



中国科学院科技战略咨询研究院  
Institutes of Science and Development, Chinese Academy of Sciences



# 科教算力共享

王晓明, 吴静, 李宏, 孙翊, 吕佳龄, 刘昌新

中国科学院科技战略咨询研究院



## 一、算力全景图

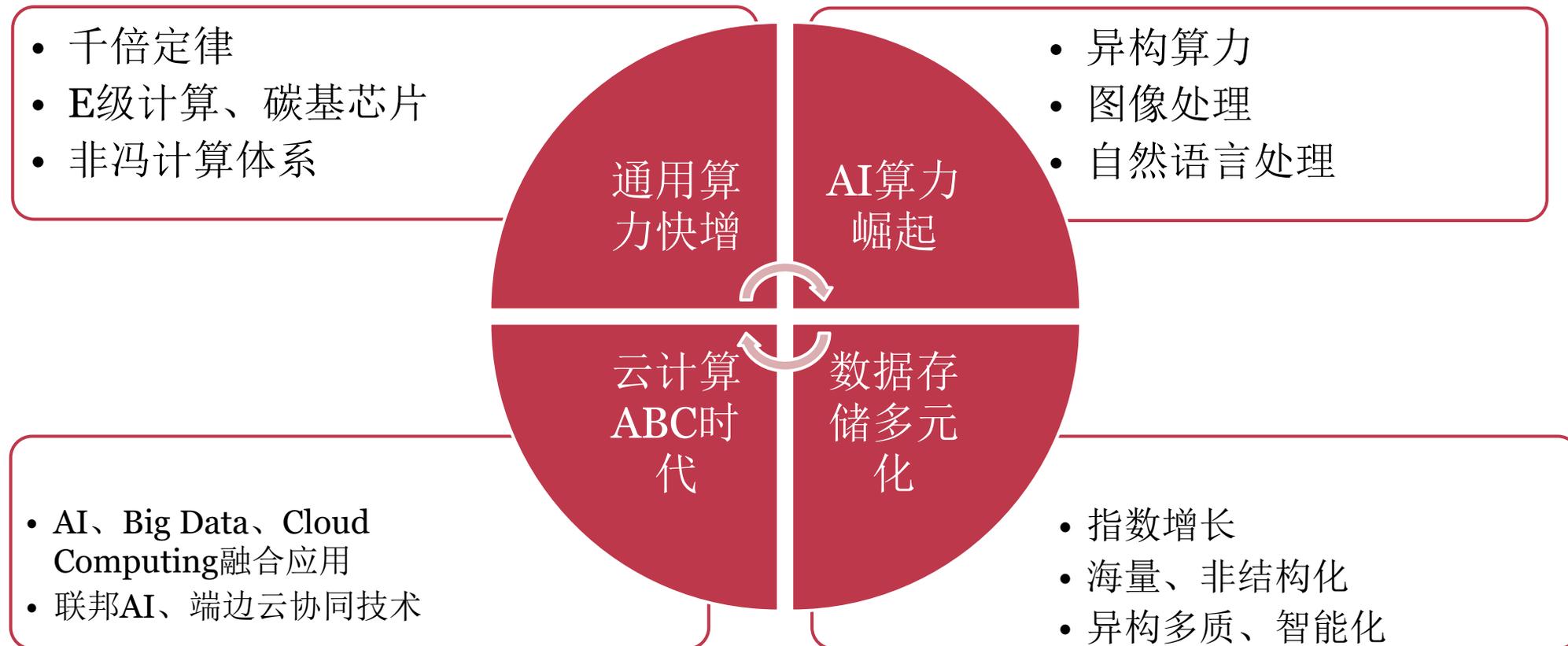
## 二、科研算力需求

## 三、教育算力需求

## 四、未来发展思路与建议

# 算力的发展态势

- **算力：** 计算能力，主要指数据收集、传输、计算和存储能力。
- **算力相关设施：** 大数据中心、超算中心/云计算中心、大科学中心、AI等等





# 算力发展态势



- 国产芯片成为国产超级计算新主流
- 但发展路径较多，有待优化

	龙芯	申威	海光	飞腾	鲲鹏	兆芯
指令集	MIPS	Alpha	X86	ARM	ARM	X86
指令集来源	永久授权+自研	永久授权	授权	授权	授权	授权
主要产品	龙芯 3A3000/3B3000	SW410、 SW26010	Dhyana	FT-2000/64	鲲鹏920	ZX-C、ZX-A 系列
支持的桌面及服务 器操作系统	Linux	Linux等	Windows、 Linux	多种操作系统	欧拉	Windows、 Linux
所属公司	龙芯中科技术有限公司	江南计算机所	海光信息技术有限公司	天津飞腾信息技术有限公司	华为技术有限公司	上海兆芯集成电路有限公司

- 美国没有像中国一样以政府为主部署的国家计算中心网络
- 各大企业有各自的独立算力布置，如IBM、微软等
- 超级计算机设施“蓝色纽约”位于纽约的布鲁克海文国家实验室（BNL），是纽约计算科学中心（NYCCS）的核心
- “蓝色纽约”是BNL与IBM成功合作的例子，其软硬件及操作体系被IBM控制和支持
- “蓝色纽约”一直在寻求学术界与工业界的合作，计算技术改进开发吸引了新业务，将先进的科学研究转化为了广泛的商业成功





# 美国的算力设施



## • 美国主要算力基础设施

所属实验室	研发设施英文名称	中文翻译
<b>Argonne National Laboratory</b>	<a href="#">Transportation Research &amp; Analysis Computing Center</a>	交通研究与分析 计算中心
<b>Brookhaven National Laboratory</b>	<a href="#">New York Blue</a>	蓝色纽约计算科学中心
<b>Brookhaven National Laboratory</b>	<a href="#">New York Center for Computational Sciences</a>	纽约计算科学中心
<b>Fermilab</b>	<a href="#">Advanced Computing Technologies</a>	先进计算技术设施
<b>Lawrence Livermore National Laboratory</b>	<a href="#">High Performance Computing Innovation Center</a>	高性能计算创新中心
<b>Pacific Northwest National Laboratory</b>	<a href="#">Computational Science Facility (CSF)</a>	计算科学设施 (CSF)
<b>Pacific Northwest National Laboratory</b>	<a href="#">Northwest Institute for Advanced Computing (NIAC)</a>	西北先进计算研究所

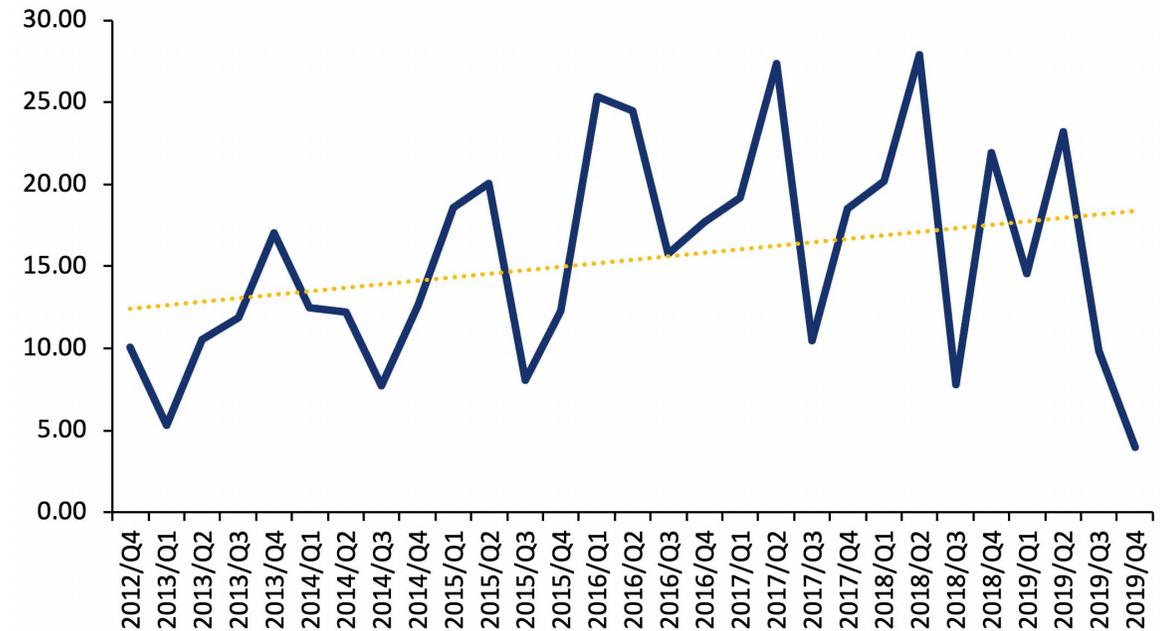
## • 科学专用算力基础设施

所属实验室	研发设施英文名称	中文翻译
<b>Argonne National Laboratory</b>	<a href="#">Argonne Leadership Computing Facility (ALCF)</a>	阿贡前沿计算设施 (ALCF)
<b>Brookhaven National Laboratory</b>	<a href="#">National Nuclear Data Center</a>	国家核数据中心
<b>Lawrence Berkeley National Laboratory</b>	<a href="#">National Energy Research Scientific Computing Center (NERSC)</a>	国家能源研究科学计算中心 (NERSC)
<b>Oak Ridge National Laboratory</b>	<a href="#">Oak Ridge Leadership Computing Facility (OLCF)</a>	橡树岭前沿计算设施 (OLCF)

## □ 欧洲高级计算伙伴关系设施（PRACE）是欧盟数字经济领域的分布式算力建设

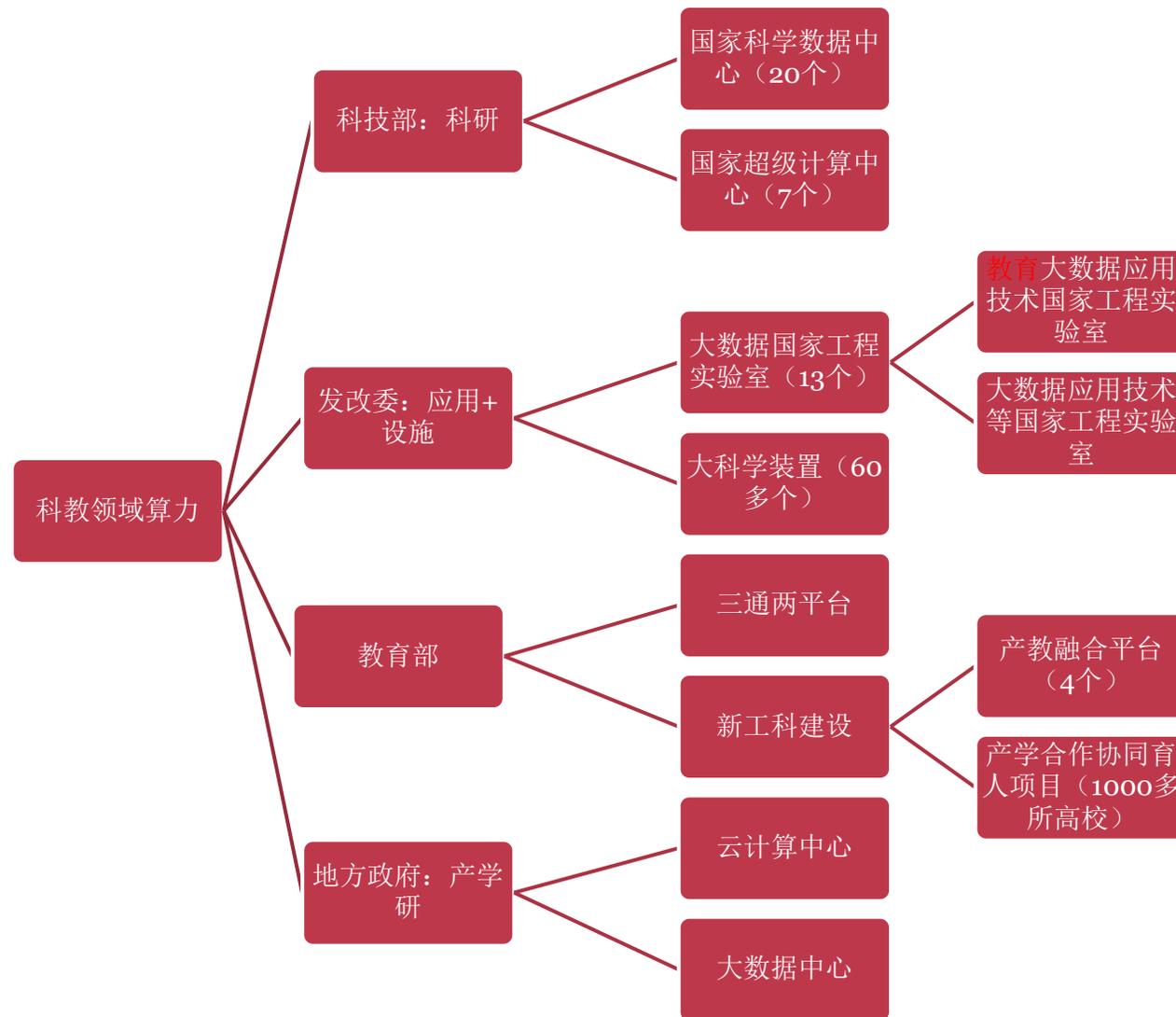
- **2006-2010**建设费为5亿欧元，建成后年运行费**0.60**亿欧元
- **2010-2019**年改造升级资金**1.32**亿欧元，其中**1.25**亿欧元由欧盟委员会提供

- PRACE的入口由**5**个成员国提供（西班牙、意大利、瑞士、德国、法国）
- 通过评审，学术界和工业界的科学家和研究人员就可以使用其系统
- PRACE部署的系统不断更新和升级，达到了全球**HPC**技术的顶峰
- PRACE的高性能计算算力也为中小企业提供支持，纳入商业模式



PRACE算力中中小企业所占用的比例

- 我国在国家层面出台了《科学数据管理办法》和《国家科技资源共享服务平台管理办法》
- 部署了国家超算中心、各国家科学数据中心、大科学装置等
- 此外，各科研院所和高校也拥有自己的小型超算中心资源



我国科教领域算力布局



# 国家超算中心



云计算中心	成立时间	规模	应用领域
天津云计算中心	2009年	占用房屋面积约8,500平方米，共建有2个大型机房共约4,000平米	生物医药、石油地震勘探数据处理、动漫与影视渲染、新材料新能源、高端装备设计与仿真、航空航天、流体力学、天气预报、气候预测、海洋环境模拟分析
深圳云计算中心	2009年	运算速度达每秒1,271万亿次。配备高达17.2PB的海量存储及来源于各大运营商、教育网的丰富网络带宽资源	大规模科学计算和工程仿真、动漫渲染等计算业务，同时以其强大的数据处理和存储能力为社会提供云计算服务
长沙云计算中心	2011年	一期工程规划建筑面积30000平方米	为气象、国土、水利、卫生/医疗、交通等公共服务部门提供了高性能的计算平台服务。
济南云计算中心	2011年	采用自主处理器构建千万亿次超级计算机系统的国家	面向海洋科学、现代农业、油气勘探、气候气象、药物筛选、金融分析、信息安全、工业设计、动漫渲染等领域提供计算和技术支持服务，承接国家、省部等重大科技或工程项目。
广州云计算中心	2013年	总建筑面积42332平方米，其中机房及附属用房面积约17500平方米	是助推战略性新兴产业发展、支撑国家创新型城市和智慧广州建设的重大战略性基础设施，成为融高性能计算、海量数据处理、信息管理服务于一体的世界一流超算中心。

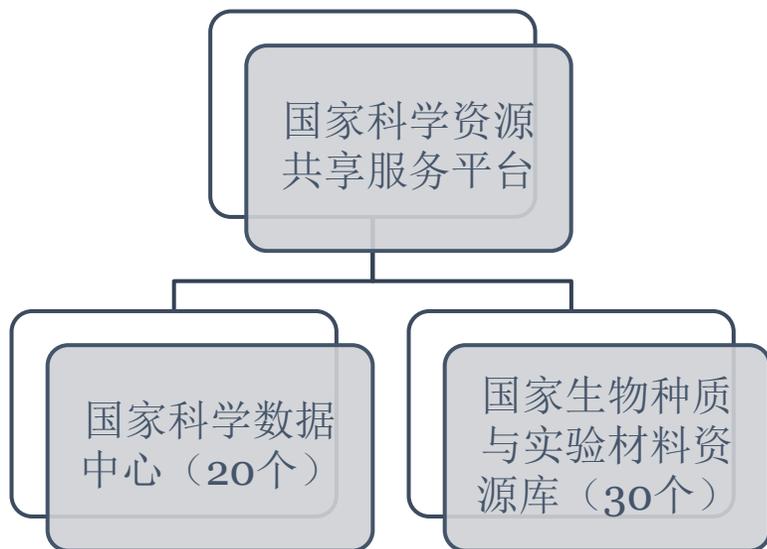


# 国家科学数据中心



**大数据中心：**用以存储和操作海量数据的中心。以IO操作、增删查改、任务调度为主。

**国家科学数据中心**是由国家按照统一标准和规划设立的大型数据管理设施，代表国家对于特定领域（或交叉领域）的科学数据进行统筹管理和服务。



- 中国科学资源共享平台

序号	国家平台名称	依托单位	主管部门	序号	国家平台名称	依托单位	主管部门
1	国家高能物理科学数据中心	中国科学院高能物理研究所	中科院	11	国家冰川冻土沙漠科学数据中心	中国科学院寒区旱区环境与工程研究所	中科院
2	国家基因组科学数据中心	中国科学院北京基因组研究所	中科院	12	国家计量科学数据中心	中国计量科学研究院	市场监管总局
3	国家微生物科学数据中心	中国科学院微生物研究所	中科院	13	国家地球系统科学数据中心	中国科学院地理科学与资源研究所	中科院
4	国家空间科学数据中心	中国科学院国家空间科学中心	中科院	14	国家人口健康科学数据中心	中国医学科学院	卫生健康委
5	国家天文科学数据中心	中国科学院国家天文台	中科院	15	国家基础学科公共科学数据中心	中国科学院计算机网络信息中心	中科院
6	国家对地观测科学数据中心	中国科学院遥感与数字地球研究所	中科院	16	国家农业科学数据中心	中国农业科学院农业信息研究所	农业农村部
7	国家极地科学数据中心	中国极地研究中心	自然资源部	17	国家林业和草原科学数据中心	中国林业科学研究院资源信息研究所	林草局
8	国家青藏高原科学数据中心	中国科学院青藏高原研究所	中科院	18	国家气象科学数据中心	国家气象信息中心	气象局
9	国家生态科学数据中心	中国科学院地理科学与资源研究所	中科院	19	国家地震科学数据中心	中国地震台网中心	地震局
10	国家材料腐蚀与防护科学数据中心	北京科技大学	教育部	20	国家海洋科学数据中心	国家海洋信息中心	自然资源部

国家发改委首次明确了新型基础设施的范围及内容，主要包括：信息基础设施、融合基础设施和创新基础设施，算力基础设施则是创新基础设施的重点建设内容之一。

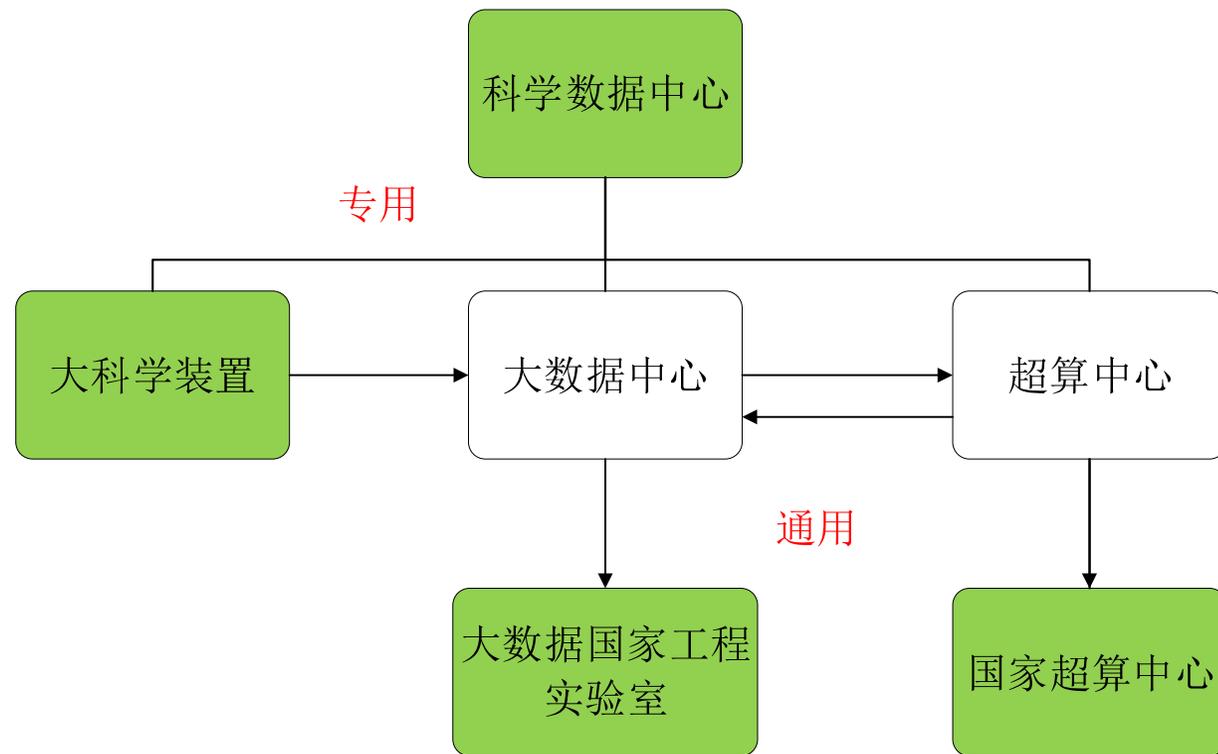
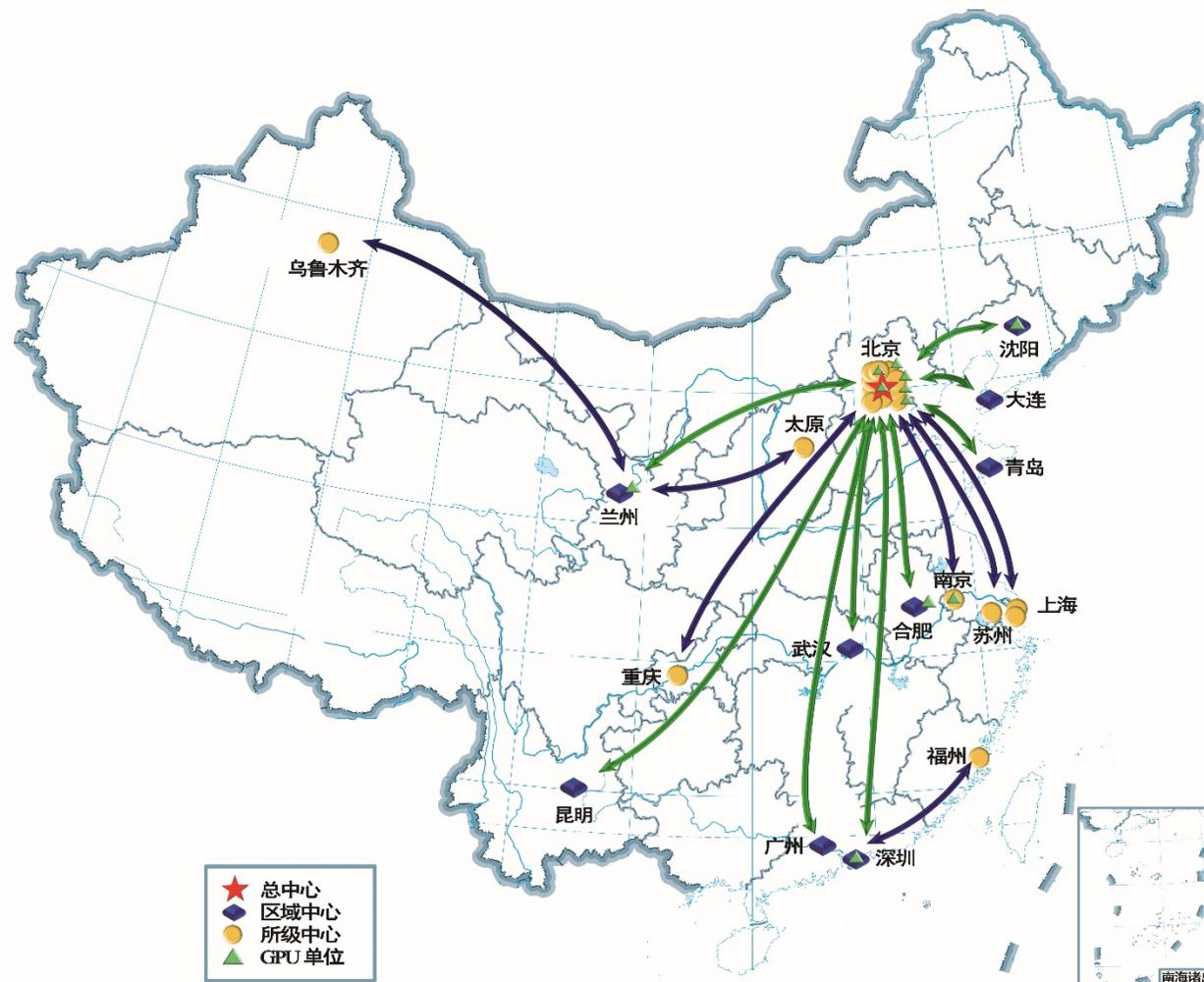


图 我国几个重要算力载体关系

- 中国科学院超级计算环境（ScGrid）由总中心、分中心、所级中心组成的三层架构网格计算环境
- 还连接了中国科学院院内多家单位的GPU计算集群，聚合计算能力超过315PF
- 系统使用率超过 50%，核心算力为“元”超级计算系统
- 软件体系主要包括：MPI（信息处理界面）、各类数学库、各类工具软件、CUDA（计算统一设备架构）、各类应用软件、针对社会用户的付费应用软件



中国科学院超级计算网络



## 一、算力全景图

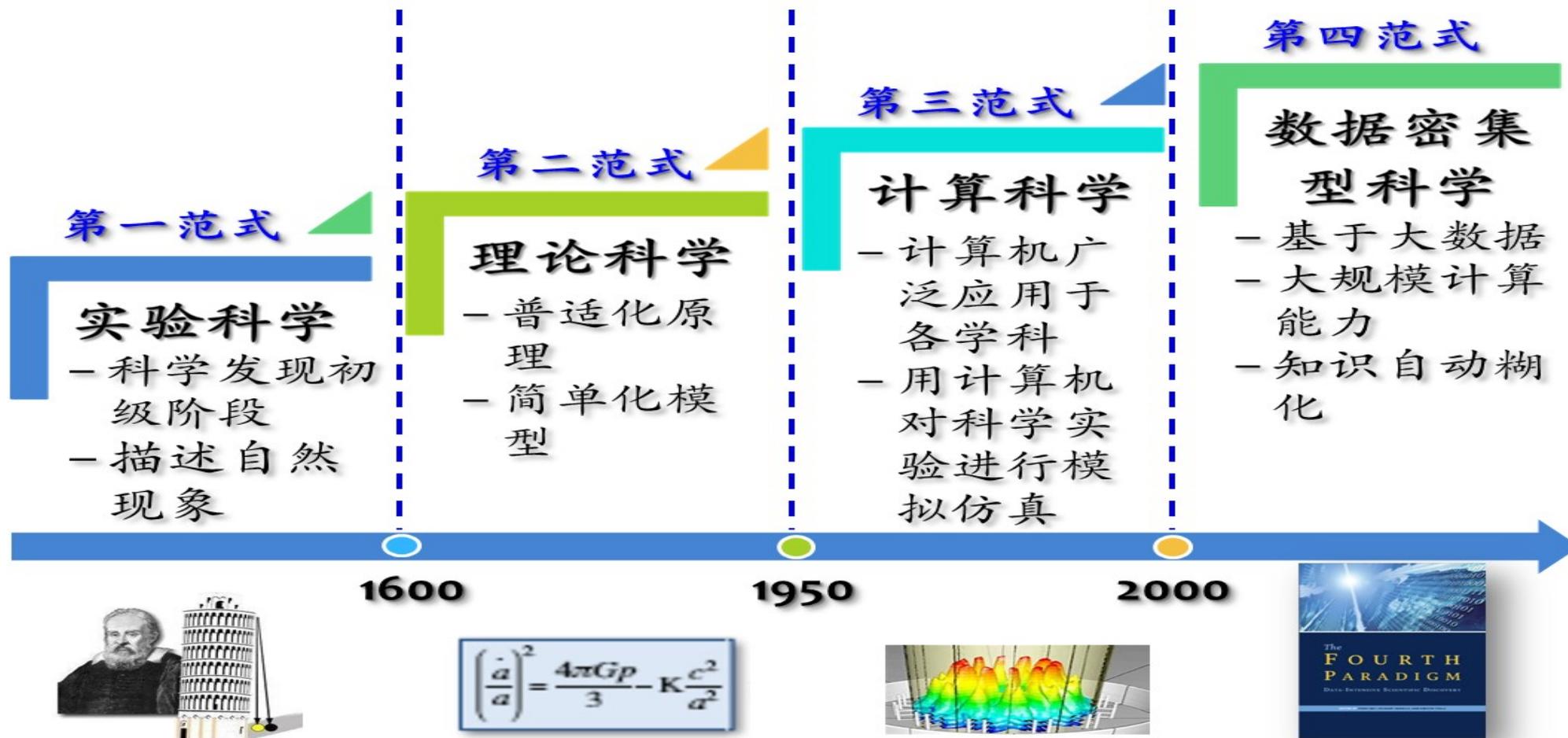
## 二、科研算力需求

## 三、教育算力需求

## 四、未来发展思路与建议

# 科研范式演进下的算力需求发展态势

图灵奖得主吉姆·格雷(Jim Gray) 2007年提出科学范式





## 1. 模型复杂度提高带来的计算需求

随着计算资源和技术的升级迭代，未来模型复杂度的深化将亟待超级计算资源和技术的支撑



## 2. 模型模拟尺度精细化带来计算需求

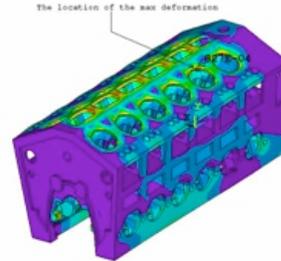
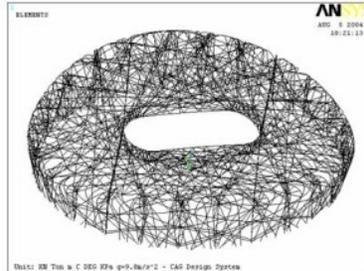
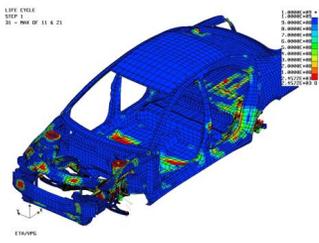
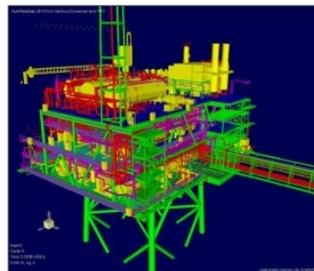
- 未来随着E超级计算平台的研发，各学科模拟研究的分辨率都将逐步精细化，从而引发新的算力需求。

- 格点QCD（量子色动力学）最大的格点体系大小为 $1283 \times 256$ ，计算规模大约为十万或数十万核；
- 未来如果规模提高到 $2563 \times 512$ ，则计算规模将增大到数百万核，必须使用E级计算



## 3. 高性能计算是助力科学研究和工程技术进步的关键

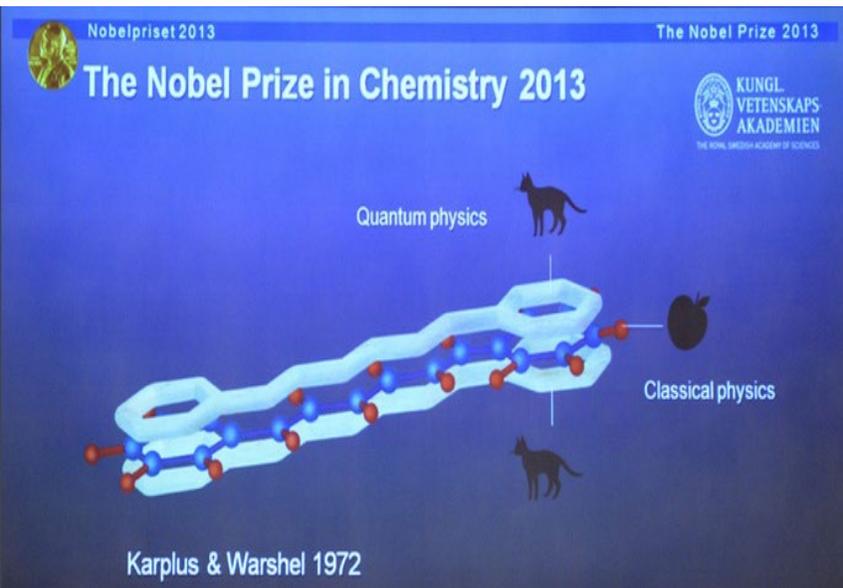
高性能计算可运用有效的算法，快速完成科学研究、工程设计、金融、工业以及社会管理等领域内具有数据密集型、计算密集型和I/O（数据输入输出）密集型的计算



```
ANSYS 5.7
OCT 27 2001
13:36:57
NODAL SOLUTION
STEP=1
SUB =1
TIME=1
USUM
TOP
RSTX=0
SXC = .827E-04
SNC = .827E-04
0
.919E-05
.184E-04
.276E-04
.367E-04
.459E-04
.551E-04
.643E-04
.735E-04
.827E-04
```

## 2013年诺贝尔化学奖

- 化学反应发生的速度堪比光速。多尺度复杂化学系统模型翻开了化学史的“新篇章”，实现对化学反应的每个步骤进行追踪。

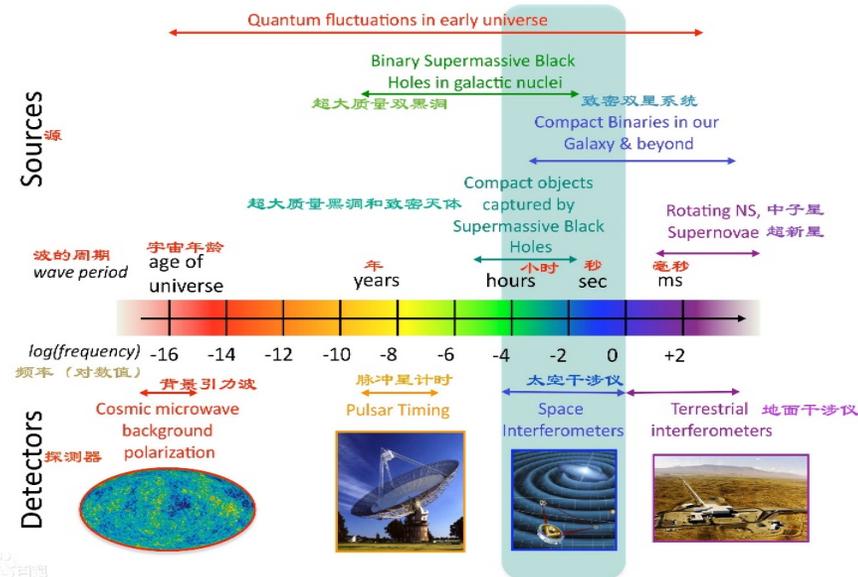


## 2017年诺贝尔物理奖

- 引力波的发现，激光干涉仪前四个月的观测数据量达到**500TB**，数据分析消耗**17亿CPU小时**的计算资源。



## The Gravitational Wave Spectrum



2019年人类首张黑洞照片：3.5PB数据，2万CPU小时计算资源。

- ▶ 密集型数据使传统人工处理方法不再有效，而必须使用更为智能的、自动的、高效的现代数据挖掘技术。



## 算力资源将更加倾向于商品化

未来关于算力共享租赁将逐渐成为一大产业，当前的科研算力租赁模式下的资金主要来源于国家科研项目支持。

## 算力资源呈现出个性化的需求

不同研究领域的科研人员算力需求构成千差万别，即便是同一领域，由于其研究问题的差异，也会表现出不同的硬件和软件资源的需求差异



# 第四范式下数据管理和访问的需求



## 数据格式需求

数据来自不同学科、不同行业，既包括结构化数据，也包括非结构化数据，或者半结构化数据，要实现这些不同类型数据的整合使用需要有统一的数据标准和数据接口，促进数据的融合



## 数据存储需求

科学数据的长期保持和共享将成为重要的环节，以助于数据的重用和验证。面对海量数据，未来将形成数据全生命周期管理，建立对大数据数据集创建、存储、分析、发布、共享的数据管理计划（DMP）



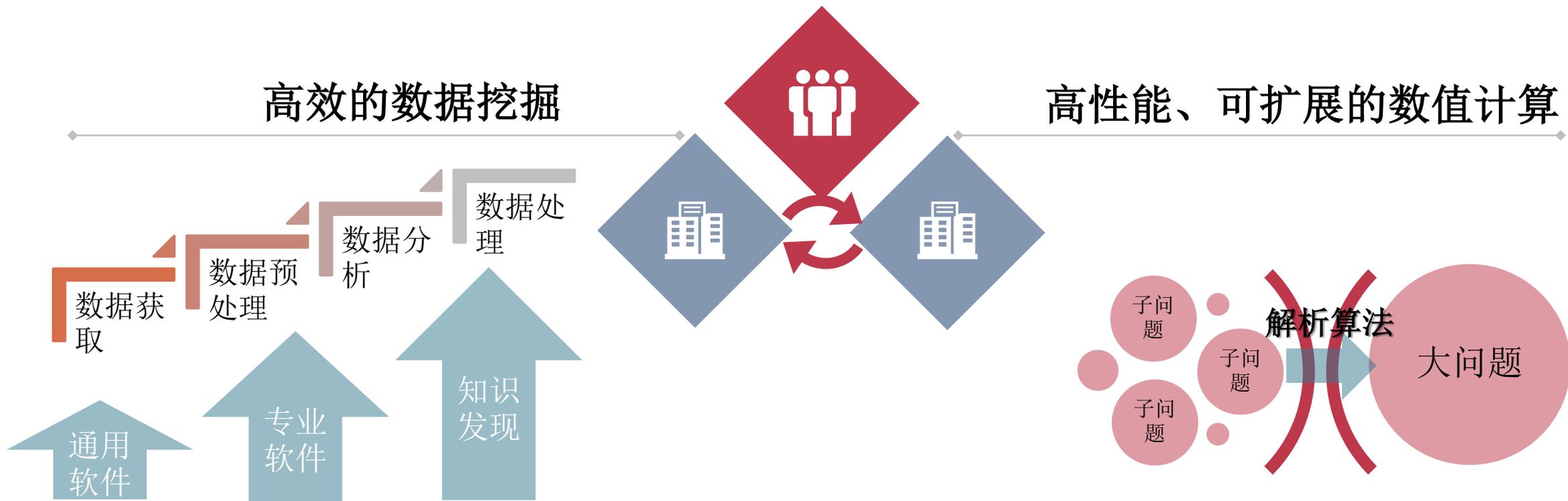
## 数据远程访问需求

第四范式下数据集变得巨大，达到不可能仅仅用文件传输协议（FTP）来传输数据或按给定模式搜索数据文件，这要求数据密集型计算在海量存储和高性能计算平台上实现，需要从远程提供接口和服务

## 算法将构成国家的未来竞争力

高效的数据挖掘

高性能、可扩展的数值计算





## 知识发现

如计算流体力学、计算材料学、计算电磁学

## 协同计算

多学科综合设计，如重工设计

## 超算 算力

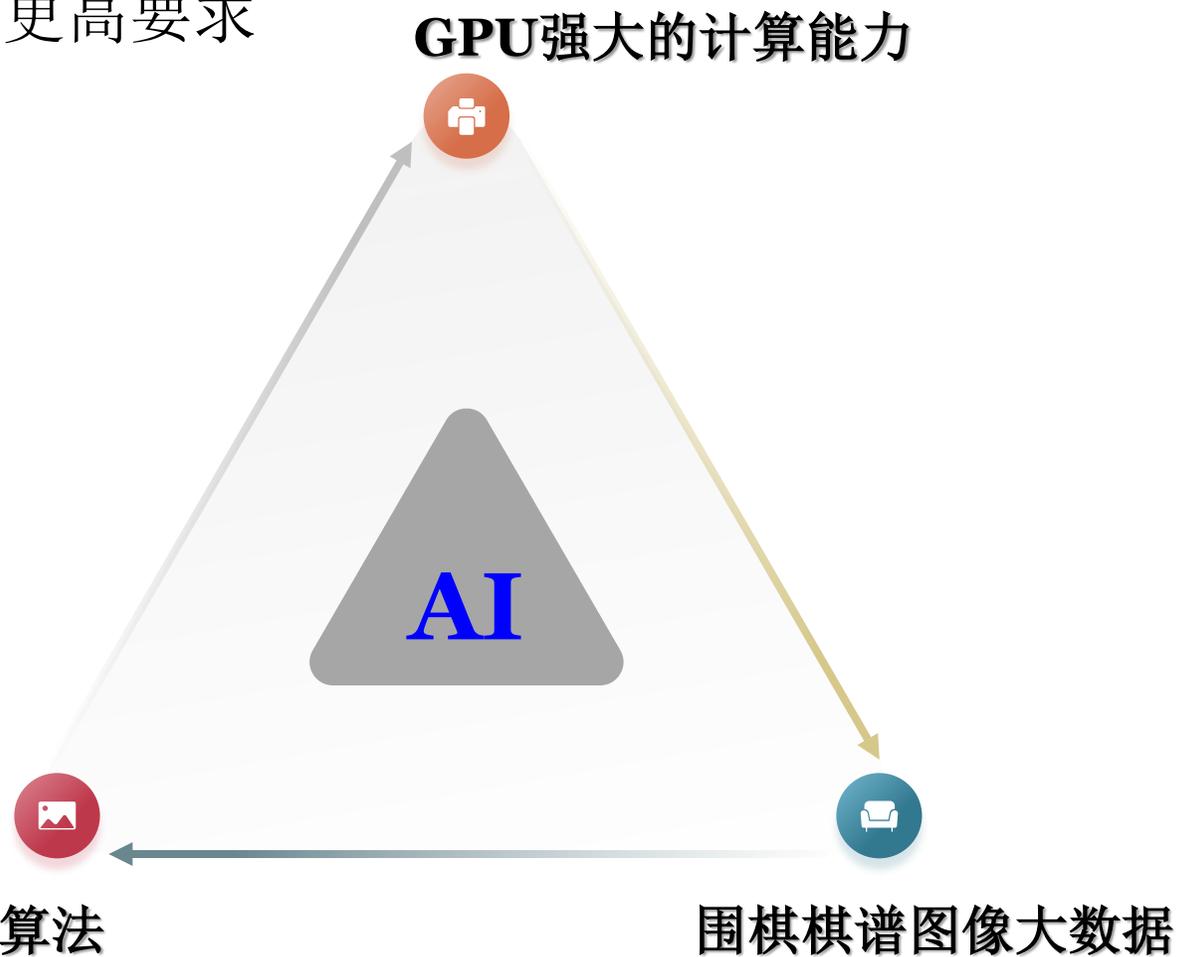
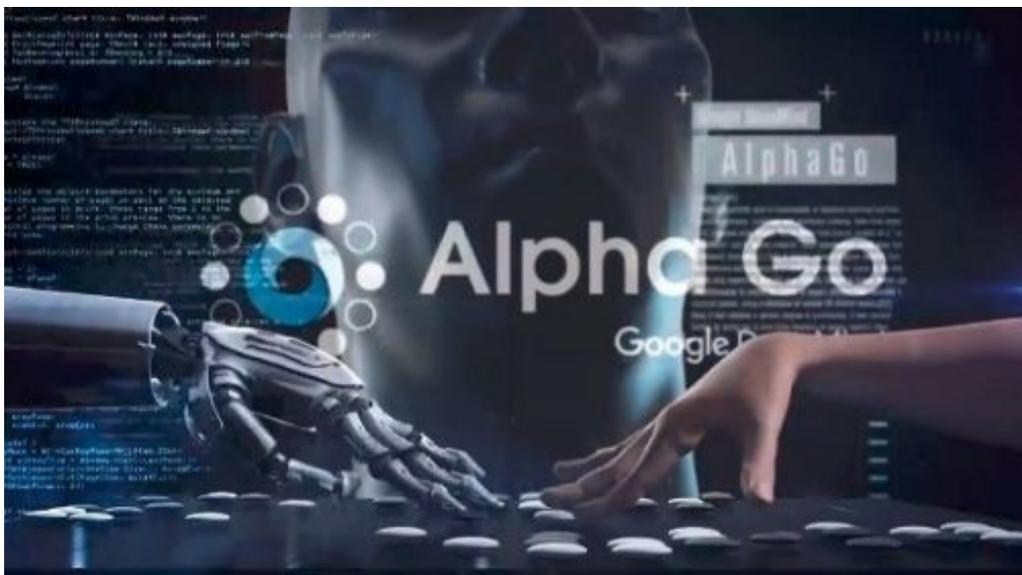
## 仿真计算

如基于仿真的医学实践、数字城市模拟、核电、煤矿仿真工具

## 数据密集型应用

高能物理实验数据处理、遥感数据处理、生物信息学、RFID数据挖掘

AI 技术的应用对数据存储和计算资源也提出更高要求



- 斯坦福大学在深度学习领域测试：12颗NVIDIA GPU可以提供相当于2000颗CPU的深度学习性能！

不同领域的AI应用方向

科学领域	AI应用方向
气候	极端天气检测
天体物理	光晕发现
等离子物理	磁重联跟踪
材料科学	合成材料预测
天文观测	瞬态检测
生物成像	聚类



谷歌、微软、IBM、Facebook等全球IT巨头都投巨资加速超大规模人工智能计算平台的研发



从发展趋势上看，在服务器平台层面开启的新一轮计算模式变革拉开帷幕，是人工智能产业正式走向成熟的拐点



# 我国科研领域算力需求的痛点问题及发展方向



- 算力总体支撑较好，但存在结构性供需错位
- 整体价格较贵，科研人员计算经费紧张
- 科研算力的软件生态严重缺乏，可用软件不多
- 用户操作友好性不强，好用的软件不多
- 网络传输与计算能力不匹配的矛盾较为突出
- 科研人员的数字技术基础差异显著



## 科研算力需求的主要三个方面

硬件基础设施建设——建立多层次的科研基础设施运营架构，协调多方主体关系，充分发挥各方积极性

软件基础设施建设——积极促进专业科研单位与社会组织合作，加快专业软件和通用软件开发

科研数字基础设施运营——建立开放、共享、创新的理念，充分发挥基础设施的社会价值



## 建立人工智能研发与应用服务平台

- 异构计算资源、多源异构数据场景
- 分布式智能计算体系
- 建立开放共享大数据平台
- 发展在线众包新模式
- 隐私安全防护

## 发展人工智能基础软件包

算法理论与学科基础理论模型结合

机器学习技术深入学科应用

构建AI软件生态（汇聚平台、软件开源社区、代码托管与协作平台、软件评估专家系统等）



## 一、算力全景图

## 二、科研算力需求

## 三、教育算力需求

## 四、未来发展思路与建议



## 教育领域算力共享的基本态势：主要发达国家在教育领域提升算力的布局

世界大多发达国家通过终身教育立法推进本国科技人力资源发展，教育（包括继续教育）信息化是信息化时代构建学习型社会和终身学习体系的基本技术途径。

2019年6月，Mary Meeker发布《Internet Trends 2019》，指出在线教育数量庞大，在规模上大幅增长，利用互联网开展终身学习，员工重新培训的参与度更高。如Udemy（全球继续教育）遵循“用户是知识生产者和接受者双重角色”理念，为用户提供3000门工作技能课程并打造企业解决方案。

美国教育领域高度重视信息技术对教育变革的推动作用，教育信息化研究和实践一直走在世界前沿，并为世界各国教育信息化建设和发展提供可借鉴的经验和示范。美国提出《国家教育技术计划（NETP2016）》为美国教育信息化未来五年发展指明了方向，融合学习、教学、领导力、评价、基础设施五大领域构建了“为未来做准备的学习”的基本框架。

## 教育领域算力共享的基本态势：我国高校近年来提升算力的布局

对高校算力提升布局的分析，要围绕高等院校在国家创新系统中的功能定位而展开：学术研究、人才培养。除去与科研领域的共性需求外，还有其特殊性：

第一，算力基础设施条件的制约。由于高校不具有大科学装置，在算力基础设施方面存在“先天制约”，因此无法生成、收集海量数据；一般来说只能获取小规模数据。

第二，现有科研经费管理制度对这种购买服务的活动尚不具备完善的制度规范。为部分克服上述问题，一些高校采取了由学校统一采购软件及其服务，再“分发”给有需求的老师的处理方式。大部分老师根据实际需求有限度使用，这一方面可以大大减轻课题组的科研成本，同时也提高了科研设施的使用效率。但是，这只是少部分机构和课题组所采用的的权宜之计，根据现有课题经费管理办法及经费预算科目，多数情况下，课题组难以列支该笔费用。



## 教育领域算力共享的基本态势：中科院教育信息化体系布局

- 中科院在信息化发展之初就非常重视教育信息化建设，目前已形成覆盖全院的全生命周期一体化教育信息化服务体系，建设了教育全过程管理、自主学习服务、科教资源融合共享的信息化系统。
- 调研中发现，中科院教育信息化体系布局目前存在的问题和需求包括：
  - ① 由于教育决策支持还处于初级探索阶段，基于大数据的教育决策支持能力有待进一步加强；
  - ② 自主学习服务偏重传统模式，需构建以学习测量和自适应学习为主体的学习服务平台，为学习者提供更精准的学习策略；
  - ③ 科教资源融合共享还需进一步探索数字化学习资源开放与服务的模式和机制。



## 我国教育领域科研活动算力共享的应用场景与需求

总体上，教育领域科研活动算力提升与共享，目前主要存在建设和完善基础设施及其相关硬件设备，以及理顺现有资源共享机制的需求。

**第一，技术维度。**近年来“网上校园”“网上课堂”的形式与内涵不断丰富，例如信息化教育服务平台、教学资源存储共享平台、教学信息发布与交流空间以及围绕教学活动建设的网上办事空间等。随着需求的快速复杂化，对算力提升所需的基础设施建设提出了极高要求。

**第二，政策与机制维度。**亟需在制度与机制层面建立起院校两级协同治理的新治理体系、构建不同于传统教学方式的新教学空间并塑造新的围绕教学而展开的学术空间与平台、突破现有的有关科教资源共享与应用的体制机制束缚。



↑↗  
新冠疫情期间，哈佛在线  
教学场景。  
如何建构扁平化的深入教-  
学双方“深度卷入”关系？

□ 从传统的“教-学”单向关系到去中心化、扁平的“深度卷入”，是一种社会关系的重构。

- ① 用户并发数，即同时访问服务器站点的连接数，对瞬间带宽的支持能力要求极高。高并发系统不仅需要良好的架构和技术选型，也需要对实现高并发系统的业务场景进行全链路测试。
- ② 在由不同关系模式所主导的不同教学场景下，对算力的需求存在较大差异。随着AR（增强现实）、VR（虚拟现实）等技术被应用于教学场景，会对算力提出更高要求。



③ 在教学资源产生端，根据不同的课程内容展示模式，对算力的需求会有较大的不同。

在教学资源产生端，根据不同的课程内容展示模式，对算力的需求会有较大不同。随着附加信息的增多及不断复杂化，从文字到音频到视频，随着编码难度增加，对算力要求也必然增加。

在资源分配端，例如流服务器，接入端越多，对算力的需求就越大。在服务器数量一定的情况下，接入端越多，单个服务器处理进程的效率就要越高。对于接入端数量不确定的场景，通常需要采取算力资源自适应可拓展的架构方案。

对于资源的接受端，根据解析度和信息流压缩率，对解压缩还原信息所需的算力要求也有所不同。

新教学资源获取和教学形式对算力提出的部分新要求：

- ① 为实现增强现实（AR）所需要的图像识别、知识数据库支撑以及信息图层叠加显示的能力；为虚拟现实（VR）功能，额外需要三维建模以及实时确定物体位置姿态并显示的能力。
- ② 除课件展示外，远程教学所需要的交互功能，比如实时留言，对话，以及在线批改作业的机器人，在线远程虚拟机，远程编译运算，远程工程建模和仿真测试等。



## 一、算力全景图

## 二、科研算力需求

## 三、教育算力需求

## 四、未来发展思路与建议



## □ 建立权威和有影响力的科学数据品牌

- 目前，主要有影响力的数据库与数据集主要掌握在发达国家。如材料、大气海洋、遥感观测等领域数据库集中在美、日、欧这几个国家。

研究方向	权威数据库	所属国家
化学领域	TOXNET、Cambridge Structural Database、SciFinder、IRISPubChem等	美、英、德、荷
遥感观测领域	MODIS, GOSAT, ALOS, Landsat, NOAA, Google Earth, Sentinel, 等	美、欧、日
大气海洋领域	气象台站数据, WAS, TOGA, CFSR, ECMWF, NCEP2等	美、欧、日
能源领域	IAEA-NDS, AIAA, ASME等	美、欧、日
材料领域	NOMAD, ICSD, NIST Standard Reference Data等	美、日、欧



# 我国科研算力共享思路



## □ 发展自主可控的科学数据与高性能计算软件

- 目前大部分科学计算软件都是源自国外。国内自主可控软件以**ANSYS**为代表的不多见。

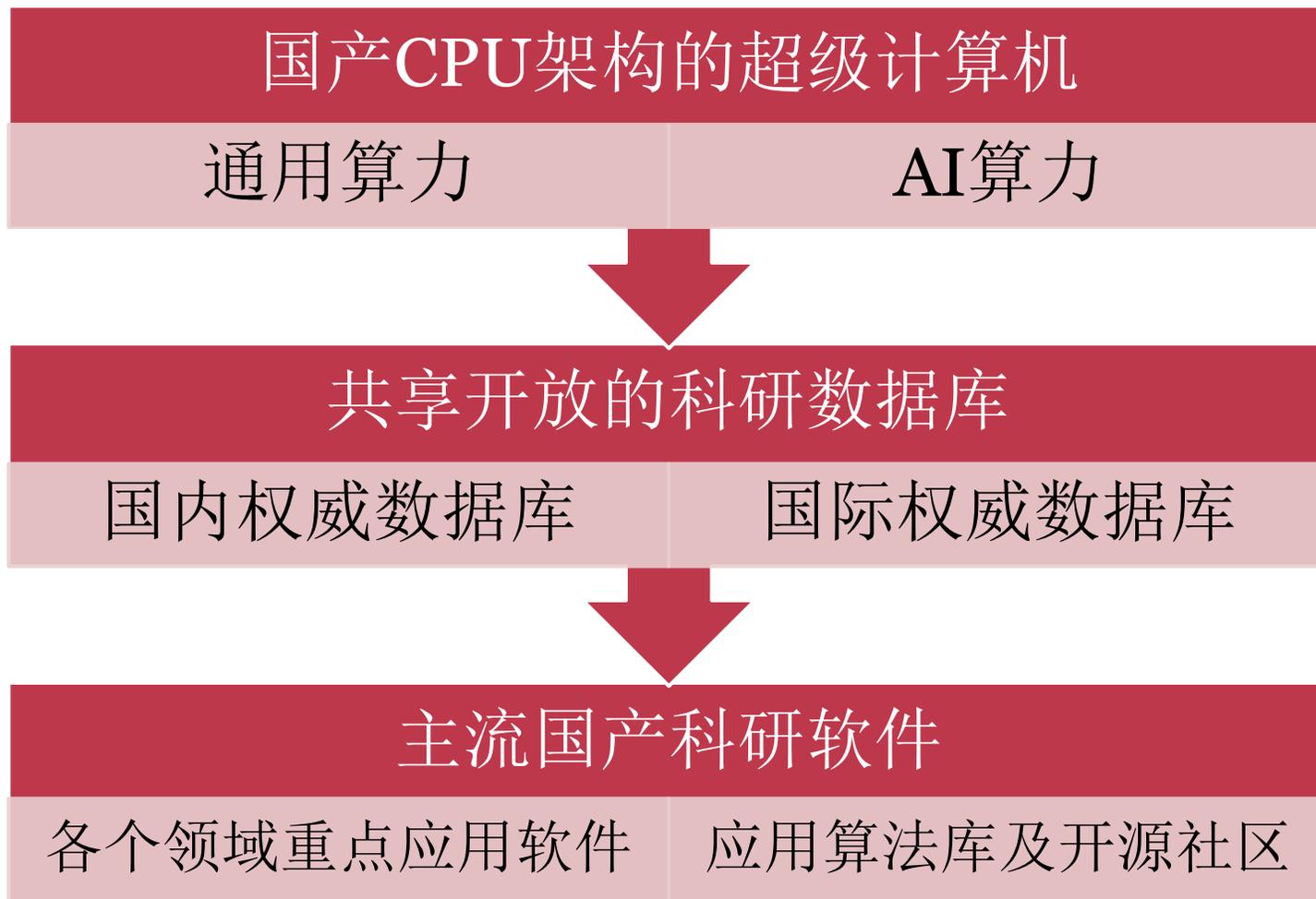
学科领域	软件主要方向	典型代表软件名称
数学与物理	数值计算	MATLAB、Fortran、Python
	有限元分析	ANSYS（中国）、MATLAB、ADINA、ABAQUS
	计算流体力学	ANSYS（中国）、EDEM、Comsol、Xflow
化学	量子化学	Gaussian、ADF
	分子动力学	CHARMM、AMBER
天文学	宇宙演化模拟	AREPO、ATHENA、Illustris、Cubep3m
地学	遥感与地理信息系统	Erdas、ArcGis、Envi
	气候模式	MM5、CCM、WRF
海洋	海洋模式	OGCM、MOM4、HYCOM、LICOM（中国）



# 我国科研算力共享思路

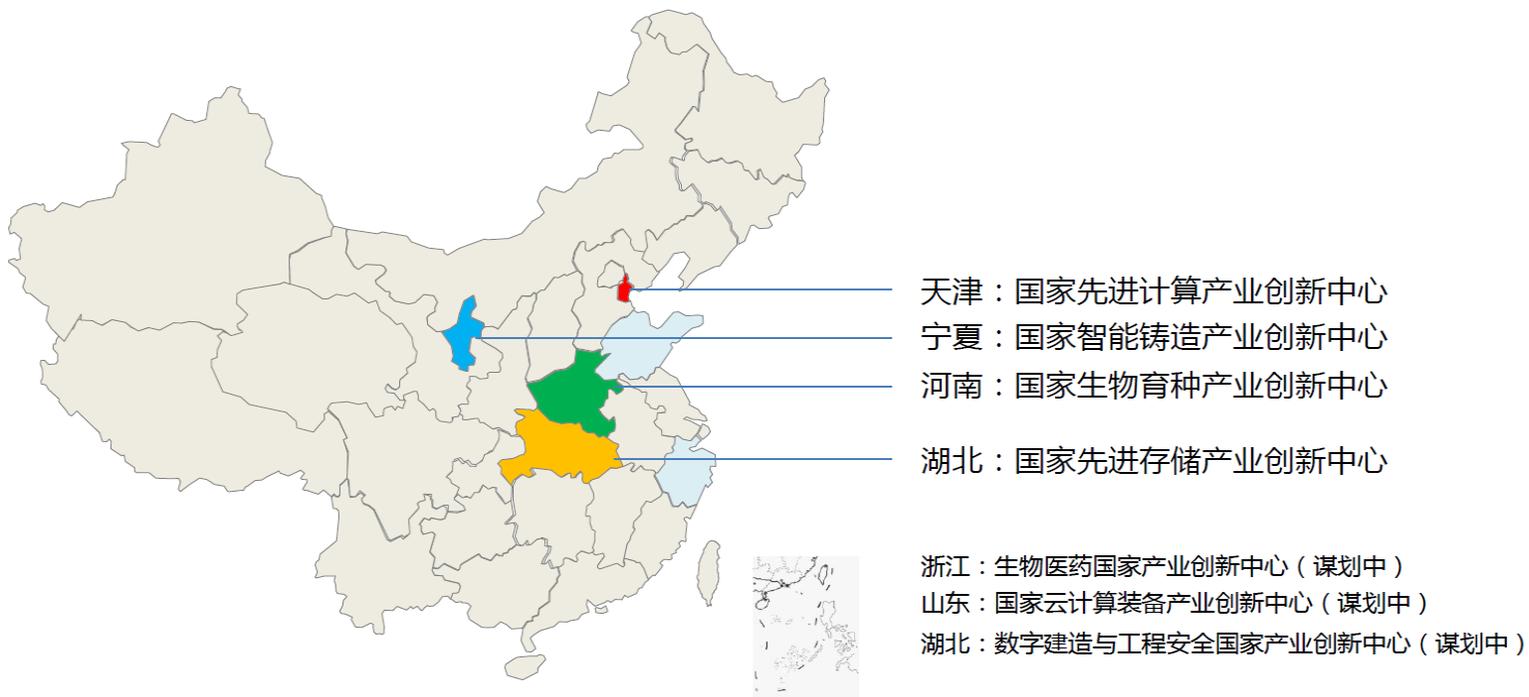


□ 以国产高性能计算机为主线，大力促进算力应用生态建设



## □ 建设国家先进计算科教创新中心

在算力产业中，国家先进产业创新中心已经落户天津，先进存储产业创新中心落户湖北。产业创新中心对推动我国算力设施和产业发展发挥了重要作用。为推进先进计算产业中心的进一步在垂直领域的发展，建议在国家先进产业创新中心分设国家先进计算科教创新中心，直接面向科研教育领域的算力发展。算力共享的发展思路可以作为先进计算科教创新中心的一个重要发展方向。





## 算力共享纳入“十四五”规划

- 创新基础设施规划内容
- 模式优化与布局

## 实施算力共享项目

- 通用算力共享项目
- AI算力共享项目

## 完善资金保障

- 建设资金
- 使用资金
- 维护资金



---

**敬请批评指正！**  
**谢谢！**