厦门大学计算机科学系研究生课程

《大数据技术基础》

第11章 云数据库

（2013年新版）

林子雨

厦门大学计算机科学系

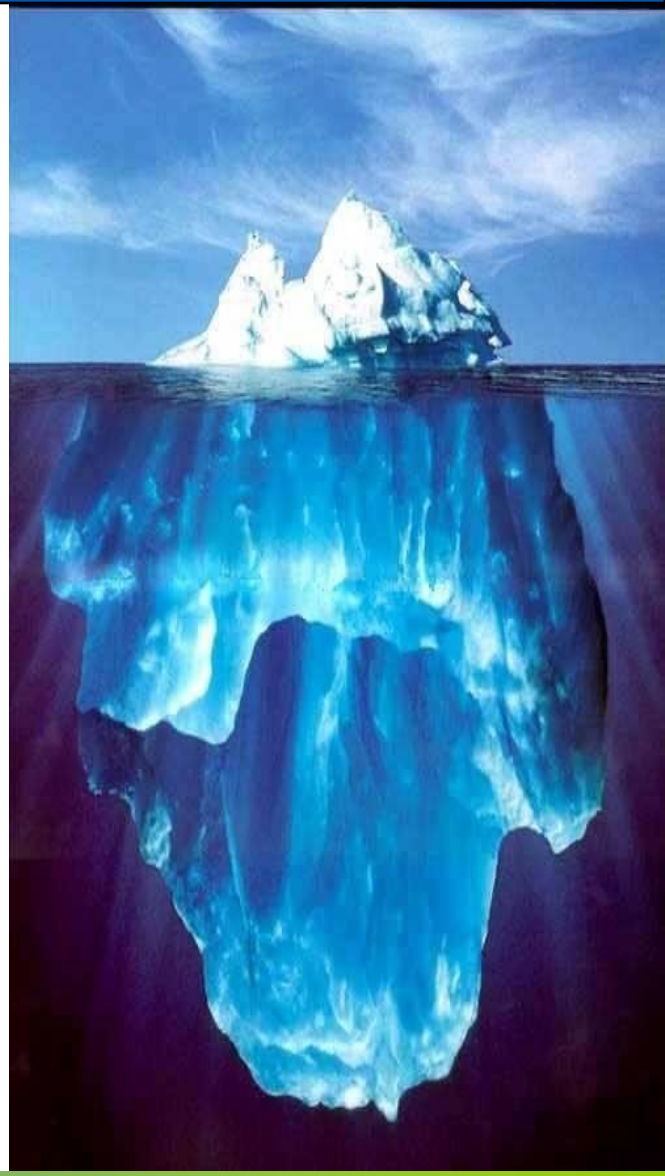
E-mail: ziyulin@xmu.edu.cn

主页：http://www.cs.xmu.edu.cn/linziyu

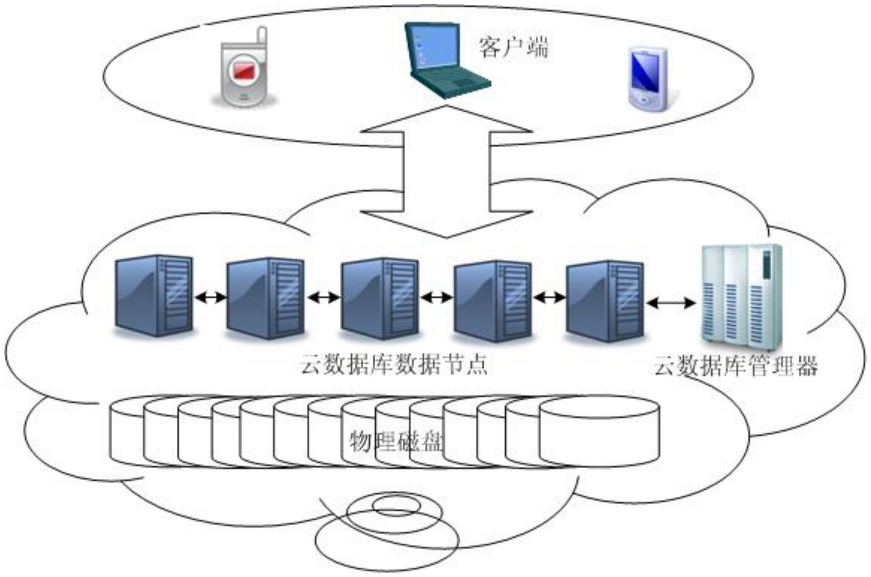
提纲

- 云数据库概述
- 云数据库的特性
- 云数据库是海量存储需求的必然选择
- 云数据库与传统的分布式数据库
- 云数据库的影响
- 云数据库产品
- 数据模型
- 数据访问方法
- 编程模型

本讲义PPT存在配套教材，由林子雨通过大量阅读、收集、整理各种资料后编写而成
下载配套教材请访问《大数据技术基础》2013班级网站：http://dblab.xmu.edu.cn/node/423



云数据库概念和特点



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

云数据库与传统的分布式数据库

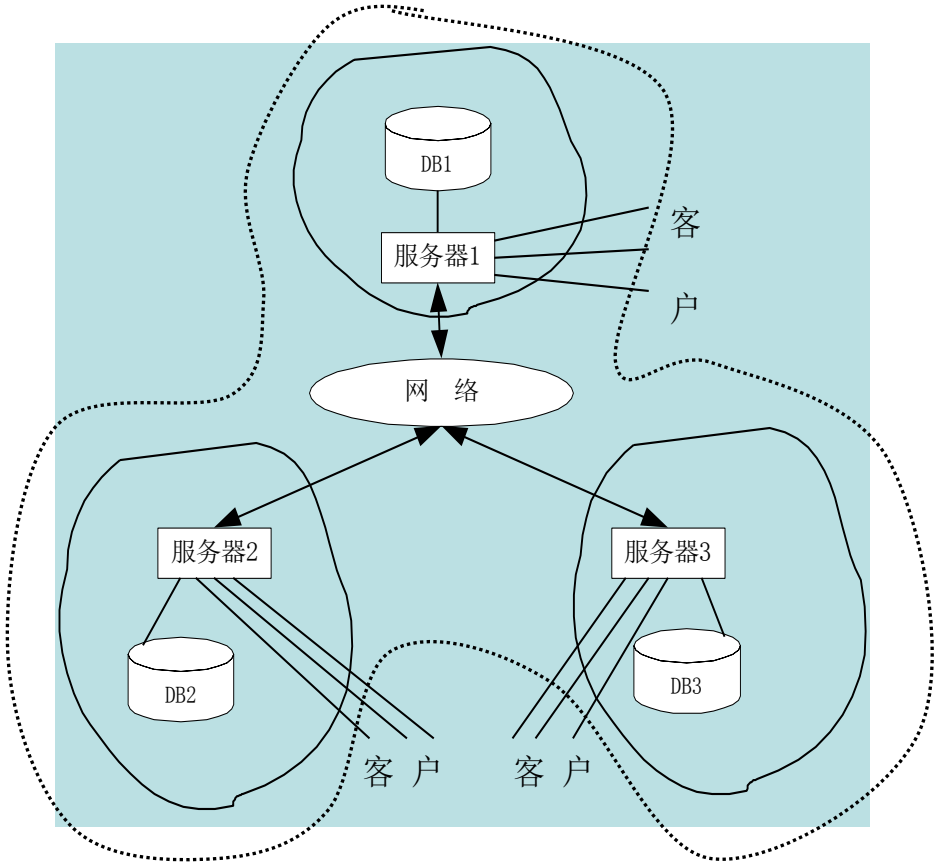


图 分布式数据库系统示意图

分布式数据库概念

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。

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。

云数据库的影响

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

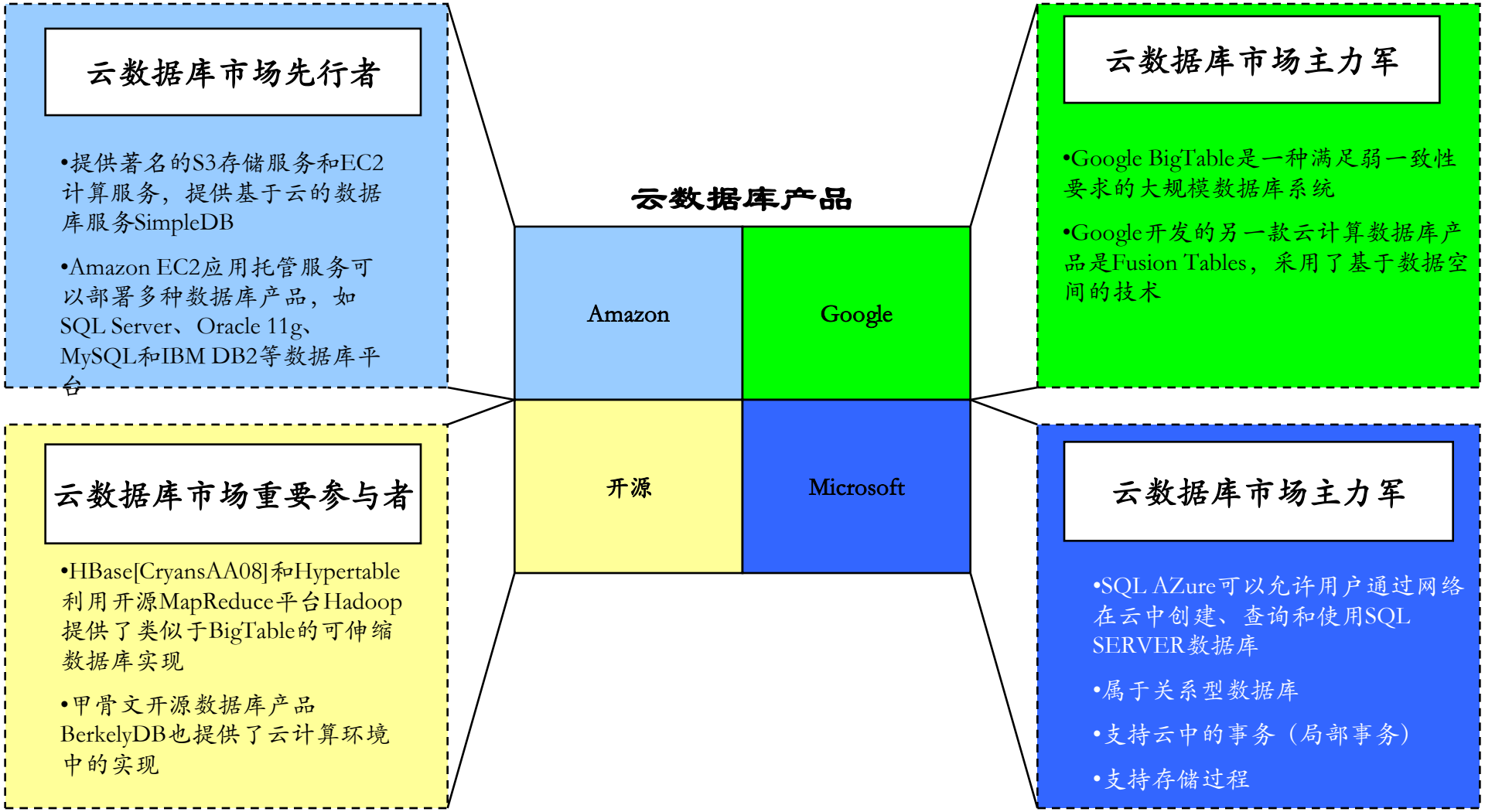
•新的云数据库厂商开始出现Vertica和EnterpriseDB

云数据库产品

- 传统的数据库厂商：Teradata、Oracle、IBM DB2和Microsoft SQL Server；
- 涉足数据库市场的云供应商：Amazon、Google和Yahoo；
- 新兴小公司：Vertica、LongJump和EnterpriseDB。

企业	产品
Amazon	SimpleDB、RDS
Google	BigTable、FusionTable、GoogleBase
Microsoft	Microsoft SQL Azure
Oracle	Oracle Cloud
Yahoo!	PNUTS
Vertica	Analytic Database v3.0 for the Cloud
EnerpriseDB	Postgres Plus in the Cloud
开源项目	HBase、Hypertable
其他	EnerpriseDB、FathomDB、ScaleDB、Objectivity/DB、M/DB:X

云数据库产品


云数据库市场先行者
•提供著名的S3存储服务和EC2计算服务，提供基于云的数据库服务SimpleDB
•Amazon EC2应用托管服务可以部署多种数据库产品，如SQL Server、Oracle 11g、MySQL和IBM DB2等数据库平台
云数据库市场主力军
•Google BigTable是一种满足弱一致性要求的大规模数据库系统
•Google开发的另一款云计算数据库产品是Fusion Tables，采用了基于数据空间的技术
云数据库产品
Amazon
Google
开源
Microsoft
云数据库市场重要参与者
•HBase[CryansAA08]和Hypertable利用开源MapReduce平台Hadoop提供了类似于BigTable的可伸缩数据库实现
•甲骨文开源数据库产品BerkelyDB也提供了云计算环境中的实现
云数据库市场主力军
•SQL AZure可以允许用户通过网络在云中创建、查询和使用SQL SERVER数据库
•属于关系型数据库
•支持云中的事务（局部事务）
•支持存储过程

云数据库领域的研究问题

云数据库领域的研究问题-数据模型

键/值模型

BigTable

一个BigTable实际上就是一个稀疏的、分布的、永久的多维排序图，它采用行键（row key）、列键（column key）和时间戳（timestamp）对图进行索引。图中的每个值都是未经解释的字节数组。

行键

■BigTable在行键上根据字典顺序对数据进行维护。对于一个表而言，行区间是根据行键的值进行动态划分的。每个行区间称为一个Tablet，它是负载均衡和数据分发的基本单位，这些Tablet会被分发到不同的数据服务器上。

列键

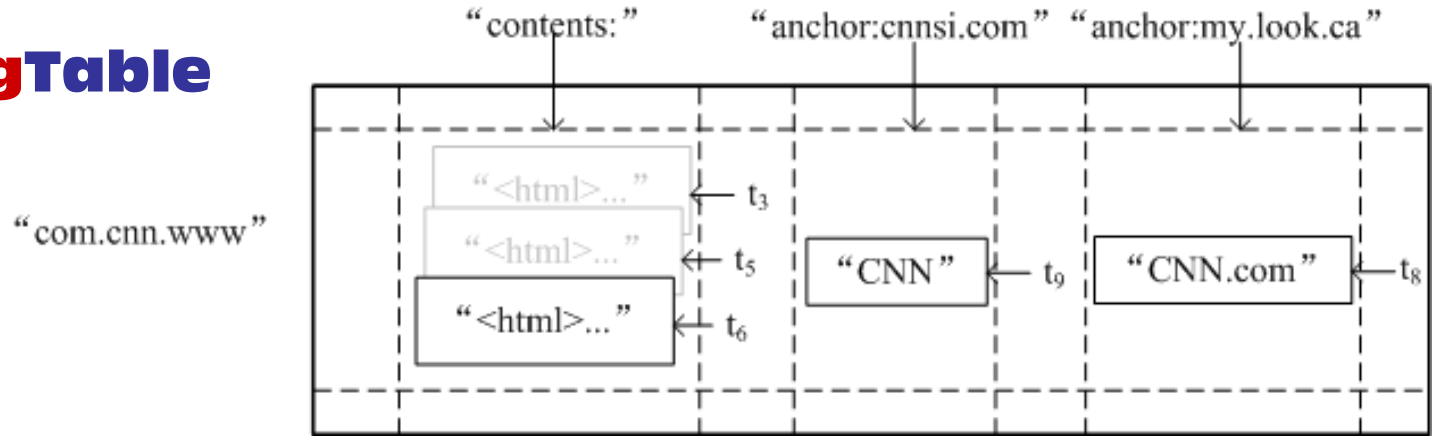
■被分组成许多“列家族”的集合，它是基本的访问控制单元。存储在一个列家族当中的所有数据，通常都属于同一种数据类型，这通常意味着具有更高的压缩率。数据可以被存放到列家族的某个列键下面，但是，在把数据存放到这个列家族的某个列键下面之前，必须首先创建这个列家族。在创建完成一个列家族以后，就可以使用同一个家族当中的列键。

时间戳

■在BigTable中的每个单元格当中，都包含相同数据的多个版本，这些版本采用时间戳进行索引。BitTable时间戳是64位整数。一个单元格的不同版本是根据时间戳降序的顺序进行存储的，这样，最新的版本可以被最先读取。

云数据库领域的研究问题-数据模型

BigTable



Row Key	Timestamp	Column Family					
		contents:	…	anchor:cnnsi.com	…	anchor:my.look.ca	…
com.cnn.www	t9			CNN			
	t8					CNN.com	
	t7	…		…		…	
	t6	<html>…					
	t5	<html>…					
	t4	…		…		…	
	t3	<html>…					

云数据库领域的研究问题-数据模型

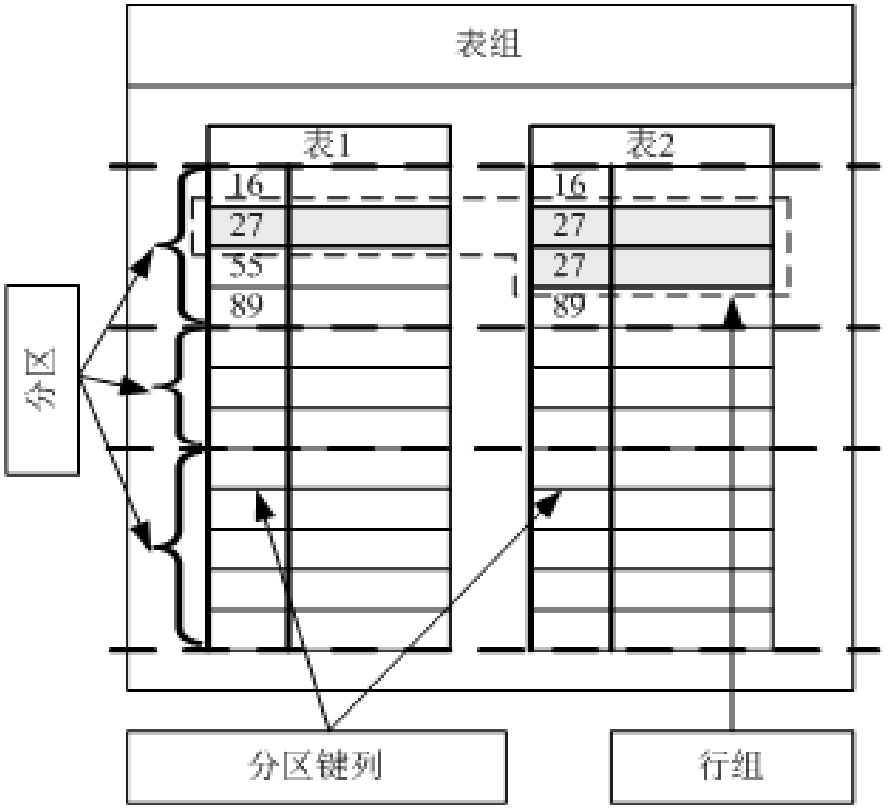
关系模型

■表：一个表是一个逻辑关系，它包含一个分区键，用来对表进行分区。

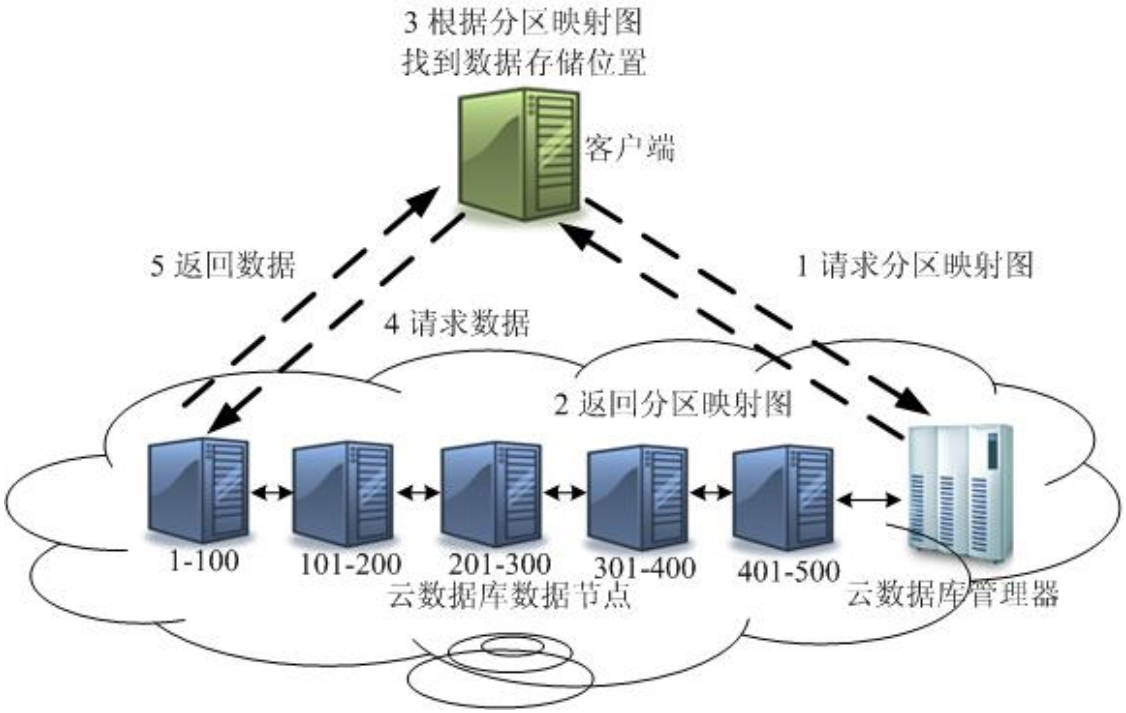
■表组：具有相同分区键的多个表的集合，称为表组。

■行组：在表组中，具有相同分区键值的多个行的集合，称为行组。一个行组中包含的行，总是被分配到同一个数据节点上。每个表组会包含多个行组，这些行组会被分配到不同的数据节点上。

■数据分区：一个数据分区包含了多个行组。因此，每个数据节点都存储了位于某个分区键值区间内的所有行。



云数据库领域的研究问题-体系架构



数据访问方法

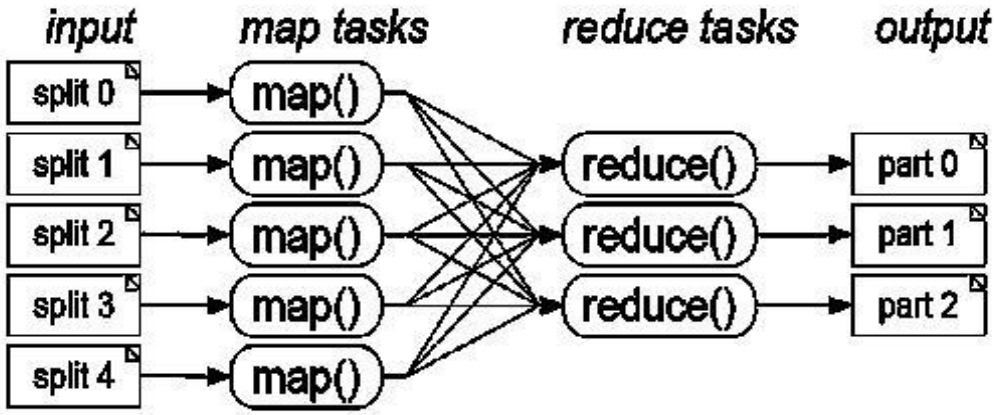
- 1、客户端首先向管理器请求一份分区映射图
- 2、管理器向客户端发送分区映射图
- 3、客户端在映射图中根据键值找到所需数据的存储位置
- 4、客户端到指定的数据节点请求数据
- 5、由该数据节点把数据返回给客户端

实际上，为了改进性能，同时也为了避免管理器的性能瓶颈，通常会在客户端缓存常用的分区映射图，这样，客户端在很多情况下不用与管理器交互就可以直接访问相应的数据节点。

云数据库领域的研究问题-编程模型

MapReduce

- Map/Reduce计算流程



表一 Map 和 Reduce 函数

函数	输入	输出	说明
Map	<k1, v1>	List(<k2,v2>)	1. 将小数据集进一步解析成一批 <key,value> 对，输入 Map 函数中进行处理。 2. 每一个输入的 <k1,v1> 会输出一批 <k2,v2>。 <k2,v2> 是计算的中间结果。
Reduce	<k2,List(v2)>	<k3,v3>	输入的中间结果 <k2,List(v2)> 中的 List(v2) 表示是一批属于同一个 k2 的 value

云数据库领域的研究问题-编程模型

MapReduce 在MapReduce环境下执行两个关系的联接操作

- 假设关系R(A,B)和S(B,C)都存储在一个文件中。
- 为了联接这些关系，必须把来自每个关系的各个元组都和一个key关联，这个key就是属性B的值。
- 可以使用一个Map进程集合，把来自R的每个元组(a,b)转换成一个key-value对，其中的key就是b，值就是(a,R)。注意，这里把关系R包含到value中，这样做使得我们可以在Reduce阶段，只把那些来自R的元组和来自S的元组进行匹配。
- 类似地，可以使用一个Map进程集合，把来自S的每个元组(b,c)，转换成一个key-value对，key是b，value是(c,S)。这里把关系名字包含在属性值中，可以使得在Reduce阶段只把那些来自不同关系的元组进行合并。
- Reduce进程的任务就是，把来自关系R和S的具有共同属性B值的元组进行合并。这样，所有具有特定B值的元组必须被发送到同一个Reduce进程。
- 假设使用k个Reduce进程。这里选择一个哈希函数h，它可以把属性B的值映射到k个哈希桶，每个哈希值对应一个Reduce进程。每个Map进程把key是b的key-value对，都发送到与哈希值h(b)对应的Reduce进程。Reduce进程把联接后的元组(a,b,c)，写到一个单独的输出文件中。

主讲教师和助教

主讲教师：林子雨

单位：厦门大学计算机科学系
E-mail: ziyulin@xmu.edu.cn
个人网页：http://www.cs.xmu.edu.cn/linziyu
数据库实验室网站：http://dblab.xmu.edu.cn

助教：赖明星

单位：厦门大学计算机科学系数据库实验室2011级硕士研究生（导师：林子雨）
E-mail: mingxinglai@gmail.com
个人主页：http://mingxinglai.com

欢迎访问《大数据技术基础》2013班级网站：http://dblab.xmu.edu.cn/node/423
本讲义PPT存在配套教材《大数据技术基础》，请到上面网站下载。

Department of Computer Science, Xiamen University, Sep, 2013