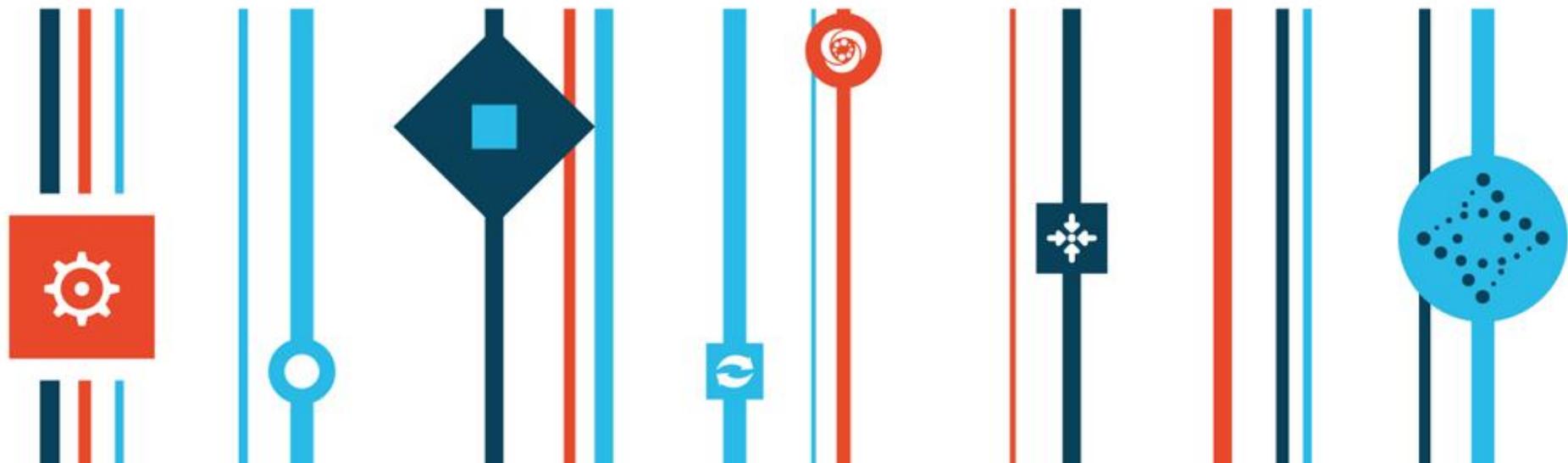


云计算在高性能计算中心的最佳实践

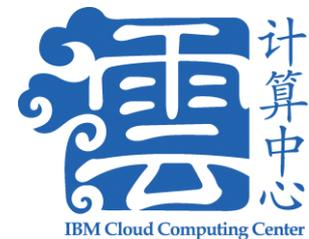
曹凡 IBM云计算中心架构师

梁毅 北京工业大学计算机学院副教授



IBM全球云计算中心

- IBM把云计算视为一项重要战略，成立专门的云计算部门，并已在全球建立十一个**云计算中心**，为客户提供云计算技术支持和服务，提供全面的**端到端云计算解决方案**。
 - 现场设计实施云计算中心的基础架构
 - 提供云计算的高技能的人力资源支持
 - 提供下一代数据中心服务的培训
 - 快速部署和实施云计算的概念验证及试运行



IBM大中华区云计算中心

计世资讯的调查显示：**IBM**是国内客户第一想到的云计算解决方案提供商

- 以**IBM**中国研究院和软件开发中心的数千研发人员为后盾
- 以推动云计算在国内的发展为目标
- 结合国内客户的实际需求和业务特点
- 为客户提供端到端解决方案，充分整合**IBM**各产品线的优势



中国中化集团公司
SINOCHEM CORPORATION



北京工业大学
BEIJING UNIVERSITY OF TECHNOLOGY



中国·东营
www.dongying.gov.cn



中国移动通信
CHINA MOBILE

2010年

- 进一步深化云计算与行业应用的紧密结合
- 积极探索云计算在商业智能和物联网方面的应用前景

2009年

- 推出面向行业的云计算“6+1”解决方案
- 实现云计算在国内各行业的全面开花
- 加入由电子学会主办的云计算专家委员会

2008年成立

- 在中国建立第一个云计算中心——无锡

目录

- 高性能计算中心发展
- 云架构的高性能计算中心
- 北工大高性能计算中心的成功应用

高性能计算应用不同/需求不同

侧重处理能力的高性能计算机 (CPU Intensive)



IBM BladeCenter HS22

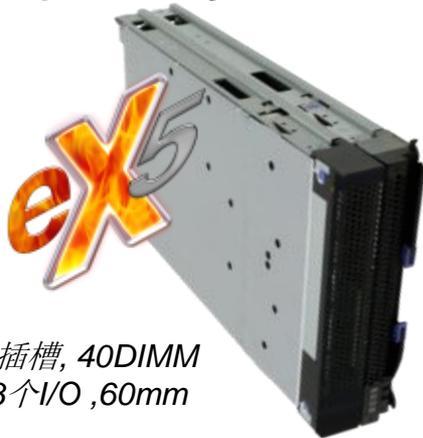
15分钟搜索10亿多文件

IO吞吐量: **134+GB**每秒

实测最大的文件系统大小为~**2PB**

川庆物探: **1PB, 4GB**每秒

侧重内存的高性能计算机 (Memory Intensive)



2插槽, 40DIMM
8个I/O, 60mm

IBM BladeCenter HX5

看到:

32cores 2.53 128GB内存, 内存不足

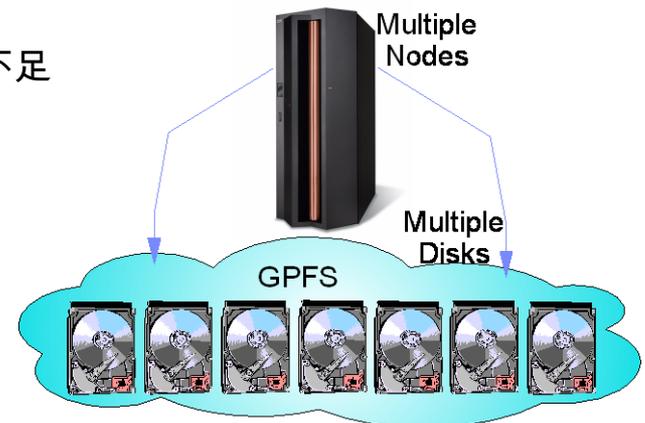
SWAP频繁导致宕机

局部仿真20-30小时完成

可MPI扩展

侧重内存的高性能计算

侧重I/O的高性能计算机 (I/O Intensive)



IBM General Parallel File System (GPFS)



高性能计算中心的挑战

智揽云海 云领未来
2010 IBM 云计算高峰论坛

- 应用模式复杂
 - 商用、开源、自开发
 - 多种平台需求：Linux、Windows、UNIX
 - 串行、并行
- 管理复杂
 - 作业管理
 - 应用平台管理
 - 数据管理
 - License管理
 - 用户管理
- 资源共享的需求
 - 开发、生产
 - 网格平台的需求
 - 跨WAN的资源共享
- 能源消耗巨大

什么是云计算

- 云计算是一种通过计算管理分配的方式**共享资源**的计算，计算资源可以动态部署、动态调优、动态收回。
- 在云计算基础设施中，各种计算资源被连接在一起形成统一的**资源池**，这些资源会被动态的分配给不同的应用和服务，满足它们在不同时刻的需求。



云的优点:

云 = 更少的投资
云 = 动态的规模
云 = 灵活和高效率

高性能计算云方案架构





高性能计算云服务模式

智揽云海 云领未来
2010 IBM 云计算高峰论坛



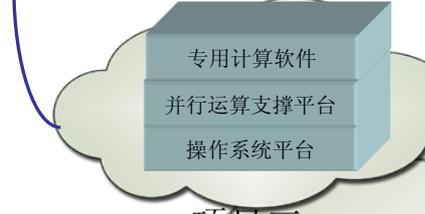
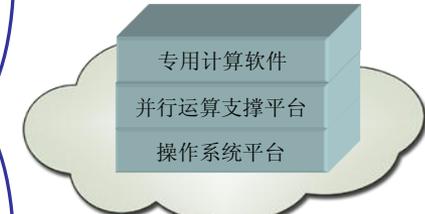
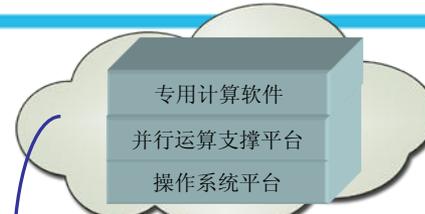
各研究人员



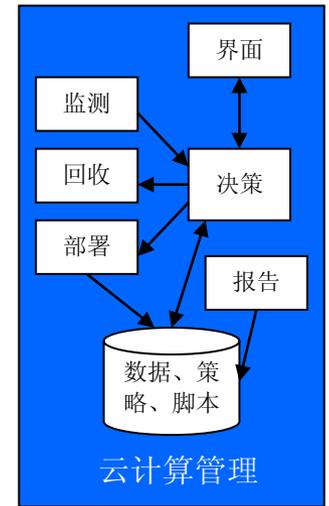
公众使用人群



云计算服务Portal



动态产生



北京

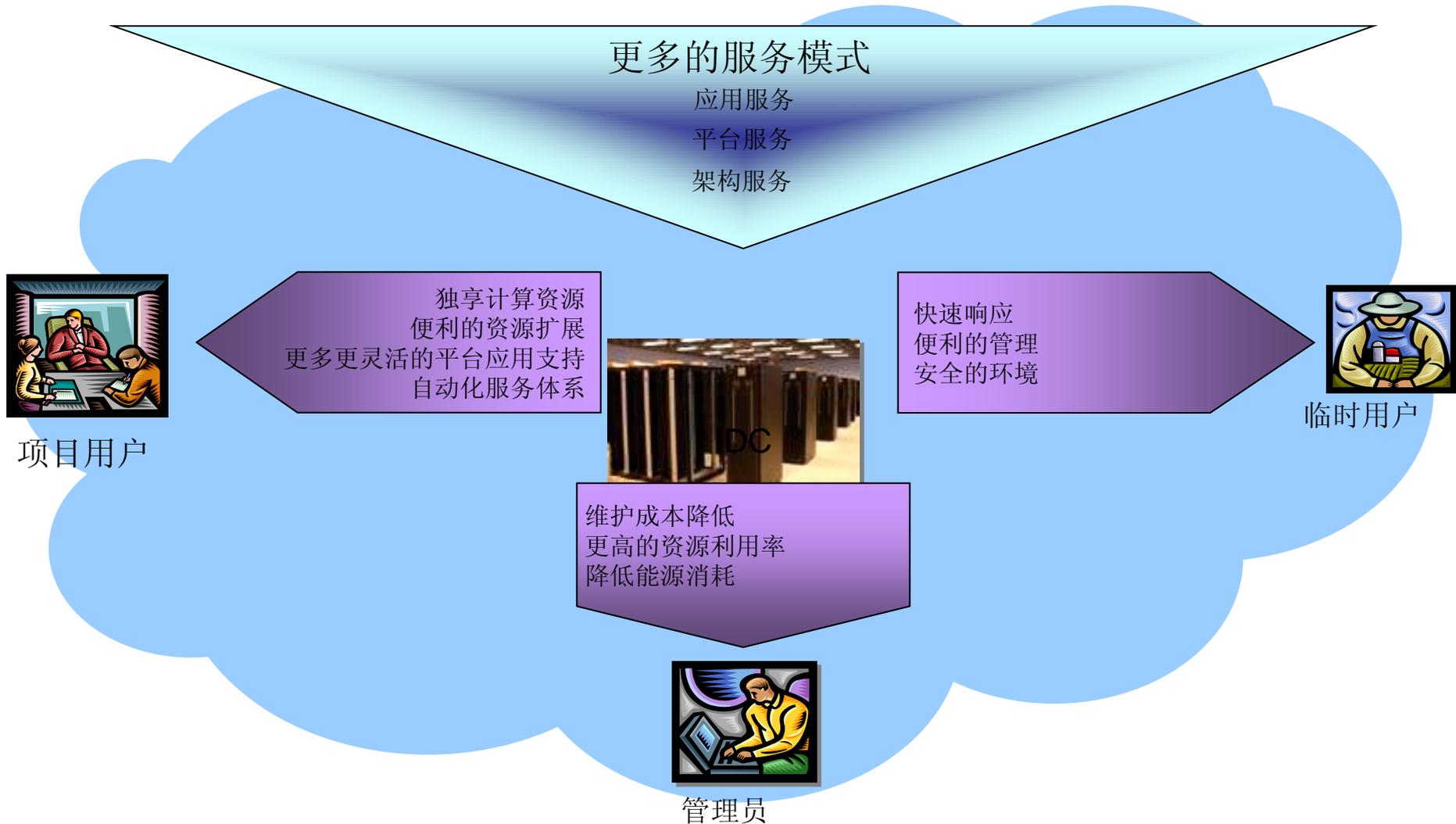


上海

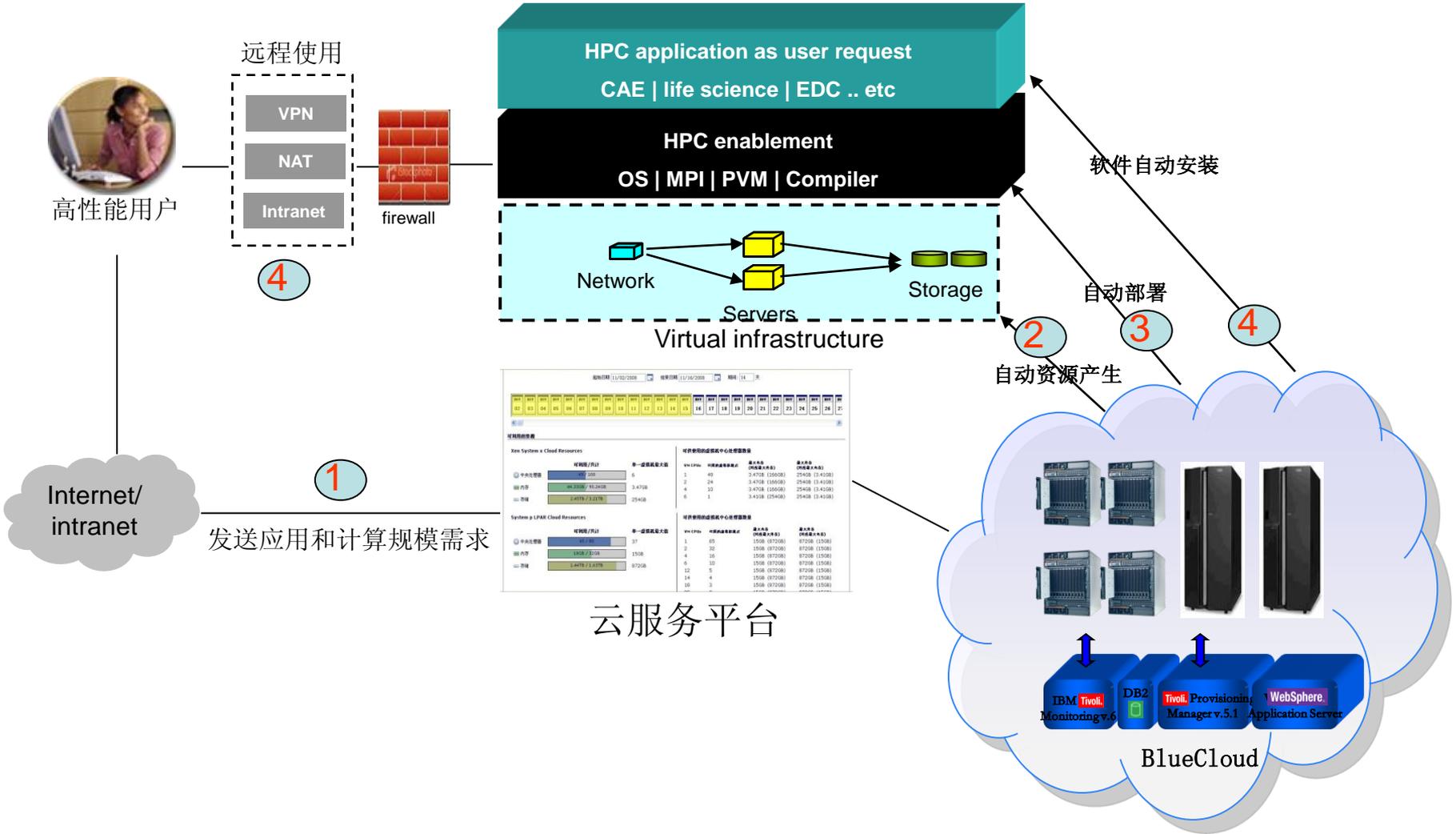


其它地区

所有计算资源



服务流程



- 适合对象
 - 高性能计算中心
 - 多用户、多学科研究平台
 - 研究院集中数据中心
- 服务模式
 - 公众服务平台
 - 基于研究课题动态分配
 - 基于研究团队资源分配
- 优点
 - 为研究者提供独立环境
 - 快速响应各种研究需求
 - 支撑多种研发平台

IBM云计算技术在北京工业大学 高性能计算平台的应用

报告人：梁毅

2010-06-11

2009-12-25

北京工业大学2008级硕士开题答辩



提纲

1

北京工业大学高性能计算平台简介

2

IBM云成功案例——
北京工业大学并行计算大赛

- 建设目标

- ☆ 服务教学科研

- ☆ 开展科学研究

- ☆ 支撑服务北京

- 高性能计算平台建设规模及技术路线

- ☆ 遵循分区规划、统一管理的建设思路

- ☆ 采用基于云计算技术的高性能计算中心新一代解决方案

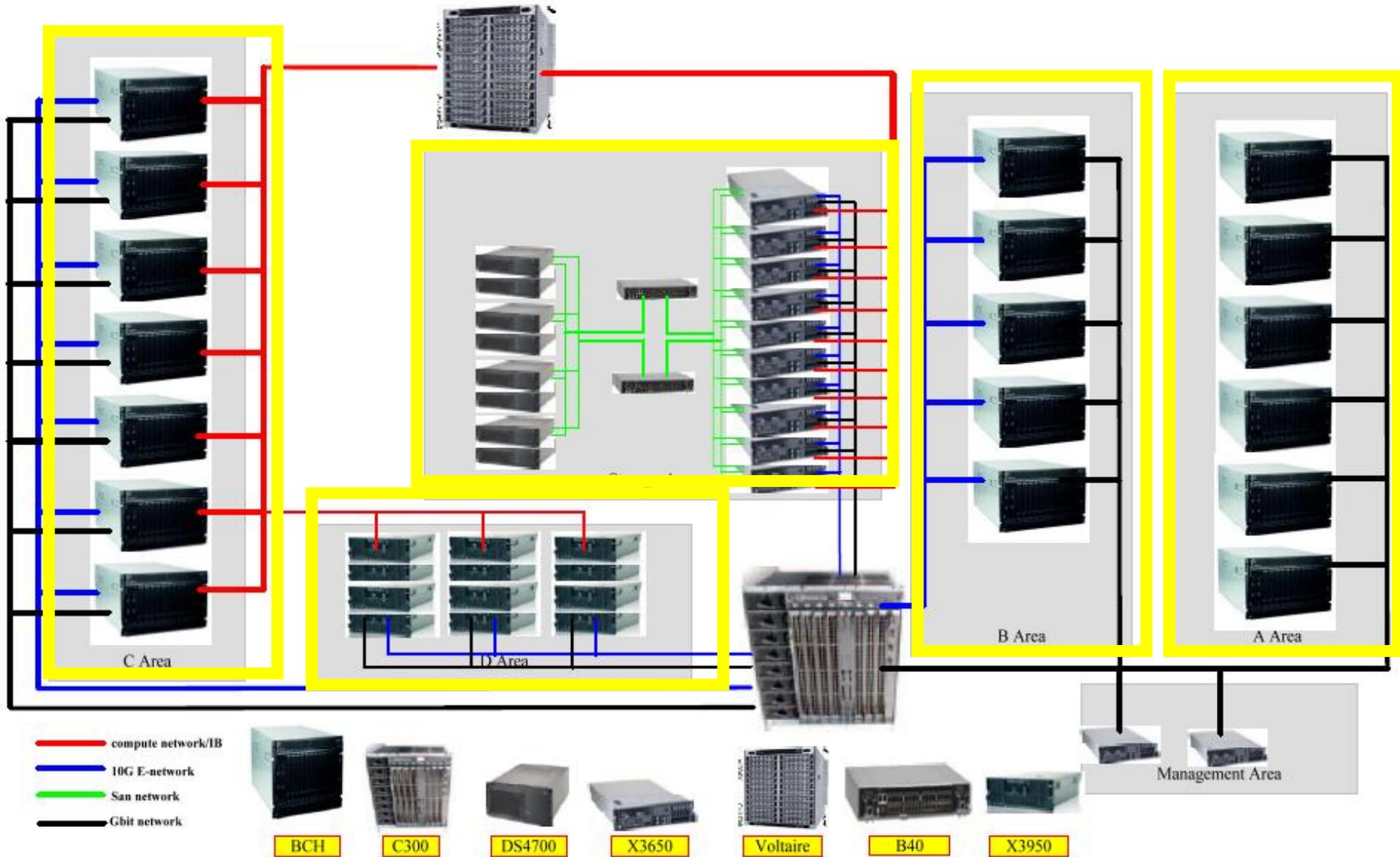
- ☆ 总计算能力23TF1ops, 总存储能力40TB, 位居全国高校第三。

北京工业大学高性能计算平台

硬件基础设施概况

计算资源	252台IBM HS21刀片服务器
	3台IBM X3950M2大内存机架服务器
存储资源	4台IBM TotalStorage DS4700-70A磁盘阵列
	基于SAN架构的存储网络系统
网络资源	一套 Voltaire 20Gb/s Infiniband高性能网络
	一套Force10 万兆以太网
	一套Force10 千兆以太网

北京工业大学高性能计算平台



北京工业大学高性能计算平台



提纲

1

北京工业大学高性能计算平台简介

2

IBM云成功案例---
北京工业大学并行计算大赛

北京工业大学并行计算大赛

- 共10个学院, 29个参赛小组
- 自选课题, 86%来源于实际项目需求
- 应用类型多样, 并行环境需求异构
- 所需的机群节点规模各异

北京工业大学并行计算大赛

• 并行环境需求

学院	应用	软件环境
电 控	基于Hadoop的图像检索与地理信息查询系统的并行算法设计与实现	Linux + MPI + NetBean + JDK + Hadoop + Hbase + Zookeeper + tomcat + geoserver + postgresql
生 命	生物分子相互作用复杂网络的并行模块划分方法	Linux + MPI + Boost_1_34_1
激 光	飞秒激光与等离子的相互作用机制的数值模拟	Linux + MPI + Pvm
建 工	建筑工程有限元分析软件OpenSees的并行优化	Linux + MPI + OpenMPI + OpenMP + Opensees
机 电	多孔介质化-力耦合问题的并行算法研究	Linux + MPI + Fortran90
机 电	喷涂中液滴形成和撞击的并行计算模拟	Linux + MPI
机 电	变速箱减振降噪优化设计程序的并行化	Windows XP + MPI
数 理	并行计算在识别飞行物着陆点的中应用	Windows xp + MPI + Vc++6.0
材 料	大规模锂电池生产管理中优化组合解决方案	Linux + MPI + Gcc
计算机	基于大规模数据库的人脸识别研究	Linux + MPI + Opencv

北京工业大学并行计算大赛

- 基于高性能平台A区，采用IBM云计算技术，提供并行应用调试/运行环境。

	虚拟机个数	单个虚拟机硬件配置			单个虚拟机软件配置				
		CPU	内存	硬盘	操作系统	集群配置	并行环境	监控	作业调度
头节点	1	1*2.83 GHz	2GB	30 GB	Windows xp/ Windows 2003/ RHEL 5.4-32bit/ RHEL 5.4-64bit/ RHEL 5.2-32bit	ssh/nfs	Mpich 2/ OpenMP /Hadoop	ITM Monitoring Agent	Torque-- server
计算节点	7	1*2.83 GHz	1GB	10 GB	Windows xp/ Windows 2003/ RHEL 5.4-32bit/ RHEL 5.4-64bit/ RHEL 5.2-32bit	ssh/nfs	Mpich2 / OpenMP /Hadoop	ITM Monitoring Agent	Torque-- client

北京工业大学并行计算大赛

• 蓝云使用展示---环境部署、资源供应

新项目

1. 浏览可使用的基本系统并选择日期

2. 选择服务器并且配置软件

3. 递交请求

选定的日期

起始日期 06/10/2010 结束日期 06/24/2010 期间 14 days

完成输入新项目

项目名称:

Group07_linuxHPC

描述:

你的虚拟机

 Xen RedHat Linux 5.4 x86_32 IBM Tivoli Monitoring Agent mpich2-1.2.1 - Linux torque-2.3.7-server - Linux	1 CPU unit (1 vcpu) 2GB Memory 30GB Disk (incl. 2GB swap)
 x 7 Xen RedHat Linux 5.4 x86_32 IBM Tivoli Monitoring Agent mpich2-1.2.1 - Linux torque-2.3.7-client - Linux	1 CPU unit (1 vcpu) 1024MB Memory 10GB Disk (incl. 2GB swap)

选择项目类型: linuxHPC

Configure linuxHPC Project

Linux High Performance Computing Cluster

OS version	RedHat5.4_32
Head Node Physical CPU size / vm	1
Head Node VCPU number / vm	1
Head Node Memory size / vm	2048 M
Head Node Disk size / vm	30 G
Slaver Node Physical CPU size / vm	1
Slaver Node VCPU number / vm	1
Slaver Node Memory size / vm	1024 M
Slaver Node Disk size / vm	10 G
MPI version	mpich2-1.2.1
Job scheduler	torque-2.3.7
Number of computing nodes	7
Enable monitoring	<input checked="" type="checkbox"/>

北京工业大学并行计算大赛

• IBM云使用展示---环境部署、资源供应

The screenshot displays the IBM Cloud Computing Center interface. At the top, the browser address bar shows the URL `http://10.3.250.1:9080/cloud/#1275910955722`. The page title is "IBM Cloud Computing Center".

项目详细资料

项目名: Group07-linuxPC
客户: group07
项目状态: 正在部署中... **0**
起始日期: 2010-6-7

用户: group07-admin (10-6-7 下午7:44)
请求的服务器计数: 8
结束日期: 2010-6-21

项目类型: linuxHPC
运行中的服务器计数: 0
期间: 14天

项目基础设施

名字	硬件设置	系统基础形象	状态
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...
Xen System x Local Disk VM	1.0CPU (1 vcpu) - 2048MB 内存 - 20GB 硬盘 (包括 2048MB 交换)	Xen RedHat Linux 5.4 x86_32	正在部署...

增加/删除服务器 变更项目日期 终止项目 删除项目 显示报告 刷新 上一页

北京工业大学并行计算大赛

• IBM云使用展示--- 并行环境使用

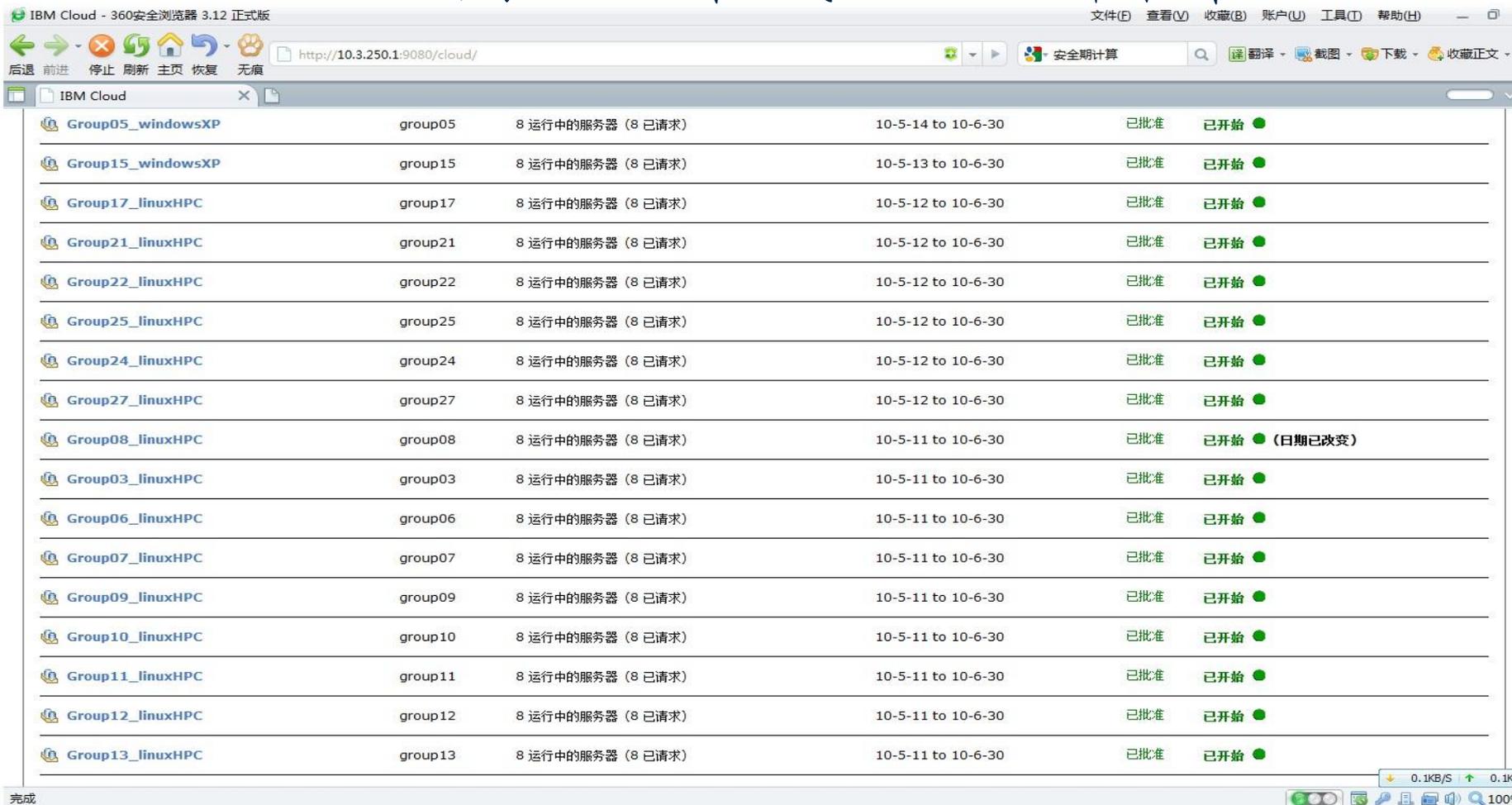
The screenshot displays the IBM Cloud management console. On the left, a sidebar shows project details for 'Group07_linuxHPC-5.4'. The main area is divided into several sections:

- 项目详细资料:** Group07_linuxHPC-5.4, 客户 group07, 项目状态 已开始, 起始日期 2010-5-30.
- 项目基础设施:** A table listing virtual machines (v0A-3-251-47 to v0A-3-251-45) with their hardware specifications (1.0 CPU, 1536MB memory, 15GB disk).
- 系统信息:** A table with columns for IP, OS type, Pool/Type, and Admin Password. The entry for v0A-3-251-45 is highlighted with a red box.
- 硬件设置:** A table listing hardware specifications for each VM.
- 另外的软件:** Lists installed software like mpich2-1.2.1 and IBM Tivoli Monitoring Agent.
- 实时监控:** Shows real-time resource usage: CPU 2%, Memory 853 MB free, Storage 3.02 GB free.

In the background, a terminal window shows a root login session on v0A-3-251-45, and a desktop environment window shows a Linux desktop with icons for Computer, root's Home, and Trash.

北京工业大学并行计算大赛

• IBM云使用展示—在线项目统计和管理



The screenshot displays the IBM Cloud project management interface. The browser window title is "IBM Cloud - 360安全浏览器 3.12 正式版". The address bar shows "http://10.3.250.1:9080/cloud/". The page content is a table listing project groups with columns for group name, ID, server status, dates, approval status, and start status.

Group Name	Group ID	Server Status	Dates	Approval	Start Status
Group05_windowsXP	group05	8 运行中的服务器 (8 已请求)	10-5-14 to 10-6-30	已批准	已开始 ●
Group15_windowsXP	group15	8 运行中的服务器 (8 已请求)	10-5-13 to 10-6-30	已批准	已开始 ●
Group17_linuxHPC	group17	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group21_linuxHPC	group21	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group22_linuxHPC	group22	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group25_linuxHPC	group25	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group24_linuxHPC	group24	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group27_linuxHPC	group27	8 运行中的服务器 (8 已请求)	10-5-12 to 10-6-30	已批准	已开始 ●
Group08_linuxHPC	group08	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ● (日期已改变)
Group03_linuxHPC	group03	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group06_linuxHPC	group06	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group07_linuxHPC	group07	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group09_linuxHPC	group09	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group10_linuxHPC	group10	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group11_linuxHPC	group11	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group12_linuxHPC	group12	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●
Group13_linuxHPC	group13	8 运行中的服务器 (8 已请求)	10-5-11 to 10-6-30	已批准	已开始 ●

北京工业大学并行计算大赛

• IBM云使用展示—基于项目组的资源配置明细

The screenshot displays the IBM Cloud Computing Center interface. The browser address bar shows the URL: <http://10.3.250.1:9080/cloud/#1275828093712>. The page title is "IBM Cloud Computing Center". The user is logged in as "ca_test".

项目详细资料

项目名称: **Group07_linuxHPC-5.4**

客户: **group07** 用户: **ca_test** (10-5-30 上午12:22) 项目类型: **linuxHPC**

项目状态: **已开始** 请求的服务器计数: **8** 运行中的服务器计数: **8**

起始日期: **2010-5-30** 结束日期: **2010-6-30** 期间: **31天**

[项目](#) 已批准

项目基础设施

名字	硬件设置	系统基础形象	状态
<input type="checkbox"/> v0A-3-251-47(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-42(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-44(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-46(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-49(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-43(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-41(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份
<input type="checkbox"/> v0A-3-251-45(head)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的 备份

增加/删除服务器 改变项目日期 终止项目 删除项目 显示报告 刷新 上一页

北京工业大学并行计算大赛

• IBM云使用展示—资源状态实时监控

The screenshot displays the IBM Cloud console interface for monitoring five virtual machines (VMs). Each VM card provides a comprehensive overview of its configuration and current status.

名字	硬件设置	系统基础形象	状态
v0A-3-251-47(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的
系统信息 IP: 10.3.251.47 OS 类型: Xen RedHat Linux 5.4 x86_32 Pool / 类型: Xen System x Local Disk (xen) 管理员密码: 10gV08S0 管理名称/IP: v0A-3-251-47 / 10.3.251.47			
另外的软件 mpich2-1.2.1 - Linux IBM Tivoli Monitoring Agent			
实时监控 CPU 使用: 36 % 内存 自由: 283 MB 存储 自由: 6.40 GB			
远程控制 开机 关机 重启 重设密码			
获取物理主机信息 for v0A-3-251-47(slaver)...			
v0A-3-251-42(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的
系统信息 IP: 10.3.251.42 OS 类型: Xen RedHat Linux 5.4 x86_32 Pool / 类型: Xen System x Local Disk (xen) 管理员密码: Z9x6!wCq 管理名称/IP: v0A-3-251-42 / 10.3.251.42			
另外的软件 mpich2-1.2.1 - Linux IBM Tivoli Monitoring Agent			
实时监控 CPU 使用: 43 % 内存 自由: 436 MB 存储 自由: 6.69 GB			
远程控制 开机 关机 重启 重设密码			
获取物理主机信息 for v0A-3-251-42(slaver)...			
v0A-3-251-44(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的
系统信息 IP: 10.3.251.44 OS 类型: Xen RedHat Linux 5.4 x86_32 Pool / 类型: Xen System x Local Disk (xen) 管理员密码: 8ypvG3PL 管理名称/IP: v0A-3-251-44 / 10.3.251.44			
另外的软件 mpich2-1.2.1 - Linux IBM Tivoli Monitoring Agent			
实时监控 CPU 使用: 31 % 内存 自由: 331 MB 存储 自由: 6.07 GB			
远程控制 开机 关机 重启 重设密码			
获取物理主机信息 for v0A-3-251-44(slaver)...			
v0A-3-251-46(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的
v0A-3-251-49(slaver)	1.0CPU (1 vcpu) - 1536MB 内存 - 15GB 硬盘 (包括 2048MB 交换) [修改]	Xen RedHat Linux 5.4 x86_32	可用的

北京工业大学并行计算大赛

- 总结

- 共计部署于**84**个刀片服务器
- 虚拟机群规模**8-100**个节点
- 在线同时管理虚拟机群数最大为**32**个
- 虚拟机群部署时间约为**30**分钟
- 通过**IBM**云提供的细粒度资源供给功能，仅使用了**50-65%**的硬件资源，满足所有**29**个参赛小组的高性能资源需求。

THANK
YOU!

2009-12-25
2010-06-11

北京工业大学2008级硕士开题答辩

