Computer Science and Application 计算机科学与应用, 2013, 3, 396-400
http://dx.doi.org/10.12677/csa.2013.39068 Published Online December 2013 (http://www.hanspub.org/journal/csa.html)

HPC-Cloud Advantages in the Field of Scientific Research

Xue Huang, Liutao Zhao, Yi Jin

Department of Beijing Computing Center, Beijing
Email: huangxue@bcc.ac.cn, zhaolt@bcc.ac.cn, jinyi@bcc.ac.cn

Received: Nov. 29th, 2013; revised: Dec. 18th, 2013; accepted: Dec. 25th, 2013

Abstract: High-performance computing is an important service aid and plays an irreplaceable role in the field of scientific research in China. In this dissertation, it proposes a preliminary solution——high-performance computing cloud platform, for the bottleneck problem in the field of high-performance computing when providing computing service, and summarizes the present situation of high-performance computing cloud platform at home and abroad. Main functions of HPC-Cloud system are designed according to the characteristics that apply in the field of scientific research. It also shows the advantages of high-performance computing cloud platform according to a typical case in solving problems of intensive computing applications.

Keywords: HPC; Cluster; HPC-Cloud

HPC-Cloud 在科学研究领域的优势

黄　雪，赵琉涛，金　翊

北京市计算中心，北京
Email: huangxue@bcc.ac.cn, zhaolt@bcc.ac.cn, jinyi@bcc.ac.cn

收稿日期：2013 年 11 月 29 日；修回日期：2013 年 12 月 18 日；录用日期：2013 年 12 月 25 日

摘　要：高性能计算是我国科研领域重要的服务工具，起着不可替代的作用。本文就高性能计算为科学研究领域提供计算服务时遇到的瓶颈问题，提出了初步的解决方案——高性能计算云平台，总结了高性能计算云平台的国内外现状。针对科学研究领域应用的特点设计了 HPC-Cloud 系统的主要功能，并以典型案例表明高性能计算云平台在解决计算密集型应用问题上的优势。

关键词：高性能计算；集群；高性能计算平台

1. 引言

高性能计算(HPC)一直以来都是人们比较关注的领域，其作为科学研究的重要手段被广泛应用于分子物理、分子生物学、高能物理、石油勘探等众多领域。在经历了几十年的发展，HPC 作为实验和理论以外的第三大科学研究手段，甚至被当作一个国家综合国力的主要评价标准。

HPC China 2013 在广西桂林于 10 月 31 号落下帷幕，期间众多学者对于高性能计算相关内容的演讲，使我们看到了在未来计算机发展的方向，和所面临的问题[1]。随着计算效率、计算数据量级的不断突破，高性能集群的管理面临前所未有的挑战：

体系架构——HPC 的体系架构在加速演变中，从传统的 X86 架构到如今的 GPU、MIC 架构[2]，都在不断提升着 HPC 的计算性能。但这也同时加大了对系

统资源管理、作业调用等的难度。

资源管理——高性能集群中的资源是固定的，当多用户并发使用时，只能通过系统管理员手动调整资源分配，这使得系统利用率与用户服务期望值不能达到一致。即增大了管理难度，又难以保证每个用户的服务质量。

并行应用——高性能计算的应用研发水平，目前远远落后于集群的计算水平，有些应用甚至还不能实现并行计算。不同的应用对集群的体系架构、编程模型、优化方法等都需要不同的处理方法。

以上问题反应出，有效的管理在高性能集群上的工作负载，提高用户生产效率，充分利用高性能计算的优势才是解决问题的关键。云计算的资源共享、弹性扩展等特点正好可以解决高性能计算遇见的瓶颈问题。云计算的优势主要体现在以下几点：

- 公共云 IaaS[3]平台使用虚拟化技术，在一台物理机供多个用户使用，实现资源有效隔离，避免用户之间发生资源冲突或抢占问题。
- 云计算的动态扩展功能可实现水平扩展和垂直扩展，小到用户的资源调整，大到集群规模的变化都可以轻松应对。这样的功能正好可以满足不同用户在不同时期对资源的不同要求。
- 一般的高性能集群都是部署在单位内部，仅供内网使用。但是云计算则是有效利用互联网技术，使用户可以不受办公地点的限制，随时随地登录使用资源。

HPC 云是云计算在 IaaS 层的体现，用户在云端即可完成软件应用、硬件资源租用等服务。高性能计算云平台提供基础资源服务的同时，更注重为用户带来平台化的标准服务，为用户定制规范的业务流程，使用户在统一的平台化服务上得到高性能计算的高效率计算服务。所以，HPC-Cloud[4]关注的重点是 PaaS 层服务，这样，不仅可以解决高性能集群资源管理方面的问题，还能在业务流程、使用规范、资源调整等方面带来更优质的解决方案。

2. HPC-Cloud 国内外现状

云计算被分为三个层次：基础设施即服务(IaaS)、平台即服务(PaaS)和软件即服务(SaaS)。我国的 HPC 主要是由地方政府主导的超算中心进行研究的，对工业化与信息化的融合承载着重要的作用。我国现有国家级超算中心 4 家：天津中心、深圳中心、济南中心和长沙中心；另外还有近 7 家的地方性超算中心。各超算中心都有自己的 HPC 业务，IaaS 层云平台目前只有深圳超算中心拥有。各超算中心都意识到 HPC 云的重要性，纷纷都在加紧研究。现阶段，IaaS 层的基础是最好的，PaaS 其次，最后 SaaS 还没有真正开始。从超级计算中心模式到 HPC 云模式，最大的变化是要能够提供一个更简单、方便、易用的计算环境，这一目标主要由 PaaS 层实现。这一层扮演主要角色的是作业调度系统，它能有效的对资源和用户作业进行监控、调度和管理。这是 PaaS 层需要解决的。

国外的超算中心与国内的超算中心有很大区别，他们大多是有研究机构组建的，客户群比较单一，基本上只对本机构的人开放。目前高性能计算云做的比较不错的有亚马逊的 AWS HPC[5]、荷兰的 SARA、美国的 MIT、欧洲空间局(ESA)等。AWS HPC 优势在于根据任务需求进行资源的弹性配置，并使工作负载加速执行；SARA 则是针对那些不能在超级计算中心运行但却可以在本地集群、工作站或者 PC 运行的 HPC 应用，例如那些需要专用或者定制函数库的应用软件，基于 OpenNebula；MIT 用来在 Amazon 的 EC2 上配置和部署高性能计算环境，包括了 Ubuntu、NFS、OpenMPI、SSH 选项、作业调度系统(SGE)和高性能数学库 ATLAS 等；欧洲空间局主要是观测并分析获得银河系的三维地图，其间需要大量的数据处理计算。ESA 目前正和合作厂商一起，研究利用 Amazon 的 EC2/S3 进行相关计算。

3. HPC-Cloud 功能介绍

我国 HPC 业务主要服务于科学研究，用户群多为科研院所、高校实验室等，计算任务分为工程计算、生物医药、仿真模拟等几个区域。在高性能计算机群环境下，用户提交并行计算作业在机群内部的机器上运行，用户作业的行为是未知的，并且是不可控制的[6]，分配给每一个用户的资源也是动态变化的，因此用户管理就需要更加灵活。针对用户群体的特征，以及 HPC 集群发展中遇到的问题，设计了 HPC 云平台的主要业务功能，用以解决上述问题。

HPC-Cloud 功能主要包含作业管理、作业调度、

账户管理以及资源监控、自动化配置和高可用性等六个方面。

1) 作业管理

在网站上用户可以选择不同的模板，模板是根据任务的类型、规模分别制作的。用户直接点击下拉菜单选择队列、计算的核心数，甚至还可以有选择性的选择计算节点。任务输出路径、内存、任务开始/终止时间等信息会清晰的显示在页面上。保存提交作业后，在作业列表中会显示由用户提交的所有作业信息，包括作业的状态、作业号、提交的脚本名、作业的所有者、运行时间等。这些信息有助于帮助用户更加详细的了解作业，判断作业是否正常。

除了上述提到的作业管理功能，还有检索功能。当用户提交的作业数量逐渐增多时，查找某一特定作业的难度也是在逐渐增大。通过关键字检索，帮助用户快速定位目标作业。

2) 作业调度

作业调度功能可以帮助用户查看集群中可用的资源，以及作业实时的计算状态。用户根据对提交的作业的初始认知，结合网站上反馈的信息，即可初步判断任务是否正常。如果得出任务运行异常的结论，可选择结束任务，查看输出文件，从而确定任务异常的原因。若不能自行解决，可选择在线邮件通知方式告知后台，也可以选择在线窗口与技术人员进行沟通。这样避免了计算资源的浪费，为按需付费用户提供了更合理的资源管理依据。

3) 账户管理

用户的账户由系统管理员统一管理，普通账户在登录后拥有对自己账户下文件的读、写、共享、删除等权限，可修改账户密码，但不能修改用户名。对于系统文件，一般情况下是拥有读的权限。账户中的使用者信息，包括使用者姓名、单位、联系方式、创建时间等都是可以被查询、修改的。

4) 资源监控

集成监控软件系统，对 IT 资源进行统一监控。可集成 NagI/Os、catti、ganglia 等监控管理软件供用户选择，同时也要支持通过定制开发插件方式集成用户现有的管理与监控系统。

5) 自动化配置

根据用户资源申请，动态部署账户所需的基础环境、软件应用等，并提供不同作业模板；系统资源自动管理功能可以自动化的对集群设备进行增加和配置，通过自动化的流程将设备添加到资源池中，便于管理与配置。

6) 高可用性

平台提供用户应用的高可用性，支持 HA、N + 1 和 N + M 的方式，提供一对一、多对一和多对多的备份方式，以保证用户应用的高可用性。同时，为实现高可用性，平台支持实体机到虚拟机、虚拟机到虚拟机的迁移等功能。

4. HPC-Cloud 科学工程领域典型案例

科学工程计算是 HPC-Cloud 的典型应用领域。科学工程领域包含高能物理、分子化学、生命科学等领域。下面是北京某研究所在做分子与材料模拟研究时提交的一个脚本作业，应用软件为分子力学模拟计算常用软件 amber，amber 属于计算密集型应用。当用户在高性能集群中提交 amber 的循环作业，脚本如下：

```
APP_NAME=scinormal
NP_PER_NODE=12
MY_MPI_TYPE=openmpi
MY_MPI_HOME=/share/software/openmpi/icc-13.0.1/openmpi-1.6.3-icc
#任务调用的核心数，为 NP_PER_NODE 的整数倍将获得更好的计算效率，该参数根据任务计算量大小常调整
NP=48
RUN=""
#vasp：RUN="vasp"
#lammps:RUN="lmp_openmpi < relax.in > relax.log"
#gaussian09:RUN="g09 in.gjf"
#GMX:RUN="mdrun input.file"
#fortan:RUN="run.file"
#minimize sander run scripes
mpirun -np $NP /share/home/users/amber12/bin/sander.MPI -O -i min1.in -p cmp.prmtop -c cmp.inpcrd -ref cmp.inpcrd -r cmps_min1.rst -o cmps_min1.out
mpirun -np $NP /share/home/users/amber12/bin/sander.MPI -O -i min2.in -p cmp.prmtop -c cmps_
```

min1.rst -ref cmps_min1.rst -r cmps_min2.rst -o cmps_min2.out

mpirun -np $NP /share/home/users/amber12/bin/sander.MPI -O -i min3.in -p cmp.prmtop -c cmps_min2.rst -ref cmps_min2.rst -r cmps_heat0.rst -o cmps_min3.out

##seven heat sander run scripes

num=1

mu=0

enu=1

emu=0

while (($num < 8))

do

mpirun -np $NP /share/home/users/amber12/bin/sander.MPI -O -i heat$num.in -p cmp.prmtop -c cmps_heat$mu.rst -r cmps_heat$num.rst -o cmps_heat$num.out -x cmps_heat$num.mdcrd

gzip -9 cmps_heat$num.mdcrd

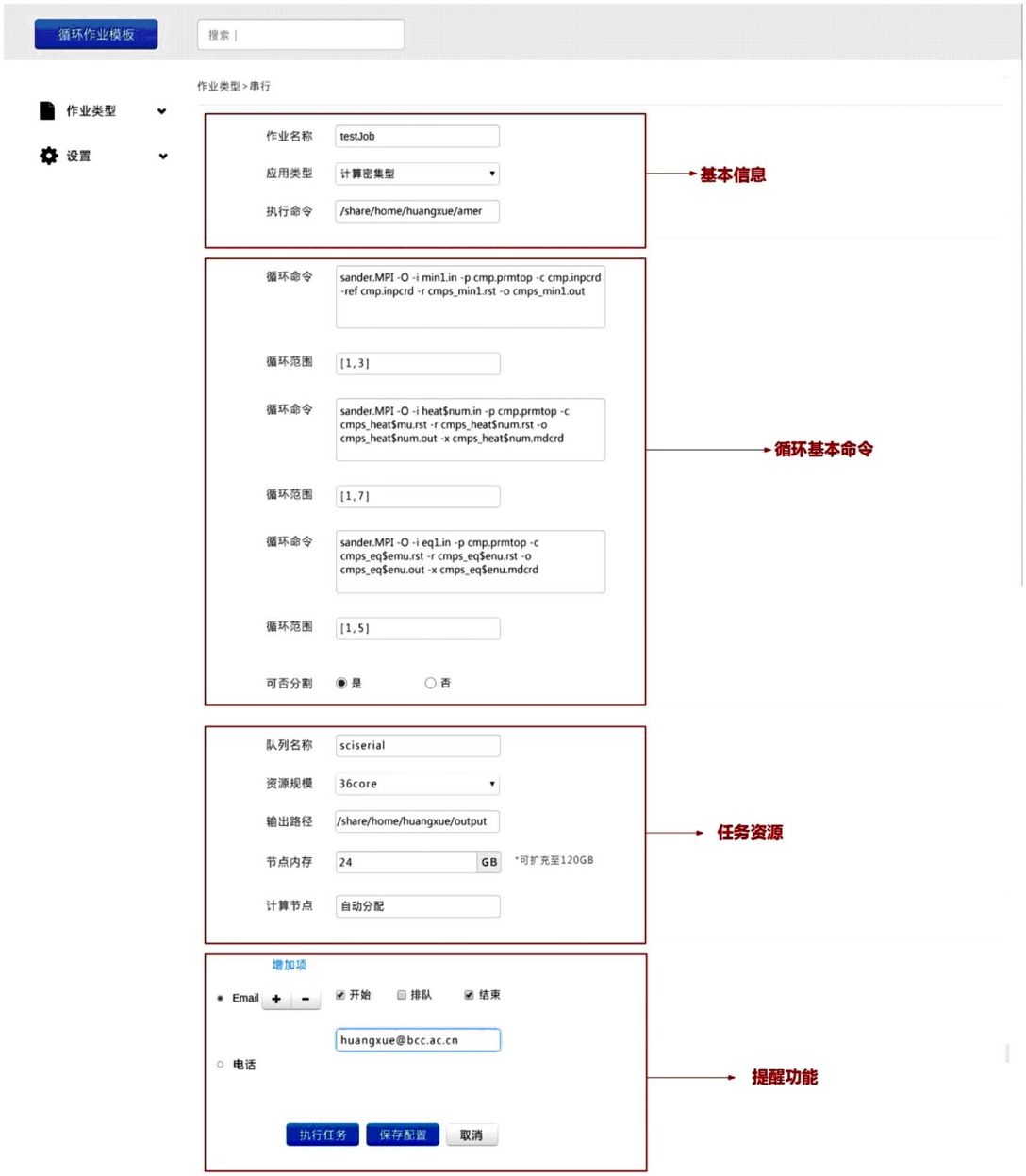

Figure 1. Template of cycle operation
图 1. 循环作业模板脚本

```
echo "heat$num.in done"
let num+=1
let mu+=1
done
mv cmps_heat7.rst cmps_eq0.rst
#
##MD sander run scripes
#
while (( $enu < 6 ))
do
mpirun -np $NP /share/home/users/amber12/bin/sander.MPI -O -i eq1.in -p cmp.prmtop -c cmps_eq$emu.rst -r cmps_eq$enu.rst -o cmps_eq$enu.out -x cmps_eq$enu.mdcrd
gzip -9 cmps_eq$enu.mdcrd
let enu+=1
let emu+=1
done
```

这样类型的作业脚本在高性能集群中提交，存在两方面问题。一方面脚本中申请的资源是 4 个计算节点共 48core，amber 作业不适用在跨节点的并行模式下进行计算。作业提交后的计算效率是很低的，单核 CPU 的利用率不足 50%，跨节点计算的效率就更是低之又低。也就是说，原本可以在 1 个小时内结束的任务，采用此脚本，可能 2 天才能完成。两一方面，由于计算效率低，提交的计算命令多，任务会长期占用某 4 个计算节点，造成资源的低利用率，这与高性能集群高利用率的原则是相悖的，所以循环作业是不被允许的。

但是，在 HPC-Cloud 系统里，作业管理、作业调度、资源监控以及自动化配置等功能可以解决循环作业的问题。用户只需要简单的几个步骤，就可以提交一份多次循环的作业：

登录－选择脚本模板－添加脚本参数－保存提交－查看计算结果

用户在登录到 HPC-Cloud 系统内，直接选择循环作业脚本并填写作业信息即可，具体请见图 1。

5. 结束

针对高性能集群在提供科学计算服务时遇到的问题，本文初步提出了高性能计算云平台的解决方案，并列举了云计算技术在解决高性能集群问题中可发挥的优势及功能介绍。通过科学计算区用户的典型案例，从传统集群方式与 HPC 云模式对比出 HPC-Cloud 在处理用户常见问题上的优势。文中关于 HPC 云平台的功能设计，对研究服务其他领域的 HPC 云平台的也有一定的借鉴意义。

6. 致谢

感谢北京市计算中心提供的云计算环境，感谢云平台事业部领导、同事们的帮助。

参考文献 (References)

[1] Dostor (2013) HPC China 2013 三天大会内容汇总. IT 专家网. http://server.ctocio.com.cn/317/12764317.shtml

[2] 王恩东, 张清, 沈铂, 张广勇 (2012) MIC 高性能计算编程指南. 中国水利水电出版社, 北京.

[3] 张帅 (2009) 浅谈云计算如何落地? IT 专家网. http://security.ctocio.com.cn/securitycomment/388/8873388.shtml

[4] it 专家网 (2010) 让多核更加强大云计算催生 HPC 新模式. 中关村在线. http://server.zol.com.cn/206/2062526.html

[5] 新浪科技 (2011) 亚马逊: AWS 云计算服务盈利能力稳固. http://server.zdnet.com.cn/server/2011/0908/2056763.shtml

[6] 李云春, 霍建同, 王汉文, 杨秀梅, 郑剑 (2013) 高性能计算机群分布式强制访问控制技术可行性研究. HPC China 论文集, 272-278.