



中国科大超级计算中心及主要超算系统简介

李会民

中国科学技术大学超级计算中心

2014年4月8日



1 超级计算中心简介

- 建设背景
- 机构及人员设置
- 职责及可提供的服务
- 超算中心主页

2 超算中心超算系统简介

- IBM JS22刀片集群
- 联想深腾7000G GPU集群
- 刀片及胖节点超级计算系统
- ChinaGrid高性能计算集群
- 招标中的超算系统等

3 联系信息



- 2002年底，学校决定建设第五公共服务平台，即科学计算平台（超级计算平台，简称超算平台）
- 2003年10月初步建成中国科大-中国惠普高性能计算联合实验室，后经几年的不断发展扩建，逐步建设成为现在的中国科学技术大学超级计算中心（简称超算中心）
- 本着**推动科学计算、促进人才培养**的建设方针，中心面向校内所有**有高性能计算需求**的科研院系、实验室、教师、学生**提供高性能计算服务**



- 校五大**公共实验中心**之一
- 设有**专家委员会**:
 - 主要负责对超算中心的建设、运行、管理和发展等提出指导和咨询意见，对超算中心的重大问题，包括软硬件建设和运行资金的筹措与资源的使用等做出决策和决定
 - 组长：陈国良院士
 - 成员：丁泽军、李京、杨寿保、杨金龙、陆夕云、陈华平、陈晓非、傅尧
- 挂靠校**网络信息中心**
 - 主任：**李京**（网络信息中心主任）
 - 副主任：**张焕杰**（网络信息中心副主任）
- 专职人员
 - 李会民：计算物理博士，主管
 - 张运动：计算机在读硕士，主要负责系统日常管理、开户及用户支持
 - 沈瑜：物理化学博士，主要负责应用软件方面的技术支持、与用户合作进行应用开发



职责及可提供的服务 I

负责超算平台的运行管理和维护、开发工作，规划超算的硬件平台与软件资源建设，为用户提供超算及相关技术支持

- 计算平台:

- 为科学计算、工程计算等提供所需要的超算资源
- 当前**免费对校内师生开放**，用户作业排队运行
- **缴费用户可以独占计算节点随时使用，无需与其他用户一起排队**
- 新超算系统建成后，将在**保留部分免费资源的情况下，用户需付费使用超算资源**，以避免由于免费导致的资源浪费，保证真正有需求用户的科研活动

- 对用户提供技术支持:

- 用户程序的编译、安装和调试
- 用户程序运行中的遇到的问题等
- 用户自己超算系统的建设及维护

- 与用户合作开发:

- 与用户合作进行计算模拟等方面合作
- 协助用户与校外超算方面的单位进行合作开发



- 托管服务: <http://scc.ustc.edu.cn/tgfw/>
 - 协助解决部分校内院系、实验室自建超算系统所面临的占地空间、电力、制冷、设备维护、管理人员等难题
 - 提供约20个服务器专用机柜用于托管用户的计算服务器、存储等设备
 - 超算机房按照高标准建设，单机柜可支持功耗约25KW的高密度发热需求，主要适合托管高发热密度的计算服务器及配套设备：
 - 计算服务器类型：刀片式服务器、机架式服务器¹
 - 高发热密度计算服务器：电源模块为N+N冗余，一般不接UPS，可直接两路市电，一路市电停电后，另一路市电可以完全支撑其运行的
 - 高发热密度计算服务器的配套设备：如存储阵列等，如设备不支持双路市电直接供电但非常重要，可以申请通过120KVA UPS供电
 - 收费：政策尚未最终确定
- 超算用户在计算成果产生后的相关论文等中有义务致谢说明：
 - 中文：“本论文的数值计算得到了中国科学技术大学超级计算中心的计算支持和帮助。”
 - 英文：“The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of University of Science and Technology of China.”

¹不接受塔式的或组装机等



网址: <http://scc.ustc.edu.cn>

- **系统平台**: 系统平台介绍
- **新闻公告**: 超算中心公告
- **业界动态**: 超算方面的新闻动态
- **培训信息**: 用户培训信息
- **计算项目**: 现有的部分用户项目
- **运行监控**: Ganglia系统负载及LSF作业监控
- **用户申请**: 申请账户
 - **登录软件下载**
 - **VPN账户申请**: VPN用于从校外登录超算系统
- **资料手册**: C/C++ Fortran编译器、MPI、OpenMP、CUDA、Linux、专业软件等以及培训资料
- **联系方式**
- **超算中心门户**: <http://scc.ustc.edu.cn/i>, 登录后可以查看自己的作业运行统计等



1 超级计算中心简介

- 建设背景
- 机构及人员设置
- 职责及可提供的服务
- 超算中心主页

2 超算中心超算系统简介

- IBM JS22刀片集群
- 联想深腾7000G GPU集群
- 刀片及胖节点超级计算系统
- ChinaGrid高性能计算集群
- 招标中的超算系统等

3 联系信息



- 运行中的超算系统 (CPU: 3656核, 43.64万亿次/秒; GPU: 216万
亿次/秒):
 - **ChinaGrid高性能计算集群**: 704 CPU核, 14.64万亿次/秒, 56Gbps
InfiniBand高速计算网
 - **刀片及胖节点超算系统**: 1528 CPU核, 15.3万亿次/秒, 千兆以太网
 - **联想深腾7000G GPU集群**: CPU: 800核, 7.2万亿次/秒; GPU 216万
亿次/秒; 20Gbps InfiniBand高速计算网
 - **IBM JS22刀片集群**: 112 CPU核, 1.8万亿次/秒, 10Gbps InfiniBand高
速计算网
 - **联想深腾1800集群** (一般不再接受新用户): 512 CPU核, 4.7万亿
次/秒, 千兆以太网
- **招标中的超算系统** (预计下学期可正式使用): 2680 CPU核, 60万亿
次/秒, 56Gbps InfiniBand高速计算网
- 已退役的超算系统
 - HP Superdome服务器集群
 - HP Rx2600服务器集群
 - 基于龙芯2号国产万亿次高性能计算机KD-50-I



IBM JS22刀片集群

- 计算节点: 28个JS22刀片, 每刀片两颗4.0GHz IBM POWER6 64位双核处理器 (共112 CPU核), 16GB内存, 146GB硬盘
- 浮点计算能力: 1.8万亿次/秒
- 计算网络: 10Gbps InfiniBand高速计算网
- 存储: 2TB IBM DS4800
- 操作系统: AIX 6.1
- 编译器: IBM XL C/C++/Fortran编译器
- 并行环境: IBM Parallel Environment(PE), 支持节点内OpenMP等共享内存并行, 节点内和节点间MPI并行
- 数学函数库: 数学加速子系统MASS、基本线性代数函数库BLAS、工程与科学子程序库ESSL及其并行版本PESSL
- 作业管理: TWS LoadLeveler
- 启用时间: 2008年4月
- 适合应用: 对主频和通信要求高的应用程序, 如VASP、SIESTA等; 如为二进制版本, 需为AIX和POWER平台的, 源代码的基本不受此限制; 本系统较特殊, 请了解后再申请



- **计算节点**: 100个, 每节点**两颗2.26GHz Intel XEON E5520 64位四核处理器 (共800核)**, **16GB内存**, **500GB硬盘**, 其中:
 - 16个节点配置有一块**NVIDIA C2050 GPU卡**
 - 18个节点配置有两块**NVIDIA C1060 GPU卡**
 - 10个节点配置有两块**NVIDIA GTX295 GPU卡**
- **浮点计算能力**: **CPU: 7.2万亿次/秒; GPU: 216万亿次/秒**
- **计算网络**: **20Gbps InfiniBand高速计算网**和**千兆以太网**
- **存储**: 2TB
- **磁盘限额**: **50GB/用户**, 如需更大空间, 请提出申请
- **操作系统**: CentOS 5.5 x86_64
- **编译器**: **Intel和GNU C/C++/Fortran编译器, NVidia CUDA编译器**
- **数学函数库**: Intel MKL(BLAS, LAPACK, ScaLAPACK, FFT, VML, VSL, Data Fitting Library, Extended Eigensolver)



- 并行环境: **Open MPI**和**CUDA**, 支持:
 - MPI等分布式内存程序: 节点内和节点间
 - OpenMP等共享内存程序: 节点内
 - CUDA GPU并行: GPU节点
- 作业管理: Platform LSF
- 启用时间: 2010年1月
- 适合应用:
 - 不支持串行作业
 - **CUDA GPU**计算
 - 对计算网络要求高的大规模**MPI**并程序, 如VASP、SIESTA、Fluent



- 队列信息²:

- CPU计算队列:

- **small**: 可运行12~64进程并行作业, 提交作业时需加参数-q small
- **large**: 可运行64~128进程并行作业, 提交作业时需加参数-q large

- GPU计算队列³:

- **c2050**: 运行于node29~node44节点, 每节点1颗C2050 GPU卡, 提交作业时加参数-q c2050
- **c1060**: 运行于node1~node16节点, 每节点2颗C1060 GPU卡, 提交作业时加参数-q c1060
- **gtx295**: 运行于node19~node28节点, 每节点2颗GTX295 GPU卡, 提交作业时加参数-q gtx295

²队列会根据运行情况调整, 请注意登录提示

³非GPU作业不得使用GPU队列



刀片及胖节点超级计算系统 I

主要由曙光CB60-G、CB60-G2、联想B710等刀片式服务器和曙光A950、曙光I620r-G、曙光I420r-G、浪潮TS850、联想万全R630G7、HP DL580 G7等机架式服务器构成，共**102个计算节点1512CPU核心**，总双精度峰值计算能力为**15.30万亿次/秒**：

- **管理和用户登录节点**：
 - 节点名为sugon
 - 用户主登录节点，可以进行编译与通过作业调度系统提交作业，**禁止直接在此节点上运行作业**
 - 曙光I620机架式服务器一台，2颗主频2.40GHz的Intel XEON E5620 x86_64 4核CPU（共8核），8GB DDR3 1333MHz内存
- **存储**：13TB曙光DS600-F20 SAS磁盘阵列，用户主目录/home所用
- **磁盘配额**：**50GB/用户**，如需更大空间，请提出申请
- **计算网络**：**千兆以太网**



- 刀片或机架式计算节点：用户都可使用

节点名	数量	CPU	内存(GB)
node1-node70	40	Intel XEON X5650, 共12核, 主频2.66GHz	24
node71-node80	10	Intel XEON X5650, 共12核, 主频2.66GHz	48
node93-node102	11	Intel XEON E5620, 共8核, 主频2.4GHz	16

- 胖计算节点：更多CPU核大共享内存节点，数量有限，有些需特殊申请

节点名	数量	CPU	内存(GB)
node81-node83	3	Intel XEON E7540, 共48核, 主频2.0GHz	128
node84-node85	2	Intel XEON E7-8837, 共64核, 主频2.66GHz	256
node86-node87	2	AMD Opteron 8431, 共48核, 主频2.40GHz	128
node88	1	AMD Opteron 6168, 共24核, 主频1.90GHz	32
node89-node90	2	Intel XEON E7430, 共16核, 主频2.13GHz	64
node91-node92	2	Intel XEON E7-4807, 共24核, 主频1.87GHz	96

- 操作系统：x86_64架构的64位CentOS 6.5 Linux
- 编译器：Intel、PGI、GNU等C/C++ Fortran编译器
- 数值函数库：Intel MKL



- 并行环境: Open MPI和Intel MPI, 支持:
 - MPI等分布式内存程序: 节点内和节点间
 - OpenMP等共享内存程序: 节点内
- 资源管理和作业调度: Platform LSF
- 常用公用软件安装目录: */opt*, 请自己查看有什么软件, 有些软件需要在自己`~/.bashrc`等配置文件中设置后才可以使用的
- 建设时间: 2010年-2012年
- 适合应用:
 - 串行作业
 - 对计算网络要求不高的应用, 如MATLAB、VASP和SIESTA等应用
 - 某些需要单节点大内存多CPU核的应用, 如Gaussian等



● 队列信息⁴:

- **serial**: 1~2进程作业队列, 提交作业时需加参数-q serial, 每用户可最多同时运行24个serial队列作业
- **normal**: 仅运行10~12 CPU核的作业, 运行于node1~node80其中之一节点, 提交作业时需加参数-q normal
- **long**: 可运行大于12 CPU核的多节点并行作业, 运行于node1~node80节点, 提交作业时需加参数-q long, 每用户最多可运行120进程的作业
- **mem48**: 可运行每进程需求内存较大但一个作业总需求内存小于48GB的作业, 运行于node71~node80某一节点上, 提交作业时需加参数-q mem48, 队列优先级较高
- **mem64**: 可运行每进程需求内存较大但一个作业总需求内存小于64GB的作业, 运行于node89~node90某一节点上, 提交作业时需加参数-q mem64, 队列优先级较高
- **mem96**: 可运行每进程需求内存较大但一个作业总需求内存小于96GB的作业, 运行于node91~node92某一节点上, 提交作业时需加参数-q mem96, 队列优先级较高



- **fat48**: 运行大共享内存作业，每节点48 CPU核，提交作业时需加参数-q fat48 -R "rusage[mem=**]", **为以MB为单位的每进程内存数，**需特殊申请使用权限**
- **fat64**: 运行大共享内存作业，每节点64 CPU核，提交作业时需加参数-q fat64 -R "rusage[mem=**]", **为以MB为单位的每进程内存数，**需特殊申请使用权限**
- **8cpu**: 作业运行在node93~node102节点上，作业设置请使用进程数为8的倍数，提交作业时需加参数-q 8cpu

⁴大内存服务器，CPU单核比不上普通队列的，主要供对内存有大需求的作业使用，请勿乱用



由浪潮NF5260M3机架式服务器和浪潮NX5440、曙光TC4600、联想B710等刀片服务器构成，共一个管理和用户登录节点、一个存储节点及**44个计算节点704个CPU核心**，总双精度峰值计算能力为**14.64万亿次/秒**：

- **管理和用户登录节点**：
 - 节点名为chinagrid
 - 用户主登录节点，可以进行编译与通过作业调度系统提交作业，**禁止直接在此节点上运行作业**
 - 浪潮NF5260M3机架式服务器一台，两颗64位主频2.60GHz的Intel XEON E5-2670 x86_64 8核CPU，共16核，32GB内存
- **存储**：可用空间9TB，浪潮SA5212H2机架式存储服务器，6块2TB SATA硬盘
- **磁盘配额**：**50GB/用户**，如需更大空间，请提出申请



- 刀片计算节点：
 - 节点名: node1-node44
 - 各节点配置: 两颗64位主频2.6GHz的Intel XEON E5-2670 x86_64 8核CPU (共16核)、32GB内存及300GB SAS硬盘
- 计算网络: 56Gbps InfiniBand高速计算网和千兆以太网
- 管理网络: 千兆以太网
- 操作系统: x86_64架构的64位CentOS 6.5 Linux
- 编译器: Intel、PGI和GNU等C/C++ Fortran编译器
- 数值函数库: Intel MKL
- 并行环境: Intel MPI和Open MPI等, 支持:
 - MPI等分布式内存程序: 节点内和节点间
 - OpenMP等共享内存程序: 节点内
- 资源管理和作业调度: IBM Platform LSF
- 常用公用软件安装目录: /opt, 有些软件需要在自己~/.bashrc等配置文件中设置后才可以使用的
- 启用时间: 2013年3月



- 适合应用:
 - 不支持串行作业
 - 对计算网络要求高的大规模MPI并行应用，如VASP、SIESTA和Fluent
- 队列信息:
 - **normal**: 每用户最多可同时运行16 CPU的作业，单作业最少8 CPU，最多16 CPU，所有用户在此队列最多同时使用160 CPU核
 - **long**: 每用户最多可同时运行128 CPU的作业，单作业最少24 CPU，最多128 CPU，所有用户在此队列最多同时使用704 CPU核
 - **normal**和**long**队列同一用户同时最多使用160作业进程数，作业最大运行时间15天，**bqueues -l**查看各队列设置的详细说明



● 招标中的超算系统

- 招标预算：700万元
- 开标时间：4月23日
- 安装调试完成时间：预计7月
- 正式使用：预计下学期初
- 计算能力：2680 CPU核心，60万亿次/秒
- 主要硬件配置：
 - 计算节点：134台刀片，每刀片两颗主频2.8GHz的Intel E5 2680V2 10核CPU（共20核），64GB内存，300GB SAS硬盘
 - 存储：裸容量67TB，并行文件系统，IO带宽高于1GBps
 - 计算网络：56Gbps全线速InfiniBand高速计算网
- 适合应用：对计算网络要求高的大规模MPI并行应用，如VASP，SIESTA，Fluent

● 其它准备建设的超算系统：数量不多，主要为大家提供一个学习测试平台，等有了足够应用再大批量扩充

- Intel MIC：MIC(Many Integrated Core)架构芯片则专门执行高度并行的计算密集型任务，Intel XEON Phi
- GPGPU：GPGPU全称General Purpose GPU，即通用计算图形处理器，适合高度并行的计算密集型任务，NVIDIA Tesla GPU



- 办公室：科大东区新图书馆一楼东侧126室
- 日常联系：
 - 电话：0551-63602248
 - 公共邮箱：sccadmin@ustc.edu.cn
 - QQ用户交流群：8355136
- 工作人员：
 - 李会民：0551-63600316, hmli@ustc.edu.cn
 - 张运动：0551-63602248, zhangzyd@ustc.edu.cn
 - 沈瑜：0551-63602248, shenyu@ustc.edu.cn
- 主页：<http://scc.ustc.edu.cn>



问题?