



# 大数据技术公开课

## 《大数据概念、技术与应用》

2015年10月13日 山东大学

## 第6讲 云数据库



山东大学公开课主页

林子雨 博士/助理教授

厦门大学计算机科学系

厦门大学云计算与大数据研究中心

E-mail: [ziyulin@xmu.edu.cn](mailto:ziyulin@xmu.edu.cn) ▶▶

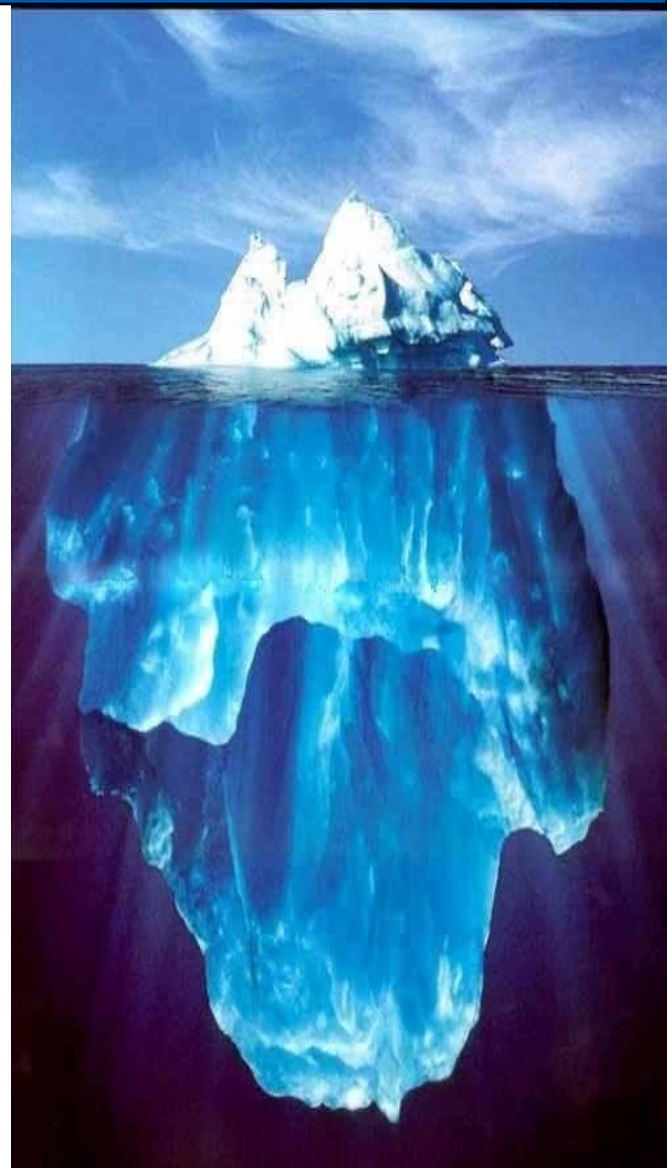
主页: <http://www.cs.xmu.edu.cn/linziyu>





# 提纲

- 6.1 云数据库概述大数据的产生和应用
- 6.2 云数据库产品
- 6.3 云数据库系统架构





# 6.1 云数据库概述

- 6.1.1 云计算是云数据库兴起的基础
- 6.1.2 云数据库概念
- 6.1.3 云数据库的特性
- 6.1.4 云数据库是个性化数据存储需求的理想选择
- 6.1.5 云数据库与其他数据库的关系



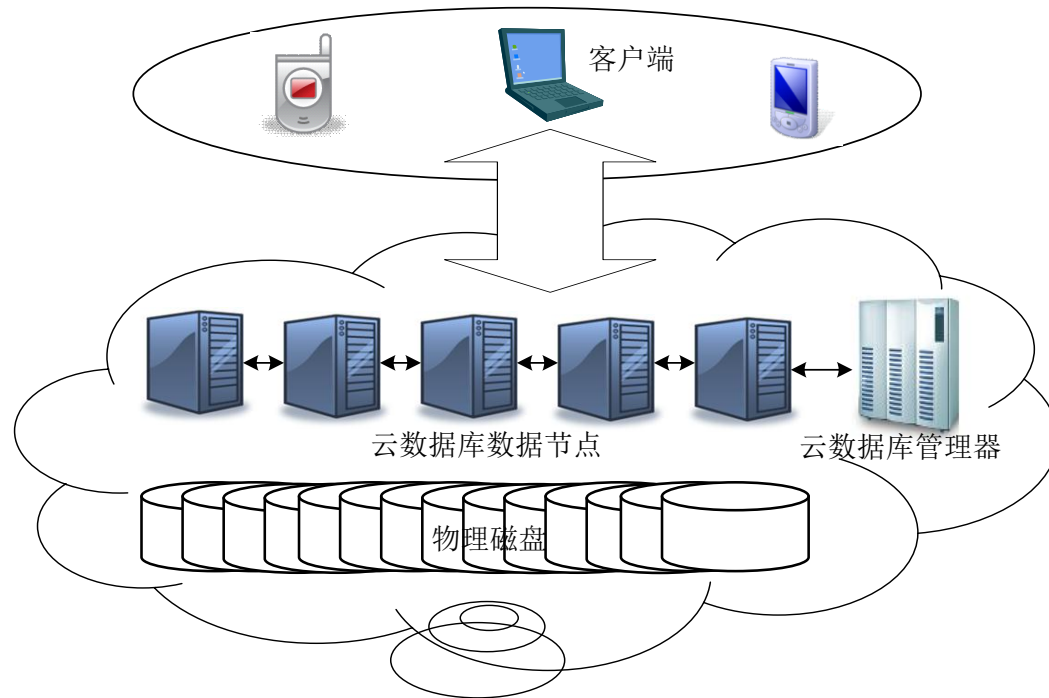
# 6.1.1 云计算是云数据库兴起的基础

表6-1 传统的软件使用方式和云计算方式的比较

	传统方式	云计算方式
获得软件的方式	自己投资建设机房，搭建硬件平台，购买软件在本地安装	直接购买云计算厂商的软件服务
使用方式	本地安装，本地使用	软件运行在云计算厂商服务器上，用户在任何有网络接入的地方都可以通过网络使用软件服务
付费方式	需要一次性支付较大的初期投入成本，包括建设机房、配置硬件、购买各种软件（操作系统、杀毒、业务软件等）	零成本投入就可以立即获得所需的IT资源，只需要为所使用的资源付费，多用多付，少用少付，极其廉价
维护成本	需要自己花钱聘请专业技术人员维护	零成本，所有维护工作由云计算厂商负责
获得IT资源速度	需要耗费较长时间建设机房、购买和安装调试设备系统	随时可用，购买服务后立即可用
共享方式	自己建设，自给自足	云计算厂商建设好云计算服务平台后，同时为众多用户提供服务
维修速度	出现病毒、系统崩溃等问题时，需要自己聘请IT人员维护，很多普通企业的IT人员技术能力有限，碰到一些问题甚至需要寻找外援，通常不能立即解决	出现任何系统问题时，云计算厂商都会凭借其专业化团队给出及时响应，确保云服务的正常使用
资源利用率	利用率较低，投入大量资金建设的IT系统，往往只供企业自己使用，当企业不需要那么多IT资源时，就会产生资源浪费	利用率较高，每天都可以为大量用户提供服务；当存在闲置资源时，云计算管理系统会自动关闭和退出多余资源；当需要增加资源时，又会自动启动和加入相关资源
用户搬迁时的成本	当企业搬家时，原来的机房设施就要作废，需要在新地方重新投入较大成本建设机房	企业无论搬迁到哪里，都可以通过网络重新零成本立即获得云计算服务，因为，资源在云端，不在用户端，用户搬迁不会影响到IT资源的分布
资源可拓展性	企业自己建设的IT基础设施的服务能力通常是有上限的，当企业业务量突然增加时，现有的IT基础设施无法立即满足需求，就需要花费时间和金钱购买和安装新设备；当业务高峰过去时，多余的设备就会闲置，造成资源浪费	云计算厂商可以为企业提供近乎无限的IT资源（存储和计算等资源），用户想用多少都可以立即获得，当用户不使用时，只需退订多余资源，不存在任何资源闲置问题



## 6.1.2 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。云数据库是在云计算的大背景下发展起来的一种新兴的共享基础架构的方法，它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，让软、硬件升级变得更加容易，同时，也虚拟化了许多后端功能。云数据库具有高可扩展性、高可用性、采用多租形式和支持资源有效分发等特点。



## 6.1.3 云数据库的特性

云数据库具有以下特性：

- (1) 动态可扩展
- (2) 高可用性
- (3) 较低的使用代价
- (4) 易用性
- (5) 高性能
- (6) 免维护
- (7) 安全

表6-2 腾讯云数据库和自建数据库的比较

	自建数据库	腾讯云数据库
数据安全性	开发者自行解决，成本高昂	15种类型备份数据，保证数据安全
服务可用性		99.99%高可靠性
数据备份		0花费，系统自动多时间点数据备份
维护成本		0成本，专业团队7x24小时帮助维护
实例扩容		一键式直接扩容，安全可靠
资源利用率		按需申请，资源利用率高达99.9%
技术支持		专业团队一对一指导、QQ远程协助开发者



## 6.1.4 云数据库是个性化数据存储需求的理想选择

企业类型不同，对于存储的需求也千差万别，而云数据库可以很好地满足不同企业的个性化存储需求：

- 首先，云数据库可以满足大企业的海量数据存储需求
- 其次，云数据库可以满足中小企业的低成本数据存储需求
- 另外，云数据库可以满足企业动态变化的数据存储需求

到底选择自建数据库还是选择云数据库，取决于企业自身的具体需求

- 对于一些大型企业，目前通常采用自建数据库
- 对于一些财力有限的中小企业而言，IT预算比较有限，云数据库这种前期零投入、后期免维护的数据库服务，可以很好满足它们的需求





## 6.1.5云数据库与其他数据库的关系

- 从数据模型的角度来说，云数据库并非一种全新的数据库技术，而只是以服务的方式提供数据库功能
- 云数据库并没有专属于自己的数据模型，云数据库所采用的数据模型可以是关系数据库所使用的关系模型（微软的**SQL Azure**云数据库、阿里云**RDS**都采用了关系模型），也可以是**NoSQL**数据库所使用的非关系模型（**Amazon Dynamo**云数据库采用的是“键/值”存储）
- 同一个公司也可能提供采用不同数据模型的多种云数据库服务
- 许多公司在开发云数据库时，后端数据库都是直接使用现有的各种关系数据库或**NoSQL**数据库产品





## 6.2 云数据库产品

- 6.2.1 云数据库厂商概述
- 6.2.2 Amazon的云数据库产品
- 6.2.3 Google的云数据库产品
- 6.2.4 Microsoft的云数据库产品
- 6.2.5 其他云数据库产品



## 6.2.1 云数据库厂商概述

表6-3 云数据库产品

企业	产品
Amazon	Dynamo、SimpleDB、RDS
Google	Google Cloud SQL
Microsoft	Microsoft SQL Azure
Oracle	Oracle Cloud
Yahoo!	PNUTS
Vertica	Analytic Database v3.0 for the Cloud
EnterpriseDB	Postgres Plus in the Cloud
阿里	阿里云RDS
百度	百度云数据库
腾讯	腾讯云数据库



## 6.2.2 Amazon的云数据库产品

- Amazon是云数据库市场的先行者。Amazon除了提供著名的S3存储服务和EC2计算服务以外，还提供基于云的数据库服务SimpleDB和Dynamo
- SimpleDB是Amazon公司开发的一个可供查询的分布式数据存储系统，AWS（Amazon Web Service）上的第一个NoSQL数据库服务，集合了Amazon的大量AWS基础设施
- SimpleDB采用了“键/值”存储，它主要是服务于那些不需要关系数据库的Web开发者
- SimpleDB存在一些明显缺陷，比如存在单表限制、性能不稳定、只能支持最终一致性等
- Dynamo吸收了SimpleDB以及其他NoSQL数据库设计思想的精华，旨在为要求更高的应用设计，这些应用要求可扩展的数据存储以及更高级的数据管理功能
- DynamoDB使用固态硬盘，实现恒定、低延迟的读写时间，旨在扩展大容量同时维持一致的性能
- Amazon RDS（Amazon Relational Database Service）是Amazon开发的一种Web服务，它可以让用户在云环境中建立、操作关系型数据库（可以支持MySQL和Oracle等数据库）



## 6.2.3 Google的云数据库产品

- Google Cloud SQL是谷歌公司推出的基于MySQL的云数据库
- 使用Cloud SQL，所有的事务都在云中，并由谷歌管理，用户不需要配置或者排查错误
- 由于数据在谷歌多个数据中心中复制，因此它永远是可用的
- 谷歌还提供导入或导出服务，方便用户将数据库带进或带出云
- 谷歌使用用户非常熟悉的MySQL，带有JDBC支持（适用于基于Java的App Engine应用）和DB-API支持（适用于基于Python的App Engine应用）的传统MySQL数据库环境，因此，多数应用程序不需过多调试即可运行，数据格式对于大多数开发者和管理员来说也是非常熟悉的
- Google Cloud SQL还有一个好处就是与Google App Engine集成



## 6.2.4 Microsoft的云数据库产品

SQL Azure具有以下特性：

- 属于关系型数据库：支持使用TSQL（Transact Structured Query Language）来管理、创建和操作云数据库
- 支持存储过程：它的数据类型、存储过程和传统的SQL Server具有很大的相似性，因此，应用可以在本地进行开发，然后部署到云平台上
- 支持大量数据类型：包含了几乎所有典型的SQL Server 2008的数据类型
- 支持云中的事务：支持局部事务，但是不支持分布式事务

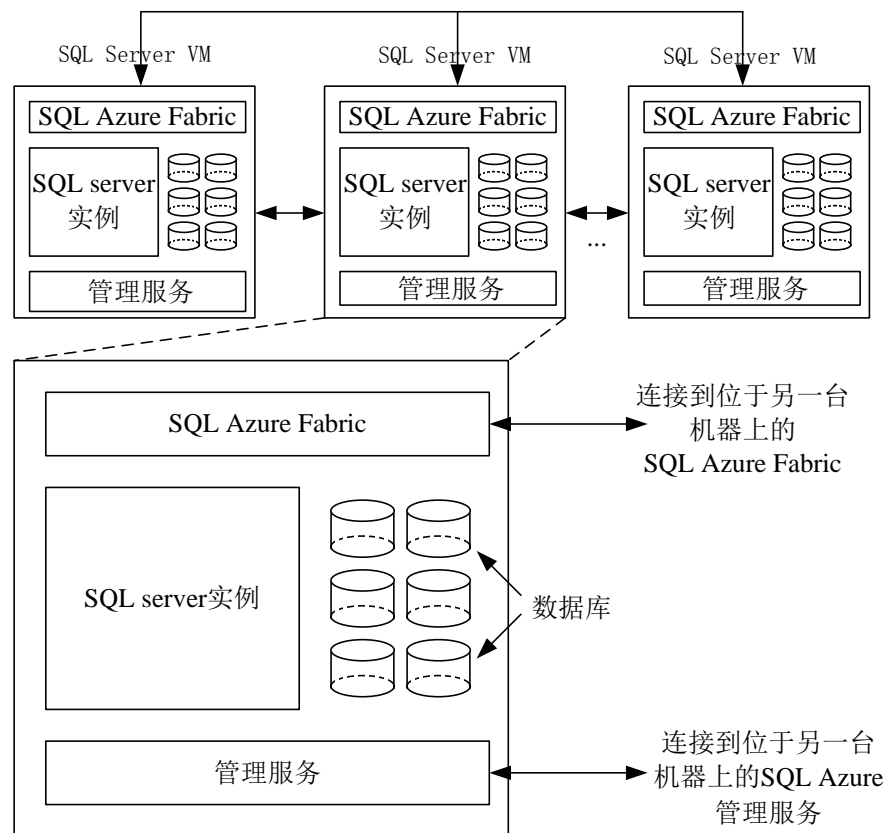


图6-2 SQL Azure的体系架构



# 本讲小结

- 介绍了云数据库的相关知识。云数据库是在云计算兴起的大背景下发展起来的，在云端为用户提供数据服务，用户不需要自己投资建设软硬件环境，只需要向云数据库服务供应商购买数据库服务，就可以方便、快捷、低成本地实现数据存储和管理功能
- 云数据库具有动态可扩展、高可用性、低成本、易用性、大规模并行处理等突出优点，是大数据时代企业实现低成本的大规模数据存储的理想选择
- 云数据库市场有很多代表性的产品可供选择。**Amazon**是云数据库市场的先行者，谷歌和微软公司都开发了自己的云数据库产品，都在市场上形成了自己的影响力



# 主讲教师



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>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革委员会副局长。中国高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度厦门大学奖教金获得者。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，编著出版中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》并成为畅销书籍；主讲厦门大学计算机系本科生课程《数据库系统原理》和研究生课程《分布式数据库》《大数据技术基础》。具有丰富的政府和企业信息化培训经验，曾先后给中国移动通信集团公司、福州马尾区政府、福建省物联网科学研究院、石狮市物流协会、厦门市物流协会等多家单位和个人开展信息化培训，累计培训人数达2000人以上。





# 大数据学习教材推荐



扫一扫访问教材官网

《大数据技术原理与应用——概念、存储、处理、分析与应用》，由厦门大学计算机科学系林子雨博士编著，是中国高校第一本系统介绍大数据知识的专业教材。

全书共有13章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：  
<http://dblab.xmu.edu.cn/post/bigdata>



Principles and Applications of Big Data Technology - Big Data Conception, Storage, Processing, Analysis and Application

林子雨 编著



The background of the slide features several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. On the left side, two people are shown in profile, facing each other. The overall scene suggests a group of people in a meeting or a social gathering.

**Thank You!**

Department of Computer Science, Xiamen University, October, 2015