**Assignment 15.3:**

**Hadoop Deployment Layout:**

We need to divide Hadoop up into packages that can be independently upgraded. The list of packages should include:

* Hadoop Common - Common including the native code and required jar files.
* HDFS Client - HDFS jars, scripts, and shared libraries.
* HDFS Server - jsvc executable
* Yarn Client - Yarn client jars and scripts
* Yarn Server - Yarn server jars and scripts
* MapReduce - MapReduce jars, scripts, and shared libraries
* LZO - LZ0 codec from github.com/omally/hadoop-gpl-compression
* Metrics - Plugins for Chukwa and Ganglia

Packages from other teams will include:

* Pig
* Hive
* Oozie client
* Oozie server
* Howl client
* Howl server

These packages should be deployable with RPM on RedHat. We also need a package that depends on a version of each of these packages. In general, we can generate tarballs in the new deployment layout. Note that some packages, like Pig, which are user facing, will have 2 versions installed in a given deployment. This will be accomplished by modifying the package name and the associated binaries to include the version number.All of the following paths are based on a prefix directory that is the root of the installation. Our packages must support having multiple Hadoop stack installation on a computer at the same time. For RPMs, this means that the packages must be relocatable and honor the prefix option.

Path Configurations:

Path can be configured at compile phase or installation phase. For RPM, it takes advantage of the --relocate directive to allow path reconfiguration at install phase. For Debian package, path is configured at compile phase.

Build phase parameter:

* package.prefix - Location of package prefix (Default /usr)
* package.conf.dir - Location of configuration directory (Default /etc/hadoop)
* package.log.dir - Location of log directory (Default /var/log/hadoop)
* package.pid.dir - Location of pid directory (Default /var/run/hadoop)