

HDFS 分布式文件系统设计目标

- 1、 硬件错误 由于集群很多时候由数量众多的廉价机组成,使得硬件错误成为常态 (磁盘损坏 屌丝机)
- 2、 数据流访问 所有应用以流的方式访问数据,设置之初便是为了用于批量的处理数据,而不是低延时的实时交互处理
- 3、 大数据集 典型的 HDFS 集群上面的一个文件是以 G 或者 T 数量级的,支持一个集群当中的文件数量达到千万数量级
- 4、 简单的相关模型 假定文件是一次写入,多次读取的操作 (假设 不会频繁写)
- 5、 移动计算比移动数据便宜 一个应用请求的计算,离它操作的数据越近,就越高效 (代码到数据)
- 6、 多种软硬件的可移植性