



云计算技术及应用

莫同

motong@ss.pku.edu.cn



北京大学



第1讲 云计算概述

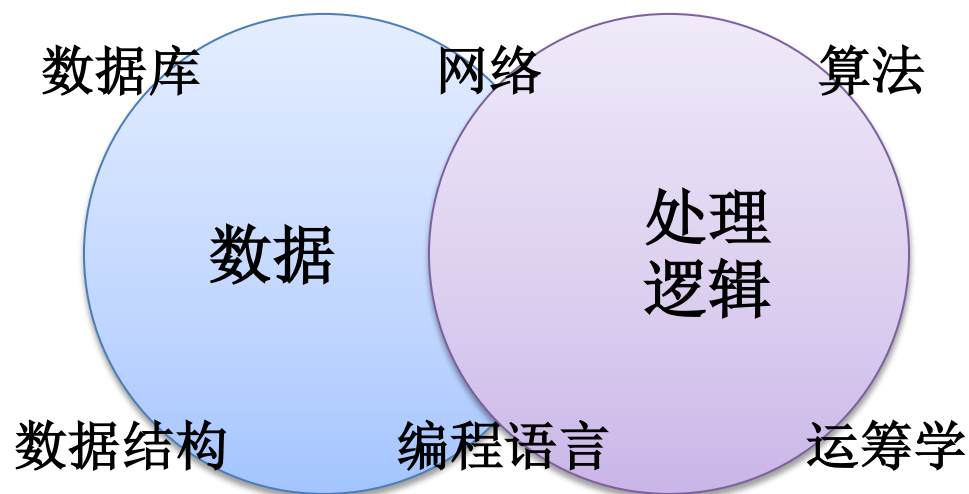
莫同

motong@ss.pku.edu.cn



北京大学

- 什么是计算机？
- 计算机有什么用？
- 如何理解IT系统？



- 当数据量大了之后怎么办？



- 图书管理系统/学生管理系统

数据表中的数据越来越多

用户越来越多

查询出来的结果过大

读写操作的区别

...

如何做一个真正可用的系统？



- 小测验：你的计算机基础知识掌握的怎么样？

1、点击快捷方式之后到内容出现之间发生了什么

2、输入链接地址之后到网页刷新之间发生了什么

4、从计算机角度，如何证明“你”是“你”

3、微信朋友发一条消息到你看到这条消息之间发生了什么



- 安全问题
 - 棱镜门
 - CSDN泄密门
 - 酒店开房记录泄露事件
 - 好莱坞影星私密照泄露事件



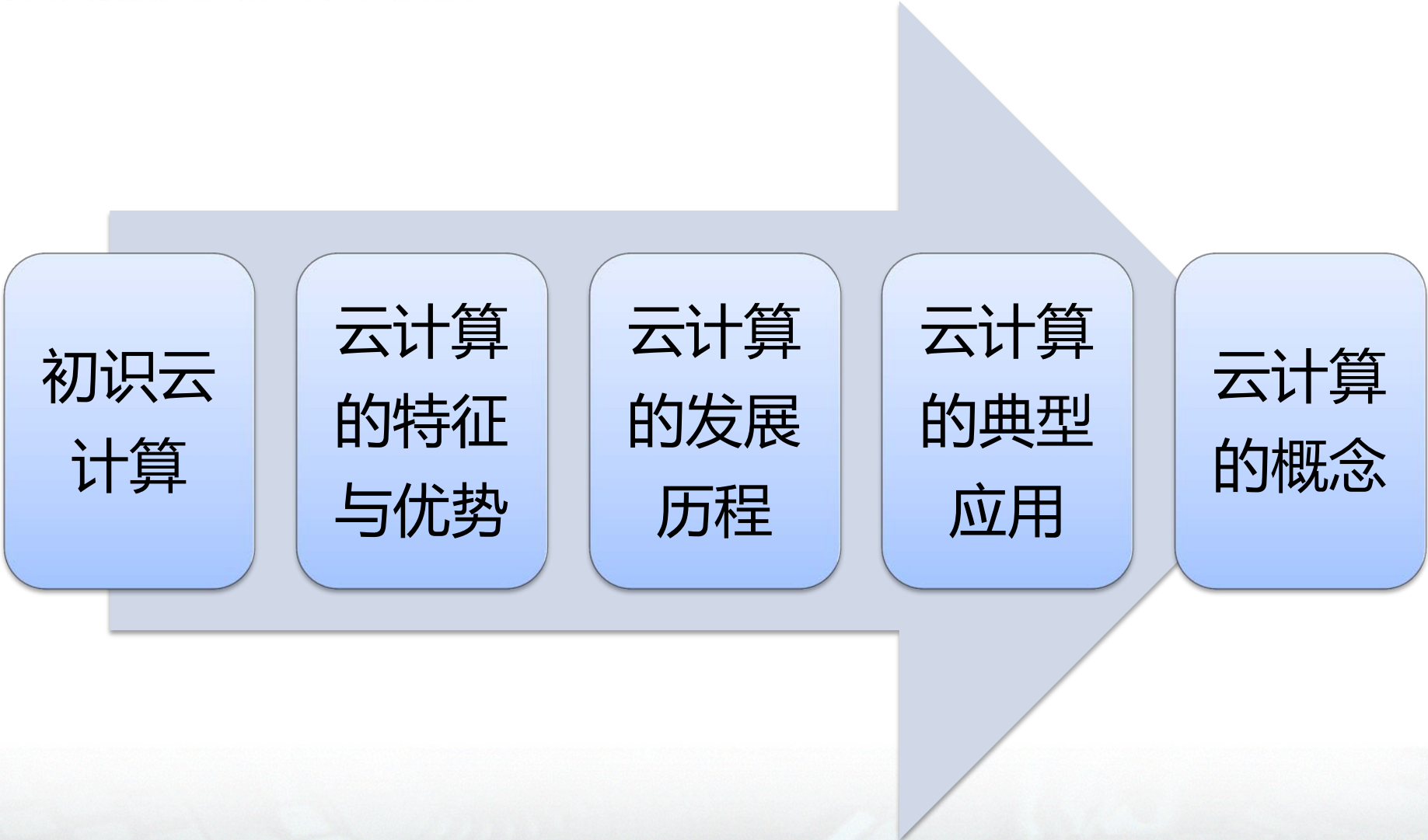
- 程序员的出路在何方？
 - 无休止的代码
 - 功能实现
 - 调试bug
 - 测试
 - 提升效率
- 架构：机制的设计
- 思考：使用SSH方式开发学生管理系统，如何解决数据增长问题？



- 如何来定位云计算？
 - 全新的技术理念？
 - 已有技术的进一步发展？
- 如何来定位云计算课？
 - 分布式并行前提下的计算机课？
 - 处理问题的思想的转变？



内容提要



初识云计算

云计算的特征与优势

云计算的发展历程

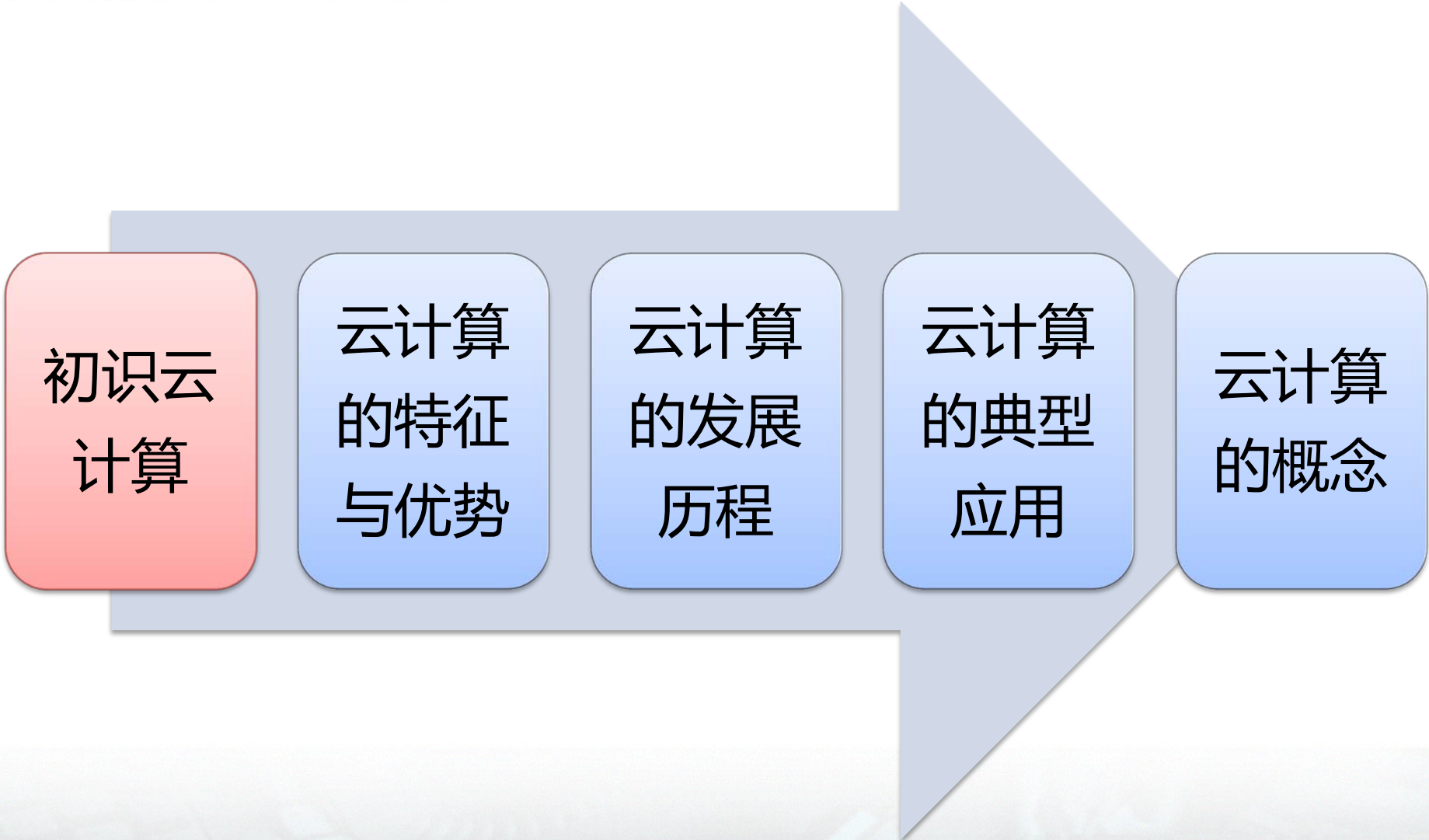
云计算的典型应用

云计算的概念



北京大学

内容提要



初识云计算

云计算
的特征
与优势

云计算
的发展
历程

云计算
的典型
应用

云计算
的概念

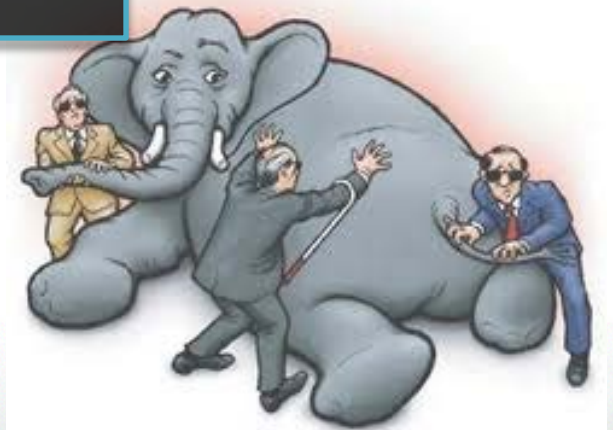


北京大学

- 什么是云计算？

- 5 people might have 8 answers !

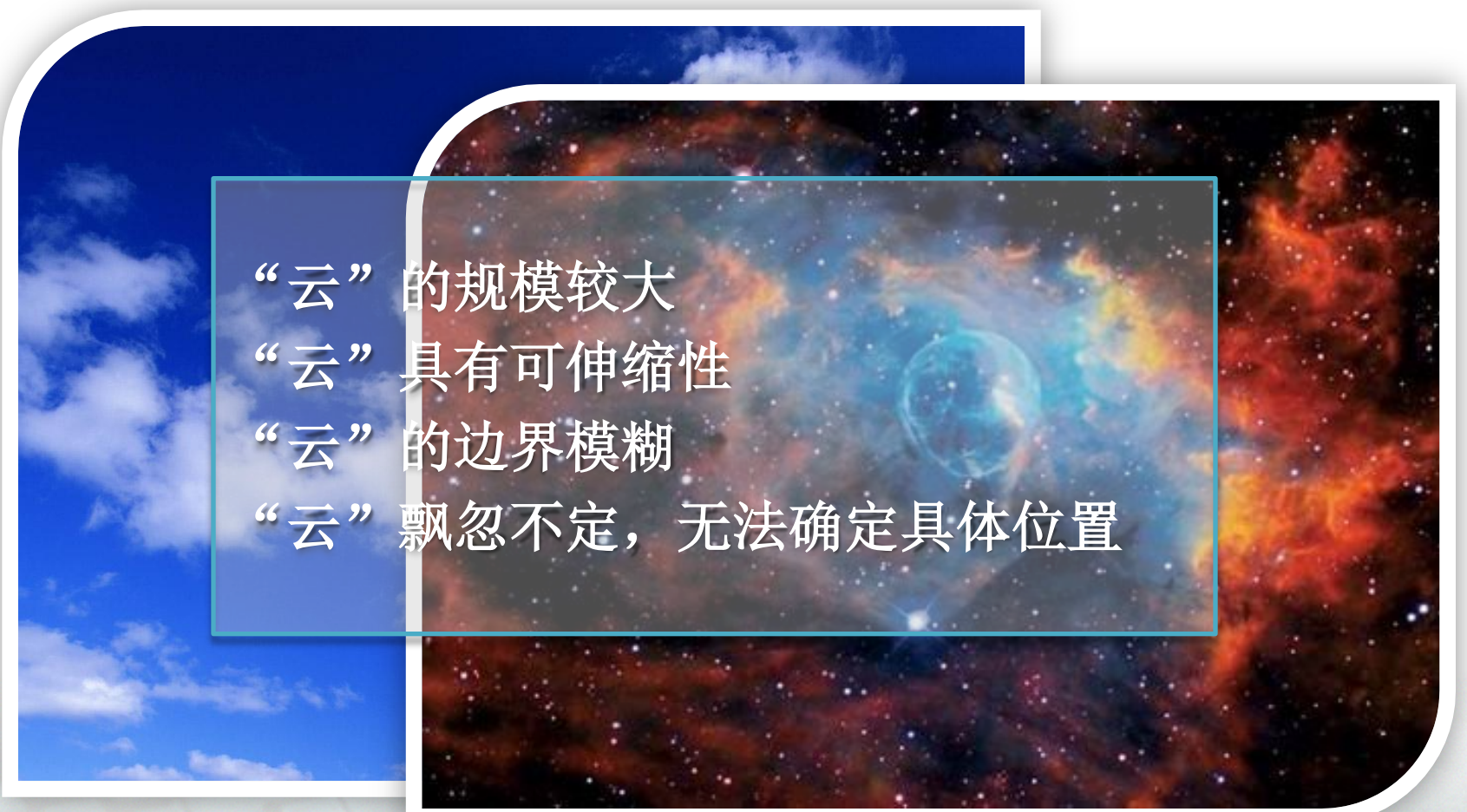
- 分布式计算
 - 并行计算
 - 效用计算
 - 网格计算
 - 大规模存储 一夜春风，满天xx云
 - 海量数据处理
 - XaaS (IaaS/PaaS/SaaS...)
 - 云电视/云手机/云服务/云xx...



- 云计算至今仍缺乏统一规范明确的定义
- 云计算为什么叫云计算？
 - 云计算=云+计算



— 计算很好理解，那么什么是云？

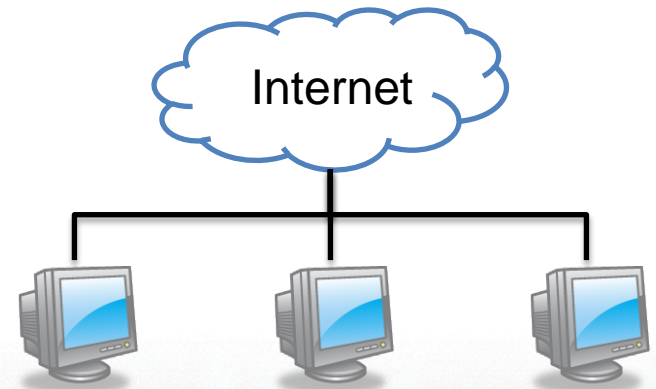


“云”的规模较大
“云”具有可伸缩性
“云”的边界模糊
“云”飘忽不定，无法确定具体位置



初识云计算

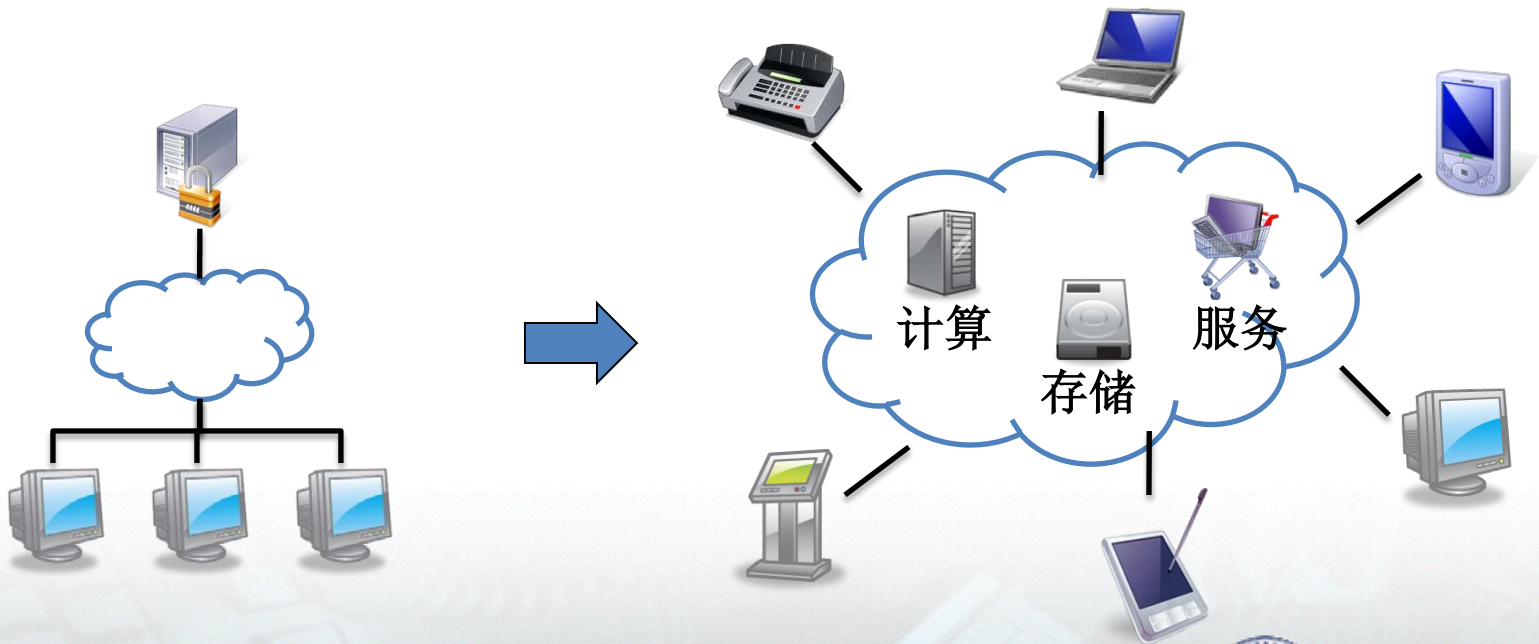
- 起初，人们更加关心服务器端和客户端，网络部分对用户透明，因此在绘制互联网示意图时，习惯用一朵云来抽象表示互联网



北京大学

初识云计算

- 随着互联网发展，接入设备种类增多，网络的作用除了接入、路由，还包括计算、存储、服务等元素，云变得充实、具体化、复杂。



北京大学

初识云计算

- 如何理解云计算？
 - Google
 - 12306
 - VMware
 - GAE/SAE



北京大学

- Google

- 抓取互联网数据
 - 网页、图片...
- 数据线下分析处理
 - 网页解析、分类消重合并、关键词提取...
- 信息存储
 - 高性能大容量数据存储、数据组织方式
- 在线搜索
 - 搜索算法、内容展现



北京大学

- Google

- 抓取的数据量巨大
- 海量数据的分析处理
- 海量数据的高效存取
- 如此大规模的数据处理，没有云计算的支持很难实现



北京大学

- Google

- 进一步的思考

- 互联网上有多少公共数据？
 - 这些数据包含了多少有用的信息
 - 得不到我们想要的信息的原因是什么？
 - 能力不够？
 - 真的不够吗？

- 云计算为我们提供了“无限”的廉价分析能力

- 封装成服务，提供分析处理能力



北京大学

初识云计算

- www.12306.cn
 - 迅速跻身世界网站点击排名100强
 - 史上最牛电子商务网站
 - 大量用户同时登录造成网站瘫痪
 - 问题何在？如何解决？



北京大学

- www.12306.cn
 - 软件架构
 - 硬件投资
 - 适应多少用户的访问？
 - 1.3亿峰值vs百万常态？
 - 适应峰值必然造成大量浪费
 - 适应常态无法应对峰值
 - 有没有一种弹性的解决方案？

是否似曾相识？
电子商务网站？



- www.12306.cn
 - 12306：火车票电子商务网站
 - 12306vs亚马逊
 - 亚马逊将剩余硬件资源封装成服务对外提供
 - 亚马逊也是著名的云服务提供商之一

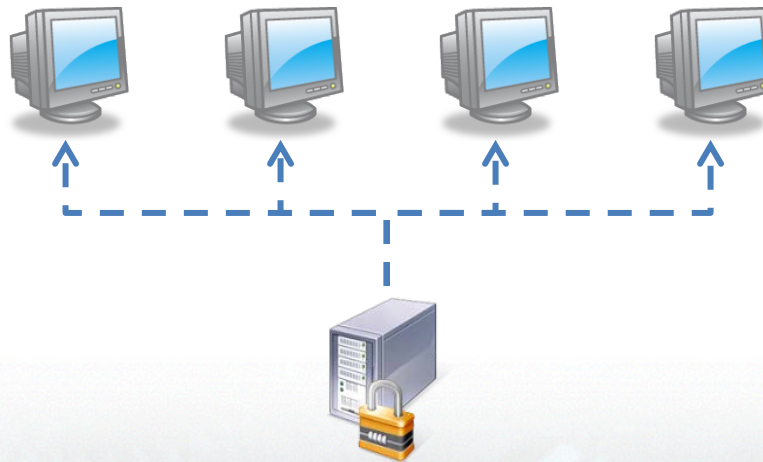
笑谈：铁道部到淘宝上开个店吧~



- VMware
 - 如何快速建立计算环境
 - 操作系统
 - 应用软件
 - GHOST ?



- VMware
 - 虚拟机
 - 一台高性能机器上虚拟出多台机器



- VMware
 - 虚拟机
 - 进一步扩展
 - 多台机器上虚拟出多台机器



- VMware
 - 动态的管理硬件资源
 - 完整的IT解决方案（硬件/软件）
 - 无需繁琐的配置管理
 - 便捷高效的资源分配
 - 封装成服务提供



- GAE/SAE
 - 如何快速的支持应用的开发部署
 - 应用开发
 - 应用部署
 - 申请域名
 - 购买服务器
 - 安装调试软件
 - 系统维护
 - 计算能力扩展
 - ...



- GAE/SAE
 - 有没有好的解决方案？
 - 专注于开发，而不是部署
 - 联想一下淘宝
 - 开一个新的店
 - 得到新的URL
 - 网站运维交给淘宝
 - 专注于内容

应用开发需要一个类似的平台





初识云计算

- GAE/SAE
 - GAE——Google App Engine
 - SAE——Sina App Engine



北京大学

• 四个案例

Google

- 搜索引擎——大规模数据存储与分析处理

12306

- 电子商务——资源的弹性配置管理

Vmware

- 镜像——IT资源管理与快速配置

SAE/GAE

- 应用开发平台——快速部署与实施



- 为什么我们需要云计算？
- 或者说，云计算技术是怎么来的？

源自实际的服务需求？



- 硬件需求
 - 变买为租
 - 成本压力
 - 只需要使用权，不注重所有权
 - 能力弹性
 - 硬件能力需求是弹性的
 - 希望能够按需扩展
 - 可靠性保障
 - 需要可靠但自身缺少相关保障技术
 - 管理维护
 - 需要管理维护但又无法承担成本



- 平台需求（开发者需求）
 - 变买为租
 - 成本压力
 - 只需要使用权，不注重所有权
 - 配置管理
 - 安装、配置过程繁琐
 - 各种开发环境的冲突
 - 需要平台式一站式服务
 - 培训
 - 快速应对应用使用者的个性化新应用需求



- 应用需求

- 更多的应用选择

- 个性化
 - 需要更多提供商参与

- 更方便的应用获取渠道

- 应用商店
 - 试用

- 参考信息

- 他人的评价
 - 购买量
 - 开发者口碑

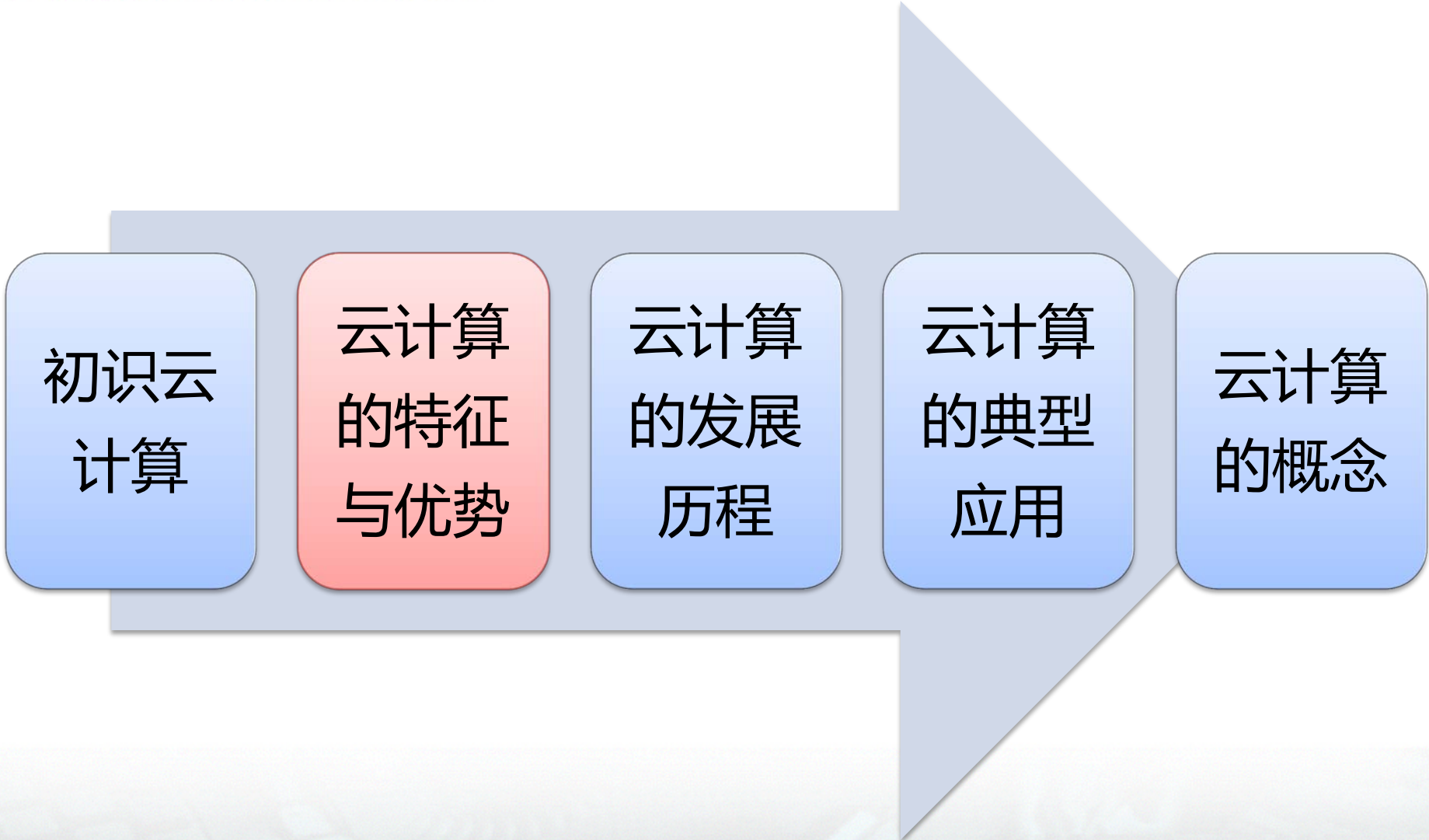


- 云计算是万能灵药？
- 什么前提下我们才需要云计算？
- 我们建设云计算中心、数据中心为了什么？

云不是重点，业务才是重点
云是业务的支撑手段



内容提要



初识云计算

云计算
的特征
与优势

云计算
的发展
历程

云计算
的典型
应用

云计算
的概念



北京大学



云计算的特征

- 硬件和软件都是资源
 - 通过互联网以服务方式获得
 - 资源不仅包括处理器机时、网络带宽等物理范畴，还包括软件平台、Web服务和应用程序等
 - 传统自给自足的IT模式->网络化分工专业、协同配合模式
 - 将IT交由他人管理



北京大学



云计算的特征

- 资源可以根据需要进行动态扩展与配置
 - 按需分配，需要多少使用多少
 - 取之不尽用之不竭
 - 避免无谓的浪费
- 按用计费、无需管理
 - 租用式，用多少交多少钱
 - 只负责使用，无需管理



北京大学



云计算的特征

- 物理上分布式共享，逻辑上以单一整体呈现
 - 计算密集型需要并行计算提高效率
 - 地域上的分布式
 - 合作共享机制
 - 实现机制对用户不可见



北京大学

- 优化产业布局

- 成本关键由硬件成本转为电力成本、散热成本
- 分散的、高能耗模式转变为集中的、资源友好的模式
- 自给自足资源作坊转变为规模效应的工业化资源工厂





云计算的优势

- 推进专业分工
 - 专业的大型云计算厂商
 - 实力雄厚的科研技术团队推动硬件技术革新
 - 具有丰富知识和经验的维护管理团队和配套软件
 - 产业链中的新产业契机



北京大学



云计算的优势

- 提升资源利用率
 - 更加高效的解决常规业务
 - 更好的应对突发任务或事件
 - 更加平衡的资源分配和负载



北京大学



云计算的优势

- 减少初期投资
 - 更少的IT基础设施投入
 - 更少的软件投入
 - 更少的人力投入
 - 更短的培训周期
 - 更灵活的转型支持



北京大学

- 降低管理开销
 - 服务化管理
 - 无需自己建立维护管理团队
 - 按需应变的解决方案
 - 知识、经验更加丰富的管理队伍
 - 通过业务模块的可配置提高系统的灵活性





对云参与者的影响

- 硬件提供商

- 并未影响硬件提供商的业务
- 使用门槛的降低带来更多的用户，硬件能力的总需求增加
- 面向更加专业的顾客
- 促进硬件产品的技术革新



北京大学



对云参与者的影响

- 基础软件提供商-操作系统、中间件
 - 影响巨大
 - 新的模式
 - 操作系统的升级
 - 中间件厂商的新契机
 - 技术的革新



北京大学



对云参与者的影响

- 云提供商
 - 最大的受益者
 - 高门槛：丰富的硬件、软件和行业经验
 - 核心垄断地位，竞争者少，但十分激烈



北京大学



对云参与者的影响

- 云服务提供商
 - 前景广阔
 - 为眼光卓越的精英们提供创业机会
 - 小软件厂商的转型目标
 - 业务转型快，创新和变革为核心



北京大学



对云参与者的影响

- 应用提供商
 - 低门槛，竞争者多，竞争激烈
 - 以创新赢得市场
 - 面向更加专业的顾客
 - 淘汰率极高



北京大学



对云参与者的影响

- 个人用户
 - 更多的基于互联网的服务
 - 无需购买高性能电脑、复杂软件
 - 更低的费用，更多的选择
 - 更加安全的数据保护



北京大学



对云参与者的影响

- 企业机构用户
 - 更短的IT项目支持
 - 更少的IT投入和维护成本
 - 遗留系统迁移和保密成为巨大的问题



北京大学

内容提要

初识云计算

云计算的特征与优势

云计算的发展历程

云计算的典型应用

云计算的概念



北京大学

云计算的发展历程

Super
Computer



云计算的发展历程

Super
Computer

Cluster
Computing

集群计算

- ❖ 将一组松散的计算机硬件通过软件集成、连接起来，紧密的协作完成计算工作
- ❖ PVM 、 MPI
- ❖ 1960~
- ❖ 相对于超级电脑有较高的性价比



云计算的发展历程

Super
Computer

Cluster
Computing

Distributed
Computing

分布式计算

- 把需要进行大量计算的工程数据分割成小块，由多台计算机分别计算，然后上传计算结果，将结果统一合并得出最终结果。
- Remote procedure calls(RPC)



云计算的发展历程

Super
Computer

Cluster
Computing

Distributed
Computing

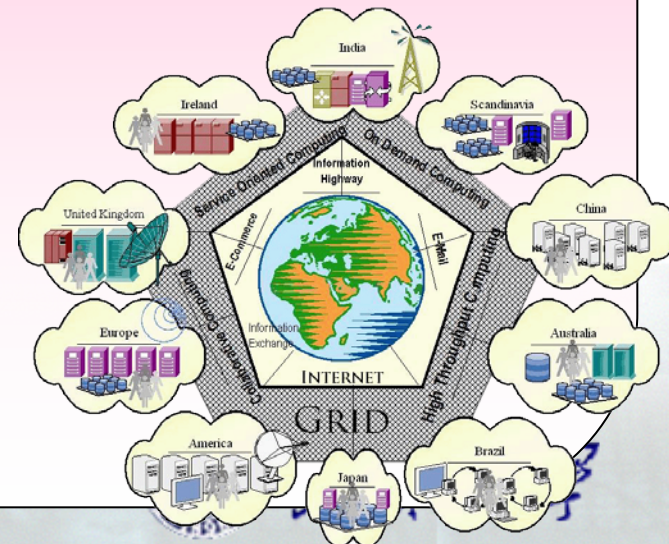
Grid
Computing

网格计算

- ❖ 将大量异构计算机的未用资源（**CPU**时间和磁盘空间）作为嵌入在分布式系统基础设施中的一个虚拟的计算机集群，为解决大规模的计算问题提供一个模型。

Globus

- ❖ 1990~



云计算的发展历程

Super
Computer

Cluster
Computing

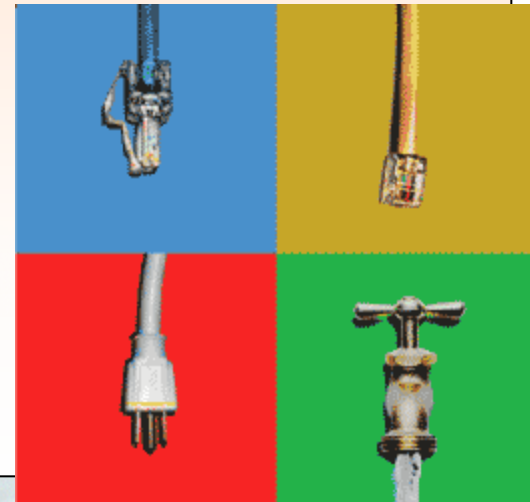
Distributed
Computing

Grid
Computing

Utility
Computing

效用（公用）计算

- ❖ 提倡一种理想化的企业信息架构，让IT服务模仿公共服务的方式进行，如水、电力、煤气的供应方式，“用多少付多少”以及“按需即用”
- ❖ From IBM,



云计算的发展历程

Super
Computer

Cluster
Computing

Distributed
Computing

Grid
Computing

Utility
Computing

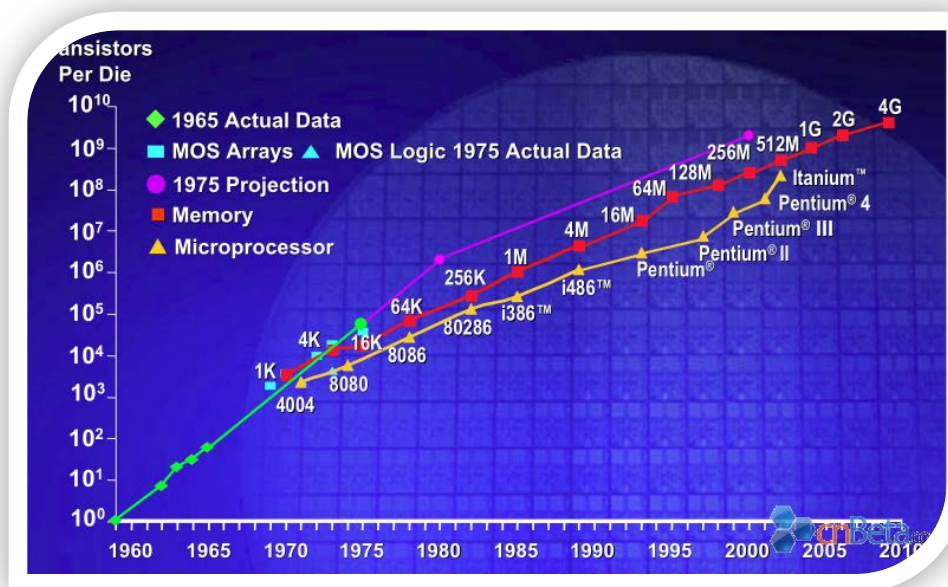
Cloud
Computing

云计算

- ❖ 通过网络将庞大的运算处理程序自动拆分成若干个较小的子程序，再交给多个服务器所组成的庞大系统经过搜索、运算分析后将处理结果返回给用户
- ❖ 云---网络
- ❖ Google：GFS、MapReduce 及BigTable

- 芯片与硬件技术

- 摩尔定律：集成电路芯片上所集成的电路的数目，每隔18个月就翻一番，同时性能也提升一倍



- 硬件能力的激增、成本的大幅下降，使得独立运作的公司集中客观的硬件能力实现规模效益成为可能



- 资源虚拟化

- 资源在云端，需要被统一的管理
- 异构硬件、兼容性问题
- 虚拟化技术
 - 开发人员只需关注业务逻辑，无需考虑底层资源的供给与调度问题
 - 硬件资源可以被有效的细粒度分割和管理，以服务方式提供硬件和软件资源成为可能
 - 单点崩溃不会影响全局
 - 易创建性使得使用更加方便，提高自身可靠性和可用性



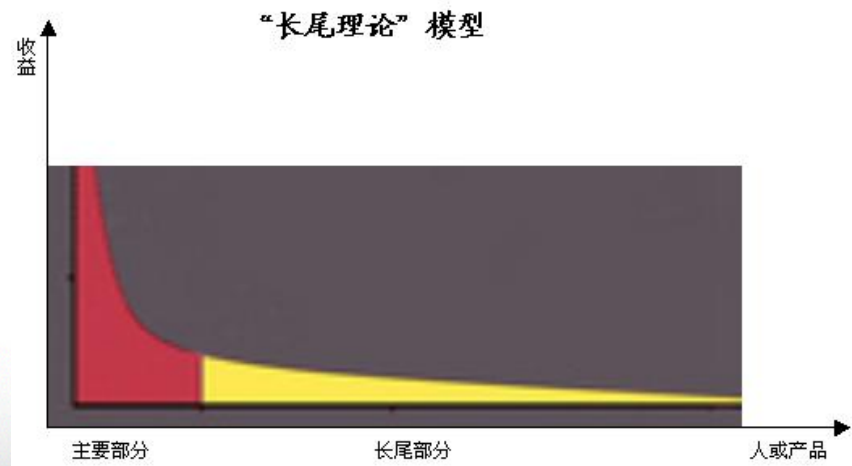
- 面向服务的架构SOA
 - 开放式数据模型
 - 统一通信标准
 - 更加丰富的服务
 - 更加松散耦合、灵活的IT架构
 - 转变了人们对IT系统的认识



- 软件即服务SaaS

- 转变了人们使用服务的方式
- 使得终端用户熟悉服务的交互模式
- 改变了IT界的商业模式
- 实力雄厚的大公司负责基础设施，小企业通过创新挖掘充满潜力的市场

- “长尾理论”：



- 互联网技术
 - 基础设施
 - 多种接入方式
 - 更加广阔的覆盖
 - 带宽和可靠性得到大幅提升
 - 使得IT新模式的稳定性、可靠性、安全性、可用性、灵活性、可管理性、自动化程度、节能环保等得到保障



- Web2.0技术

- 用户从信息的获得者变成信息的贡献者
- 博客（微博）、内容聚合、百科全书（Wiki）、社交网络、对等网络（P2P）
- 富互联网应用（Rich Internet Application）成为网络应用发展的趋势
- 改变了人们的生活方式
- 为云计算提出了内在需求



内容提要

初识云计算

云计算的特征与优势

云计算的发展历程

云计算的典型应用

云计算的概念



北京大学

- 案例一：
 - 美国国家档案馆公布1993-2001年希拉里作为第一夫人期间的白宫日程档案
 - 低质量的PDF需要被转换成可检索的文件格式
 - 华盛顿邮报现有的计算能力转换1页要30min
 - 新闻时效性
 - Amazon EC2在9小时内提供了1407小时的虚拟服务器机时，将全部档案转换完毕
 - 只需支付144.62美元



- 案例二：
 - Giftag是与好友分享购物清单的Web2.0应用
 - 以插件形式安装在浏览器上
 - 一经推出便广泛流行，服务器不堪重负
 - 迁移到Google App Engine平台
 - 通过GAE可伸缩的计算性能来响应用户请求
 - 实现初创应用向稳定、持续增长的服务过渡



- 案例三：
 - 哈根达斯需要一个CRM解决方案
 - 让分布在世界各地的员工进行沟通协作
 - 可根据不同的需求进行灵活配置
 - 通过Salesforce CRM在不到6个月上线
 - 无需构建自己的计算中心



- 案例四：
 - 云计算课程需要一个课程网站
 - 各地学生共享课件
 - 公网可访问
 - 传统方式需要建站、备案、维护等
 - 通过SAE在不到2小时上线
 - 费用极低

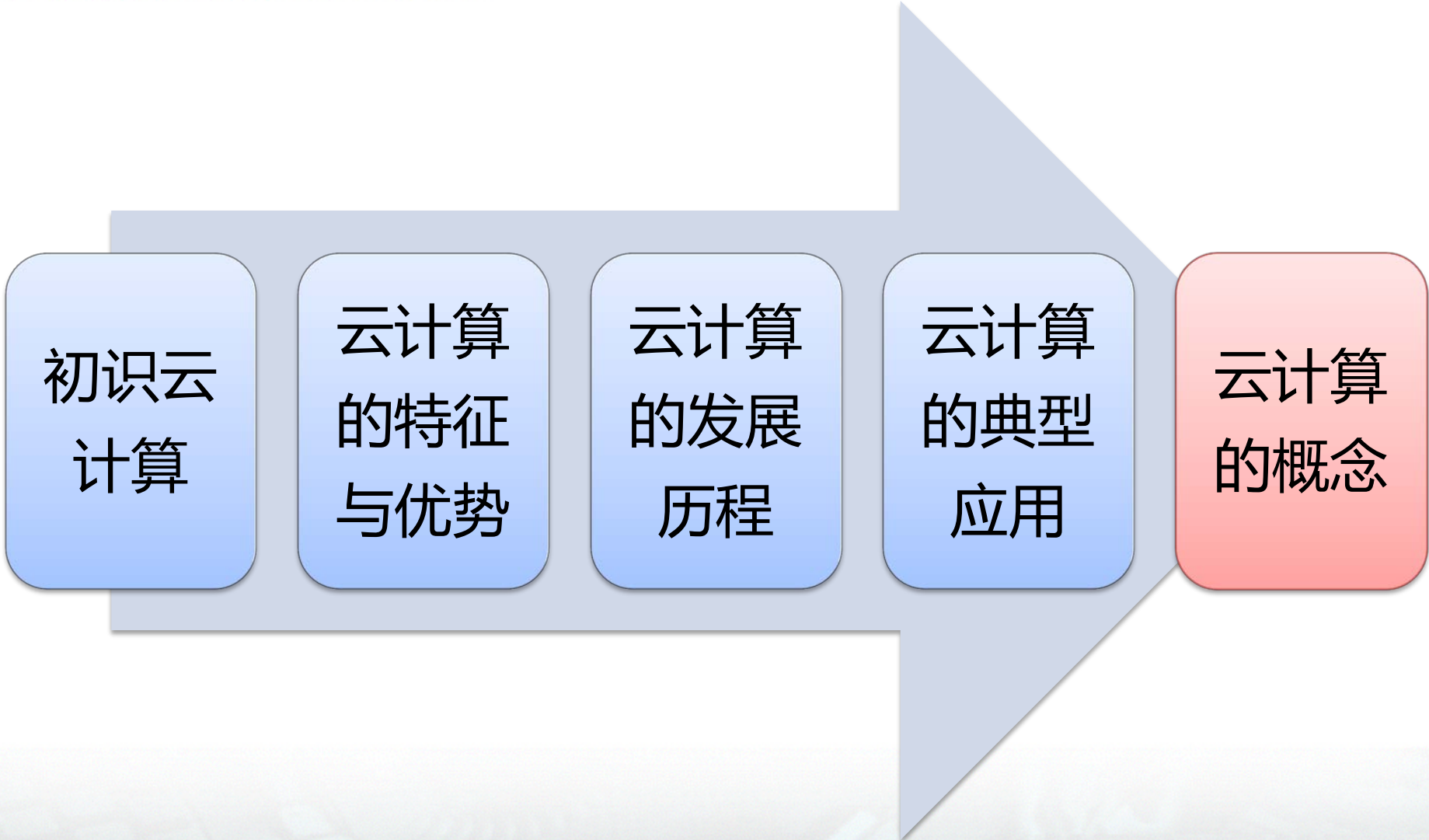


典型应用

- ❖ Google
- ❖ Amazon
- ❖ IBM
- ❖ Microsoft
- ❖ Facebook
- ❖ Twitter
- ❖ 百度
- ❖ 阿里
- ❖ 腾讯
- ❖ 新浪
- ❖ ...



内容提要



初识云计算

云计算
的特征
与优势

云计算
的发展
历程

云计算
的典型
应用

云计算
的概念



北京大学



云计算的概念

- 云计算是一种能够将动态伸缩的虚拟化资源通过互联网以服务的方式提供给用户的计算模式。

-wiki

- XaaS+ pay-as-you-go computing

-Google

- 云计算是一种新的应用模式在探索中
云计算是一种新的技术模式
云计算是一种新的编程理念
云计算是一种共享的网络交付信息服务的模式，云服务的用户看到的只是应用，而不用关心相关基础设施的具体实现。
-IBM



北京大学

• 云计算的分类

基础设施云

- 为用户提供底层的、接近于直接操作硬件资源的服务接口
- 用户直接通过接口获取计算和存储能力，几乎没有逻辑限制
- 仅提供计算和存储等基础功能，应用的实现需用户自行完成

平台云

- 为用户提供应用的托管平台
- 应用的开发和部署需遵守平台的规则和限制，如语言、编程框架、数据存储模型等

应用云

- 避免无谓的浪费





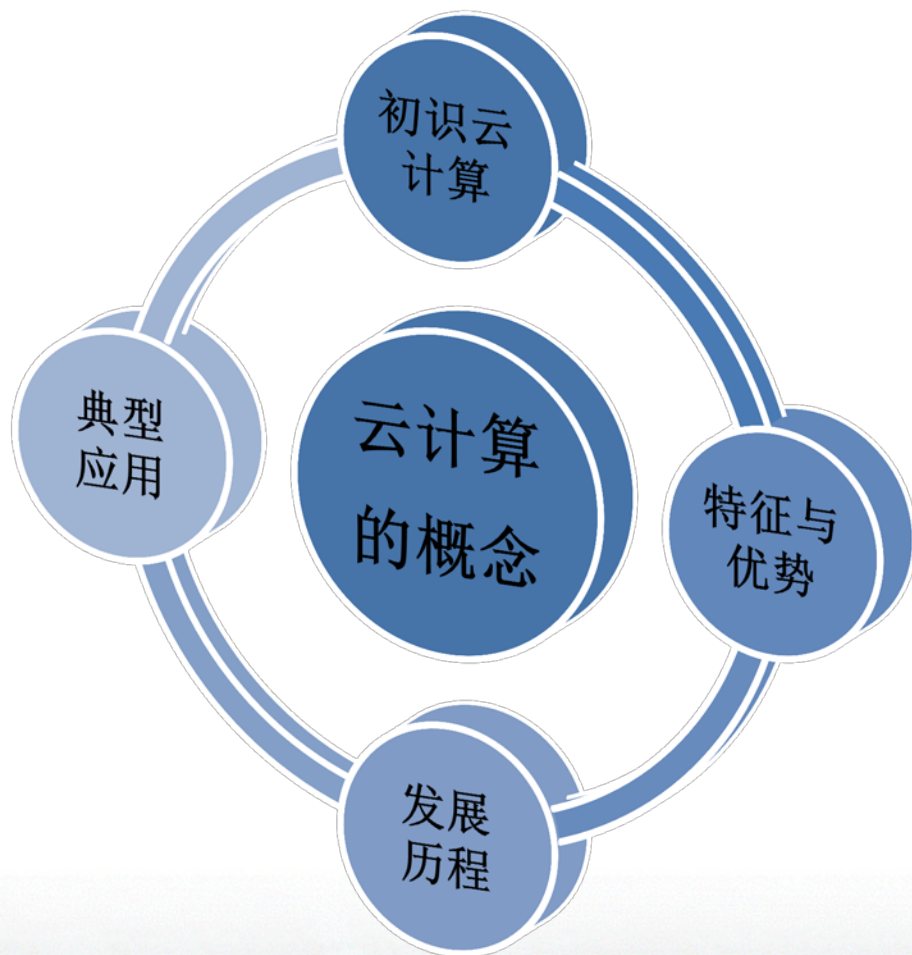
云计算的概念

- 服务类型：基础设施云、平台云、应用云

分 类	服务类型	灵活性	难易程度
✓基础设施云	接近原始的计算存储能力	高	难
✓平台云	应用的托管环境	中	中
✓应用云	特定功能的应用	低	易

- 服务方式：公有云、私有云、混合云





- 初识云计算
 - Google
 - www.12306.cn
 - VMware
 - GAE/SAE
- 特征与优势
 - 四个特征
 - 五点优势
 - 七类参与者的好处
- 发展历程
 - 六步历程
 - 产生动因
- 典型应用
 - 应用
 - 云服务



课后作业

- 思考题：
 - 什么是云计算
 - 云计算的优势
 - 云计算的动因
- 查找资料：
 - 业界主流的云计算产品
 - 典型的云计算应用



北京大学