CMP406 Task 2 Log Book

2003115 Robbie Sangster

## Creating Cluster

1. Firstly the vm was opened in VMWare and a skeleton was created to be used as the base for the other nodes called hadoopworker1 through 3 with the original being the hadoopmaster node.

A screenshot of a computer

Description automatically generated

## Confiuration

1. Each node had its hostname changed to distinguish each node.



1. Each nodes hosts file was changed to include all node ip addresses

A screenshot of a computer

Description automatically generated

1. Each nodes netplan installer config file was updated with the nodes ip address.

A computer screen shot of a computer

Description automatically generated

This was done on each node with the ip address changing for each node.

1. SSH keys were generated in the master node and were saved on each worker node.



## Node Confiuration

The following files were configured for Hadoop.

Hdfs-site.xml

A screen shot of a computer program

Description automatically generated

Core-site.xml

A computer screen shot of a code

Description automatically generated

Yarn-site.xml

A screen shot of a computer program

Description automatically generated

Mapred-site.xml

A screen shot of a computer program

Description automatically generated

Workers

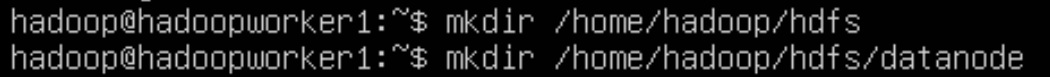
A black screen with white text

Description automatically generated

The files were then ssh to each worker node as well creating a datanode directory.

A screen shot of a computer

Description automatically generated



The master node was format.



And each datanode.



## Working Nodes

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

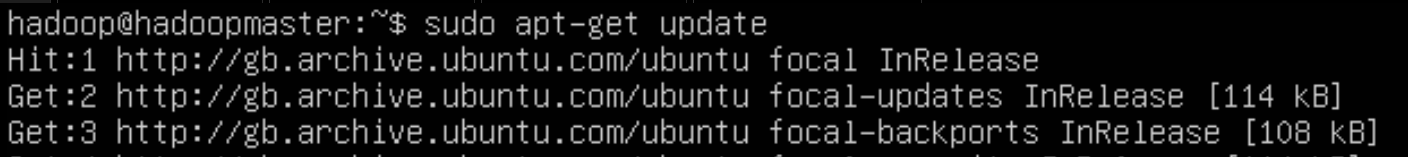
### Pi Example

A screenshot of a computer

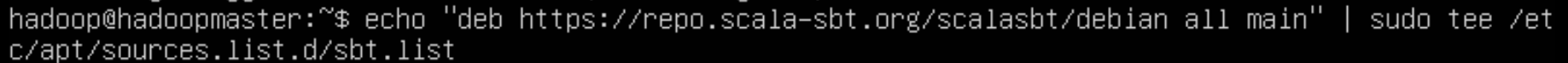
Description automatically generated

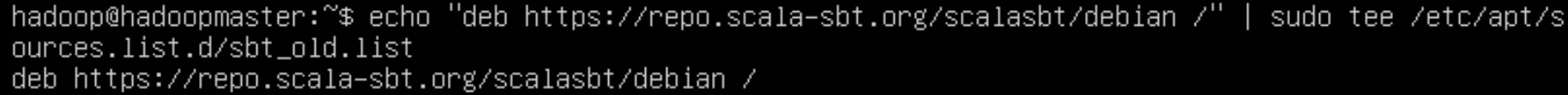


## SBT Setup









A screenshot of a computer

Description automatically generated

## Spark setup

The following lines were added to the .bashrc file.

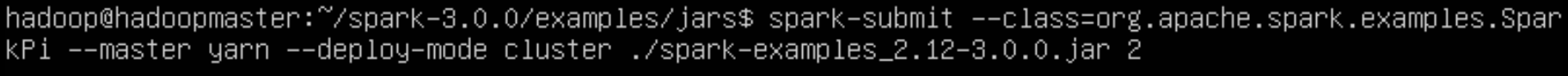
A screenshot of a computer screen

Description automatically generated

And then sourced.



To test spark we will upload the examples files using spark-submit



A screenshot of a computer

Description automatically generated

## Scala Application

The scala application was built in cloudera to test before being moved to the cluster.

Language.sbt file

A screenshot of a computer

Description automatically generated

Language.scala file

A white paper with black text

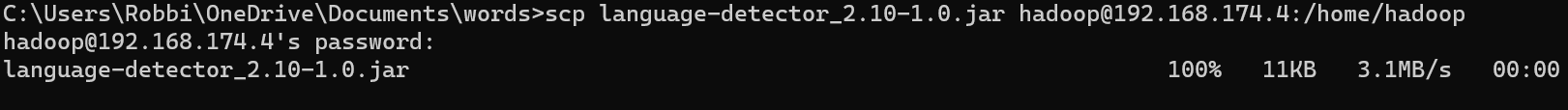
Description automatically generated

## Running Spark

The books, word files and jar file Gutenberg were scp from the host machine to the master.

A computer screen with white text

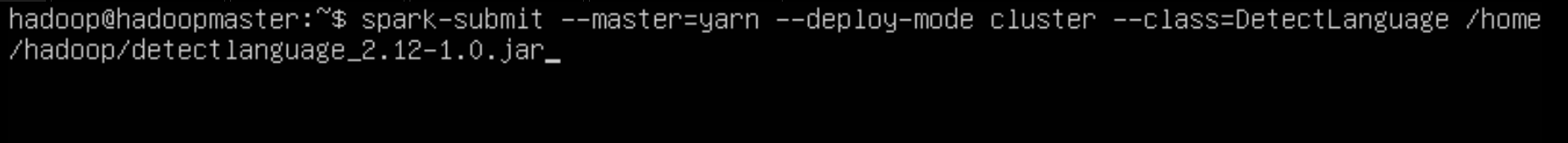
Description automatically generated



These were put in the Hadoop dfs directory.



## Running Application



A black screen with white text

Description automatically generated