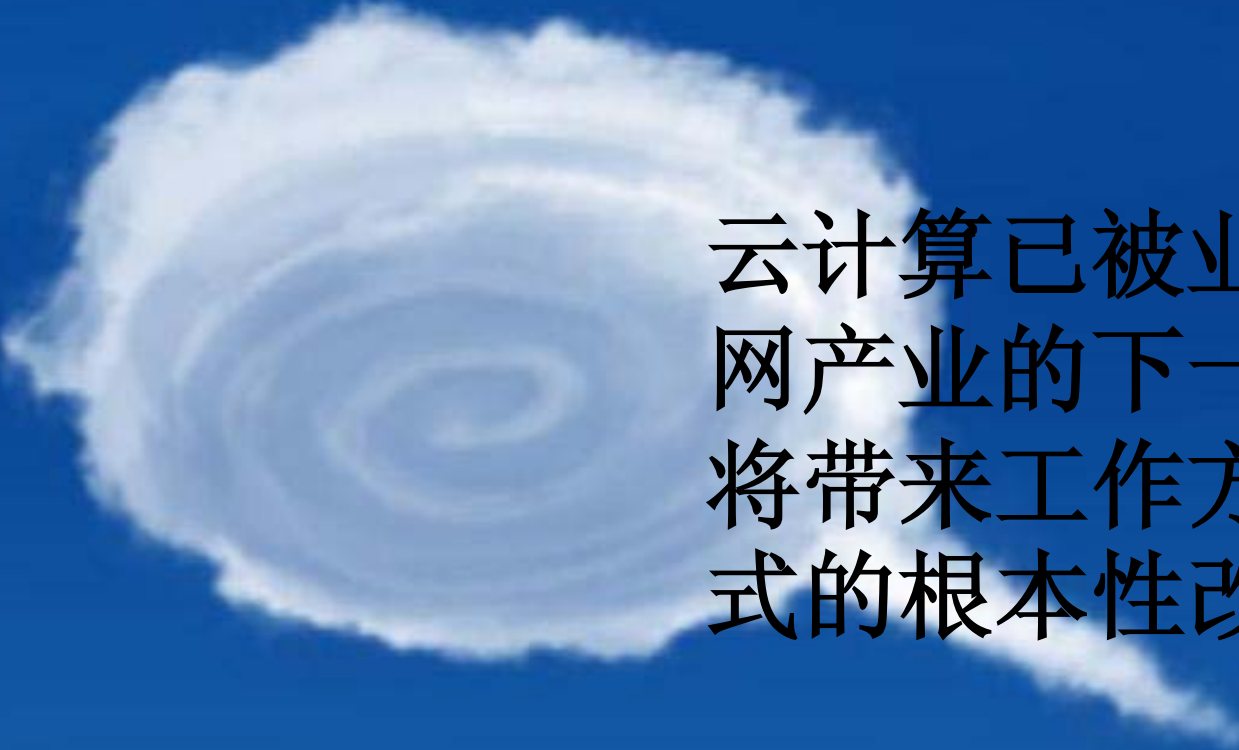


# 云计算技术原理和应用发展

报告人：\*\*\*

2010年\*月\*日

# 云计算：Cloud Computing



云计算已被业界视为互联网产业的下一次革命，它将带来工作方式和商业模式的根本性改变

**云时代已经到来！**

# 主要内容

**01**

云计算概念及影响

**02**

云计算的主要服务形式和典型应用

**03**

云计算的核心技术

**04**

典型云计算平台介绍

**05**

云计算实验：云脑系统

**06**

如何进入云计算领域

# 云计算概念及影响

云计算（**Cloud Computing**）是由分布式计算（**Distributed Computing**）、并行处理（**Parallel Computing**）、网格计算（**Grid Computing**）发展来的，是一种新兴的商业计算模型。

中国网格计算、云计算专家刘鹏给出如下定义：“云计算将计算任务分布在大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算力、存储空间和各种软件服务”。

云计算模式即为电厂集中供电模式，它的最终目标是将计算、服务和应用作为一种公共设施提供给公众，使人们能够像使用水、电、煤气和电话那样使用计算机资源。



# 云计算的主要服务形式和典型应用

云计算的主要服务形式有：**SaaS(Software as a Service)**，**PaaS(Platform as a Service)**，**IaaS(Infrastructure as a Service)**。

## •(1)软件即服务(SaaS)

**SaaS**服务提供商将应用软件统一部署在自己的服务器上，用户根据需求通过互联网向厂商订购应用服务，服务提供商根据客户所定软件的数量、时间的长短等因素收费，并且通过浏览器向客户提供软件的模式。

这种服务模式的优势是，由服务提供商维护和管理软件、提供软件运行的硬件设施，用户只需拥有能够接入互联网的终端，即可随时随地使用软件。**客户不再像传统模式那样花费大量资金在硬件、软件、维护人员，只需要支出一定的租赁服务费用，通过互联网就可以享受到相应的硬件、软件和维护服务。**对于小型企业来说，**SaaS**是采用先进技术的最好途径。

目前，**Salesforce.com**是提供这类服务最有名的公司，**Google Docs**，**Google Apps**和**Zoho Office**也属于这类服务。

## •平台即服务(PaaS)

把开发环境作为一种服务来提供。这是一种分布式平台服务，厂商提供开发环境、服务器平台、硬件资源等服务给客户，用户在其平台上定制开发自己的应用程序并通过其服务器和互联网传递给其他客户。

**PaaS**能够给企业或个人提供研发的中间件平台，提供应用程序开发、数据库、应用服务器、试验、托管及应用服务。

**Google App Engine, Salesforce的force.com平台，八百客的800APP**是**PaaS**的代表产品。



# 云计算的主要服务形式和典型应用

## 基础设施即服务(IaaS)

**IaaS**即把厂商的由多台服务器组成的“云端”基础设施，作为计量服务提供给客户。它将内存、I/O设备、存储和计算能力整合成一个虚拟的资源池为整个业界提供所需要的存储资源和虚拟化服务器等服务。这是一种托管型硬件方式，用户付费使用厂商的硬件设施。

**Amazon Web服务（AWS），IBM的BlueCloud**等均是**将基础设施作为服务出租**。

**IaaS**的优点是用户只需低成本硬件，按需租用相应计算能力和存储能力，大大降低了用户在硬件上的开销。

**GoogleDocs**类似于微软的**Office**的一套在线办公软件。用户只需一台接入互联网的计算机和浏览器即可在线处理和搜索文档、表格、幻灯片，并可以通过网络与他人分享并设置共享权限。

<http://docs.google.com>

**Google AppEngine**使用户可以使用**Python**和**Java**在**Google**的基础架构上开发和部署运行自己的应用程序。每个**Google AppEngine**应用程序可以使用达到**500MB**的持久存储空间及可支持每月**500万**综合浏览量的带宽和**CPU**，并且可根据用户的访问量和数据存储需要的增长轻松扩展。

其它与“云”相关应用：云安全、云存储、云下载



# 云计算的核心技术

- 编程模型
- 海量数据分布存储技术
- 海量数据管理技术
- 虚拟化技术
- 云计算平台管理技术

- ## 编程模型

为了使云计算环境下的编程十分简单，**Google**开发了**java**、**Python**、**C++**编程工具**MapReduce**，它是一种简化的分布式编程模型和高效的任务调度模型，用于大规模数据集（大于**1TB**）的并行运算。

**MapReduce**模式的思想是将要执行的问题分解成**Map**(映射)和**Reduce**(化简)的方式，先通过**Map**程序将数据切割成不相关的区块，分配（调度）给大量计算机处理，达到分布式运算的效果，再通过**Reduce**程序将结果汇整输出。

## ●海量数据分布存储技术

**GFS**是一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用。

一个**GFS**集群由一个主服务器和大量的块服务器构成，并被许多客户访问。主服务器存储文件系统所有的元数据，包括名字空间、访问控制信息、从文件到块的映射以及块的当前位置。它也控制系统范围的活动，如块租约管理，孤儿块的垃圾收集，块服务器间的块迁移。主服务器定期通过**HeartBeat**消息与每一个块服务器通信，给块服务器传递指令并收集它的状态。**GFS**中的文件被切分为**64MB**的块并以冗余存储，每份数据在系统中保存**3**个以上备份。

客户与主服务器的交换只限于对元数据的操作，所有数据方面的通信都直接和块服务器联系，这大大提高了系统的效率，防止主服务器负载过重。



# 数据管理技术

云计算需要对分布的、海量的数据进行处理、分析，因此，数据管理技术必需能够高效的管理大量的数据。云计算系统中的数据管理技术主要是**Google**的**BT(BigTable)**数据管理技术和**Hadoop**团队开发的开源数据管理模块**HBase**。

**BT**是建立在**GFS, Scheduler, Lock Service**和**MapReduce**之上的一个大型的分布式数据库，与传统的关系数据库不同，它把所有数据都作为对象来处理，形成一个巨大的表格，用来分布存储大规模结构化数据。

**Google**的很多项目使用**BT**来存储数据，包括网页查询，**Google earth**和**Google**金融。这些应用程序对**BT**的要求各不相同：数据大小（从**URL**到网页到卫星图象）不同，反应速度不同（从后端的大批处理到实时数据服务）。对于不同的要求，**BT**都成功的提供了灵活高效的服务。

# 虚拟化技术

通过虚拟化技术可实现软件应用与底层硬件相隔离，它包括将单个资源划分成多个虚拟资源的裂分模式，也包括将多个资源整合成一个虚拟资源的聚合模式。虚拟化技术根据对象可分成存储虚拟化、计算虚拟化、网络虚拟化等，计算虚拟化又分为系统级虚拟化、应用级虚拟化和桌面虚拟化。

## 云计算平台管理技术

云计算资源规模庞大，服务器数量众多并分布在不同的地点，同时运行着数百种应用，如何有效的管理这些服务器，保证整个系统提供不间断的服务是巨大的挑战。

云计算系统的平台管理技术能够使大量的服务器协同工作，方便的进行业务部署和开通，快速发现和恢复系统故障，通过自动化、智能化的手段实现大规模系统的可靠运营。



# 典型云计算平台介绍

- **Google** 的云计算平台

**Google**公布的云计算技术主要由**MapReduce**、**Google**文件系统(**GFS**)、**BigTable**组成。

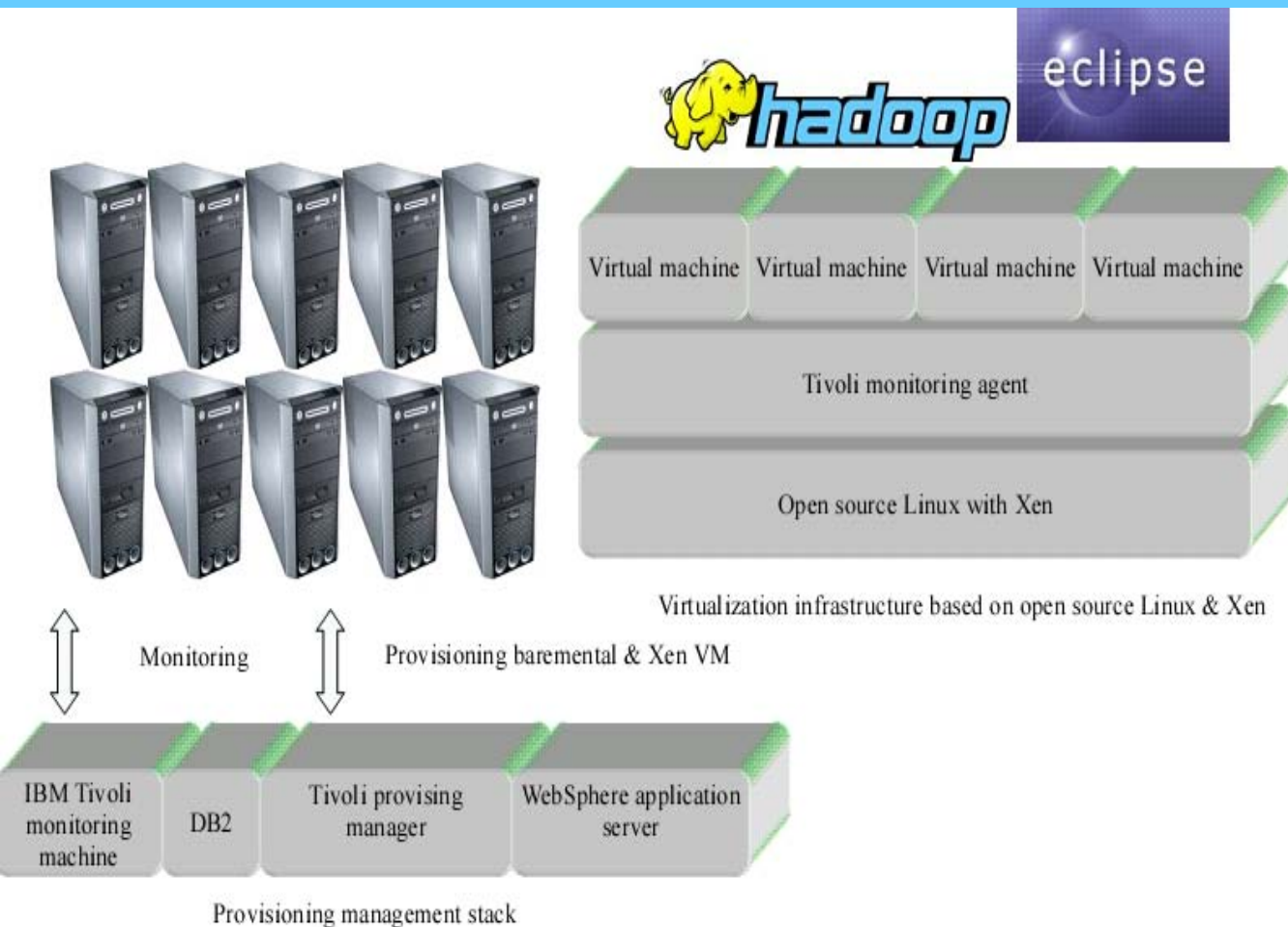
**Google** 还构建其他云计算组件，包括一个领域描述语言以及分布式锁服务机制等。**Sawzall**是一种建立在**MapReduce**基础上的领域语言，专门用于大规模的信息处理。**Chubby**是一个高可用、分布式数据锁服务，当有机器失效时，**Chubby**使用**Paxos**算法来保证备份。





# • IBM“蓝云” 计算平台

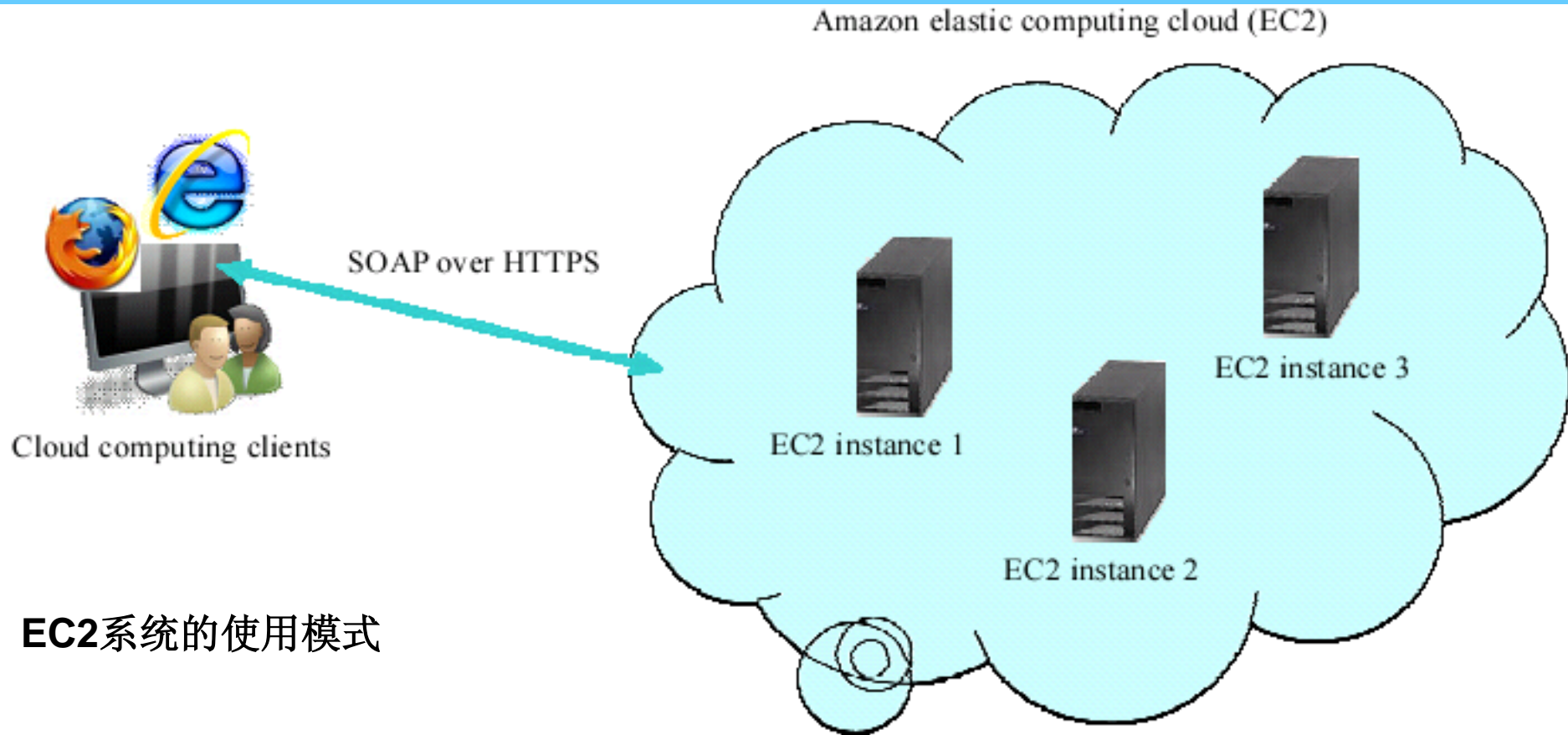
“蓝云” 基于IBM Almaden研究中心的云基础架构，采用了Xen和PowerVM虚拟化软件，Linux操作系统映像以及Hadoop软件(Google File System以及MapReduce的开源实现)。IBM已经正式推出了基于x86芯片服务器系统的“蓝云”产品。



由图可知，“蓝云”计算平台由一个数据中心、**IBM Tivoli**部署管理软件(**Tivoli provisioning manager**)、**IBM Tivoli**监控软件(**IBM Tivoli monitoring**)、**IBM WebSphere**应用服务器、**IBM DB2**数据库以及一些开源信息处理软件和开源虚拟化软件共同组成。

# Amazon的弹性计算云

Amazon将自己的弹性计算云建立在公司内部的大规模集群计算的平台上，而用户可以通过弹性计算云的网络界面去操作在云计算平台上运行的各个实例(instance)。用户使用实例的付费方式由用户的使用状况决定，即用户只需为自己所使用的计算平台实例付费，运行结束后计费也随之结束。



# 云计算系统实例——云脑

云登陆实现

云文件系统实现

云存储实现

云监控实现

云计算器实现



**机群系统：共5台电脑，一台作为主服务器和监控服务器，两台作为应用节点服务器和存储节点服务器，两台作为计算节点服务器；**

**基础软件：Linux、MPICH2.0（免费的消息传递编程接口系统）、JDK（Java运行环境）、Tomcat、MySQL数据库；**



- ✓ 精通linux操作系统，熟悉linux环境下环境搭建和编程；
- ✓ 精通UML、XML、JavaScript脚本语言、Java语言及相关开发工具、多线程编程、网络编程，深刻理解算法设计、数据结构和面向对象思想；
- ✓ 精通数据库原理和数据库应用，熟练掌握MySql、Oracle等两种以上数据库管理系统；
- ✓ 熟悉Google GFS/Bigtable/Mapreduce原理，熟悉对开源框架Hadoop的配置和使用；
- ✓ 熟悉虚拟化技术，如Xen,Vmware,KVM等，熟悉分布式虚拟化；
- ✓ 深入理解云计算体系架构、分布式计算、网格计算、并向计算、分布式存储、计算机网络和存储结构；

谢谢