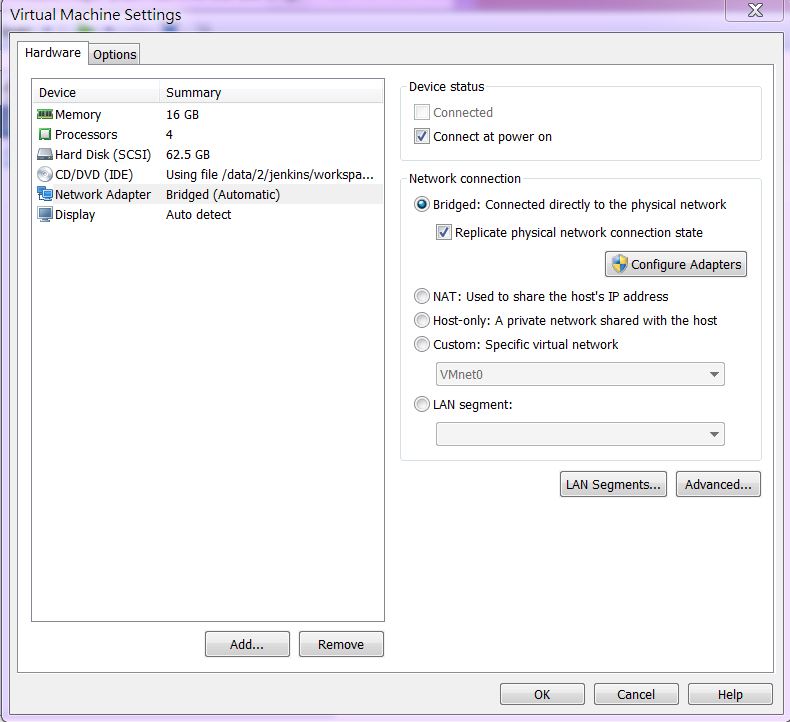
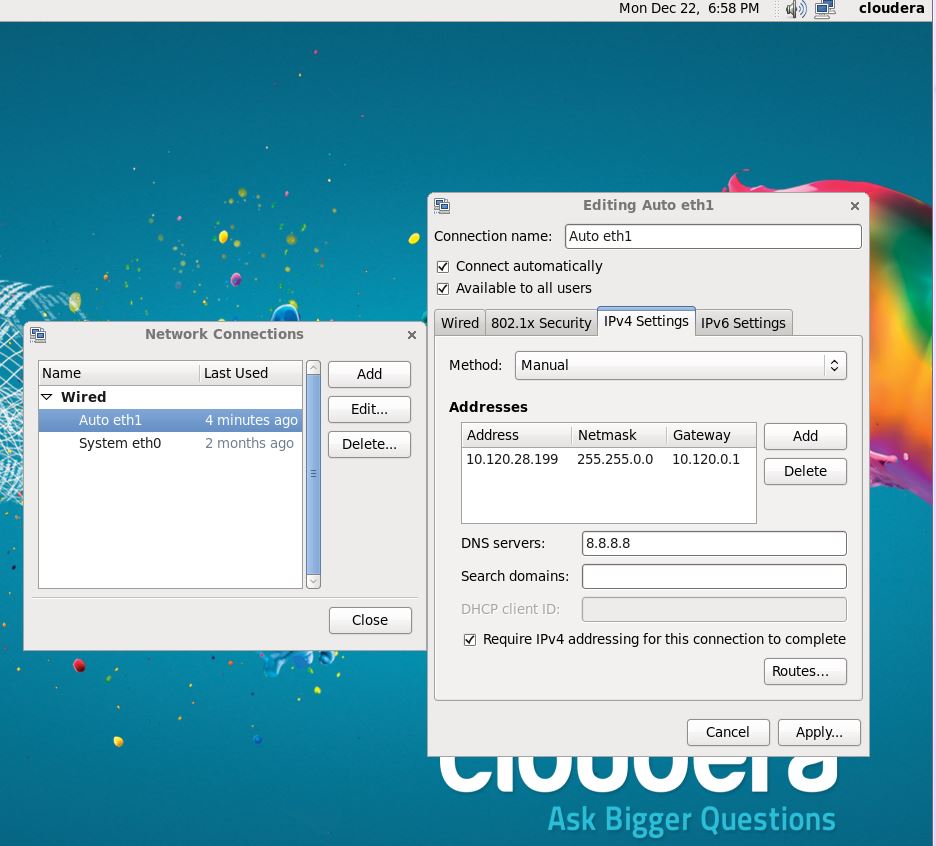
* 1. **(事前設定)** 開一個新VM, Network Adapter Bridged(Automatic)



* 1. **(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

#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**

# yum install libcurl curl curl-devel

# export HADOOP\_CMD=/usr/bin/hadoop

# export HADOOP\_STREAMING=/usr/lib/hadoop-0.20-mapreduce/contrib/streaming/hadoop-streaming.jar

# export JAVA\_HOME=/usr/java/jdk1.7.0\_67-cloudera

# R CMD javareconf

# sudo R

> install.packages(c("RCurl","httr"))

> install.packages(c("codetools", "Rcpp", "RJSONIO", "bitops", "digest", "functional", "stringr", "plyr", "reshape2", "rJava", "caTools","devtools"))

> install.packages(c("rjson","bit64"))

> install.packages(c("R.methodsS3","Hmisc","dplyr"))

> install\_github("RevolutionAnalytics/memoise")

> install\_github("hadley/lazyeval")

> q()

**2-3. Download rmr2**

# wget –no-check-certificate http://goo.gl/Y5ytsm

# R CMD INSTALL rmr2\_3.3.0.tar.gz

**2-4.** **Download rhdfs**

# wget –no-check-certificate https://github.com/RevolutionAnalytics/rhdfs/blob/master/build/rhdfs\_1.0.8.tar.gz?raw=true

# R CMD INSTALL rhdfs\_1.0.8.tar.gz

**2-4.1** **Download ravro**

# wget –no-check-certificate https://github.com/RevolutionAnalytics/ravro/blob/1.0.4/build/ravro\_1.0.4.tar.gz?raw=true

# R CMD INSTALL ravro\_1.0.4.tar.gz

**2-4.2 Download plyrmr**

# wget --no-check-certificate http://goo.gl/pHmFV3

# R CMD INSTALL plyrmr\_0.5.0.tar.gz

**2-4.3 Download rhbase**

~~請參照安裝Thrift手冊進行安裝thrift~~

（仿照下列網頁）

http://www.geedoo.info/installed-on-the-cloudera-hadoop-cdh-r-and-rhadoop-rhdfs-rmr2-rhbase-rhive.html

# wget http://archive.apache.org/dist/thrift/0.9.0/thrift-0.9.0.tar.gz

# tar -xvf thrift-0.9.0.tar.gz

# cd thrift-0.9.0

# yum -y install automake libtool flex bison pkgconfig gcc-c++ boost-devel libevent-devel zlib-devel python-devel ruby-devel openssl-devel

# ./configure

# make

# make install

檢查版本，如為0.9.0則為成功

# thrift -version

下載rhbase

# wget --no-check-certificate https://github.com/RevolutionAnalytics/rhbase/blob/master/build/rhbase\_1.2.1.tar.gz?raw=true

# cp /usr/local/lib/libthrift\* /usr/lib

# cp /usr/local/lib/libthrift\* /usr/lib64

# R CMD INSTALL rhbase\_1.2.1.tar.gz

> hb.init(serialize="raw")

> if (names(hb.list.tables())=="mytable") {

hb.delete.table("mytable")

}

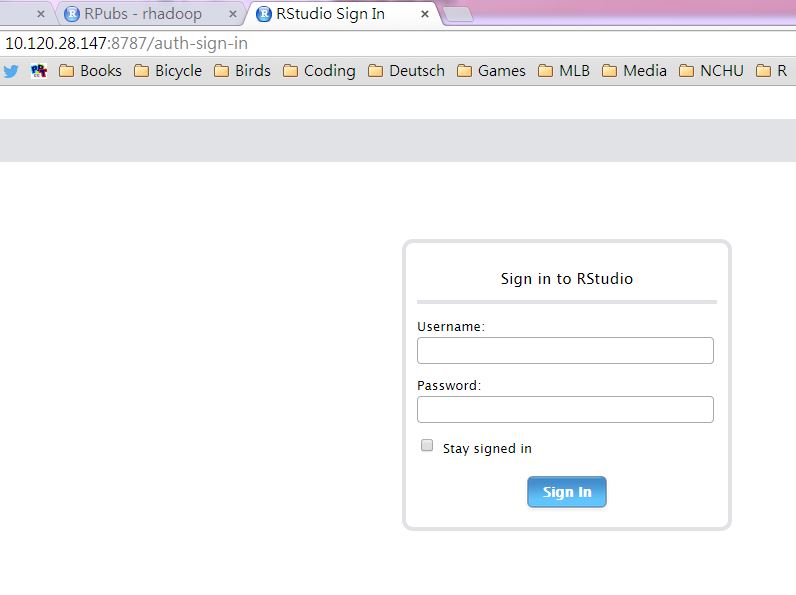
> hb.new.table("mytable", "x","y","z",opts=list(y=list(compression='GZ')))

# 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("/")

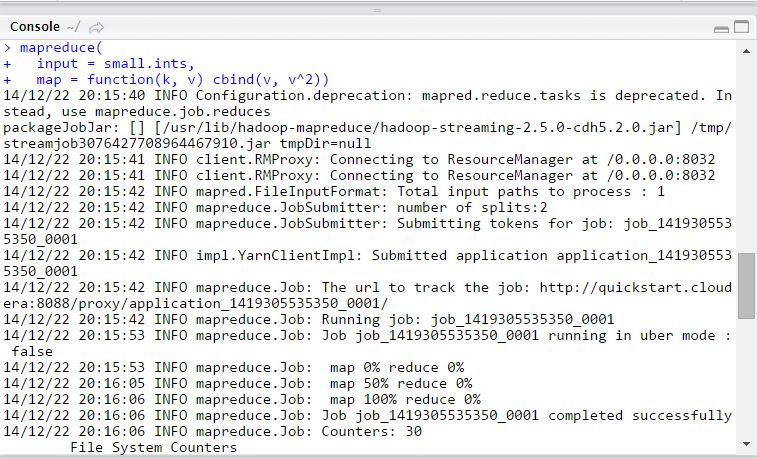
#test rmr

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

mapreduce(

input = small.ints,

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



> q()

