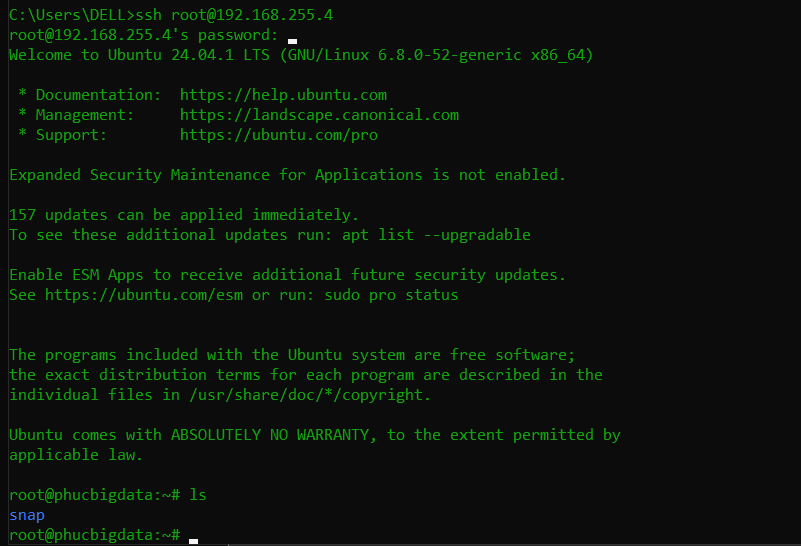
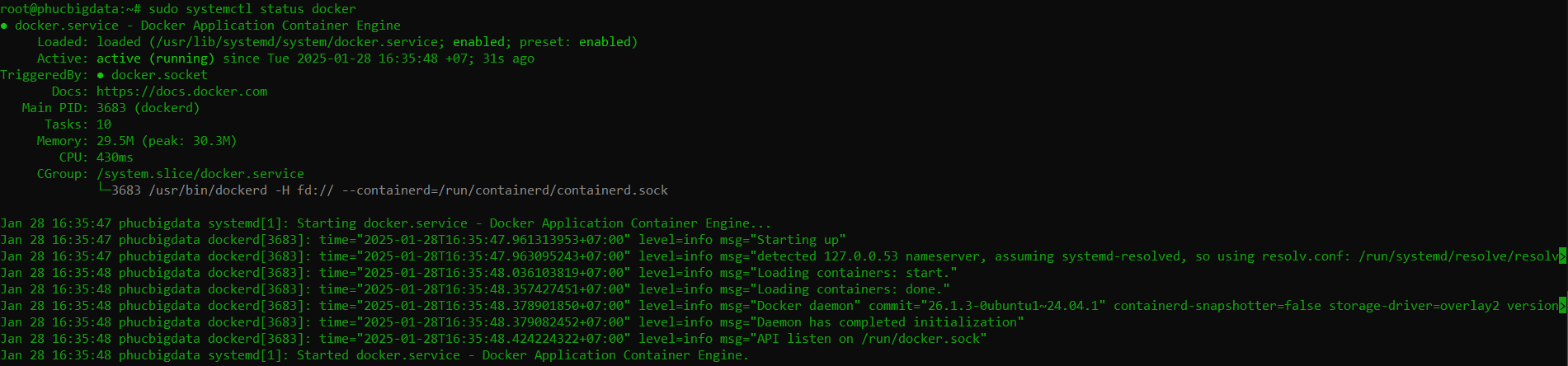
**Lab 2**

**The Process on VMware phucbigdata with ip 192.168.255.4**



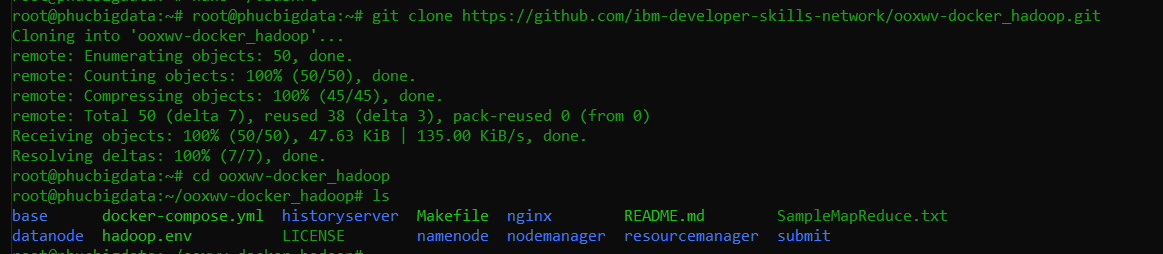
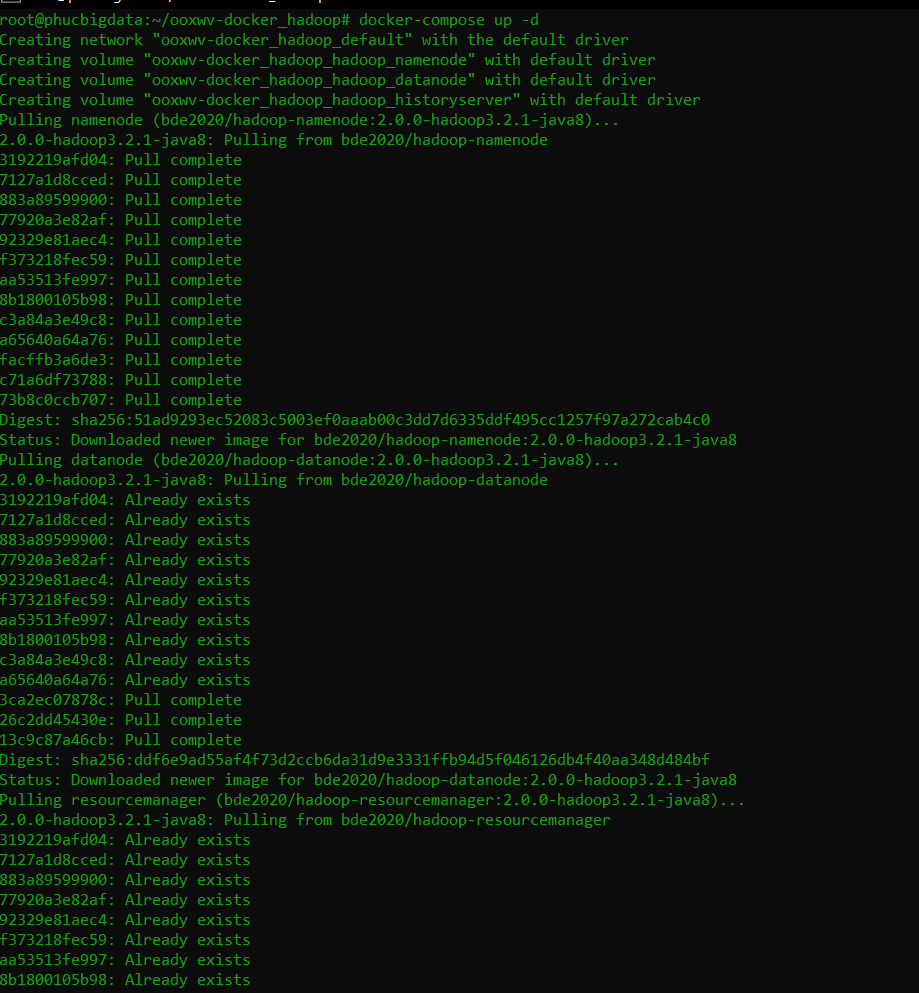
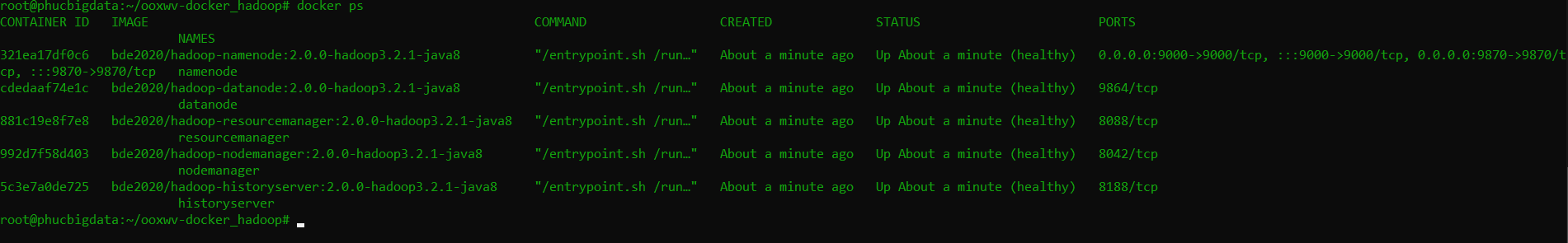
1. **Set up Docker on linux**

* **sudo apt update && sudo apt upgrade -y**
* **sudo apt install -y docker.io**
* **sudo systemctl start docker**
* **sudo systemctl enable docker**
  + 
  + 

1. **Set up Docker Compose**

* **sudo apt install curl**
* **sudo curl -L "https://github.com/docker/compose/releases/download/1.29.2/docker-compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose**
* **sudo chmod +x /usr/local/bin/docker-compose**
* **which docker-compose**
  + 
* **nano ~/.bashrc**
  + 
* **source ~/.bashrc**
* **docker-compose –version**
  + 

1. **Hadoop Cluster Deployment**

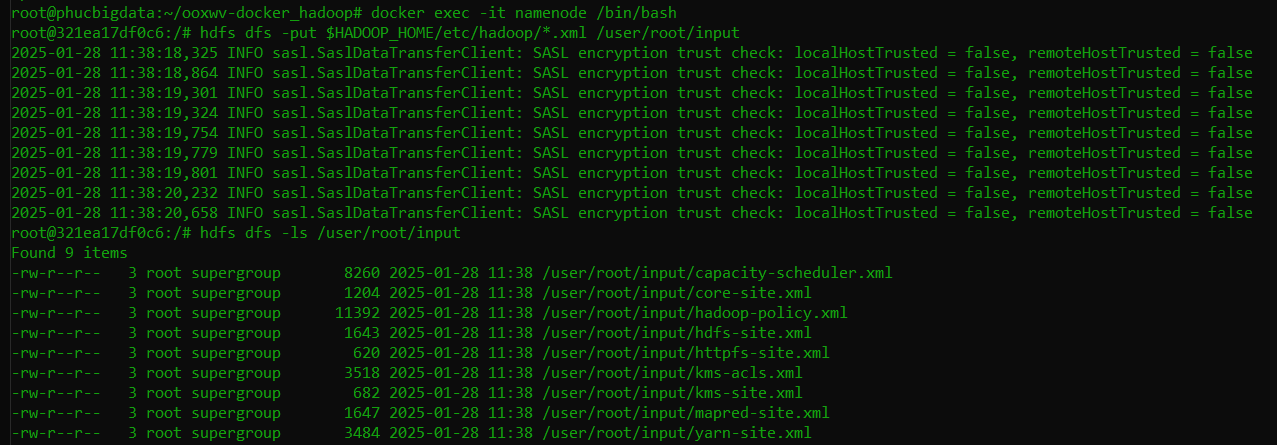
* **git clone** [**https://github.com/ibm-developer-skills-network/ooxwv-docker\_hadoop.git**](https://github.com/ibm-developer-skills-network/ooxwv-docker_hadoop.git)
* **cd ooxwv-docker\_hadoop**
  + 
* **Compose Docker containers:**
  + **docker-compose up -d**
  + **docker ps**
    - 
    - 

1. **Test Hadoop Docker**

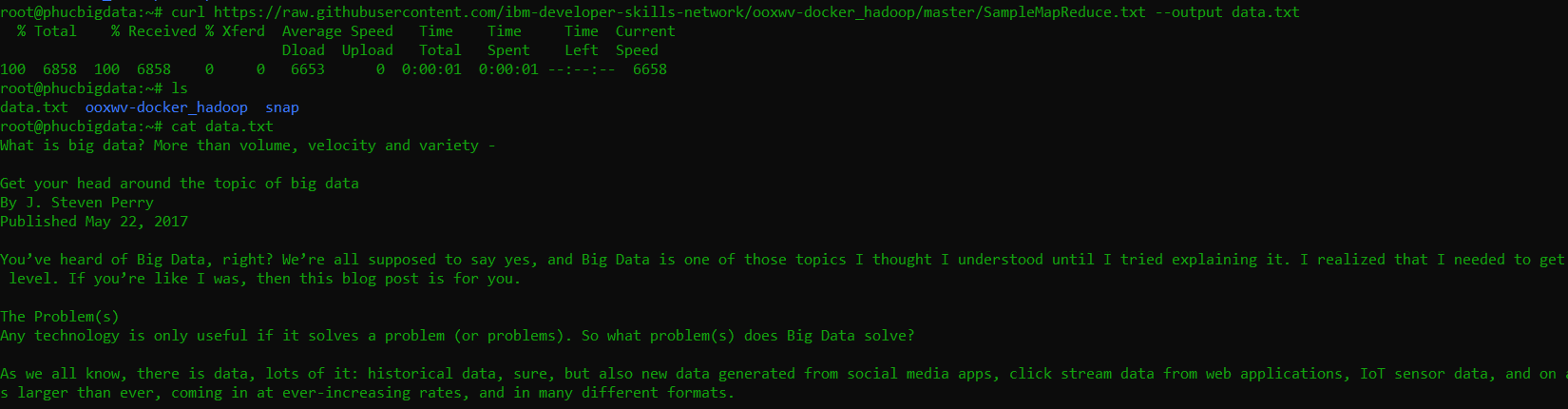
* 

1. **Exercise on HDFS Docker**

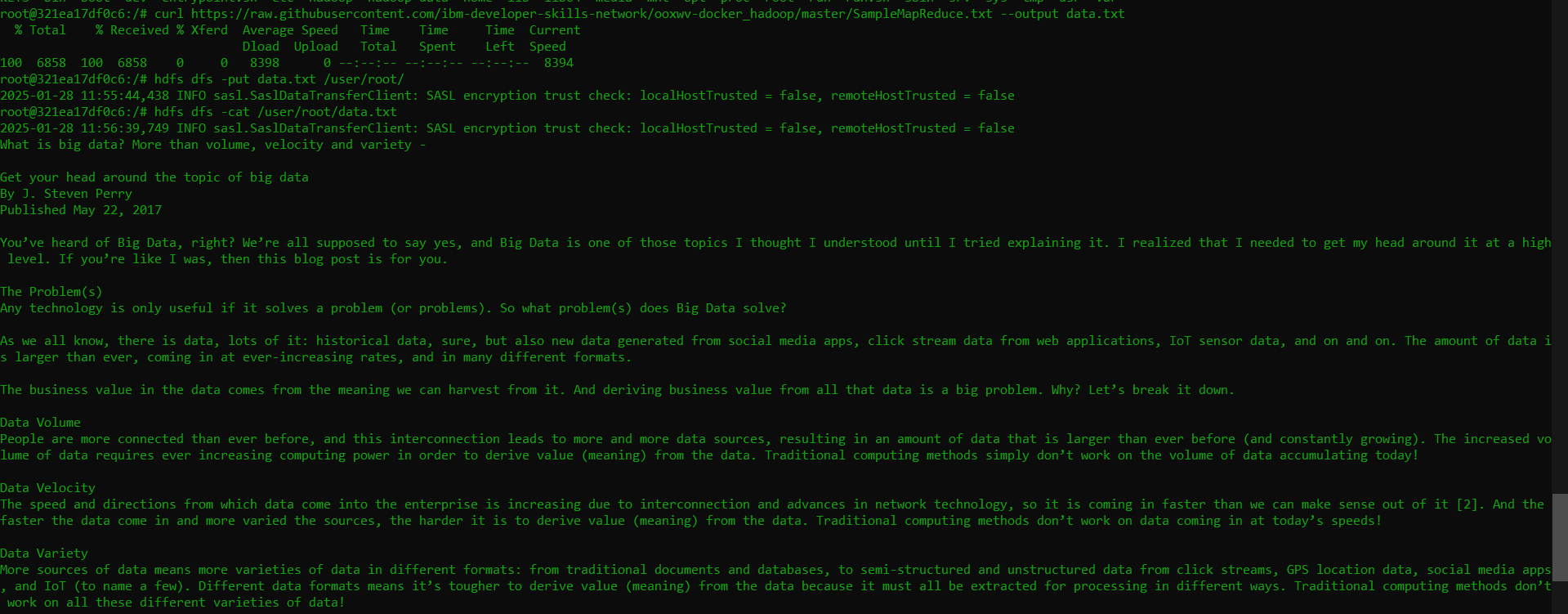
**Must log to docker namenode first**

* **Create folder /user/root/input on HDFS**
  + **hdfs dfs -mkdir -p /user/root/input**
* **Copy all Hadoop configuration files to the input directory**
  + **hdfs dfs -put $HADOOP\_HOME/etc/hadoop/\*.xml /user/root/input**
  + **hdfs dfs -ls /user/root/input**
    - 

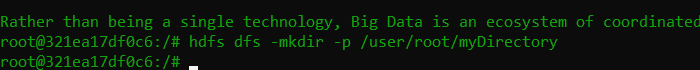
1. **Install file data.txt**

* **curl** [**https://raw.githubusercontent.com/ibm-developer-skills-network/ooxwv-docker\_hadoop/master/SampleMapReduce.txt --output data.txt**](https://raw.githubusercontent.com/ibm-developer-skills-network/ooxwv-docker_hadoop/master/SampleMapReduce.txt%20--output%20data.txt)
  + 

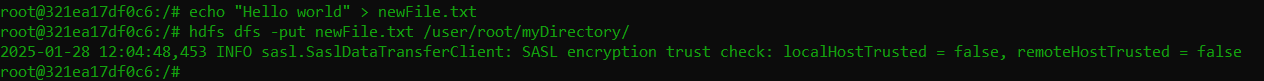
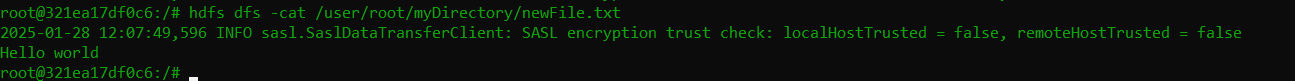
1. **Save data.txt to HDFS**

* **hdfs dfs -put data.txt /user/root/**
* **hdfs dfs -cat /user/root/data.txt**
  + 

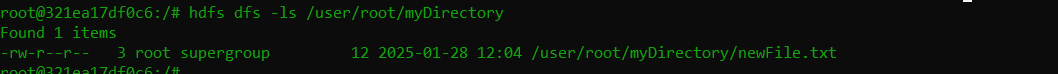
1. **Create new folder on HDFS**

* **hdfs dfs -mkdir -p /user/root/myDirectory**
  + 

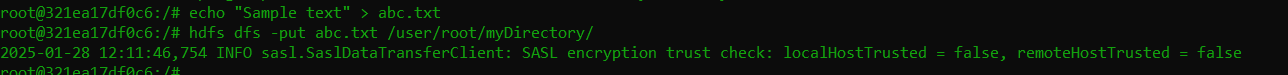
1. **Create the newfile.txt on new folder myDirectory**

* **echo "Hello world" > newFile.txt**
* **hdfs dfs -put newFile.txt /user/root/myDirectory/**
  + 
* **hdfs dfs -cat /user/root/myDirectory/newFile.txt**
  + 

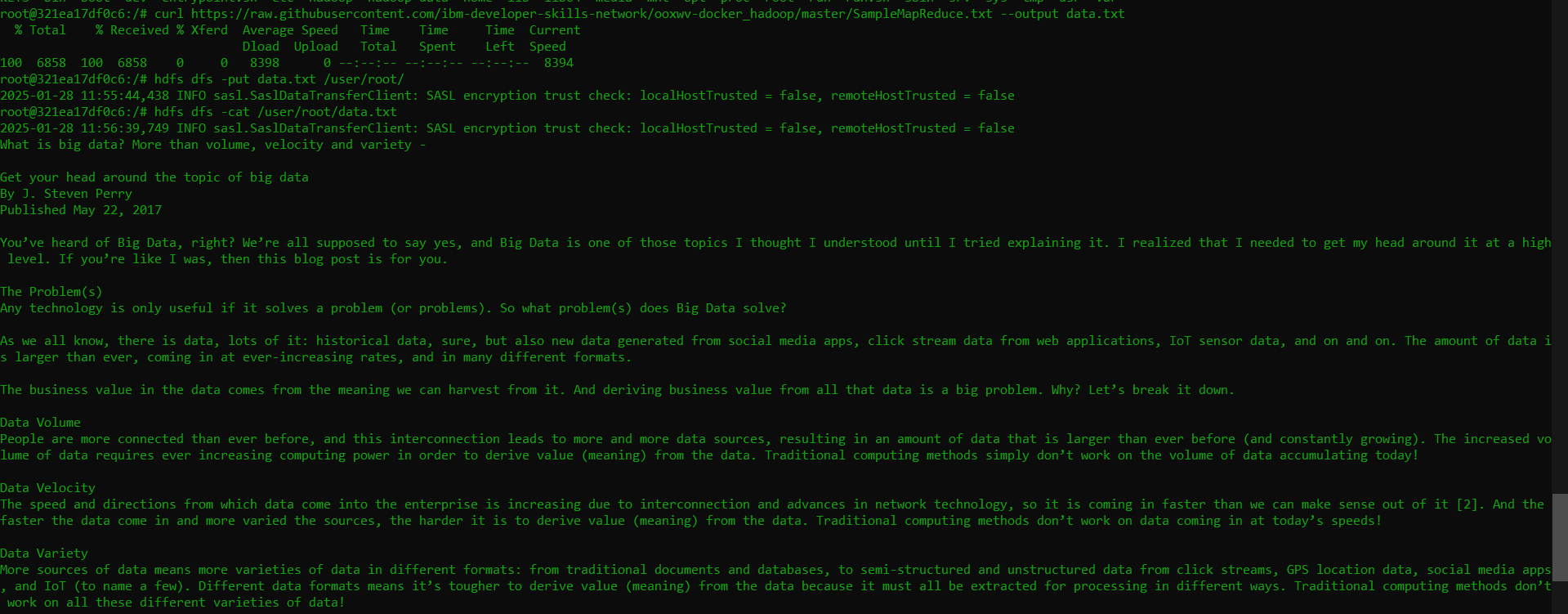
1. **List files and directories in /user/root/myDirectory**

* **hdfs dfs -ls /user/root/myDirectory**
  + 

1. **Create abc.txt file and delete this file**

* **echo "Sample text" > abc.txt**
* **hdfs dfs -put abc.txt /user/root/myDirectory/**
  + 
* **hdfs dfs -rm /user/root/myDirectory/abc.txt**
  + 

1. **Upload files from local system to HDFS**

* **hdfs dfs -put data.txt /user/root/**
* **hdfs dfs -cat /user/root/data.txt**
  + 

1. **Download files from server and save to HDFS**

* **curl** [**https://www.example.com/file.txt --output server\_file.txt**](https://www.example.com/file.txt%20--output%20server_file.txt)
* **hdfs dfs -put server\_file.txt /user/root/**
  + 