Atmosphere
RESEARCH GROUP

Big Data Step-by-Step

Boston Predictive Analytics Big Data Workshop

Microsoft New England Research & Development Center, Cambridge, MA

Saturday, March 10, 2012

http://atms.gr/bigdata0310

by **Jeffrey Breen**

President and Co-Founder
Atmosphere Research Group
email: jeffrey@atmosgrp.com
Twitter: @JeffreyBreen

Big Data Infrastructure

Part 1: Local VM

Code & more on github:

https://github.com/jeffreybreen/tutorial-201203-big-data

Overview

- Download and install a virtual machine containing a configured and working version of Hadoop
- Install R
- Copy some data into the HDFS
- Test our installation by running some small Hadoop jobs
- Extra credit: install RStudio

Thank you, Cloudera

- Cloudera's Hadoop Demo VM provides everything you need to run small jobs in a virtual environment
- Hadoop 0.20 + Flume, HBase, Hive, Hue, Mahout, Oozie, Pig, Sqoop, Whirr, Zookeeper
- Based on CentOS 5.7 & available for VMware, KVM and VirtualBox:

  https://ccp.cloudera.com/display/SUPPORT/Cloudera%27s+Hadoop+Demo+VM

- Older versions came with training exercises, but fortunately they're still available on github:

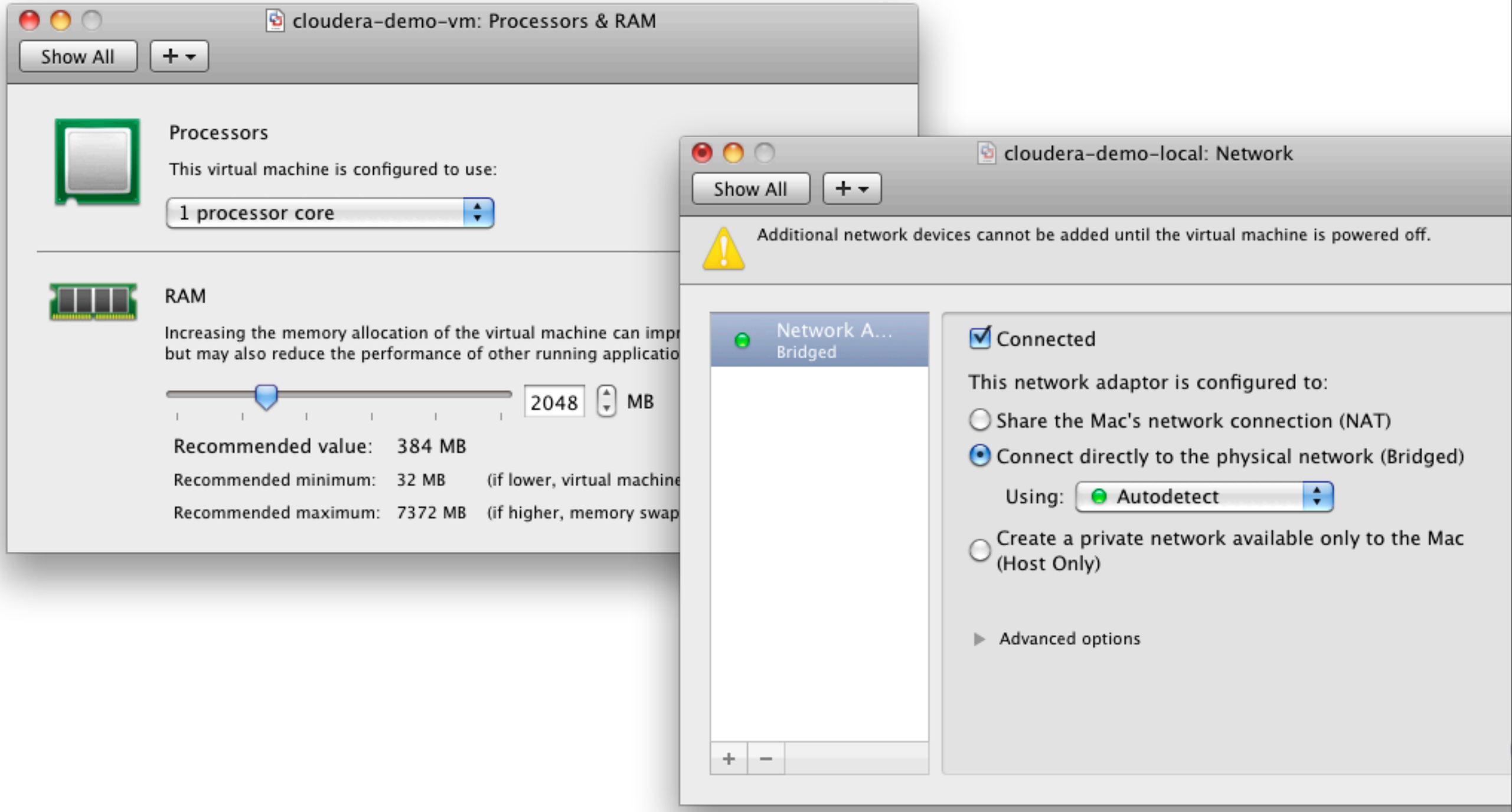
  https://github.com/cloudera/cloudera-training

- Provides a common base which we will use for our later cluster, etc. work

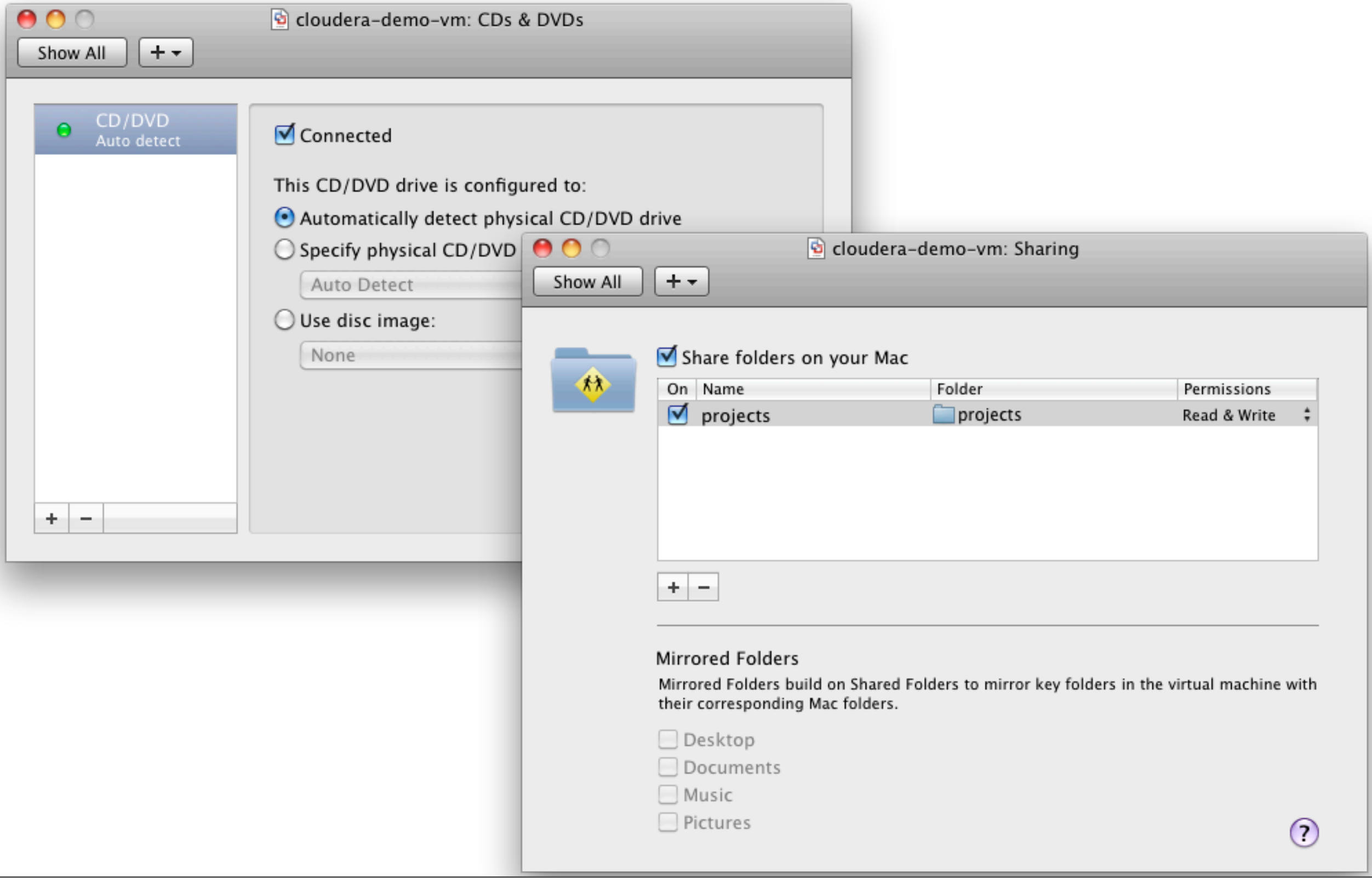
A couple of tweaks

- Give it more RAM
  - uses 1GB by default
  - not configured with a swap file
- Use Bridged networking vs. NAT or Host-only
  - Virtual machine will get its own IP address on your network
  - Experienced DNS errors with whirr while sharing an IP
- Extras: Set up shared folders & add a CD-ROM
  - Shared folders make it easy to share data & code between your computer and the VM
  - Add a CD-ROM drive if you want to install VMware tools or any ISO file

Important


cloudera-demo-vm: Processors & RAM
Show All
Processors
This virtual machine is configured to use:
1 processor core
RAM
Increasing the memory allocation of the virtual machine can impr
but may also reduce the performance of other running applicatio
2048
MB
Recommended value: 384 MB
Recommended minimum: 32 MB (if lower, virtual machine
Recommended maximum: 7372 MB (if higher, memory swap
cloudera-demo-local: Network
Show All
Additional network devices cannot be added until the virtual machine is powered off.
Network A...
Bridged
Connected
This network adaptor is configured to:
Share the Mac's network connection (NAT)
Connect directly to the physical network (Bridged)
Using: Autodetect
Create a private network available only to the Mac (Host Only)
Advanced options

Nice to have


cloudera-demo-vm: CDs & DVDs
Show All
CD/DVD
Auto detect
Connected
This CD/DVD drive is configured to:
Automatically detect physical CD/DVD drive
Specify physical CD/DVD
Auto Detect
Use disc image:
None
cloudera-demo-vm: Sharing
Show All
Share folders on your Mac
On
Name
Folder
Permissions
projects
projects
Read & Write
Mirrored Folders
Mirrored Folders build on Shared Folders to mirror key folders in the virtual machine with their corresponding Mac folders.
Desktop
Documents
Music
Pictures

Yes, it's that easy

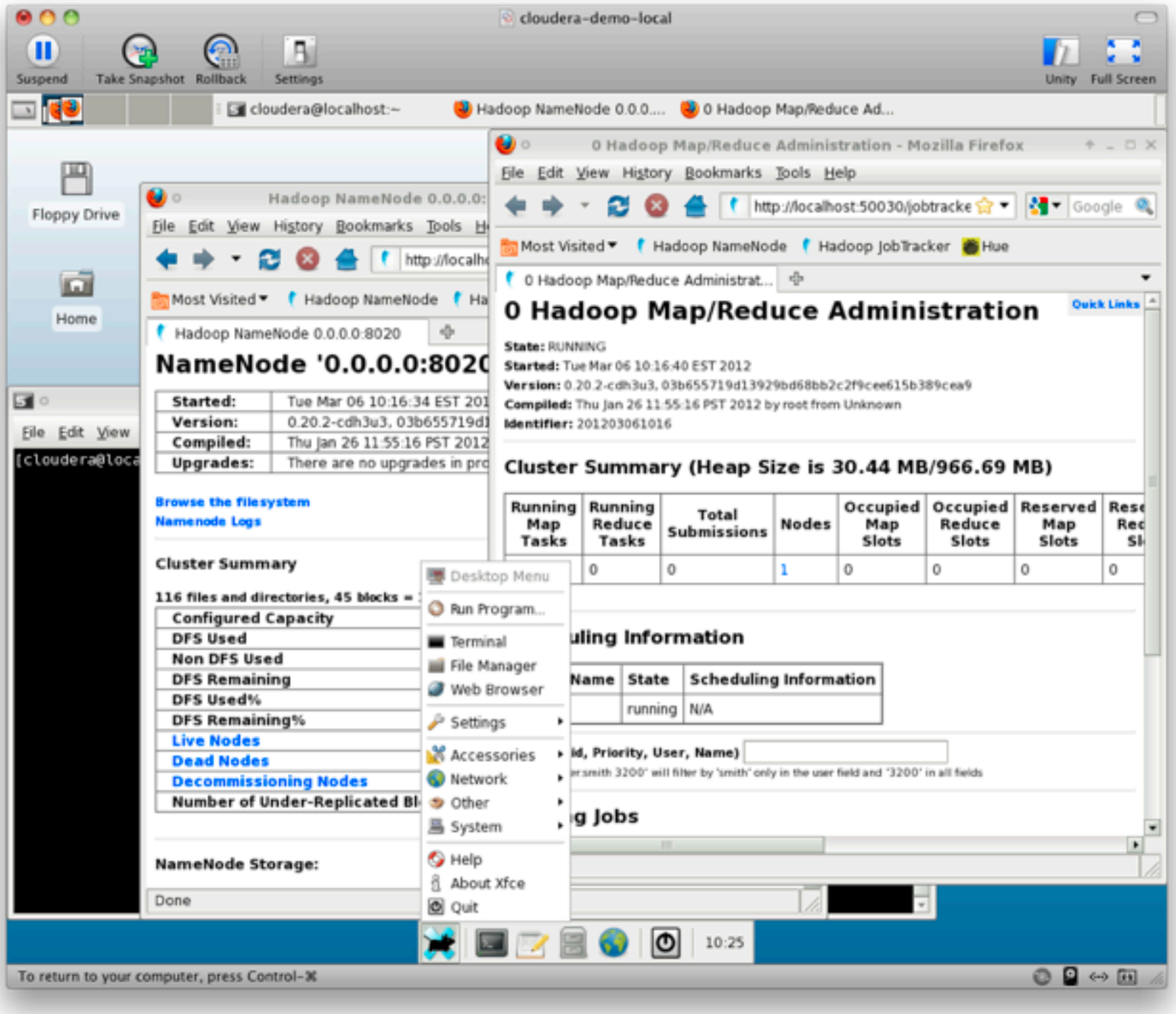
Boot VM and log in as "cloudera". (Password = "cloudera" too)

Execute as root with "`sudo`"

"`sudo su -`" for root shell

Hadoop already running

Firefox contains bookmarks to admin pages



Well, almost.

- Install VMware tools and link to shared folder on host PC

```
$ sudo mkdir /mnt/vmware
$ sudo mount /dev/hda /mnt/vmware
$ tar zxf /mnt/vmware/VMwareTools-8.4.7-416484.tar.gz
$ cd vmware-tools-distrib/
$ sudo ./vmware-install.pl
$ ln -s /mnt/hgfs/projects/tutorial-201203-big-data/ ~/.
```

- Install handy utilities (wget, git)

```
$ sudo yum -y install wget git
```

- Install EPEL repository

```
$ sudo rpm -Uvh http://dl.fedoraproject.org/pub/epel/5/x86_64/epel-release-5-4.noarch.rpm
```

- Install R from EPEL

```
$ sudo yum -y install R
```

- Set Hadoop environment variables *(workaround for CDH3u3 VM)*

```
$ sudo ln -s /etc/default/hadoop-0.20 /etc/profile.d/hadoop-0.20.sh
$ cat /etc/default/hadoop-0.20 | sed 's/export //g' > ~/.Renviron
```

Warning: Pages of fast-scrolling gibberish to follow

But it's all going to be OK

```
[cloudera@localhost ~]$ sudo mkdir /mnt/vmware
[cloudera@localhost ~]$ sudo mount /dev/hda /mnt/vmware
mount: block device /dev/hda is write-protected, mounting read-only
[cloudera@localhost ~]$ tar zxf /mnt/vmware/VMwareTools-8.4.7-416484.tar.gz
[cloudera@localhost ~]$ cd vmware-tools-distrib/
[cloudera@localhost vmware-tools-distrib]$ sudo ./vmware-install.pl
Creating a new VMware Tools installer database using the tar4 format.

Installing VMware Tools.

In which directory do you want to install the binary files?
[/usr/bin]

What is the directory that contains the init directories (rc0.d/ to rc6.d/)?
[/etc/rc.d]

What is the directory that contains the init scripts?
[/etc/rc.d/init.d]

In which directory do you want to install the daemon files?
[/usr/sbin]

In which directory do you want to install the library files?
[/usr/lib/vmware-tools]

The path "/usr/lib/vmware-tools" does not exist currently. This program is
going to create it, including needed parent directories. Is this what you want?
[yes]

In which directory do you want to install the documentation files?
[/usr/share/doc/vmware-tools]

The path "/usr/share/doc/vmware-tools" does not exist currently. This program
is going to create it, including needed parent directories. Is this what you
want? [yes]

The installation of VMware Tools 8.4.7 build-416484 for Linux completed
successfully. You can decide to remove this software from your system at any
time by invoking the following command: "/usr/bin/vmware-uninstall-tools.pl".

Before running VMware Tools for the first time, you need to configure it by
invoking the following command: "/usr/bin/vmware-config-tools.pl". Do you want
this program to invoke the command for you now? [yes]

Initializing...

Making sure services for VMware Tools are stopped.

Stopping VMware Tools services in the virtual machine:
   Guest operating system daemon:                          [  OK  ]
   Virtual Printing daemon:                                [  OK  ]
   Unmounting HGFS shares:                                 [  OK  ]
   Guest filesystem driver:                                [  OK  ]


Found a compatible pre-built module for vmmemctl.  Installing it...


Found a compatible pre-built module for vmhgfs.  Installing it...
```

```
[cloudera@localhost ~]$ sudo yum -y install wget git
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * base: mirror.symnds.com
 * epel: mirror.symnds.com
 * extras: mirrors.einstein.yu.edu
 * updates: mirror.symnds.com
epel                                                                                                                | 3.4 kB     00:00
epel/primary_db                                                                                                     | 3.7 MB     00:01
Setting up Install Process
Resolving Dependencies
There are unfinished transactions remaining. You might consider running yum-complete-transaction first to finish them.
The program yum-complete-transaction is found in the yum-utils package.
--> Running transaction check
---> Package git.x86_64 0:1.7.4.1-1.el5 set to be updated
--> Processing Dependency: perl-Git = 1.7.4.1-1.el5 for package: git
--> Processing Dependency: perl(Error) for package: git
--> Processing Dependency: perl(Git) for package: git
---> Package wget.x86_64 0:1.11.4-2.el5_4.1 set to be updated
--> Running transaction check
---> Package perl-Error.noarch 1:0.17010-1.el5 set to be updated
---> Package perl-Git.x86_64 0:1.7.4.1-1.el5 set to be updated
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================================================================
 Package                        Arch                        Version                             Repository                 Size
================================================================================================================================
Installing:
 git                            x86_64                      1.7.4.1-1.el5                       epel                      4.5 M
 wget                           x86_64                      1.11.4-2.el5_4.1                    base                      582 k
Installing for dependencies:
 perl-Error                     noarch                      1:0.17010-1.el5                     epel                       26 k
 perl-Git                       x86_64                      1.7.4.1-1.el5                       epel                       28 k

Transaction Summary
================================================================================================================================
Install       4 Package(s)
Upgrade       0 Package(s)

Total download size: 5.1 M
Downloading Packages:
(1/4): perl-Error-0.17010-1.el5.noarch.rpm                                                          |  26 kB     00:00
(2/4): perl-Git-1.7.4.1-1.el5.x86_64.rpm                                                            |  28 kB     00:00
(3/4): wget-1.11.4-2.el5_4.1.x86_64.rpm                                                             | 582 kB     00:00
(4/4): git-1.7.4.1-1.el5.x86_64.rpm                                                                 | 4.5 MB     00:01
--------------------------------------------------------------------------------------------------------------------------------
Total                                                                                    2.6 MB/s | 5.1 MB     00:02
warning: rpmts_HdrFromFdno: Header V3 DSA signature: NOKEY, key ID 217521f6
epel/gpgkey                                                                                         | 1.7 kB     00:00
Importing GPG key 0x217521F6 "Fedora EPEL <epel@fedoraproject.org>" from /etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL
Running rpm_check_debug
Running Transaction Test
Finished Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing     : wget                                                                                                1/4
  Installing     : perl-Error                                                                                          2/4
  Installing     : git                                                                                                 3/4
  Installing     : perl-Git                                                                                            4/4

Installed:
  git.x86_64 0:1.7.4.1-1.el5                                    wget.x86_64 0:1.11.4-2.el5_4.1

Dependency Installed:
  perl-Error.noarch 1:0.17010-1.el5                                perl-Git.x86_64 0:1.7.4.1-1.el5

Complete!
```

```
[cloudera@localhost ~]$ sudo rpm -Uvh http://dl.fedoraproject.org/pub/epel/5/x86_64/epel-release-5-4.noarch.rpm
Retrieving http://dl.fedoraproject.org/pub/epel/5/x86_64/epel-release-5-4.noarch.rpm
warning: /var/tmp/rpm-xfer.CPJMIi: Header V3 DSA signature: NOKEY, key ID 217521f6
Preparing...                ########################################### [100%]
   1:epel-release           ########################################### [100%]
[cloudera@localhost ~]$ sudo yum -y install R
Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * base: mirror.symnds.com
 * epel: mirrors.einstein.yu.edu
 * extras: mirrors.einstein.yu.edu
 * updates: mirror.symnds.com
Setting up Install Process
Resolving Dependencies
There are unfinished transactions remaining. You might consider running yum-complete-transaction first to finish them.
The program yum-complete-transaction is found in the yum-utils package.
--> Running transaction check
---> Package R.x86_64 0:2.14.1-1.el5 set to be updated
--> Processing Dependency: libRmath-devel = 2.14.1-1.el5 for package: R
--> Processing Dependency: R-devel = 2.14.1-1.el5 for package: R
--> Running transaction check
---> Package R-devel.x86_64 0:2.14.1-1.el5 set to be updated
--> Processing Dependency: R-core = 2.14.1-1.el5 for package: R-devel
--> Processing Dependency: zlib-devel for package: R-devel
--> Processing Dependency: tk-devel for package: R-devel
--> Processing Dependency: texinfo-tex for package: R-devel
--> Processing Dependency: tetex-latex for package: R-devel
--> Processing Dependency: tcl-devel for package: R-devel
--> Processing Dependency: pcre-devel for package: R-devel
--> Processing Dependency: libX11-devel for package: R-devel
--> Processing Dependency: gcc-gfortran for package: R-devel
--> Processing Dependency: gcc-c++ for package: R-devel
--> Processing Dependency: bzip2-devel for package: R-devel
---> Package libRmath-devel.x86_64 0:2.14.1-1.el5 set to be updated
--> Processing Dependency: libRmath = 2.14.1-1.el5 for package: libRmath-devel
--> Running transaction check
---> Package R-core.x86_64 0:2.14.1-1.el5 set to be updated
--> Processing Dependency: xdg-utils for package: R-core
--> Processing Dependency: cups for package: R-core
--> Processing Dependency: libgfortran.so.1()(64bit) for package: R-core
---> Package bzip2-devel.x86_64 0:1.0.3-6.el5_5 set to be updated
---> Package gcc-c++.x86_64 0:4.1.2-51.el5 set to be updated
--> Processing Dependency: gcc = 4.1.2-51.el5 for package: gcc-c++
--> Processing Dependency: libstdc++-devel = 4.1.2-51.el5 for package: gcc-c++
---> Package gcc-gfortran.x86_64 0:4.1.2-51.el5 set to be updated
--> Processing Dependency: libgmp.so.3()(64bit) for package: gcc-gfortran
---> Package libRmath.x86_64 0:2.14.1-1.el5 set to be updated
---> Package libX11-devel.x86_64 0:1.0.3-11.el5_7.1 set to be updated
--> Processing Dependency: xorg-x11-proto-devel >= 7.1-2 for package: libX11-devel
--> Processing Dependency: libXau-devel for package: libX11-devel
--> Processing Dependency: libXdmcp-devel for package: libX11-devel
---> Package pcre-devel.x86_64 0:6.6-6.el5_6.1 set to be updated
---> Package tcl-devel.x86_64 0:8.4.13-4.el5 set to be updated
---> Package tetex-latex.x86_64 0:3.0-33.13.el5 set to be updated
--> Processing Dependency: tetex-dvips = 3.0 for package: tetex-latex
--> Processing Dependency: tetex = 3.0 for package: tetex-latex
--> Processing Dependency: netpbm-progs for package: tetex-latex
---> Package texinfo-tex.x86_64 0:4.8-14.el5 set to be updated
--> Processing Dependency: texinfo = 4.8-14.el5 for package: texinfo-tex
---> Package tk-devel.x86_64 0:8.4.13-5.el5_1.1 set to be updated
---> Package zlib-devel.x86_64 0:1.2.3-4.el5 set to be updated
--> Running transaction check
```

Pretty impressive for cut-and-pasting a few commands, eh?

Test Hadoop with a small job

Download my fork of Jonathan Seidman's sample R code from github

```
$ mkdir hadoop-r
$ cd hadoop-r
$ git init
$ git pull git://github.com/jeffreybreen/hadoop-R.git
```

Grab first 1,000 lines from ASA's 2004 airline data

```
$ curl http://stat-computing.org/dataexpo/2009/2004.csv.bz2 | bzcat \
   | head -1000 > 2004-1000.csv
```

Make some directories in HDFS and load the data file

```
$ hadoop fs -mkdir /user/cloudera
$ hadoop fs -mkdir asa-airline
$ hadoop fs -mkdir asa-airline/data
$ hadoop fs -mkdir asa-airline/out
$ hadoop fs -put 2004-1000.csv asa-airline/data/
```

Run Jonathan's sample streaming job

```
$ cd airline/src/deptdelay_by_month/R/streaming
$ hadoop jar /usr/lib/hadoop/contrib/streaming/hadoop-streaming-*.jar \
  -input asa-airline/data -output asa-airline/out/dept-delay-month \
  -mapper map.R -reducer reduce.R -file map.R -file reduce.R
```

```
[cloudera@localhost hadoop-r]$ head -2 2004-1000.csv
Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,
CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,Carrier
Delay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay
2004,1,12,1,623,630,901,915,UA,462,N805UA,98,105,80,-14,-7,ORD,CLT,599,7,11,0,,0,0,0,0,0,0
[cloudera@localhost hadoop-r]$ tail -2 2004-1000.csv
2004,1,25,7,857,900,1441,1446,UA,484,N457UA,224,226,208,-5,-3,PDX,ORD,1739,5,11,0,,0,0,0,0,0,0
2004,1,26,1,903,900,1524,1444,UA,484,N554UA,261,224,200,40,3,PDX,ORD,1739,25,36,0,,0,0,0,40,0,0

[cloudera@localhost hadoop-r]$ cd airline/src/deptdelay_by_month/R/streaming

[cloudera@localhost streaming]$ hadoop jar /usr/lib/hadoop/contrib/streaming/hadoop-streaming-*.jar \
> -input asa-airline/data -output asa-airline/out/dept-delay-month \
> -mapper map.R -reducer reduce.R -file map.R -file reduce.R
packageJobJar: [map.R, reduce.R, /var/lib/hadoop-0.20/cache/cloudera/hadoop-unjar4442605735512091493/] [] /tmp/
streamjob2138397329652275361.jar tmpDir=null
12/03/06 15:28:15 WARN snappy.LoadSnappy: Snappy native library is available
12/03/06 15:28:15 INFO util.NativeCodeLoader: Loaded the native-hadoop library
12/03/06 15:28:15 INFO snappy.LoadSnappy: Snappy native library loaded
12/03/06 15:28:15 INFO mapred.FileInputFormat: Total input paths to process : 1
12/03/06 15:28:17 INFO streaming.StreamJob: getLocalDirs(): [/var/lib/hadoop-0.20/cache/cloudera/mapred/local]
12/03/06 15:28:17 INFO streaming.StreamJob: Running job: job_201203061110_0001
12/03/06 15:28:17 INFO streaming.StreamJob: To kill this job, run:
12/03/06 15:28:17 INFO streaming.StreamJob: /usr/lib/hadoop-0.20/bin/hadoop job  -Dmapred.job.tracker=0.0.0.0:8021 -kill
job_201203061110_0001
12/03/06 15:28:17 INFO streaming.StreamJob: Tracking URL: http://0.0.0.0:50030/jobdetails.jsp?
jobid=job_201203061110_0001
12/03/06 15:28:18 INFO streaming.StreamJob:  map 0%  reduce 0%
12/03/06 15:28:37 INFO streaming.StreamJob:  map 100%  reduce 0%
12/03/06 15:29:15 INFO streaming.StreamJob:  map 100%  reduce 100%
12/03/06 15:29:18 INFO streaming.StreamJob: Job complete: job_201203061110_0001
12/03/06 15:29:18 INFO streaming.StreamJob: Output: asa-airline/out/dept-delay-month

[cloudera@localhost streaming]$ hadoop fs -ls asa-airline/out/dept-delay-month
Found 3 items
-rw-r--r--   1 cloudera supergroup          0 2012-03-06 15:29 /user/cloudera/asa-airline/out/dept-delay-month/_SUCCESS
drwxr-xr-x   - cloudera supergroup          0 2012-03-06 15:28 /user/cloudera/asa-airline/out/dept-delay-month/_logs
-rw-r--r--   1 cloudera supergroup         33 2012-03-06 15:29 /user/cloudera/asa-airline/out/dept-delay-month/
part-00000

[cloudera@localhost streaming]$ hadoop fs -cat asa-airline/out/dept-delay-month/part-00000
2004   1     973   UA    11.55293
```

Install RHadoop's **rmr** package

- RHadoop is an open source project sponsored by Revolution Analytics and is one of several available to make it easier to work with R and Hadoop
  - The **rmr** package contains all the mapreduce-related functions, including generating Hadoop streaming jobs and basic data exchange with HDFS
- First install prerequisite packages (run R as root to install system-wide)

```
$ sudo R
> install.packages( c('RJSONIO', 'itertools', 'digest'),
      repos='http://cran.revolutionanalytics.com')
```

- Download the latest stable release (1.2) from github

```
$ wget --no-check-certificate https://github.com/downloads/RevolutionAnalytics/RHadoop/rmr_1.2.tar.gz
```

- Install the package from the tar file

```
$ sudo R CMD INSTALL rmr_1.2.tar.gz
```

- Test that it loads

```
$ R
> library(rmr)
Loading required package: RJSONIO
Loading required package: itertools
Loading required package: iterators
Loading required package: digest
```

Test **rmr** with the airline example

- Runs same analysis as streaming example, but using **rmr**'s abstractions

```
$ cd
$ cd hadoop-r/airline/src/deptdelay_by_month/R/rmr/
$ export HADOOP_HOME=/usr/lib/hadoop
$ R
[...]
> source('deptdelay-rmr12.R')
```

- It will fail because our HDFS input paths don't match, but it did load all the functions so we can easily kick off the job by hand:

```
> df = from.dfs(deptdelay("asa-airline/data", "asa-airline/out/deptdelay-
month-rmr"), to.data.frame=T)
[...]
> colnames(df) = c('year', 'month', 'count', 'airline', 'mean.delay')
> df
      year month count airline       mean.delay
rmr.key 2004     1   973      UA 11.5529290853032
```

```
> df = from.dfs(deptdelay("asa-airline/data", "asa-airline/out/deptdelay-month-rmr"),
to.data.frame=T)

packageJobJar: [/tmp/RtmpZAckHy/rhstr.map4da957c5e126, /tmp/RtmpZAckHy/
rhstr.reduce4da938d5ffcb, /tmp/RtmpZAckHy/rmr-local-env, /tmp/RtmpZAckHy/rmr-global-env, /
var/lib/hadoop-0.20/cache/cloudera/hadoop-unjar6746493936124492 55/] [] /tmp/
streamjob8188313657687081754.jar tmpDir=null
12/03/06 16:28:57 WARN snappy.LoadSnappy: Snappy native library is available
12/03/06 16:28:57 INFO util.NativeCodeLoader: Loaded the native-hadoop library
12/03/06 16:28:57 INFO snappy.LoadSnappy: Snappy native library loaded
12/03/06 16:28:57 INFO mapred.FileInputFormat: Total input paths to process : 1
12/03/06 16:28:58 INFO streaming.StreamJob: getLocalDirs(): [/var/lib/hadoop-0.20/cache/
cloudera/mapred/local]
12/03/06 16:28:58 INFO streaming.StreamJob: Running job: job_201203061110_0003
12/03/06 16:28:58 INFO streaming.StreamJob: To kill this job, run:
12/03/06 16:28:58 INFO streaming.StreamJob: /usr/lib/hadoop/bin/hadoop job  -
Dmapred.job.tracker=0.0.0.0:8021 -kill job_201203061110_0003
12/03/06 16:28:58 INFO streaming.StreamJob: Tracking URL: http://0.0.0.0:50030/
jobdetails.jsp?jobid=job_201203061110_0003
12/03/06 16:28:59 INFO streaming.StreamJob:  map 0%  reduce 0%
12/03/06 16:29:21 INFO streaming.StreamJob:  map 100%  reduce 0%
12/03/06 16:29:46 INFO streaming.StreamJob:  map 100%  reduce 100%
12/03/06 16:29:55 INFO streaming.StreamJob: Job complete: job_201203061110_0003
12/03/06 16:29:55 INFO streaming.StreamJob: Output: asa-airline/out/deptdelay-month-rmr
>
> colnames(df) = c('year', 'month', 'count', 'airline', 'mean.delay')
> df
        year month count airline       mean.delay
rmr.key 2004     1   973      UA 11.5529290853032
```

Extra Credit: Install RStudio

- Current download link and instructions at
  http://rstudio.org/download/server

```
$ wget http://download2.rstudio.org/rstudio-server-0.95.262-x86_64.rpm
$ sudo rpm -Uvh rstudio-server-0.95.262-x86_64.rpm
```

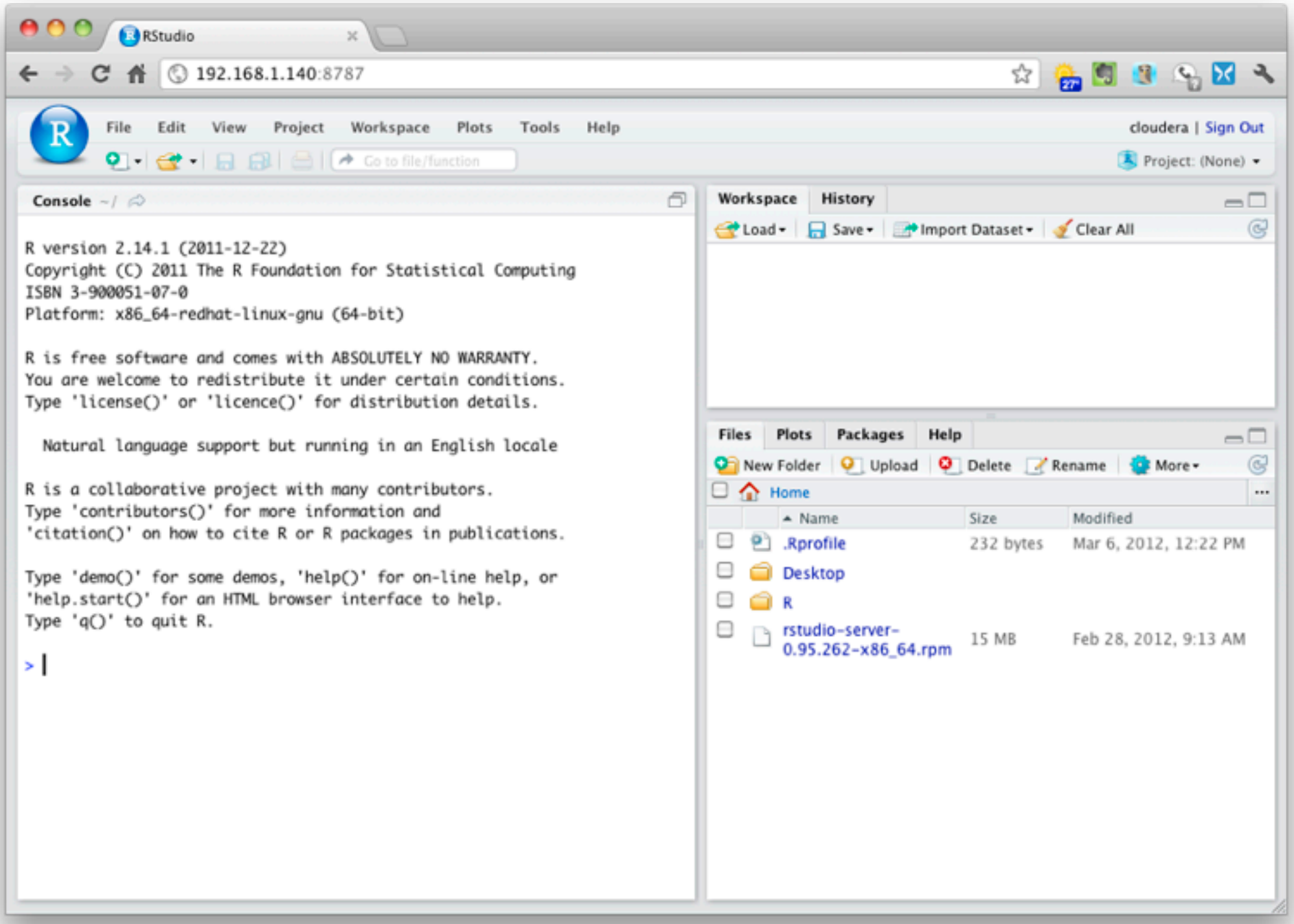
- Find IP address with ifconfig

```
$ ifconfig
```

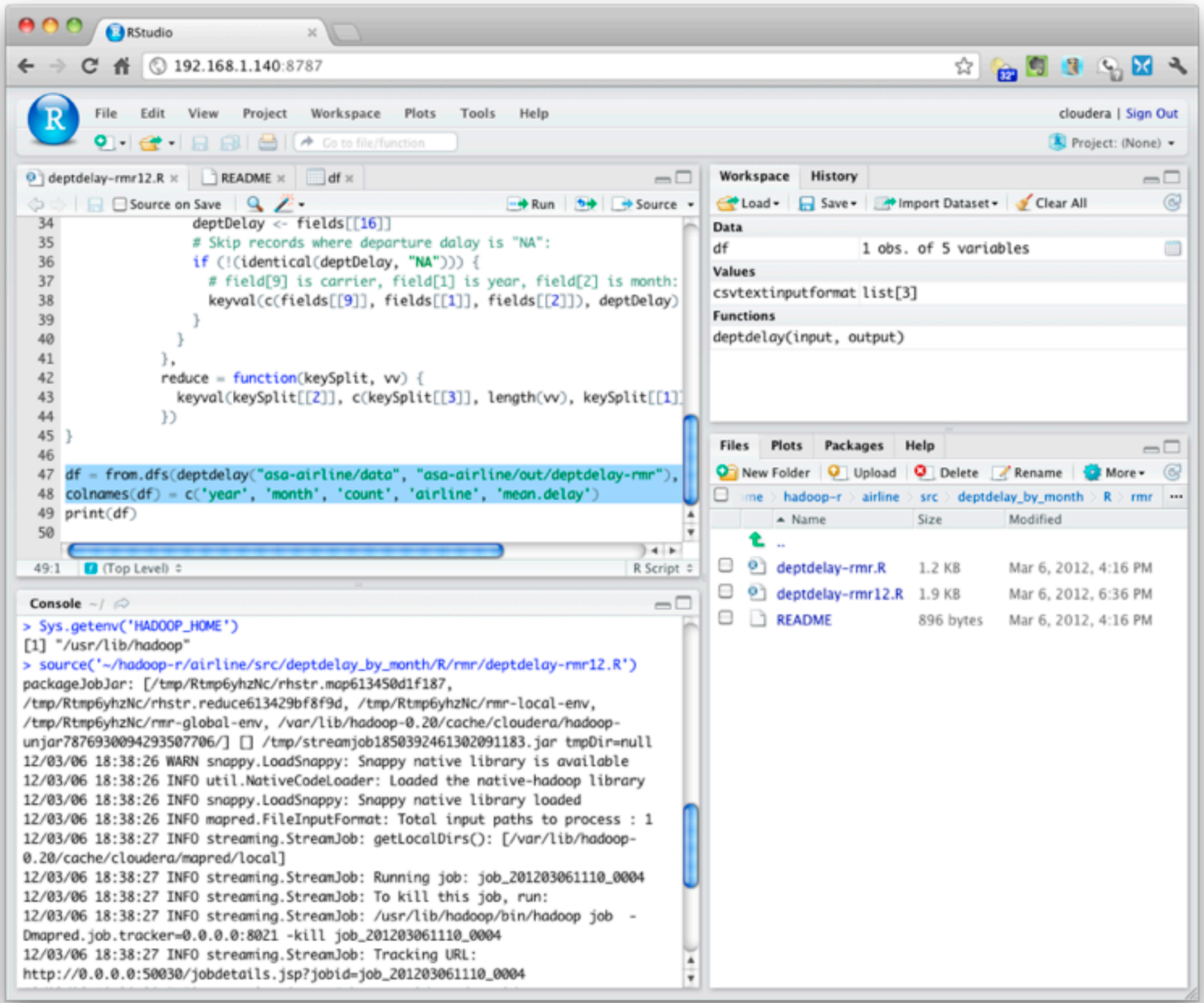
- Access from browser via port 8787
  - e.g., **http://192.168.1.140:8787/**

```
[cloudera@localhost ~]$ wget http://download2.rstudio.org/rstudio-server-0.95.262-x86_64.rpm
--2012-03-06 12:14:24--  http://download2.rstudio.org/rstudio-server-0.95.262-x86_64.rpm
Resolving download2.rstudio.org... 216.137.39.181, 216.137.39.217, 216.137.39.222, ...
Connecting to download2.rstudio.org|216.137.39.181|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 15748959 (15M) [application/x-redhat-package-manager]
Saving to: `rstudio-server-0.95.262-x86_64.rpm'

100%[======================================================================================================>] 15,748,959  1.83M/s   in 7.2s

2012-03-06 12:14:31 (2.09 MB/s) - `rstudio-server-0.95.262-x86_64.rpm' saved [15748959/15748959]

[cloudera@localhost ~]$ sudo rpm -Uvh rstudio-server-0.95.262-x86_64.rpm
Preparing...                ########################################### [100%]
   1:rstudio-server         ########################################### [100%]
rsession: no process killed
Starting rstudio-server: [  OK  ]
[cloudera@localhost ~]$ ifconfig
eth0      Link encap:Ethernet  HWaddr 00:0C:29:4B:77:1D
          inet addr:192.168.1.140  Bcast:192.168.1.255  Mask:255.255.255.0
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:75039 errors:0 dropped:0 overruns:0 frame:0
          TX packets:36742 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:104953280 (100.0 MiB)  TX bytes:3061577 (2.9 MiB)
          Interrupt:59 Base address:0x2000

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
          RX packets:78954 errors:0 dropped:0 overruns:0 frame:0
          TX packets:78954 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:14608044 (13.9 MiB)  TX bytes:14608044 (13.9 MiB)
```

RStudio Success



RStudio + **rmr** works too



Next up:
Running R & RStudio on Amazon EC2