



云数据库

单 位：重庆大学计算机学院



大数据、云计算带来了什么变化？

主要学习目标

- 云数据库



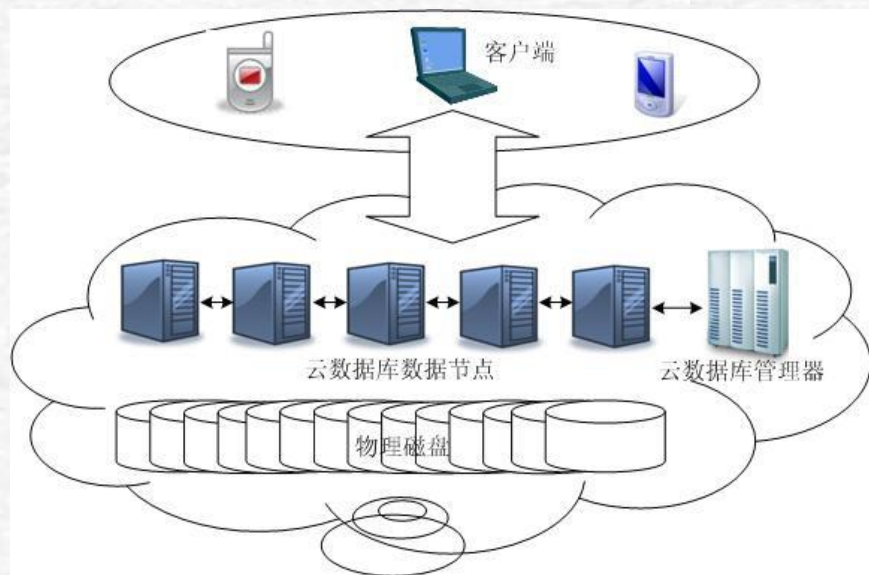
思考问题

- 自己配置数据库服务器和使用网络上的数据库服务器相比，优势和劣势是什么？

一 云数据库

- 云计算的迅猛发展使得数据库部署和虚拟化在“云端”成为可能。
- 云数据库就是数据库部署和虚拟化在云计算环境下，通过计算机网络提供数据管理服务的数据库。
- 云数据库可以共享基础架构，极大地增强了数据库的存储能力，消除了人员、硬件、软件的重复配置。
- Gartner预测，云将主导数据库市场的未来，到2022年，75%的数据库将被部署或迁移至云平台，只有5%的数据库会考虑部署在本地。（2019）

一 云数据库



云数据库应用示意图

在云数据库应用中，客户端不需要了解云数据库的底层细节，所有的底层硬件都已经被虚拟化，对客户端而言是透明的，它就像在使用一个运行在单一服务器上的数据库一样，非常方便容易，同时又可以获得理论上近乎无限的存储和处理能力。

云数据库概念

• 云数据库是部署和虚拟化在云计算环境中的数据库

云数据库特点

- 动态可扩展
- 高可用性
- 较低的使用代价
- 易用性
- 大规模并行处理

Cloud Database

海量存储需求的必然选择

一 云数据库

- 云数据库将传统的数据库系统配置在“云上”，有专门的云服务提供商进行这些“云上”数据库系统的管理和部署工作。
- 用户需要通过付费的方式就能获取数据库服务。
- 不同于传统的数据库，云数据库通过计算存储分离、存储在线扩容、计算弹性伸缩来提升数据库的可靠性和可用性。
- 代表性的云数据库：亚马逊Aurora，微软SQL Azure，**华为的 GaussDB（for mysql）**。

一 云数据库

云数据库与自建数据库对比

对比项	云数据库	自购服务器搭建数据库服务
服务可用性	高可用架构提供高可用性。	需自行保障，自行搭建主备复制，自建RAID等。
数据可靠性	自动主备复制、数据备份、日志备份等。	需自行保障，自行搭建主备复制，自建RAID等。
系统安全性	防DDoS攻击，流量清洗；及时修复各种数据库安全漏洞。	自行部署，价格高昂；自行修复数据库安全漏洞。
数据库备份	自动备份。	自行实现，但需要寻找备份存放空间以及定期验证备份是否可恢复。
软硬件投入	无软硬件投入，按需付费。	数据库服务器成本相对较高，对于SQL Server还需支付许可证费用。
系统托管	无托管费用。	每台2U服务器每年超过5000元（如果需要主备，两台服务器需超过10000元/年）。
维护成本	无需运维。	需招聘专职DBA来维护，花费大量人力成本。
部署扩容	即时开通，快速部署，弹性扩容。	需硬件采购、机房托管、机器部署等工作，周期较长。
资源利用率	按实际结算，100%利用率。	由于业务有高峰期和低峰期，资源利用率很低。

云数据库的影响

Cloud

影响

Database

1、极大地改变企业管理数据的方式

- Forrester Research 分析师 Noel Yuhanna 指出，18%的企业正在把目光投向云数据库。

- 中小企业会更多地采用云数据库产品，但是，对于大企业而言，云数据库并非首选，因为大企业通常自己建造数据中心。

2、催生新一代的数据库技术

- 第一代是20世纪70年代的早期关系数据库
- 第二代是80到90年代的更加先进的关系模型
- 第三代的数据库技术，要求数据库能够灵活处理各种类型的数据，而不是强制让数据去适应预先定制的数据结构。

- 从数据模型设计方式来看，已经有些产品（比如SimpleDB、HBase、Dynamo、BigTable）放弃传统的行存储方式，而采用键/值存储，从而可以在分布式的云环境中获得更好的性能。

3、数据库市场份额面临重新分配

- 此前，Teradata、Oracle、IBM DB2、Microsoft SQL Server、Sybase等传统数据库厂商垄断市场

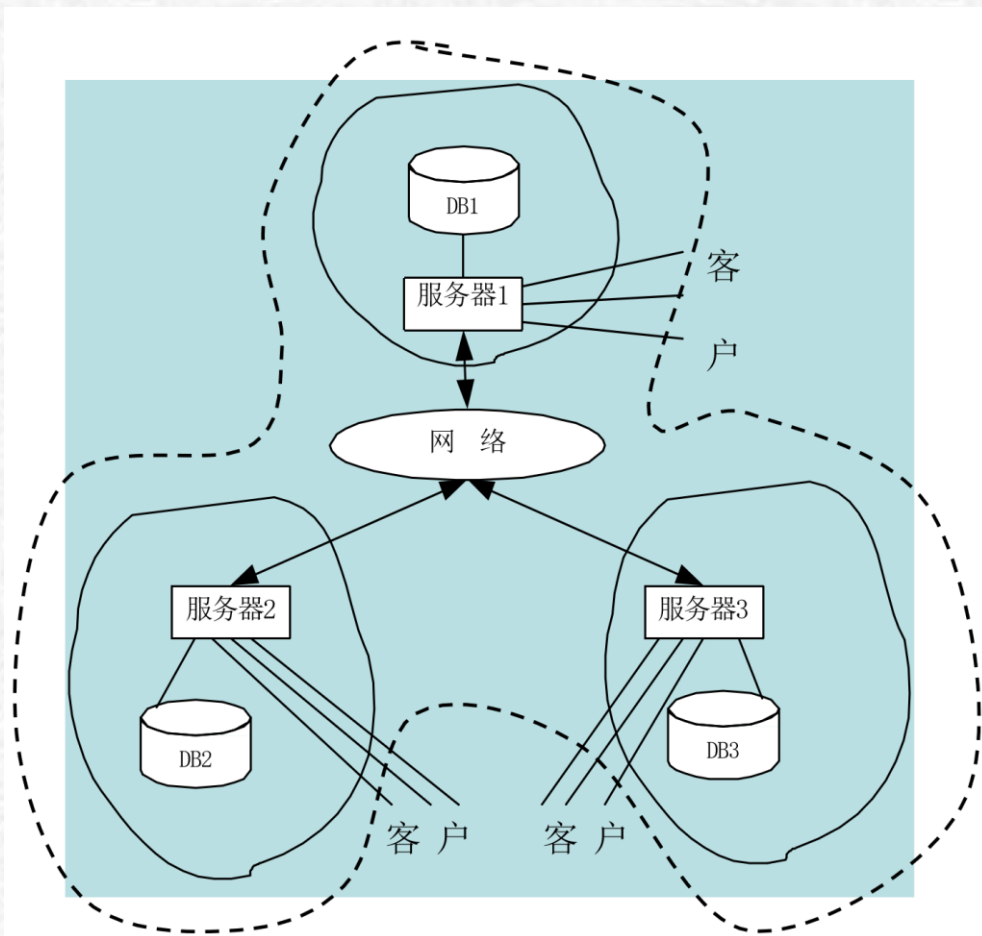
- Amazon和Google扮演引领者角色

- 新的云数据库厂商开始出现
Vertica和EnterpriseDB

二 云数据库的服务

- 灵活的使用数据库。云数据库服务可以根据需要进行购买，如果有一个应用需要在晚上运行，白天不运行，云数据库就可以很容易实现。
- 硬件和相关的基础设施由服务提供商提供。
- 提供VPC (virtual private cloud) 服务。
- 云服务库提供商会提供多个位置的服务。灾难恢复、负载均衡。
- 云供应商提供了扩展服务、提供容错和执行负载均衡的先进方法。
- 多界面服务（窗口、命令行、API）
- 安全服务。
- 监控和性能工具。

三 云数据库与传统的分布式数据库



分布式数据库是计算机网络环境中各场地或节点上的数据库的逻辑集合。逻辑上它们属于同一系统，而物理上它们分散在用计算机网络连接的多个节点/场地，并统一由一个分布式数据库管理系统管理。

三 云数据库与传统的分布式数据库

■ 云数据库和分布式数据库的共同点

- 云数据库和传统的分布式数据库有着相似的地方， 比如，都把数据存放到不同的节点上。

■ 云数据库和分布式数据库的区别

- 分布式数据库在可扩展性方面是无法和云数据库相比的：
由于需要考虑数据同步和分区失败等开销，前者随着节点的增加，会导致DDB性能快速下降。而云数据库则具有很好的可扩展性，因为后者在设计的时候，就已经避免了许多会影响到可扩展性的因素，比如采用更加简单的数据模型、对元数据和应用数据进行分离以及放松对一致性的要求等等。

四 云数据库的类型

- 托管云数据库
- 云原生数据库

云原生数据库实例

华为 Cloud Native 数据库设计原则



解耦

- 计算与存储分离；
- 主从解耦。



近数据计算下推 (NDP)

- IO密集型负载下推到存储节点完成，比如：redo处理，页重构。



充分利用云存储 的能力

- 存储层实现独立容错和自愈服务；
- 共享访问（单写多读）；



发挥 SSD 的优势

- 避免随机写带来的写放大，减少磨损，减少小时延；
- 充分利用 SSD 的随机读性能。

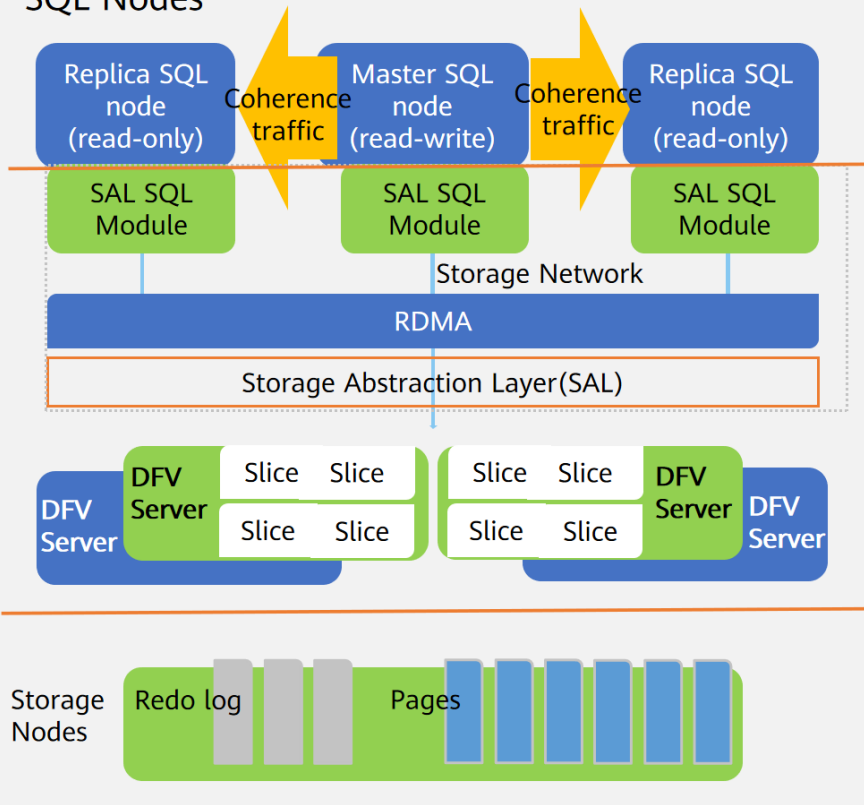


性能瓶颈已经从 计算和存储转向 网络

- 减少网络流量；
- 采用新的网络技术和硬件，比如：RDMA。

GaussDB (for MySQL) 云数据库概览

SQL Nodes



主节点

- 处理所有的写
- 写 WAL 日志 (Redo)

数据分区

- 数据页以 slice 组织, 分布在各个 slice 服务
- Slice 冗余保存
- 日志按 slice 分发

只读实例

- 处理只读请求
- 快速升主
- 快速增/删

Slice 服务

- 单数据库可以分配多个 Slice, 访问隔离
- 存储和处理日志记录
- 维护和重构数据页
- 服务数据页读请求

SQL Nodes

SQL Node

Trans
action

Buffer
Pool

SQL
Parse

SQL Node

SQL Node

Storage Node

Slice Service

Log
Write

Log
Replay

Page
Manager

Log
Read

Crash
Recovery

Log
Replication

Storage Node

Slice
Service

.....

Log
Replication

Storage Node

Slice
Service

.....

Log
Replication

- 管理客户端连接、解析 SQL 请求、分析执行计划并执行和管理事务隔离级别

- 一主多读

- 主库和只读松耦合

- 主库和只读通信少

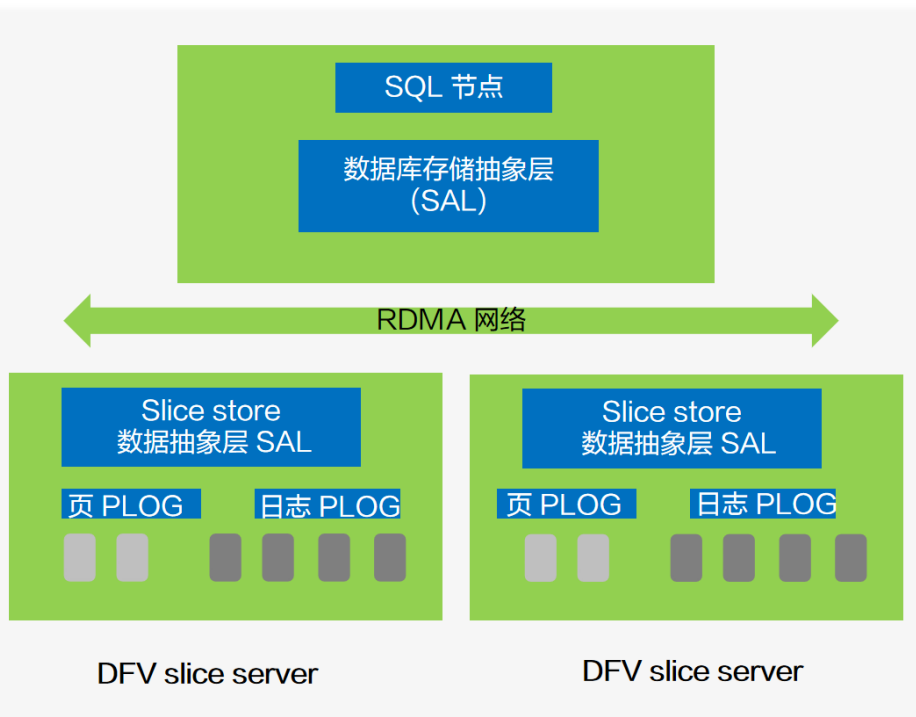
- 快速 Failover

- Query result cache

- Query plan cache

- Online DDL

存储抽象层 (SAL)



- 存储抽象层 (SAL)是逻辑层
- 将数据存储和 SQL 前端、事务、查询执行等进行隔离
- 由在 SQL 节点上执行的公共日志模块和存储节点上执行的 slice 节点组成
- SAL 将所有数据页基于 {spaceID, pageID} 划分为 slice
- 横向扩展 - 随着数据库规模的增长, 可用资源 (存储、内存) 随着 slice 的创建按比例增长.
- 数据本地化 - 数据密集型操作是在存储节点上由 slice 服务执行



随堂小测试

- 什么是云数据库？
- 云数据库和分布式数据库的区别是什么？

课堂小结和作业安排

- 基本知识：
 - 云数据库概念
 - 云数据库服务
 - 云数据库类型
- 延展性学习：
 - 云数据库的管理
- 作业
无。