ll可以查看需要操作文件夹的权限

studo 可以申请这次操作所需的权限

studo chown 所有者名称:所有者的主 文件目录

hadoop的文件目录

bin可执行文件

etc配置文件

lib库，本地库，用于调用非java方法，而是操作系统的部分

sbin可执行脚本

share所有的jar包

指定MR运行在YARN上

FDFS的设计理念：

硬件常态故障。

吞吐量比相应时间更重要

支持大文件，GB到TB的大小文件

一次写入多次读取的模型

移动代码代价比移动数据代价低

可移植性

FDFS

是一个文件系统，目录树

是一个分布式的

主从架构

分块存储，hadoop会把大于128m的文件切分存储

命名空间

namenode负责管理数据

目录文件分块位置信息叫做元数据，有namenode负责维护

datanode负责存放数据

文件的block具体存储有datanode节点负责，每一块block可以在多个datanode上。datanode需要定时向namenode汇报自己持有的block信息。

一次写入多次读出，且文件不支持修改。

修改不方便，高吞吐低响应，网络开销大。不能当做网盘使用。

hadoop提供了shell命令的客户端

hadoop fs <args>

hadoop集群的结构：

HDFS集群：

NameNode、DataNode、SecondaryNameNode

YARN集群:

ResourceManager、NodeManager

MapReduce框架：运行在yarn集群上的编程框架，负责海量数据的分布式计算。

InputFormat帮助我们把数据分割成行数据