



超级计算平台建设

李会民

hmli@ustc.edu.cn

中国科学技术大学 超级计算中心

2014-07-18



1 要不要建?

2 超算系统建设

3 联系信息

为什么建？



- 教学、科研需要

为什么建？



- 教学、科研需要
 - 培养学生、科学计算等

为什么建？



- 教学、科研需要
 - 培养学生、科学计算等
- 面子、跟风

租还是建？



- 自建：



- 自建：
 - 优点：网速快、对外宣传、申请经费、培养学生、可按需要调整



- 自建：

- 优点：网速快、对外宣传、申请经费、培养学生、可按需要调整
- 缺点：成本高，对建设人员要求高

租还是建？



- 自建：
 - 优点：网速快、对外宣传、申请经费、培养学生、可按需要调整
 - 缺点：成本高，对建设人员要求高
- 租

租还是建？



- 自建：
 - 优点：网速快、对外宣传、申请经费、培养学生、可按需要调整
 - 缺点：成本高，对建设人员要求高
- 租
 - 优点：省钱、省心、省人



- 自建：

- 优点：网速快、对外宣传、申请经费、培养学生、可按需要调整
- 缺点：成本高，对建设人员要求高

- 租

- 优点：省钱、省心、省人
- 缺点：受对方控制，网络慢



- 经费支持



- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保



- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保
- 电费：



- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保
- 电费：
 - 制冷所需要的电费一般按照70%匹配

- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保
- 电费：
 - 制冷所需要的电费一般按照70%匹配
 - 一台300W的服务器一年的电费约：
$$0.5653\text{元/kwh} \times 0.3\text{kw} \times (1+0.7) \times 24\text{H} \times 365\text{Day} = 2525\text{元/年}$$



- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保

- 电费：

- 制冷所需要的电费一般按照70%匹配

- 一台300W的服务器一年的电费约：

- $$0.5653\text{元/kwh} * 0.3\text{kw} * (1 + 0.7) * 24\text{H} * 365\text{Day} = 2525\text{元/年}$$

- 用户和利用率：硬件5年生命周期，最好两三年内能用满

- 经费支持

- 维保费：计算机服务器一般为3年或5年维保，空调、UPS等维保

- 电费：

- 制冷所需要的电费一般按照70%匹配

- 一台300W的服务器一年的电费约：

- $$0.5653\text{元/kwh} * 0.3\text{kw} * (1 + 0.7) * 24\text{H} * 365\text{Day} = 2525\text{元/年}$$

- 用户和利用率：硬件5年生命周期，最好两三年内能用满

- 经费持续性投入

- 电：
 - 是否用UPS？空调系统如停电，那么温度很快就会上升上来；只对服务器加UPS用处不大
 - 双路市电直接供电
- 湿度：加湿除湿，相对湿度 40% ~ 70%之间，防静电，防止结露，加湿除湿
- 温度：一般控制在22°C左右
 - 上送风：适合小型低密度机房
 - 下送风：3-6KW/机柜
 - 冷热通道：2-30KW/机柜
 - 机柜级制冷：10-50KW/机柜
- 承重：地板集中载荷应 $\geq 3550N$ ，均布载荷 $\geq 16500N/m^2$
- 防静电：避免静电对电子器件的损坏
- 防雷：
- 洁净度：通常对于直径大于 $5\mu m$ 的尘埃，要求其浓度小于 $3 * 10^4$ 粒
- 消防：有管网或无管网
- 动力环境监控：温适度、电力、火警等

- 外包

- 外包
- 自主维持

- 外包
- 自主维持
 - 系统管理

- 外包
- 自主维持
 - 系统管理
 - 用户技术支撑

- 外包
- 自主维持
 - 系统管理
 - 用户技术支撑
 - 应用开发



1 要不要建?

2 超算系统建设

3 联系信息



- 用户应用特点
- 未来3-5年需求：并行规模、总计算量、存储需求

按照实际应用特性选型，Linpac值仅供参考

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU



按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU

- CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>



按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU

- CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
- 协处理器: 是否存在足够的应用能利用起来

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU

- CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
- 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU

- CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
- 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)
 - 通信要求高: InfiniBand (高带宽: 56Gbps; 低延迟: $< 1\mu s$)

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)
 - 通信要求高: InfiniBand (高带宽: 56Gbps; 低延迟: $< 1\mu s$)
- 操作系统: 极少Windows, 一般为Linux (RHEL、CentOS、SuSE)

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)
 - 通信要求高: InfiniBand (高带宽: 56Gbps; 低延迟: $< 1\mu s$)
- 操作系统: 极少Windows, 一般为Linux (RHEL、CentOS、SuSE)
- C/C++、Fortran编译环境: GCC、Intel、PGI等编译器

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)
 - 通信要求高: InfiniBand (高带宽: 56Gbps; 低延迟: $< 1\mu s$)
- 操作系统: 极少Windows, 一般为Linux (RHEL、CentOS、SuSE)
- C/C++、Fortran编译环境: GCC、Intel、PGI等编译器
- 数值函数库: BLAS、LAPACK、ScaLAPACK、Intel MKL、AMD ACML等

按照实际应用特性选型，Linpack值仅供参考

- CPU、MIC、GPU
 - CPU: 架构(X86_64、Intel Itanium、IBM POWER)、主频、核数、Cache, <http://www.spec.org/>
 - 协处理器: 是否存在足够的应用能利用起来
 - MIC: Intel Many Core, Intel Xeon Phi协处理器
 - GPU: NVIDIA、AMD
- 内存: 一般内存与CPU核数比为2GB:1, 是否有大内存需求
- 存储及IO能力: 容量、并发IO读写能力
 - 硬盘: SATA or SAS
 - 文件系统: 普通NFS文件系统 or 并行文件系统
- 网络:
 - 通信要求低: 千兆以太网 (低成本)
 - 通信要求高: InfiniBand (高带宽: 56Gbps; 低延迟: $< 1\mu s$)
- 操作系统: 极少Windows, 一般为Linux (RHEL、CentOS、SuSE)
- C/C++、Fortran编译环境: GCC、Intel、PGI等编译器
- 数值函数库: BLAS、LAPACK、ScaLAPACK、Intel MKL、AMD ACML等
- MPI: MPICH、Open MPI、Intel MPI、LAM MPI



一般是通过招标采购

一般是通过招标采购

- 确定招标参数

- 多与主流厂家交流，综合几家的建议参数并修改
- 关键性参数一定写明确
- 非关键性参数不要写太多
- 利用某些参数屏蔽掉某些不符合要求的产品
- 避免被某家厂商控制，尽量让主流厂家都能参与
- 小心恶意低价中标，然后延期供货
- 明确违约责任

一般是通过招标采购

● 确定招标参数

- 多与主流厂家交流，综合几家的建议参数并修改
- 关键性参数一定写明确
- 非关键性参数不要写太多
- 利用某些参数屏蔽掉某些不符合要求的产品
- 避免被某家厂商控制，尽量让主流厂家都能参与
- 小心恶意低价中标，然后延期供货
- 明确违约责任

● 评标

- 不建议最低价中标；建议价格分占一定比例，技术分占一定比例
- 超算系统是一个整体，不仅仅是硬件的简单堆叠
- 一些评委对服务器等通用硬件很了解，但对超算等一些特有硬件及系统等不了解



- 性能测试：
 - CPU、内存、IO、网络等标准测试程序测试
 - 用户应用实际运行
- 压力测试：满负荷长时间运行，查看稳定性及持续性能



- 中国科大超级计算中心:
 - 电话: 0551-63602248
 - 信箱: sccadmin@ustc.edu.cn
 - 主页: <http://scc.ustc.edu.cn>
 - 办公室: 中国科大东区新图书馆一楼东侧126室
- 李会民:
 - 电话: 0551-63600316
 - 信箱: hmli@ustc.edu.cn
 - 主页: <http://hmli.ustc.edu.cn>