**Preparation:** First, I made sure that my Ubuntu Linux system was accurate for the project. I used the liks provided to download and install Ubuntu.

**Java Installation:** Hadoop requires Java to run, so I installed OpenJDK, a free and open-source implementation of the Java Platform:

$ sudo add-apt-repository ppa:openjdk-r/ppa  
$ sudo apt-get update  
$ sudo apt-get install openjdk-8-jre

**Hadoop Download:** Next, I downloaded the Hadoop distribution. I used a tool like wget to download it. For example:

$ cd ~  
$ wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.4/hadoop-  
2.7.4.tar.gz  
$ tar -xzf hadoop-2.7.4.tar.gz

**SSH Setup:** Hadoop relies on SSH for secure communication between nodes. Since SSH is not already set up, I generated an SSH key pair and copied the public key to localhost

$ sudo apt-get install openssh-server

$ ssh-keygen -t rsa -P ""

**Environment Variables:** I needed to set up some environment variables for Hadoop. I opened the .bashrc file for editing. Then, I added the following lines at the end of the file:

export HADOOP\_INSTALL="/home/user/hadoop-2.7.4"  
export PATH=$PATH:$HADOOP\_INSTALL/bin

export PATH=$PATH: $HADOOP\_INSTALL/sbin

**Configuration:** Hadoop requires some configuration to run. I navigated to the Hadoop configuration directory and edited the hadoop-env.sh file. I set the JAVA\_HOME variable in this file to match the Java installation path:

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-i386/

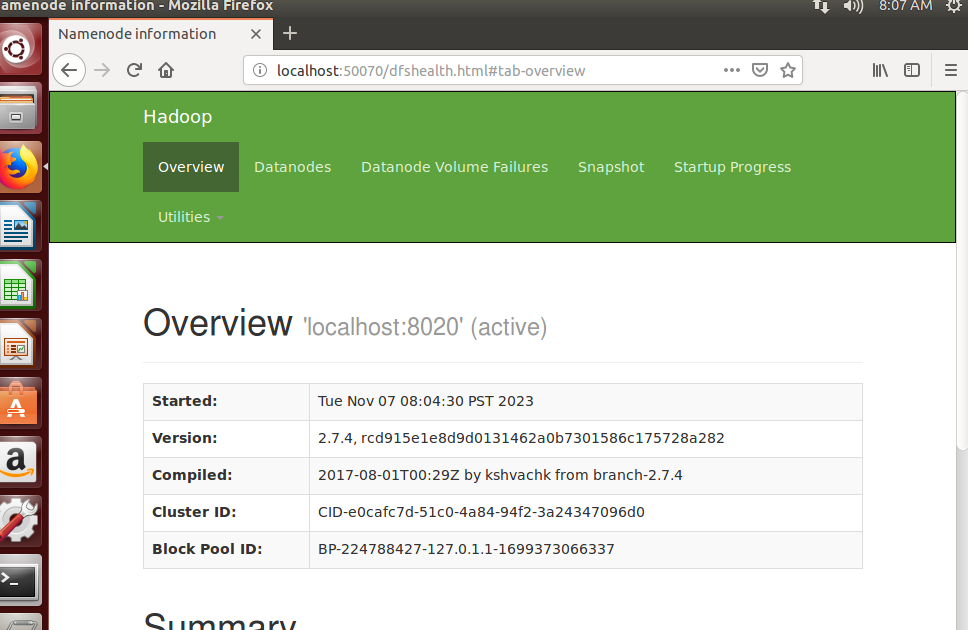
**Hadoop Cluster Setup:** Finally, I configured Hadoop for a single-node setup, which is the simplest configuration for testing and learning. I edited the core-site.xml, hdfs-site.xml, and mapred-site.xml files to set properties such as the Hadoop file system and job tracker address, as per the examples listed in the instructions.

With the setup and configuration complete, I was ready to start using Hadoop on my Ubuntu Linux machine:

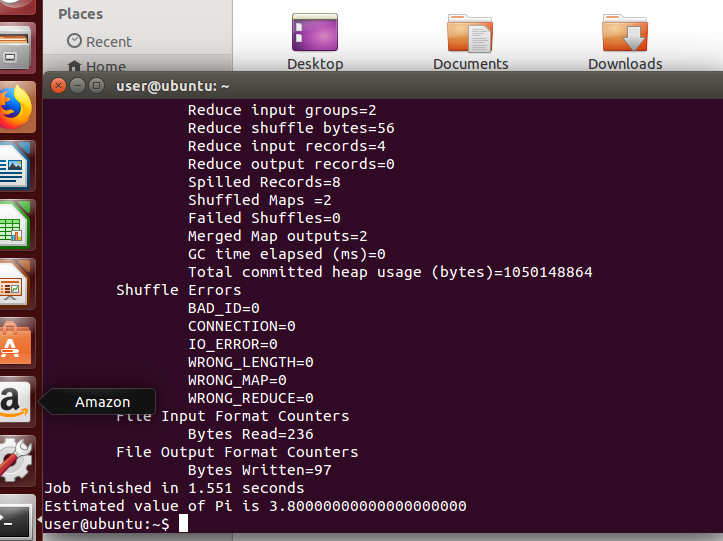
$ hdfs namenode –format

$ start-all.sh

From there, I ran the Pi job, as well as the teragen job, with the screenshots posted below:



Pi job:



Teragen input:



Teragen output:

