学习总结

1812975 曹沛鑫 2020.6.26

1. 伪分布式集群的搭建

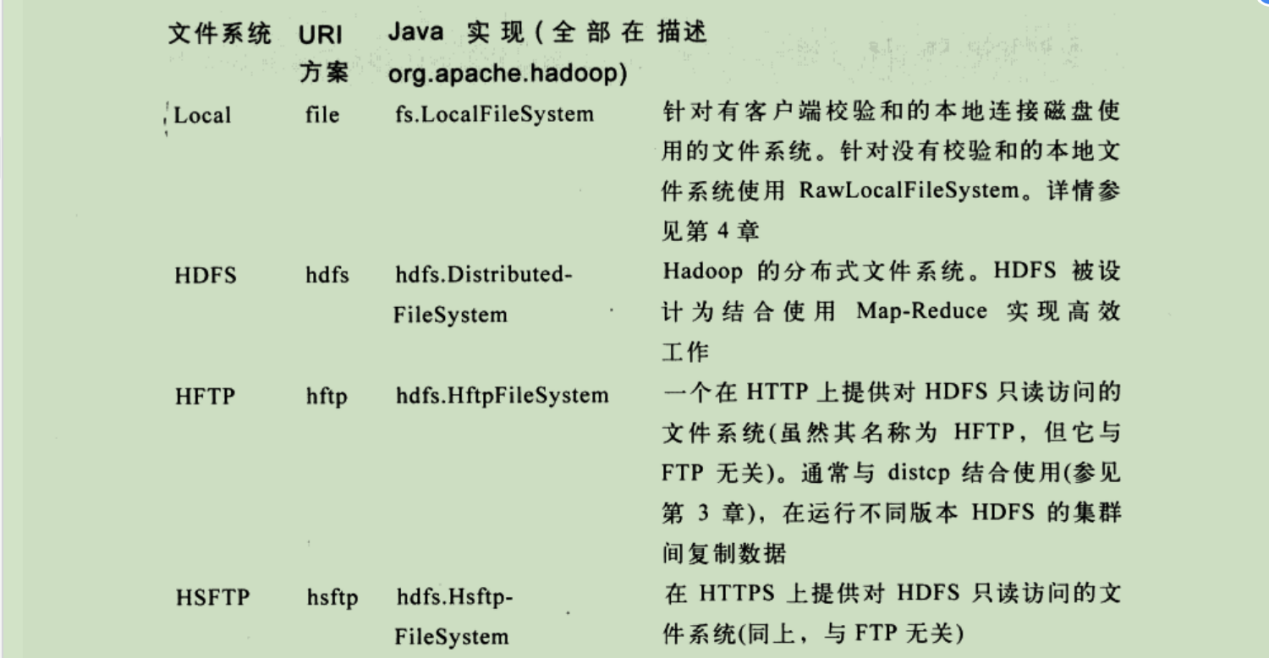
这种模式也是在一台单机上运行，但用不同的[Java](http://lib.csdn.net/base/java" \o "Java 知识库" \t "https://www.cnblogs.com/wooluwalker/p/_blank)进程模仿分布式运行中的各类结点: (NameNode,DataNode,JobTracker,TaskTracker,SecondaryNameNode)

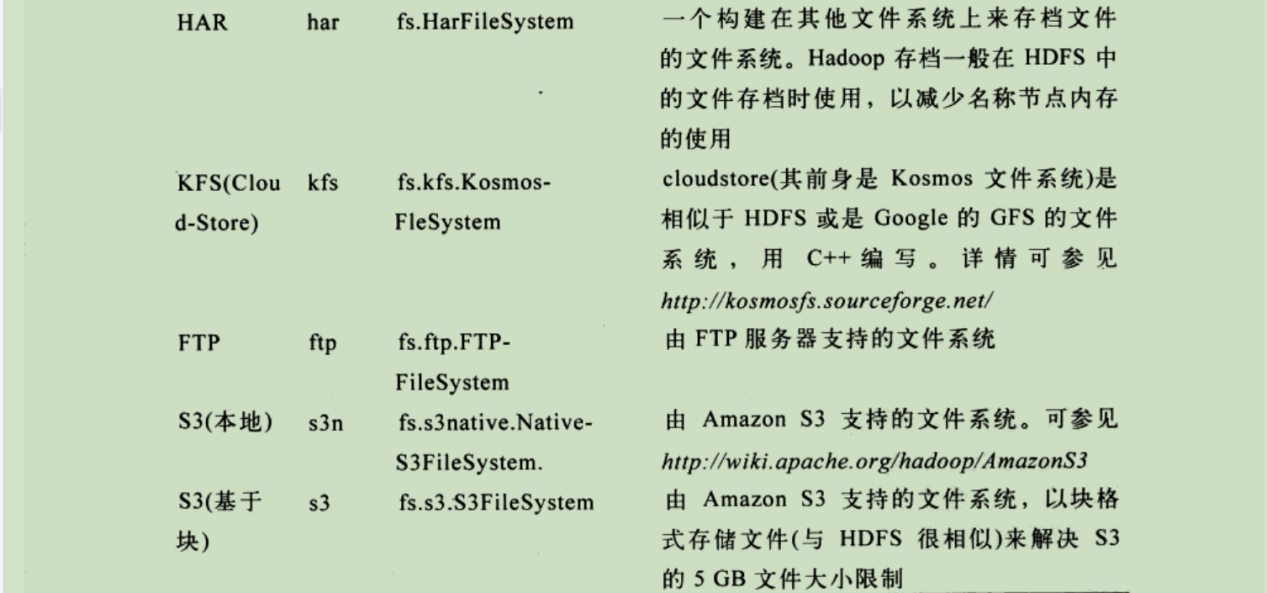
分布式运行中的这几个结点的区别：

1. 从分布式存储的角度来说，集群中的结点由一个NameNode和若干个DataNode组成,另有一个SecondaryNameNode作为NameNode的备份。
2. 从分布式应用的角度来说，集群中的结点由一个JobTracker和若干个TaskTracker组成，JobTracker负责任务的调度，TaskTracker负责并行执行任务。TaskTracker必须运行在DataNode上，这样便于数据的本地计算。JobTracker和NameNode则无须在同一台机器上。一个机器上，既当namenode，又当datanode,或者说 既 是jobtracker,又是tasktracker。没有所谓的在多台机器上进行真正的分布式计算，故称为"伪分布式"。开启多个进程模拟完全分布式，但是并没有真正提高程序执行的效率

在搭建的过程与单机模式没有太大区别，主要是分布式的搭建，明天我会继续大前天留下的进度继续和小组成员尝试。

1. Hadoop文件系统的具体实现





三、总结

1.搭建Hadoop伪分布式配置java环境时，输入java -version 报错，经查询后得知内存不足，在/etc/profile里增加一个系统环境变量：\_JAVA\_OPTIONS=-Xmx512M，保存后即可。

2.Hadoop的应用有点模糊，接下来两天会继续实践

3.复习了git的使用

4.参考CSNDN：<https://blog.csdn.net/qq_33689414/article/details/79223186?utm_medium=distribute.pc_relevant_t0.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase&depth_1-utm_source=distribute.pc_relevant_t0.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase>

<https://blog.csdn.net/DanBo_C/article/details/104832904>