**15.3)Explain Hadoop Deployment Layout in Brief.**

Standard deployment layout for Hadoop:

With increased complexity and evolving Hadoop ecosystem, having standard deployment layout ensures better integration between Hadoop sub-projects. By making the installation process easier, we can lower the barrier to entry and increase Hadoop adoption.

Packages:

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 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.

Deployment:

It is important to have a standard deployment that results from installing the packages regardless of the package manager. Here are the top level directories and a sample of what would be under each.

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)