云计算

第7讲

云存储

任桐炜，李传艺

南京大学软件学院

2017-10-18

目的

- 构建上层应用的基础

- 对外提供存储服务

非结构化数据存储

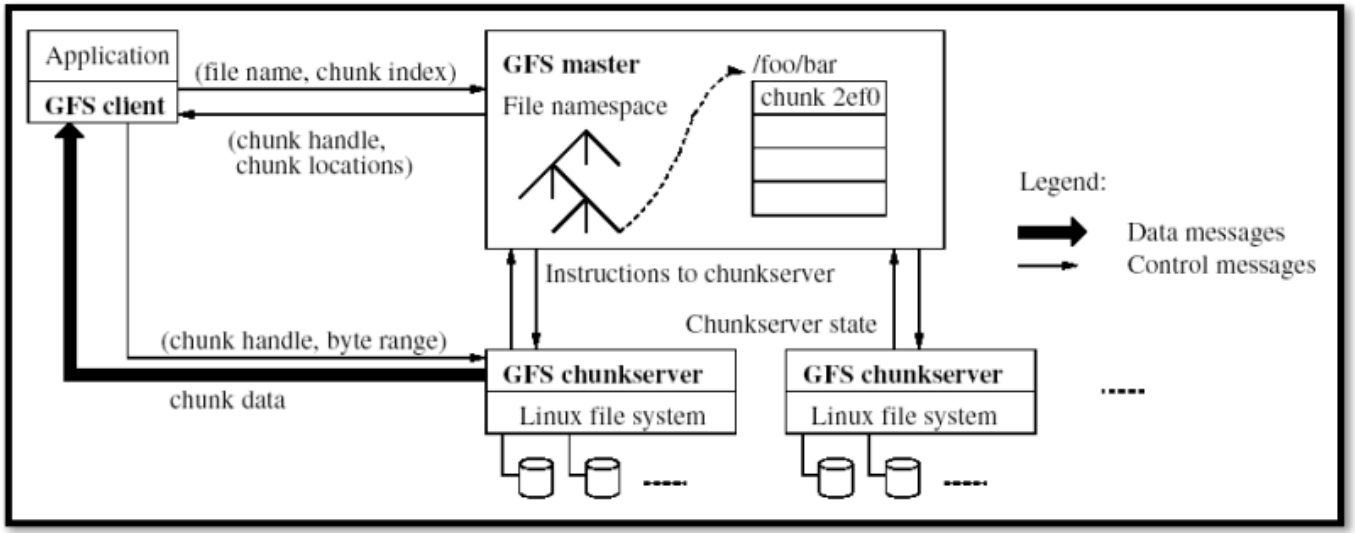
- 非结构化数据
  - 文本、图像、音频、视频……
- 假设
  - 海量的大尺寸数据（文件尺寸是GB或者TB量级）
  - 依靠廉价、不可靠的硬件（硬件出错是正常而非异常）
  - 对响应时间要求不高

解决思路

- 磁盘存储中如何处理大小不一的文件？
  - 分块
- 分块的大小？
  - 根据应用来确定
- 硬件出错造成数据丢失？
  - 通过冗余来提高可靠性（多个副本存放在不同服务器上）

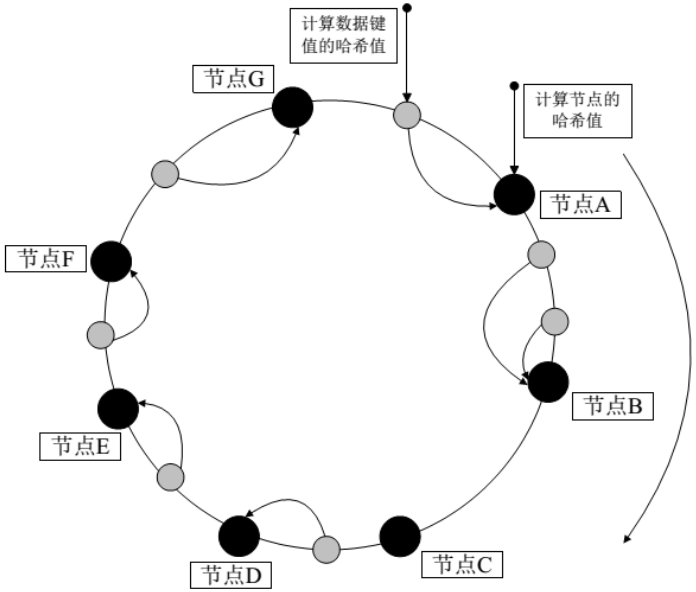
分块的管理

- 主从模式（**Google GFS**）
  - 优点：方便添加服务器和负载平衡，不存在一致性问题
  - 缺点：单点故障，性能瓶颈



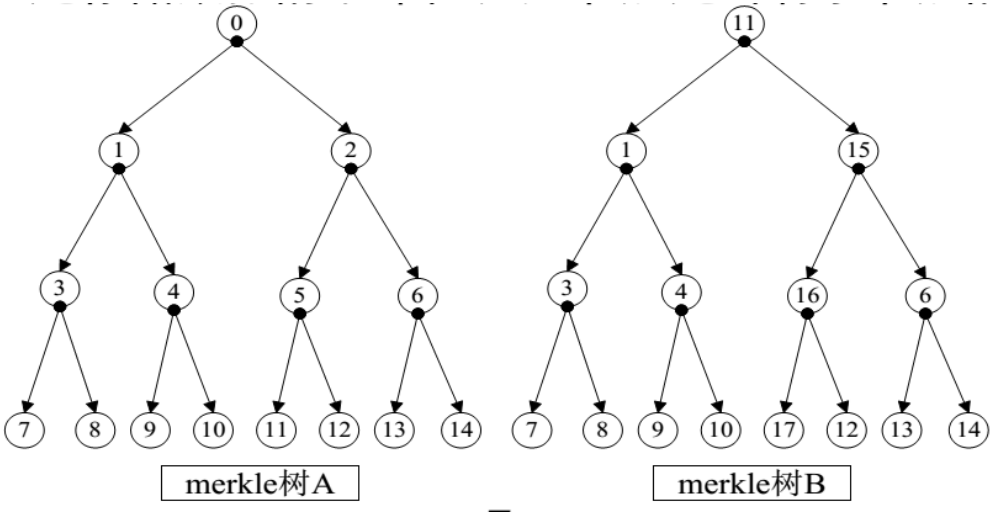
分块的管理（续）

- 无中心模式（**Amazon Dynamo**）
  - 优点：不存在单点故障和性能瓶颈
  - 缺点：负载均衡，一致性问题



分块出错

- 主从模式
  - 每个块进一步分成若干个小块，每个小块有一个校验码
- 无中心模式
  - **Merkle哈希树技术**
    - 叶节点是存储数据的哈希值，父节点是所有子节点的哈希值



服务器出错

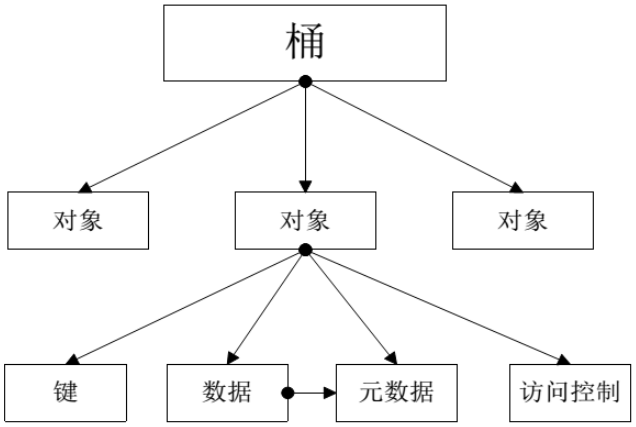
- 主从模式
  - 主节点出错
    - 无响应
    - 通过日志容错
  - 块服务器出错
    - 管理软件监测或心跳机制（主动报告）
    - 恢复服务器上每个块
- 无中心模式
  - 闲聊机制（每个服务器定期随机向另一个服务器发送消息）
  - 恢复服务器上的每个块

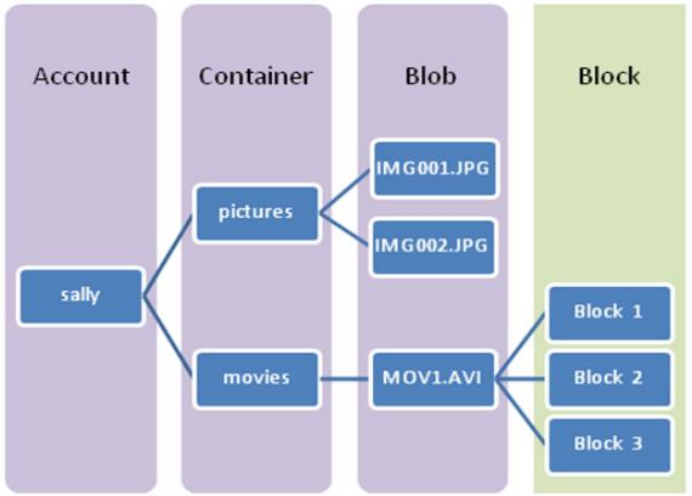
数据一致性

- 强一致性模型
  - 要求任何时刻所有的数据副本一致
- 最终一致性模型
  - 只要最终所有的数据副本一致
  - 牺牲一致性来提高可靠性和可用性

非结构化数据存储服务

- 简单存储服务S3（Amazon）
  - 架构在Dynamo上
  - 结构：桶，对象（键）
  - 数据一致性：最终一致性模型
- Blob（Microsoft）
  - 所有数据资源用URI方式标记
  - 数据模型：账户=>容器 =>Blob=>Block





结构化数据存储

- 商业数据库的问题
  - 无法满足应用需求
  - 难以部署
- 假设
  - 支持多个种类的数据
  - 响应海量的服务请求
  - 具有很好的可扩展性

解决思路

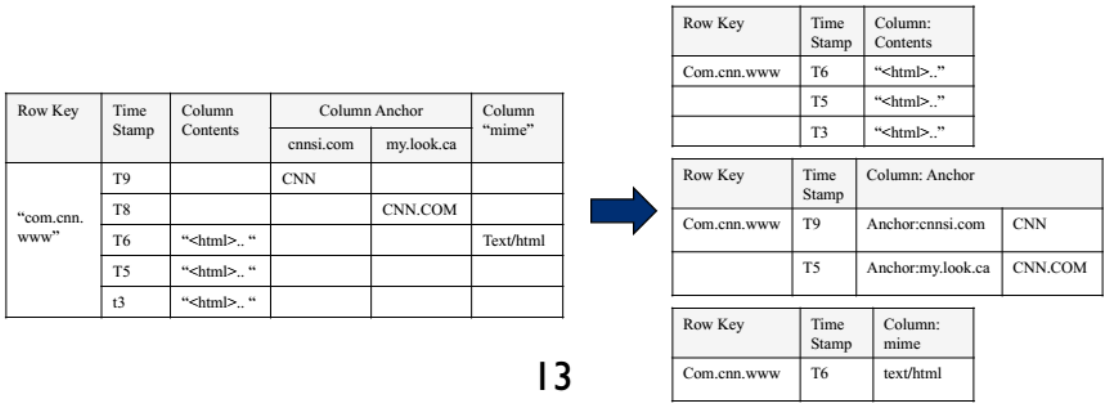
- 如何适应不同数据类型
  - 不考虑存储数据的具体类型
- 如何解决可扩展性？
  - 以非结构化数据存储机制为基础
- 如何实现数据的分块？
  - 将表分割成子表或者转换成其它易于分割的形式
  - 可以解决数据稀疏的问题

数据模型

- **BigTable（Google）**

```
(row:string, column:string, time:int64)->string
```

  - 行：表中的数据根据行关键字按词典序排序
  - 列：按照列族存储，每个族中的数据属于同一个类型
  - 时间戳：保存不同时期的数据
  - 物理划分：表 => 子表 => SSTable文件



Row Key	Time Stamp	Column Contents	Column Anchor		Column "mime"
			cnnsi.com	my.look.ca	
"com.cnn. www"	T9		CNN		
	T8			CNN.COM	
	T6	"<html>.. "			Text/html
	T5	"<html>.. "			
	t3	"<html>.. "			

Row Key	Time Stamp	Column: Contents
Com.cnn.www	T6	"<html>.."
	T5	"<html>.."
	T3	"<html>.."

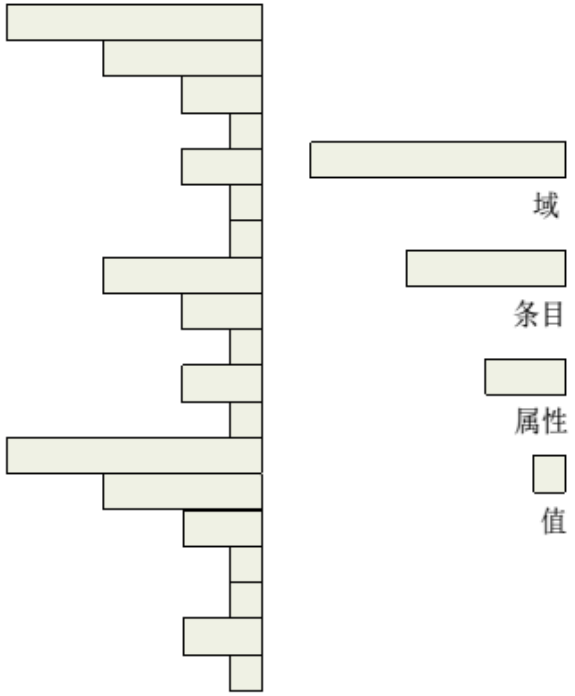
Row Key	Time Stamp	Column: Anchor	
Com.cnn.www	T9	Anchor:cnnsi.com	CNN
	T5	Anchor:my.look.ca	CNN.COM

Row Key	Time Stamp	Column: mime
Com.cnn.www	T6	text/html

13

数据模型（续）

- **SimpleDB（Amazon）**
  - 树状结构
  - 域：数据库操作的基本单位
  - 条目：域内命名唯一，不需要事先定义模式
  - 属性：条目的特征
  - 值：允许多值属性



数据模型（续）

- **Table数据模型（Microsoft）**
  - **账户 => Table => 实体 => 属性**
  - **物理划分：**
    - **通过PartitionKey和RowKey来唯一标识一个实体**
    - **分割时应考虑负载均衡和高效查询**

Partition Key Document Name	Row Key Version	Property 3 Modification Time	.....	Property N Description	
Examples Doc	V1.0	8/2/2007	.....	Committed version	Partition 1
Examples Doc	V2.0.1	9/28/2007		Alice's working version	
FAQ Doc	V1.0	5/2/2007		Committed version	
FAQ Doc	V1.0.1	7/6/2007		Alice's working version	Partition 2
FAQ Doc	V1.0.2	8/1/2007		Sally's working version	

云存储服务

- 云存储服务
  - 专注于向用户 提供以互联网为基础的在线存储服务。
  - 用户无 需考虑存储容量、存储设备类型、数据存储位置以及数据的可用性、可靠性和安全性等繁琐的底层技术细节,根据需要付费就可以从云存储服务提供商那里获得近乎无限大的存储空间和企业级的服务质量。
- 三个基本特征（**ITProPortal.com**）
  - 分布于网络（互联网或局域网）
  - 易于扩展
  - 易于管理

推动因素

- 数据量的快速持续增长
  - 个人或企业拥有的存储设备无法满足待存储数据量的需求
- 数据同步和异地访问
  - 不同设备之间的数据同步
  - 异地的协同工作和数据访问
- 数据备份
  - 防止数据丢失

影响因素

- 网络宽带
  - 需要实现大容量数据传输来提供便利的云存储
- 应用存储
  - 通过在存储设备中集成应用软件功能来提高性能和效率
- 集群技术、分布式文件系统和网络计算技术
  - 需要实现各个存储设备之间的协同工作
- 网络存储安全技术
  - 保证数据传输安全及数据不会丢失
- 存储管理技术
  - 多地域、多厂商、多硬件设备之间的传输管理

行业标准

- Overview of Cloud Storage
- Common Operations
- Interface Specification
- Data Objects
- Container Objects
- Domain Objects
- Queue Objects
- Capability Objects
- Exported Protocols
- Snapshots
- Serialization/Deserialization
- Metadata
- CDMI Logging
- Retention and Hold Management

存在的问题

- 客户对云存储和自身需求了解不足
  - 系统是否具有“无限”扩展的需求
  - 软硬件升级的费用是否具有边际效益
  - 存储系统升级是否会产生显著影响
  - 数据备份和数据灾难所产生的成本
- 市场定位模糊
  - 大中型企业不愿舍弃原有的IT设施
  - 小型企业无法承担云存储数据中心的维护费用

存在的问题（续）

- 安全感的缺失
  - 可控性不强
    - 断网、断电，时延过长，数据迁移，取回数据
  - 法律保护不够
    - 如果别人的数据遭到突击搜查，我的数据也会被搜查么？
- 公有云与私有云之争
  - 公有云：数据托管
    - 承担存储容量租金和带宽使用费，不需要硬件或技术知识
  - 私有云：所有数据完全由内部IT员工控制
    - 需要硬件和软件，不承担存储容量的租金和带宽使用费

谢 谢