***Installation of R Server Studio along with RHadoop***

=========================**R Server Studio Installation**=========================

1. Update the CRAN Package list in the ***sources.list*** file, from where the R would be downloaded.

*# echo \# R Packages >> /etc/apt/sources.list*

*# echo deb http://ftp.iitm.ac.in/cran/bin/linux/ubuntu trusty/ >> /etc/apt/sources.list*

2. Install the R-Base and gdebi-core package

*# sudo apt-get update*

*# sudo apt-get install r-base*

*# sudo apt-get install gdebi-core*

3. Download and install the R-Server-Studio and install it.

*# wget https://download2.rstudio.org/rstudio-server-0.99.903-amd64.deb*

*# sudo gdebi rstudio-server-0.99.903-amd64.deb*

4. Update the .bashrc file for R

*# echo export HADOOP\_CMD=/usr/local/hadoop/bin/hadoop >> /home/hduser/.bashrc*

*# echo export HADOOP\_STREAMING=/usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.7.2.jar >> /home/hduser/.bashrc*

*# echo export LD\_LIBRARY\_PATH=/usr/lib/jvm/java-8-oracle/lib/amd64:/usr/lib/jvm/java-8-oracle/jre/lib/amd64/server >> /home/hduser/.bashrc*

5. Update CRAN site in ***Rprofile.site*** file, for the download of R

*# echo 'options(repos=structure(c(CRAN="http://ftp.iitm.ac.in/cran/")))' >> /etc/R/Rprofile.site*

6. Verify the installation of rstudio-server

*# sudo rstudio-server verify-installation*

*=======================***RHadoop Installation***=====================*

7. Install Rjava for Rhadoop installation

*# sudo apt-get install r-cran-rjava*

*# R CMD javareconf*

8. Install dependency packages for setting up of Rhadoop

*# Rscript -e 'install.packages(c("rJava"))'*

*# Rscript -e 'install.packages(c("functional"))'*

*# sudo apt-get install r-cran-rcpp*

*# sudo apt-get install r-cran-reshape2*

*# Rscript -e 'install.packages(c("RJSONIO", "bitops", "digest", "stringr","dplyr","R.methodsS3", "caTools","Hmisc"))'*

9. Download Rhdfs and install on the client node which would run R using hduser account.

*# su hduser*

*$ wget https://github.com/RevolutionAnalytics/rhdfs/raw/master/build/rhdfs\_1.0.8.tar.gz*

*$ Rscript -e 'install.packages("rhdfs\_1.0.8.tar.gz",repos=NULL,type="source")'*

10. Download RMR and install it using hduser account

*# su hduser*

*$ wget https://github.com/RevolutionAnalytics/rmr2/releases/download/3.3.1/rmr2\_3.3.1.tar.gz*

*$ Rscript -e 'install.packages("rmr2\_3.3.1.tar.gz",repos=NULL,type="source")'*

For installation of R-MapReduce(RMR), every node in the cluster should be installed with R and dependency packages

============================In all the slave nodes===========================

1. Install the R-Base package

*# sudo apt-get update*

*# sudo apt-get install r-base*

2. Update the .bashrc file for R

*# echo export HADOOP\_CMD=/usr/local/hadoop/bin/hadoop >> /home/hduser/.bashrc*

*# echo export HADOOP\_STREAMING=/usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.7.2.jar >> /home/hduser/.bashrc*

*# echo export LD\_LIBRARY\_PATH=/usr/lib/jvm/java-8-oracle/lib/amd64:/usr/lib/jvm/java-8-oracle/jre/lib/amd64/server >> /home/hduser/.bashrc*

3. Update CRAN site in ***Rprofile.site*** file, for the download of R

*# echo 'options(repos=structure(c(CRAN="http://ftp.iitm.ac.in/cran/")))' >> /etc/R/Rprofile.site*

4. Install Rjava for Rhadoop installation

*# sudo apt-get install r-cran-rjava*

*# R CMD javareconf*

5. Install dependency packages for setting up of Rhadoop

*# Rscript -e 'install.packages(c("rJava"))'*

*# Rscript -e 'install.packages(c("functional"))'*

*# sudo apt-get install r-cran-rcpp*

*# sudo apt-get install r-cran-reshape2*

*# Rscript -e 'install.packages(c("RJSONIO", "bitops", "digest", "stringr","dplyr","R.methodsS3", "caTools","Hmisc"))'*

6. Download RMR and install it from hduser account

*# su hduser*

*$ cd /home/hduser/*

*$ wget https://github.com/RevolutionAnalytics/rmr2/releases/download/3.3.1/rmr2\_3.3.1.tar.gz*

*$ Rscript -e 'install.packages("/home/hduser/rmr2\_3.3.1.tar.gz",repos=NULL,type="source")'*

================= Run R Studio from any machine using the Hostname=============

**Open the web-browser from the client node where the R would run from web browser**

**<172.25.3.7:8787>**

**Login using the username: hduser password:Admin123**

**Run the following code below:-**

> Sys.setenv(HADOOP\_CMD="/usr/local/hadoop/bin/hadoop")

> Sys.setenv(HADOOP\_STREAMING="/usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-2.7.2.jar")

> ints = to.dfs(1:100)

> calc = mapreduce(input = ints, map = function(k, v) cbind(v, 2\*v))

> from.dfs(calc)