

高性能计算与云计算

第一讲 引言

胡金龙, 董守斌

jlhu@scut.edu.cn

华南理工大学计算机学院
广东省计算机网络重点实验室

Communication & Computer Network Laboratory (CCNL)

主要内容

- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务

本课程是关于什么？

- 如何使得计算机可以**更快更方便地解决更大规模**的问题（bigger problems faster）
- 直观的想法：可以用多台计算机
 - 如果一个程序在一台计算机上需要100个小时才能运行完，如果使用100台计算机，可能只要10个小时
 - 一台计算机只能处理2GB的数据集，如果使用100台计算机，可能可以处理200GB的数据集
- 如何能做到？

什么是高性能计算？

- 高性能计算（HPC, High Performance Computing）
 - 研究如何分解一个**巨大规模**的问题，并分配给多个计算机进行处理，并把这些计算结果综合起来得到最终的结果
- 高性能计算的等价词
 - 并行计算(Parallel Computing)
 - 高端计算(High-end Parallel Computing)
 - 超级计算(Super Computing)

为什么需要高性能计算？

- 人类对计算及性能的要求是无止境的
 - 从系统的角度：集成系统资源，以满足不断增长的对性能和功能的要求
 - 从应用的角度：适当分解应用，以实现**更大规模或更细致**的计算

超算—国家战略制高点

- 超级计算机是国家科研的基础工具，对提升综合国力具有战略意义。已成为国家科技创新力的象征，世界各国争夺的一个战略制高点。
- 美国对中国实施超算禁运等封锁政策。毫无疑问，“**国之重器**”不能受制于人。

The screenshot shows the official website of the Chinese government. The header features the Chinese National Emblem, the text "中华人民共和国中央人民政府" (The Central People's Government of the People's Republic of China), and the English translation "The Central People's Government of the People's Republic of China". Below the header is a navigation bar with links to "网站首页" (Home Page), "今日中国" (Today's China), "中国概况" (Profile of China), "法律法规" (Law and Regulations), "公文公报" (Official Documents and Bulletins), "政务互动" (Government Affairs Interaction), "政府建设" (Government Construction), "工作动态" (Work Dynamics), "人事任免" (Personnel Appointments and Removals), and "新闻发布" (Press Release). The main content area displays a news article titled "习近平对天河二号超级计算机系统研制成功作重要指示" (President Xi Jinping Issues Important Instructions on the Successful Development of the Tianhe-2 Supercomputer System). The article is dated June 18, 2013, at 18:40, and is attributed to Xinhua News Agency. It includes a summary of the instructions and a quote from President Xi. At the bottom of the page are links for font size adjustment, printing, and closing the window.

新华社北京6月18日电（记者 李宣良、白瑞雪）中共中央总书记、国家主席、中央军委主席习近平，近日对国防科技大学研制成功天河二号超级计算机系统作出重要指示，对取得这一成绩表示热烈祝贺，向参加系统研制任务的全体同志致以诚挚的问候。

习近平指出，天河二号超级计算机系统研制成功，标志着我国在超级计算机领域已走在世界前列。他希望同志们总结经验，再接再厉，坚持以我为主，勇于自主创新，不断强化前沿技术研究，为推动我国科技进步、建设创新型国家作出更大贡献。

科技安全--核心技术自己掌握

只有把核心技术掌握在自己手中,才能真正掌握竞争和发展的主动权,才能从根本上保障国家经济安全、国防安全和其他安全。

——2014年6月9日,习近平在中国科学院第十七次院士大会、中国工程院第十二次院士大会上的讲话



Rank	System	Cores	Rmax (TFlop/s)	Peak (TFlop/s)	Power (kW)
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,299,072	415,530.0	513,854.7	28,335
2	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371

“神威·太湖之光”超级计算机是由中国自主研制的超级计算机,是目前全球第四快的超级计算机。

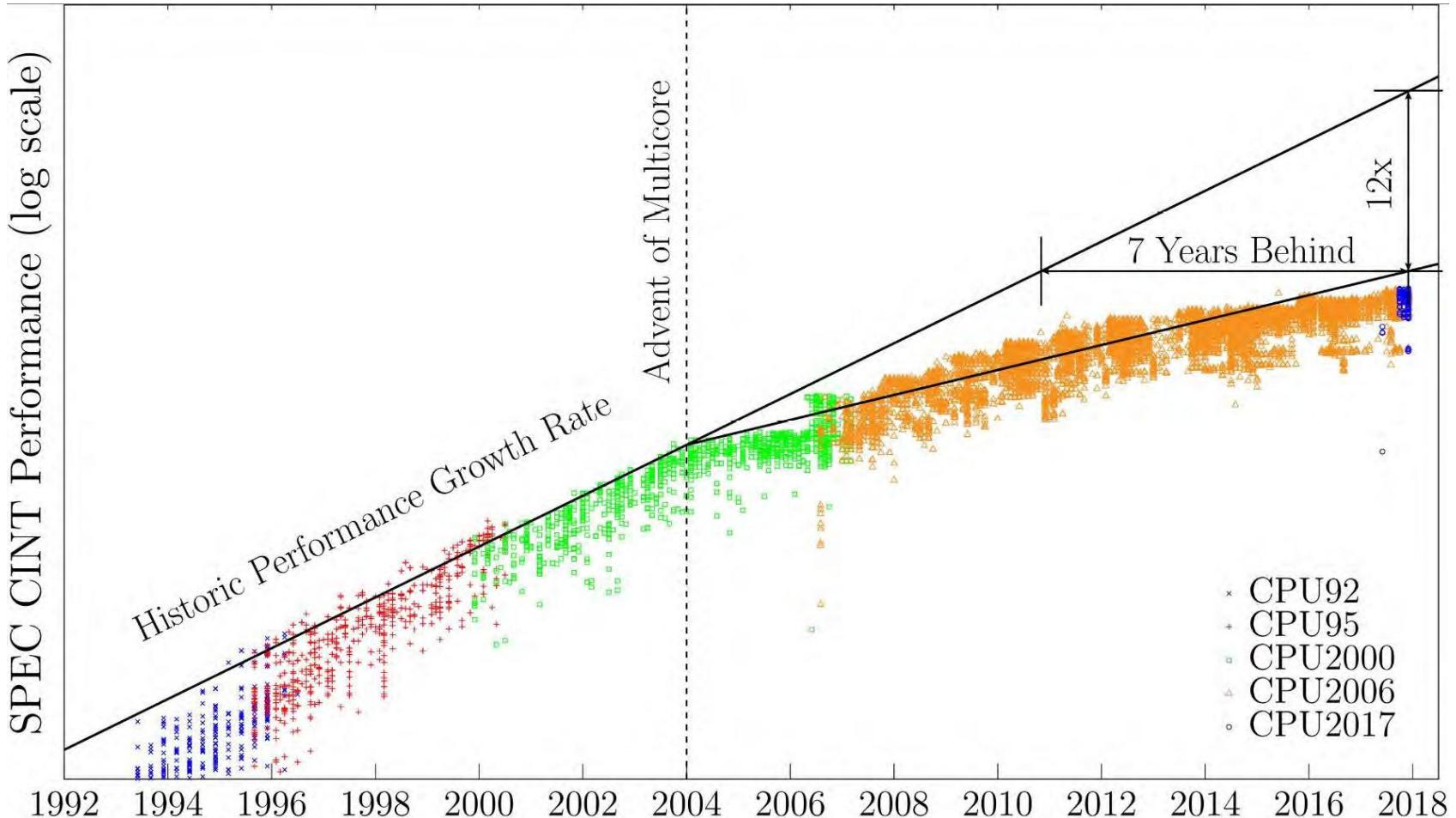
做得更快的方法

- 三种提高性能的方法
 - 努力工作 (Work Harder)
 - Increase microprocessor performance
 - 工作得更有效率 (Work smarter)
 - Better algorithm
 - 获取帮助 (Getting help)
 - Parallel processing

单核CPU的瓶颈

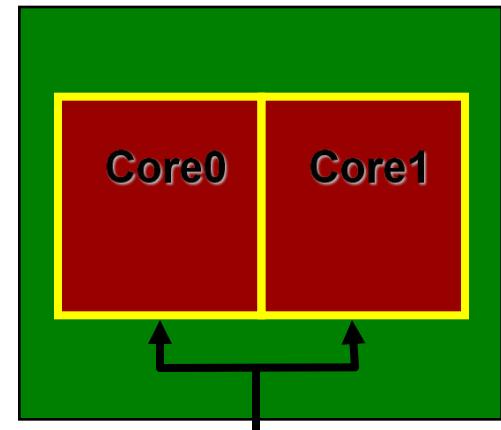
- CPU是基于低延迟的设计
 - 复杂的逻辑控制单元，提供分支预测的能力来降低延迟
 - 大量晶体管用于缓存，增大缓存从而提升缓存命中率
 - 加强算术运算单元ALU来降低延迟，执行双精度浮点运算只需1~3个时钟周期
- 由于散热、晶体管尺寸等影响，近年来，频率提升接近停止

摩尔定律已失效（2004年）



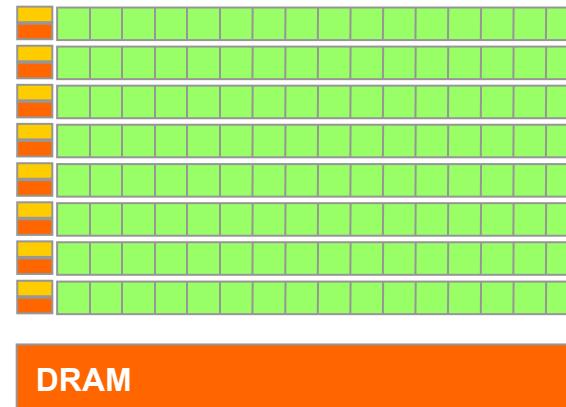
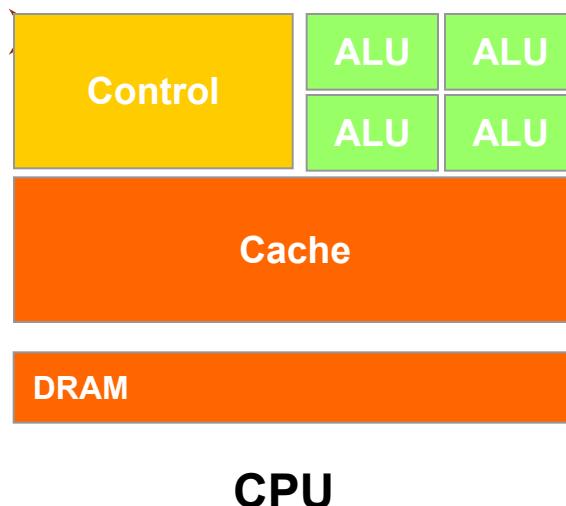
多核时代来临

- AMD在2006年第一个推出了双核处理器，这种处理器的计算单元相互独立，但它们将共享CPU的一、二级缓存
- SC11展示的Intel Knight Corner (百亿/TeraFLOP级别的处理器，50 cores on 22nm Chip)，称为集成多核架构处理器MIC (Many Integrated Cores)
 - 产品系列：Intel Xeon Phi (至强酷核)



GPU多核

- 基于高吞吐量的设计
 - 更高的浮点运算能力：大量的计算单元，可以支撑大量数据的并行计算
 - 更高的访存速度：通过并发访存及线程切换掩盖存储器访问延迟



多核挑战软件开发

- 多核的运算速度一定会比单核的CPU快吗？
- 不一定
- 对底层体系结构不了解的话，无法充分利用硬件性能
- 要想发挥多核功能，设计的软件就首先要能做**并行计算**
 - 简单的例子：排序、搜索
- 多核的出现使得高性能计算进入普及的时代

大数据时代来临

- 社交网络的逐渐成熟，移动带宽迅速提升，云计算、物联网应用更加丰富，更多的传感设备、移动终端接入到网络，由此产生的数据及增长速度将比历史上的任何时期都要多，都要快



数据来源：艾瑞咨询

大数据时代需要高性能计算 --从“计算机”发展的视角

- Computers as “**computers**” (1930s -1990s)
 - Computer architecture (CPU chips, cache, DRAM, and storage)
 - Operating systems (both open sources and commercial)
 - Compilers (execution optimizations)
 - Databases
 - Standard scientific computing software
- Computers as “**networks**” (1990s – 2010s)
 - Internet capacity updates:
 - 281 PB (1986), 471 PB (1993, +68%), 2.2 EB (2000, +3.67 times), 65 EB (2007, 29 times), 667 EB (2013, 9 times)
 - Wireless infrastructure
- Computers as “**data centers**” (starting 21st century)
 - Everything is digitized and saved in daily life and all other applications
 - Time/space creates big data: short latency and unlimited storage space
 - Data-driven decisions and actions

大数据时代需要高性能计算

- 当待处理数据量巨大，只有分布在成百上千/成千上万个节点上并行计算才能在可接受的时间内完成



这是个简单的工作吗？

大规模数据处理面临的困难

- 大规模PC集群可靠性很差

- 1节点的MTBF（Mean time between failures） = 3年
- 1000个节点的MTBF = 1天
- 商用网络 =低带宽

运行系统（Runtime System）：

- 良好可扩展性
- 良好的容错能力

- 并行/分布式程序开发、调试困难

- 数据如何划分
- 任务如何调度
- 任务之间的通信
- 错误处理，容错...

编程模型（Programming Model）：

- 一定的表达能力
- 很好的简单易用性

大数据时代的高性能计算—云计算

- 云计算是一种新兴的共享基础架构的方法，通过互联网将资源以“按需服务”的形式提供给用户
- 利用互联网连接的数据中心和服务器进行**高效计算**和**信息存取**的系统,使计算能力可以象电能一样提供给客户（高度可扩展）
- 不同于以往的高性能计算：
 - 它除了提供大规模分布式计算外，
 - 还以组织和管理数据为核心工作之一，它获取并且维护持续变化的数据集
 - 提供**存储**以及方便操作数据的**编程模式**

云计算 (Cloud Computing)

A Likely Scenario



“云计算”——将所有的计算资源集中起来，并由软件实现自动管理（无需人为参与），并提供应用服务

主要内容

- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务

课程内容

- 高性能计算系统及其结构模型
 - 并行计算机的系统结构模型，对称多处理机（SMP）、大规模并行处理机（MPP）、集群系统（Cluster）和并行计算的性能评测；
- 并行算法设计
 - 并行算法的一般设计策略、基本设计技术和一般设计过程
- 并行程序的设计原理与方法
 - 并行程序设计基础
 - 共享存储编程和分布存储编程
 - GPU编程
 - 并行程序设计环境与工具
- 云计算
 - 分布式大规模数据处理，云存储，云计算平台
- 高性能计算和云计算应用及发展趋势

课程要求

- 课堂讲授+上机实践
 - 共享存储/分布存储并行编程
 - GPU编程/MapReduce编程
- 了解和掌握高性能计算和云计算的基本原理、技术及最新研究成果，具有高性能计算和云计算的理论基础和实践能力
- 应用高性能计算技术解决实际问题
- 评分：平时作业 (10%)，实验 (40%)，期末笔试(50%)

教学日历

周次	日期	题目
1	9/7	引言
2	9/17	并行计算机体系结构
3	9/21	并行计算性能评测
4	9/28	并行算法设计(1)
6	10/12	并行算法设计(2)
7	10/19	并行程序设计基础
8	10/26	共享存储编程
9	11/2	消息传递编程

周次	日期	题目
10	11/9	GPU编程
11	11/16	Map/Reduce编程
12	11/23	云计算平台
13	11/30	实验1： 共享存储/分布存储编程
14	12/7	
15	12/14	实验2： GPU编程 /MapReduce编程
16	12/21	

注：其中第5周国庆放假

学习用书

- 课本：
 - 陈国良， 并行计算——结构. 算法. 编程（第三版）， 高等教育出版社， 2011
- 参考书：
 - Georg Ha等著， 张云泉等译， 高性能科学与工程计算， 机械工业出版社， 2014
 - 林伟伟等， 云计算与大数据技术理论及应用， 清华大学出版社， 2019

课程学习辅助系统

- 华南理工大学的教学在线 & QQ群
 - <http://eonline.jw.scut.edu.cn/meol/index.do>



主要内容

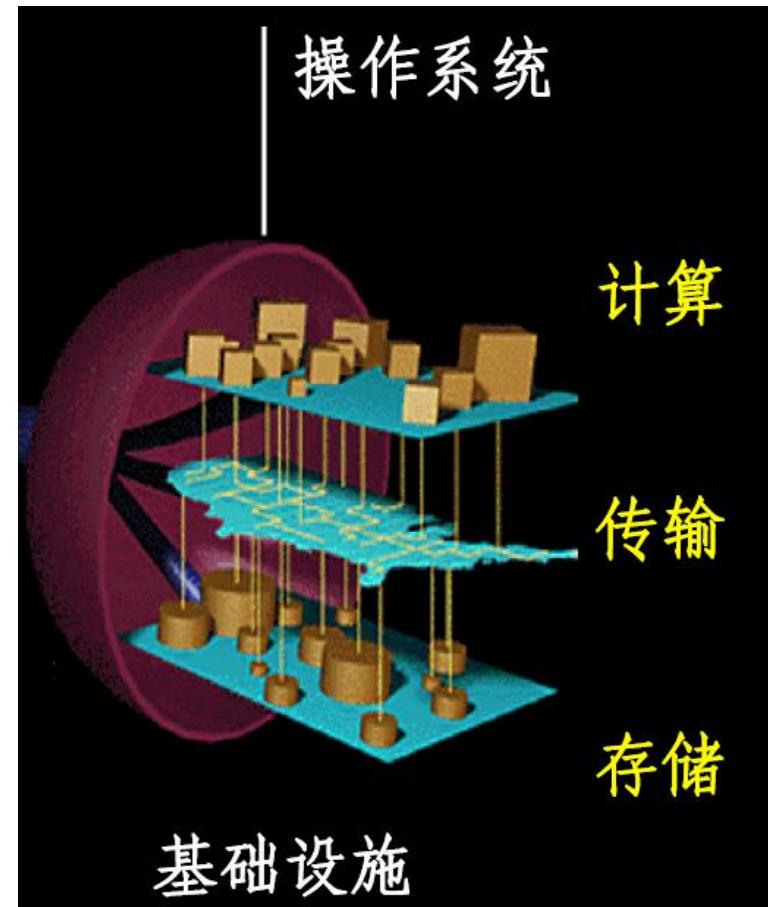
- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务

高性能计算相关领域

- **高性能计算和通信** (High-Performance Computing and Communications: HPCC)
 - 分布式高性能计算、高速网络和Internet的使用
- **并行计算** (Parallel Computing)
 - 使用多处理器系统的高性能计算
- **分布式高性能计算** (Distributed High Performance Computing)
 - 对等 (Peer to Peer, P2P) 计算
 - 网格计算 (Grid Computing)
- **云计算** (Cloud Computing)
 - 共享基础架构，将巨大的系统池连接在一起以提供各种服务

高性能计算和云计算的基础

- **计算 (Compute)** : 数据处理的能力
 - CPU主频, CPU/核数
 - 浮点计算能力 (Flops/s)
- **存储 (Storage)** : 数据存储的能力
 - 缓存、内存、硬盘、磁带等
 - 每秒读写的字节数 (Mbytes/s)
- **通信 (Communications)** : 数据通信的能力
 - 内部网络、局域和广域网络
 - 每秒传输的比特数(Bits/s)



基本测度单位

	计算单位	描述	存储单位	描述	通信单位	描述
Mega 10^6	Mflop/s	每秒百万次浮点计算操作	MByte	兆字节	Mbits/s	每秒兆比特
Giga 10^9	Gflop/s	每秒10亿次	GByte	吉字节	Gbits/s	每秒千兆比特
Tera 10^{12}	Tflop/s	每秒万亿次	TByte	太字节	Tbis/s	每秒太比特
Peta 10^{15}	Pflop/s	每秒1千万亿次	PByte	拍字节	Pbis/s	10^{15} bits/sec
Exa 10^{18}	Eflop/s	每秒100亿亿次	EByte	艾字节	Ebis/s	10^{18} bits/sec
Zetta 10^{21}	Zflop/s	每秒十万亿亿次	ZByte	泽字节	Zbis/s	10^{21} bits/sec
Yotta 10^{24}	Yflop/s	每秒1亿亿亿次	YByte	尧字节	Ybis/s	10^{24} bits/sec

- Flops (floating point operations) : 浮点计算操作

高性能计算机

- 由多个计算单元组成，运算速度快、存储容量大、可靠性高的计算机系统
- 也称为：巨型计算机、超级计算机
- 目前任何高性能计算和超级计算都离不开使用并行技术，所以高性能计算机肯定是并行计算机。



Flynn分类

- 基于指令（**instruction**）和数据流（**data streams**）(1972)
 - 单指令单数据流：SISD (**S**ingle **I**nstruction **s**tream over a **S**ingle **D**ata **s**tream)
 - 单指令多数据流： SIMD (**S**ingle **I**nstruction **s**tream over **M**ultiple **D**ata **s**treams)
 - 多指令单数据流： MISD (**M**ultiple **I**nstruction **s**treams over a **S**ingle **D**ata **s**tream)
 - 多指令多数据流： MIMD (**M**ultiple **I**nstruction **s**treams over **M**ultiple **D**ata **s**treams)
- 普及程度：
 - MIMD > SIMD > MISD

SISD (Single Instruction Stream Over A Single Data Stream)

- SISD

➤ 通用的串行机

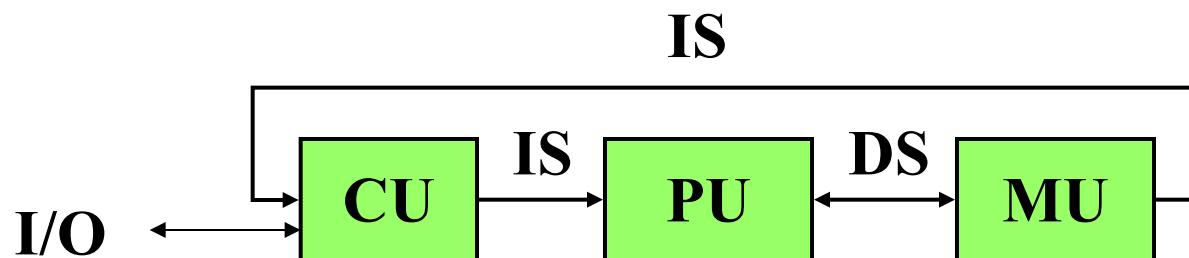
IS : 指令流 (Instruction Stream)

DS : 数据流 (Data Stream)

CU : 控制单元 (Control Unit)

PU : 处理单元 (Processing Unit)

MU : 存储单元 (Memory Unit)



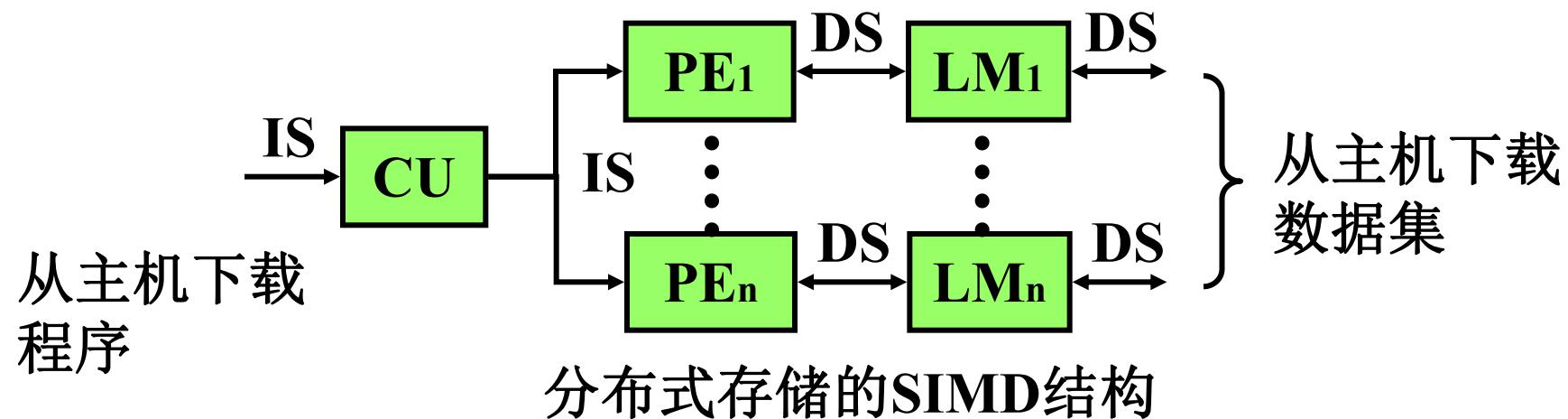
SIMD (Single Instruction Stream Over Multiple Data Streams)

- SIMD

- 矢量机 (Vector computers)
- 专用计算机

PE : 处理模块 (Processing Element)

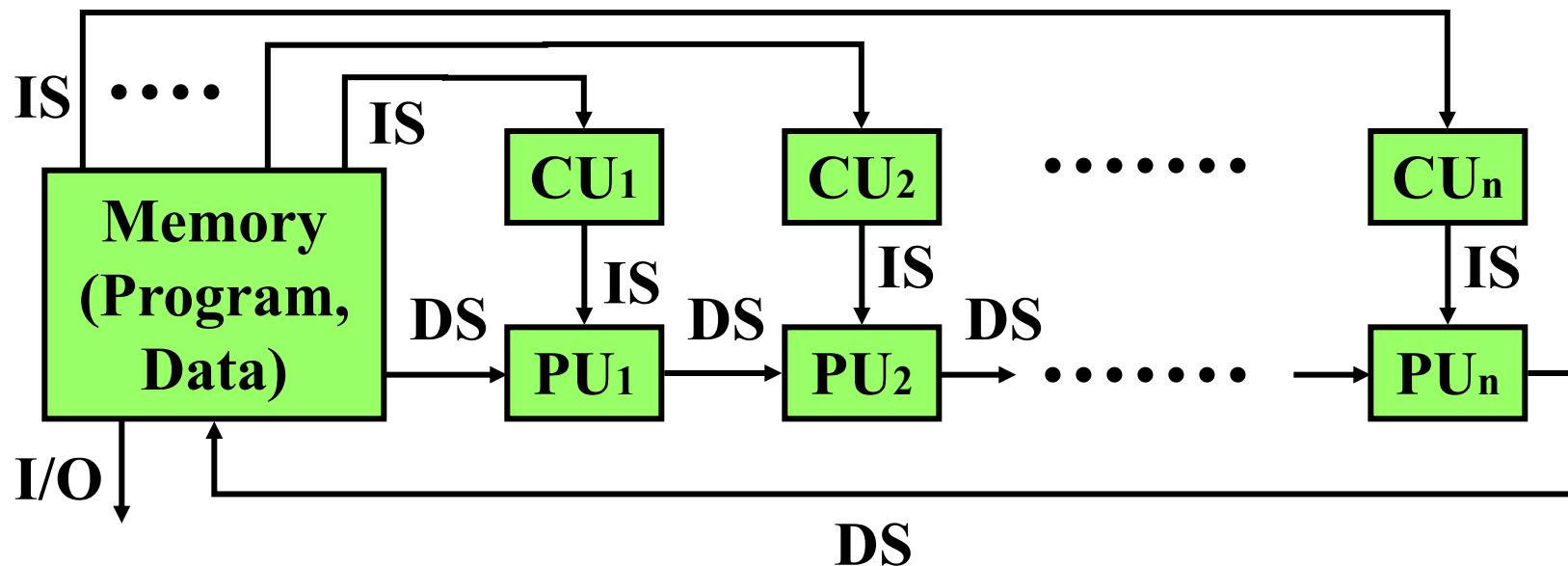
LM : 本地存储 (Local Memory)



MISD (Multiple Instruction Streams Over A Single Data Streams)

- MISD

- 处理器阵列（Processor arrays）、脉动式阵列（systolic arrays）
- 专用计算机

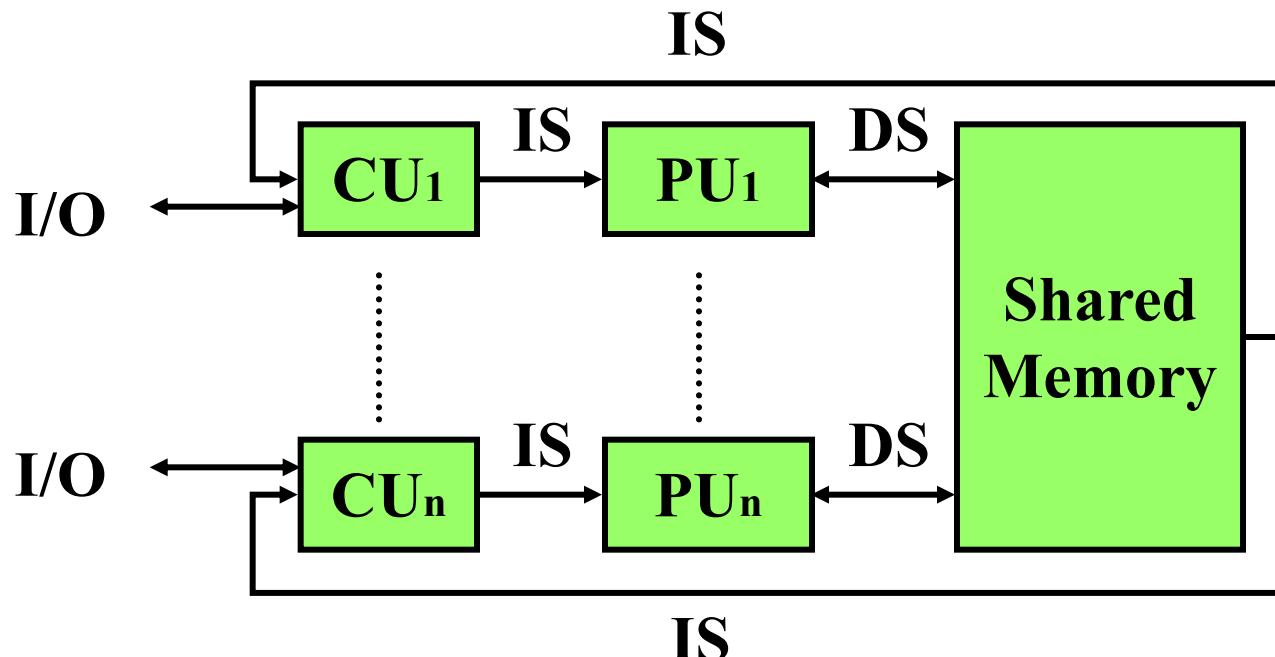


MISD结构（脉动式阵列）

MIMD (Multiple Instruction Streams Over Multiple Data Stream)

- MIMD

- 通用的并行计算机



共享存储的MIMD结构

并行计算机体系结构

- 大部分并行计算机都是**MIMD**系统
 - **PVP:** Parallel Vector Processor
 - **SMP:** Symmetric Multiprocessors
 - **MPP:** Massively Parallel Processors
 - **DSM :** Distributed Shared Memory (DSM)
 - 集群 (Cluster)
 - 分布式系统 (Distributed Systems)
- 可分为两种架构
 - 多处理器 (Multiprocessors) 架构/**共享存储架构**
 - 多计算机 (Multicomputers) 架构/**分布存储架构**

多处理器结构

- 多处理器：共享存储空间(Shared Address Space Architecture)

- PVP (Parallel Vector Processor)

- 高性能的矢量（向量）处理器通过高带宽的交叉开关（crossbar switch）连接在一起

- SMP (Symmetric Multiprocessor)

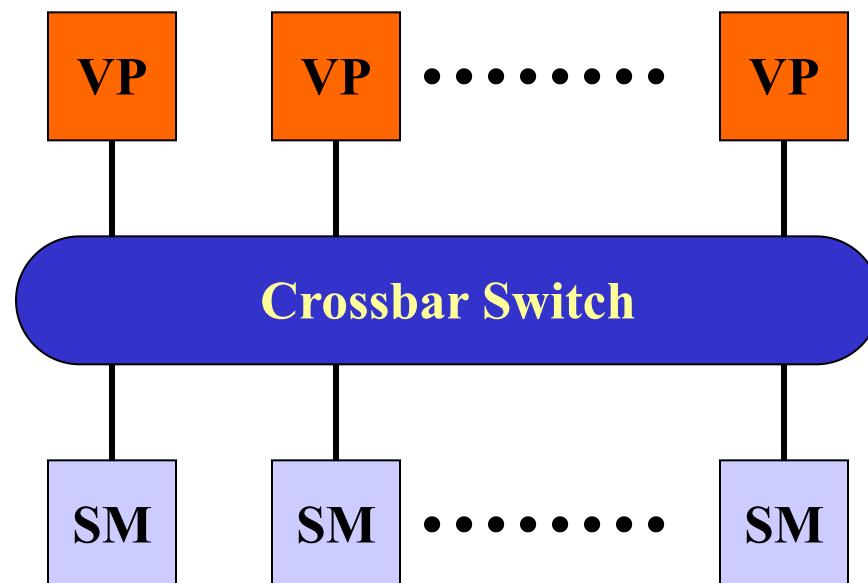
- 商业微处理器（COTS: commercial off-the-shelf）通过高速总线（bus）或交叉开关（crossbar switch）

- DSM (Distributed Shared Memory)

- 与SMP类似，但存储是物理分布在每个节点的

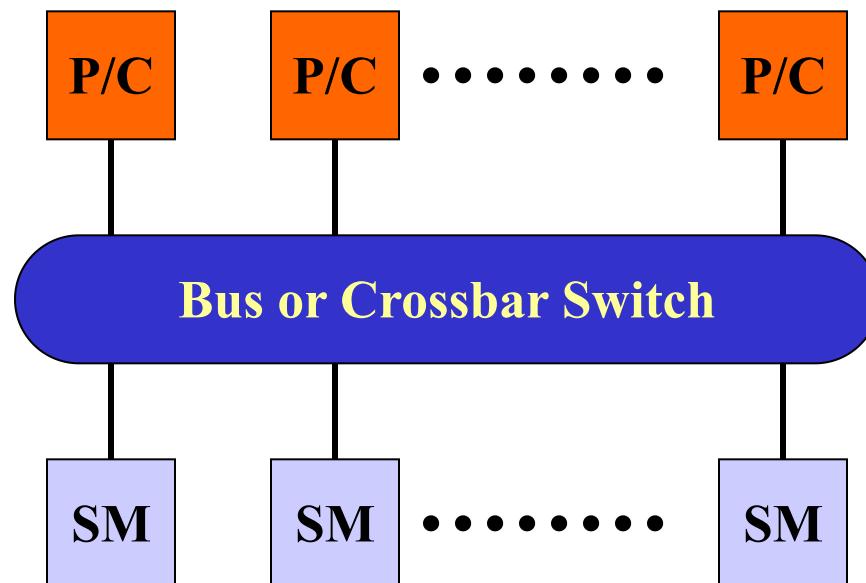
PVP (Parallel Vector Processor)

VP : 矢量处理器 (Vector Processor)
SM : 共享存储 (Shared Memory)



SMP (Symmetric Multi-Processor)

P/C : 微处理器和缓存
(Microprocessor and Cache)

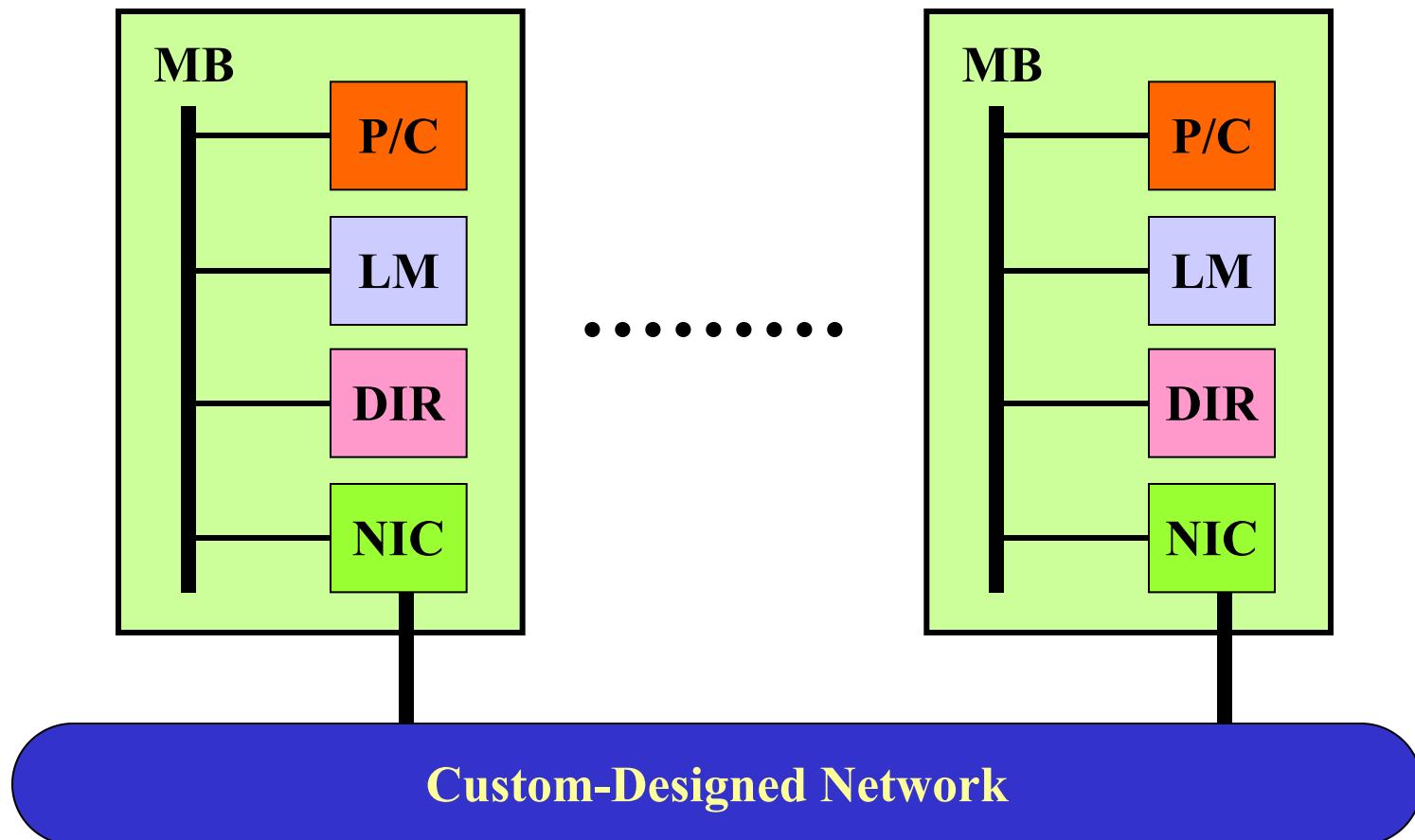


DSM (Distributed Shared Memory)

MB : 存储总线 (Memory Bus) ,

DIR : 缓存目录 (Cache Directory)

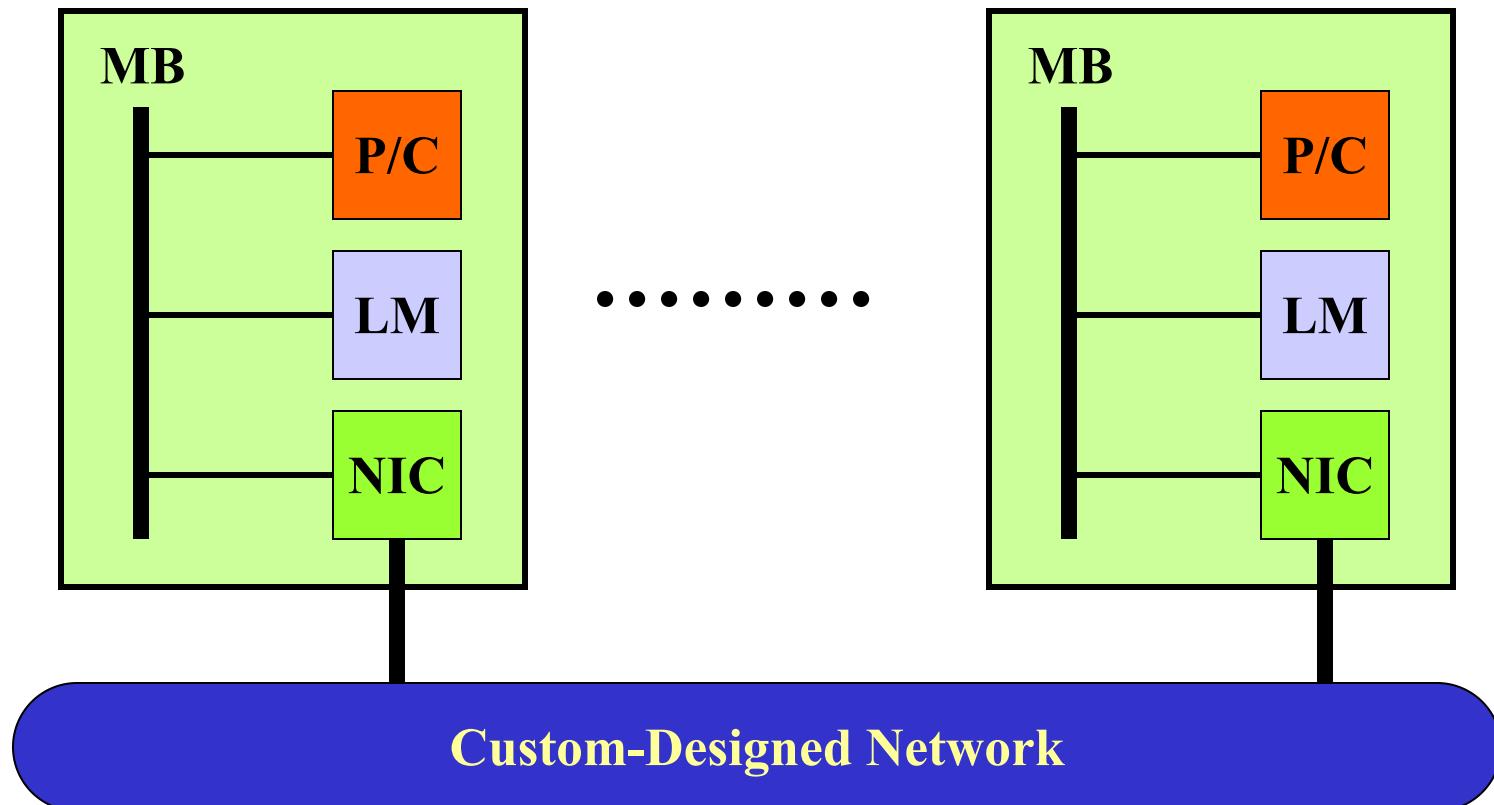
NIC : 网络接口电路 (Network Interface Circuitry)



多计算机架构

- 多计算机架构: 消息传递结构 (Message Passing Architecture)
- MPP (Massively Parallel Processing)
 - 处理器总数目> 1000
- Cluster
 - 系统中的每个节点的处理器少于 16个
- Constellation
 - 系统中的每个节点的处理器多于16个

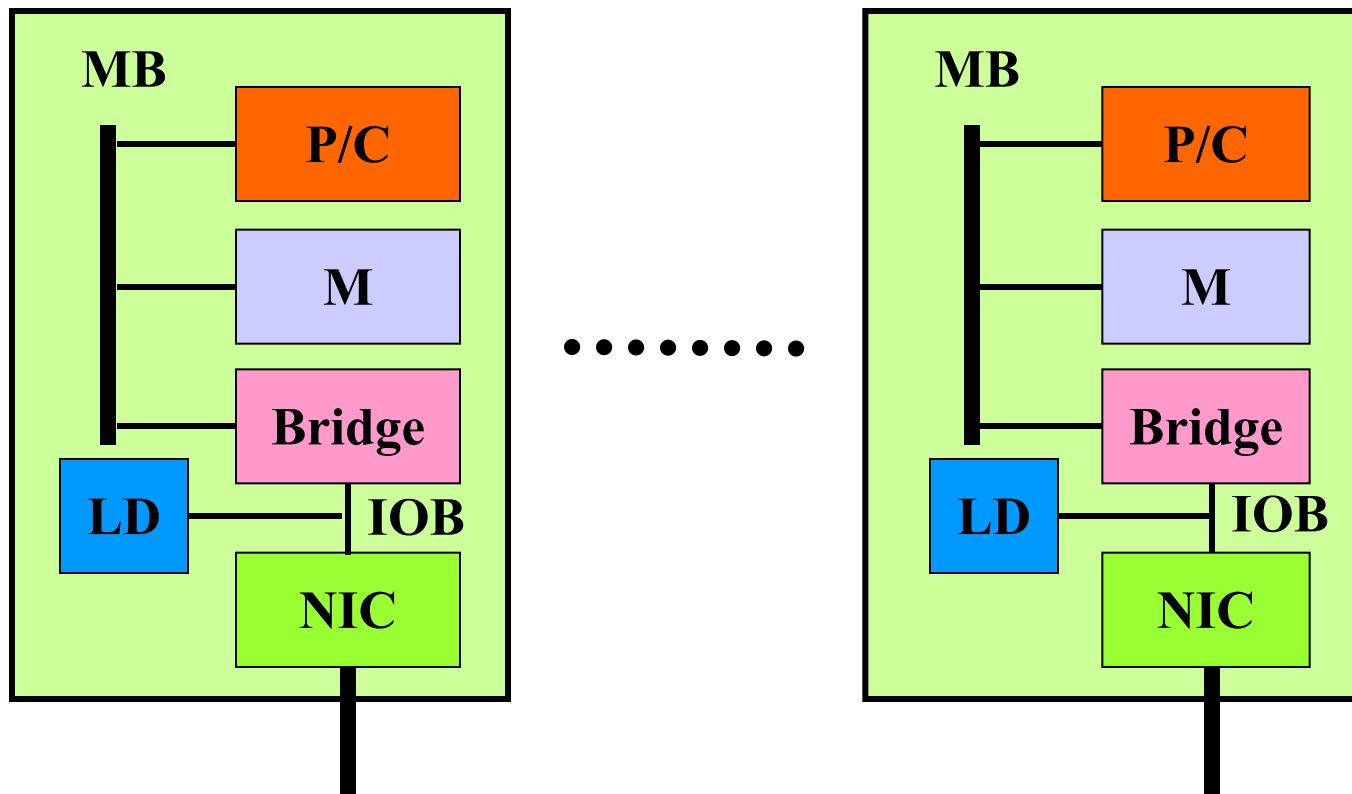
MPP (Massively Parallel Processing)



Cluster

LD : 本地硬盘 (Local Disk)

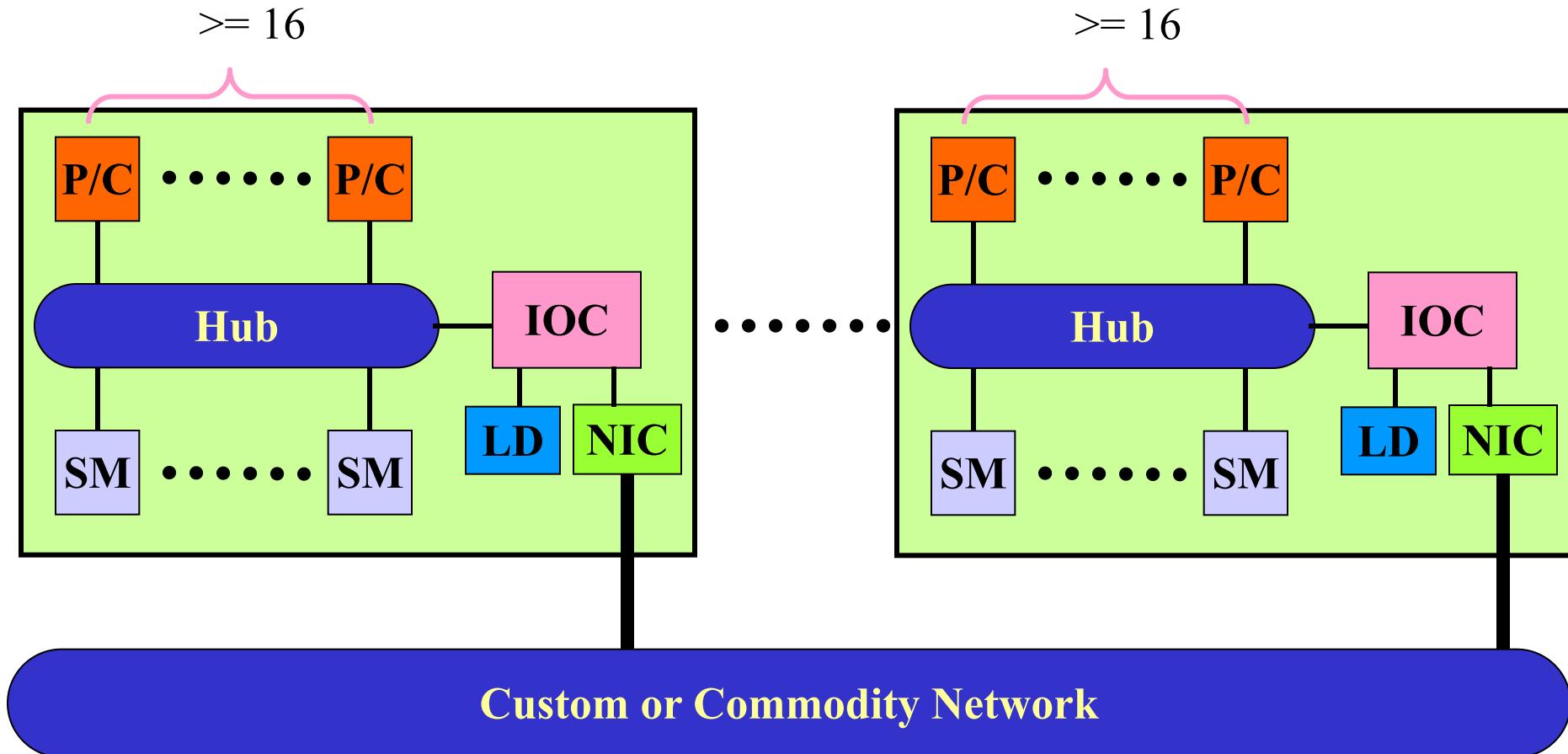
IOB : I/O总线 (I/O Bus)



Commodity Network (Ethernet, ATM, Myrinet, VIA)

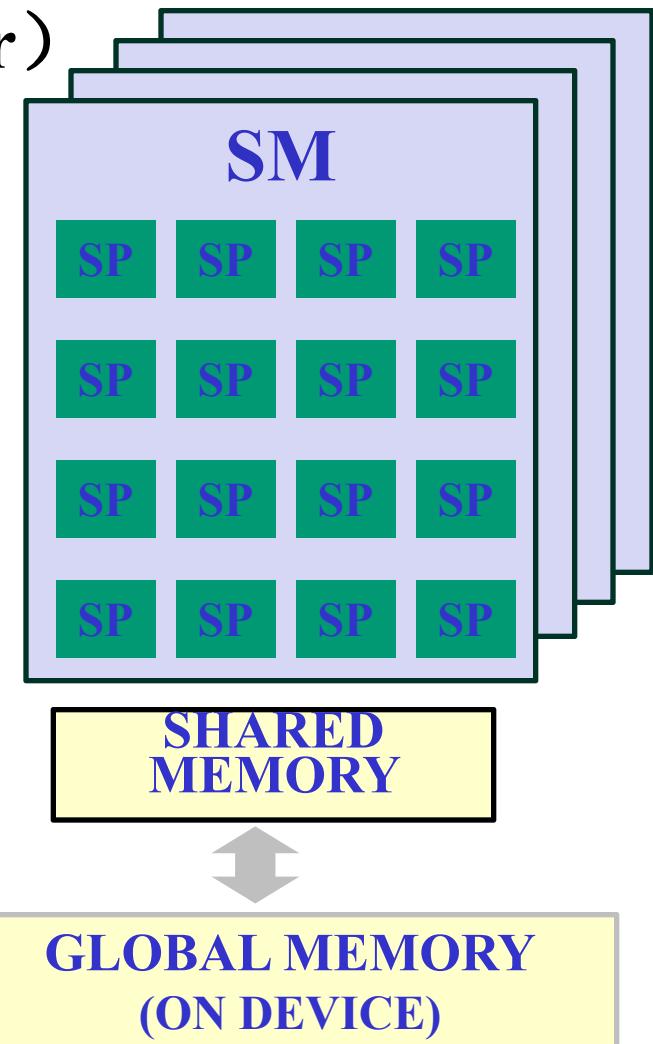
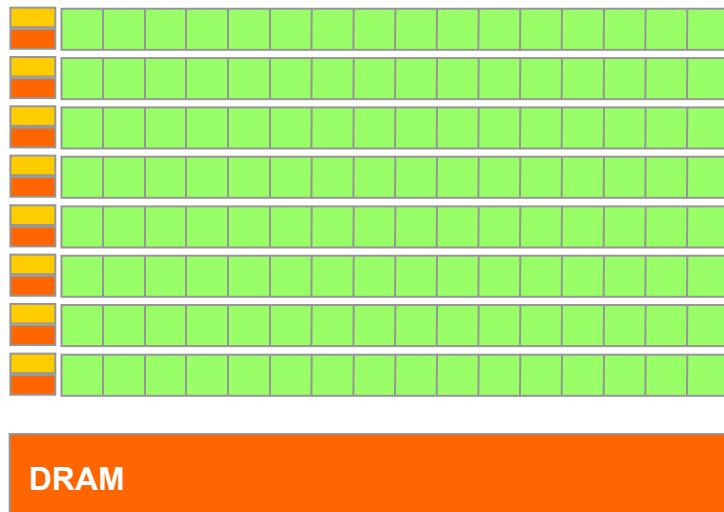
Constellation

IOC: I/O控制器 (I/O Controller)

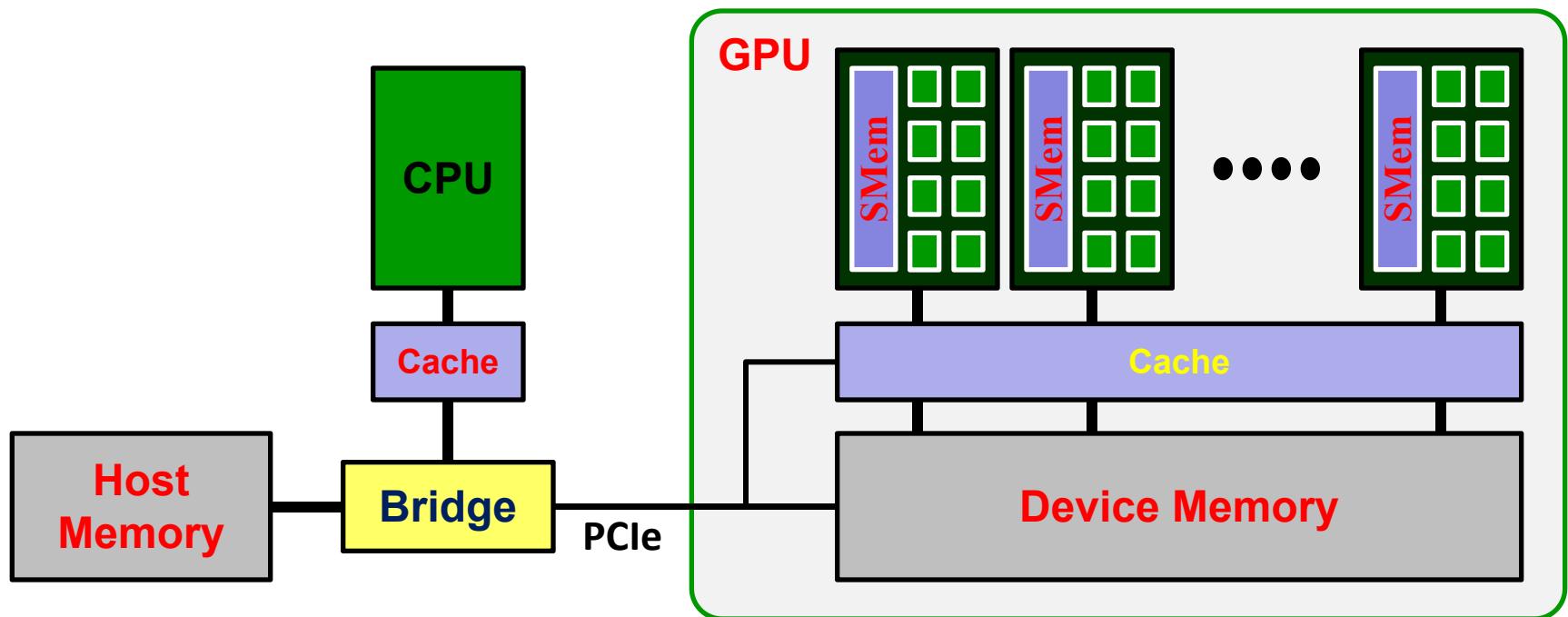


GPU体系结构

- SM (streaming multiprocessor)
：流多处理器
- SP: 流处理器



异构多核系统



主要内容

- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务

高性能计算的发展阶段

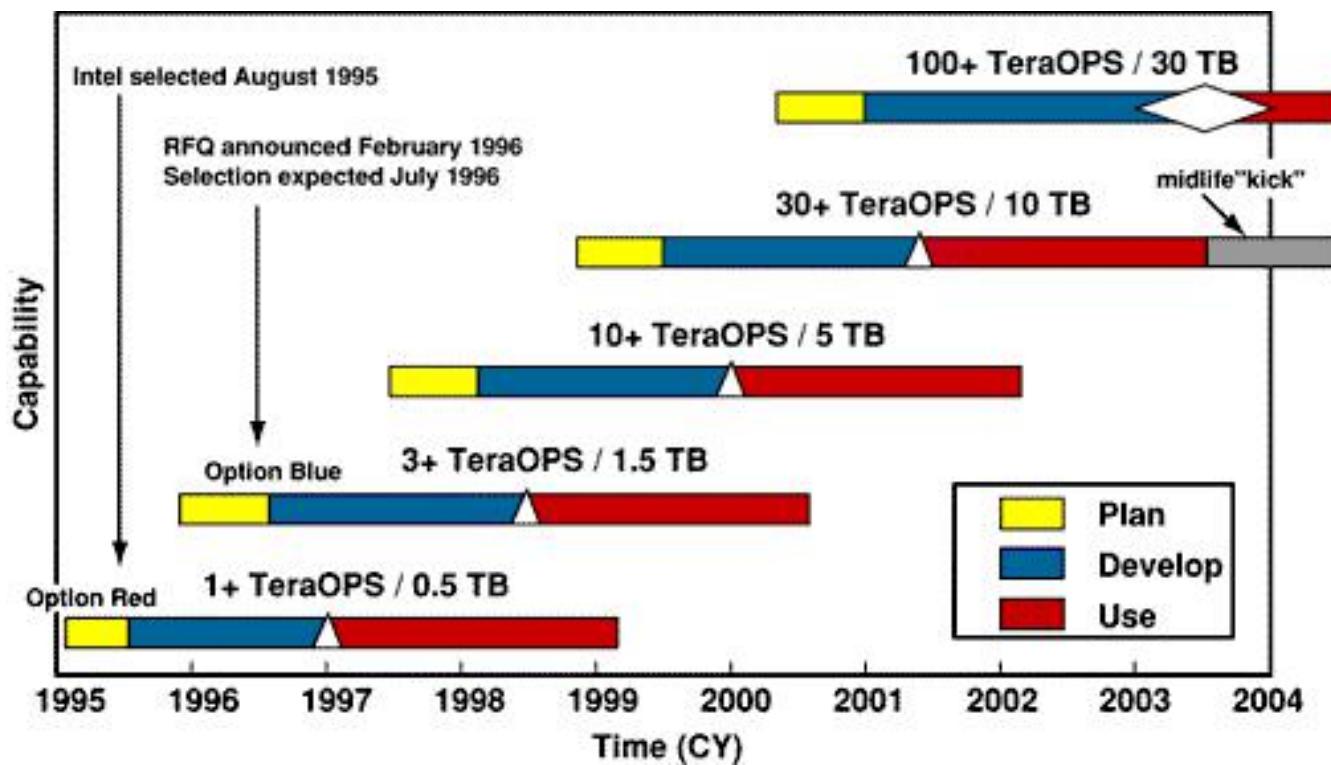
- 专用时代
 - 包括向量机，MPP系统，SMP系统。称为“专用”，是指它们的组成部件如CPU板，内存板，I/O板，操作系统，I/O系统都是专门设计的，属于高端系统，用户群窄小
- 普及时代
 - 商品化趋势使得可批量生产的商品部件接近了高性能计算机专有部件
 - 标准化趋势使得这些部件之间能够集成一个系统中，其中X86处理器、以太网、内存部件、Linux都起到决定性作用
 - 高性能计算机价格下降，应用门槛降低，应用开始普及
 - 以集群为代表

世界HPC计划

- 1993年美国HPCC（High Performance Computing & Communication，高性能计算和通信计划）：3T性能目标
- 1996年美国ASCI（Accelerated Strategic Computing Initiative，加速战略计划创新）计划：
 - 为主要设备供应商制造超大（Teraflop）计算机系统，主要用于仿真核武器实验
- 2000年HPCC二期，1Pflops (10¹⁵)计算能力
 - 显示并应用高性能计算环境来扩展我们的认识和能力，预测影响地球、太阳系和宇宙的物理、化学和生物过程
- 欧洲：ESPRIT, Alvey, Parallel Applications Programme, Europort, Fourth Framework etc
- 日本：日本的HPC计划，造就了许多大型的并行矢量机（NEC, Hitachi, Fujitsu）
- 澳大利亚APSC（Partnership for Advanced Computing）计划：提供HPC的设备和培训

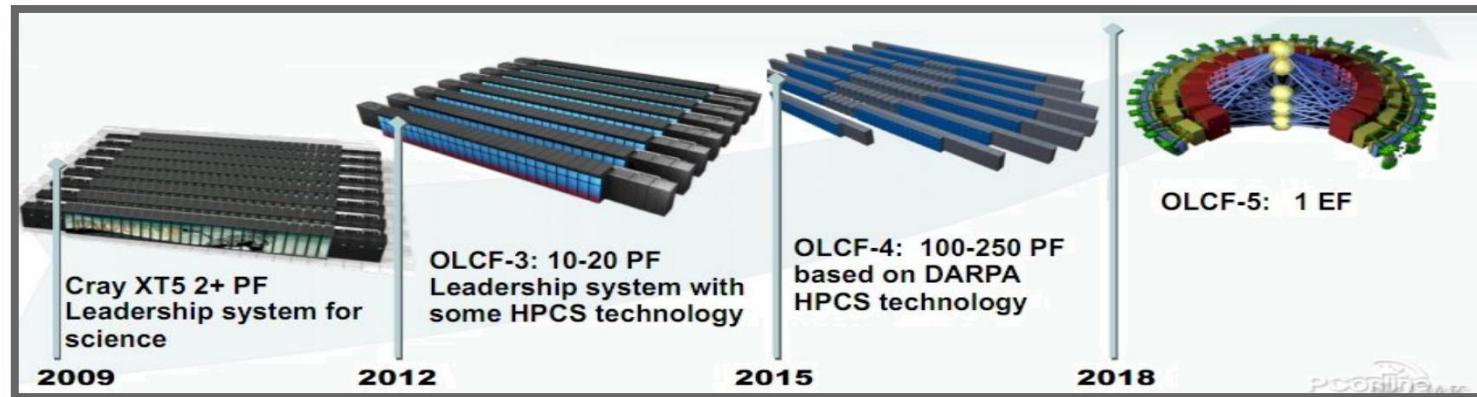
美国ASCI计划

- ASCI: Accelerated Strategic Computing Initiative (<http://www.llnl.gov/asci/>)



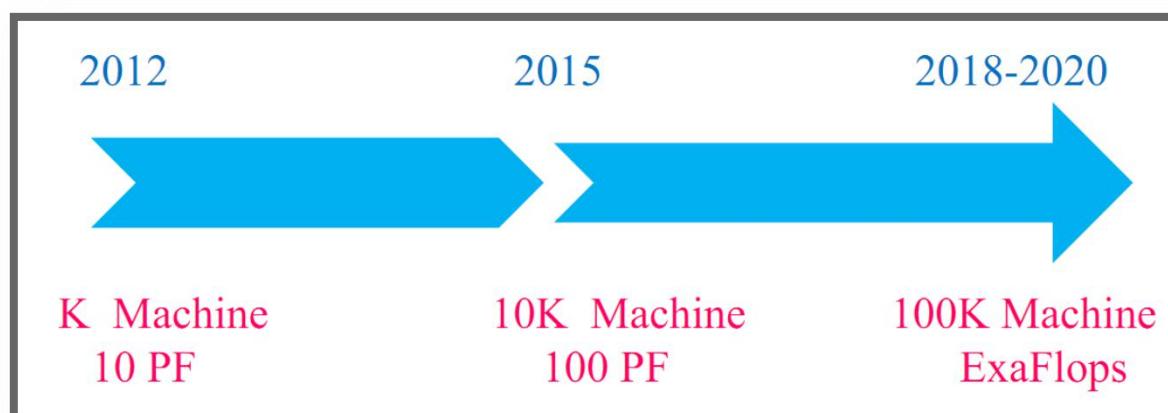
美国E级计划

- 美国总统奥巴马在“Strategy for American Innovation”计划中，将E级计算列为21世纪美国最主要技术挑战
- 美国国防部高级研究计划局DARPA提出：研究新的计算架构和编程模型，于2018年完成E级原型系统
 - 推迟到2020年以后
 - 目前技术还无法完成指标：20MW+ExaFlops



日本HPC计划

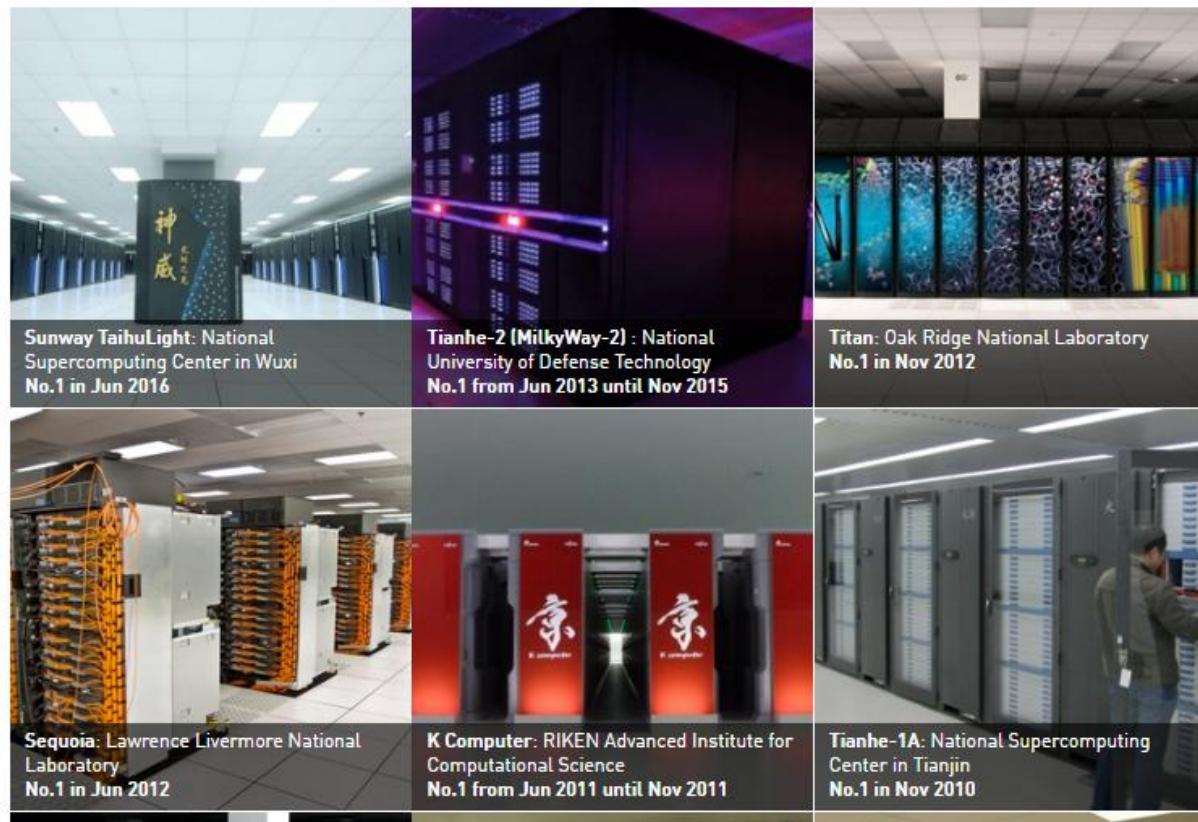
- 日本政府在K系统后
 - JST-CREST (Core Research for Evolutional Science and Technology) 计划
 - SDHPC (Workshop on Strategic Development of High Performance Computers) 计划
 - FS (Feasibility Study) 计划



TOP 500 (www.top500.org)

TOP #1 SYSTEMS

In the last 20 years, the following systems made it to the top of the TOP500 lists:



列出世界500强超级计算机，每年更新两次

TOP 10 (2020.6)

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,299,072	415,530.0	513,854.7	28,335
2	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
5	Tianhe-2A - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000, NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
6	HPC5 - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, Dell EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252

R_{max} : 实测性能 (Tflops)

R_{peak} : 理论峰值 (Tflops)

Power : 电源消耗 (KW)

中国高性能计算机发展（1）

- 1983年，“银河 I 号”巨型计算机研制成功，运算速度达每秒1亿次
- 1984年，中国第一台10亿次巨型银河计算机 II 型通过鉴定
- 1995年，曙光1000大型机通过鉴定，其峰值可达每秒25亿次
- 2000年，“神威-I”高性能计算机问世，峰值达每秒384亿次。我国成为继美国、日本之后，世界上第三个具备研制高性能计算机能力的国家。（**1998年11月美国已研制出每秒3.1万亿次机**）
- 2002年8月，中科院第一台每秒万亿次的超级计算机联想深腾1800问世。在当年TOP500中排名43。（**2001年美国已研制出每秒12.8万亿次机**）
- 2003年11月联想公司研制的“联想深腾6800高性能计算平台”，系统峰值5.3万亿次，当年Top500排名14（**2002年4月，日本NEC公司研制出当时世界上运算速度最快的超级计算机“地球模拟器”，运算峰值为40万亿次**）

中国高性能计算机发展（2）

- 2004年6月曙光公司的研制的“曙光4000A”，系统峰值10万亿次，当年Top500排名10
- 2008年8月，曙光百万亿次超级计算机“曙光5000”研制成功，成为第二个可研制**百万亿次**超级计算机的国家
- 2009年10月，国防科大“天河一号”研制成功，在Top500排名第五
- 2010年5月，曙光“星云”研制成功，在Top500排名第二，成为第二个可研制**千万亿次**超级计算机的国家
- 2010年11月，国防科大“天河一号A”研制成功，在**Top500排名第一**
- 2011年6月，国防科大“天河一号A”在Top500排名第二
- 2012年9月，**广州超算中心**启动
- 2013年6月～2015年11月，广州超算的“天河二号”（每秒33.86千万亿次的浮点运算速度）在**Top500连续排名第一**（六连冠）
- 2016年6月～2017年6月，神威太湖之光在**Top500排名第一**
- 2019年6月，神威太湖之光和天河二号分列三、四名
- 2020年6月，神威太湖之光和天河二号分列四、五名

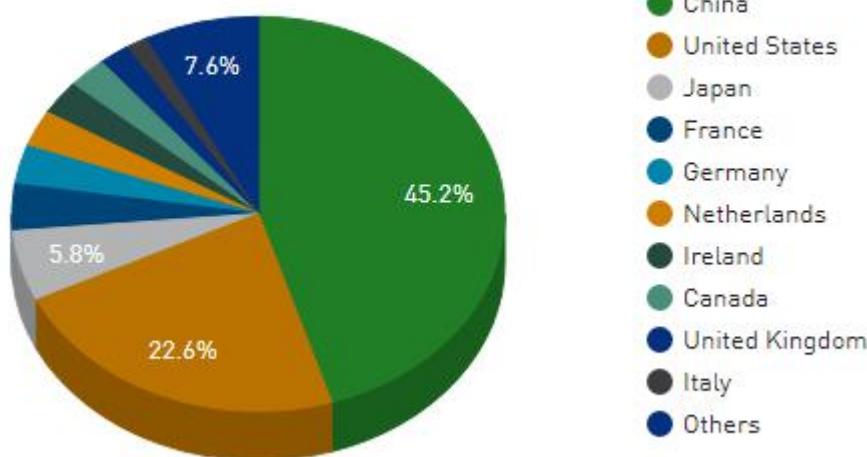
广州天河二号

- 运算速率：理论峰值**54.9PFLOPS**，实际峰值
33.86PFLOPS（3.39亿亿次浮点运算）
- 处理器
 - 16,000个运算节点，每节点配备两个Xeon E5 12核心的中央处理器、三个Xeon Phi 57核心的协处理器，共312万个计算核心
 - 中央处理器：时钟频率为2.2GHZ的Xeon E5-2692 12，峰值性能**0.2112TFLOPS**
 - 协处理器：英特尔集成众核架构的Xeon Phi 31S1P协处理器，运行时钟为1.1GHz，峰值性能为**1.003TFLOPS**
- “天河二号”占地720平方米，造价1亿美元。运算一小时，相当于全国13亿人同时用计算器计算一千年。其系统的存储总容量相当于600亿册每册10万字的书籍。
- 在架构方面，天河二号使用了自主研发的网络系统（TH Express-2）和操作系统（麒麟操作系统）。



各个国家的性能分布 (2020.6)

Countries System Share



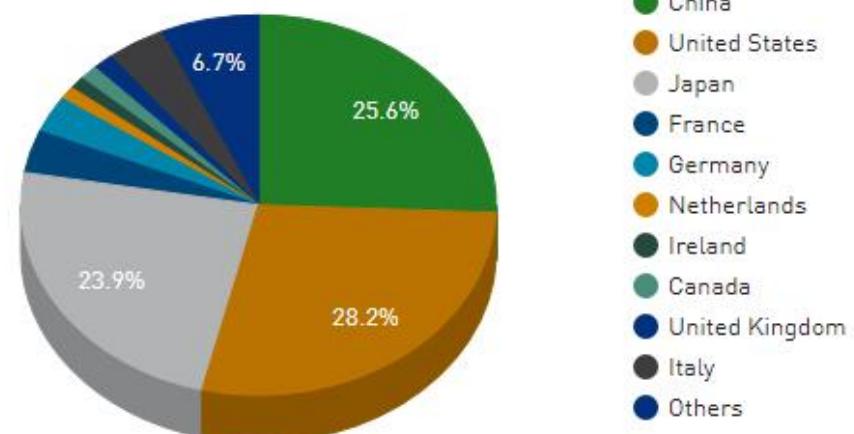
系统数量

美国: 113 (22.6%)

中国: 226 (45.2%)

日本: 29 (5.8%)

Countries Performance Share



性能占比

美国: 621,655,590 (28.2%)

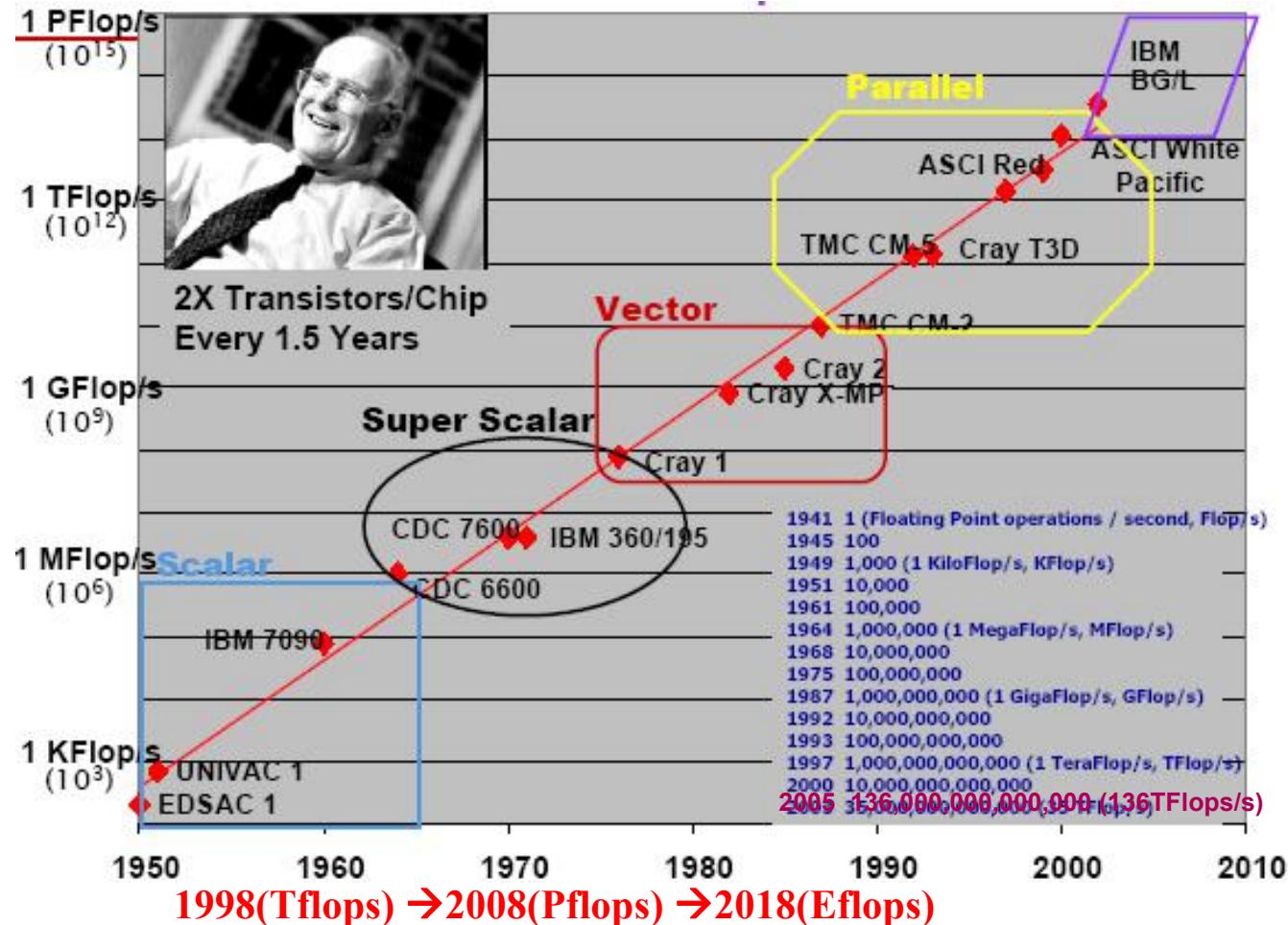
中国: 565,553,102 (25.6%)

日本: 527,607,512 (23.9%)

透过Top500看HPC的发展趋势

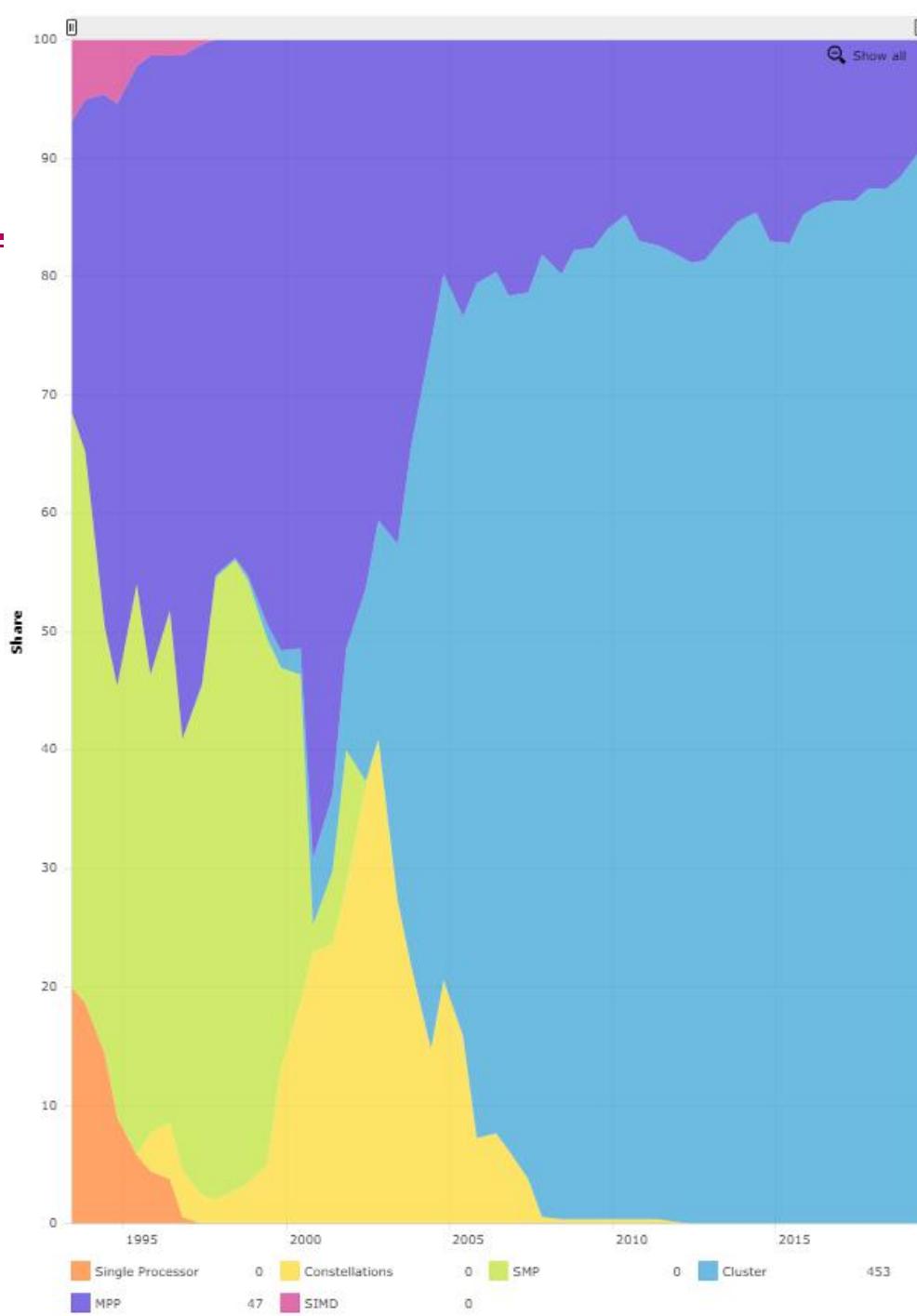


HPC的发展与摩尔定律



世界最快计算机每10年1000倍性能提升

高性能计算机系统体系结构的发展趋势



体系结构相对稳定，Cluster 占有率不断提高

**Cluster: 90.6%, MPP: 9.4%
(2019.6)**

**Cluster: 92.22%, MPP: 7.78%
(2020.6)**

主要内容

- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务
 - 高性能计算
 - 云计算

高性能计算的应用领域

- 科学
 - 天气预报
 - 天体物理
 - 生物：生物形态学， 基因，蛋白质折叠、药物设计
 - 计算生物学
 - 计算材料科学与纳米科学
- 工程
 - (汽车) 碰撞仿真
 - 半导体设计
 - 地震及结构建模
 - 计算流体力学 (飞机设计)
 - 燃烧学 (工程设计)
- 商业
 - 金融和经济构模
 - 事务处理
 - Web服务
 - 搜索引擎
 - 电子商务
 - 网络游戏
 - 动漫/影视制作
- 安全
 - 核武器—通过仿真 (simulations) 来做实验
 - 密码 (Cryptography)

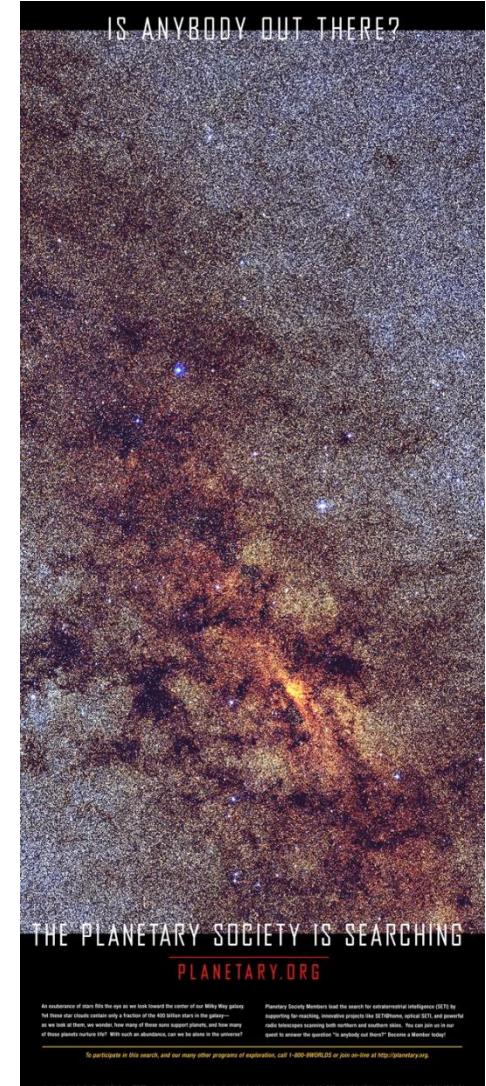
密码解密

- DES的56比特密钥有 $2^{56} = 7.2 \times 10^{16}$ 个可能的值，但强力搜索（brute force search）可能破译。
- 1997年1月28日，美国的RSA数据安全公司在RSA安全年会上公布了一项“秘密密钥挑战”竞赛，其中包括悬赏1万美元破译密钥长度为56比特的DES。美国克罗拉多洲的程序员Verser从1997年2月18日起，用了96天时间，在Internet上数万名志愿者的协同工作下，成功地找到了DES的密钥，赢得了悬赏的1万美元
- 1998年7月电子前沿基金会（EFF）使用一台25万美元的电脑在56小时内破译了56比特密钥的DES
- 1999年1月RSA数据安全会议期间，电子前沿基金会用22小时15分钟就宣告破解了一个DES的密钥
- 在现有的计算水平下，DES已经被宣布为不安全的密码

寻找星外文明：SETI@home

- SETI (Search for Extraterrestrial Intelligence) : 利用全球联网的计算机共同搜寻地外文明的科学实验计划
- 通过大规模并行计算完成来自其它宇宙文明社会电波信号的灵敏搜索
- SETI@home主要集中在检测窄频段信号，根据频段对数据进行分块，这些分块在本质上是相互独立的
- 对太空一个位置的观察得到的结果和另外一个位置得到的结果是相互独立的
- 因此可以把很大的数据集分成大量的小块，每一个计算机能够比较快的分析出其中的一块

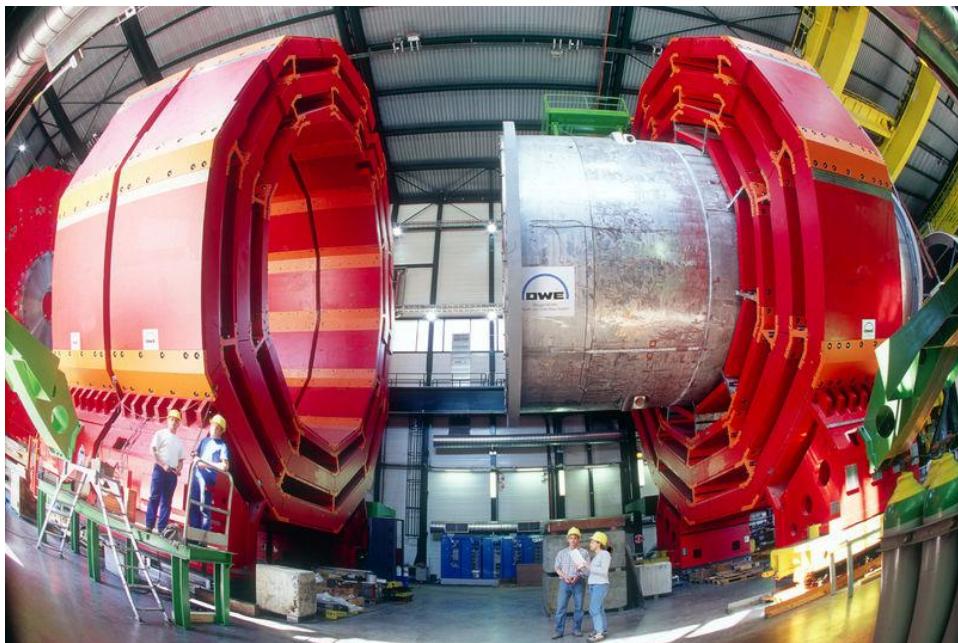
<http://setiathome.berkeley.edu/>



大规模数据分析

欧洲核子研究中心（CERN）

- 干涉重力波天文台（Interferometer Gravitational Wave Observatory : LIGO）：
由于大规模如星球的突然移动造成的空间和时间的微小扭曲
1TB/day (1024 GB/day), Year-long experiments

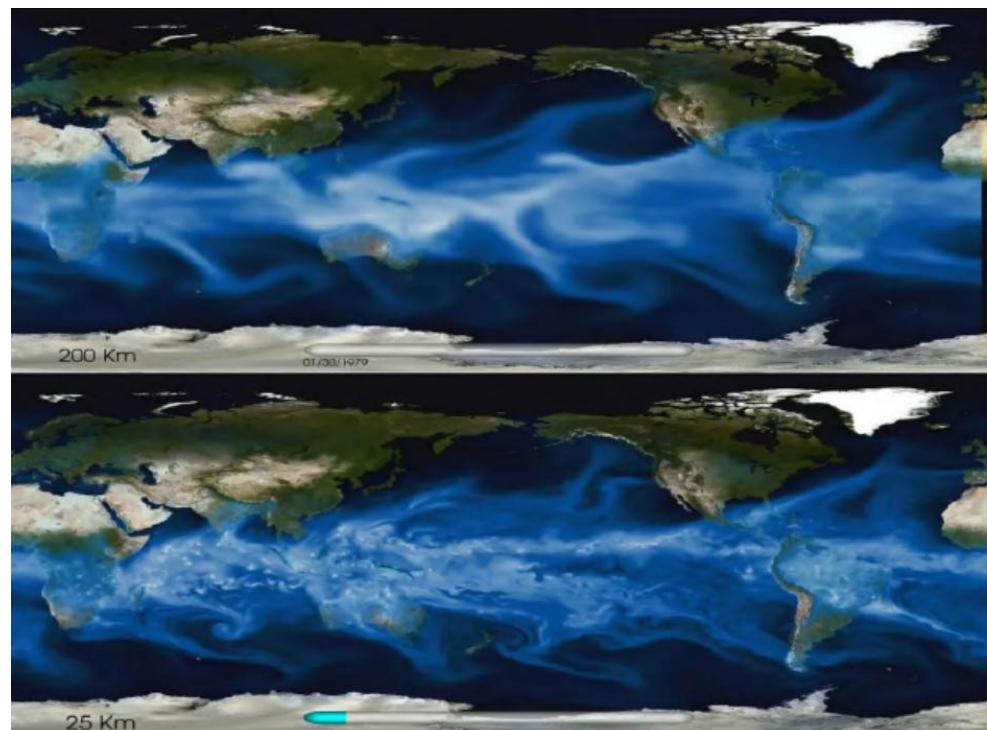


- 粒子探测器（The Compact Muon Solenoid）：用来寻找新粒子

**10 GB/sec!!!
Many PB/year (1024 TB/year)**

天气预报

- 根据大气实际情况，在一定初值和边值条件下，通过大型计算机作数值计算，求解天气变动过程中的流体力学和热力学方程组，来预测一段时间内大气的运动状态和天气现象
- 计算需求估计：
 - 云层模型的网格粒度需要精细到1.5 km以下，模拟时间达到真实时间的1/1000以下，需要200 **PFLOPS** 峰值性能和10TB以上内存



搜索引擎



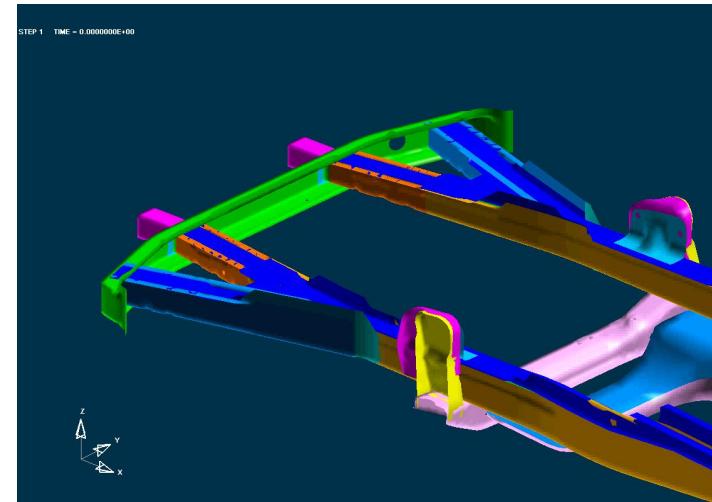
一个简单的检索涉及到.....

- 200+ 处理器
- 200+ TB数据库
- 10^{10} 总的时钟周期
- 0.1秒响应时间
- 5¢ 广告收入



工程高性能计算

- 药物设计
- 新材料
- 石油勘探
- 流体动力学设计
- 碰撞仿真
- 强度设计
- 温度场/电磁场优化
- 建筑
- 工业设计
- 电力调度
-



动漫与影视创作

项目	长片动画影片
制作人员数量	80~100人
影片时长	90分钟
画面数量	90分钟×60秒×24帧=129600
每帧存储空间	12 MB
每部影片存储空间	2 TB
每帧计算时间	60分钟~300分钟
总计算量	5400天（129600小时）~27000天（648000小时）
200节点可完成数	27天~135天
废片率（重复计算）	每帧 3~25次

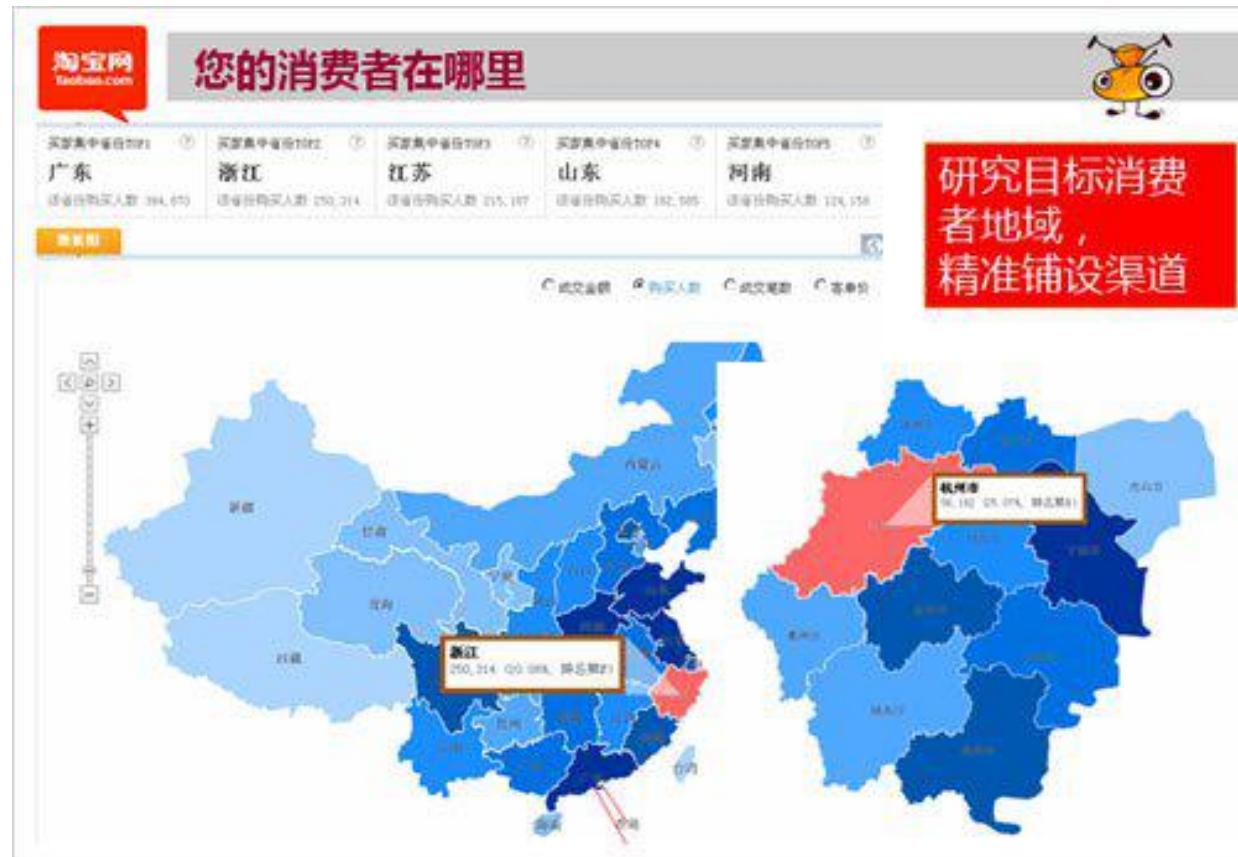
变形金刚：虚拟机器人造型，虚拟场景以及现场实拍特效完美的合成



- 渲染农场（Render Farm）：分布式并行集群计算系统

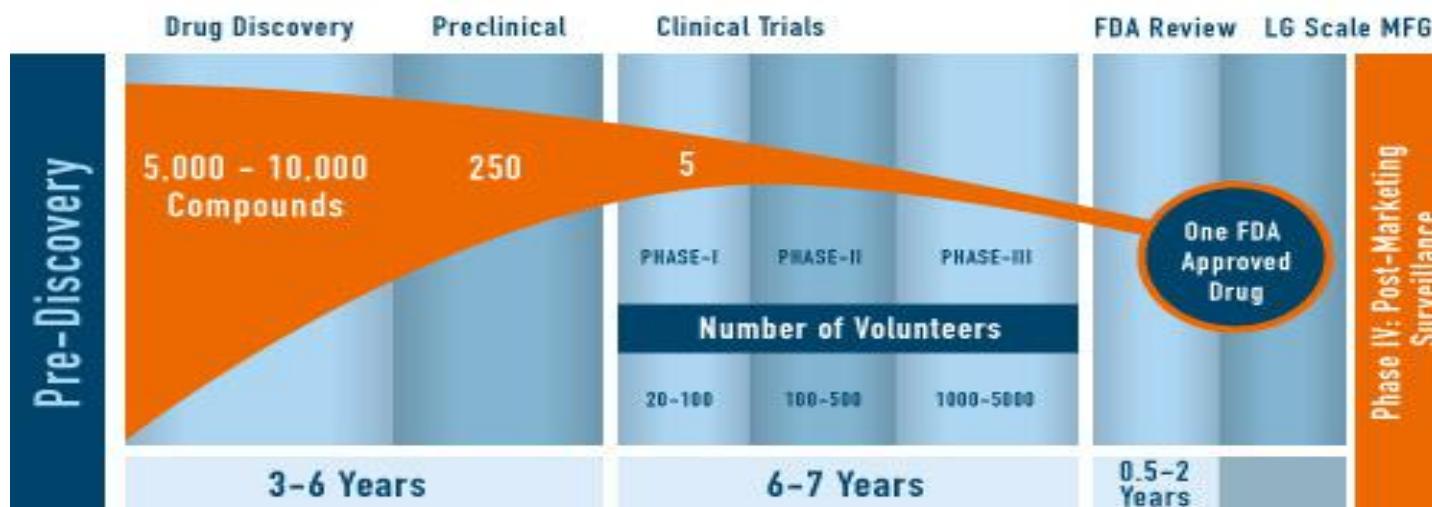
商业智能

- 决策支持
- 风险监测
- 数据挖掘
- 供应链优化
- 用户建模
- 产品推荐



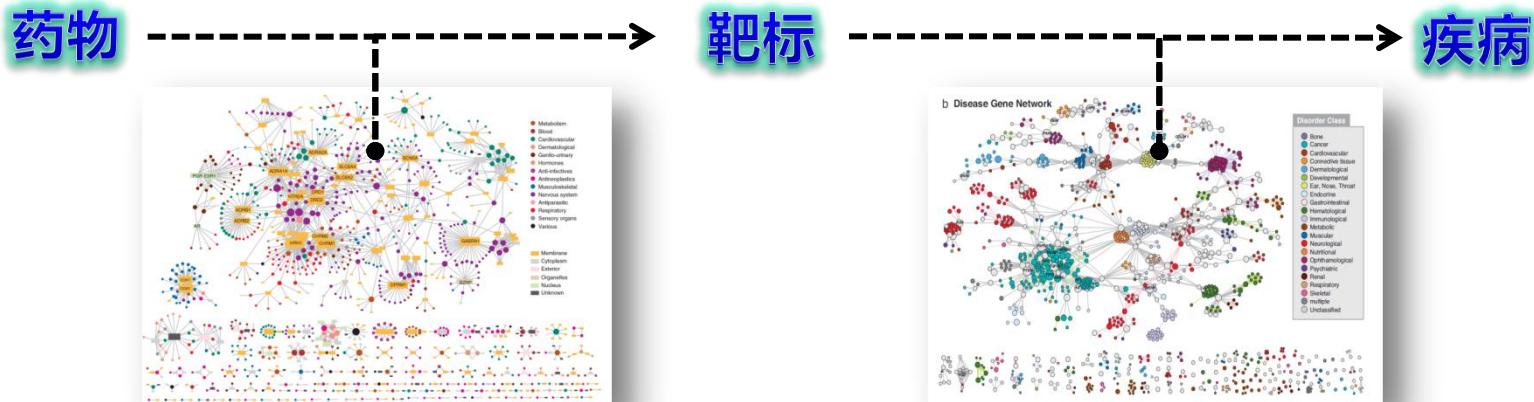
淘宝数据魔方
(淘宝网：每天产生数据量7T)

计算药学



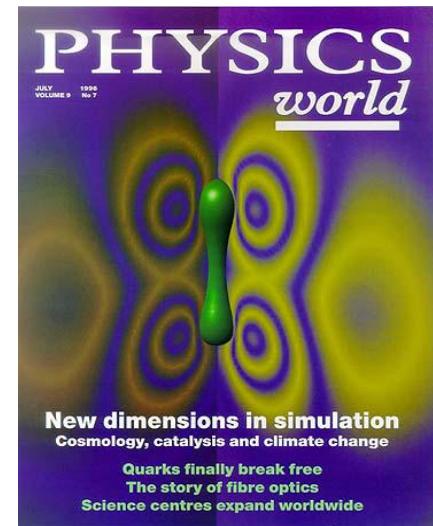
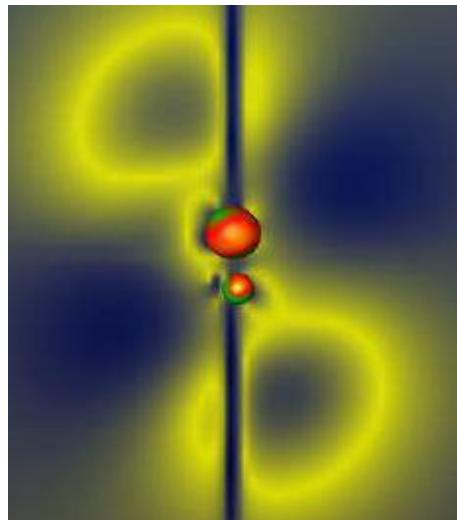
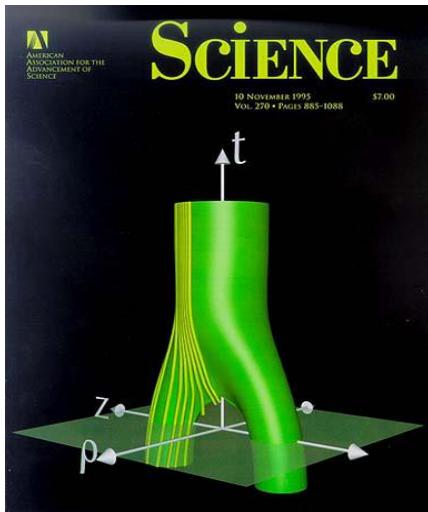
漫长和高风险的药物研发过程

药物重定向：药物-靶标-疾病的复杂关系

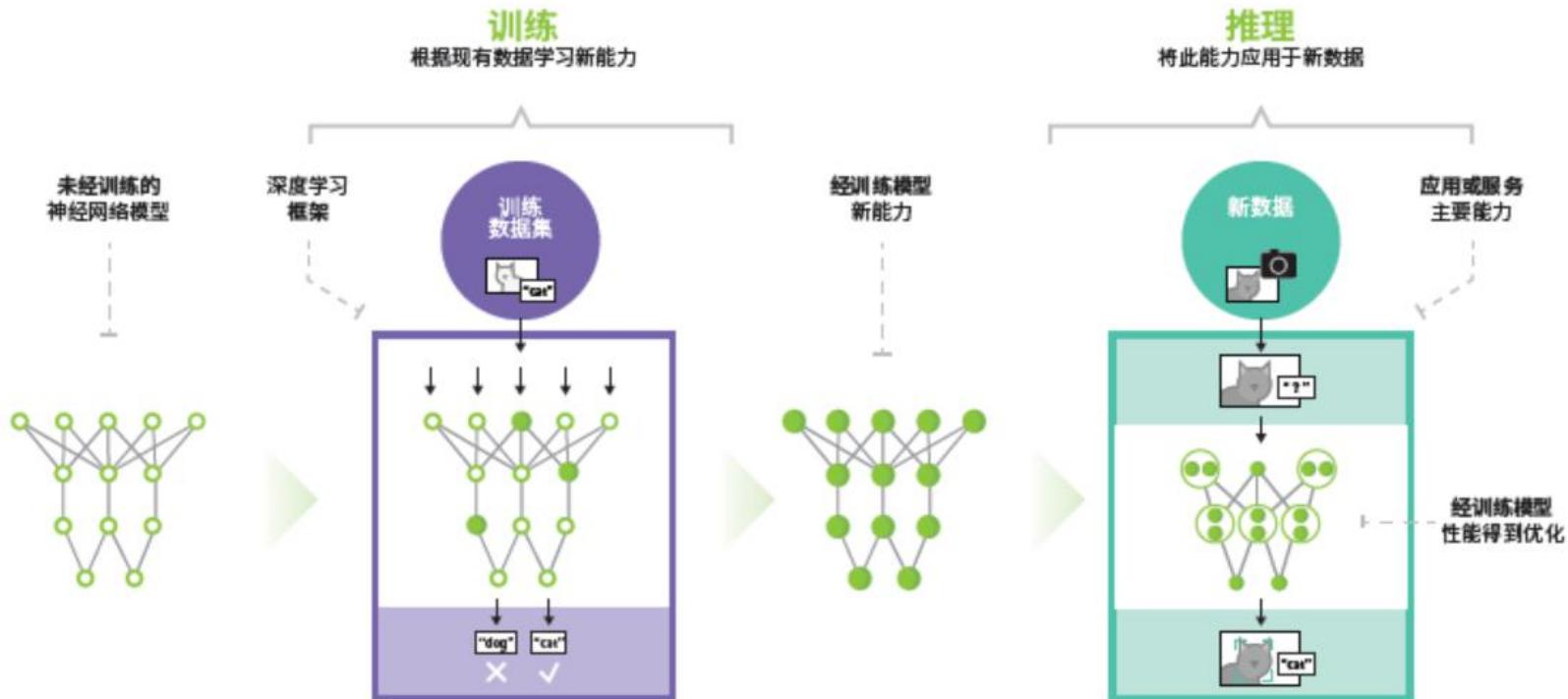


计算科学

- 随着高性能计算的兴起和发展，各科技领域与高性能计算的结合产生了一系列计算性的学科分支：
 - 计算物理、计算化学、计算生物学、计算流体力学、计算地质学、计算气象学(数值天气预报)、计算材料学、计算金融学、计算语言学、计算医学等
 - 还出现了社会计算、情感计算、经济计算等研究领域



GPU加速深度学习



人工智能需要加速计算。在摩尔定律逐渐趋缓的时代，为满足日益增长的深度学习需求，加速器可提供必要的数据处理能力。张量处理是提高深度学习训练和推理性性能的核心技术。

高性能计算应用分类

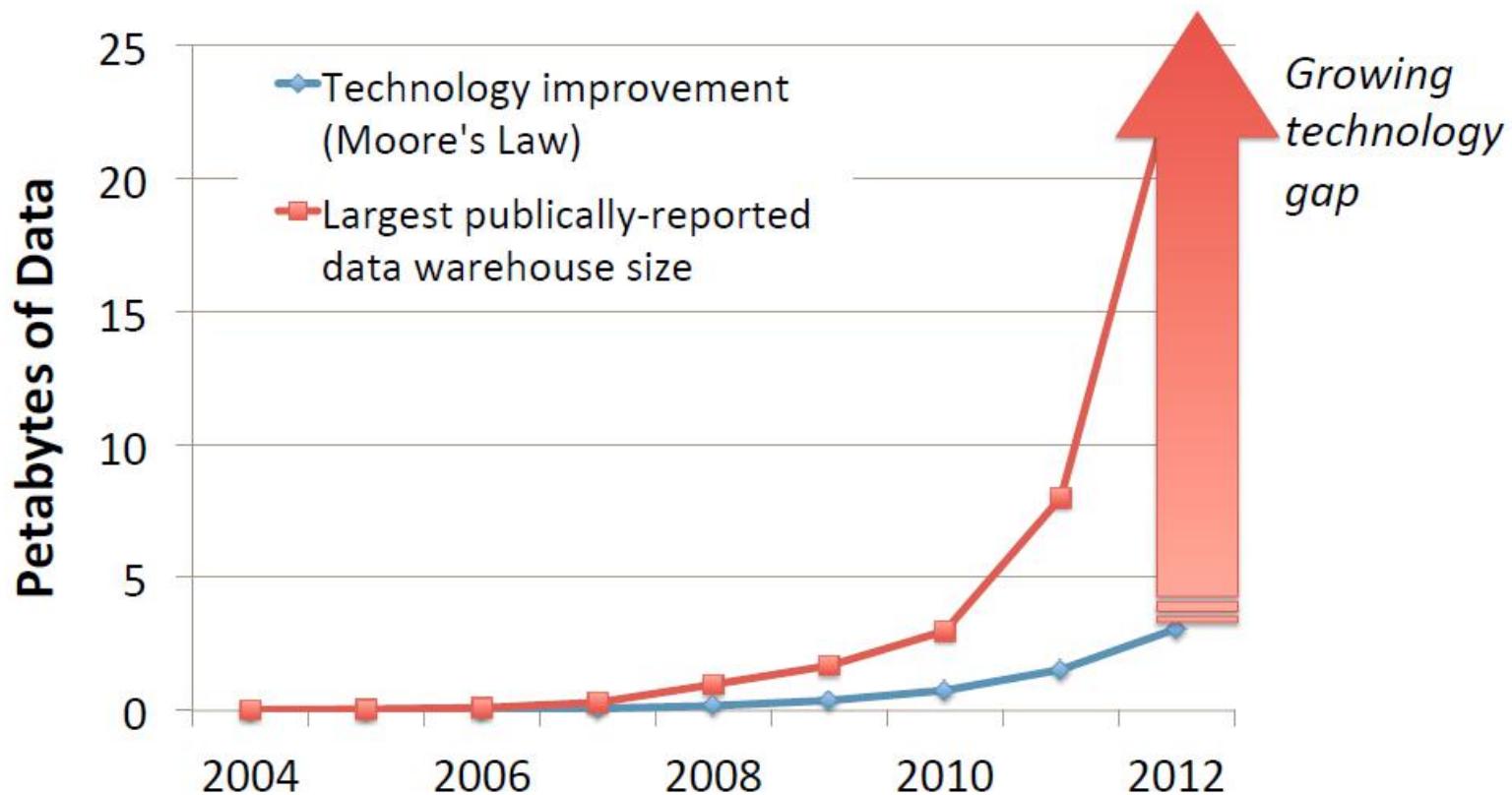
- **计算密集型（Compute-Intensive）应用**
 - 大型科学工程计算与数值模拟
- **数据密集型（Data-Intensive）应用**
 - 搜索引擎、数字图书馆、数据仓库、数据挖掘和计算可视化等
- **网络密集型（Network-Intensive）应用**
 - 协同工作、遥控和远程医疗诊断等

主要内容

- 什么是高性能计算与云计算?
- 课程介绍
- 术语与定义
- 高性能计算系统的发展现状
- 应用及服务
 - 高性能计算
 - 云计算

大数据时代的高性能计算

数据增长速度高于处理器速度的增长



WinterCorp Survey, www.wintercorp.com

大数据对高性能计算提出的挑战

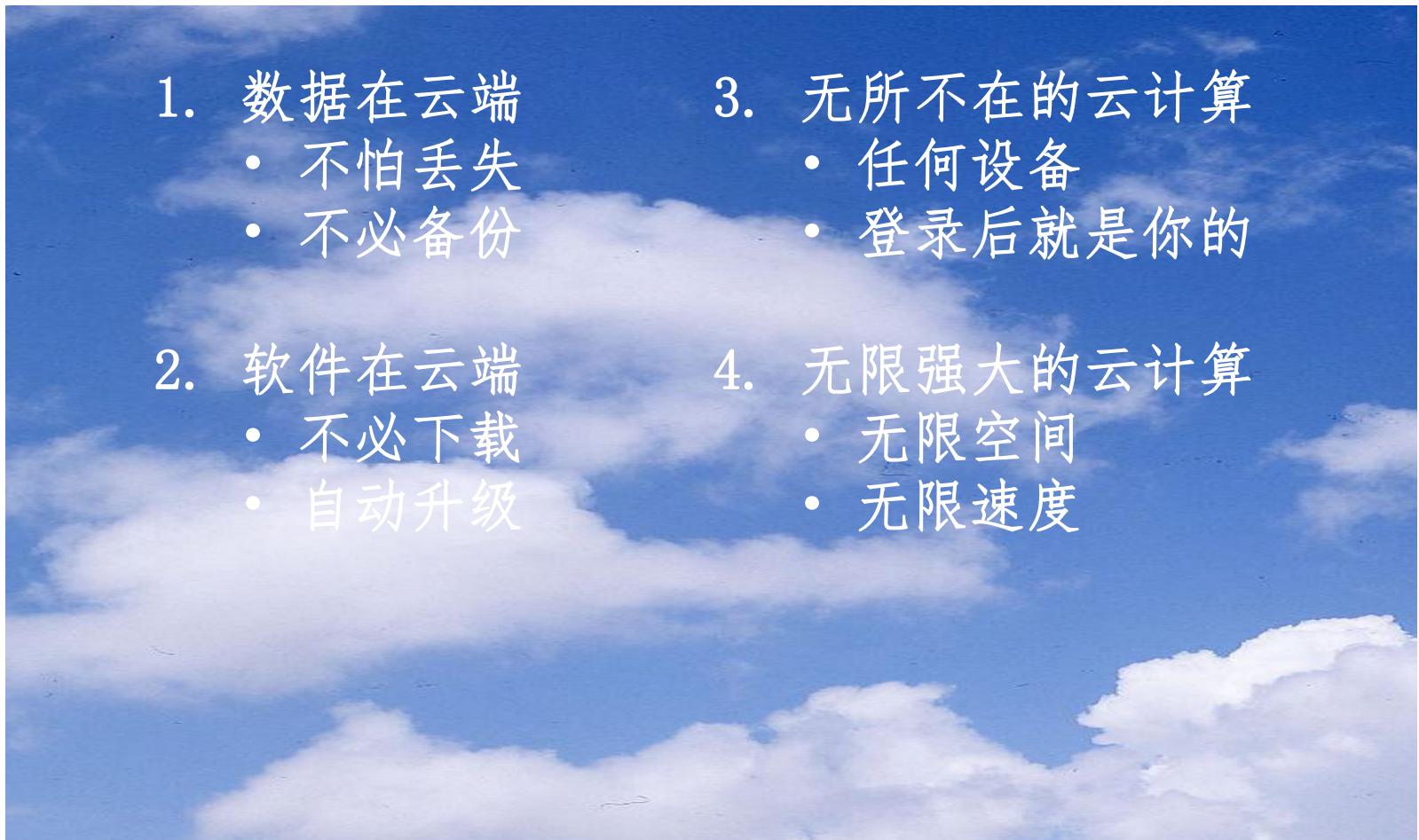
- 计算模式:
 - 如何进行并行分布式计算?
 - 如何分发待处理数据?
 - 如何处理并行分布式计算中的错误?
- 算法及实现:
 - 并行算法的设计
 - 并行/分布式编程
- 软件体系结构:
 - 容错, 高并发, 可扩展
 - 存储

新的计算模式?

什么是云计算？

- 云计算是一种新兴的共享基础架构的方法，通过互联网将资源以“按需服务”的形式提供给用户
- 利用互联网连接的数据中心和服务器进行**高效计算**和**信息存取**的系统,使计算能力可以向电能一样提供给客户（高度可扩展）
- 不同于以往的高性能计算：
 - 它除了提供大规模分布式计算外，
 - 还以组织和管理数据为核心工作之一，它获取并且维护持续变化的数据集
 - 提供**存储**以及方便操作数据的**编程模式**

云计算的理念



1. 数据在云端

- 不怕丢失
- 不必备份

3. 无所不在的云计算

- 任何设备
- 登录后就是你的

2. 软件在云端

- 不必下载
- 自动升级

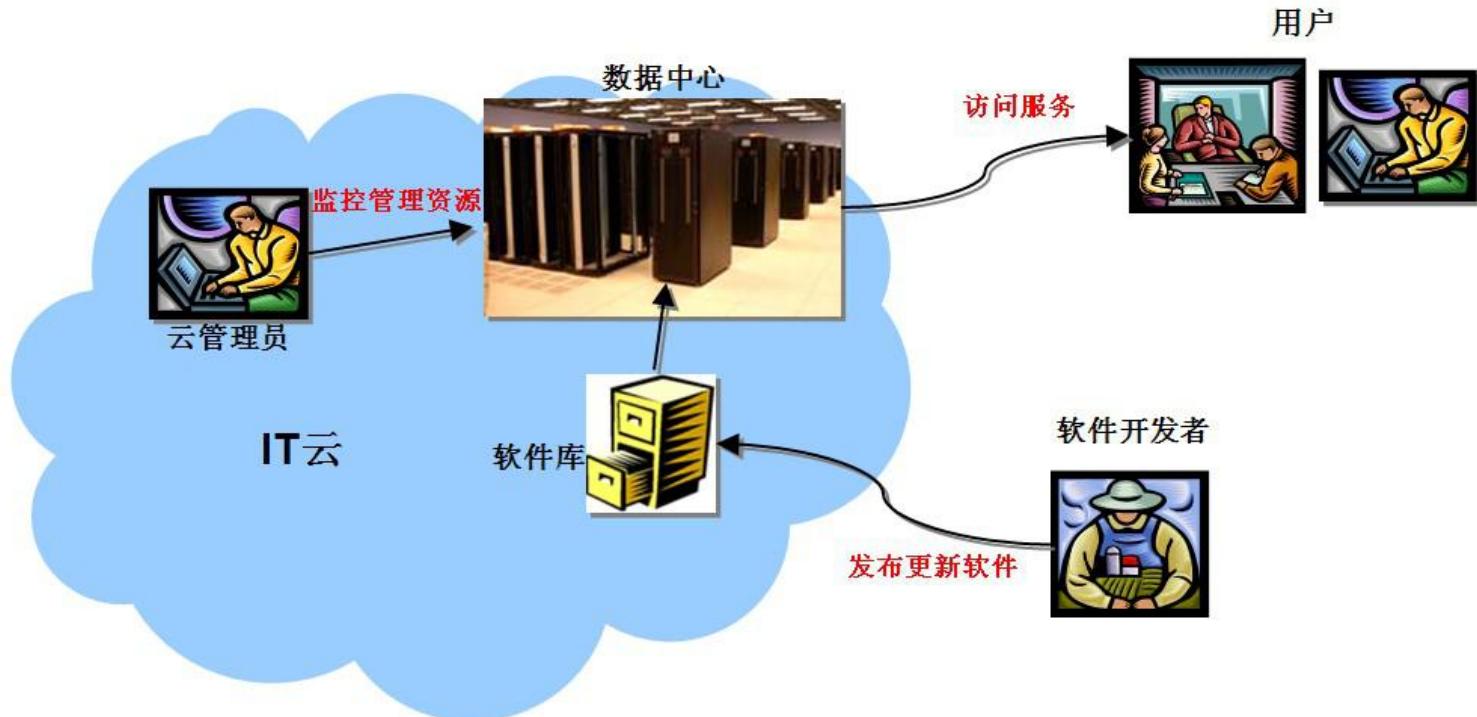
4. 无限强大的云计算

- 无限空间
- 无限速度

云计算的特点

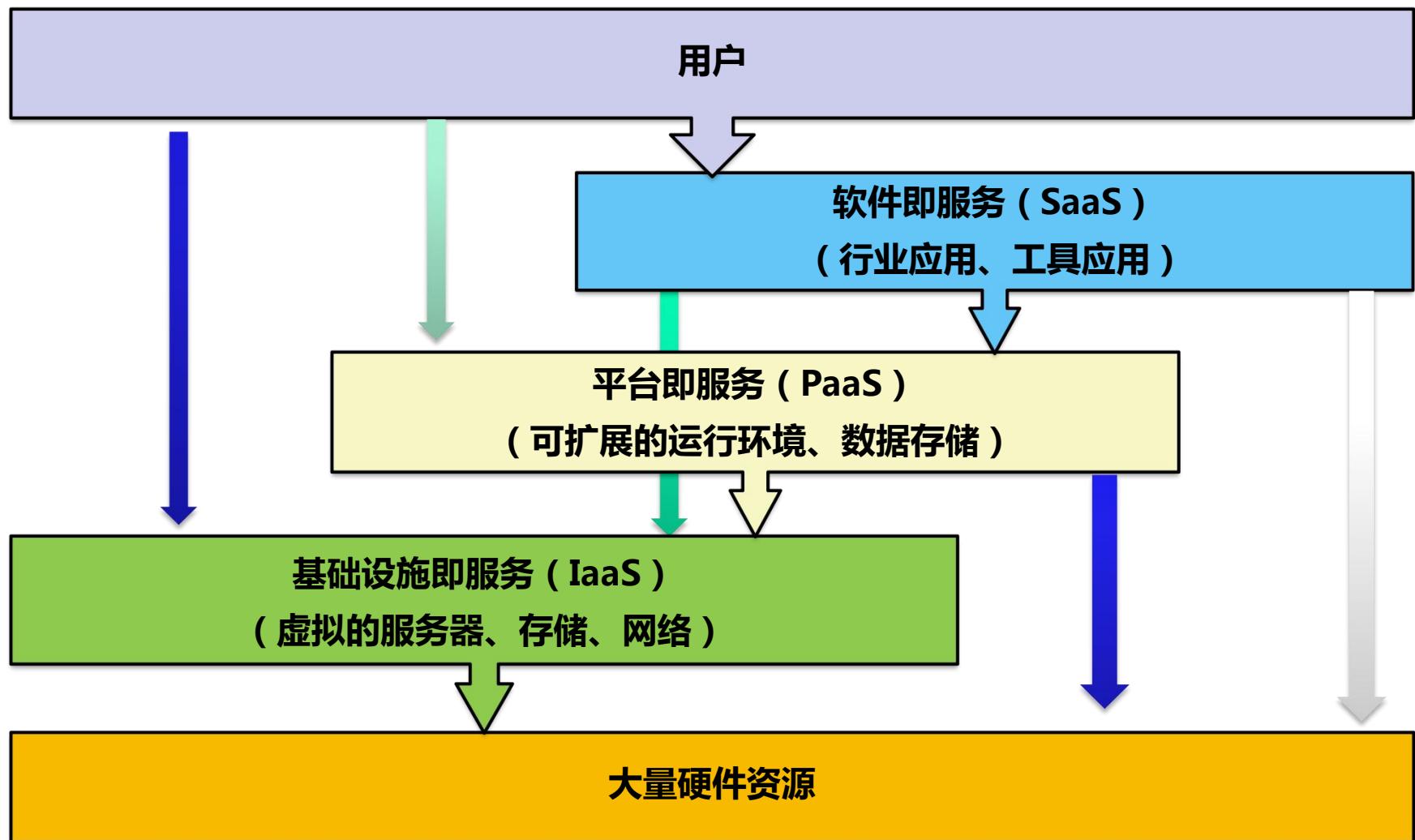
- **分布式：**
 - 使用的物理节点是分布的
- **虚拟化，这是云计算最强调的特点**
 - 每一个应用部署的环境和物理平台是没有关系的。通过虚拟平台进行管理达到对应用进行扩展、进行迁移、进行备份，种种操作通过虚拟化层次完成
- **动态可扩展**
 - 通过动态的扩展虚拟化的层次达到对以上应用进行扩展的目的。把各种IT资源，软件、硬件、操作系统、存储网络所有要素都虚拟化，放在云计算平台中统一管理

云计算应用场景



有了云计算，用户无需自购软、硬件，甚至无需知道是谁提供的服务，只关注自己真正需要什么样的资源或者得到什么样的服务。

云计算服务的层次



云服务的模式—IaaS

- 基础设施即服务
(Infrastructure as a Service)
- 基础设施从哪里来: **自己建造**
 - 机房
 - 服务器
 - 网络
 - 配套设置和管理
 - 供电, 散热, 容灾



云服务的模式—IaaS

- 基础设施即服务
(Infrastructure as a Service)
- 基础设施从哪里来: **IaaS**
 - 机房
 - 服务器
 - 网络
 - 配置设置 ➢ 管理
 - 供电, 散热, 容灾



云服务的模式—PaaS

- 平台即服务 (Platform as a Service)
- 系统平台怎么办: **自己搭建**
 - 操作系统
 - 应用软件
 - 环境配置
 - 日常维护



云服务的模式—PaaS

- 平台即服务 (Platform as a Service)
- 系统平台怎么办: **PaaS**

➤ 操作系统
➤ 应用软件
➤ 环境配置
➤ 日常维护



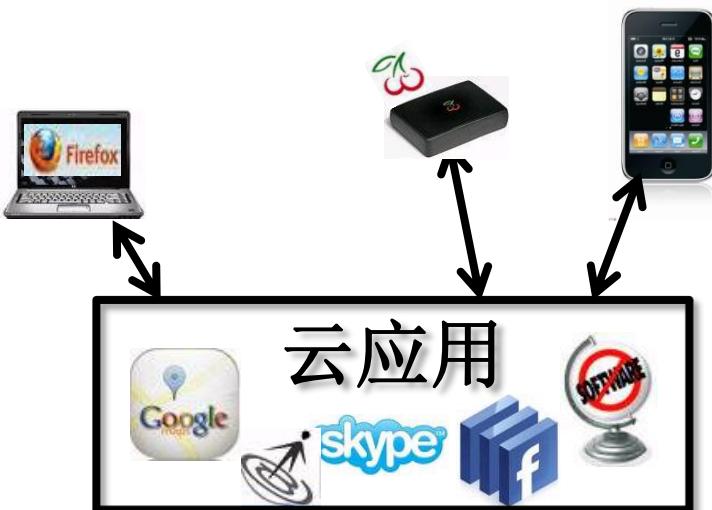
云服务的模式—SaaS

- 软件即服务 (Software as a Service)
- 应用软件:自己自足
 - 内部开发，内部使用
 - 数据难以交换，易丢失
 - 应用难交互，不规范

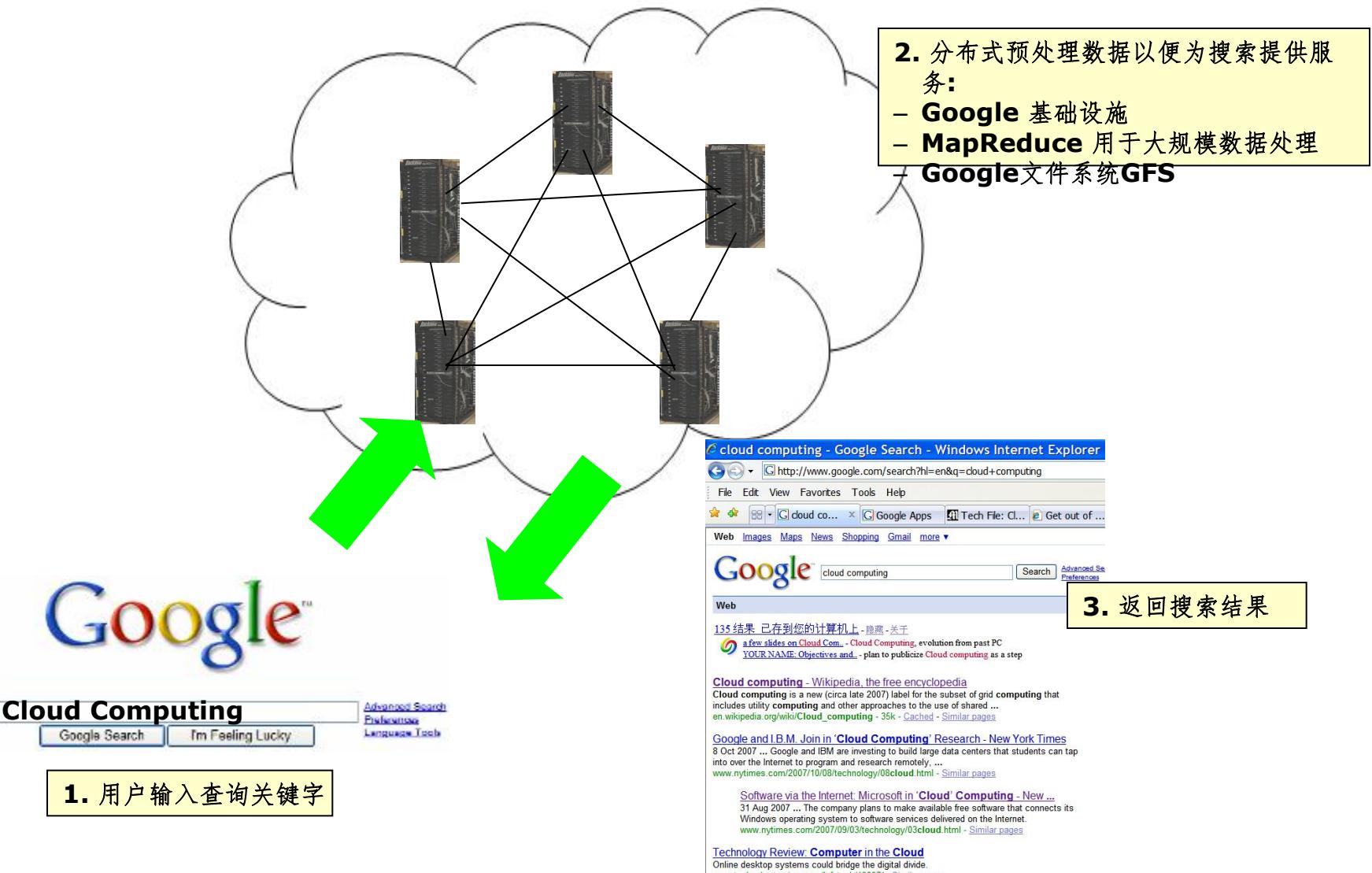
云服务的模式—SaaS

- 软件即服务 (Software as a Service)

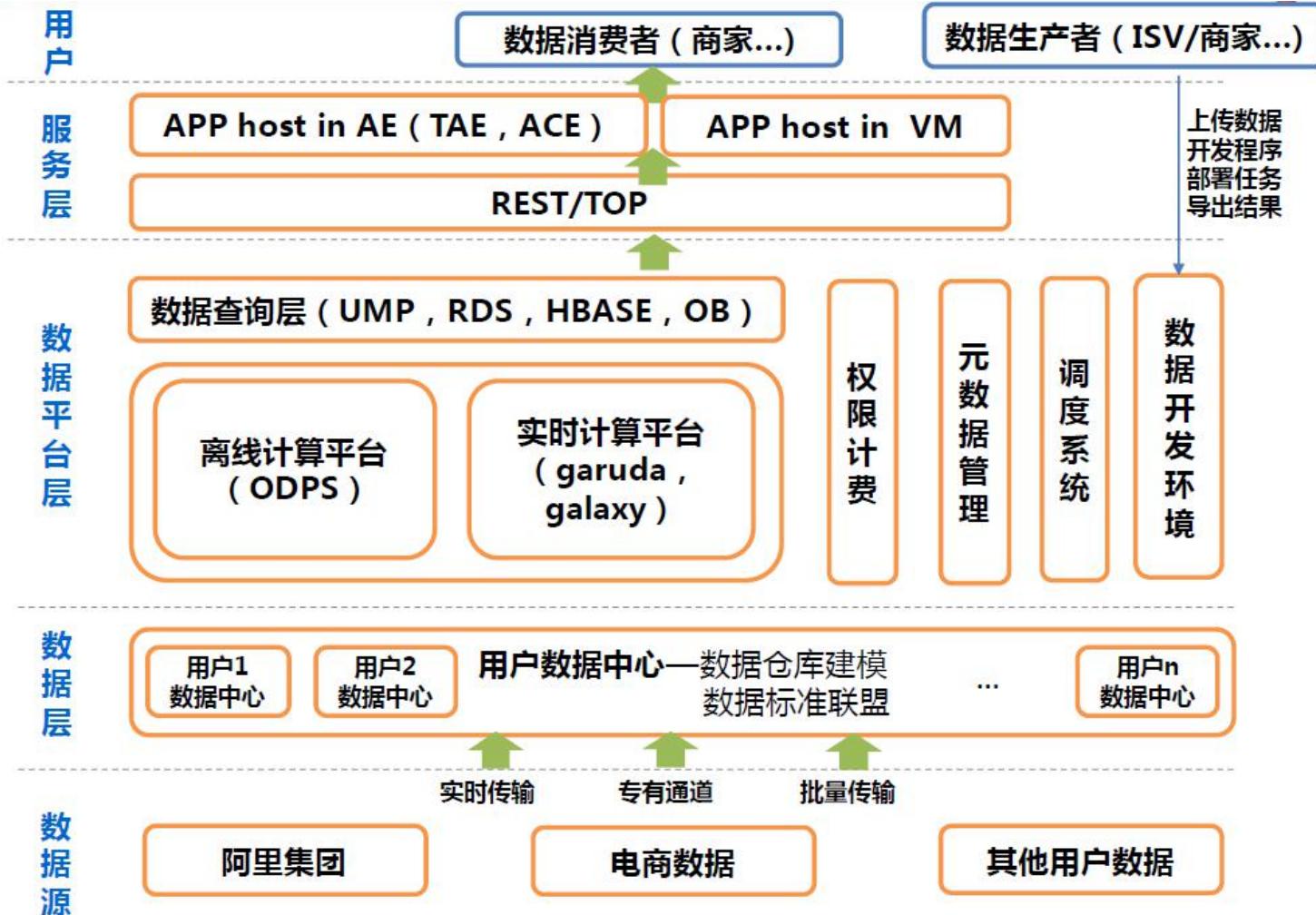
- 应用软件: **SaaS**
 - 内部开发, 内部使用
 - 数据难以交换, 易丢失
 - 应用难交互, 不规范



例子： Google搜索

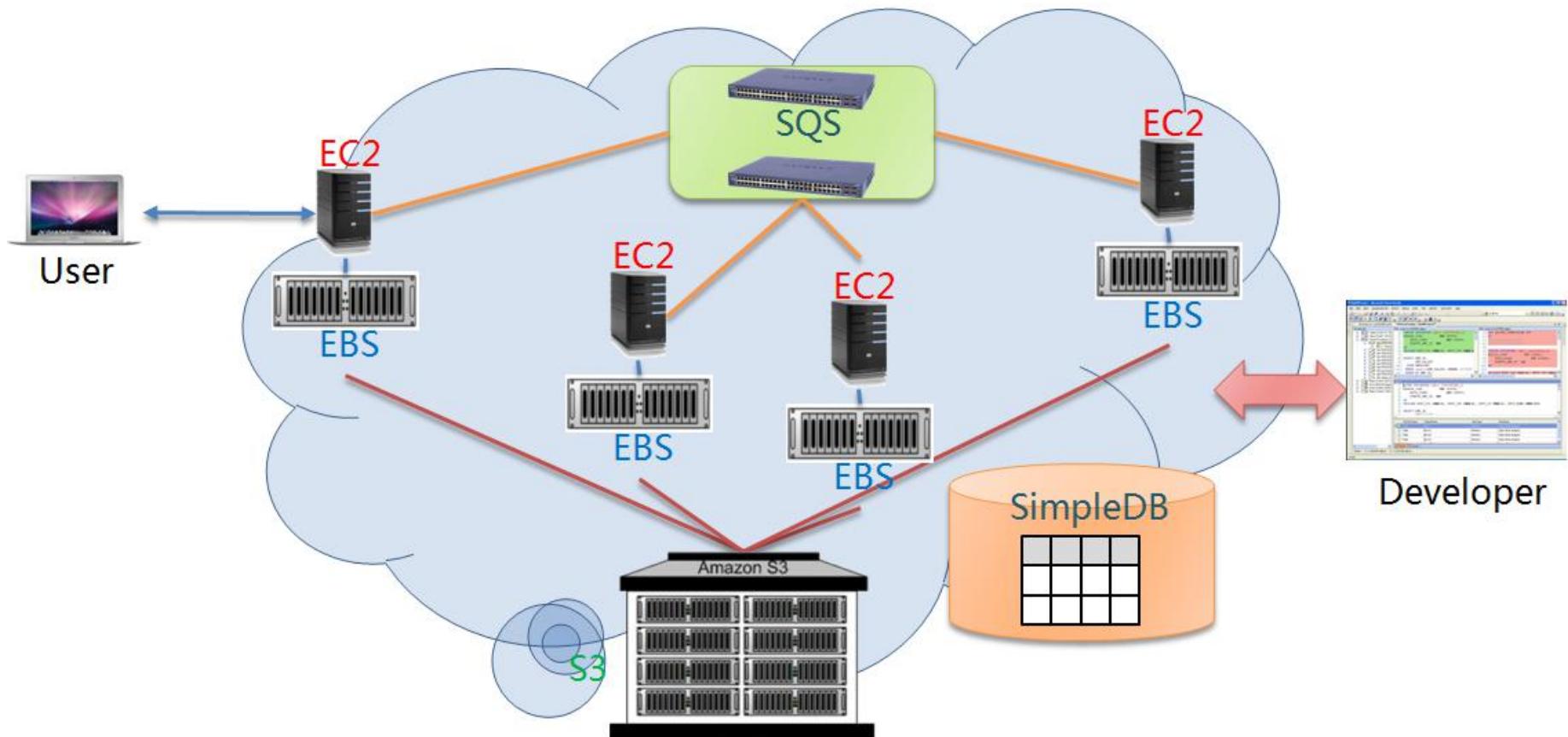


例子：阿里御膳房



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

例子：Amazon弹性云 (EC2)



- **EC2:** Amazon EC2是一种云基础设施服务，用户根据业务的需求自由地申请或者终止资源使用
- **S3:** 云存储服务

EC2计算实例的价格

按时收费 (On-Demand Instances)

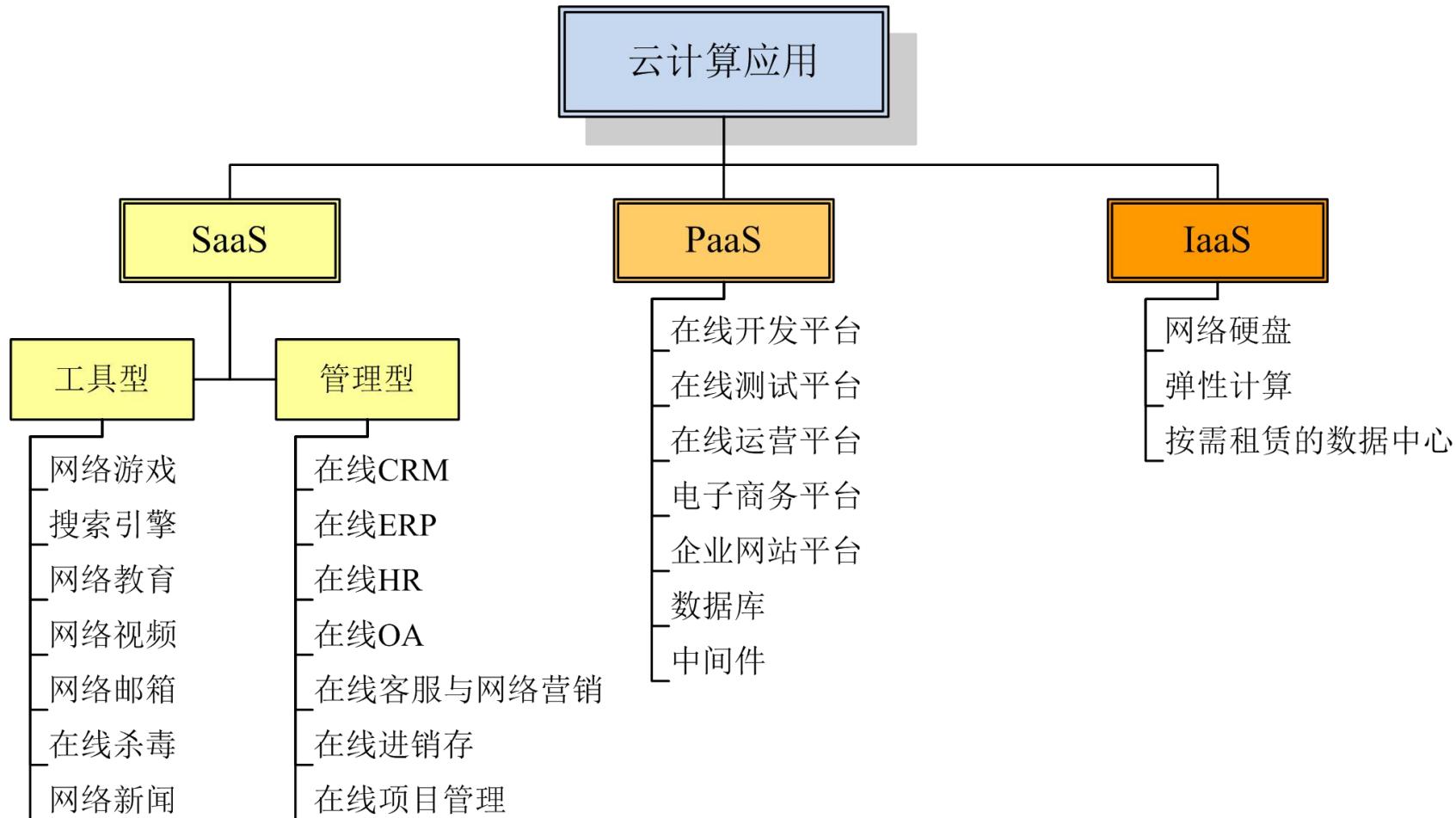
US – N. Virginia	US – N. California	EU – Ireland
Standard On-Demand Instances	Linux/UNIX Usage	Windows Usage
Small (Default)	\$0.085 per hour	\$0.12 per hour
Large	\$0.34 per hour	\$0.48 per hour
Extra Large	\$0.68 per hour	\$0.96 per hour
High-Memory On-Demand Instances	Linux/UNIX Usage	Windows Usage
Double Extra Large	\$1.20 per hour	\$1.44 per hour
Quadruple Extra Large	\$2.40 per hour	\$2.88 per hour
High-CPU On-Demand Instances	Linux/UNIX Usage	Windows Usage
Medium	\$0.17 per hour	\$0.29 per hour
Extra Large	\$0.68 per hour	\$1.16 per hour

EC2的应用案例：纽约时报



- 使用亚马逊云计算服务在不到24个小时的时间里处理了1100万篇文章
 - 累计花费240美元
 - 如果用自己的服务器，需要数月和多得多的费用

云计算的应用版图



云计算带来的革新

- 新的计算模式：
 - Internet为中心的计算
 - 海量， 并行扩展
- 新的应用模式：
 - 新的连接方式
 - 更好的信息利用方式
- 新的商业模式：
 - 开放租赁的软件平台
 - 一切都是服务

总结：高性能计算与云计算的共同点

- 都是大规模系统，相同的核心技术和挑战：
 - 分布式计算、集群、高密度计算等技术
 - 高速互连、存储分层、异构多核处理器
 - 并行/分布式编程
 - 系统可靠性和恢复能力
 - 机柜、冷却、能耗效率

总结：高性能计算与云计算的区别

- 应用：
 - 高性能计算：主要面向科学计算、工程模拟、动漫渲染等领域，漫渲染等领域，大多属于**计算密集型**的应用
 - 云计算：主要是在Web2.0、社交网络、企业IT建设和信息化等领域，以**数据密集型、I/O密集型**应用为主
- 技术：
 - **网络**：HPC需要特制的高速互联网络
 - **虚拟化**：HPC几乎不用虚拟化技术，而在企业私有云中，虚拟化却是一个最基础的技术。
- 用户：
 - 高性能计算：政府部门/大企业，专业人士
 - 云计算：具有良好的用户界面，隐含复杂的计算逻辑，面向各种企业及普通用户

总结：两者的关系

- 对高性能计算而言，云计算并不是一个新的概念。已经发展近30年的超级计算中心也是一种早期的云计算模式：昂贵的计算资源集中部署，多个领域的用户通过互联网远程使用计算服务并依据使用量支付费用
- 高性能计算需要极低的任务间通信延迟，目前的云模型并不支持顶尖的超级计算
- 云计算的易用性对传统的高性能计算的计算模式带来巨大影响，传统的排队批处理方式很难实现按需即时响应的科学计算，**On-demand**的云计算给高性能计算提供了更易交互的计算模式
- 都是基于大规模系统—高性能计算机

课程小结

- 术语和定义
 - HPC, HPCC, Parallel Computing, Distributed Computing, Cloud Computing
 - Cornerstone:Compute, Storage,Communication
 - Units of measurement
 - SISD, SIMD, MISD, MIMD
 - PVP, SMP, MPP, DSM, Cluster, Constellation
- 高性能计算的需求及应用
- 云计算的内涵及服务层次
 - IaaS、PaaS、SaaS
- 发展趋势
 - 体系结构、处理器、应用领域等

推荐读物和网站

- 阅读：
 - 《并行计算—结构、算法、编程》第一章
- 网站：
 - TOP500 Supercomputer Sites
<http://www.top500.org>

下一讲

- 并行计算机体系结构
 - 《并行计算—结构、算法、编程》第1, 2章