Step	Commands	Done (Y/N)?
set VM		
Update OS on newly created VM	sudo apt-get update	
check if java exists	java -version	
install jdk	sudo apt-get install default-jdk	
install jre	sudo apt-get install default-jre	
create separate users & groups	sudo groupadd hdp1313	
	sudo adduser hdp1313 -ingroup hdp1313	
	sudo adduser hdfs1313 -ingroup hdp1313	
	sudo usermod -aG hdp1313 root	
	sudo adduser mprd1313 -ingroup hdp1313	
add to sudo group	sudo usermod -aG sudo hdfs1313	
	sudo usermod -aG sudo hdp1313	
	sudo usermod -aG sudo mprd1313	
update /etc/security/limits.conf	vi /etc/security/limits.conf	
check kernel parameters	sysctl -a	
	sudo vi /etc/sysctl.conf	
	vm.swappiness=0	
	vm.overcommit_memory=1	
	vm.overcommit_ratio=50	
	sudo sysctl -p	
check disk format	df -Th	
check if ssh is installed	which ssh	
install if not installed	sudo apt-get install openssh-server	
	sudo ss -lnp | grep sshd	
	ssh-keygen -t rsa -b 2048	
	cat id_rsa.pub >> authorized_keys	
test ssh to localhost	ssh root@localhost	
download hadoop	wget https://www-us.apache.org/dist/hadoop/common/hadoop-2.7.7/hadoop-2.7.7.tar.gz	
unpack	tar -zxf hadoop-2.7.7.tar.gz	
	sudo chown -R root:root hadoop-2.7.7	
	sudo ln -s hadoop-2.7.7 hadoop	
find where is java	update-alternatives --config java	
set variables in .bashrc	export HADOOP_HOME=/usr/local/hadoop	
	export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop	
	export HADOOP_COMMON_HOME=/usr/local/hadoop	
	export HADOOP_HDFS_HOME=/usr/local/hadoop	
	export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64	
	export CLASSPATH=$CLASSPATH:$JAVA_HOME/lib:$JAVA_HOME/jre/lib	
	export PATH=$JAVA_HOME/bin:$JAVA_HOME/jre/bin:$PATH:$HADOOP_HO	
	export HADOOP_VERSION=2.7.7	
source bashrc	. ~/.bashrc	
test - check versions	java -version	
	hadoop version	
create data directories	mkdir -m 775 /home/hdfs1313/data	
	chown hdfs1313:hdp1313 /home/hdfs1313/data	

	mkdir -m 775 -p /hdata/data	
	chown -R hdfs1313:hdp1313 /hdata	
	chmod -R 775 /hdata	
add in core-site.xml	<property> <name>hadoop.tmp.dir</name> <value>/h	
add in mapred-site.xml	<property><name>mapreduce.framework.name</name><value>yarn</value><	
add in hdfs-site.xml	<property> <name>dfs.replication</name> <value>1</	
add in yarn-site.xml	<property> <name>yarn.nodemanager.aux-services</name>	
add in hadoop-env.sh	export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64	
format namenode	hdfs namenode -format	
start hadoop	start-dfs.sh	
	jps	
	start-yarn.sh	
	jps	
Configure users	hdfs dfs -chmod -R 777 /tmp	
	hdfs dfs -mkdir -p /user/hdp1313	
	hdfs dfs -chown -R hdp1313:hdp1313 /user/hdp1313	
	hdfs dfsadmin -refreshUserToGroupMappings	
Prepare for a test job		
download initial data	mkdir -m 755 ~/scripts	
create hdfs source directory	hdfs dfs -mkdir /userdata	
	hdfs dfs -mkdir /userdata/input	
copy source file onto hdfs	hdfs dfs -put -f /home/hdp1313/src/iliad.csv /userdata/input	
execute test job	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop*examples*.jar w	
Discuss observations		
rerun with a variation	file="/home/hdp1313/src/iliad.csv" curl https://www.gutenberg.org/ebooks/6130.	
execute test job	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop*examples*.jar w	
rerun after fixing	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop*examples*.jar w	

References	Comments
	refer specifications in presentation. Use GUI
a-compliant-cloud-	must be same version on all VMs, down to the micro releasethis is defa
	as root; add "@hdp1313 hard nofiles 32768", no quotes
	as root
	check ssh ports if already installed and if required
	generate keys under /root/.ssh. If running as hadoop user, do it with hac
	in .ssh
	ensure works passwordless
	From under /usr/local
	in /usr/local
	use the value received from command in line 30. Use path until before
)ME/bin:$HADOOP_HOME/sbin:.	
	7 (notice the sub release numbers)
	2.7.7.

ıome/hdfs1313/data</value>	/usr/local/hadoop/etc/hadoop
</property>	/usr/local/hadoop/etc/hadoop
'value> </property>	/usr/local/hadoop/etc/hadoop
<value>mapreduce_s	/usr/local/hadoop/etc/hadoop
export JAVA_HOME=${JAV/	what happens if you let it stay at default
check return code, understar	What happens to the namenode when this is done?What happens to th
check output	
check processes	
check output	
check processes	what is the difference from previous output
	run as root
	in your home directory
	save as big_file.sh under scripts dir
	as root
vordcount /userdata/input /us	observe output
	discuss output; save it to a file for later reference
.txt.utf-8 > $file	what is the variation here?
vordcount /userdata/input /us	what happens when this is run? Why does this happen like this? how to fix this?
vordcount /userdata/input /us	compare output with previous output
	discuss changes, and relate them to theory

ault JDK, check others for other versions. Additionally, cloudra RPM needs Oracle Java RPM due to depend

doop user

jre

e datanode when this is done?

ency

KlhTxx1209$n

		Done (Y/N)?
Step	Commands	pri-node
set VM		
edit /etc/hosts file	which are to be part of hadoop cluster	
Update OS on newly created VM	sudo apt-get update	
check if java exists	java -version	
install jdk	sudo apt-get install default-jdk	
install jre	sudo apt-get install default-jre	
create separate users & groups	sudo groupadd hdp1313	
	sudo adduser hdp1313 -ingroup hdp1313	
	sudo adduser hdfs1313 -ingroup hdp1313	
	sudo usermod -aG hdp1313 root	
	sudo adduser mprd1313 -ingroup hdp1313	
add to sudo group	sudo usermod -aG sudo hdfs1313	
	sudo usermod -aG sudo hdp1313	
	sudo usermod -aG sudo mprd1313	
update /etc/security/limits.conf	vi /etc/security/limits.conf	
check kernel parameters	sysctl -a	
	sudo vi /etc/sysctl.conf	
	vm.swappiness=0	
	vm.overcommit_memory=1	
	vm.overcommit_ratio=50	
	sudo sysctl -p	
check if ssh is installed	which ssh	
install if not installed	sudo apt-get install openssh-server	
	sudo ss -lnp | grep sshd	
	ssh-keygen -t rsa -b 2048	
	cat id_rsa.pub >> authorized_keys	
	test ssh to self on localhost	
copy ssh keys from pri-node to others		
test ssh from pri-node to d-node-a		
copy ssh keys from d-node-a to pri-node		NA
test ssh from d-node-a to pri-node		NA
check disk format	df -Th	
download hadoop	wget https://www-us.apache.org/dist/hadoop/common/hadoop-2.7.7/hadoop-2.7.7.tar.gz	
unpack	tar -zxf hadoop-2.7.7.tar.gz	
	sudo mv hadoop-2.7.7 /usr/local	
	sudo chown -R root:root hadoop-2.7.7/	
	sudo ln -s hadoop-2.7.7 hadoop	
find where is java	update-alternatives --config java	
set variables in .bashrc	export HADOOP_HOME=/usr/local/hadoop	
	export HADOOP_CONF_DIR=/usr/local/hadoop/etc/hadoop	
	export HADOOP_COMMON_HOME=/usr/local/hadoop	
	export HADOOP_HDFS_HOME=/usr/local/hadoop	

	export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64	
	export CLASSPATH=$CLASSPATH:$JAVA_HOME/lib:$JAV	
	export PATH=$JAVA_HOME/bin:$JAVA_HOME/jre/bin:$PA	
	export HADOOP_VERSION=2.7.7	
test - check versions	java -version	
	hadoop version	
create masters file	vi /usr/local/hadoop/etc/hadoop/masters	
update slaves file	vi /usr/local/hadoop/etc/hadoop/slaves	
create data directories	mkdir -m 775 /home/hdfs/data	
	chown hdfs:hadoop /home/hdfs/data	
	mkdir -m 775 -p /hdata/data	
	chown -R hdfs:hadoop /hdata	
	chmod -R 775 /hdata	
add in core-site.xml	<property> <name>hadoop.tmp.dir</name>	
add in mapred-site.xml	<property><name>mapreduce.framework.name</name><va	
add in hdfs-site.xml	<configuration> <property> <name>dfs.replica	
update yarn-site.xml on master	<property> <name>yarn.nodemanager.aux-se	
update yarn-site.xml on slave	<property> <name>yarn.nodemanager.aux-se	
add in hadoop-env.sh	export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64	
format namenode	hdfs namenode -format	
start hadoop	start-dfs.sh	
	jps	
	start-yarn.sh	
	jps	
Prepare for a test job		
	execute big_file.sh	
create hdfs source directory	hdfs dfs -mkdir /userdata	
	hdfs dfs -mkdir /userdata/input	
copy source file onto hdfs	hdfs dfs -put -f ~/src/iliad.csv /userdata/input	
execute test job	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hado	
rerun with a variation	file="/home/akashahuja575/src/iliad.csv" curl https://www.gu	
execute test job	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hado	
rerun after fixing	hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hado	

(Y/N)?		
d-node-a	References	Comments
		refer specifications in presentation. Use GUI
	a-compliant-cloud-	must be same version on all VMs, down to the micro releasethis is defa
		as root; add "@hadoop hard nofiles 32768", no quotes
		as root
		check ssh ports if already installed and if required
		generate keys under /root/.ssh
NA		
NA		
		as root
		in /usr/local

/A_HOME/jre/lib		
TH:$HADOOP_HOME/bin		
		7 (notice the sub release numbers)
		2.7.7.
		final content must be:pri-node (or same dns as entered in /etc/hosts for
		final content must be:pri-noded-node-a (or same dns as entered in /etc/
<value>/home/hdfs1313/data</v		Observe the dfs.replication.factor is now changed to the number of nod
alue>yarn</value></property>		
tion</name>	<value>3</value>	Observe the dfs.replication.factor is now changed to the number of nod
NA	NA	##### is server name of master node
ervices</name>	<value>mapred	##### is server name of master node
	export JAVA_HOME=${JAV/	what happens if you let it stay at default
NA	check return code, understar	What happens to the namenode when this is done?What happens to th
NA	check output	All discussed deamons shall run
	check processes	
NA	check output	
	check processes	Only datanode and nodemanager must run
NA		
NA		
NA		
NA		
NA		observe output
		discuss output; save it to a file for later reference
NA		what is the variation here?
NA		what happens when this is run? Why does this happen like this? how to fix this?
NA		compare output with previous output
		discuss changes, and relate them to theory

ault JDK, check others for other versions. Additionally, cloudra RPM needs Oracle Java RPM due to depend

master node)

/hosts for master and slave node)

lesHow is this different than a single node install? Is that difference necessary?##### is server name of mas

les

e datanode when this is done?

ency

ster node