1/运行mr程序出错

connecting to resoucemanager

retrying ....

retrying .....

原因是没有启动yarn或者启动失败

2/初始化工作目录结构

hdfs namenode -format 只是初始化了namenode的工作目录

而datanode的工作目录是在datanode启动后自己初始化的

3/datanode不被namenode识别的问题

namenode在format初始化的时候会形成两个标识：

blockPoolId：

clusterId：

新的datanode加入时，会获取这两个标识作为自己工作目录中的标识

一旦namenode重新format后，namenode的身份标识已变，而datanode如果依然

持有原来的id，就不会被namenode识别

4/datanode下线后多久看到效果

datanode不是一下线就会被namenode认定为下线的，有一个超时时间

5/关于副本数量的问题

副本数由客户端的参数dfs.replication决定（优先级： conf.set > 自定义配置文件 > jar包中的hdfs-default.xml）