Assignment-15.3:

Explain Hadoop Deployment Layout in Brief.

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:

1. Hadoop Common - Common including the native code and required jar files.
2. HDFS Client - HDFS jars, scripts, and shared libraries.
3. HDFS Server - jsvc executable
4. Yarn Client - Yarn client jars and scripts
5. Yarn Server - Yarn server jars and scripts
6. MapReduce - MapReduce jars, scripts, and shared libraries
7. LZO - LZ0 codec from github.com/omally/hadoop-gpl-compression
8. 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. Note that all of the packages are installed "flattened" into the prefix directory. For compatibility reasons, we should create "share/hadoop" that matches the old HADOOP\_HOME and set the HADOOP\_HOME variable to that.

The steps the in the deployment of the Hadoop cluster

1) Setting up Hadoop cluster

- choosing the cluster hardware

-Hadoop distributions

- choosing os for Hadoop cluster.

2) Installing and configuring Hadoop

-configuring the os.

-setting up the name node.

3) Configuring the Hadoop ecosystem.

-scoop

-hive

-impala

-pig

-sqoop

-oozie

4) Securing the Hadoop installation

-security overview

-HDFS security

-map reduce security etc,.

5) Monitoring Hadoop cluster

-monitoring Hadoop using ganglia

6) Deploying Hadoop on cloud

- Amazon elastic map reduce

**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)

Install phase parameter:

rpm -i hadoop-[version]-[rev].[arch].rpm \

--relocate /usr=/usr/local/hadoop \

--relocate /etc/hadoop=/usr/local/etc/hadoop \

--relocate /var/log/hadoop=/opt/logs/hadoop \

--relocate /var/run/hadoop=/opt/run/Hadoop

