

How to start

1. Installing Virtual Box

Install VirtualBox from <https://www.virtualbox.org/wiki/Downloads>

2. Downloading Pre-built VirtualBox Image from

https://drive.google.com/file/d/0B4ZS_Ivy-NhcNGIham51clcySmc/view?usp=sharing

Or from rush.ccr.buffalo.edu(FTP) under:

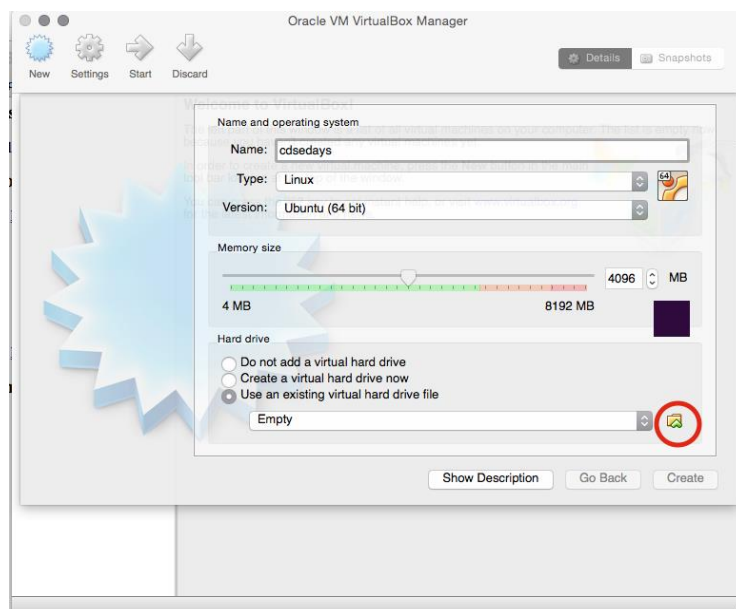
[/gpfscourses/cse587/spring2015/data/hw5/vubuntu.vdi.zip](http://gpfscourses/cse587/spring2015/data/hw5/vubuntu.vdi.zip)

3. Importing Pre-built Image in Virtual Box

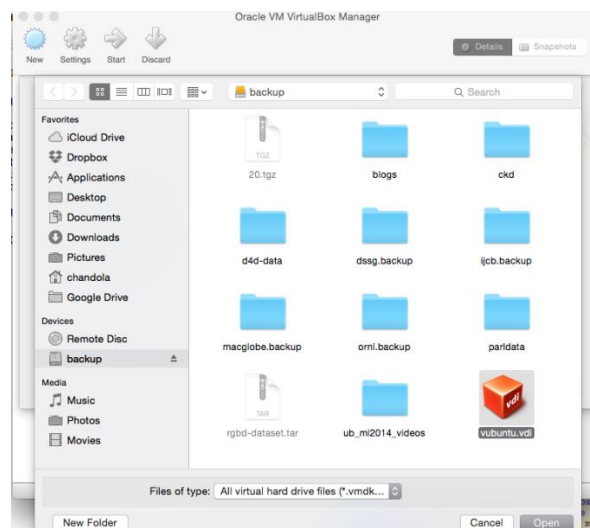
Download/copy and unzip the VDI file to your local machine.

Start VirtualBox and create a new virtual machine.

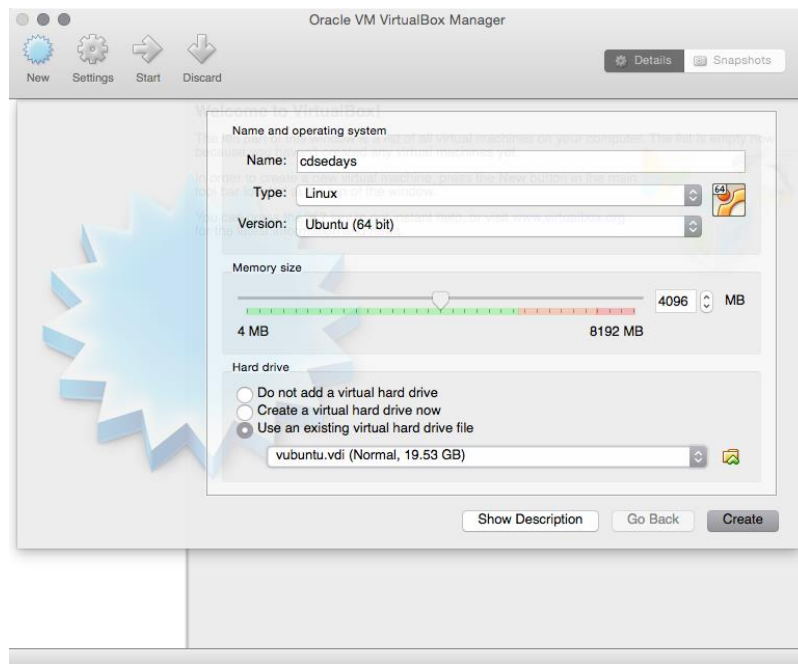
Recommended memory size = 4096 bytes.



When you're asked for a hard disk image, select Use existing hard disk and click on the small icon on the right.



Select the pre-built image.



4. There might have an error while starting Ubuntu:

If you see that, press **f** to fix the issue, and wait until the boot menu comes out, choose the first one "Ubuntu" to continue.

5. The login credentials for the VM are:

Username: **guest**

Password: **cdsedays@123**

- Copy the file "update-script.sh" into VM at **/home/guest/Desktop**
Change permission using:

```
guest@guestvb:~$ chmod 775 /home/guest/Desktop/update-script.sh
```
- In the VM, press **CTRL+ALT+T** to open terminal and init and run everything:

```
guest@guestvb:~$ /home/guest/Desktop/update-script.sh
```

You can download this file from:

<http://www.cse.buffalo.edu/~chandola/workshops/cdsedays-2014/update-script.sh>

Or type the following content:

```
#!/bin/bash

echo "cleaning old files..."

rm -r /home/guest/cdse/accumulo/hadoop-1.0.4/hdfs_data_dir/*
rm -r /home/guest/cdse/accumulo/hadoop-1.0.4/hdfs_name_dir/*
rm -r /home/guest/cdse/accumulo/zookeeper-3.3.6/zoo_data_dir/*

echo "setting proper permissions..."
chmod g-w /home/guest/cdse/accumulo/hadoop-1.0.4/hdfs_data_dir
chmod g-w /home/guest/cdse/accumulo/hadoop-1.0.4/hdfs_name_dir

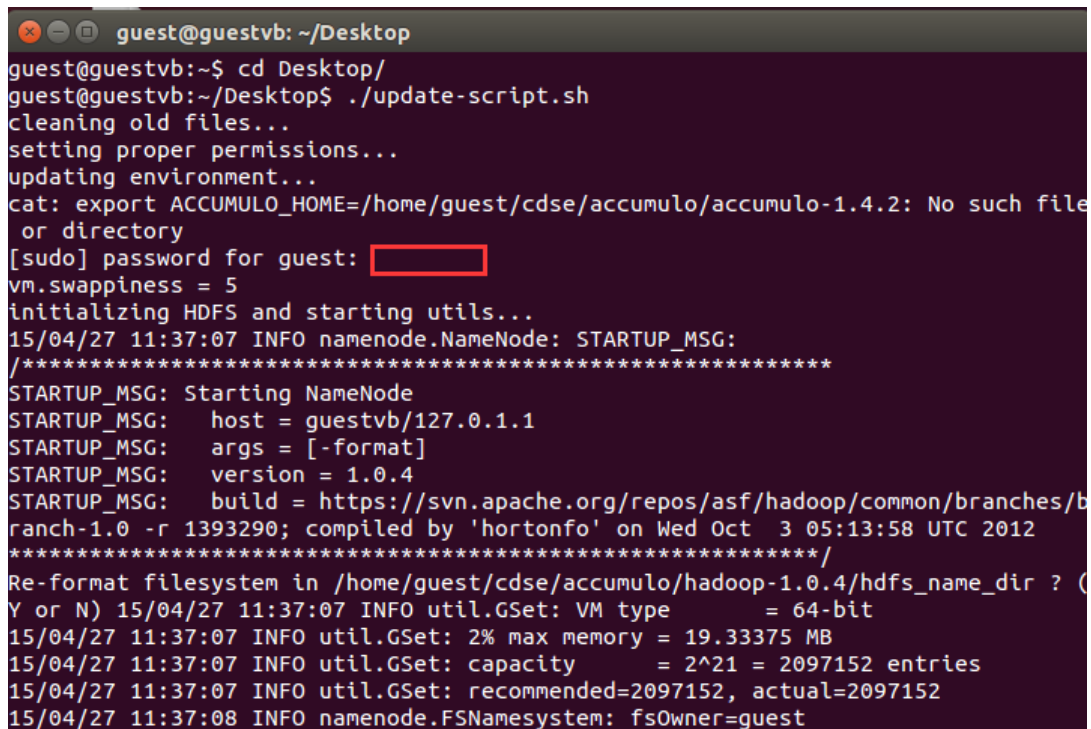
echo "updating environment..."
cat "export ACCUMULO_HOME=/home/guest/cdse/accumulo/accumulo-1.4.2" >> ~/.bashrc
sudo sysctl vm.swappiness=5

echo "initializing HDFS and starting utils..."
/home/guest/cdse/accumulo/hadoop-1.0.4/bin/hadoop namenode -format <<< '$Y\n'
zkServer.sh start
/home/guest/cdse/accumulo/hadoop-1.0.4/bin/start-all.sh

sleep 5

echo "initializing and starting Accumulo"
/home/guest/cdse/accumulo/accumulo-1.4.2/bin/accumulo init <<< '$acc\nacc\nacc\n'
/home/guest/cdse/accumulo/accumulo-1.4.2/bin/start-all.sh
```

- If you have any problem to run the script “**update-script.sh**”, please restart VM completely. If success, type in the password again, you can see the first few lines looks like:



```
guest@guestvb: ~/Desktop
guest@guestvb:~$ cd Desktop/
guest@guestvb:~/Desktop$ ./update-script.sh
cleaning old files...
setting proper permissions...
updating environment...
cat: export ACCUMULO_HOME=/home/guest/cdse/accumulo/accumulo-1.4.2: No such file
or directory
[sudo] password for guest: 
vm.swappiness = 5
initializing HDFS and starting utils...
15/04/27 11:37:07 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = guestvb/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 1.0.4
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.0 -r 1393290; compiled by 'hortonfo' on Wed Oct  3 05:13:58 UTC 2012
*****/
Re-format filesystem in /home/guest/cdse/accumulo/hadoop-1.0.4/hdfs_name_dir ? (
Y or N) 15/04/27 11:37:07 INFO util.GSet: VM type           = 64-bit
15/04/27 11:37:07 INFO util.GSet: 2% max memory = 19.33375 MB
15/04/27 11:37:07 INFO util.GSet: capacity           = 2^21 = 2097152 entries
15/04/27 11:37:07 INFO util.GSet: recommended=2097152, actual=2097152
15/04/27 11:37:08 INFO namenode.FSNamesystem: fsOwner=guest
```

And the last few lines looks like:

```
Starting tablet servers and loggers .... done
Starting logger on localhost
Starting tablet server on localhost
2015-04-27 11:37:35,267 [server.Accumulo] INFO : Attempting to talk to zookeeper
2015-04-27 11:37:35,682 [server.Accumulo] INFO : Zookeeper connected and initial
ized, attempting to talk to HDFS
2015-04-27 11:37:38,857 [server.Accumulo] INFO : Connected to HDFS
Starting master on localhost
Starting garbage collector on localhost
Starting monitor on localhost
Starting tracer on localhost
guest@guestvb:~/Desktop$
```

- Then to go the following path and start to run examples
guest@guestvb:~\$ **cd /home/guest/cdse/accumulo/accumulo-1.4.2**

Example 1 - Hello World

<https://accumulo.apache.org/1.4/examples/helloworld.html>

Under path **/home/guest/cdse/accumulo/accumulo-1.4.2**

Login Accumulo shell:

```
guest@guestvb:~$ bin/accumulo shell -u root -p acc
```

Create a table called 'hellotable':

```
root@acc> createtable hellotable
```

Logout Accumulo shell:

```
root@acc> quit
```

Launch a Java program that inserts data with a BatchWriter:

```
guest@guestvb:~$
```

```
bin/accumulo org.apache.accumulo.examples.simple.helloworld.InsertWithBatchWriter acc  
guestvb root acc hellotable
```

you should see 50K entries from browser:

```
localhost:50095
```

To view the entries, go back to the shell to scan the table:

```
root@acc> table hellotable
```

```
root@acc hellotable> scan
```

Example 2 - Visibility, Authorizations, and Permissions

<https://accumulo.apache.org/1.4/examples/visibility.html>

- Creating a new user and Granting permissions to a user

```
root@acc> createuser username
Enter new password for 'username': 123456
Please confirm new password for 'username': 123456
root@acc> user username
Enter password for user username: 123456
username@acc> createtable vistest
06 10:48:47,931 [shell.Shell] ERROR: org.apache.accumulo.core.client.AccumuloSecurityException:
Error PERMISSION_DENIED - User does not have permission to perform this action
username@acc> userpermissions
System permissions:
Table permissions (!METADATA): Table.READ
username@acc>

username@acc> user root
Enter password for user root: acc
root@acc> grant -s System.CREATE_TABLE -u username
root@acc> user username
Enter password for user username: 123456
username@acc> createtable vistest
username@acc> userpermissions
System permissions: System.CREATE_TABLE
Table permissions (!METADATA): Table.READ
Table permissions (vistest): Table.READ, Table.WRITE, Table.BULK_IMPORT, Table.ALTER_TABLE,
Table.GRANT, Table.DROP_TABLE
username@acc vistest>
```

- Inserting data with visibilities

```
username@acc vistest> insert row f1 q1 v1 -l A
username@acc vistest> insert row f2 q2 v2 -l A&B
username@acc vistest> insert row f3 q3 v3 -l apple&carrot|broccoli|spinach
06 11:19:01,432 [shell.Shell] ERROR: org.apache.accumulo.core.util.BadArgumentException: cannot
mix | and & near index 12 apple&carrot|broccoli|spinach
      ^
username@acc vistest> insert row f3 q3 v3 -l (apple&carrot)|broccoli|spinach
username@acc vistest>
```

- Scanning with authorizations and setting authorizations for a user

```
username@acc vistest> setauths -s A
06 11:53:42,056 [shell.Shell] ERROR: org.apache.accumulo.core.client.AccumuloSecurityException:
Error PERMISSION_DENIED - User does not have permission to perform this action
username@acc vistest>
```

```
username@acc vistest> user root
Enter password for user root: acc
root@acc vistest> setauths -s A -u username
root@acc vistest> user username
Enter password for user username: 123456
username@acc vistest> scan -s A
row f1:q1 [A]      v1
username@acc vistest> scan
row f1:q1 [A]      v1
username@acc vistest>
```

The default authorizations for a scan are the user's entire set of authorizations.

```
username@acc vistest> user root
Enter password for user root: acc
root@acc vistest> setauths -s A,B,broccoli -u username
root@acc vistest> user username
Enter password for user username: 123456
username@acc vistest> scan
row f1:q1 [A]      v1
row f2:q2 [A&B]     v2
row f3:q3 [(apple&carrot)|broccoli|spinach] v3
username@acc vistest> scan -s B
username@acc vistest>
```

If you want, you can limit a user to only be able to insert data which they can read themselves. It can be set with the following constraint.

```
username@acc vistest> user root
Enter password for user root: acc
root@acc vistest> config -t vistest -s table.constraint.1=org.apache.accumulo.core.security.
VisibilityConstraint
```

```
root@acc vistest> user username
Enter password for user username: 123456
username@acc vistest> insert row f4 q4 v4 -l spinach
```

Constraint Failures:

ConstraintViolationSummary(constrainClass:org.apache.accumulo.core.security.VisibilityConstraint, violationCode:2, violationDescription:User does not have authorization on column visibility, numberOfViolatingMutations:1)

```
username@acc vistest> insert row f4 q4 v4 -l spinach | broccoli
username@acc vistest> scan
row f1:q1 [A]      v1
row f2:q2 [A&B]     v2
row f3:q3 [(apple&carrot)|broccoli|spinach] v3
row f4:q4 [spinach|broccoli]      v4
username@acc vistest>
```

Example 3 – MapReduce, WordCount

<https://accumulo.apache.org/1.4/examples/mapred.html>

Under path **/home/guest/cdse/accumulo/accumulo-1.4.2**

```
guest@guestvb:~/cdse/accumulo/accumulo-1.4.2$
```

```
hadoop fs -copyFromLocal /home/guest/cdse/accumulo/accumulo-1.4.2/README /input
```

```
guest@guestvb:~/cdse/accumulo/accumulo-1.4.2$ hadoop fs -ls /
```

Found 3 items

```
drwxr-xr-x - guest supergroup 0 2015-04-27 11:37 /accumulo
```

```
-rw-r--r-- 1 guest supergroup 12449 2015-04-27 12:44 /input
```

```
drwxr-xr-x - guest supergroup 0 2015-04-27 11:37 /tmp
```

```
guest@guestvb:~/cdse/accumulo/accumulo-1.4.2$ accumulo shell -u root -p acc
```

Shell - Apache Accumulo Interactive Shell

- - version: 1.4.2

- instance name: acc

- instance id: c963a61a-902c-4d70-877f-2636670cf392

- - type 'help' for a list of available commands

```
root@acc> createtable wordCount
```

```
root@acc wordCount> setiter -class org.apache.accumulo.core.iterators.user.SummingCombiner -p 10 -t wordCount -majc -minc -scan
```

SummingCombiner interprets Values as Longs and adds them together. A variety of encodings (variable length, fixed length, or string) are available

-----> set SummingCombiner parameter all, set to true to apply Combiner to every column, otherwise leave blank. if true, columns option will be ignored.: **true**

-----> set SummingCombiner parameter columns, <col fam>[:<col qual>]{,<col fam>[:<col qual>]} escape non-alphanum chars using %<hex>.: **<press enter>**

-----> set SummingCombiner parameter lossy, if true, failed decodes are ignored. Otherwise combiner will error on failed decodes (default false): <TRUE|FALSE>: **<press enter>**

-----> set SummingCombiner parameter type, <VARLEN|FIXEDLEN|STRING|fullClassName>: **STRING**

```
root@acc wordCount> quit
```

```
guest@guestvb:~/cdse/accumulo/accumulo-1.4.2$
```

```
bin/tool.sh lib/examples-simple-1.4.2.jar org.apache.accumulo.examples.simple.mapreduce.
```

```
WordCount acc guestvb /input wordCount -u root -p acc
```

```
15/04/27 12:58:16 INFO mapred.JobClient: Running job: job_201504271137_0003
```

```
15/04/27 12:58:17 INFO mapred.JobClient: map 0% reduce 0%
```

```
15/04/27 12:58:34 INFO mapred.JobClient: map 100% reduce 0%
```

```
15/04/27 12:58:39 INFO mapred.JobClient: Job complete: job_201504271137_0003
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Counters: 18
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Job Counters
```

```
15/04/27 12:58:40 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=16683
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Launched map tasks=1
```

```
15/04/27 12:58:40 INFO mapred.JobClient: Data-local map tasks=1
15/04/27 12:58:40 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=0
15/04/27 12:58:40 INFO mapred.JobClient: File Output Format Counters
15/04/27 12:58:40 INFO mapred.JobClient: Bytes Written=0
15/04/27 12:58:40 INFO mapred.JobClient: FileSystemCounters
15/04/27 12:58:40 INFO mapred.JobClient: HDFS_BYTES_READ=12541
15/04/27 12:58:40 INFO mapred.JobClient: FILE_BYTES_WRITTEN=29267
15/04/27 12:58:40 INFO mapred.JobClient: File Input Format Counters
15/04/27 12:58:40 INFO mapred.JobClient: Bytes Read=12449
15/04/27 12:58:40 INFO mapred.JobClient: Map-Reduce Framework
15/04/27 12:58:40 INFO mapred.JobClient: Map input records=284
15/04/27 12:58:40 INFO mapred.JobClient: Physical memory (bytes) snapshot=104456192
15/04/27 12:58:40 INFO mapred.JobClient: Spilled Records=0
15/04/27 12:58:40 INFO mapred.JobClient: CPU time spent (ms)=1310
15/04/27 12:58:40 INFO mapred.JobClient: Total committed heap usage (bytes)=62717952
15/04/27 12:58:40 INFO mapred.JobClient: Virtual memory (bytes) snapshot=629846016
15/04/27 12:58:40 INFO mapred.JobClient: Map output records=1853
15/04/27 12:58:40 INFO mapred.JobClient: SPLIT_RAW_BYTES=92
```

- Go back to scan the table

```
guest@guestvb:~/cdse/accumulo/accumulo-1.4.2$ accumulo shell -u root -p acc
```

```
Shell - Apache Accumulo Interactive Shell
```

```
- - version: 1.4.2
```

```
- instance name: acc
```

```
- instance id: c963a61a-902c-4d70-877f-2636670cf392
```

```
- - type 'help' for a list of available commands
```

```
root@acc> table wordCount
```

```
root@acc wordCount> scan
```

```
count:20080906 []      120
"addprinc count:20080906 []      1
"bin/accumulo count:20080906 []      2
"bin/stop-all.sh" count:20080906 []      1
"createtable count:20080906 []      1
"links": count:20080906 []      1
"logger.dir.walog" count:20080906 []      1
"masters" count:20080906 []      1
"mvn count:20080906 []      3
"native-standalone". count:20080906 []      1
"scan" count:20080906 []      1
"slaves" count:20080906 []      1
"standalone". count:20080906 []      1
"table count:20080906 []      1
```

```
.....
.....
```