ORAAH 2.3.1 Installation

Note: Since release 2.3.0, Oracle R Connector for Hadoop (ORCH) has been renamed as Oracle R Advanced Analytics for Hadoop (ORAAH) to reflect the addition of a breadth of work-horse predictive analytic techniques optimized for the Hadoop platform.

Hadoop cluster setup.

Until the ORCH 2.2.1 release, installation of ORCH (and dependent ORE) packages was only required on the client node from where the R user interacts with the Hadoop deployment. In particular, there was no installation requirement on the actual cluster nodes of Hadoop.

Starting from 2.3.0 release, ORAAH 2.3.0 now has 2 additional dependent ORE packages (in addition to the already existing 2 ORE packages dependency) that need to be installed on the client node and all Hadoop nodes. Thus, the set of ORE dependent packages are:

- OREbase
- OREstats
- OREmodels
- OREserver

Further, the OREserver package needs to be installed on all cluster nodes where mappers / reducers will run. The installation instructions on the client node are:

```
R --vanilla CMD INSTALL OREbase_1.4_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL OREstats_1.4_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL OREmodels_1.4_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL OREserver_1.4_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL ORCHcore_2.3.0_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL ORCHstats_2.3.0_R_x86_64-unknown-linux-gnu.tar.gz
R --vanilla CMD INSTALL ORCH_2.3.0_R_x86_64-unknown-linux-gnu.tar.gz
```

The instructions for installing OREserver on all nodes of the cluster are:

```
dcli -C -f OREserver_1.4_R_x86_64-unknown-linux-gnu.tar.gz -d /tmp/ \
     OREserver_1.4_R_x86_64-unknown-linux-gnu.tar.gz
dcli -C " R --vanilla CMD INSTALL /tmp/OREserver_1.4_R_x86_64-unknown-linux-gnu.tar.gz"
```

If the OREserver package has NOT been installed on any of the cluster nodes where a mapper/reducer tasks could run, then 2 of the analytics, namely, "orch.glm" and "orch.lm" will error out. The following error can be seen in the mapper task log file:

```
Error in loadNamespace(name) : there is no package called 'OREserver':
Calls: source ... tryCatch -> tryCatchList -> tryCatchOne -> <Anonymous>
In addition: Warning message:
In library(package, lib.loc = lib.loc, character.only = TRUE, logical.return = TRUE,  :
 there is no package called 'OREserver'_
Execution halted
```

If this error is encountered please verify/reintall OREserver package using dcli command line utility on

all Hadoop cluster task tracker nodes.

Installing with non Oracle's R distribution.

In case you choose to do not use Oracle's R distribution you will need to install one additional library "libOrdBlasLoader.so" to enable ORCH statistical packages to function correctly. If the library is not installed then you will encounter failures when running most of ORCH analytics.

If you use Oracle's R distribution (available via http://public-yum.oracle.com/ for OEL5 and OEL6) the library is already included with the distribution and no addition actions are required.

To install the library, copy the library file "libOrdBlasLoader.so", which is included in ORCH distribution zip file, into one of the directories listed in $LD_LIBRARY_PATH or (preferred way) into R's "lib" directory (default "/usr/lib64/R/lib").

```
cp /tmp/libOrdBlasLoader.so /usr/lib64/R/lib
```

The same should be done on every node of Hadoop cluster where R is installed:

```
dcli -C -f  libOrdBlasLoader.so -d /tmp/  libOrdBlasLoader.so
dcli -C "cp /tmp/libOrdBlasLoader.so /usr/lib64/R/lib"
```

Note that you need root privileges to install this library.

Other dependencies.

ORCH also has indirect dependency on "png" package. You can download it from CRAN http://cran.r-project.org/web/packages/png/index.html and install it on the client's host only.

```
R CMD INSTALL png_0.1-6.tar.gz
```