

# HDFS 概述

1. 分布式
2. commodity hardware (low cost) 去IOE
3. fault-tolerance
4. High throughput
5. big data sets

HDFS是分布式文件系统

普通文件系统 : Linux(ET2, ET4), Windows(FAT, NTFS), Mac。。。。

文件目录 : C /

存放的是文件或者文件夹

对外提供服务 : 创建, 修改, 删除, 查看, 移动

普通文件系统 vs 分布式文件系统

单价 vs 横跨n个机器

## HDFS assumptions and goals

- 1) hardware failure  
    blocksize = 128M  
    block 存在不同机器上, 3 copies
- 2) streaming data access