# 大数据背景下的新型体系结构

## 主讲人：华中科技大学计算机科学与技术学院 金海教授

# 简要介绍

大数据时代以计算为中心的数据处理如何改变？

数据爆炸式增长引发的处理时效性如何保障？

数据体量巨大造成的能耗成本如何控制？

这一切皆源于大数据的兴起，其以独特的魅力风靡全球，小到衣食住、社交娱乐，大到智能制造的实现、基因密码的破解，它以高效快速的运转速度获得各个行业的青睐，让世界焕然一新，它的存在即是人类攀爬web3.0时代高峰的又一奇迹，支撑这一奇迹发生的内存计算又如何实现呢？

金教授带我们揭开大数据的神秘面纱，探讨实现大数据运算的异构内存体系结构所面临的技术挑战和研究方向。

## 内容

1. 从数据到大数据

信息技术为数据处理提供了自动的方法和手段，推动**数据**（信息）成为继**物质**、**能源**之后的第三大战略资源。

信息技术廉价化 → 信息数据激增 → 大数据

1. 中国互联网在一分钟内会发生什么？
2. 大数据的出现标志信息化进入新的阶段

最早的信息化就是数字化。如档案管理系统、人事管理系统。

数字化（-1995） → 网络化（-2015） → 智能化（2015-） → ？？

1. 大数据不只是数量大

三维：

* Data Velocity
* Data Volume
* Data Variety

2016年双十一发生了什么？

1. 大数据的4V特性

数据庞大（用户多、分布广、总量庞大）

变化频繁（用户强交互性、跨多通道快速传播、传播行为复杂）

种类繁多（数据源多样、数据类型多样、交互方式多样）

价值高但密度低（商业分析、政府决策、舆情监控）

淘宝→新浪推送

1. 大数据定义（来自维基百科）

规模庞大，结构复杂，难以通过现有的IT技术与工具处理的数据集。

1. **大数据处理改变思维方式**

转变1：从抽样到全样

大数据数量大，数据统计特征分布不均匀，传统采样方法不适用

（如Google流感预测采用搜索数据取代取样）

转变2：从精确到非精确

大数据下精确性不再是绝对追求目标，需对宏观趋势给出快速预测

（如所搜与推荐系统只需返回前几项）

转变3：从因果到关联

仅需知其然，无需知其所以然，用于“发现事实、预测未来”

（如沃尔玛综合商业、Twitter和Facebook数据）

1. Gartner新兴IT技术成熟度曲线

云计算和大数据已看不到，说明已非新兴技术，正慢慢融入我们的生活

1. 大数据在各行各业都获得应用
2. 以电商为例的行业应用数据

结构与表达规范

增长速度快

数据处理时效高

价值密度高

安全性要求高

双11开始后10分钟内，支付峰值（9分39秒）12万笔/秒

1. **内存计算**

把数据放到内存中计算，不是说内存一定具有计算能力

内存计算的概念在70年代就有了，Why now?

* 64-bit processors can address up to 16 exabytes of data.
* DRAM production costs drop by 32% every 12 months.
* 1GB of NAND flash memory average price is 56$ cents\*
* Commodity hardware provide multi terabyte of DRAM.
* In-memory-enabling software is available and proven.
* IMC software is often embedded in products.

2013年Gartner发布的10项战略性技术，其中就有In-memory Computing。

软件最好的内存计算SAP HANA

硬件最好的内存计算Oracle’s Big Data Appliance/Exalytics

1. 现代电子计算机系统的能耗

CPU不工作时，10%的功耗；

CPU 100%工作时，50%的功耗；

无论CPU是否工作，DRAM都有20%左右的功耗。

1. DRAM刷新所带来的问题

* Energy consumption

桌面电脑的2~8GB内存功耗大概是15%

服务器的16~64GB内存功耗大概是**47%**

* Performance degradation
* QoS/predictability impact: (Long)pause time during refresh
* Refresh rate limits DRAM capacity scaling

1. 基于DRAM的内存计算模式面临的挑战

DRAM介质易失性

DRAM介质存储密度低（远不能满足大数据对内存容量TB级甚至PB级的需求）

DRAM功耗高

内存子系统成本高

1. **基于非易失存储（NVM）技术的大数据时代计算机**

非易失存储技术的兴起 → Intel-Micon 3D XPoint Technology

新型存储介质的性能不断逼近传统DRAM。

ITRS（国际半导体技术蓝图） Projection

新型：Memristor（忆阻器）、PCM、STT-RAM

传统：DRAM、Flash、HD

Intel Non-Volatile Memory Solutions

OPTANE memory闪腾

插在内存条上，一个槽375TB

1. 内存和存储技术的发展——Storage Class Memory

新型存储级内存（SCM）为内存计算带来曙光

* 可按字节寻址
* 持久存储，断电数据不丢失
* 比NAND flash读写快1000倍
* 耐久性比NAND flash高1000倍
* 静态功耗接近“零”

重构虚拟内存结构：打破I/O瓶颈

新型内存计算结构：

构建DRAM+SCM的混合层次/并行内存结构

从以计算为中心转向计算与数据相结合的全新内存计算架构

1. HP：The Machine

编程怎么办？没有硬盘，数据全都存在内存当中。

→ SNIA NVM Programming TWG

SNIA NVM Programming Model v1.1

1. 内存计算方面的实践

混合内存架构给软硬件生态系统带来巨大冲击。

* 体系结构

异构内存部件（NVM/DRAM）的组织与管理（多通道、多层次？）

多核与大内存之间不断加大的带宽鸿沟（内存级并行）

* 操作系统

大容量内存的寻址问题（大页？段式结构？）

面向混合内存的操作系统任务调度

* 数据组织

文件结构、key/value store

* 编程模型

1. 内存计算：大数据处理的机遇与挑战

基于新型非易失存储介质的高可靠、大容量、低功耗的混合内存体系结构原型系统，非易失存储介质作为内存容量在TB级。

1. 混合内存计算全系统模拟器

基于Zsim和NVMain的全系统混合模拟器，支持对多种存储介质和存储架构的模拟

开源社区：<https://github.com/CGCL-codes/SHMA>

1. 混合内存计算持久化文件系统

新型文件模式、新型内存访问

1. 混合内存中时延感知的缓存替换策略

问题：传统基于命中率的Cache替换算法在混合内存中效率低下

根源：异构内存架构下，LLC确实代价的不对称性

思路：Cache替换时兼顾局部性和替换代价，时延感知的Cache替换策略（MALRU）

提出新的评价存储系统性能的通用指标：平均访存时延Average Memory Access Time（AMAT）

Cache替换原则：（1）尽量使

1. Deca：基于数据生命周期的内存计算大数据处理系统

大内存导致垃圾回收扫描一遍内存的时间很长

→

自动分析用户自定义函数和数据类型，获得数据对象的生命周期，根据生命周期来分配和释放内存空间，减少因垃圾回收所带来的开销。

系统按照相似的生命周期透明地分解和组合数据对象，对于同一生命周期的对象一起释放内存空间。

与Apache Spark相比，可加速22.7倍到41.6倍

节省了99%的垃圾回收时间

1. Mammoth（开源）：以内存为中心的MapReduce系统

以内存为中心调度数据，细粒度分配内存，以内存为缓冲串行化硬盘

1. Landscape of Disk-Based and In-Memory Data Management Systems
2. 大数据时代解决大数据处理时效性的有效手段：内存计算
3. 值得重点研究的关键机理

* 面向大数据处理的异构层次内存的协同组织模式研究
* 基于内存计算的大数据低能耗处理环境研究
* 基于内存计算的大数据高效并行处理机制的研究

1. 未来发展趋势：领域通用计算机

社会分工：律师、医生、教师等等

→ 计算机：领域通用计算机

1. 老师赠语

“应用驱动发展，软件定义一切，硬件改变未来”

“大学本科阶段教给我们的主要是一种思维方式和思维模式，而不是某一种特定技术。”

## 种子班同学提问

1. 问：之前听您说到内存计算在硬件方面，CPU的指令集其实是要重做的，我想问的是内存计算和CPU架构的问题，CPU对内存的操作是依什么样的类型和架构来计算的？

答：我首先要澄清一下，内存计算并不一定意味着内存具有计算的功能（有一类存储器，如忆阻器本身具有计算功能），它是指整个数据放到内存里面来做计算，而不是把计算的功能由CPU交给内存来做。传统意义上的内存计算是用现有的存储器去扩充内存的容量，这样的话，第一，计算机的指令集不会发生变化；第二，要将操作系统的内存管理模块都要替换掉，因为它已经不是一种内存，而是异构内存，存储也会分层次和体系结构，我刚才讲块大小也会发生变化，替换策略也会发生变化，很多相关算法都要进行改进，但这只是操作系统的一个部分，你只要改进了这个部分，操作系统提供的接口还是原来的接口。如果你用到了忆阻器，因为它具有计算功能，所以数据写进去和都出来已经不一样了，这种情况下可能有另外一种模式去控制忆阻器进行计算，但这不一定是内存计算必须的，它是内存计算更高阶的发展。

1. 问：既然能保持提供给应用程序的接口一致，是否大部分现有的软件不需要移植都能够跑在这个新平台上？

答：有一点需要记住，任何新硬件的出现必须继承现有的软件，如果不能继承，颠覆性的硬件一定是没有前程的。这在计算机里面有血的教训，在2000年左右，出现了一个64位的服务器叫安腾服务器，现在已经没有了，就是因为它和原来的至强不兼容，现在我们用的都是至强64位。

1. 问：新的硬件结构为了能够通用的话，在性能方面肯定会有一定的消耗，那么在未来的发展，是不是还是需要对上层软件的硬件接口进行一定的修改，来完整地利用好内存计算的性能呢？

答：你说的完全正确，因为继承上层应用并不意味着要迁就，它并没有最有效地利用下层的硬件环境，新的编程模型的出现，就是为了更有效地利用下层的硬件，我们不能让现在所有的程序都重写，但可以使得以后编程的时候按照新的模型来编写，这是未来发展的趋势。所以比如以后大学再教C语言的时候，可能就要有所拓展了，不再是原来架构的C语言。

1. 问：我的理解是目前那些存储量更大的内存还是不能做计算的，而忆阻器是可以的，如果它真正实现商业化量产的话，对计算机带来的革命性改变是不是比现在内存计算更加剧烈？

答：是的。

问：这项技术和量子计算相比，您认为哪一个会首先进入人们的日常生活呢？

答：应该是内存计算，因为这个还是传统的计算过程，量子计算是个新的计算过程，我认为在未来的十到二十年里，量子计算机不会成为通用计算机，而是专用领域计算机，比如解密。

问：您能不能继续介绍一下忆阻器呢？

答：建议以后你们可以专门请光电国家实验室缪向水老师介绍忆阻器，我有一次听过他的报告之后就明白了，要我讲的话就讲不出来了。

1. 问：在人工智能这么火的时代，我们面对了非常巨大的计算量，今天您分享的主题，我的理解就是通过改变各种各样的计算机体系结构来适应计算量的要求，我想问一下，除了这个方面，在计算理论方面有没有相关的研究可以提供一个更好的计算模型，这方面的进展怎么样？能不能跟我们现在不断发展的硬件相配合起来？

答：你说的这个问题稍微超前了一点，现在还没有新的计算模式或计算理论。我们大数据曾经希望借助一种相当于计算机里面用到的统计学这样的一种数学模型去建模，但是非常可惜直到现在还没有一种模型能被大家所接受。

问：我们不断增强计算机体系结构，它所能达到的计算能力仍然只是图灵完备级别的吗？

答：应该是这样的。未来会是什么样子的话，比如人工智能，今年有很大的突破，我之前讲到的问题都是基于标签学习的，今年人工智能最大的突破就是无标签学习，可能将来它还是要跟解开我们大脑的奥秘相结合，比如为什么婴儿看过之后就记得了，而电脑没有这个功能，因为没有样本让它去学习，所以现在有一种baby learning的模式，很多理论都待突破但还没有突破。我听过一个报告，如果把人的眼睛当作一个摄像头的话，一天看16个小时睡8个小时，我们可以摄几个TB的数据，但我们的脑袋装不下几个TB的数据。我今天看过某些同学之后，可能以后走在路上遇到就知道这个人是谁，虽然叫不出名字，但我知道见过他，为什么人有这么强大的能力，因为我们对人脸的识别，只需要记十几个特征点。所以我觉得，未来学科交叉是非常非常关键的，我们不能纯计算机的人去想计算机的事，比如人工智能完全依赖于脑科学的发展，我们很容易地能做出每个人的基因图谱，但是还不能做出脑图谱。所以我们对自身的研究不断地提高，有助于我们学科的发展。

问：那么可不可以这样理解？如果未来理论上有重大突破，可以显著地降低我们对计算量的需求？

答：这是肯定的，其实量子计算就是一种理论上的突破。

## 在讲座的最后，全班同学与金海教授合影留念，大家对金教授的讲座一致好评，认为他讲得生动形象，本人也十分亲和，听完讲座之后都收获颇深。金教授也表示欢迎种子班的同学去他的实验室。