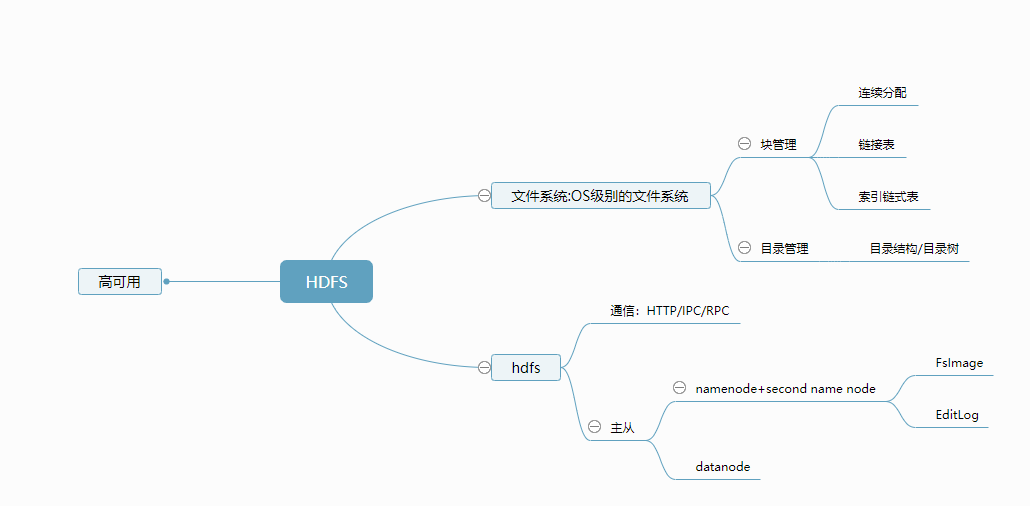
Hadoop的本质还是对操作系统中文件系统的分装，通过网络通信的方式将不同的机器，不同的节点联系起来。其设计思想与Linux操作系统大同小异，从接口名称INode可以看出，致敬了Linux。通俗一点说，分布式特点，有内存通信变为网络通信，hadoop抽象出目录结构，来保证分布式文件系统对开发者来说是透明的。他们之间主要通过socket，rpc，http等，实现注册，心跳，握手，保活，重连，续租等。主从结构中，name node（确切的说name node + second name node才是完整的name node，这两个一般不再一个节点上）为主，datanode为从。有没有想过为什么处于活跃状态的主节点只要一个，因为代码里有加锁的部分，如果是多个，就会出现不一致的问题，甚至还有一些事物问题。



但是，注意但是，hdfs目前自身好像不支持热备哟，这就是接下来我们要去看的。

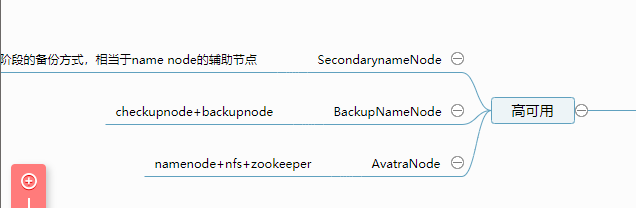
Name node

Name node

Second namenode

Namenode

Hdfs的高可用：



Hdfs的高可用并不是绝对的高可用，而且其发展也经历了很长一段时间，前期基本处于冷备状态，后来逐渐有了backup node 计划，facebook也废了解决热备问题：AvatarNode ，后来CDH也基于此种方案发展了自己的高可用。

Secondary name node：理解为checkup node辅助namenode合并生成fsimage checkup文件

Backupnode，有两部分，checkup node + backup node，checkupnode做的仍然是secondarynamenode的事情，而backup name node会直接同步name node的日志，并在本地也有一份fsimage 和 editlog(secondarynamenode没有哦)，name日志会有两个出处，一个是自己的editlog，一个是backup name node的内存，这backup name node直接在内存进行合并，不会向second name node 先把文件下载下来，然后在读到内存进行合并，在写到磁盘。而且当name node宕机的时候，可以启动backup node节点作为那么node，启动的时候注意启动命令即可。

AvatarNode：具备两个以上的那么node，但是只有一个提供服务，日志统一写到nfs上，nfs为独立节点，并且datanode的信息会同时向两个name node发送信息。当一个宕机的时候，可以通过zookeeper来启动自己实现热备。