第11章 数据安全

- 数据安全概述
- 数据备份与恢复
- 云计算技术
- 云数据安全
- 本章小结

数据安全概念

- 数据安全通常有两方面的含义：
  - 数据本身的安全：指采用现代密码算法对数据进行主动保护。
  - 数据的防护安全：采用现代信息存储手段对数据进行主动防护，如通过磁盘阵列、数据备份和异地容灾等手段保证数据的安全。

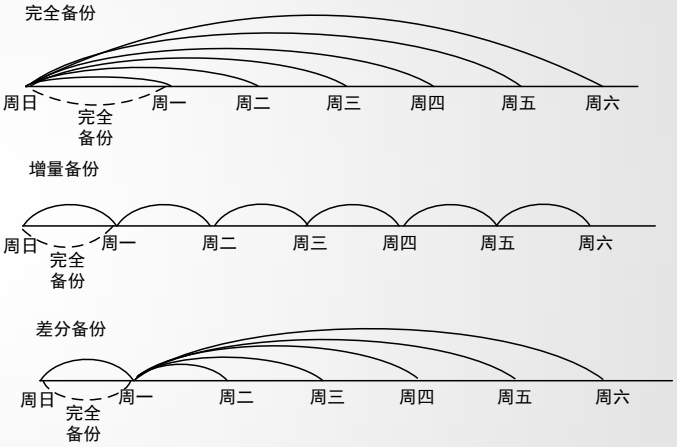
数据备份与恢复

- 数据备份(Data Backup)：将数据以某种方式加以保留，以便在系统遭受破坏或其他特定情况下，重新加以恢复的一个过程。
- 数据备份不是数据复制或数据拷贝
  - 为降低备份数据所占用的额外空间，需要改变数据格式、进行压缩等操作。
- 数据备份是为达到数据恢复和重建的目标所进行的一系列备份步骤和行为。
  - 在灾难发生前，通过对主系统进行备份并加强管理，保证其完整性和可用性。

数据备份的根本目的是重新利用，备份工作的核心是恢复。

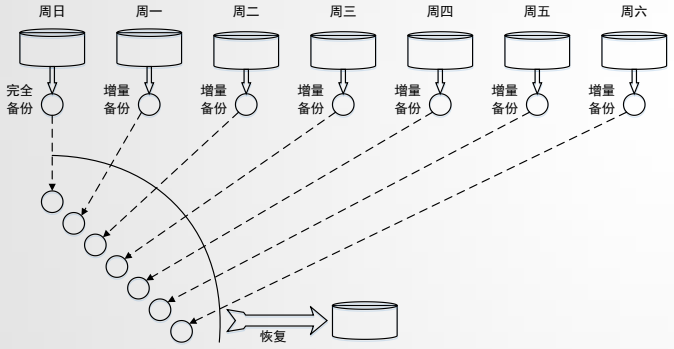
数据备份类型：按照备份策略

- 完全备份：对系统中所有的数据进行备份。
- 增量备份：只对上次备份后产生变化的数据进行备份。
- 差分备份：只对上次进行*完全备份*后产生变化的数据进行备份。

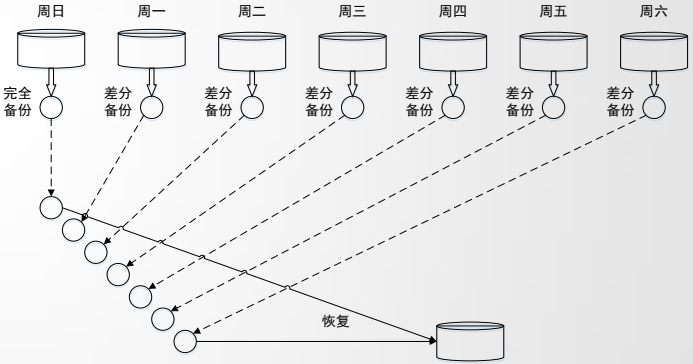


按照备份策略

- 完全备份和增量备份的结合



- 完全备份和差分备份的结合



按照备份状态

- **物理备份：**将实际物理数据从一处复制到另一处的备份，如对数据库的冷备份、热备份。
  - 冷备份（脱机备份）
  - 热备份（联机备份）
- **逻辑备份：**将某个数据库的记录读出，并将其写入到一个文件中。

按照备份层次

- 硬件级备份：通过硬件冗余来实现。
  - 双机容错
  - 磁盘双工
  - 磁盘阵列RAID
  - 磁盘镜像
- 软件级备份：与硬件容错相结合解决软件故障或人为误操作造成数据丢失。

理想的备份系统=硬件冗余+软件备份

按照备份地点

- 本地备份
  - 在本地硬盘的指定区域备份文件
  - 用于系统或人为误操作造成的数据损坏和丢失
- 异地备份
  - 备份的数据放在异地
  - 用于地域性灾难造成的数据问题

数据容灾

- 数据备份是数据容灾的基础
  - 数据备份的目的是为了系统数据崩溃时能够快速的恢复数据。
  - 数据备份只是数据容灾方案中的一种，而且它的容灾能力有限。
- 真正的数据容灾就是要避免传统冷备份所具有的先天不足，它能在灾难发生时，全面、及时地恢复整个系统
- 国际标准SHARE 78定义的容灾系统有7个层次
  - 从最简单的仅在本地进行磁盘备份，到将备份的磁盘存储在异地，再到建立应用系统实时切换的异地备份系统。

数据容灾技术

- 远程镜像技术
- 快照技术
- 互联技术
- 虚拟存储
- …

远程镜像技术

- 远程镜像又称远程复制：利用物理位置上分离的存储设备所具备的远程数据连接功能，在远程维护一套数据镜像，一旦灾难发生时，分布在异地存储器上的数据备份并不会受到波及。
  - 同步远程镜像：将本地数据以完全同步的方式复制到异地，对系统性能影响较大。
  - 异步远程镜像：以后台同步的方式进行，本地系统性能影响较小，需要解决数据的一致性问题。

快照技术

- SNIA（存储网络行业协会）对快照（Snapshot）的定义
  - 数据集合的一个完全可用拷贝，该拷贝包括相应数据在某个时间点（拷贝开始的时间点）的映像。
  - 快照可以是其所表示的数据的一个副本，也可以是数据的一个复制品。
- 快照功能
  - 能够进行在线数据恢复，当存储设备发生应用故障或者文件损坏时可以进行及时数据恢复，将数据恢复成快照产生时间点的状态。
  - 为存储用户提供了另外一个数据访问通道，当原数据进行在线应用处理时，用户可以访问快照数据，还可以利用快照进行测试等工作。

互联技术

- 基于SAN的远程复制
  - 通过光通道FC，把两个SAN链接起来，进行远程镜像。
  - 当灾难发生时，由备援数据中心替代主数据中心保证系统工作的连续性。
- 缺点
  - 实现成本高
  - 设备的互操作性差
  - 跨越的地理位置短
  - …

虚拟存储

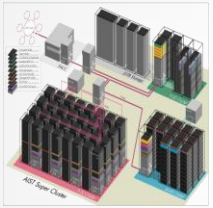
- 在物理存储系统和服务器之间增加一个虚拟层，它管理和控制所有存储并对服务器提供存储服务。
- 服务器不直接与存储硬件打交道，存储硬件的增减、调换、分拆、合并对服务器层完全透明。
  - 隐藏了复杂程度
  - 允许将现有的功能集成使用
  - 摆脱了物理容量的局限

云计算技术

- 云计算概述
- 云计算体系架构
- 云数据存储技术
- 云数据管理技术

云计算概述

并行计算



集群计算

网格计算

云计算

云计算(Cloud Computing)是继分布式计算(Distributed Computing)、并行计算(Parallel Computing)和网格计算(Grid Computing)之后的一种新的商业计算模式

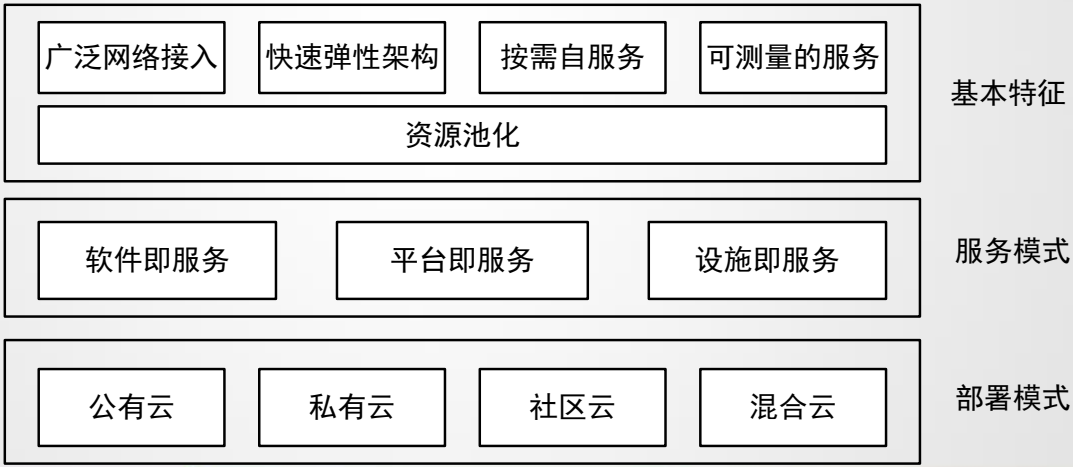
什么是云计算

- IBM技术白皮书
  - 云计算一词描述了一个系统平台或一类应用程序；该平台可以根据用户的需求动态部署、配置、重新配置以及取消服务等；云计算是一种可以通过互联网进行访问的可扩展的应用程序。
- Berkeley白皮书
  - 云计算包括互联网上各种服务形式的应用以及数据中心中提供这些服务的软硬件设施。互联网上的应用服务一直被称作软件即服务（Software as a Service, SaaS），而数据中心的软硬件设施就是云。
- ISO/IEC17788《云计算词汇与概述》DIS版
  - 云计算是一种将可伸缩、弹性、共享的物理和虚拟资源池以按需自服务的方式供应和管理，并提供网络访问的模式。云计算模式由关键特征、云计算角色和活动、云能力类型和云服务分类、云部署模型、云计算共同关注点组成。

什么是云计算

- 美国标准计算研究院NIST

  - 云计算是一种计算模式，它以一种便捷的、通过网络按需接入到一组已经配好的计算资源池，如网络、服务器、存储、应用程序和服务等。在这种模式中，计算资源将以最小的管理和交互代价快速提供给用户。



云计算特征

- 广泛网络接入：用户可从任何网络覆盖的地方，使用各种终端设备随时随地的通过互联网访问云计算服务。
- 快速弹性架构：服务的规模可快速伸缩，以自动适应业务负载的动态变化。用户使用的资源同业务的需求相一致，避免了因服务器性能过载或冗余而导致服务质量下降或资源浪费。
- 资源池化：资源以共享资源池的方式统一管理。利用虚拟化技术，将资源分享给不同用户，资源的放置、管理和分配策略对用户透明。
- 按需自服务：以服务的形式为用户提供应用程序、数据存储、基础设施等资源，并可根据用户需求，自动分配资源，而不需要系统管理员的干预。
- 可测量的服务：通过监控用户的资源使用量，并根据资源的使用情况对服务计费。

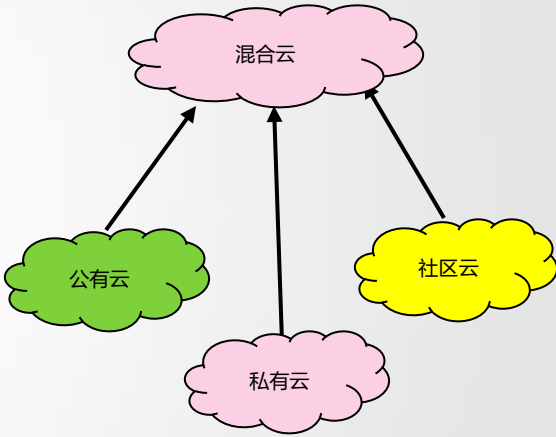
云计算分类

- 按照云计算的服务模式
  - 软件即服务(Software as a Service, SaaS)：向用户提供使用运行在云基础设施上的某些应用软件的能力。典型的应用有：Salesforce的客户关系管理系统CRM，Google的在线办公自动化软件等。
  - 平台即服务(Platform as a Service, PaaS)：为用户提供在云基础设施上部署定制应用的系统软件平台。典型的代表有：Google App Engine、Microsoft Azure等。
  - 基础设施即服务(Infrastructure as a Service, IaaS)：通过虚拟化技术来组织底层网络连接、服务器等物理设备，为用户提供资源租用与管理服务。典型的代表有：Amazon的Web服务，包括弹性计算云EC2、简单存储服务S3和结构化数据存储服务SimpleDB，IBM公司的蓝云Blue Cloud、Sun的云基础设施平台IAAS等

云计算分类

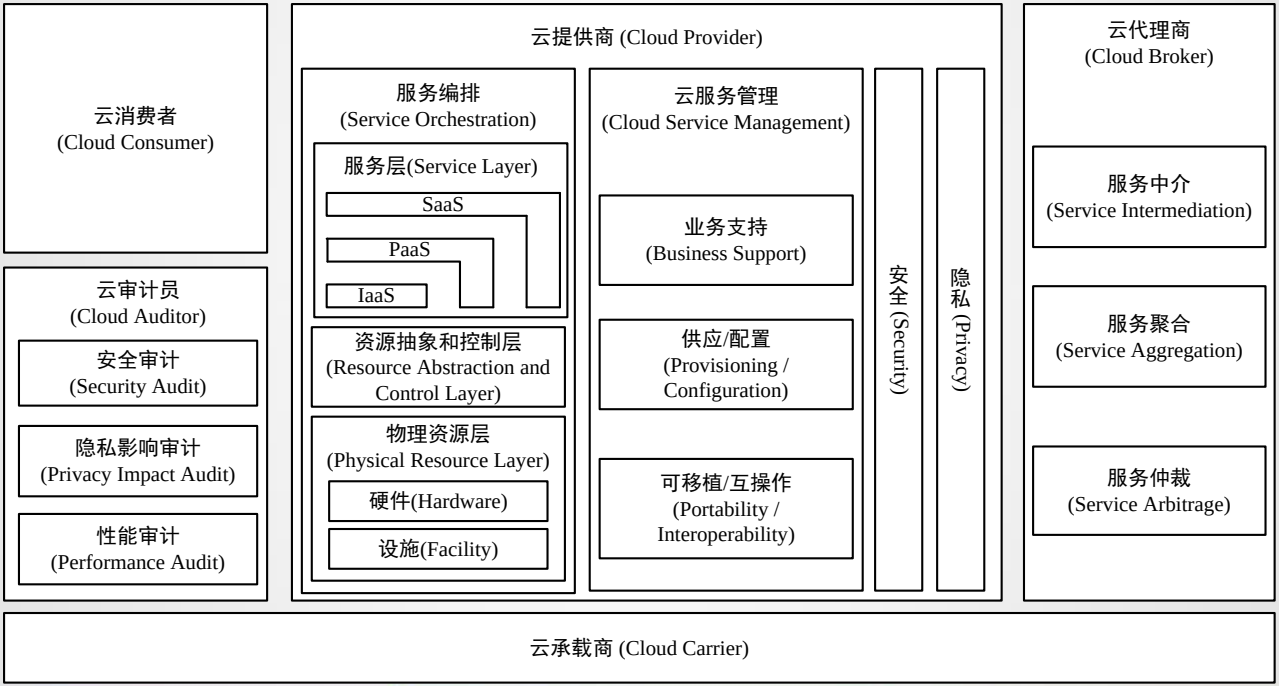
- 按照云计算的部署模式

  - 公有云(Public Cloud)：由某个组织拥有，其云基础设施向普通用户、公司或各类组织提供云服务。
  - 私有云(Private Cloud)：云基础设施特定为某个组织运行服务，可以是该组织或某个第三方负责管理，可以是场内服务(on-premises)，也可以是场外服务(off-premises)。
  - 社区云(Community Cloud)：云基础设施由若干个组织分享，以支持某个特定的社区。
  - 混合云(Hybird Cloud)：云基础设施由两个或多个云（私有云、社区云或公有云）组成，独立存在，但是通过标准的或私有的技术绑定在一起，这些技术可促成数据和应用的可移植性，如用于云之间负载分担的cloud bursting技术。



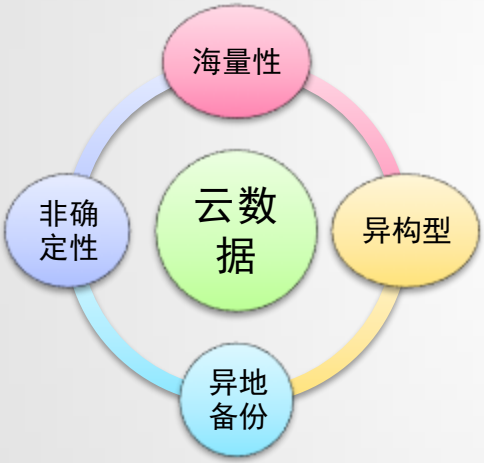
云计算体系架构

- NIST的云计算参考架构



云数据存储技术

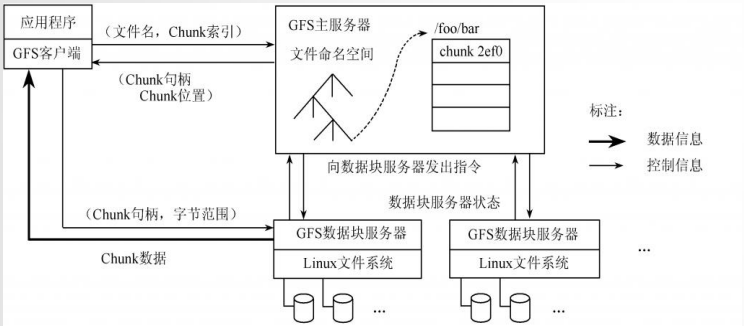
- 云中数据特点



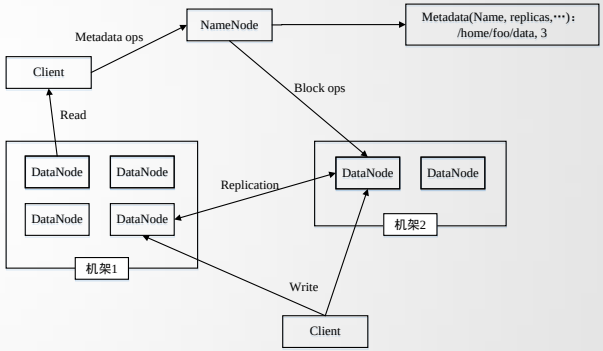
- 云计算系统中常用的数据文件存储系统
  - Google的GFS(Google File System, GFS)
  - Hadoop开发的GFS的开源实现HDFS(Hadoop Distributed File System, HDFS)
- 云计算系统中常用的数据管理技术
  - Google的BigTable数据管理技术
  - Hadoop开发的开源数据管理模块HBase

云数据文件存储系统

- GFS

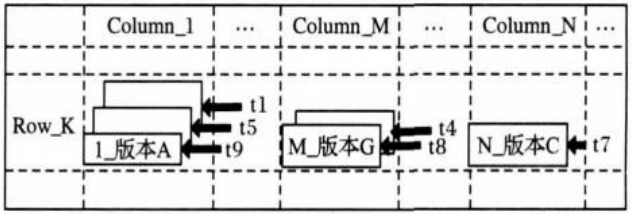


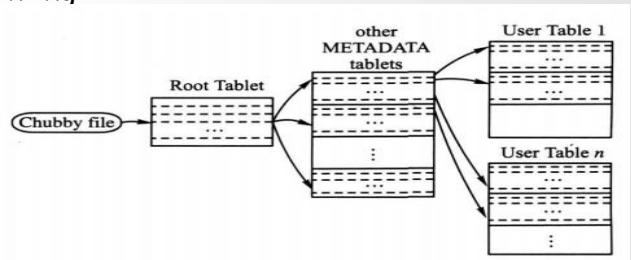
- HDFS



云数据管理技术

- BigTable
  - Bigtable是一个键值（key-value）映射。
  - Bigtable的键有三维，分别是行键（row key）、列键（column key）和时间戳（timestamp），行键和列键都是字节串，时间戳是64位整型；而值是一个字节串。
  - (row: string, column: string, time: int64)→string





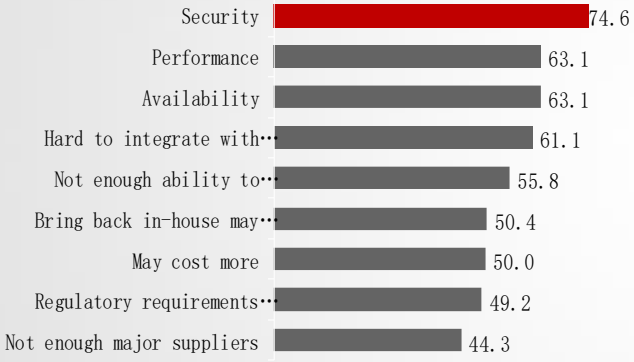
内容提纲

- 数据安全概述
- 数据备份与恢复
- 云计算技术
- 云数据安全
- 本章小结

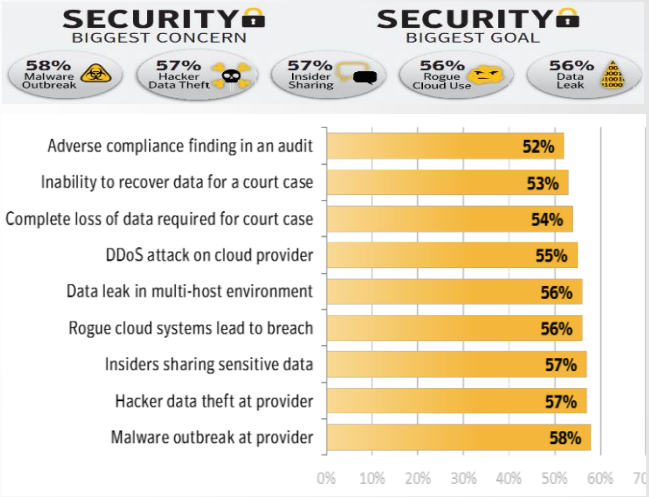
云数据安全

云计算所面临的挑战中，安全问题排在首位

75%用户在安全性上犹豫不决



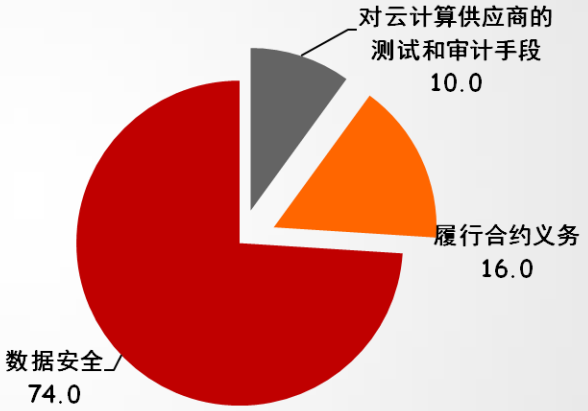
Source：IDC Enterprise Panel（国际数据公司IDC）



Source：Symantec 2011 State of cloud report(global)

云数据安全

- 数据安全
  - 传统的CIA属性。
  - 数据的非授权使用。
  - 数据和知识产权的保护。



Source：Deloitte 德勤 《隐私和数据保护调查》

云数据安全

- 云数据安全需求
- 云数据安全威胁
- 云计算中数据安全技术

云数据安全需求

- 云数据所有权与管理权相分离
  - 用户将不再知道自己的数据存储在哪里、被怎么存储的、谁在处理、有没有备份等信息。
- 云计算的通用性、虚拟性、共享性等特点，导致了传统系统中的隐私保护技术往往无法使用在云数据中。
- 如何在云计算中实现不同属性用户的数据资源安全共享是一大难题。

云数据安全威胁

- 云计算的滥用、恶用、拒绝服务攻击
- 不安全的接口和API
- 恶意的内部员工
- 共享技术产生的问题
- 数据泄漏
- 账号和服务劫持
- 未知的风险场景

云数据安全威胁

- 云计算网络层面的数据安全风险
  - 确保服务提供商传输数据的保密性及完整性（跨越公共网络）
  - 确保服务提供商对所有的资源都提供适当的访问控制，包括审计、认证和授权
  - 确保云计算中的公有云资源具备可用性（防止拒绝服务攻击）
  - 域管理来代替现有的网络层面模型
    - 域具有排他性，只允许特定角色访问该指定区域；
    - 安全域是逻辑上的隔离，主机层面上不一定是物理隔离。

云数据安全威胁

- 云计算主机层面的数据安全威胁
  - SaaS和PaaS的主机安全
  - IaaS的主机安全
- 云计算应用层面的数据安全威胁
  - 网络漏洞攻击快速增长、多种新的网络渗透方法涌现
  - Web浏览器本身的脆弱性，Web应用程序会很容易被植入恶意代码而对用户和服务提供商带来损失

云数据安全技术

- 以数据安全为主要目标的云安全架构：DSLC (Data Security Life Cycle)
  - 获得云中数据的存储、传输、处理的相关信息（监控目的）。
  - 建立数据安全生命周期，包括6个阶段：创建、存储、使用、共享、归档和销毁。
  - 对数据安全生命周期中的每个阶段均明确数据安全保护机制，将行为实施者（可以是用户、用户、系统/进行等）对数据的操作定义为functions，而安全机制则定义为controls，将所有可能的行为限制得到允许的行为范围内。
- DSLC存在的局限性是与云计算的体系结构联系不够紧密，安全机制针对性不强。

云数据安全技术

- 云计算中的身份认证技术
  - 基于PKI（公钥基础设施）的联合身份认证技术。
  - 基于身份的加密技术(IBC, Identity-Based Cryptography)。
    - 不使用证书，用户的公钥直接从用户的身份信息提取。
- 静态存储数据的保护
  - 加密数据的检索（密文检索成为一个研究热点）
  - 基于密码学的访问控制策略
  - 云数据完整性验证
  - 云数据隐私保护

云数据安全技术

■ 动态存储数据的保护

- 隔离机制
  - 沙箱机制对云应用进行隔离，如CyberGuarder（虚拟化安全保护框架，创建一个独立的软件系统的虚拟复制）。
- 访问控制模型和机制（防止用户进行非授权的访问）
  - 强制访问控制（MAC）
  - 自主访问控制（DAC）
  - 基于角色的访问控制（RBAC）
- 基于信息流模型的数据安全保护机制
  - 通过追踪系统中的数据流，防止非授权访问和机密数据泄露。

本章小结

- 掌握数据备份相关概念及实现技术
- 熟悉云计算体系结构及云数据管理方式
- 熟悉云数据面临的安全威胁
- 了解当前云数据安全主要保障措施