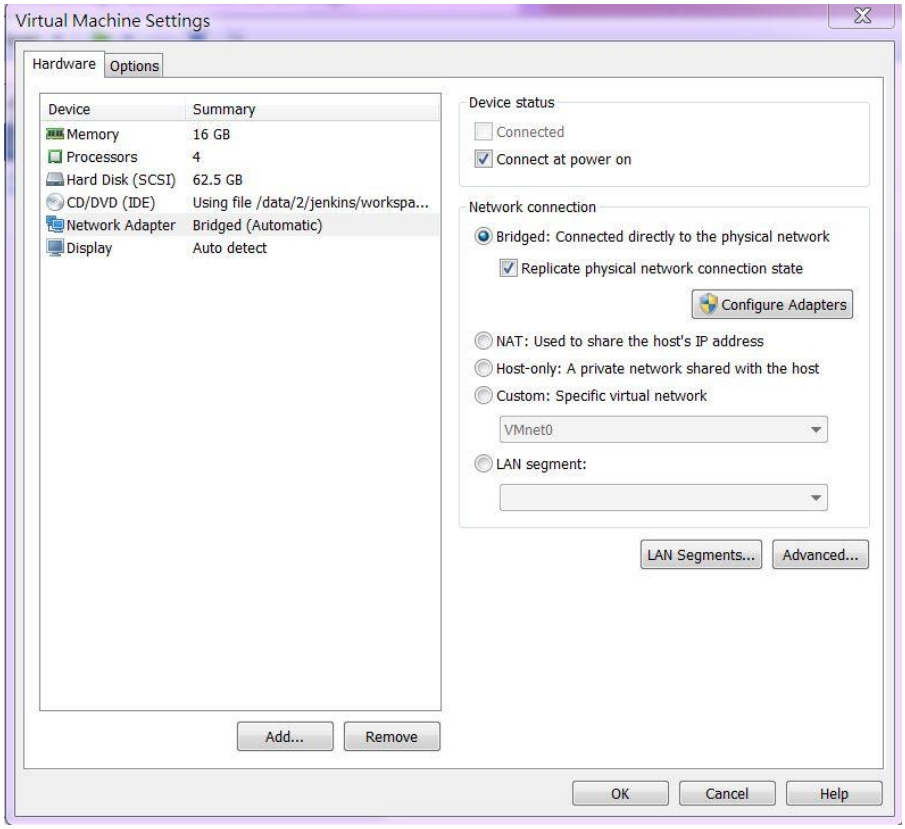
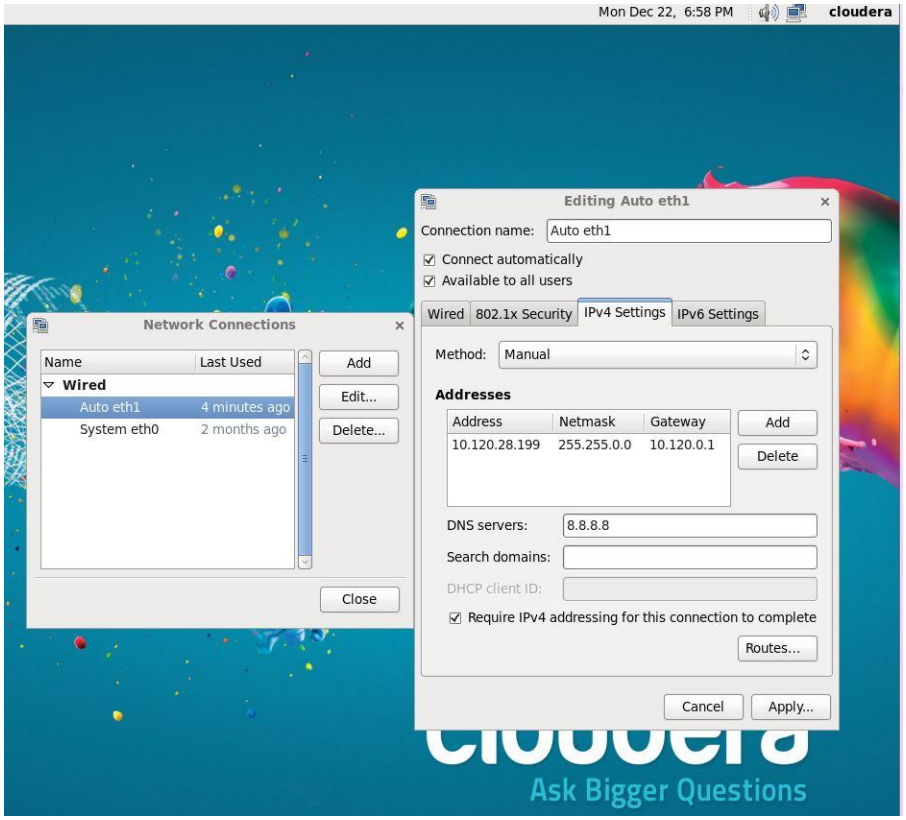
1-0. (事前下載) Download for VMWare

http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-3-x.html

1-1. (事前設定) 開一個新 VM, Network Adapter Bridged(Automatic)



1-2. (事前設定) VM 內部設定, IPv4 同 windows 網卡, Address 末 2 碼不能相同



2-1. Linux (RRO8.0.1) centos 6 版本(全程以 root 身分進行)

```
[cloudera@quickstart ~]$ su
[root@quickstart cloudera]# sudo yum clean all
# yum list make gcc gcc-gfortran
# yum install gcc-c++.x86_64
# wget -no-check-certificate
http://mran.revolutionanalytics.com/install/RRO-8.0.1-Beta-el6.x86_64.rpm
# yum --nogpgcheck localinstall RRO-8.0.1-Beta-el6.x86_64.rpm
```

2-2. 套件下載及安裝

```
# which hadoop
# export HADOOP_CMD=/usr/bin/hadoop
# locate streaming | grep jar | more
# export
HADOOP_STREAMING=/usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-str
eaming.jar
# echo $JAVA_HOME
# export JAVA_HOME=/usr/java/jdk1.7.0_67-cloudera
```

```
# R CMD javareconf
# sudo R
> install.packages(c("codetools", "Rcpp", "RJSONIO", "bitops", "digest", "fun
ctional", "stringr", "plyr", "reshape2", "rJava", "caTools"))

> q()
```

2-3. Download rmr2

```
# wget –no-check-certificate http://goo.gl/Y5ytsm
# R CMD INSTALL Y5ytsm
```

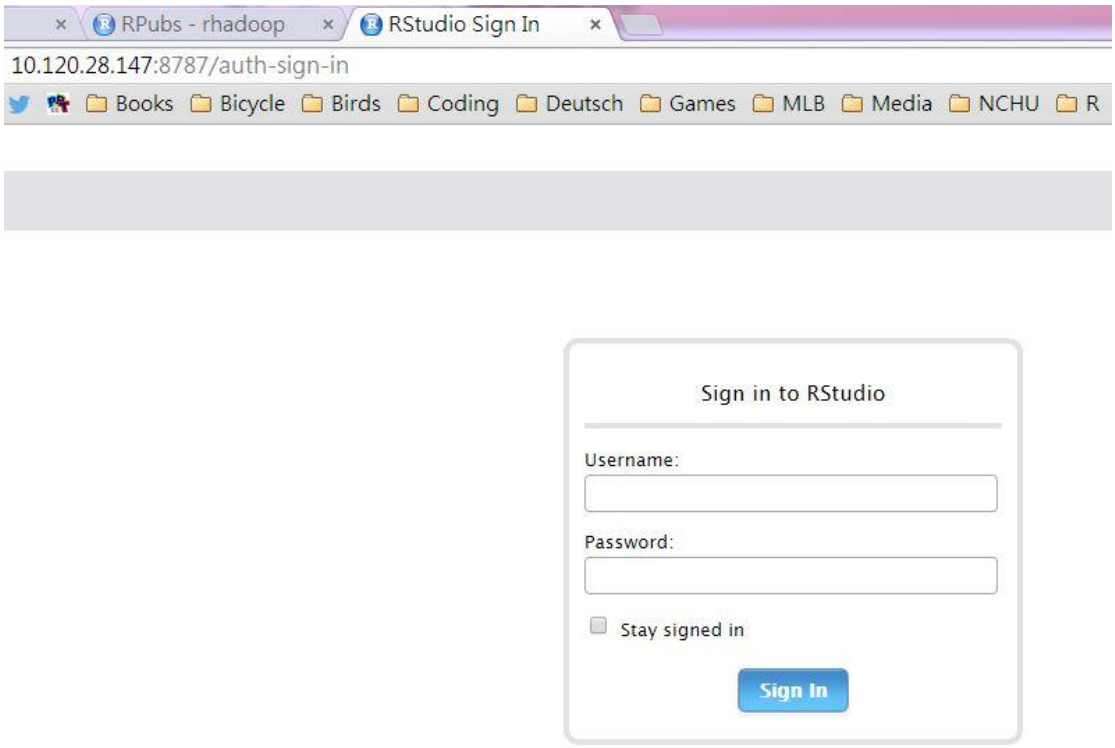
2-4. Download rhdfs

```
# wget –no-check-certificate
https://github.com/RevolutionAnalytics/rhdfs/blob/master/build/rhdfs_1.0.8.t
ar.gz?raw=true
# R CMD INSTALL rhdfs_1.0.8.tar.gz\?raw\=true
```

2-5. Download RStudio Server - RedHat/CentOS

http://www.rstudio.com/products/rstudio/download-server/

```
# sudo yum install openssl098e # Required only for RedHat/CentOS 6 and 7
# wget http://download2.rstudio.org/rstudio-server-0.98.1091-x86_64.rpm
# sudo yum install --nogpgcheck rstudio-server-0.98.1091-x86_64.rpm
# sudo rstudio-server restart
```



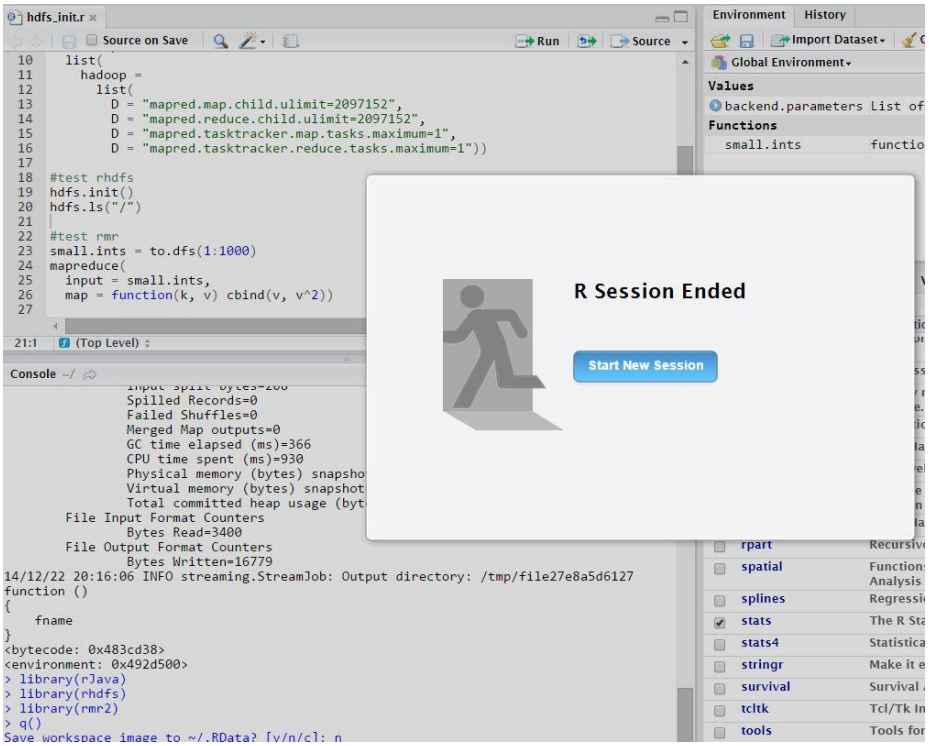
2-6. 執行 rmr2 前的系統設定(建議用 RScript 儲存，每次開啟 Rsession 時需執行)

```
Sys.setenv(HADOOP_CMD="/usr/bin/hadoop")
Sys.setenv(HADOOP_STREAMING="/usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming.jar")
Sys.setenv(JAVA_HOME="/usr/java/jdk1.7.0_67-cloudera")

library(rJava)
library(rhdfs)
library(rmr2)

backend.parameters =
  list(
    hadoop =
      list(
        D = "mapred.map.child.ulimit=2097152",
        D = "mapred.reduce.child.ulimit=2097152",
        D = "mapred.tasktracker.map.tasks.maximum=1",
        D = "mapred.tasktracker.reduce.tasks.maximum=1"))
#-- test rhdfs
hdfs.init()
```

```
hdfs.ls("/")

#-- test rmr

small.ints = to.dfs(1:1000)

mapreduce(

  input = small.ints,

  map = function(k, v) cbind(v, v^2))
```

```
Console ~/
> mapreduce(
+   input = small.ints,
+   map = function(k, v) cbind(v, v^2))
14/12/22 20:15:40 INFO Configuration.deprecation: mapred.reduce.tasks is deprecated. In
stead, use mapreduce.job.reduces
packageJobJar: [] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.5.0-cdh5.2.0.jar] /tmp/
streamjob3076427708964467910.jar tmpDir=null
14/12/22 20:15:41 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
14/12/22 20:15:41 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
14/12/22 20:15:42 INFO mapred.FileInputFormat: Total input paths to process : 1
14/12/22 20:15:42 INFO mapreduce.JobSubmitter: number of splits:2
14/12/22 20:15:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_141930553
5350_0001
14/12/22 20:15:42 INFO impl.YarnClientImpl: Submitted application application_141930553
5350_0001
14/12/22 20:15:42 INFO mapreduce.Job: The url to track the job: http://quickstart.cloud
era:8088/proxy/application_1419305535350_0001/
14/12/22 20:15:42 INFO mapreduce.Job: Running job: job_1419305535350_0001
14/12/22 20:15:53 INFO mapreduce.Job: Job job_1419305535350_0001 running in uber mode :
 false
14/12/22 20:15:53 INFO mapreduce.Job:  map 0% reduce 0%
14/12/22 20:16:05 INFO mapreduce.Job:  map 50% reduce 0%
14/12/22 20:16:06 INFO mapreduce.Job:  map 100% reduce 0%
14/12/22 20:16:06 INFO mapreduce.Job: Job job_1419305535350_0001 completed successfully
14/12/22 20:16:06 INFO mapreduce.Job: Counters: 30
        File System Counters
```

```
> q()
```



3-1 SHINY-SERVER

http://www.rstudio.com/products/shiny/download-server/

選取 RedHat/CentOS

```
# sudo R
> install.packages('shiny', repos='http://cran.rstudio.com/')
> q()
```

Download and Install

```
# wget
http://download3.rstudio.org/centos-5.9/x86_64/shiny-server-1.2.3.368-x86_64
.rpm
# sudo yum install --nogpgcheck shiny-server-1.2.3.368-x86_64.rpm

# sudo rm -r starOpeningWeekendGross
# sudo cp -r /home/cloudera/Desktop/R/starOpeningWeekendGross
/srv/shiny-server/sample-apps
```

Reference:

http://rpubs.com/ywchiu/25570

https://support.rstudio.com/hc/en-us/articles/200552306-Getting-Started

RStudio-server

http://10.120.28.xxx:8787/

shiny-server

http://10.120.28.xxx:3838/

version 20150111