hadoop提供数据存储与计算模型。

HDFS集群安装与部署：

Hdfs集群环境的搭建：下载二进制文件，分别下载1.2.1,2.0.3的版本进行查看，其中前者包含源代码，后者没有，要分别下载。Hdfs的部署，包含独立模式，伪分布式模式，分布式模式，见《hadoop权威指南附录A》，hdfs的配置conf下的配置hadoop-env.sh里面的javahome，core-site.xml, hdfs-site.xml, mapred-site.xml，配置好前两者后就可以启动namenode节点了，启动前先要namenode –format；（我这首先在配置core-site.xml的时候少写了一个“/”，但在format的过程中可以发现这个错误，但还有一点那就是我每次启动namenode都要输入密码，为什么？ 前面说到输入密码是因为在分布式的模式下，要启动远程的守护进程才会要求，但我只是伪分布式的，为什么还是要求输入呢？去除登录密码：$ ssh-keygen -t dsa -P '' -f ~/.ssh/id\_dsa   
$ cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys）

Hdfs shell命令的使用：<http://hadoop.apache.org/docs/r1.2.1/file_system_shell.html>，

./hadoop command -?（在这里你要尝试fs, dfsadmin, fsck, balancer, namenode -format）

Hdfs shell脚本的使用：我们可以查看./start-dfs.sh后的执行顺序，后面会遇到hadoop-daemon.sh, hadoop-daemons.sh, 接下来就是会启动hadoop，然后会执行java程序，这是我们最后的落脚点。

操作实验：分布模式，独立模式的实验，hdfs的shell脚本，cloudera的repo建立，实验examples；