

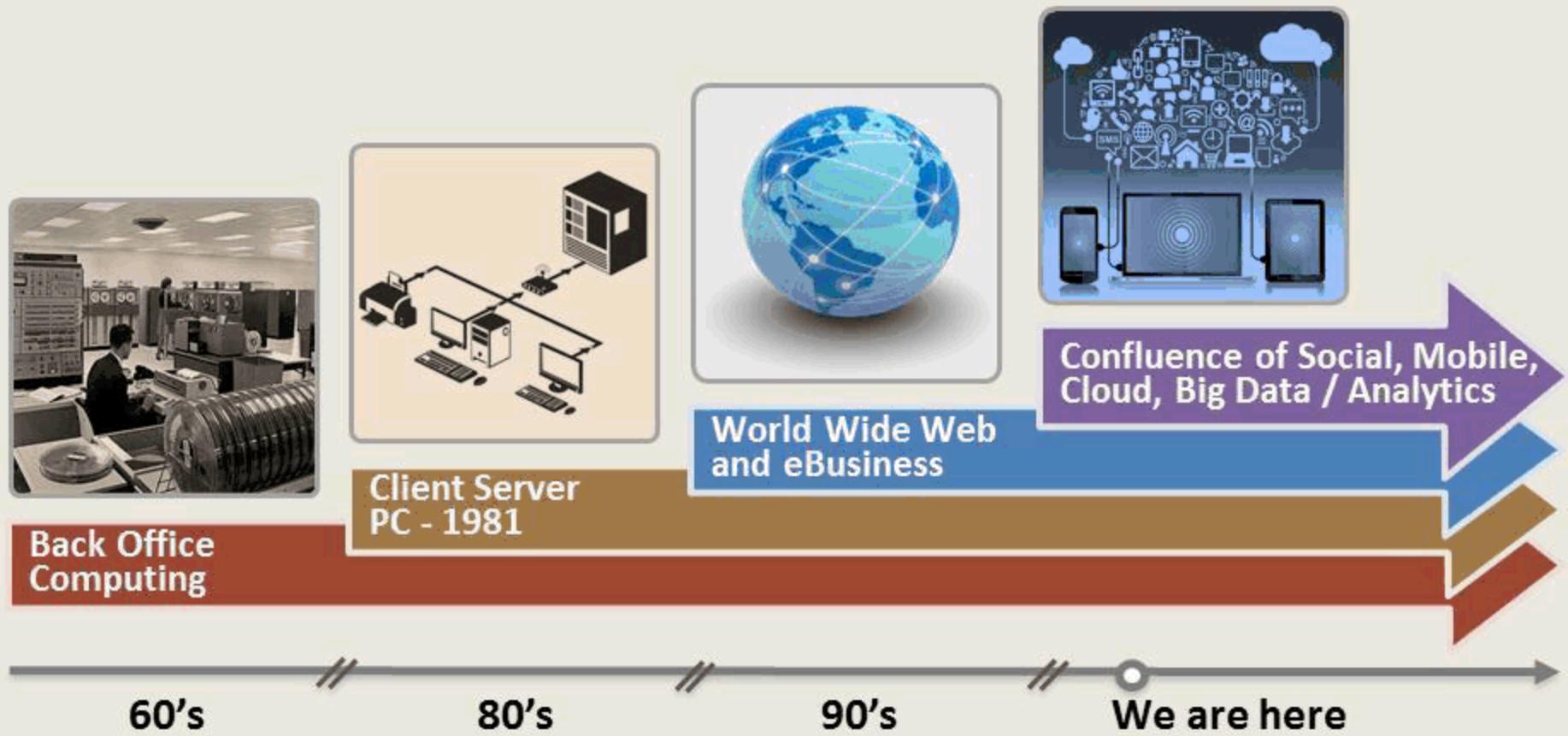
it 基础设施

--回顾

1/4/2019

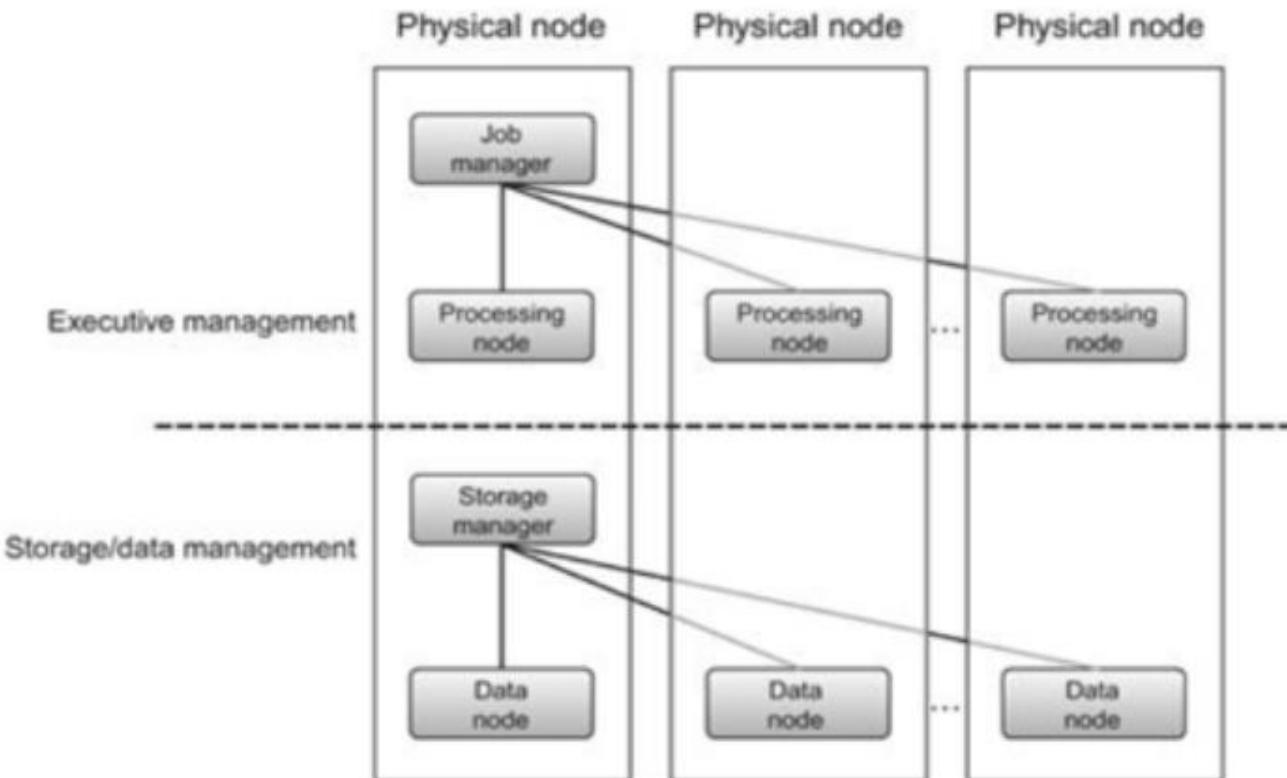
灵宗, 博士
ibm almaden 研究中心
美国加利福尼亚州圣何塞

资讯科技发展



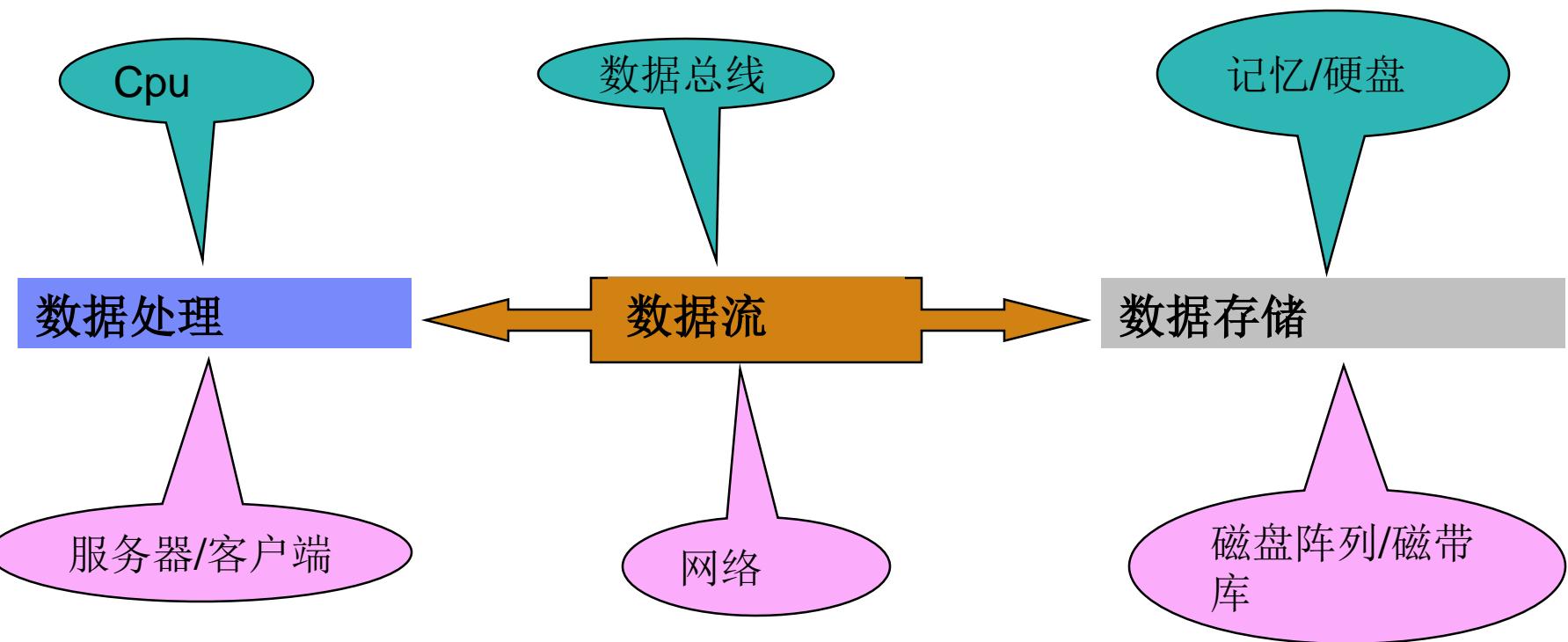
大数据的关键计算资源

- 处理能力: cpu、处理器或节点。
- 记忆
- 网络
- 存储



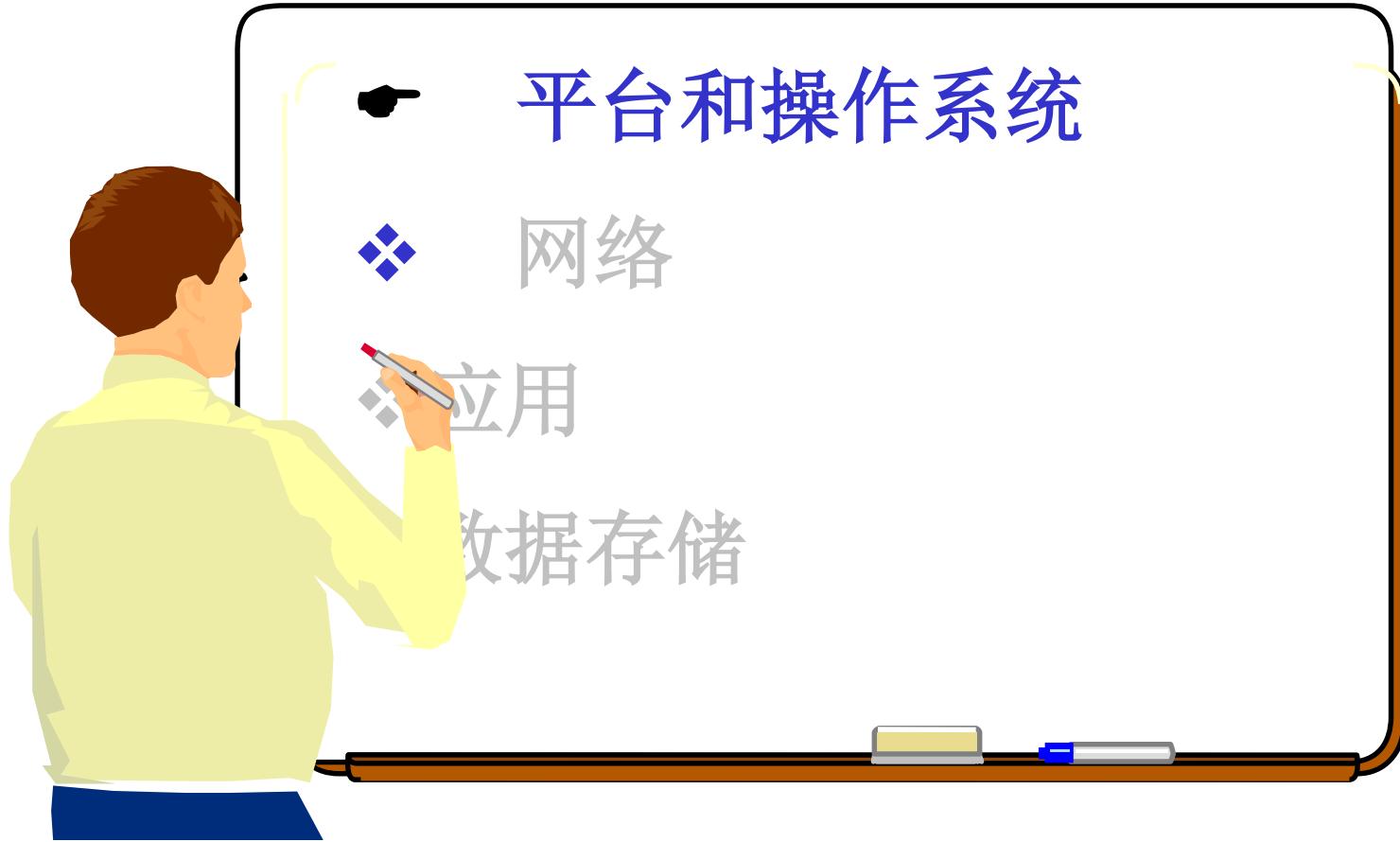
"大数据分析", david loshin, 2013年

it 组件



以数据为中心与以计算为中心

内容



计算平台 (硬件和软件)

- 电脑 (个人电脑)
- 笔记本 电脑
- 服务器
- 中档
- 主机
- 超级计算机

- 计算平台可以被看作是一个框架, 无论是在硬件还是软件, 允许软件运行。
 - 框架是一个定义的支持结构, 在这个结构中可以开发另一个项目
- 典型的平台包括计算机的建筑, 操作系统, 编程语言及相关运行时库或图形用户界面.

硬件平台示例

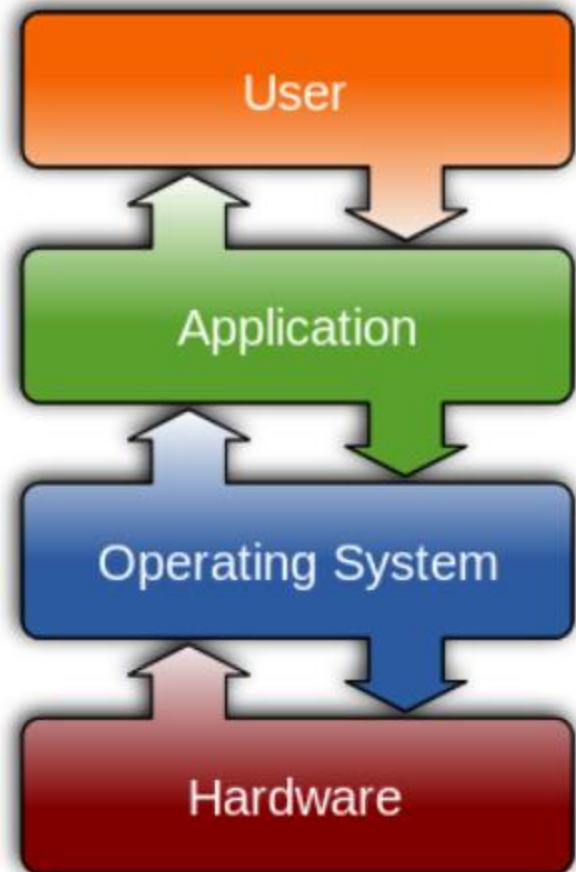
- 超级计算机架构。
- **a** 个主机计算机它的自定义操作系统, 说 **ibm zos**。
- **a** 个中端计算机它的自定义操作系统, 说一个 **ibm os/400**。
- 基于处理器的运行 **unix** 变型的机器:
 - 运行 **solaris** 操作系统的 **sun** 计算机。
 - **dec alpha** 集群在 **openvms** 下运行。
- **Macintosh**, 自定义 **apple** 电脑硬件和 **mac os** 操作系统 (现在在 **x86** 上迁移)。
 - 运行牛顿操作系统的牛顿设备, 也来自苹果。
- 商品计算机平台, 例如:
 - **wintel**, 即英特尔 **x86** 或兼容的硬件和 **windows** 操作系统。
 - **x86** 与其他类似 **unix** 的系统, 如 **bsd** 变型。
- **gumsfir** 全功能微型电脑与 **linux**。
- 在移动设备中找到的 **arm** 体系结构。
- 任何种类的电子游戏控制台。

软件平台示例

- java-jdk 和 jre
- .net 框架
- 莫兹拉·普拉·普勒和 xul 赛跑者
- adobe air
- 单

计算机操作系统 (一)

- 操作系统 (操作系统) 是一种管理计算机硬件和软件资源的软件。
- 操作系统任务包括控制和分配内存, 确定指令处理的优先级, 控制输入和输出设备, 促进联网和管理文件。
- 中。 内核是所有操作系统中最低的级别
- 大多数操作系统都包含管理图形用户界面 (windows) 的系统软件。 其他人使用 cli 或命令行界面 (unix)



http://en.wikipedia.org/wiki/Operating_system

计算操作系统 (二)

■ 操作系统充当**应用程序的主机**在机器上运行的。

- 作为**主机**, 操作系统的目的一次是处理硬件操作的细节。这样可以使应用程序不必管理这些详细信息, 并且可以更轻松地编写应用程序。
- 几乎所有的计算机, 包括掌上电脑、台式计算机、超级计算机, 甚至视频游戏机, 都使用某种类型的操作系统。
- 但是, 一些最古老的模型可能会使用**嵌入式操作系统**, 该设备可能包含在光盘或其他数据存储设备上。

■ 常见的当代操作系统包括

- 微软视窗**、**mac os**、**linux** 和 **solaris**
- 微软视窗**在台式机和笔记本电脑市场占有相当大的市场份额

■ 服务器通常在 **unix** 或类似 **unix** 的系统上运行

- aix**、**hp-ux**、**irix**、**linux**、**minix**、**ultrix** 和 **xenix**

■ 嵌入式设备市场由多个操作系统划分

案例分析: 客户服务----对 windows 服务器上的多线程应用程序的单个 cpu 使用情况进行故障排除, 使用 "复制" 命令?

大型机上的操作系统

- 整个 20世纪60年代,许多主要功能都是在操作系统领域首创的。
- 的发展。**ibm system-360**生产了一个以不同的容量和价格点提供的主机计算机系列,为此计划使用一个操作系统 **osw 360** (而不是为每个模型开发临时程序)。
- 跨整个产品线的单一操作系统的这一概念对于 **system-360** 的成功至关重要,事实上, **ibm** 目前的大型机操作系统是这个原始系统的遥远后代;为 **os360** 编写的应用程序仍然可以在现代计算机上运行。
- **os360** 开创了一些概念,在某些情况下,这些概念在主机领域之外仍看不到。例如,在 **os360** 中,当程序启动时,操作系统**跟踪所有系统资源**使用的资源包括存储、锁、数据文件等。当进程终止时,操作系统将重新声明所有资源。
- 另一个 **cp-67** 系统启动了整个操作系统系列,重点是**虚拟机的概念**.

微型计算机上的操作系统

- 第一台微型计算机不具备或不需要为大型机和小型机开发的复杂操作系统;简约的操作系统被开发,通常加载从**磁盘**被称为**显示器**.
- 一个值得注意的早期**基于磁盘的操作系统**是 cp/m, 它在许多早期的微型计算机上得到了支持, 并在 ms-dos 中被密切模仿, ms-dos 作为 ibm pc (ibm 的版本被称为 ibm dos 或 pc dos) 选择的操作系统变得非常流行, 它的继任者使微软成为世界上的最赚钱的公司。
- 在80年代的苹果电脑公司 (现为苹果公司), 放弃了其流行的苹果 ii 系列微型计算机, 推出了苹果 macintosh 电脑与创新**图形用户界面 (gui)**到 mac 操作系统。

ibm 超级电脑沃森的操作系统

watson 是一个为复杂分析而设计的工作负载优化系统, 通过集成大规模并行 power7 处理器和 ibm deepqa 软件来回答问题, 使其成为可能危险! 在三秒内提问。

watson 由 90个 ibm power 750 台服务器 (加上10个机架中的额外 ito、网络和群集控制器节点) 组成, 共有 2880 power7 处理器内核和 16 tb 的 ram。每台 power 750 服务器使用 3.5 ghz power7 8个核心处理器, 每个内核有4个线程。power7 处理器的大规模并行处理能力是与沃森 ibm deepqa 软件的理想匹配, 该软件的工作负载很容易被拆分为多个并行任务。

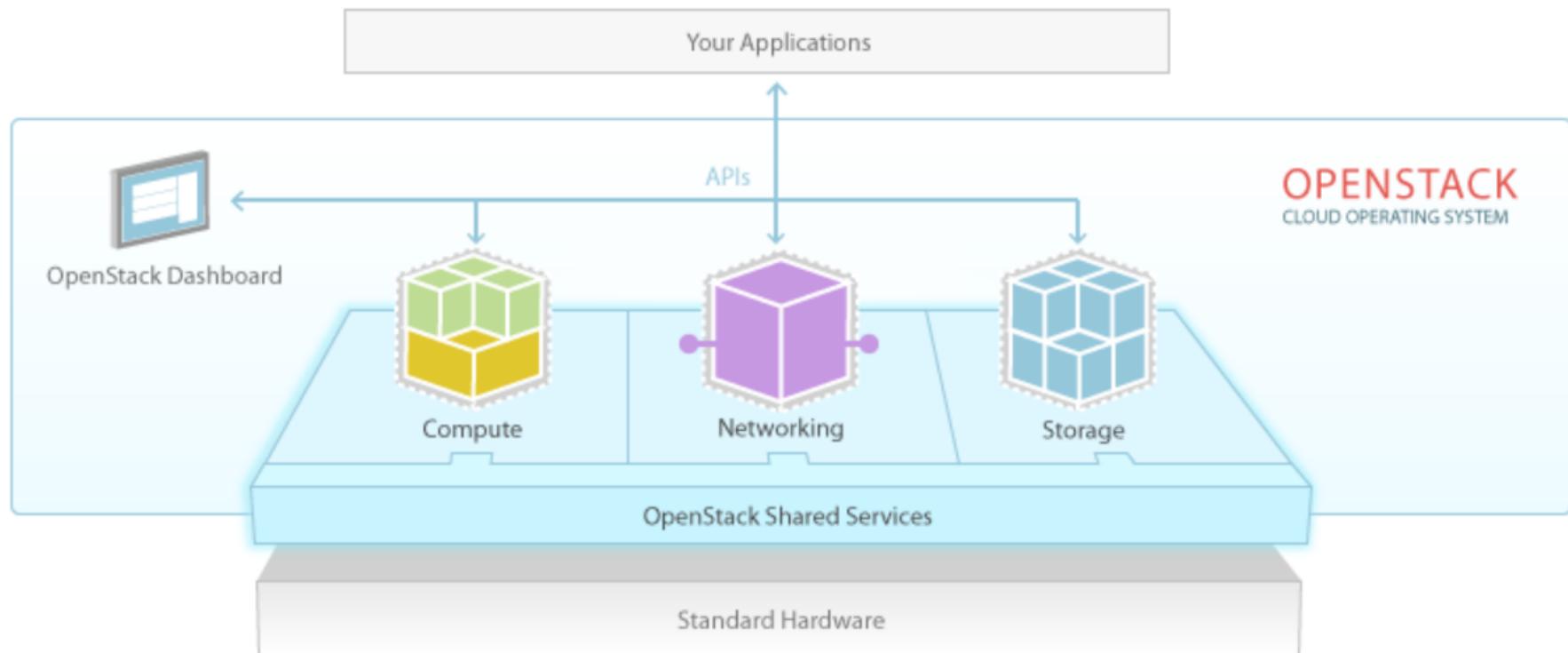




openstack
CLOUD SOFTWARE

适用于云的操作系统

开源云操作系统



<http://en.wikipedia.org/wiki/OpenStack>

<http://www.openstack.org/software/>

适用于大数据应用的操作系统

- 用于业务分析和数据仓库的 **ibm 大型机 zos**
- 中间层数据中心操作系统
- 新的英特尔数据平台提供最新版本的 **hadoop** 发行版
- 用于云计算的弹性操作系统

"分析操作系统"

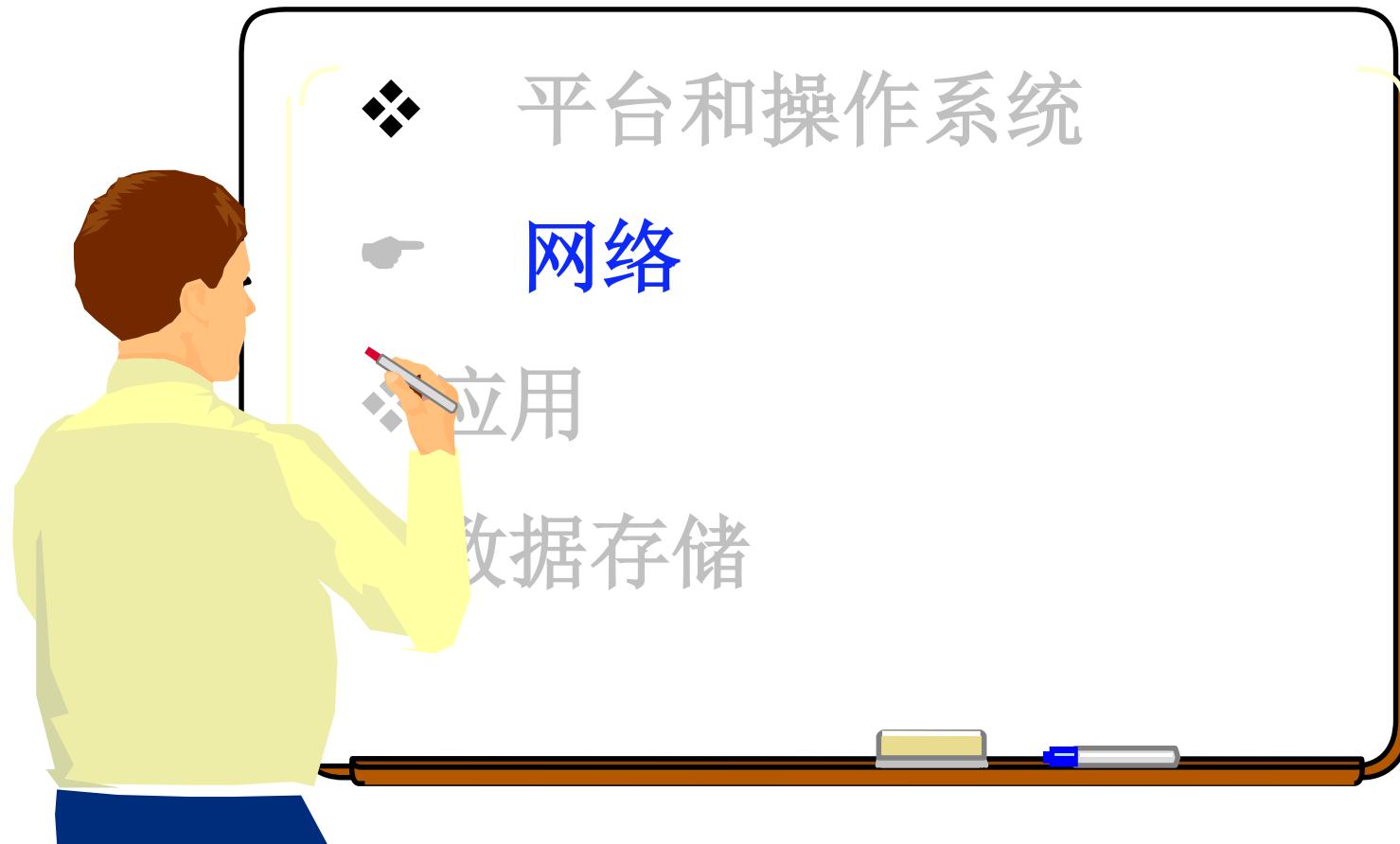
- **spark** 是一个开源分析操作系统.
- 它是实时运行迭代交互式分析的平台, 可以嵌入到应用程序中。
- **ibm** 将此引擎用作我们的分析和商务平台、**watson** 运行状况云等平台的一部分。
- 数据科学家和数据工程师今天使用了这一技术, 但像 **spark** 这样的平台最大的障碍之一是缺乏具备这些数据技能的人。作为回应, **ibm** 致力于在全球范围内教育**100万**数据科学家和数据工程师。

国家战略计算倡议

nsci 的目标:

- 加速交付一个有能力的 **exascale** 计算系统, 该系统集成了硬件和软件功能, 可提供约**100元**代表政府需求的一系列应用中, 当前的 10个 petaflop 系统的性能。
- 提高用于建模和仿真的技术基础与用于数据分析计算的技术基础之间的一致性。
- 建立, 在未来**15年**, 即使在当前半导体技术的极限 ("后摩尔定律时代") 达到极限之后, 未来的 hpc 系统也会有一条可行的前进道路。
- 通过采用整体方法解决网络技术、工作流程、向下扩展、基本算法和软件、可访问性和劳动力发展。
- 发展持久的公私合作, 以确保美国政府与工业和学术部门在最大程度上分享研究和发展进展的好处。

内容

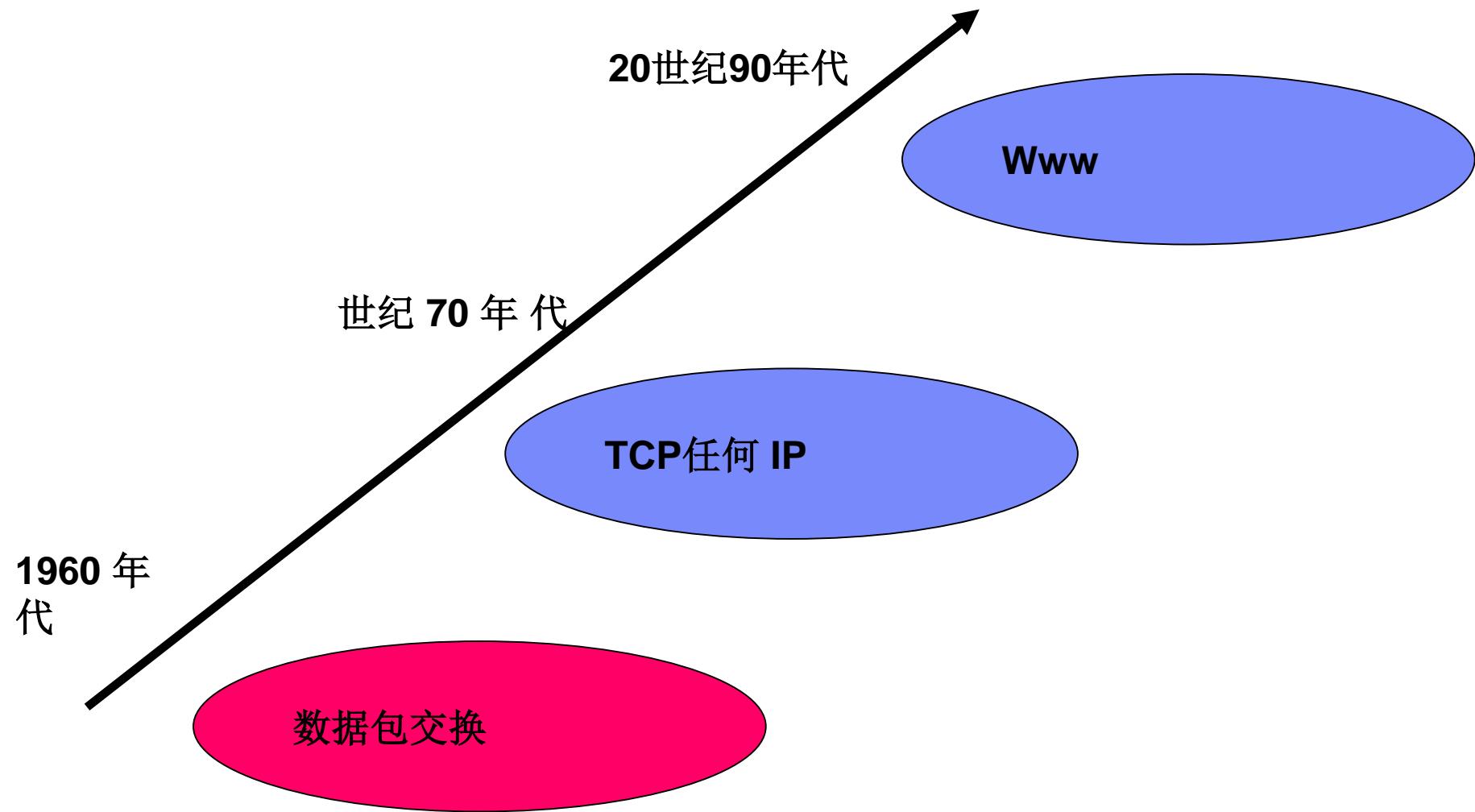


计算机网络

重要的网络概念包括：

- 枢纽
- 路由器
- 打印机
- 防火墙
- 开关
- 光纤面板
- 存储区域网络 (san)
- 服务器网络接口
- 其他局域网 (lan) 组件
- 广域网 (wan) 电路
- 城域网络 (man) 电路

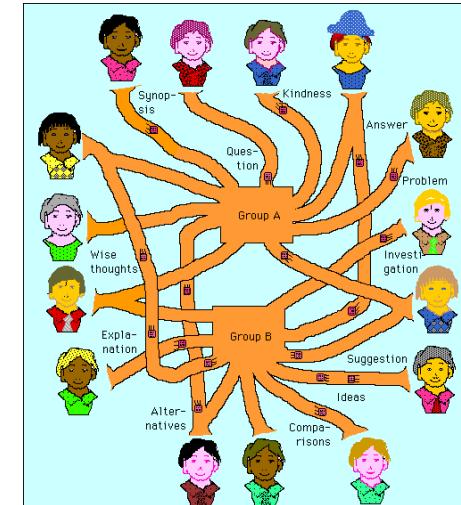
网络演进--3主要里程碑



到1970年为止, 大多数计算机**孤立**在这个意义上说, 数据或计算资源不是以电子方式链接的



孤立的岛屿

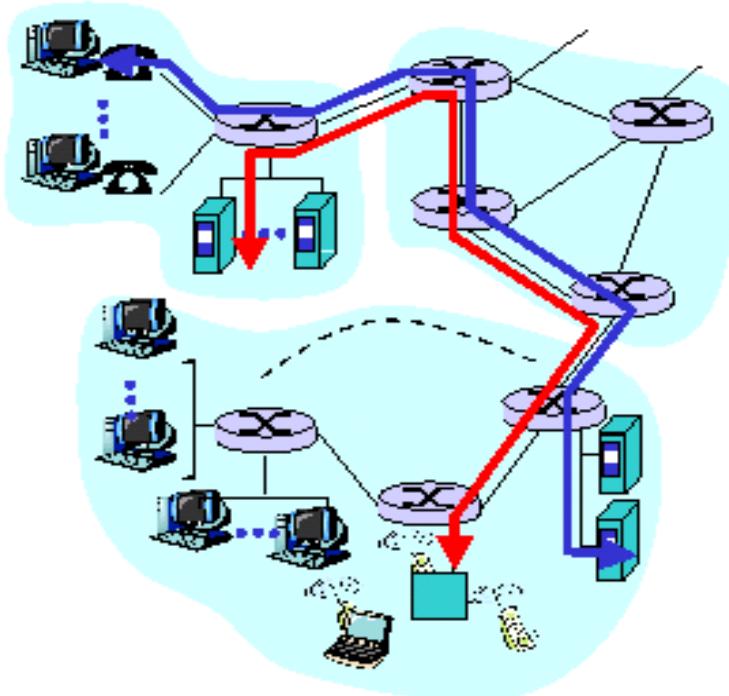


电路切换

电路切换用于电话通信。

Network Core: Circuit Switching

- End-end resources reserved for "call"
- link bandwidth, switch capacity
 - dedicated resources: no sharing
 - circuit-like (guaranteed) performance
 - call setup required



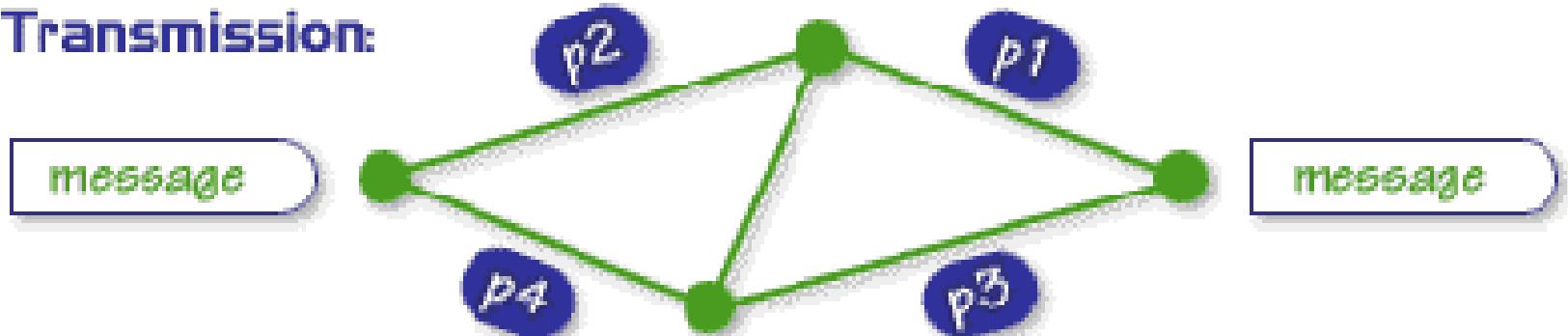
数据包交换

数据包交换用于互联网通信

Message: People look up to me for obvious reasons...

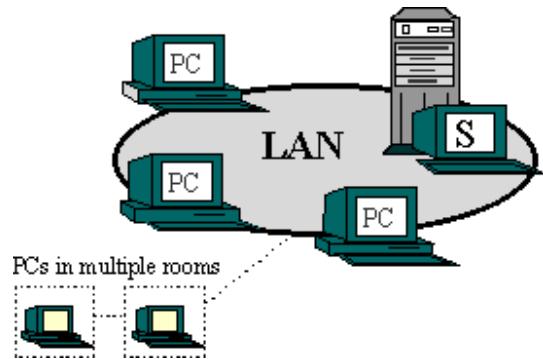
Packets: p1 People look
p2 up to me fo
p3 r obvious r
p4 easone...

Transmission:

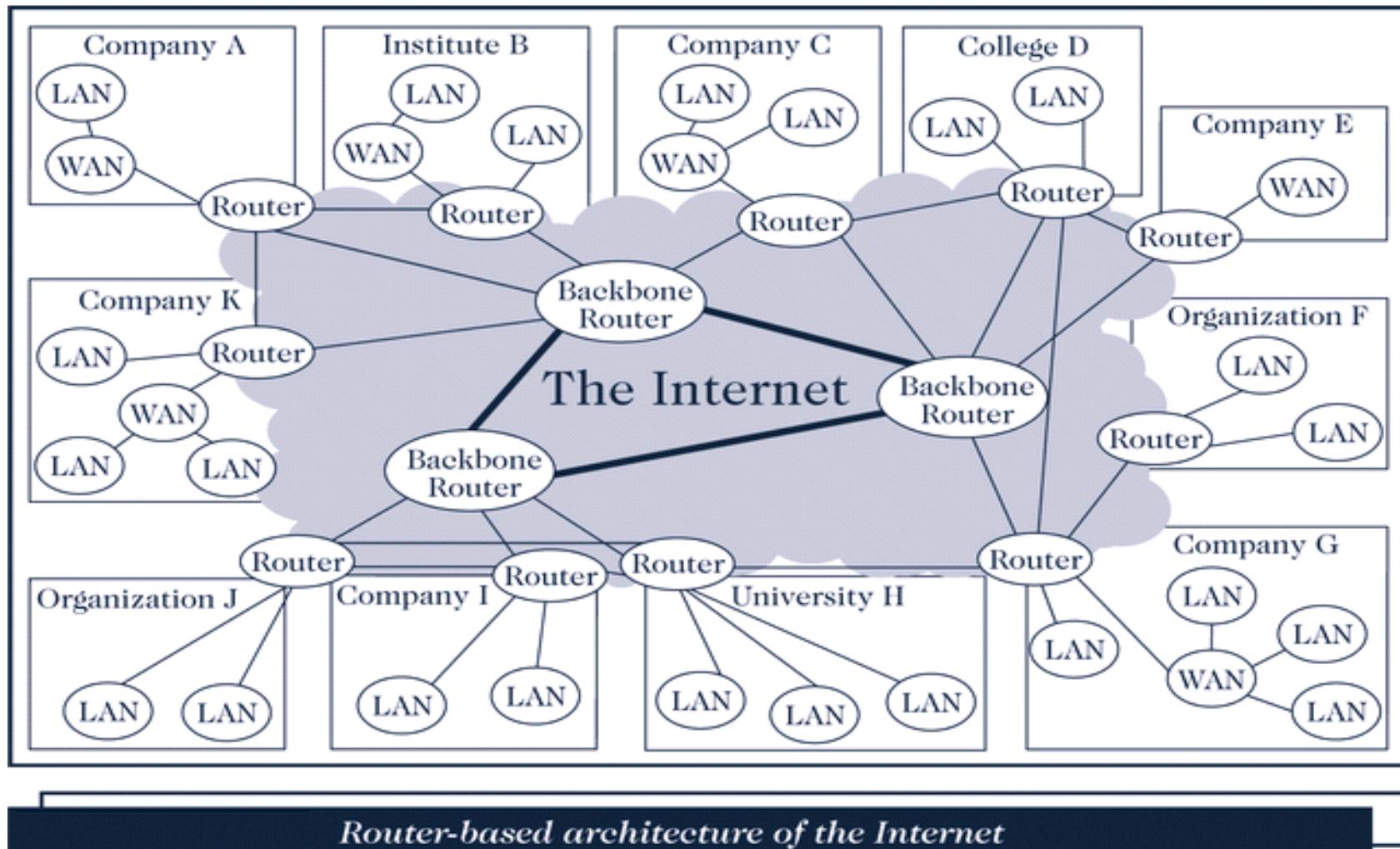


分组交换网络

- 局域网 (lan) 是连接在一起的计算机网络。
- 广域网 (wan) 是一个由很远的距离连接的计算机网络组成的网络。



基于路由器的体系结构



历史回顾

如何使这些计算机在网络上沟通彼此可靠、轻松？

电路交换与分组交换

- 20世纪60年代初

- alohanet

- arpanet

alohonet

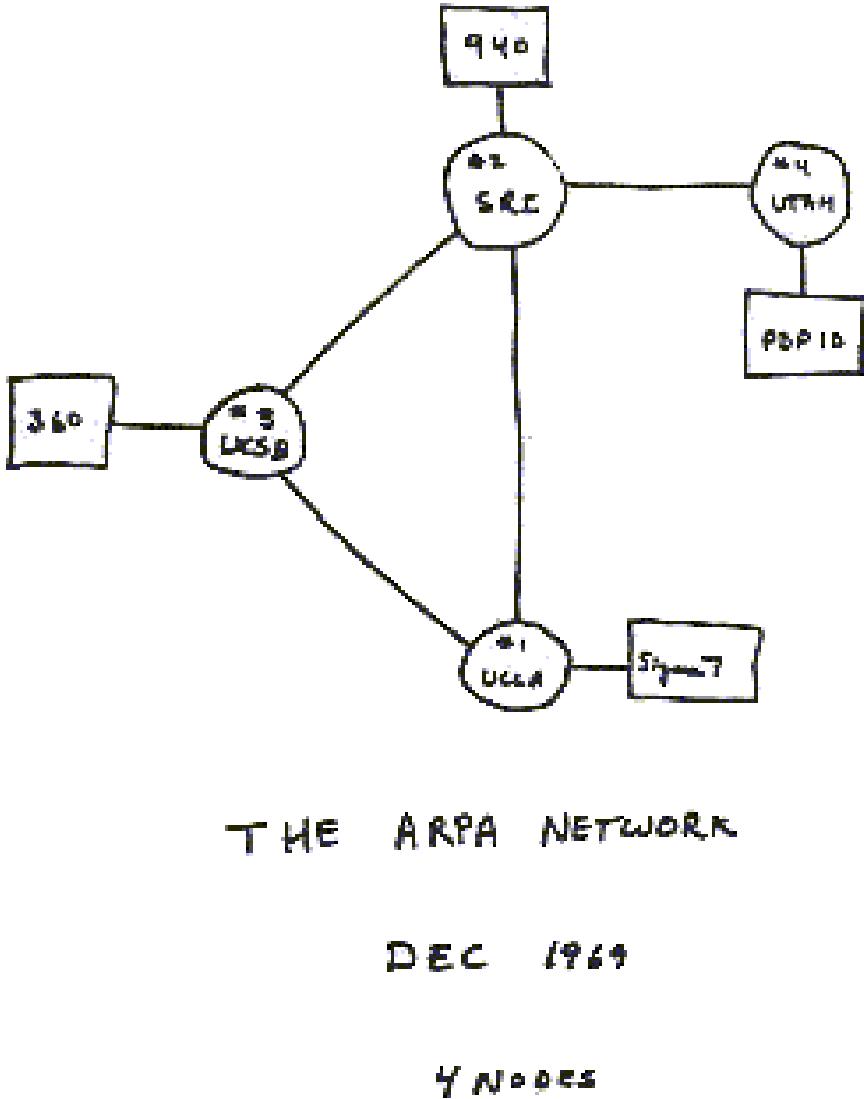
也称为阿罗哈，是夏威夷大学开发的一个开拓性计算机网络系统

。

它最初是在1970年部署的，虽然网络本身已不再使用，但网络中的核心概念之一是广泛使用的以太网的基础。



阿帕内: A大卫R研究与应用P漫游者a个里网工作



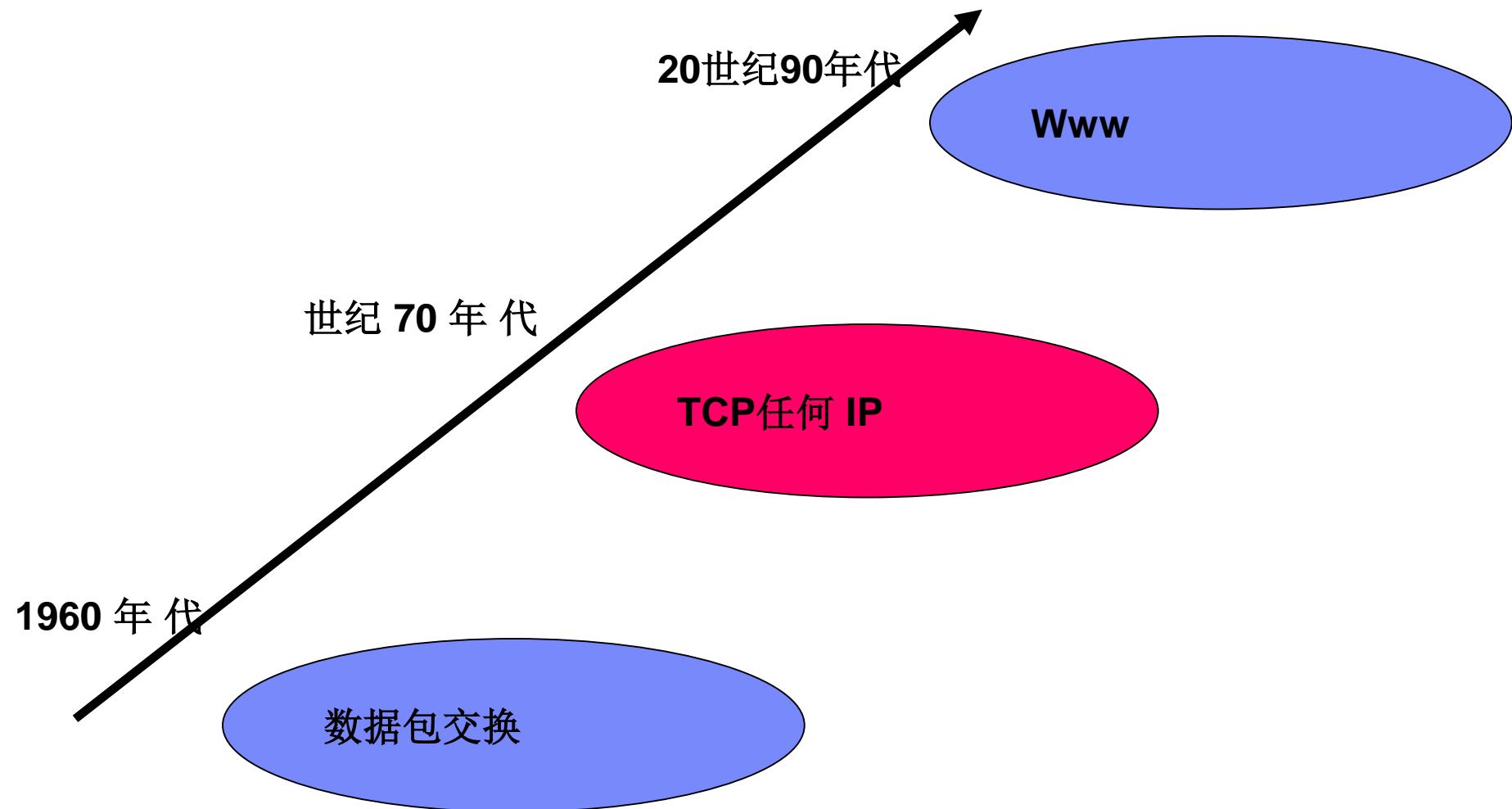
1969年在这张纸上爬来爬去

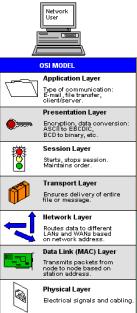
是 arpanet 的前四个节点。他们几乎没有意识到, 这是网络的开始, 每天都会有数千万人使用这个网络

4台电脑互联网投入运行

- u. california, los angeles
- u. california, 圣巴巴拉
- 斯里兰卡(斯坦福研究所)国际
- 犹他大学

网络演进--3主要里程碑





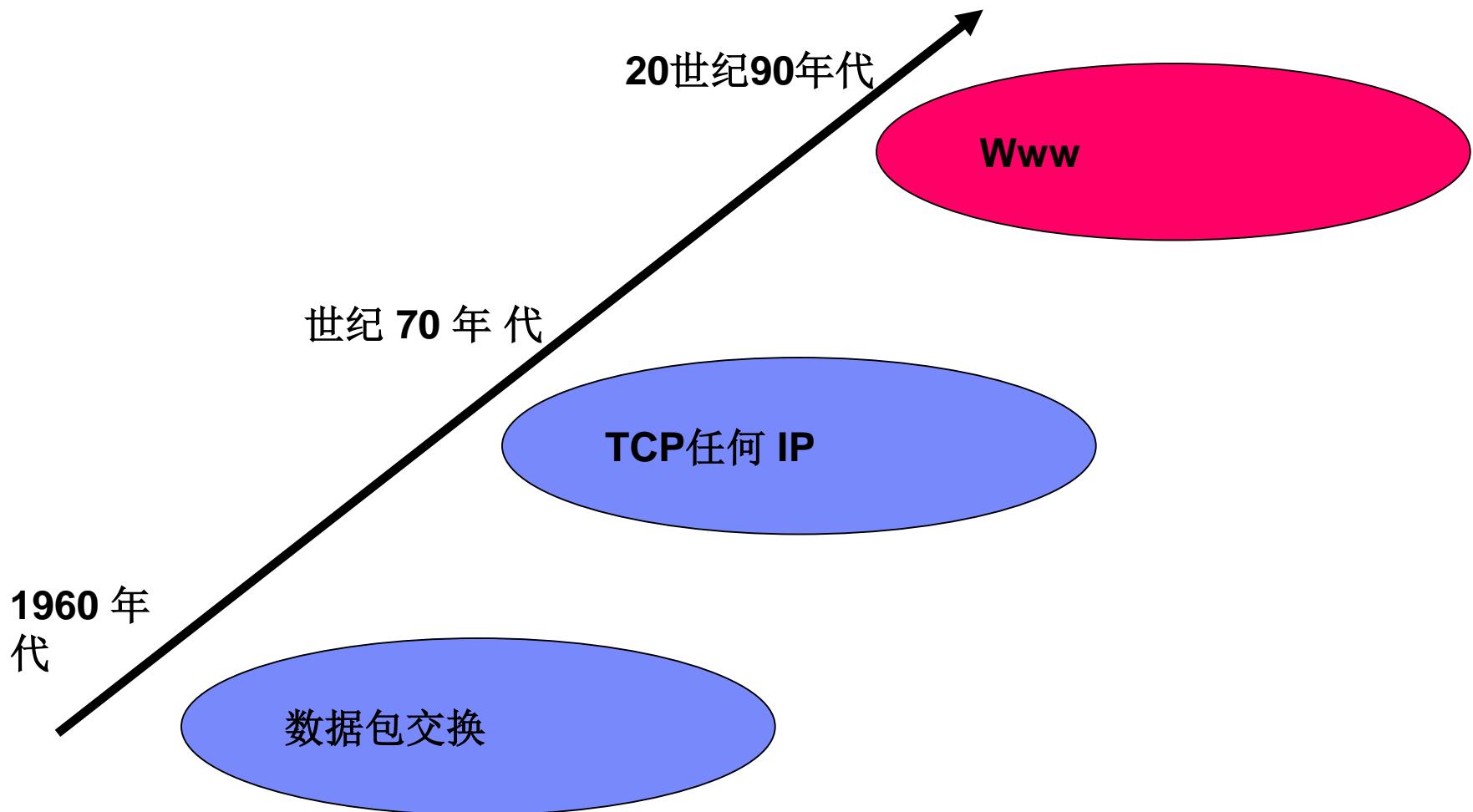
互联网协议

- 传输控制协议 (tcp) 和互联网协议 (ip) 是支持互联网操作的两种协议 (通常称为**TCP任何IP**).
- 中。**Tcp**控制**拆卸**将消息传输到数据包中, 然后再通过互联网传输, **重组**这些数据包, 当他们到达他们的目的地。
- 中。**Ip**指定**寻址详细信息**为传输的每个数据包。

这种开放式的体系结构**四个关键规则**为互联网的成功做出了贡献。

1. 独立的网络应**不要求**任何内部**变化**连接到网络。
2. **包**不到达他们的目的地必须是**重新传送**从他们的源网络。
3. **路由器**计算机的作用是**接收和转发**设备;它们不保留有关其处理的数据包的信息。
4. **无全局控制**存在于网络上。

网络演进--3主要里程碑



互联网和万维网

- **互联网**: 一个跨越世界的大型连接计算机系统
- **Www**: 互联网上具有易于使用的标准界面

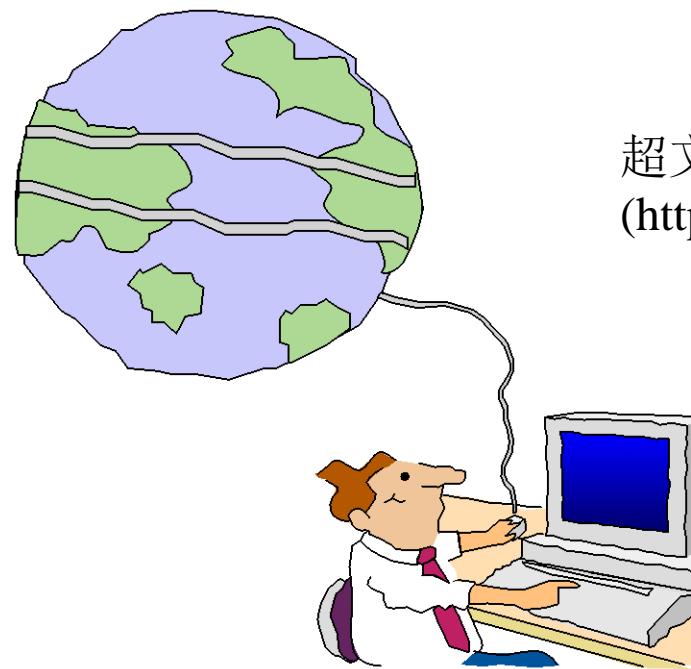
- 美国国防部的原始目标
 - 控制武器系统和转移研究档案
- 其他用途在过去的**30年**里, 互联网的迅速发展
 - 电子邮件始于1972年
 - 乌塞内(新闻组: 会员可以阅读和发布)
 - 其他应用, 如游戏
- 20世纪80年代, 随着个人电脑的出现, 学术网络和企业网络一体化的必要性变得显而易见。

互联网 主要功能

- 电子邮件 (电子邮件)
- 文件传输 (ftp)
- 远程访问
- WWW 服务 (网址)
- 公告
- 新闻



网络服务器



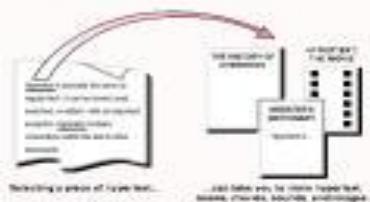
超文本传输协议
(http)

网络客户端
使用浏览器

万维网或网络

- 网络更一种思维方式关于和组织信息存储和检索比它是一个特定的技术
 - 两个关键要素
 - 超文本
 - 图形用户界面
- 网络使人们更容易使用互联网，从而创造急剧的扩张互联网社区的



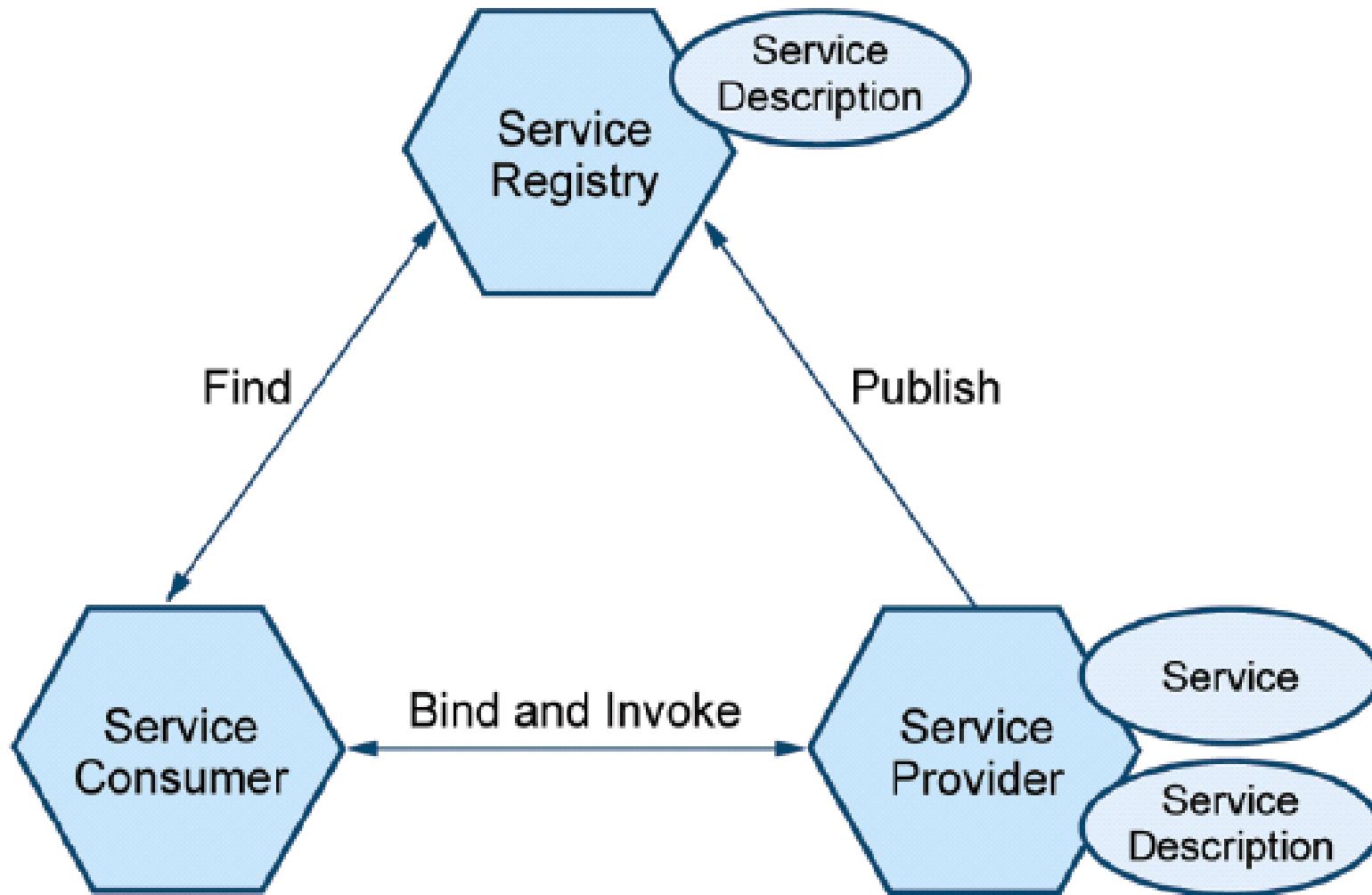


超文本

- 超文本是一个[页面链接机制](#)将连接相关的信息页面,而不考虑信息的存储位置。
- 我们都这样做 "[超发短信](#)"
— 一个很好的例子:[图书管理员](#)
- 这一概念已正式化,并在[1945年由万内瓦尔·布什](#)
- **1989:** [蒂姆·伯纳斯-李](#),在[欧洲核子研究中心 \(欧洲人权委员会\)](#),开始开发[超文本服务器](#)(现在称为 web 服务器),它存储用超文本标记语言 ([Html](#)),并允许其他计算机连接到它并读取这些文件

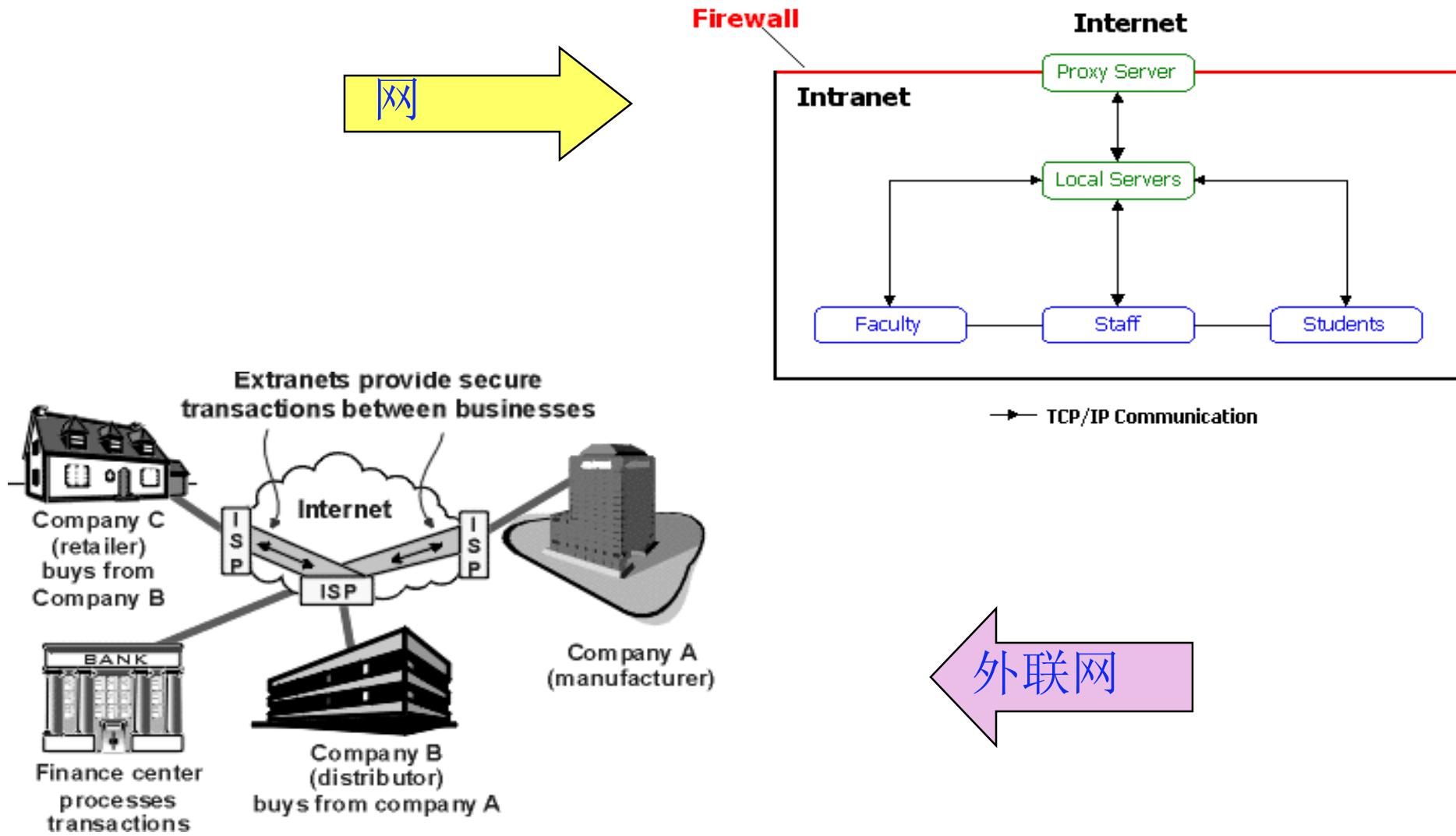
- 网页交付**
- 超文本传输协议 ([http](#)) 是一组规则,用于[提供网页](#)在互联网上。
 - [http](#) 使用[客户端/服务器模型](#)
 - 用户的网站[浏览器](#)打开一个[http 会话](#)并将网页请求发送到远程服务器。
 - 作为响应,服务器将创建一个 [http 响应消息](#),并将其发送回客户端的 [web 浏览器](#).
 - 协议名称和域名的组合称为[统一的资源定位器\(Url\)](#). 例如,<http://www.almaden.ibm.com>

网络服务模型

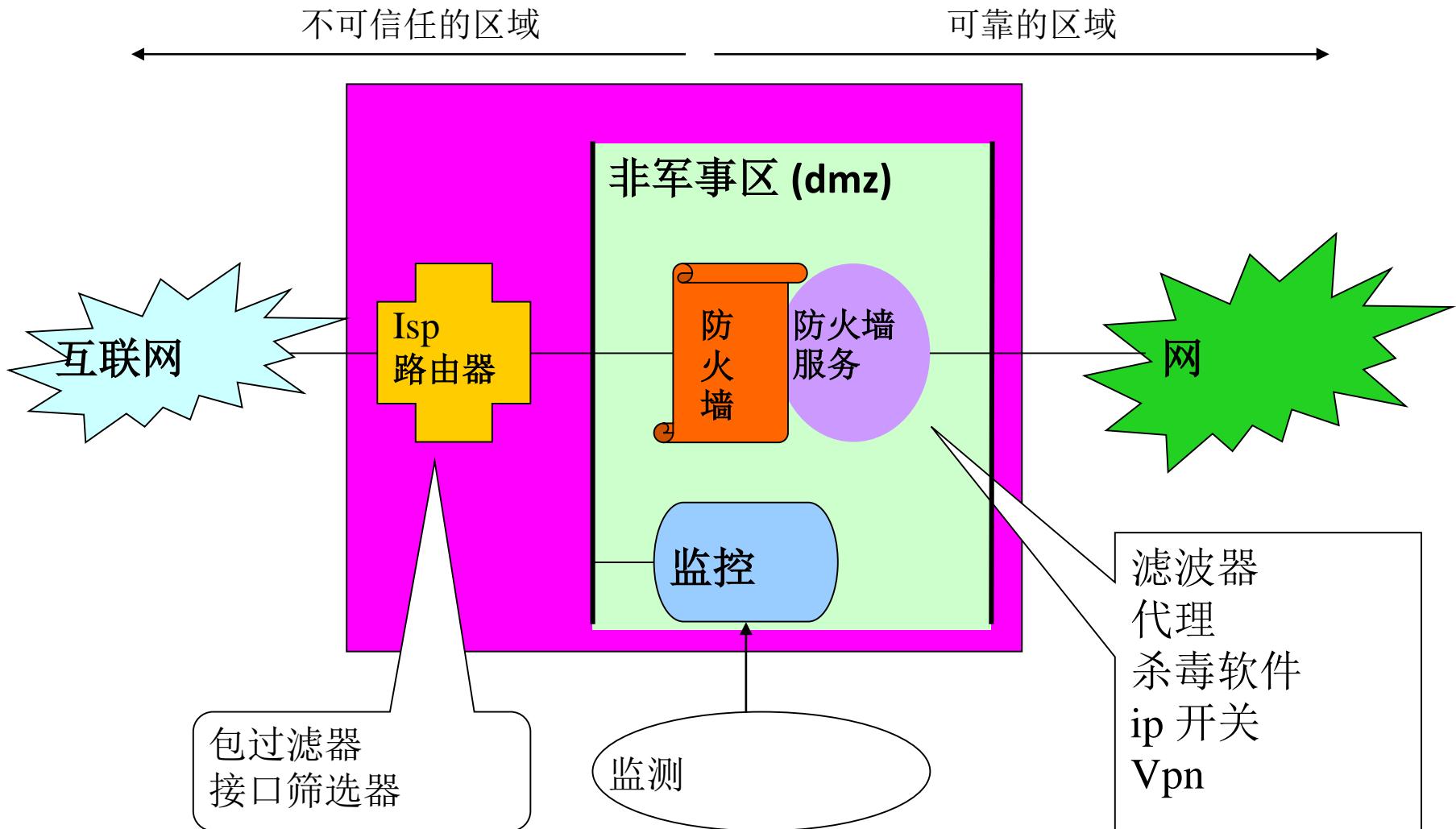


查看更多来自: web 开发技术. doc

内联网和外部网



dmz 和防火墙



案例研究: 韩国的术语 dmz

geni--探索未来的网络

■ 全球网络创新环境 (geni)

- geni 是一个虚拟实验室, 用于大规模探索未来的互联网。
- geni 创造了重要的机会来理解创新, 并改变全球网络及其与社会的互动。
- geni 在网络科学和工程的前沿开辟了新的研究领域, 并增加了产生重大社会经济影响的机会。



<http://www.geni.net/>

大数据网络的六个注意事项

对于大数据应用程序，需要

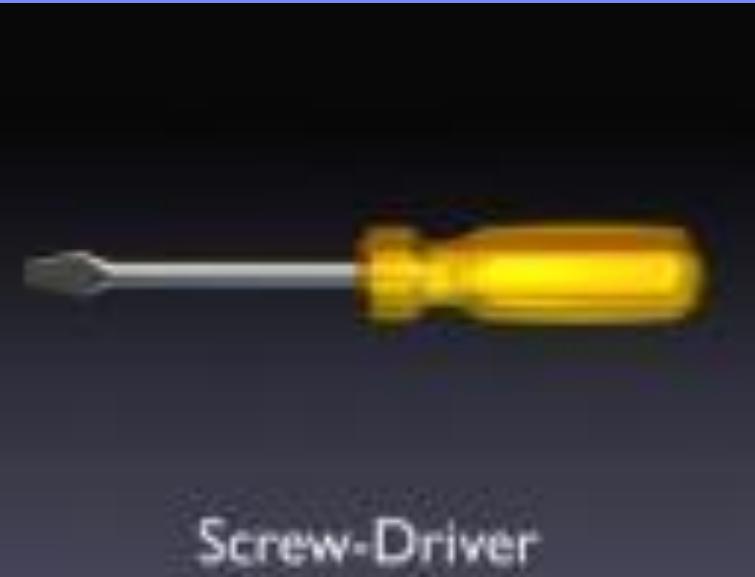
1. 对故障的网络恢复能力
2. 解决网络拥塞问题
3. 跨空间和时间的网络一致性能
4. 现在就准备好大数据可伸缩性
5. 处理大数据的网络分区
6. 大数据网络的应用感知

内容



应用工具

单一使用工具



多种使用工具



通过使用工具，是否所有的人都被创造平等？！

计算机

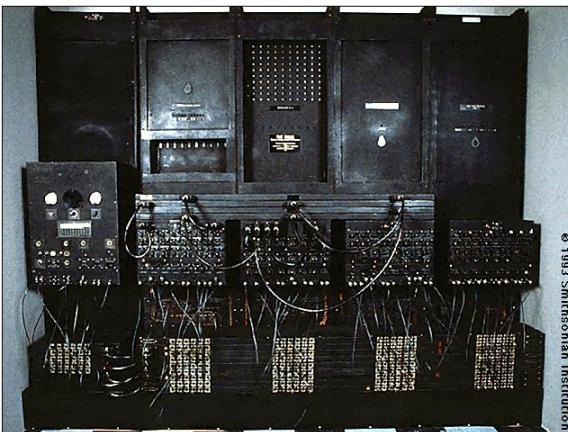
ibm 360



ibm 7090



illiac 4



Eniac

他们是做什么的？他们计算。

这就是为什么他们被称为计算机。

我们是否应该继续称它们为 "计算机"？

- (科学)计算
- 数据**处理**
- 符号**处理**
- 词**处理**
- 电子邮件**处理**
- Web**冲浪**
- 游戏
-



这真的是一个细胞**电话**？

是什么让 "电脑" 不做计算？

电脑软件

- 计算机软件是为计算机提供执行任务的能力的程序和过程。
- 有**三**主要软件类：
 - **系统软件**运行计算机硬件和计算机系统。
 - **编程软件**提供了使用不同编程语言编写计算机程序和软件的工具
 - **应用软件**

应用软件分类

- 分析软件
 - 统计包
 - 协作软件
 - 博客, 维基的
 - 以计算机为媒介的通信
 - 电子邮件, 网络浏览器
 - 商业软件
 - 客户关系管理、企业业务软件等
 - 数据库软件
 - 甲骨文、db2、微软 sql、informix、nosql
 - 娱乐、多媒体和艺术软件
 - 电子游戏, 图片编辑 sw,
- 中间件软件
 - 消息队列系列, 燕尾服
 - it 管理软件
 - 蒂沃利、ca unicenter、hp oplopen view 等
- 复杂?

为什么选择软件工程?

- 问题是**复杂性**
- 许多来源, 但**大小是关键:**
 - Unix包含**400万条线路**的代码
 - 视窗2000包含 **10^8 线**的代码

软件工程就是要管理这种**复杂性**.



软件工程知识体系

计算基础知识

- 算法和数据结构
- 计算机体系结构
- 数学基础
- 操作系统
- 编程语言

软件产品工程

- 需求工程
- 软件设计
- 软件编码
- 软件测试
- 软件操作和维护

软件管理

- 项目过程管理
- 风险与管理
- 质量管理
- 配置管理
- 流程管理
- 采购管理

软件领域

- 人工智能
- 数据库系统
- 人机交互
- 数值和符号梳子。
- 计算机模拟
- 实时系统



资料来源:

<http://www.sei.cmu.edu/pub/documents/99.reports/pdf/99tr004.pdf>

热软件工程技能

今天和明天

- visual basic--编程语言
- 微软访问–关系数据库 (小型应用程序)
- powerbuilder 应用程序开发工具–rad 数据库工具
- c 和 c++ 编程--编程语言
- sap–企业应用软件 (财务、制造、产品数据管理)
- mvs、 unix 和 windows nt–主机操作系统 (ibm、 sun、 hp、 microsoft)
- 微软交换–电子邮件服务器
- lotus nos–基于 web 的公开管理应用程序软件
- 数据库管理、管理和开发-甲骨文和 sybase 在较小程度上。 –甲骨文在 dbms 市场占据主导地位
- 因特网/万维网: 与网络有关的技能;例如, java、 activex、 cgi 和互联网安全--当然--看看 web 和电子商务的爆炸式增长
- 面向对象的技能--分析、设计、开发



好好学习 C++!



今天的软件工程

- 软件工程
 - 包括以程和产品的工程学科：
- -软件工程管理
- -软件需求分析
- -软件配置管理
- -软件设计
- -软件建设
- -软件测试
- -软件工程基础设施
- -软件工程流程
- -软件的开发和维护
- -软件质量分析



资料来源: swebok– <http://www.swebok.org>

软件工程概述

- 软件工程有着光明的前景
 - 工作机会
- 网络将是软件工程增长的主要推动力
- 成功的软件工程师将不断学习和适应新技术
 - 了解如何解决问题
 - 学习如何沟通 (口头和书面)
 - 了解如何提供服务

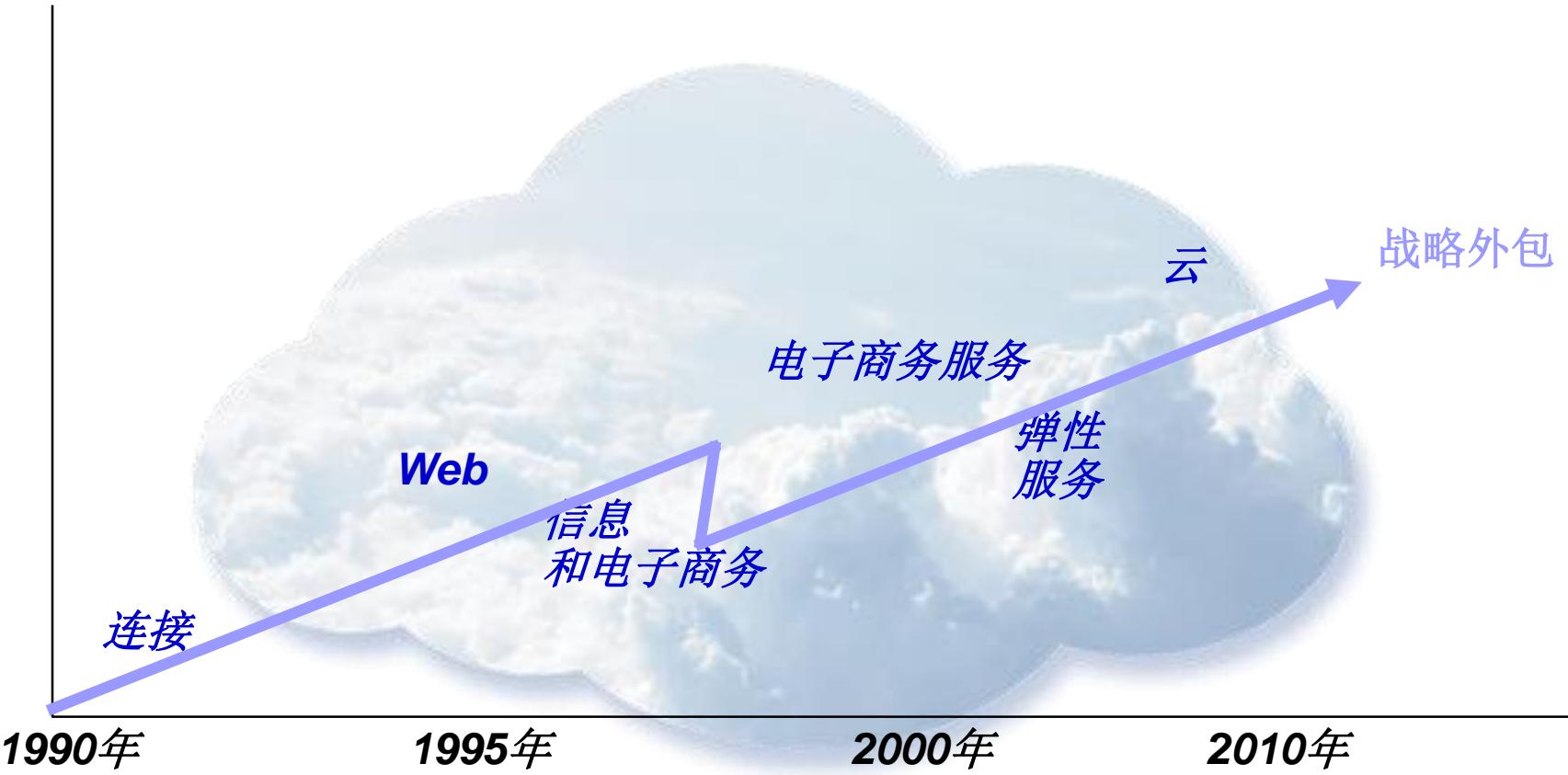


知识就是力量!

软件即服务

- 大量的企业或 "打包" 软件正在从已安装的产品转变为托管服务 (crm、hr、bi 是重要的部门, 供应链和 erp 即将到来)
- 客户不是在本地计算机上安装软件, 而是按订阅或按使用付费, 以便使用 web 浏览器访问功能
 - 谷歌还为电子邮件、电子表格和办公文档提供 **s-a-s** 应用程序 (以换取你知道什么)
- 许多 "已部署" 的软件应用程序已成为与重要的托管或 "供应商管理" **s-as-s** 组件的混合体

一组技术的最新创新



新的计算机应用程序--云服务

提供订阅服务和实例化

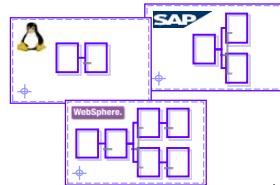
- 选择 "服务", 指定参数和 sla
- 自动实例化服务

提供创建和注册的服务

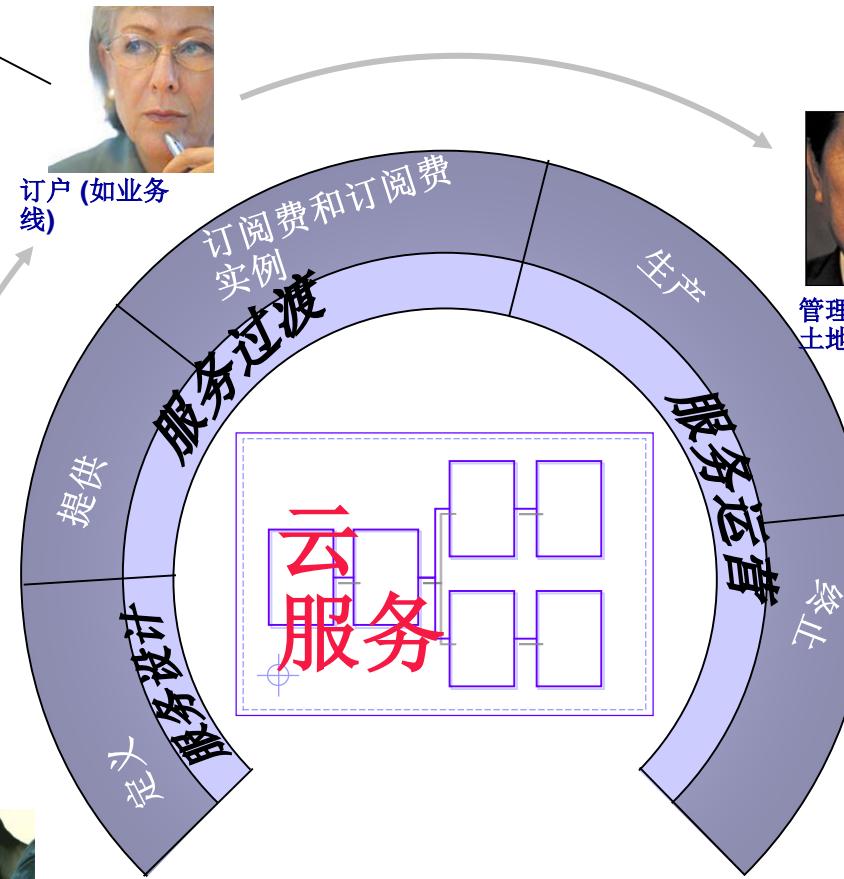
- 基于模板定义服务并将其注册到目录中

服务模板定义

- 为服务创建生成和管理计划



IBM/ISV
it 部



管理员/可持续
土地管理

利用自动化和虚
拟化, 手动或自
主执行管理计划

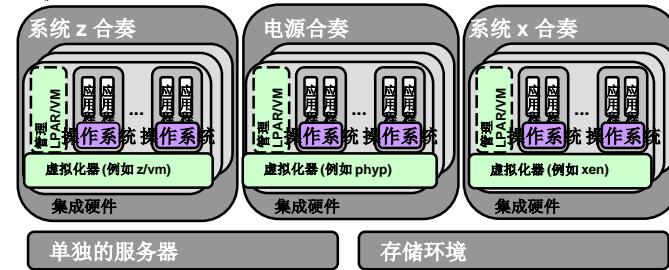
确保 sla 一致性



订户 (如业务
线)

服务实例终
止

销毁服务和
释放资源



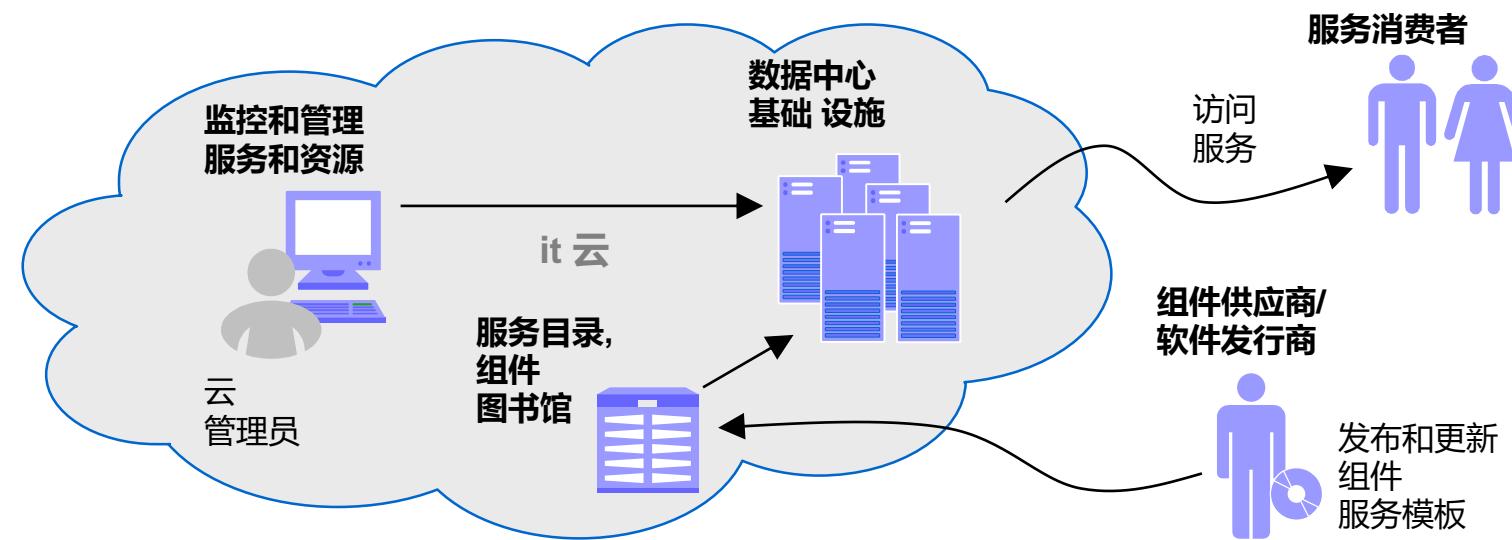
云计算

用户体验和业务模型

- 云计算是一种新兴的 it 交付方式, 其中应用程序、数据和 it 资源快速预配并提供为标准化产品户通过网络在灵活的定价模式。

基础设施管理和服务提供方法

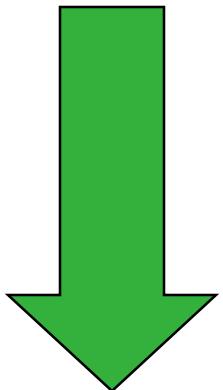
- 云计算是一种管理大量的高度虚拟化资源因此, 从管理的角度来看, 它们类似于单个大型资源。然后, 这可用于提供服务与弹性缩放.



需要架构转换

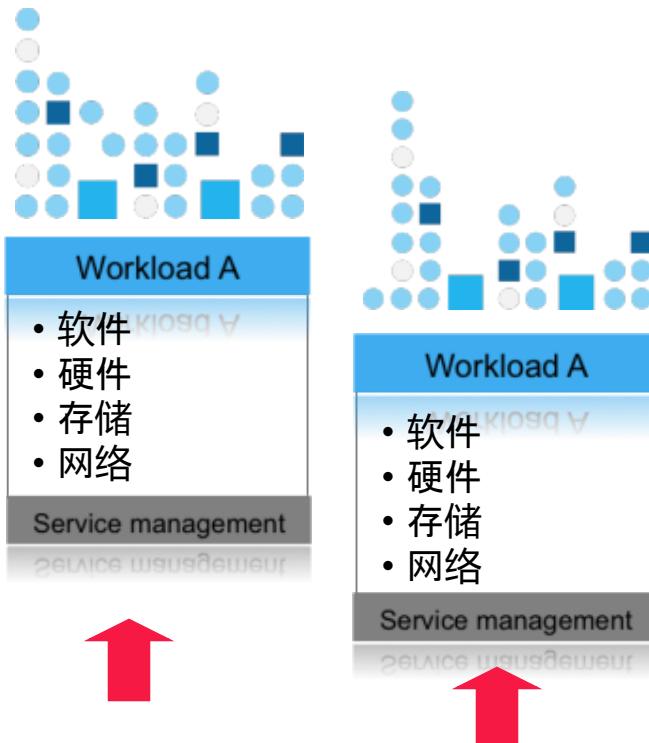
单实例 -> 多租户

- 多租户高效
 - 共享资源 (一个实例来运行它们)
- 定制
 - 通过配置进行自定义
- 可伸缩
 - 许多应用需要互联网规模



有什么区别?

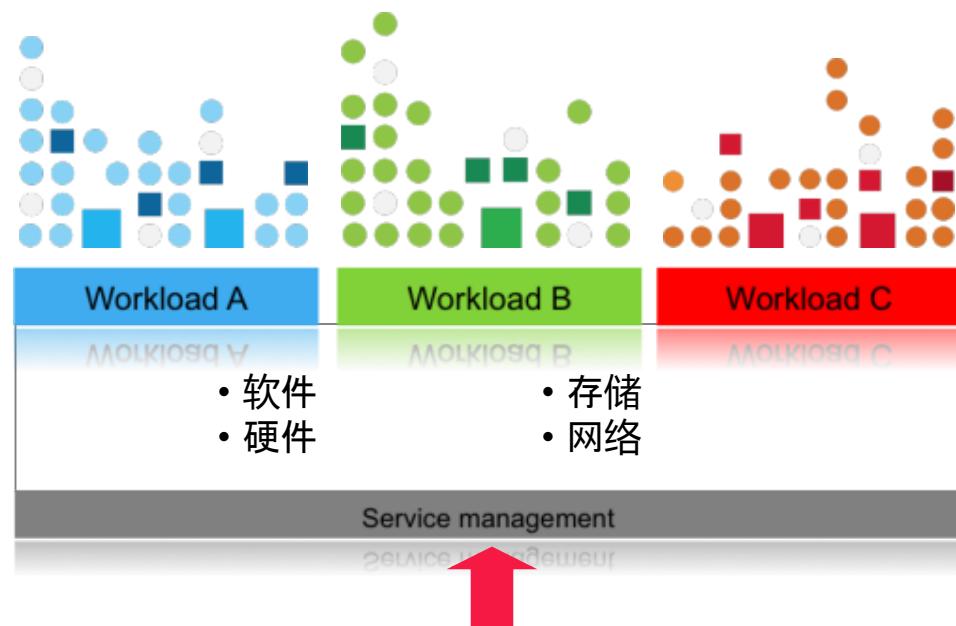
没有云计算



- 虚拟化资源
- 位置独立
- 快速的可扩展性

与云计算

- 自动化服务管理
- 标准化服务
- 自用-服务

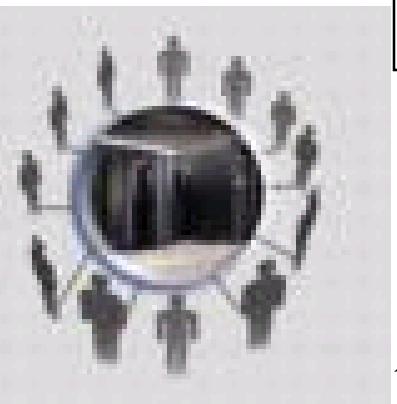


注: 从 nist、gartner、forrester 和 idc 云计算定义中提取的云计算元素

云计算的演进

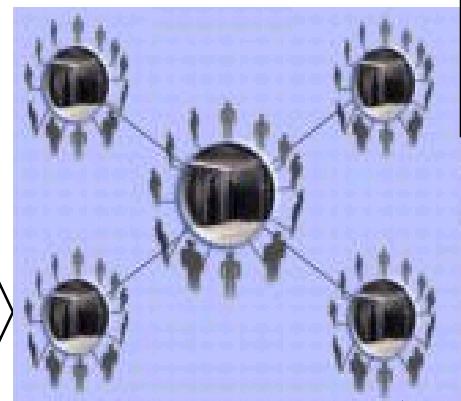


过去



司机

- 标准硬件
- 客户端服务器
- 互联网
- 电子商务



目前

司机

- 虚拟
- 高速网络
- web 2.0
- SaaS

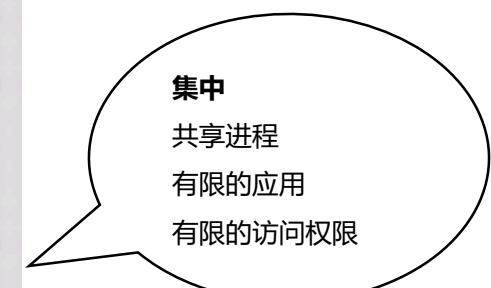


重新集中

共享基础架构
透明的服务交付
具有高带宽、低延迟的无处不在的访问

分布式

专用基础架构
应用的爆炸
无处不在的访问
计算岛屿



云改变一切

软件是如何交付的?

- trad: 在自设置的环境中本地安装
- 云: 在线随时提供服务, 而不是内部部署

如何维护部署?

- trad: 根据自己的时间线进行配置和维护
- 云: 配置和维护由服务提供商处理

更新、更改、新功能的频率?

- trad: 通常以月为单位, 可以是几年
- 云: 小批量; 可以更新多个时间/天

部署的唯一性?

- trad: 每个客户都是不同的 (网络、硬件, 存储)
- 云: 云部署是记录的部署
- trad: 单租户
- 云: 多租户
- trad: 一般与其他产品隔离, 无干扰
- 云: 嘈杂的邻居, 不可预知的背景

在失败或问题的情况下?

- trad: 确定根本原因并修复 (测试/发布修补程序)
- 云: 保持生产进行; 保持可操作性 (尽快恢复运行并记录下一次)

客户资料

- trad: 大投资后, 客户接受奋斗 "
- 云: 入境门槛低; 如果沮丧, 将去其他地方
- trad: 企业规模 (ibm 擅长此)
- 云: 小弗雷米姆, paygo (ibm 还不擅长这个)

如何测试?

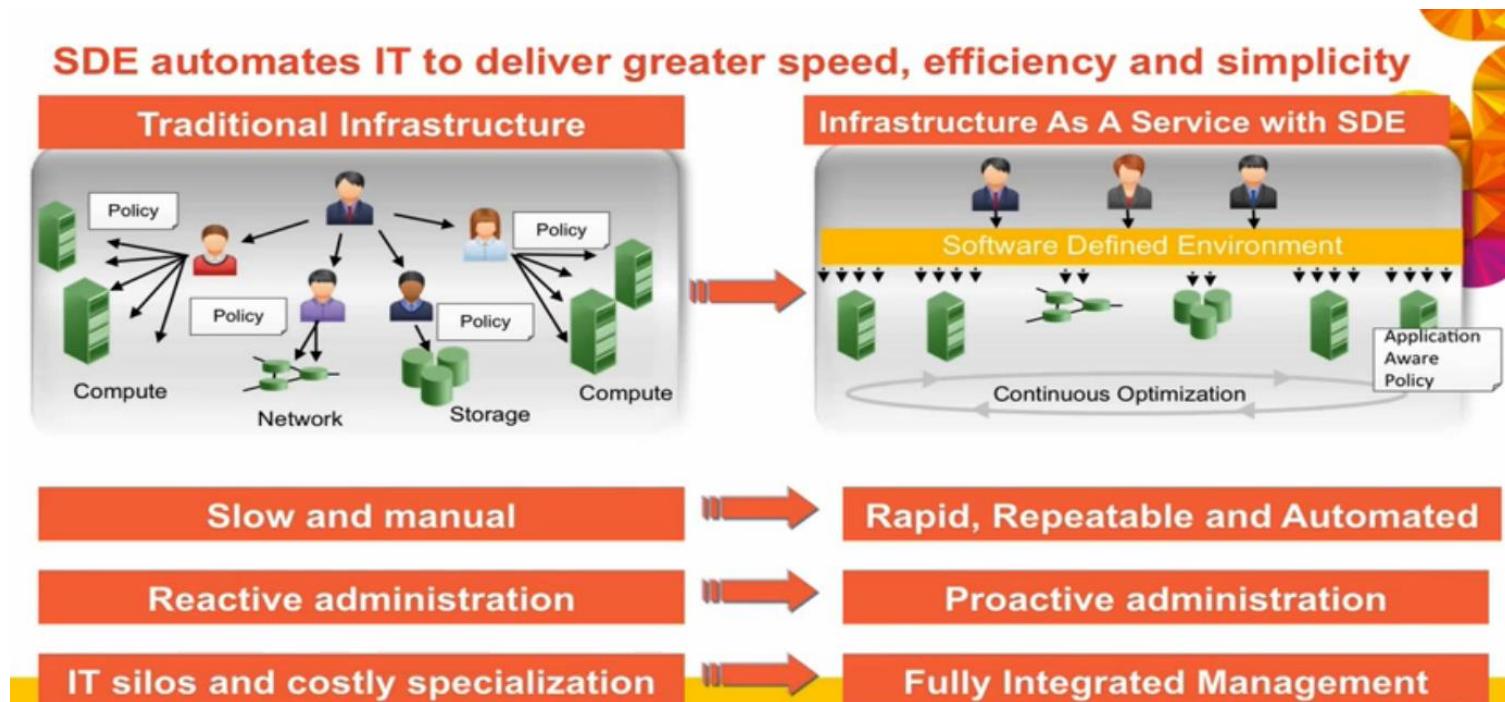
- trad: 洁净室, 隔离, 理想的条件
- 云: 嘈杂的邻居, 多租户, 肮脏的数据
- trad: 隔离建模和预测
- 云: 不可行的模型 n!组合

trad: 传统

软件定义的环境 (sde)

当包括计算、存储和网络在内的整个基础架构在云中成为软件定义和可编程时，出现了一个新的统一控制平面，它是高度可配置和完全可编程的，工作负载正在编译到它上。

这是软件定义环境 (sde) 的基础。

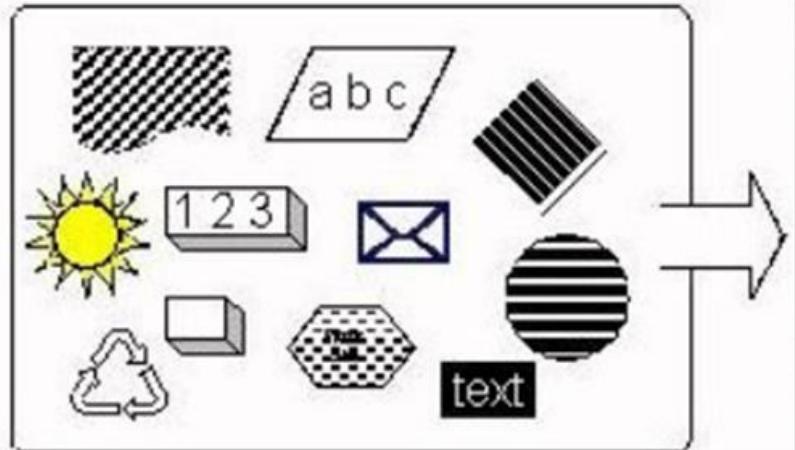


内容



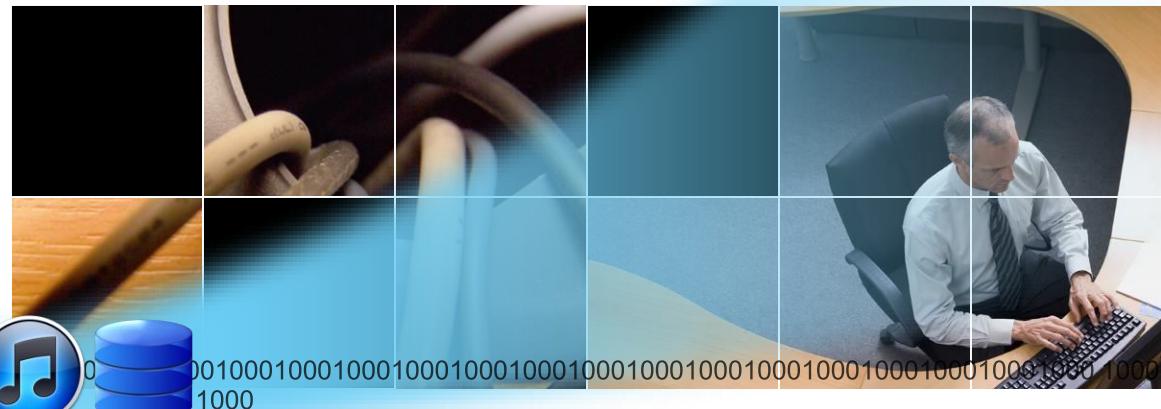
计算机中的数据

Real-World Information



Computer Data

```
01110101011010101  
10100101011010101  
01010101011010101  
01000101011010101  
01101010101001100  
00101011101100111  
10101001010101010
```



计算机数据存储

- "计算机数据存储" 一词大致指的是集成电路,磁性或光盘和/或磁带设备用于计算机系统的记录和保留数字数据在一段时间内。
- 存储通常称为大容量存储—磁盘、可移动光盘、磁带盒和其他类型的介质包括:
 - 比 ram 慢很多 (随机存取存储器)
 - 比 ram 便宜得多
 - 专为永久保留数据而设计

存储分类

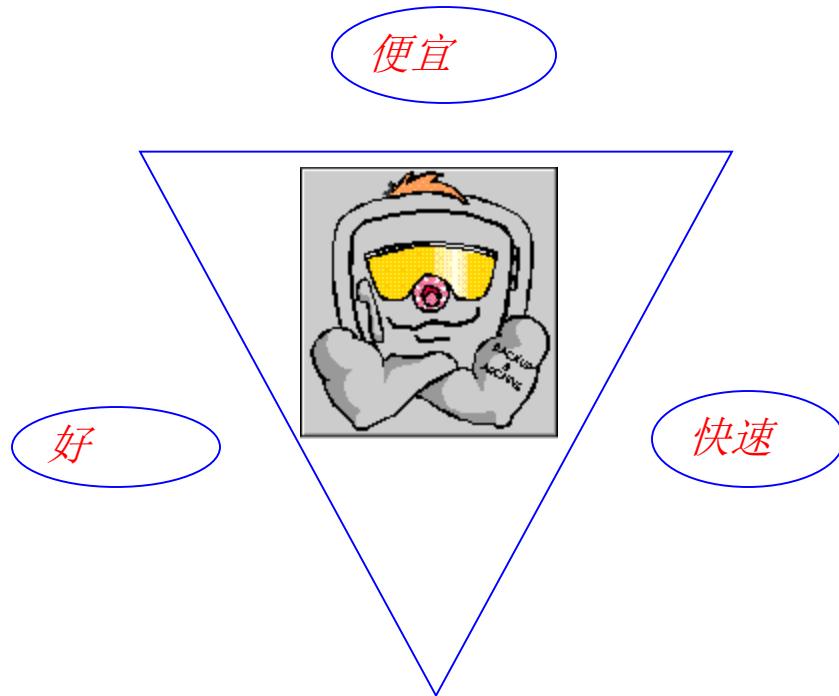
存储的特征包括分层层次结构, 或划分一级、二级、三级和离线存储或距离从中央处理器 (cpu).

- **主存储**是中央处理单元可以访问的内部存储器, 无需使用计算机的输入输出通道。
- **二次存储**是不直接连接到计算机中央处理单元的内存, 需要使用计算机的输入输出通道。 辅助存储用于保留不在活动中使用的数据。
 -
- **三级储存**或**三级内存**, 是由一个或多个存储驱动器和自动媒体库 (例如磁带库或光盘自动点唱机) 组成的计算机存储系统。



数据存储解决方案三角

- ❖ 数据存储在哪里？
- ❖ 如何访问数据？
- ❖ 如何保护数据？
- ❖ 如何管理数据？



哪一个**2** 三角形的两侧是否用于存储解决方案？

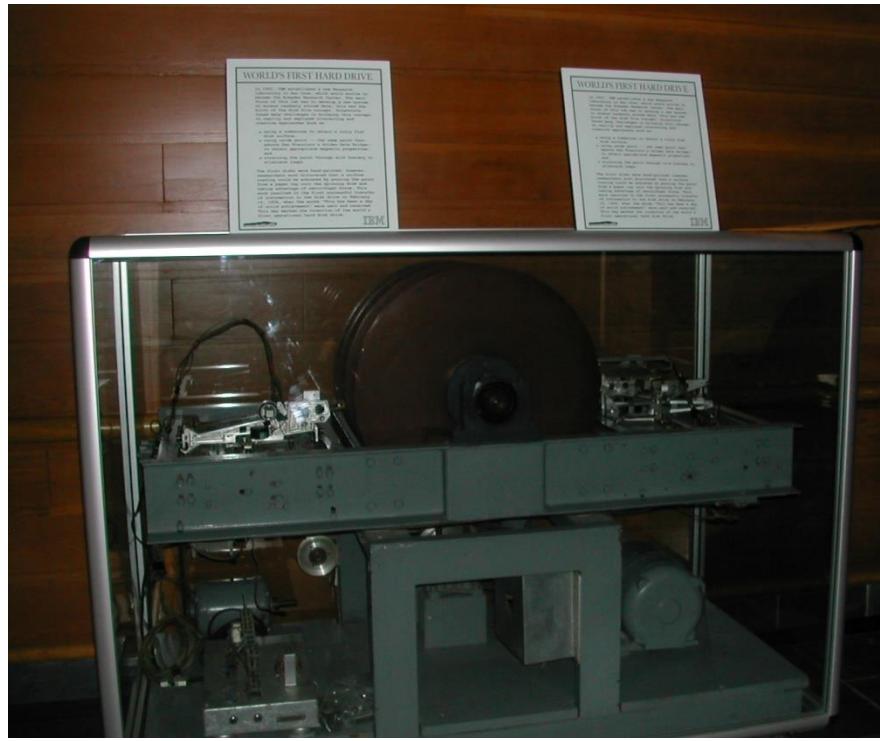
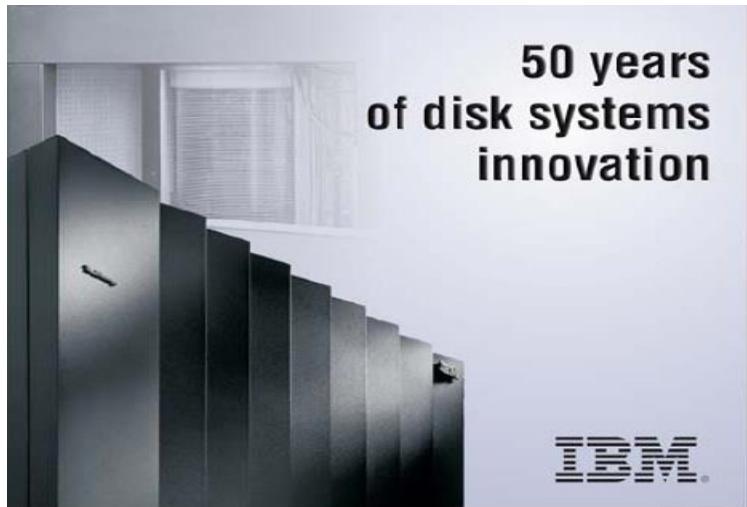
磁盘



- 在磁盘上, 数据被编码为微观磁化针在磁盘的表面。您可以记录和擦除磁盘上的数据的次数, 就像使用盒式磁带一样。磁盘有许多不同的形式:
 - **软盘**: 一个典型的5英寸软盘可以容纳360k 或 1.2 mb (兆字节)。3 1/2 英寸软盘通常存储720k、1.2 mb 或 1.44 mb 的数据。
 - **硬盘**: 硬盘可以存储在20mb 到200gb 以上的任何地方。硬盘也比软盘快10到100倍。
 - **可拆卸墨盒**: 可拆卸墨盒**硬盘**封装在金属或塑料墨盒中, 因此您可以像软盘一样取出它们。移动**墨盒**速度非常快, 尽管通常不如固定硬盘快。

世界上第一个硬盘驱动器

- ❖ ibm 350 磁盘存储单元
- ❖ 1955年-1961年
- ❖ 50 x 24 英寸
- ❖ 5 mb
- ❖ 0.0088 mbb



光盘



- 光盘通过用激光在磁盘表面燃烧微小的孔来记录数据。要读取磁盘, 另一个激光束在磁盘上发光, 并通过反射模式的变化来检测孔。
- 光盘有三种基本形式:
 - **光盘**: 大多数光盘都是只读的。当您购买它们时, 它们已经充满了数据。您可以从 **cd-rom** 读取数据, 但不能修改、删除或写入新数据。
 - **蠕虫**: 代表写一次, 读很多。 **蠕虫** 磁盘可以写一次, 然后读取任意次数;但是, 您需要一个特殊的 **worm** 磁盘驱动器将数据写入 **worm** 磁盘。
 - **可擦除光学(eo)**: **eo** 磁盘可以像磁盘一样读取、写入和擦除。

固态硬盘 (ssd)



a 个固态硬盘-电气、机械和软件与传统 (磁性) 硬盘兼容。

- 不同之处在于, 存储介质不是磁性的 (如硬盘) 或光学的 (如 cd), 而是**固态半导体**如电池支持的 ram、eprom 或其他可电擦除的 ram, 如闪存芯片。

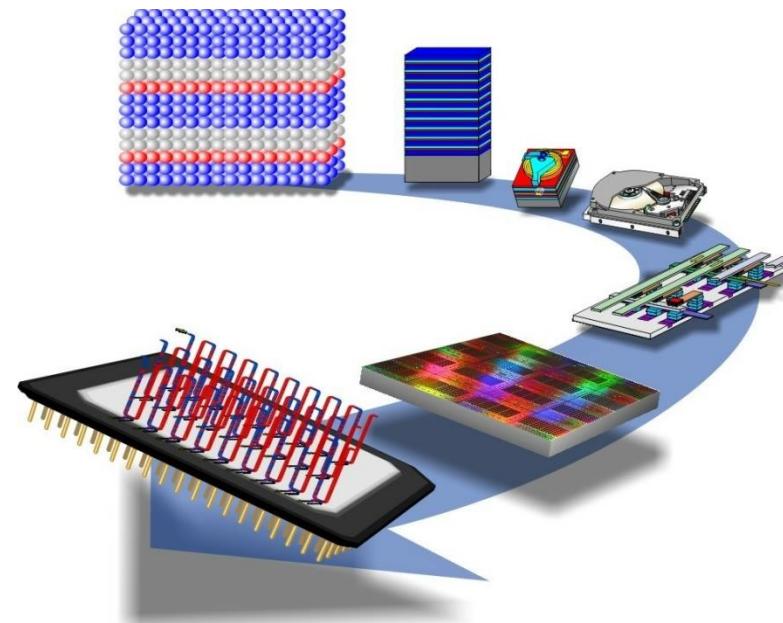
这提供了**更快** 访问时间比硬盘, 因为 **ssd** 数据可以在同一时间随机访问存储位置。固态硬盘访问时间不依赖于与旋转磁盘上的数据扇区同步的 **read\ write** 接口头。

- 固态硬盘还提供了更大的物理恢复能力, 以应对物理振动、冲击和极端温度波动。唯一的缺点是**每兆字节的更高成本**存储-尽管在某些应用中, 固态硬盘的可靠性较高, 因此它们的拥有成本比更换多个故障硬盘更便宜。此外, 在企业服务器加速应用中, **ssd** 的优点是, 与单独使用基于硬盘的 **raid** 相比, 它减少了所需的服务器数量。

赛马场记忆 ←ibm 的新发明

- 使用电子的 "自旋" 来存储数据
- 无移动部件 (硬盘)
- 无磨损机构 (ssd)

例如, 这项技术可以使 mp3 播放器等手持设备能够以更低的成本和功耗存储大约 500, 000首歌曲或大约 3, 500 部电影--比今天高出100倍。



磁带库

- 在计算机存储中,磁带库是一个存储设备,其中包含一个或多个磁带驱动器,多个插槽可容纳磁带盒(a)条形码读取器识别磁带盒和自动加载磁带的方法 (a机器人).
- 最早的例子之一是1974年宣布的 ibm 3850 大容量存储系统 (mss)。



网络存储

- 涉及通过计算机网络访问信息的任何类型的计算机存储。
- 网络存储的示例包括:
 - 网络连接存储是连接到计算机的辅助或三级存储, 另一台计算机可以通过局域网、专用广域网或在线文件存储通过 **internet** 访问该存储。
 - 网络计算机是不包含内部辅助存储设备的计算机。相反, 文档和其他数据存储在网络连接的存储上。

SAN, NAS

互联网 scsi—iscsi

iscsi是互联网 scsi (小型计算机系统接口), 这是一种基于互联网协议 (ip) 的存储网络标准, 用于连接数据存储设施, 由互联网工程任务组 (ietf) 开发。

通过通过 ip 网络传输 scsi 命令, iscsi 用于促进内联网上的数据传输, 并管理远距离存储。

iscsi 协议是通过提高存储数据传输的能力和性能, 有望帮助实现存储区域网络 (san) 市场快速发展的关键技术之一。由于 ip 网络无处不在, iscsi 可用于通过局域网 (lan)、广域网 (wan) 或 internet 传输数据, 并可启用与位置无关的数据存储和检索。

光纤通道和 **iscsi san** 历史记录

- 光纤通道技术自20世纪80年代中期以来一直存在，并在1994年获得批准。
- 光纤通道对 scsi 就像 tcp 对 ip 一样。它是用于在主机启动器和存储目标之间传递 scsi 命令的传输机制。
- 随着 **dot. com** 的信息爆炸，光纤通道 **san** 存储的普及在20世纪90年代中后期开始流行。当时，对于一个充满计算机的数据中心来说，访问集中式存储确实是唯一可行的方式。
- 光纤通道 **san** 需要一个专用的光纤存储网络，服务器中安装了特定于存储的网络交换机和主机总线适配器 (**hba**) 卡。从一开始，光纤通道 **san** 就被认为是非常昂贵而复杂的。

raid 配置

Raid, 代表**廉价磁盘的冗余阵列**, (后来称为独立磁盘的冗余阵列)是一个系统, 它使用多个硬盘驱动器显示为服务器的单个逻辑磁盘。

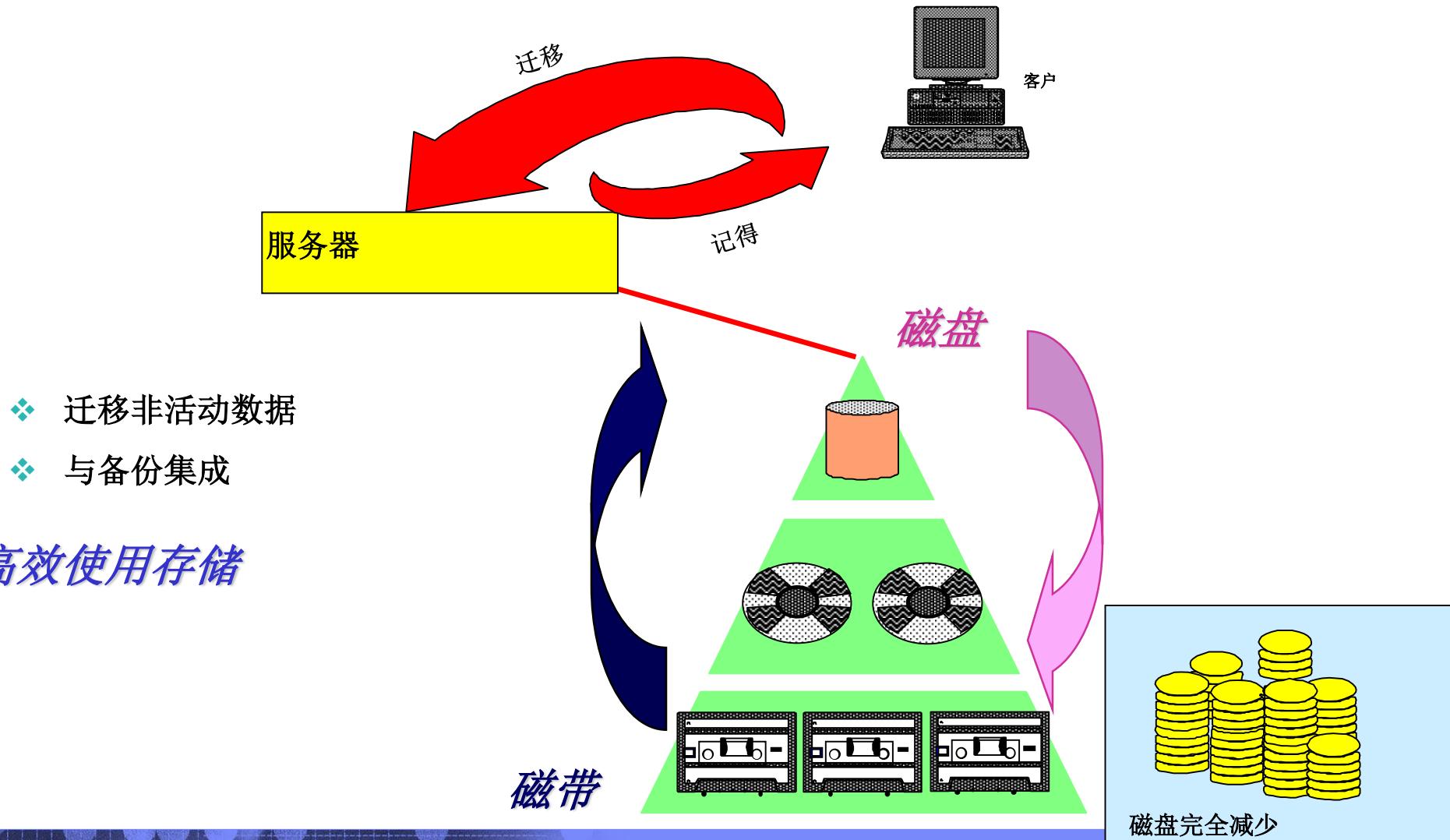
最常用的 raid 级别包括:

- raid 0: 条带, 无需数据保护 (jbod—只是一组磁盘)
- raid 1: 镜像 (100% 冗余)
- raid 3:2008 条带数据 (专用奇偶校验磁盘)
- **raid 5: 条带化 (奇偶校验均匀分布在磁盘上)**

常见的嵌套 raid 级别:

- raid 6: 带奇偶校验块分布在两个磁盘上的条带数据
- raid 01: 一组镜像的带区磁盘
- raid 10: 一组带区的镜像磁盘
- raid 30: 专用奇偶校验 raid 系统中的条带
- raid 100: 镜像条纹

存储空间管理— 分层存储管理 (hsm)



基本存储管理操作

- 备份
- 数据备份
- 数据恢复
- 灾难恢复
- 磁盘镜像
- 复制
- 归档

数据备份类型

- 快照备份
- 完整备份
- 差异化备份
- 增量备份
- 连续备份 (连续数据保护)
- 磁盘镜像

备份问题

- 备份窗口
- 恢复时间
- 保留时间
- 应用程序/数据库状态
- 备份资源
- 打开文件备份
- 数据验证

重复数据删除

- 重复数据删除一般指的是消除重复或冗余的信息。
- **成本效益**: 降低初始存储购置成本, 或延长存储容量升级之间的间隔。
- **管理效益**: 每个存储单元存储 "更多" 数据的能力, 或将在线数据保留更长时间的能力。

软件定义的存储 (sds)

- 计算机数据存储软件管理的一个不断发展的概念基于策略独立于硬件的数据存储的配置和管理。
- 软件定义的存储定义通常包括一种形式的存储虚拟化将存储硬件与管理存储基础结构的软件分开。
- 启用软件定义的存储环境的软件还可以为功能选项提供策略管理, 如**重复数据删除**、**复制**、**精简资源调配**、**快照**和**备份**。

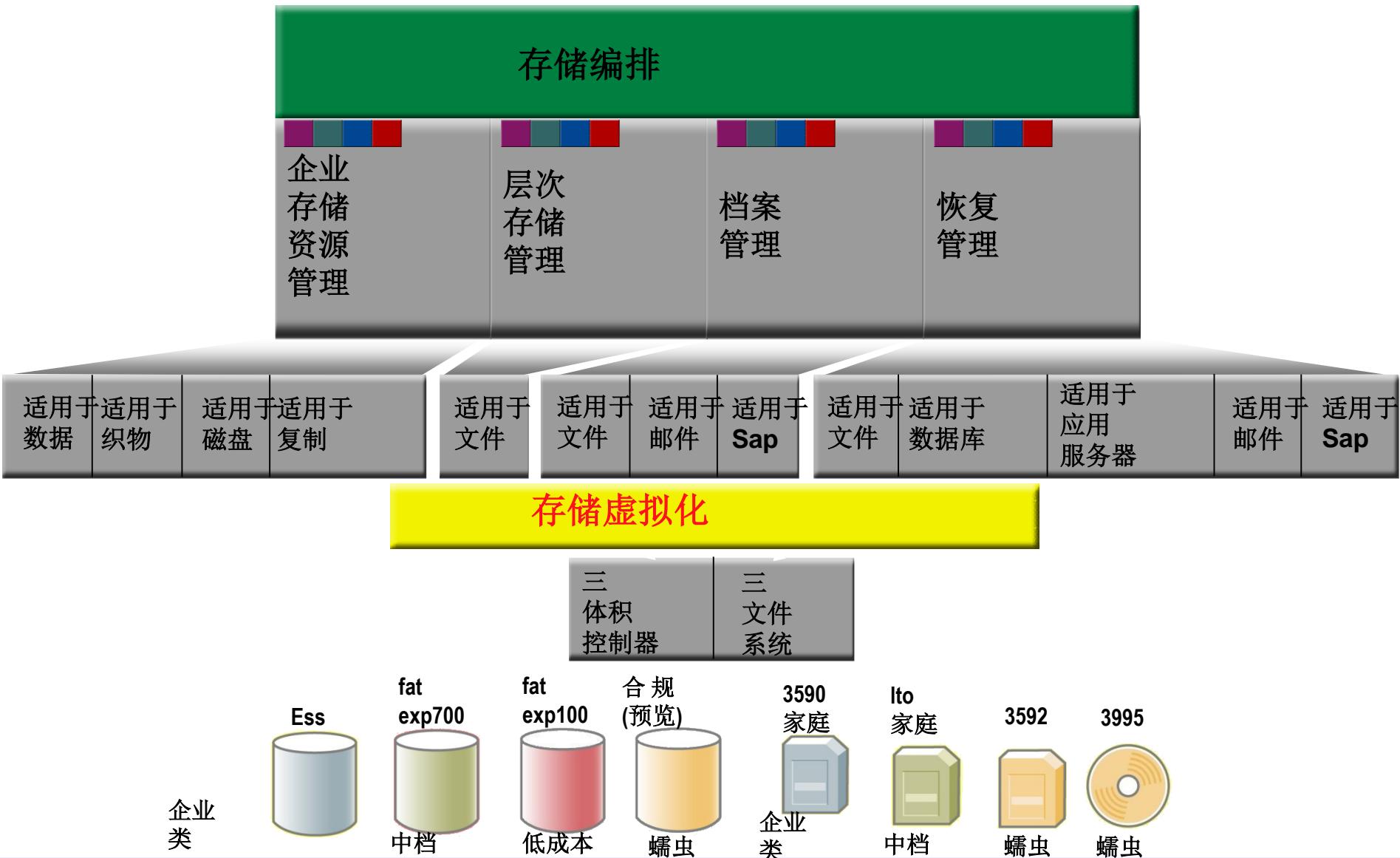
http://en.wikipedia.org/wiki/Software-defined_storage

存储管理产品

- 存储虚拟化
 - 存储 **svc (san 卷控制器)**
 - **Vmware**
 - **maid (空闲磁盘的大数组)**
- 数据管理
 - 主数据管理解决方案
 - 数据清理解决方案
 - 数据发现



ibm 存储软件产品



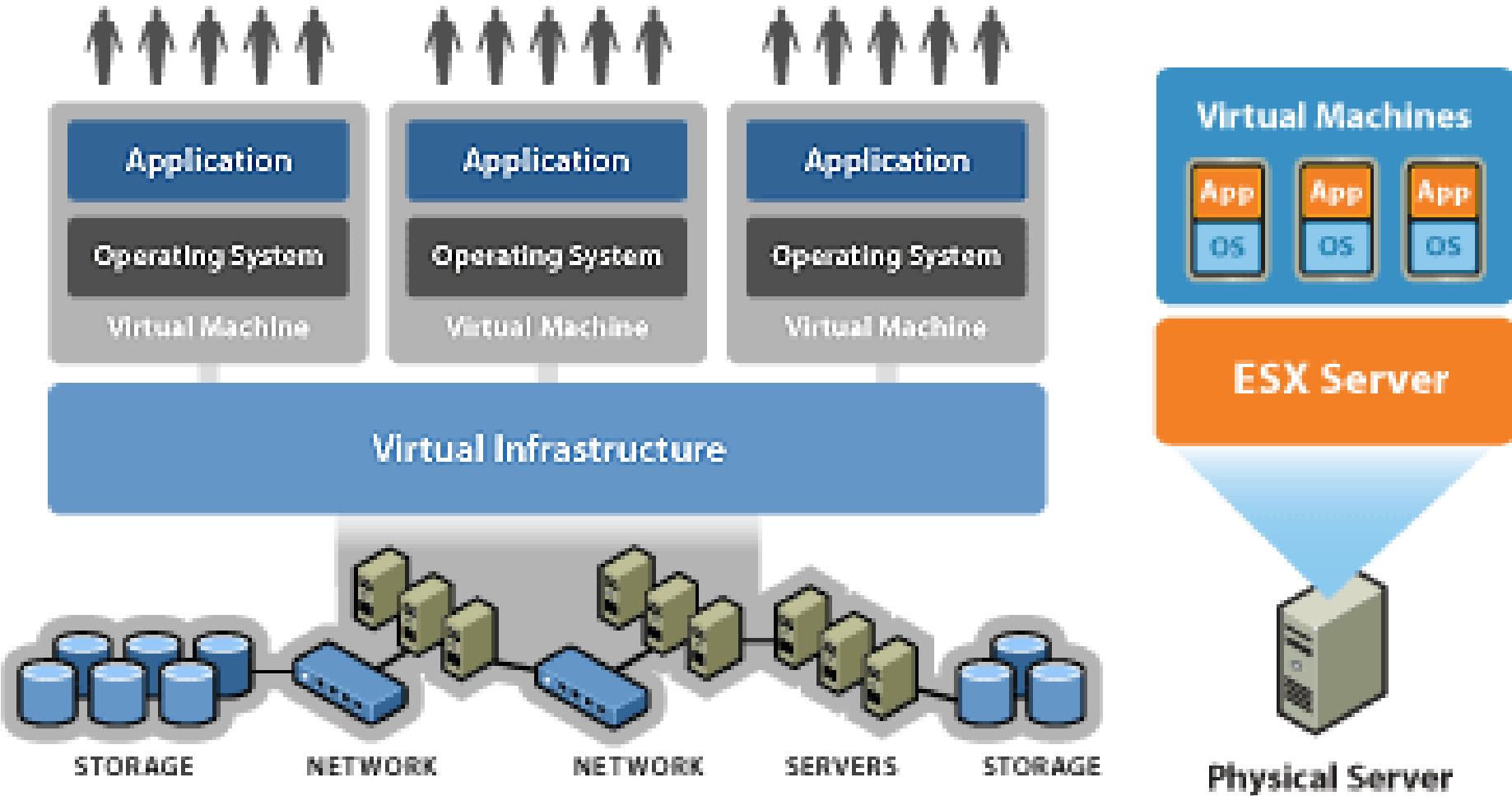
ibm 系统存储 san 卷控制器

- ibm 系统存储™san 卷控制器 (svc) 支持针对不同的异构存储资源的单点控制, 以帮助支持提高业务应用程序可用性和更高的资源利用率。目标是识别 it 基础架构中的所有存储资源, 并确保它们能够为您的业务带来优势, 并快速、高效、实时地执行, 同时避免管理成本。



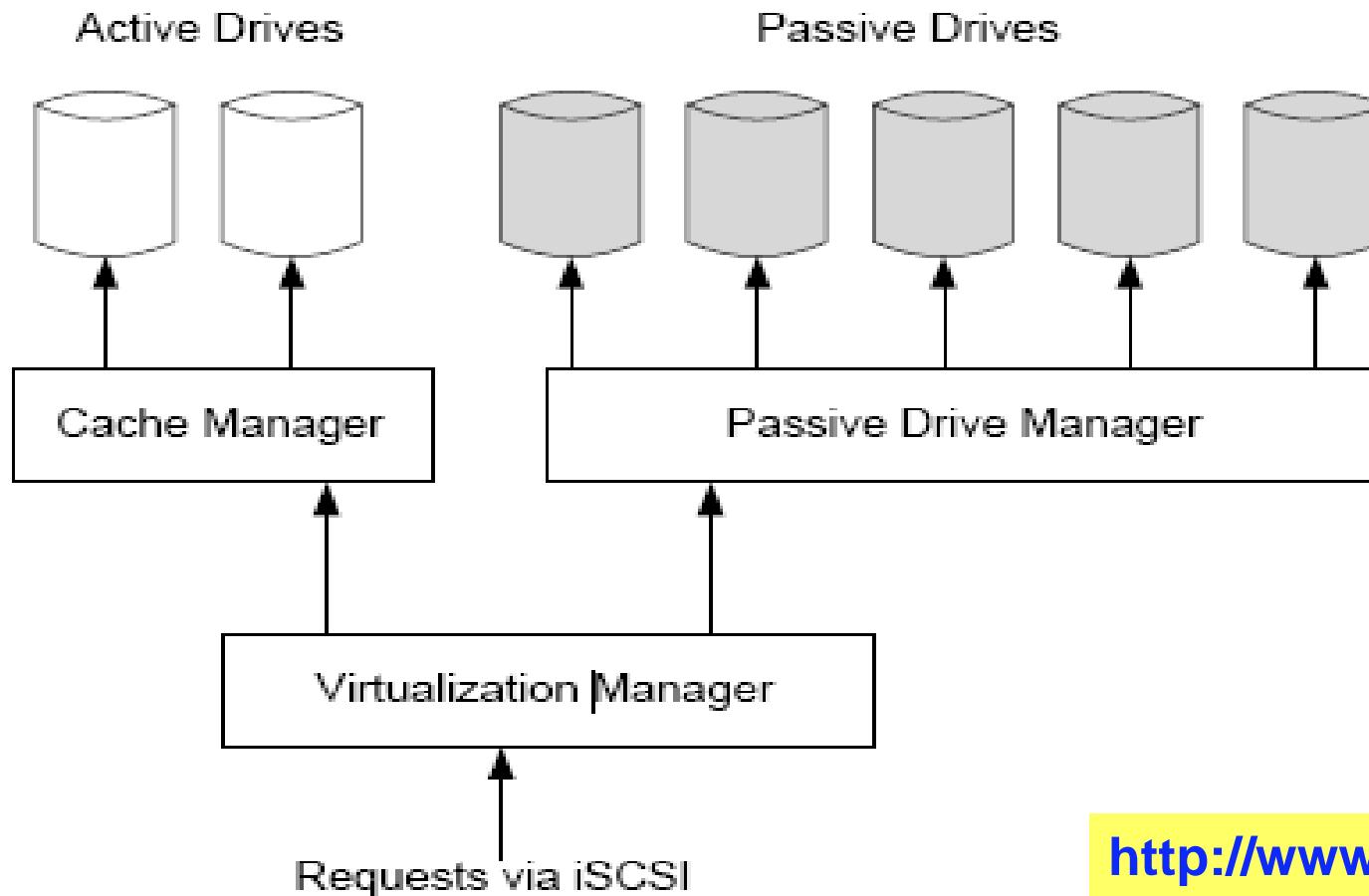
Vmware

vmware 虚拟化方法插入了薄薄的软件层直接在计算机硬件或主机操作系统上。



maid (空闲磁盘的大数组)

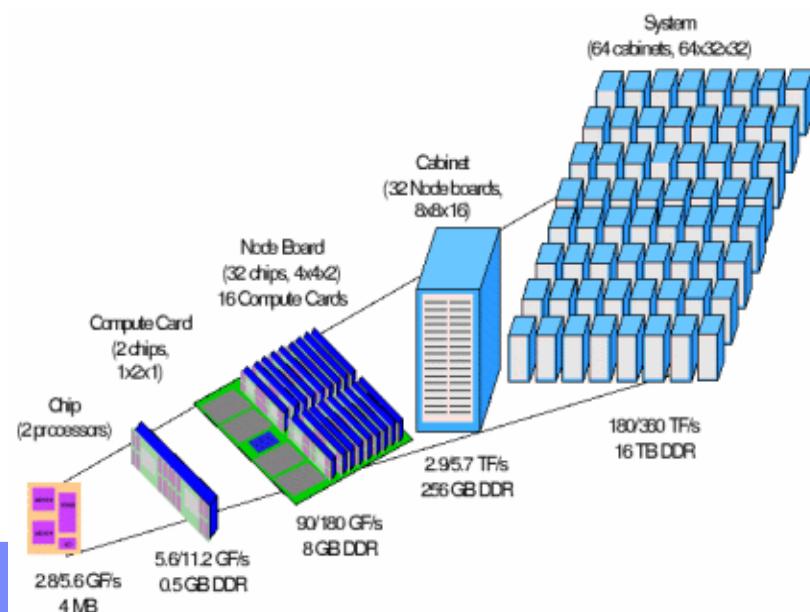
- ❖ 一种存储技术，它使用大量磁盘驱动器，其中只有那些正在使用的驱动器在任何给定时间旋转。这降低功耗延长驱动器的使用寿命。



<http://www.copansys.com/>

存储管理产品

- 存储虚拟化
 - 存储 svc (san 卷控制器)
 - Vmware
 - 网络设备
 - maid (空闲磁盘的大数组)
- 数据管理
 - 主数据管理解决方案
 - 数据清理解决方案
 - 数据发现



数据管理

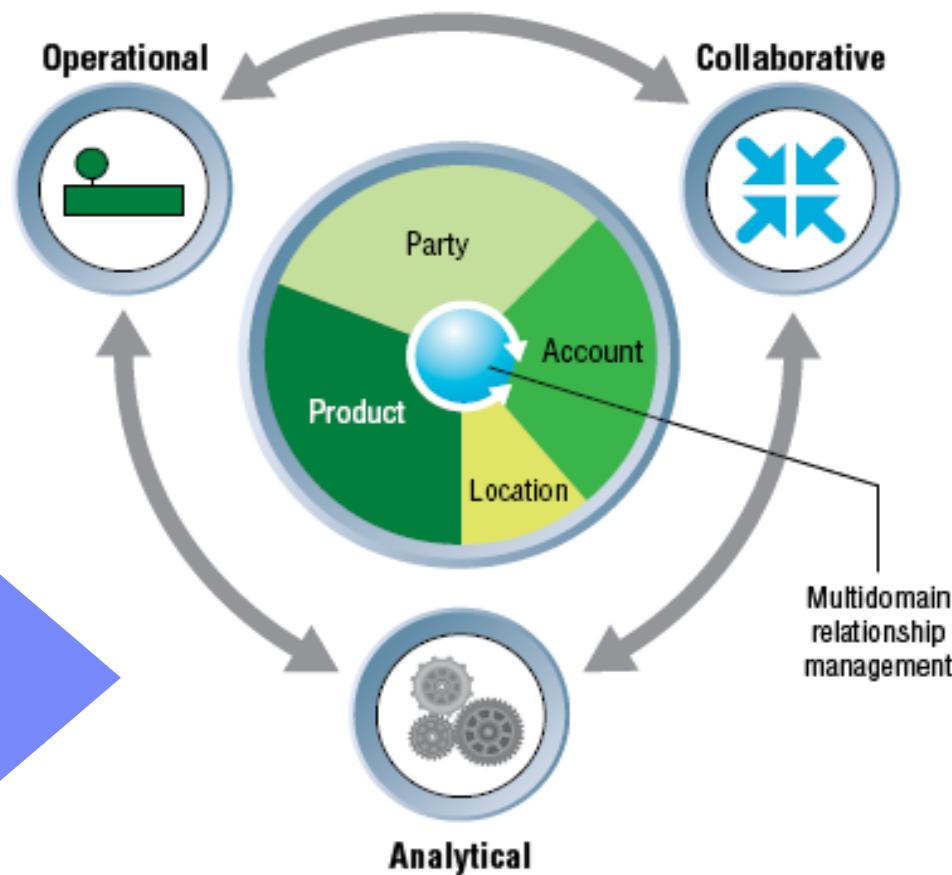
数据管理包括与将数据作为宝贵资源进行管理相关的所有学科。官方定义是 "数据资源管理是开发和执行体系结构、策略、实践和过程，以正确管理企业的整个数据生命周期需求"。这一定义相当宽泛，包括一些可能与数据管理的较低级别方面（如关系数据库管理）没有直接技术联系的专业。

- 主数据管理解决方案
- 数据清理解决方案
- 数据发现

主数据管理解决方案

- 主数据管理通过集中多个域并提供使用组织所需的多个主数据使用样式的功能，帮助组织实现业务价值。
- 主数据（有关公司客户、产品和帐户的关键业务信息）的有效管理不仅可以帮助组织提高收入收益，还可以帮助降低成本、提高战略灵活性和降低风险。

主数据可用于实现协作、操作或分析优势

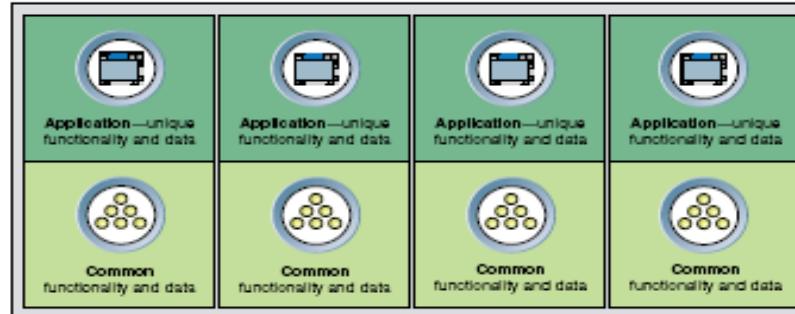


ibm 多格式主数据管理

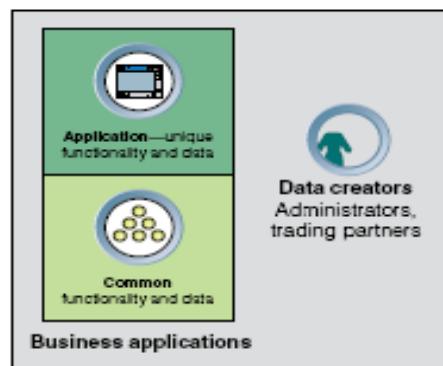
IBM Multiform MDM is characterized by multiple users and multiple usages



Operational data usage

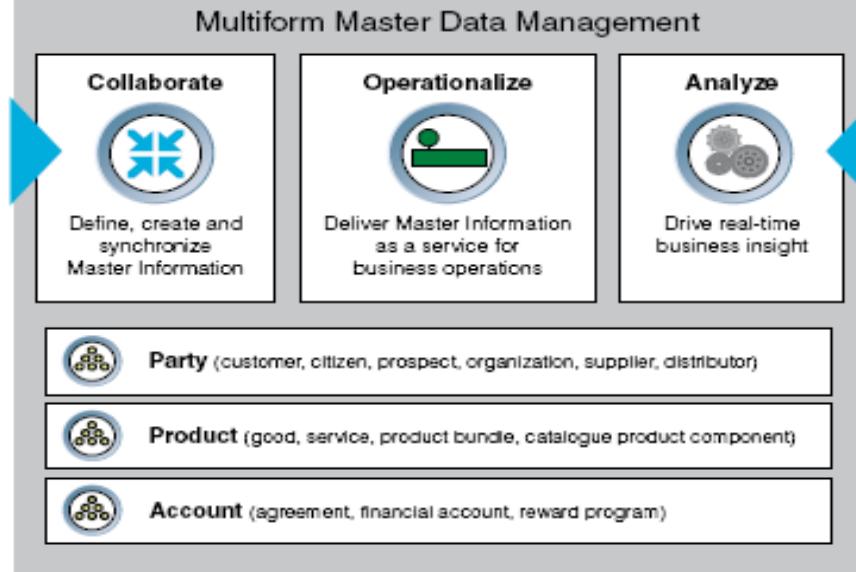


Collaborative data usage

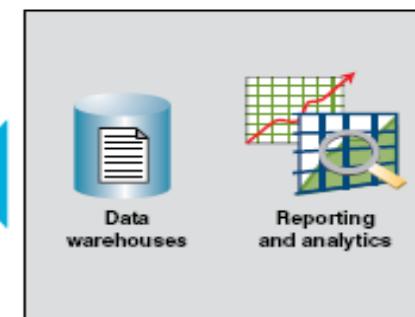


Define and create master data

Real-time SOA data access



Analytical data usage



Analyze master data

数据清理解决方案

如果需要执行以下一项或多项操作，则应考虑数据清理解决方案：

- 由于数据丢失、重复或不准确而清理数据库
- 避免从市场营销列表中导入重复项
- 集成或交叉引用不同的应用程序数据库
- 合并多个数据库
- 匹配任何 `odbc` 连接的数据库中的数据，以提高和维护数据质量

数据发现

- 由于在过去20年中创建了超过80% 的电子数据, 今天的大部
分信息可能永远不会离开它原来的格式。
- 公司和法律部门电子邮件、文字文档、电子表格和其他电子
文件中存储的数据的收集、处理、本机文件审阅和生成可能会
很快变得不堪重负。
- 然而, 任何人都不能忽视这些媒体对案件结果的重视。
- 由于这些趋势, **电子发现**已成为彻底诉讼的重要组成部分。

ibm 数据发现和查询生成器

ibm 数据发现和查询生成器**医疗保健和生命科学**是一个基于 web 的框架, 可帮助医生和研究人员制定和运行数据库搜索, 以识别和关联人口统计、实验室测试和诊断等患者数据, 以及实验室和医生的笔记。

这个功能强大的工具使具有不同级别专业知识的用户能够轻松地利用结构化和非结构化文本数据形成查询。. 用户可以:

- 聚合和查询数据, 并为自己或其他人保存和修改搜索。
- 将搜索标记为公共搜索, 以允许其他人运行相同的查询或将其修改为单独使用。
- 将查询标记为私有查询, 使其只能由开发人员或选定的团队成员访问。

软件应用程序简化了复杂的查询, 并以几种常见格式之一返回数据结果, 可以对其进行排序和排序以进行观察。该应用程序可以在多个操作系统和硬件平台上运行。

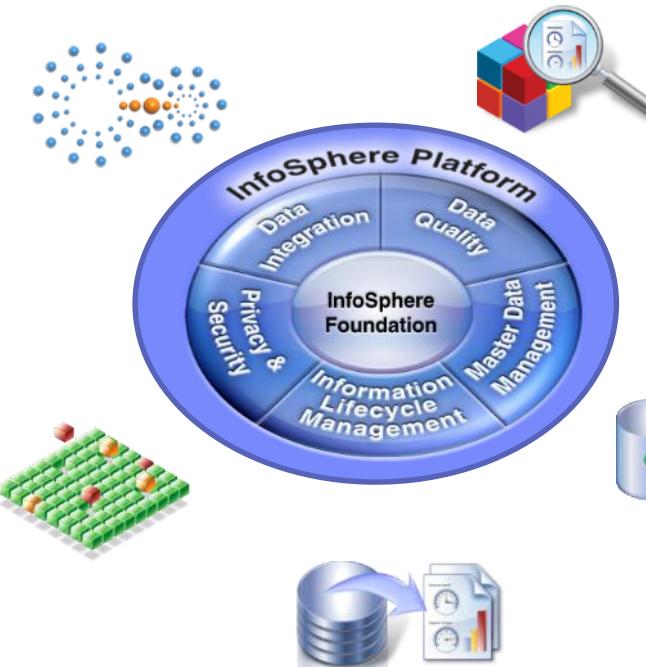
大数据集成和数据治理

整合和连接大数据

- 大数据作为一个源
- 大数据为目标
- 数据转换
- 数据移动
- 整合现有企业
- 数据谱系和影响分析
- 元数据集成
- 实时和数据联盟

保护大数据

- 活动监控
- 数据屏蔽
- 数据加密
- 按需/就地保护
- 在线保护 (w/etl 等)
- 主动检测和报警



审核和归档大数据

- 可查询的存档
- 结构化和半结构化
- 现有应用的优化连接器
- 热光飞行
- 不可变和安全的访问
- 数据冻结的自动法律保留能力

清理和验证大数据

- 准确性和实体与社会数据的匹配
- 重复数据消除和标准化
- 集成的在线清洁
- 受信任的数据仪表板和数据质量报告

掌握大数据

- 作为供应商的大数据
- 作为消费者的大数据
- 大数据和可信的黄金记录之间的链接
- 在大数据分析中利用主数据
- 极端扩展级别下的实体解析
- 概率实体匹配

新兴合作领域： 云和弹性数据服务



- 参与
- 移动
- 快速变化
- 动态
- 寿命短
- 竞争
- 企业系统网关
- 时尚

大数据存储在未来

需要能够处理容量并为分析工作提供低延迟

- **横向扩展或群集 nas.**这是文件访问共享存储, 可以扩展以满足容量或增加的计算要求, 并使用并行文件系统, 这些文件系统分布在多个存储节点上, 可以处理数十亿的文件, 而不会降低性能。随着文件系统的增长而发生。
- **对象存储.**这解决了与横向扩展 nas 相同的挑战--传统的树状文件系统在包含大量文件时变得难以操作。基于对象的存储通过为每个文件提供唯一标识符并为数据及其位置编制索引来解决此问题。这更像是在互联网上做事的 dns 方式, 而不是我们习惯的那种文件系统。
 - 对象存储系统可以扩展到非常高的容量和大量的文件, 因此对于希望利用大数据的企业来说也是另一种选择。话虽如此, 但与横向扩展 nas 相比, 对象存储是一种不太成熟的技术。

原子级别的大数据存储

- 随着大数据存储需求呈指数级增长, 用于存储数据的技术必须缩小到原子级别, 以确保其易于访问和可用。
- 未来的存储系统基于**原子级存储器**将能够存储海量的大数据。例如, 每一部电影都可以存储在指甲大小的原子级存储设备上。存储如此庞大的数据池的能力可以改变企业、工业、政府和社会每天使用数据的方式。
- 在创作这部电影的时候, ibm 的科学家们希望世界上最小的电影能激励年轻人学习科学, 从事科学和技术的职业。

原子存储器

2012年, ibm 研究科学家仅使用12个原子创建了世界上最小的磁性存储器位。

仍在试验中**原子级磁铁存储器**密度是今天硬盘驱动器和固态内存芯片的100倍。

今天, 它需要一百万个原子 存储单个数据位, 但 ibm 的科学家们正在研究原子规模的内存技术, 可以只使用1个原子存储一些数据.

一部关于大数据未来的小电影

这部名为《男孩和他的原子》的电影被吉尼斯世界纪录认证为世界上最小的电影。



ibm 的科学家们将大屏幕缩小到原子水平,通过使用独特的两吨显微镜精心移动原子,创造了一部创造纪录的电影。除了讲述一个由原子制成的男孩的故事外,这部电影还通过超过300万次的 youtube 浏览,提高了人们对 ibm 存储突破的认识,这些突破可能会扩大大数据的使用。

内容



未来, 请参见: **ibm 研究-全球技术展望**

休息一下!

