信息存储与IT管理

## IT基础设施

### 1.1数据中心

1）概念：用于部署计算机系统及一系列配套设备的设施，配套设备如通信和存储系统。

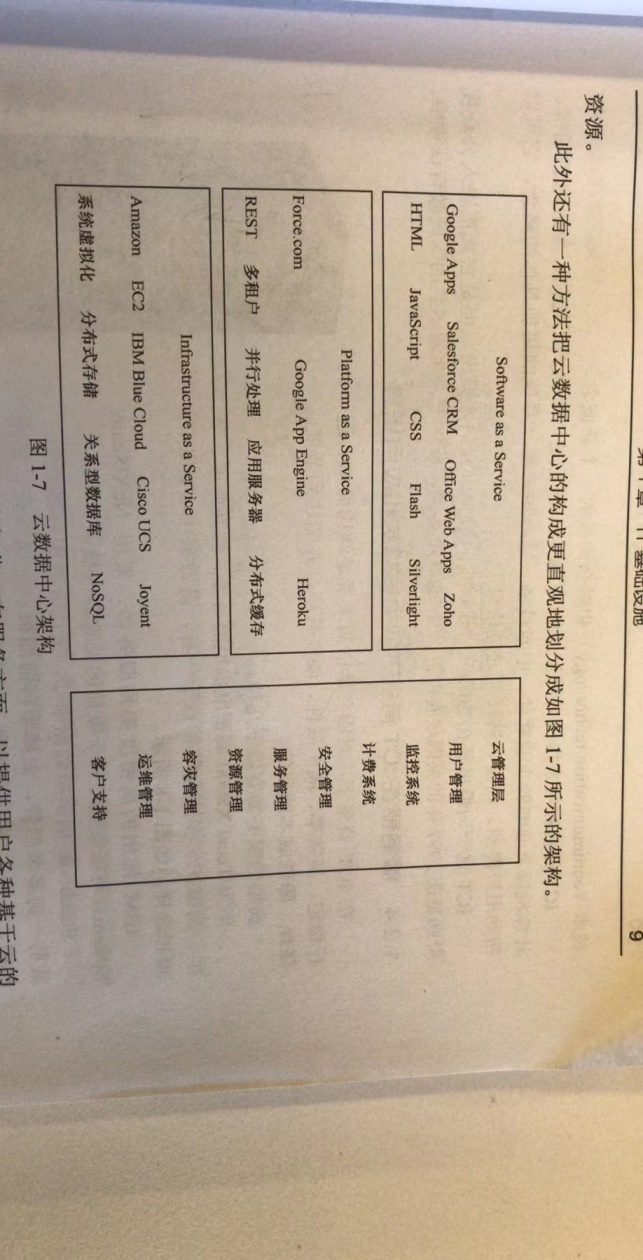
2）存储在数据中心的作用：SAN（storage area network）一种特殊的高速网络：连接网络服务器和大磁盘阵列或备份磁带库的存储设备。

3）服务器在数据中心的作用：作为用户访问数据中心的媒介，提供数据共享的服务。

4）网络在数据中心的作用：网络是信息传输、接收，共享的虚拟平台。

### 1.2云数据中心

概念：基于云计算架构的，计算、存储以及网络资源松耦合，完全虚拟化各种IT设备，模块化程度高，自动化程度高，具备绿色节能程度的新型数据中心。最主要的特点就是高度的虚拟化。



服务部分：

软件即服务（Software as a Service）SaaS:通过互联网提供软件服务的应用模式，不需要用户自己配置程序的运行环境。

平台即服务：（platform as a service）Paas:把相应的服务器平台或者开发环境作为服务提供给用户的一种商业模式

基础设施即服务：（infrastructure as service）IaaS：提供给用户所有设施使用权力的服务。

管理部分：

以云管理层为核心确保云计算数据中心的有效管理和安全稳定运行。

## 存储与应用环境

### 2.1存储的流程

1）在计算机系统中，数据从用户所看到的文件，到最终存储在底层存储设备上的信息，中间经过哪些流程

系统调用接口

文件系统

设备驱动

存储设备

a、应用程序通过系统调用访问数据。

b、文件系统处理数据请求。文件系统负责将文件以某种特定的方式存放在磁盘的数据块上。当他收到操作系统发来的文件读写请求时，便将请求转化为对磁盘上数据块的读写操作。

注意：文件系统下的磁盘指的是逻辑卷，这是一种将物理设备逻辑分区后产生的一块虚拟磁盘，这块磁盘上的地址到实际物理设备上地址的映射，是通过逻辑卷管理器来维护的。因此，对逻辑卷上数据块的操作，在经过逻辑卷管理器处理，转化为真实物理设备上的数据块操作。

c、驱动程序负责从物理设备读写数据。---驱动程序为了兼容不同的设备。

### 2.2网络存储技术

1）概念：基于网络进行存储的技术总称。（network storage technologies）

2）分类：a、直连式存储（direct attached storage,DAS存储设备直接通过总线或电缆连接到服务器。b、存储区域网络（SAN）：通过光纤集线器，光纤路由器，光纤交换机等连接设备将磁盘阵列、磁带等存储设备与相关服务器连接起来的高速专用子网。

c、网络附加存储（network attached storageNAS）：将存储设备通过以太网络拓扑结构连接到服务器上，实现文件级别的共享。

### 2.3存储设备：

最常见的机械硬盘和固态硬盘.

固态硬盘：主控芯片（负责控制数据的读写过程协调）和存储芯片（闪存（Flash芯片）和随机存储器DRAM）组成

### 2.4主机和应用

主机一般是指计算机完成其基本功能所需要的最主要硬件：中央处理器、内存、电源、主板、硬盘和风扇等。同时主机为了实现其功能，还需要一些必不可少的组件：操作系统，设备驱动，文件系统，卷管理器等。这些软硬件共同组成简单的计算机系统。

#### 2.4.1操作系统

概念（OS）：是一种用于管理计算机硬件和软件资源的程序。也是计算机的内核与基础。主要包括：进程与处理机管理、作业管理、存储管理、设备管理、文件管理。

#### 2.4.2设备驱动（device driver）

概念：是一种可以使操作系统和设备通信的特殊程序。提供了硬件与软件通信的接口。

#### 2.4.3卷管理器

概念（logical volume managerLVM）：又称为逻辑卷管理器，它是计算机管理物理存储空间的重要工具。功能是对连接到计算机的物理存储设备进行虚拟化。操作系统对虚拟的空间进行操作，由逻辑卷管理器映射到真正的磁盘上。

#### 2.4.4文件系统

1）概念：是操作系统的子系统，其实是一种软件组件。将磁盘的地址空间结构化，使应用程序能方便的处里数据对象（文件）

2）组成：与文件管理有关的软件，被管理的文件以及实施文件管理所需的数据结构。

3）功能：分配、管理和操作。建立文件，存入、读出、修改、转储文件，控制文件的存取，以及文件的撤销等。

4）分类：linux下的Ext2、Ext3、Ext4、JFS

windows系统下的Fat16、Fat32、exFat和NTFS等。

#### 2.4.5集群

将许多计算设备通过软件或硬件连接起来，组成一个大型计算机网络，并使这些计算设备高度紧密的协作完成计算任务。

### 2.5大数据应用的挑战

#### 2.5.1大数据

1）概念：任何传统的数据处理应用无法在指定时间内完成的指定任务的庞大的，或者复杂的数据集，都可以成为大数据。

2）特点：volume:数据量大：数据达到PB级。

velocity:数据处理速度快。数据实时处理有极高的要求。

variety:数据类型多样。不仅包括结构型数据（可以用关系型数据库存储的数据），还包括图片，视频，音频等非结构化或者半结构化的数据。

value价值密度低。大量数据中筛选出极少数的有价值的数据。

#### 2.5.2大数据应用对于存储的挑战

大数据的核心是提升大量数据的分析能力。

高效性：大量数据的实时访问响应时间，要求存储系统更快的软件，更大的带宽，更高效的存储算法。

可靠性：数据增加---硬盘增加---故障率增加---容错能力提升和更快的数据恢复能力。

## 服务器基础

### 3.1概述：

#### 3.1.1服务器简介

在网络上提供各种服务的高性能计算机，作为网络的重要节点；处理网络上80%的数据和信息。

#### 3.1.2服务器的特点：

1）处理能力强：具有强大的CPU,通常采用多核。向量处理器（数组处理器）是一种实现了直接操作一维数组（向量）指令集的中央处理器（CPU），这与一次只能处理一个数据的标量处理器正好相反。GPU在浮点运算和并行计算等部分计算方面，数十倍乃至于上百倍于CPU的性能。

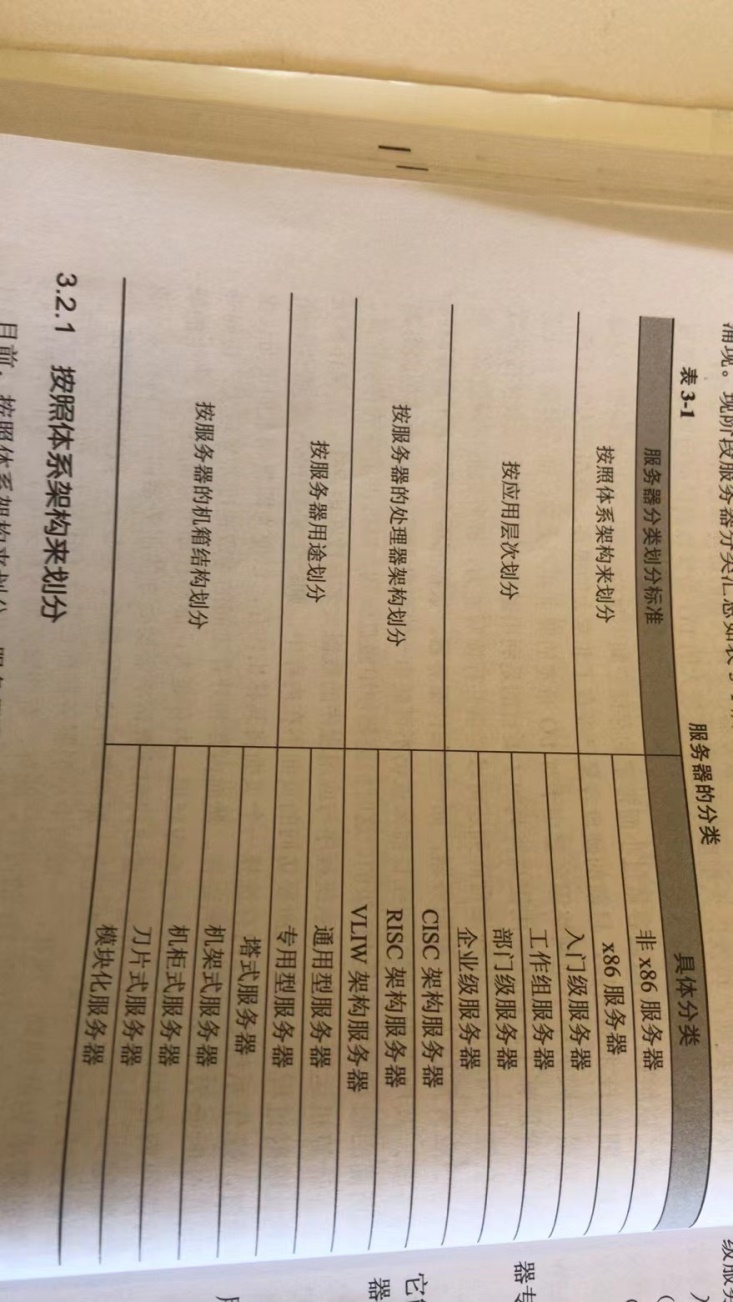
2）主板复杂并且功能强大。

3）硬盘，内存容量大，速度快并且可靠性高。采用磁盘阵列。

4）支持热拔插：热拔插是一项便于服务器部件替换的技术，它支持在服务器运行的过程中，增加或者移除服务器的一个或者多个部件，而不需要服务器重新启动。从而保证服务器的不间断运行。

5）冗余配置：部件进行冗余配置，以保证服务器长时间运行时部件发生故障后服务器能够持续运行。

#### 3.1.3服务器的分类



**按照体系架构划分：**

1. x86服务器（complex instruction set computerCISC）:复杂指令集架构服务器，即通常所讲的PC服务器，基于PC体系结构，使用Intel或其他兼容x86指令集的处理器芯片的服务器。中小企业和非关键业务。
2. 非x86服务器：包括大型机，小型机，和UNIX服务器，他们是使用精简指令集（reduced instruction set computer,RISC）处理器，并且主要采用UNIX和其他专用操作系统的服务器。价格昂贵，体系封闭，稳定性好，性能强大，主要用在金融，电信等大型企业的核心系统或者大中型企业的关键业务中。

**根据服务器在网络中的应用层次划分：**

1. 入门级服务器：设计简单，功能有限，内存小一般在10GB以内，cpu少通常只有一个。低档服务器。
2. 工作组服务器：连接的用户数限制在50台左右。低档服务器。
3. 部门级服务器：中档服务器，通常双CPU，操作系统通常为Linux
4. 企业级服务器：高档服务器，4CPU。金融，证券，交通，邮电，通信及其他大型企业。

**按照服务器的处理架构：**

CISC架构服务器：complex instruction set computer,复杂指令集服务器，一般为windows系统。

RISC架构服务器：reduced instruction set computer精简指令集服务器。中高级服务器

VLIW架构服务器：very long instruction word 超长指令字。在每个时钟周期内，相比于CISC通常只能运行1~3条指令，RISC能运行4条指令，而VLIW架构能运行20条指令。

**按照用途划分：**

专用型服务器：FTP服务器，邮件服务器。

**按照服务器的吉机箱结构进行划分：**

机架式服务器：多为专用型服务器。

#### 3.1.4处理器

常见的处理器：

CISC型CPU：Intel,AMD等厂商。

RISC型CPU：PowerPC处理器，SPARC处理器，PA-RISC处理器MIPS处理器，Alpha处理器

VLIW型CPU：

#### 3.1.5内存

概念：用来存储服务器程序和数据的部件。就是服务器主板上的存储部件，cpu直接与之沟通，存放当前正在使用的数据和程序。

内存技术：ECC（error checking and correcting）指令纠错技术

Chipkill技术，镜像内存技术（mirrored memory）;热拔插RAID内存技术。

#### 3.1.6I/O总线

总线是处理器和外部设备连接通信的通道。

分为内部总线：连接处理器与系统内部芯片。系统总线：连接主板，扩展卡等，外部总线用于连接系统与外部设备。

#### 3.1.7服务器的RAID技术

独立磁盘冗余阵列（redundant array of independent disks,RAID）:将多块独立的硬盘按照不同方式组合起来形成一个硬盘组，从而提供比单个硬盘更高的存储性能和提高数据冗余的技术。

RAID利用条带技术提高磁盘存取速度，同时使用数据冗余技术提供磁盘数据备份。

#### 3.1.8服务器安全保护技术

1）全自动备份技术：两台服务器同样数据，一台发生故障，另一台能够快速切入并恢复数据。

2）事务跟踪技术：

3）自动检验技术：快速检验服务器的故障，并对错误及时修复。

4）内存纠错技术：尽可能早的发现错误并纠正。

5）热定位技术：进行磁盘操作时，将刚写入磁盘的数据快速读取出来，并与缓冲区的原始数据进行比较。如果出现错误，说明磁盘已经损坏，需要将服务器的磁盘坏区情况记录在磁盘中的“热定位重定区”，并将数据重新写入已确定的坏区以外的区域。

6）自动重启技术：发现系统运行中出现的不可由系统恢复的错误，并重新启动。

7）网络监控技术：使用一台普通的客户机就能监控网络上所有使用的服务器。

8）故障在线修复技术：故障部件热拔插和部件在线配置。

#### 3.1.9负载均衡技术

1）分类：a、基于特定服务器软件的负载均衡：修改服务器软件的配置，比如请求重定向。

b、基于DNS的负载均衡：多个IP地址同一个域名，DNS服务器给与访问这个域名的服务器随机分配任意一个IP地址。

c、反向代理负载均衡：普通代理：客户使用代理访问多个外部web服务器，反向代理是多个客户使用它访问内部Web服务器。

d、基于NAT的负载均衡技术：（network address translation）可以实现内部地址和外部地址的转换，使不具有外部网络地址的计算机能够访问外部网络，而当外部网络中的计算机希望访问NAT连接的子网内部的某一台计算机时，NAT技术能通过外部网络中的计算机提的外部地址，将其映射到一个唯一的内部地址上从而实现外部网络与内部网络计算机之间连接。

#### 3.1.10服务器容错技术

服务器集群，双机热备份技术（切换服务器需要中断服务器运行，一般用于对业务连续性要求不高的政府，教育等行业中，但是灵活性高），单机容错技术（电信，金融，证券，医疗等行业）

#### 3.1.11服务器虚拟技术

三种服务器虚拟化技术

1. 硬件虚拟化：Vmware以及Virtual Sever
2. 并行虚拟化：可在一台服务器上运行多个操作系统，与硬件虚拟化类似，但是并行虚拟化可以对操作系统进行更改。
3. 操作系统虚拟化

#### 3.1.12服务器文件系统

1）磁盘文件系统：利用数据存储设备存储文件的文件系统

2）数据库文件系统。

3）网络文件系统。

## RAID技术及应用

### 4.1单个磁盘的不足：

1）磁盘容量有限，无法动态扩容2）不具备容错性，且只能对数据进行逐一读写，无法实现数据的并行读写。

### 4.2概述

主要是指多个独立磁盘通过一定的算法组成一个高可靠性的存储系统。

## 存储阵列技术及应用

## SAN技术及应用

## IP-SAN技术及应用

## NAS技术及应用

## 对象存储

## 存储虚拟化技术及应用

## 备份与恢复

## 容灾及应用

## 大数据存储概论

## 云计算基础

## 数据中心方案

## 16、IT运维