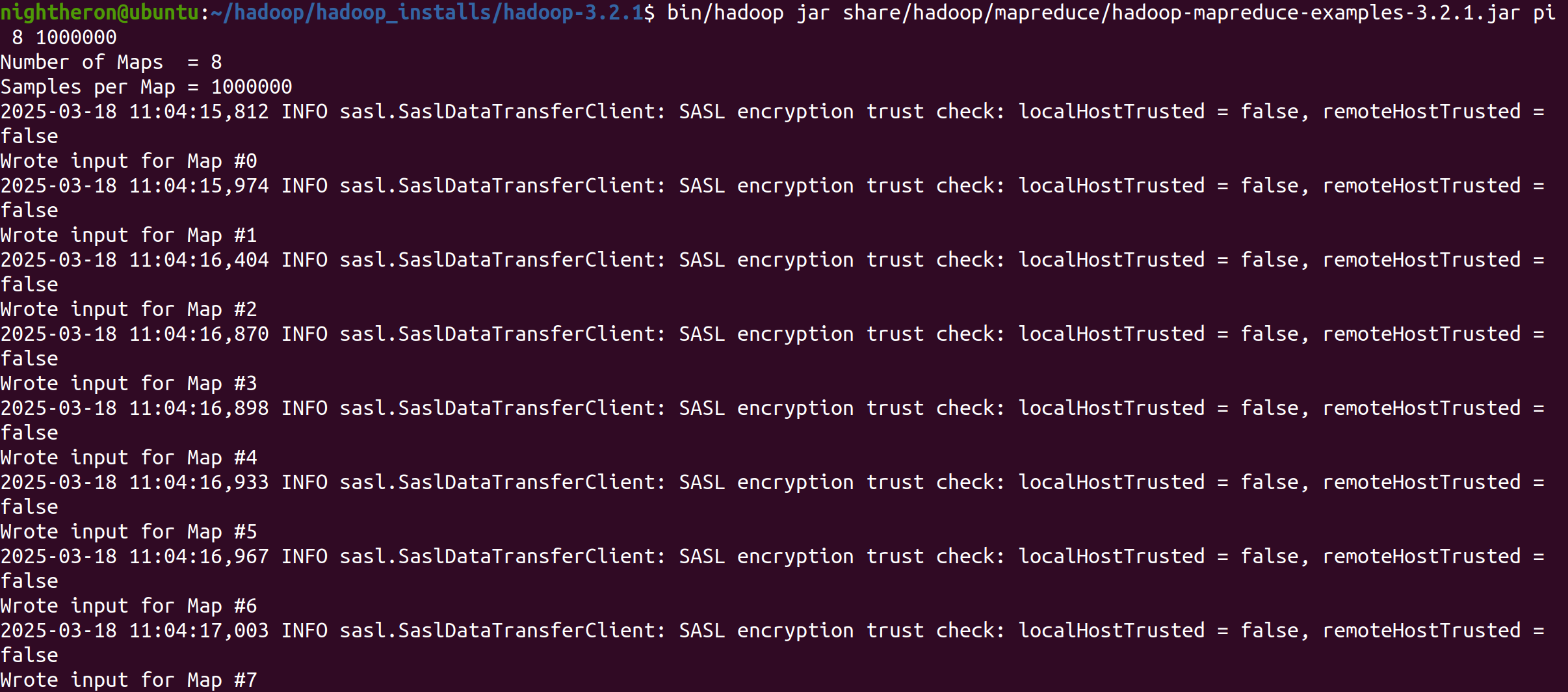


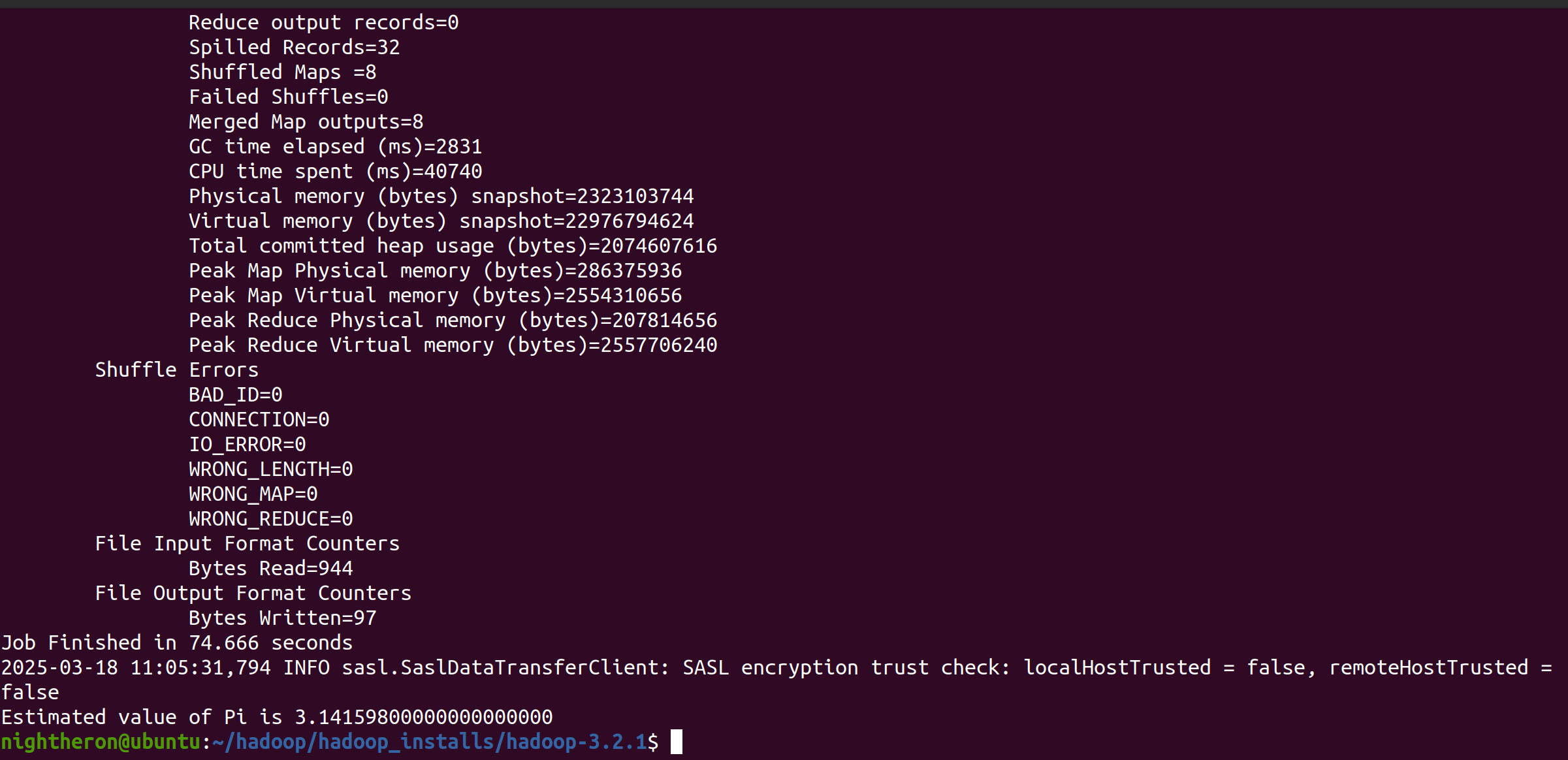
⚪查看执行结果：



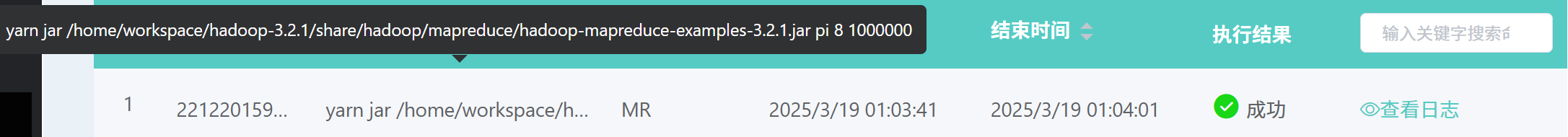
**附加实验(可选)：执行更多的Hadoop官方示例程序**

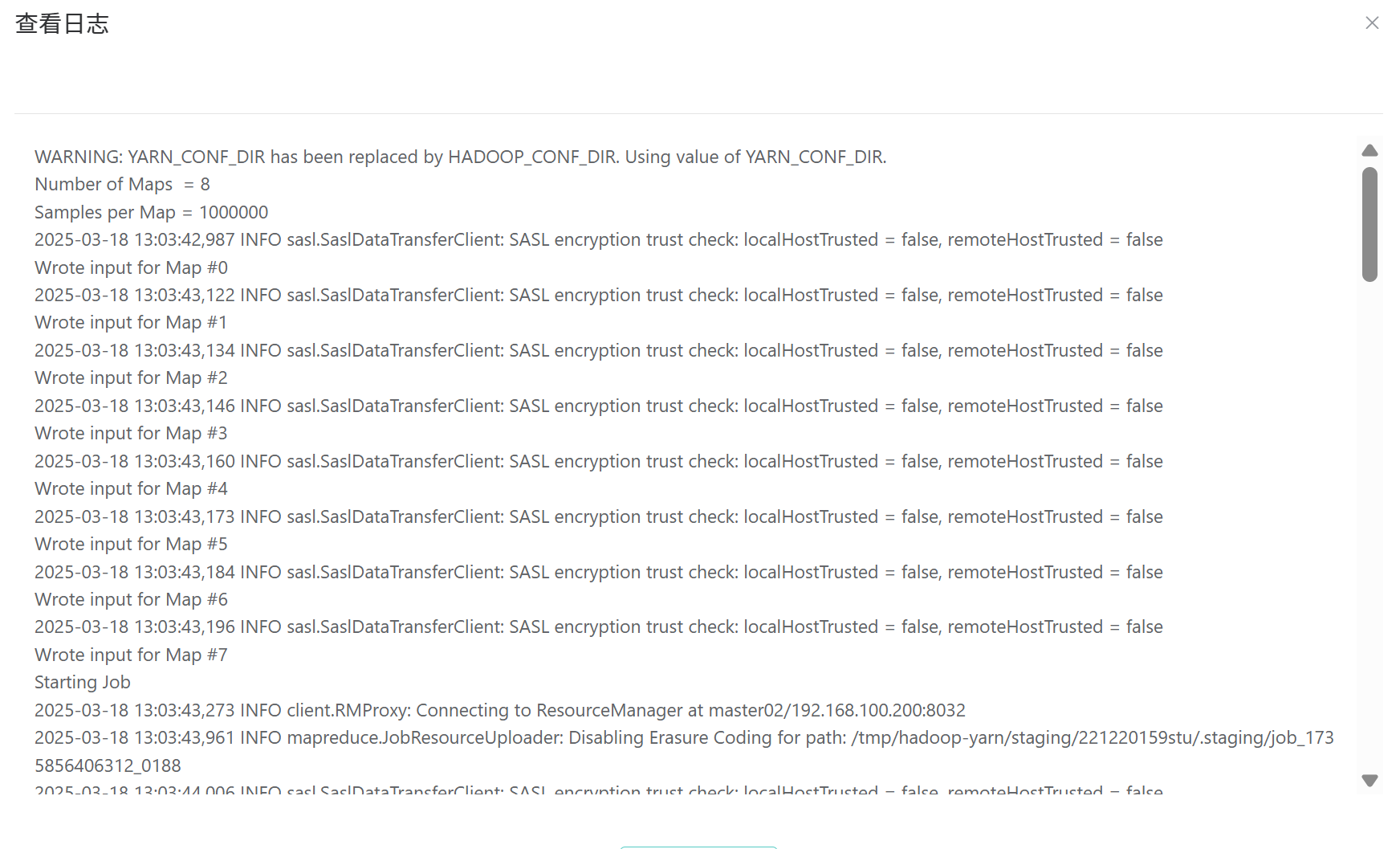
**1.本地运行pi示例程序，观察MapReduce任务执行过程(8个map任务，每个任务1,000,000个样本)**

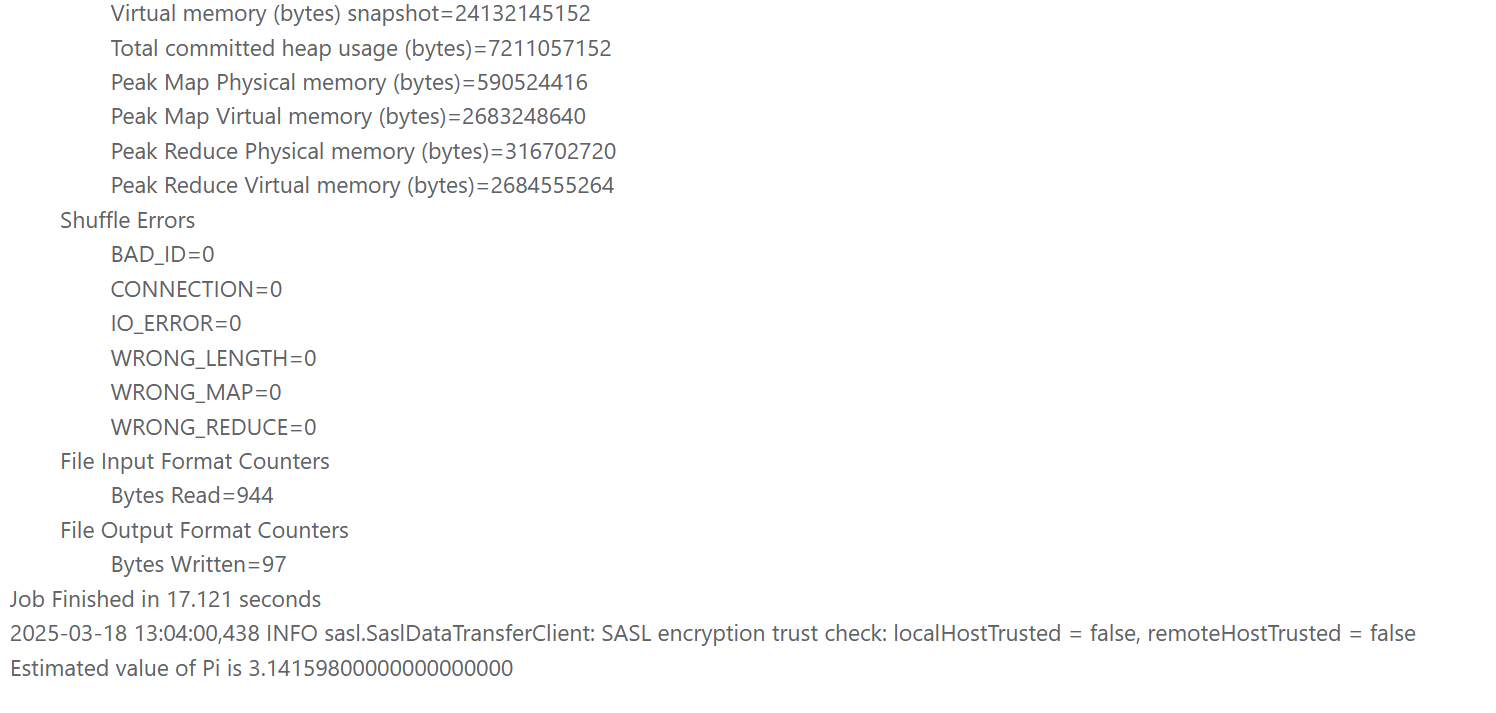




**2.实验平台上运行pi示例程序，观察MapReduce任务执行过程(8个map任务，每个任务1,000,000个样本)**



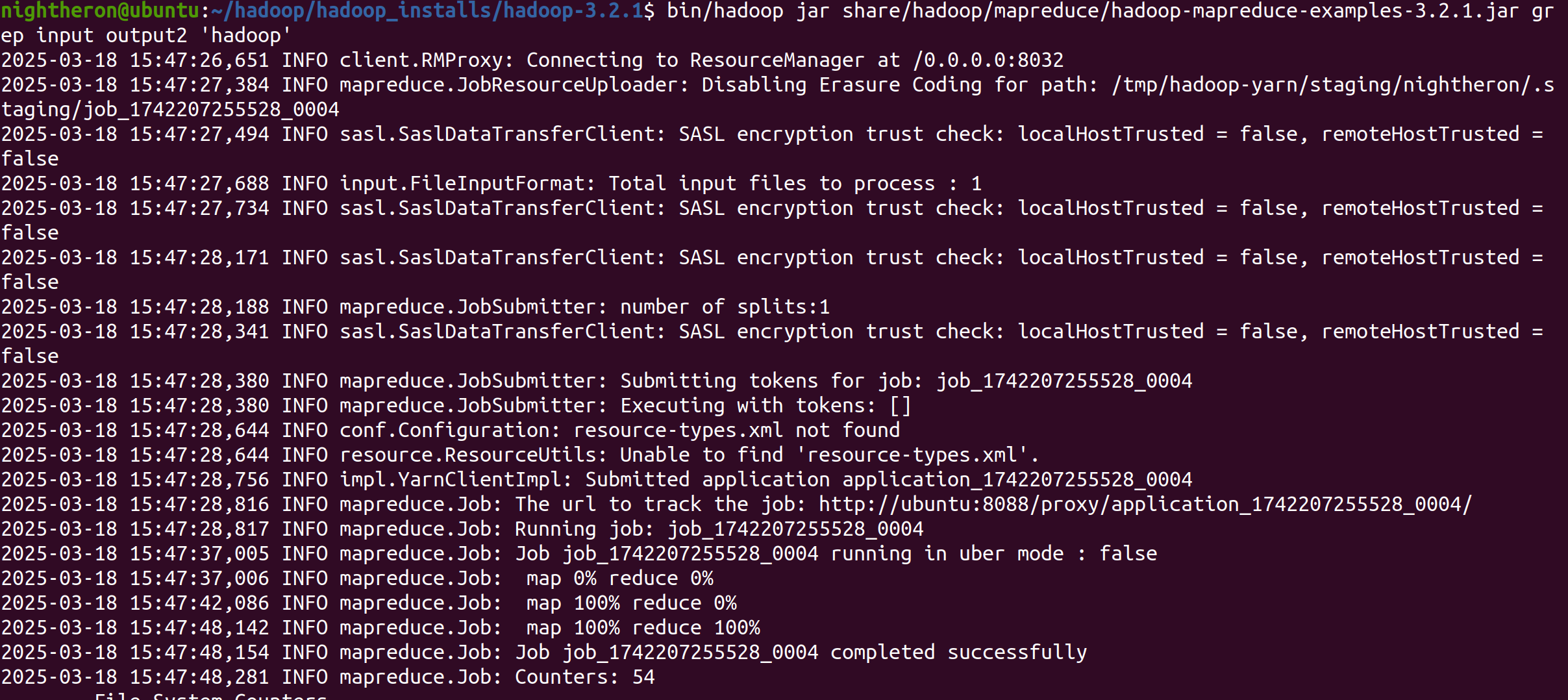




**3.比较1和2的执行时间，说明你的发现**

本地执行时间为1分16秒，实验平台执行时间为16秒，实验平台执行速度远快于本地，可见由于在集群模式下，任务会被分配到不同的节点进行并行处理，集群模式的并行处理能够显著提高执行速度。

**4.本地或实验平台上运行grep示例程序，找出给定的example.txt中"hadoop"这个单词出现的次数**



hadoop共出现3次

