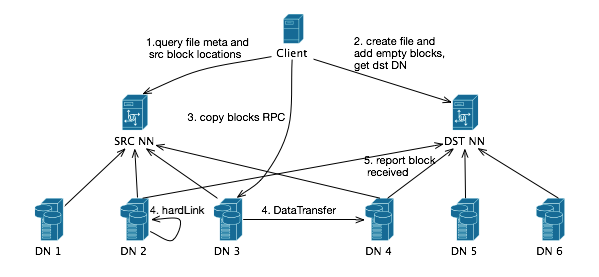
HDFS Fastcopy

——Hardlink

HDFS原始实现两个集群间的数据复制使用的是DistCp的方式，主要通过提交Job的并行读取文件数据，写入目标文件，这种方式需要使block数据经由源集群DN->Client节点->目标集群，效率较低。

Facebook实现Fastcopy来提升集群间copy的效率，基本结构如下：



步骤如图所示：

1) 查询源NS中文件的meta信息，获取源文件所有的block信息

2) 对于每个Block，获取其在原集群中的location信息

3) 对于源文件中的每个block，在目标NS中的文件上添加空的block信息

4) 对于所有的源Block，通过DN的copyBlock接口实现local copy

5)每个目标DN在完成block的copy之后向目标NS的NN中报告接收的Block

6)等待所有的block都copy完成后退出

1、File Meta的复制

FastCopy首先获取源文件的meta信息(FileStatus)和blocks locations(LocatedBlocks)信息。检查源文件是否处于构建中，LocatedBlocks.isUnderConstruction()，如果是则跳过该文件。在目标NS中创建目标文件， 副本数，permission， blockSize等信息与源文件一致，为避免目标文件已存在，默认使用覆写模式创建

2、Block的复制

对于源文件的LocatedBlocks中所有的block信息，进行Block复制：通过向目标NS中addBlock获取目标NS中block的DN 列表。对源block的DN列表和目标Block的DN列表进行排序对齐，使相同的DN在各自的列表中的位置相同。通过源Datanode的copyBlock()接口实现想目标DN的block数据复制

DataNode.copyBlock()具体分为三种情况：

1) 同一DN实例(同一节点上不同DN实例，但是目前不存在这种情况)

存在于Federation中同一DN节点服务于两个NS的情况，只需要为源Block的文件创建一个HardLink指向目标block文件

2) 通常源DN和目标DN位于不同节点上，通过网络传输block数据，这个传输的过程与Client向DN写入Block数据基本一致，因此可以直接使用DataTransfer来完成