The 4th China Cloud Computing Conference

中国电子学会 Chinese Institute of Electronics

第四届中国云计算大会

5月23-25日 北京·国家会议中心

云存储技术实践

七牛云存储 创始人
许式伟

自我介绍

- 七牛云存储
  - 创始人，CEO
- 盛大
  - 前盛大网盘负责人
  - 前盛大祥云计划（盛大云前身）负责人
- 金山
  - 前金山实验室负责人，云存储团队组建者
  - 前金山技术总监，WPS2005首席架构师

大纲

- 云存储技术
  - 数据模型
  - 可靠性
  - 可用性
  - 伸缩性
  - 性能/成本

云存储技术

- 要点
  - 本文讨论云存储技术的共性需求，要点难点，以及可能的对策。
  - 因方案细节差异可以非常之大，本文不讨论具体的某个云存储技术方案。

云存储技术

- 功能属性
  - 键值存储（KV）
  - 数据库（DB）
  - 文件系统（FS）
- 质量属性
  - 数据尺寸
    - 结构化数据：小数据，普遍一行数据不到16K
    - 非结构化：大数据，普遍在 256K 甚至M、G级别
  - 访问特征
    - 读多写少 - 优化读
    - 写多读少 - 优化写
  - 技术指标

云存储技术

- 技术指标
  - 可靠性：不丢数据
  - 可用性：随时可访问
  - 伸缩性：随着集群访问压力、数据规模的增大，性能不能有显著的降低
  - 速度：快，更好的用户体验
  - 低成本
    - 单位空间的硬件成本更具竞争力
    - 自动化运维：降低人工成本

可靠性

- 数据冗余
  - 多副本
  - RAID
  - EC
- 异地容灾
  - 多IDC备份

可靠性

- 挑战1：数据一致性
  - 数据有多份副本，必然带来一致性问题：不同副本的数据不同的时候，听谁的。
  - 数据一致性是个大问题。
  - 数据一致性和高性能是一对矛盾。
  - 有时需要容忍读到旧版本数据。
    - 但不能容忍读出来的数据，前半部分是旧的，后半部分是新的。
  - 对策：
    - 主从结构
    - 版本号（或时间戳）

可靠性

- 挑战2：数据修复
  - 当机器的磁盘损坏的时候，需要将该磁盘的数据搬到其他磁盘。
  - 关键点
    - 如何计算出所丢失的数据。
    - 如何搬数据。
    - 数据修复不能影响集群的正常工作。
      - 最好能够感知集群当前的负荷，以此适配修复的速度。
  - 技术参数
    - 数据恢复时间：影响集群可靠性的最关键指标。

可靠性

- 挑战3：如何降低成本
  - 三份副本的代价：硬件成本 x 3
  - 对策：用CPU换空间
    - RAID
    - EC

可用性

- 任何机器都可以挂掉
  - 消除集群单点
- 机房也是个问题
  - 地域问题
    - 有的机房在部分地区不可访问
  - 机房故障
    - 整个机房可能发生临时不可访问

可用性：多层次解决

- 服务器：杜绝单点
  - Load Balance
  - 主从
- DNS
- 客户端
  - 自行选择可用的机房

伸缩性

- 分摊访问压力
  - 避免Master单机热点
    - 写压力：只能将Master集群化
    - 读压力：可由Slave分担压力，亦可在Client加缓冲
  - 对策：压力转移、Load Balance
- 数据规模压力
  - 算法复杂度
  - 对策：避免出现与数据量呈线性相关的运算

速度

- 地域问题
  - 就近访问 IDC
  - 优化路由
- 架构上
  - 优化 Cache 结构，提高命中率

成本

- 降低数据冗余
  - 多副本 => RAID/EC
- 公有资源：P2P网络
  - 将成本降到趋近于 0
- 在保证可靠性的前提下省成本
  - 永远的矛盾：可靠 <=> 低成本

成本

- 降低运维成本
  - 自动化运维
  - 无单点故障
  - 优化运维流程，可以自动化的尽量自动化

Q & A

xushiweizh@gmail.com
@许式伟

第四届中国云计算大会

THANKS