Table of Contents

[1. Purpose 2](#_Toc452975387)

[2. Tools Used 2](#_Toc452975388)

[3. Apache Hadoop Single Node setup 2](#_Toc452975389)

[3.1 Prerequisites 2](#_Toc452975390)

[3.2 Installing Software 3](#_Toc452975391)

[4. Run example 3](#_Toc452975392)

[5. Templates and References 5](#_Toc452975393)

# Purpose

This document describes how to set up and configure a single-node Hadoop installation so that you can quickly perform simple operations using Hadoop MapReduce and the Hadoop Distributed File System (HDFS).

# Tools Used

|  |  |
| --- | --- |
| * Cloudera | Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data: The Enterprise Data Hub. Cloudera offers enterprises one place to store, process, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. |

# Apache Hadoop Single Node setup

## Prerequisites

* These 64-bit VMs require a 64-bit host OS and a virtualization product that can support a 64-bit guest OS.
* To use a VMware VM, you must use a player compatible with WorkStation 8.x or higher:
  + Player 4.x or higher
  + Fusion 4.x or higher
* Older versions of WorkStation can be used to create a new VM using the same virtual disk (VMDK file), but some features in VMware Tools are not available.
* The amount of RAM required varies by the run-time option you choose:

|  |  |
| --- | --- |
| **CDH and Cloudera Manager Version** | **RAM Required by VM** |
| CDH 5 (default) | 4+ GiB\* |
| Cloudera Express | 8+ GiB\* |
| Cloudera Enterprise (trial) | 10+ GiB\* |

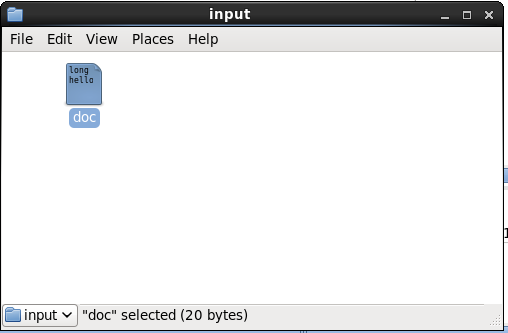
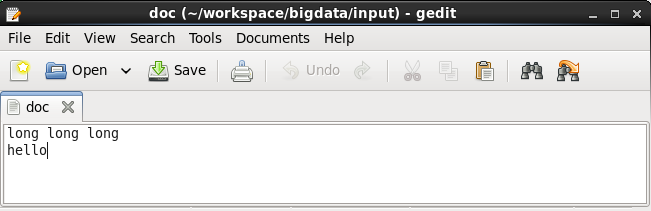
\*Minimum recommended memory. If you are running workloads larger than the examples provided, consider allocating additional memory.

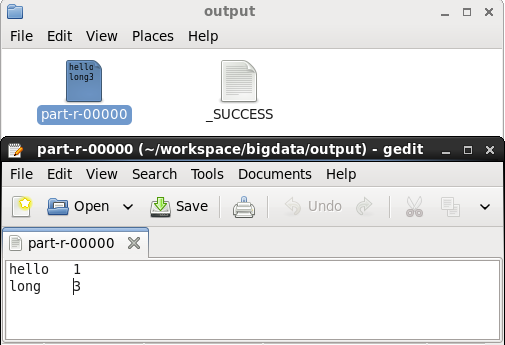
## Installing Software

* Download and install [virtual box](https://www.virtualbox.org/wiki/Downloads)
* Download cloudera virtual machine image for virtual box from [here](http://www.cloudera.com/downloads/quickstart_vms/5-7.html)
* Unzip and import cloudera virtual machine image
* Change the virtual machine setting
  + Increase RAM up to 5gb
  + Set up a shared folder between host and guest machine (It does not work well in Virtual box. Therefore, I ended up enable drag and drop files between host and guest machines)
* Test the WordCount program. I follow step by step guide from [here](http://www.cloudera.com/documentation/other/tutorial/CDH5/Hadoop-Tutorial/ht_wordcount1.html)

# Run example

* Download WordCount.java from sakai or implement WordCount by yourself.
* Create new java project by using eclipse.
* Add references:
  + File system/usr/lib/hadoop/client-0.20
  + File system/usr/lib/hadoop
  + File system/usr/lib/hadoop/lib
* Create folder input inside project and add document file with text.



* Run program.
* Check file in output folder to see result.

# Templates and References

* https://hadoop.apache.org/docs/r1.2.1/single\_node\_setup.html
* http://www.cloudera.com/documentation/enterprise/5-3-x/topics/cdh\_qs\_yarn\_pseudo.html
* http://www.cloudera.com/downloads/quickstart\_vms/5-7.html