

Discuss the Importance of the High Performance Computing User Group in the Product Analysis

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: Sep. 25th, 2014; revised: Oct. 24th, 2014; accepted: Nov. 5th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Through the user research of high performance computing, it is concluded that high performance computing is based on Web management platform solution because there is no fit closely user group of the business process that has not been widely used in this paper. Then through the analysis of the material class user groups, we analyze the user group structure and summarize the user group of characteristics and the use of common types of problems. Take HPC-Cloud platform process innovations which have got the users' authentication and recognition for example, it intends to illustrate how important it is to analyze user groups and business characteristics in the product analysis stage.

Keywords

HPC, Users, Questions, HPC-Cloud

浅谈高性能计算用户群分析 在产品分析中的重要性

黄雪, 赵疏涛, 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：2014年9月25日；修回日期：2014年10月24日；录用日期：2014年11月5日

摘要

本文通过高性能计算用户调研，得出基于Web的高性能计算管理平台解决方案因为没有紧密贴合用户群业务流程致使其不被广泛应用。继而通过材料类用户群体分析出发，分析用户群体结构，总结用户群特征以及使用中常见类问题。并以得到用户验证且认可的HPC-Cloud平台流程创新点为例，阐明用户群体以及业务特征分析对于产品分析、定义阶段的重要性。

关键词

高性能计算，用户，问题，HPC-Cloud

1. 引言

高性能计算在兴起的这些年间，服务了包括航天、气象、科研等在内的各个领域。专注于高性能计算、并行算法、体系架构、软件开发等研究方向的专家们，在各自的领域也一直有不断的创新[1]，计算效率指数的上升，体系架构的不断变换[2]，使得高性能计算承担着越来越重要的角色。但是，从使用高性能计算的用户群体[3]来讲，在关注于效率的同时，使用时的体验感也是被带入考虑的重要因素之一。如今，高性能计算的用户群体组成已经发生了翻天覆地的变化。除了专注于计算的研究机构，更多的高校实验室、企业单位也加入了使用高性能计算的行列。其中以高校用户增加最为迅猛，占据了近九成的比例。

高性能计算的客户体系，不论是高校实验室，还是研究机构或企事业单位，使用者均是科研人员，从事生物基因、环境断裂、材料科学、仿真模拟、气象预测等大数据研究。每个行业在集群上计算的任务特点各不相同，以材料类用户为例，模拟类作业并行程度较高，对CPU要求高，对内存要求相对较低。数据分析类作业则正好相反。普遍作业在128core规模以下就能解决，单核内存2GB即可。根据用户类型、作业类型以及成熟、规范的业务流程设计HPC产品，才会被市场认可。相反，目前市场上还没有稳定的商业版基于Web的HPC管理平台的原因，就是设计时只粗略提取了部分业务需求，没有进行深入的用户群体分析，忽略了用户对于产品的根本需求，导致产品生命周期短暂即逝。由此笔者进行了一次验证，调查涵盖高校、企业、科研院所等27家单位89位用户，受访用户对HPC业务的感受见图1，大部分用户放弃使用HPC业务是迫于无法承受使用中出现的过多问题导致的。图2是用户使用基于Web的高性能计算管理平台的调研，多数用户反映虽然功能基本满足，但业务流程繁杂，不成系统，使用起来有种力不从心的感觉。由图1、图2可见，使用Web技术可以解决绝大部分用户关于使用方面的问题，但是因为把握用户不到位导致产品定义不成功，因此使用效果不理想，问题比较多。

本文重点讨论现阶段高性能计算主体用户的特征，以及使用高性能计算时的常见问题，并结合用户最关心的业务需求介绍HPC-Cloud解决方案在业务流程上的创新点。最终总结用户群体分析在基于Web的高性能计算管理平台产品定义中的重要意义。

2. 用户群体特征

在笔者服务高性能计算用户两年的经验来讲，高校用户占有所有用户的95%之多。总结用户群体的科研性质，发现有如下几个特点：

用户对HPC业务使用调研

- 使用流畅，计算效果满意
- 问题较少，计算效果满意
- 问题较多，能克服坚持使用
- 问题多，放弃使用

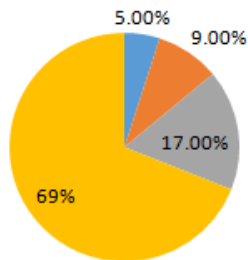


Figure 1. User research to HPC business

图 1. 用户对 HPC 业务使用调研

使用情况调查

- 基本满足业务功能，没有亮点功能
- 页面设计差，使用不流畅
- 业务流程繁杂
- 计算效果不理想

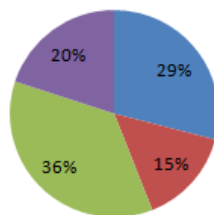


Figure 2. Investigation of the usage of similar products

图 2. 对类似产品使用情况调查

一带多：资源由实验室课题组的主要负责人使用。根据工作量大小，由负责人分配计算资源，所有使用者基本都是课题组里的学生。这种一带多模式区别于导师带学生，基本上同一账户的多个使用者之间没有有效的交流机制。所以会出现需求不统一的情况。

阶段性：每年的两个开学季，是用户任务并发量最大的时间。这个时间段里，系统会出现持续性排队现象。相对的每年寒暑假则是系统利用率最低的时期。但是用户在离校前会大量提交作业，造成临时性突发排队。

知识性：一般新接触高性能计算的用户，对高性能计算的认知程度很低，需要加大技术支持力度。而续签或者通过介绍签约的用户，通常至少会有 1 人对业务熟练，因此不需要花费太多的时间、精力重新培养。

协调性：一般研究人员多为博士研究生、硕士研究生等，用户容易沟通，较少出现沟通障碍。

针对以上高校用户的特点，在 HPC 业务工作中遇到的问题大致有以下几类：

- ◆ 一带多模式延续下去带来的问题是，当新的课题组接手 HPC 业务后，需要重复的花费时间、精力培养客户。
- ◆ 高校用户因为不熟悉 HPC 业务，所以在提问题、展示问题、阐述问题时难以做到表述清晰，并且同样的问题会重复多次。造成解答问题时不能快速、准确定位问题原因。
- ◆ 同一账户使用的人数不固定，使用账户的要求意见难以统一。造成在与技术支持沟通问题时，使问题复杂化。

以上所述可以通过有效的沟通机制、规范的问题提交模板等解决。但是用户周而复始的提问相同的问题，解决方法要从根本抓起。用户问题最大的两个特点，一是重复性。单一用户重复同样的问题。不同用户提问同样的问题；二是简单性。提交的问题大多都是简单的系统、应用使用问题。

3. 高性能计算使用隐性要求

使用高性能计算资源时，要求使用者具备的知识包括：

- 基础的 Linux 系统知识。包括但不限于基础命令的使用、权限管理、简单 shell 脚本等。
- 部分高级的 Linux 知识。包括但不限于系统安装配置、PAM 机制、软件安装、环境配置等。
- 基础的作业调度系统知识。包括但不限于作业提交、作业查询、作业资源策略等。
- 并行软件编译、使用识。包括但不限于软件的多种并行编译方法、多元化的并行作业提交方式等。

这些隐性要求是阻碍高效使用计算资源的主要原因，每个用户都至少缺乏以上两种类型知识，从而导致使用资源时不断出现使用规范类问题、基础常识类问题、应用编译类问题等多种问题[4]。例如作业脚本参数设置混乱、超线程作业占用资源、应用并行编译效率差等。

4. 解决方案

虽然通过技术支持可以解决上述问题，但时间、效果上会不尽如人意。基于 Web 的 HPC 管理平台解决方案完美地解决了这一尴尬问题。因为产品处于发展期，所以在功能、流程上设置还存在一定的不足，原因是在产品定义阶段缺少深入的用户群体分析，针对不同用户设计产品才是会被市场、用户接受的最终道理。北京市计算中心有多年材料类高性能计算服务经验，结合材料类用户群体、业务流程特点，设计了 HPC-Cloud 平台，且得到了用户的一致认可。材料类用户特点主要分两点，一是以业务规模类型多变，参数变化频繁，计算密集型应用较多等为业务特点；二是符合大众用户缺乏集群知识等行业特点。在充分考虑用户群体以及计算业务的特点，平台架构设计如图 3。

HPC-Cloud 平台创新点如下：

- 基于云计算的 SaaS 层[5]服务：HPC-Cloud 是基于公网运营的云计算平台，用户采用账户租赁模式使用集群资源，账户可预存费用，按照租赁模式按需付费。
- 业务管理流程层级清晰：平台不只高级管理员、用户两个层级，还添加了协助管理员，也称业务管理员。业务管理员负责管理用户业务流程上相关事宜，根据用户业务特点优化业务流程。
- 添加试用用户管理业务：不只是集群付费用户可使用该平台。对 HPC 业务有兴趣想要试用的用户也可以在自助模式下直接试用。每 15 天提供 10 个试用账户，试用用户可顺延领取到试用账户的用户名和密码，在有限资源内使用集群资源。
- 提高集群资源、策略透明度：集群资源[6]的使用信息、负载信息、队列策略等在资源管理模块都有明确标注，使用户深入了解集群资源，避免队列资源拥堵造成的“堵车”。
- 多种策略管理：为满足各种计算资源需要，制定了大内存队列、串行作业队列、普通队列、VIP 队列、中型内存作业、多线程作业队列等各种类型队列策略。
- 促进业务交流，拓展业务范围：对于集群业务流程的任何意见与建议，用户都可以在固定板块畅所欲言。还可以对材料类应用软件使用规范、技巧在论坛进行交流。
- 开放式登录、交互式管理：Web 方式与传统客户端方式都可使用，为不同需求的用户提供便利。同时移动客户端 APP 也可登录，满足用户随时随地关注任务情况的需求。用户可在线与技术人员进行交流，保证问题及时得到解决。

根据材料类用户的业务规模多变，特点制定了图 4 是 HPC-Cloud 业务流程，该流程除了满足用户常

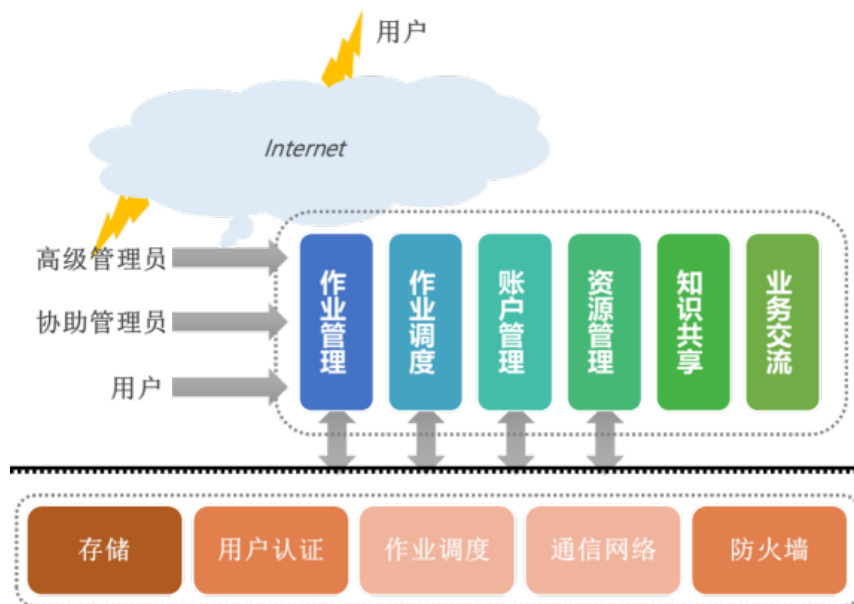


Figure 3. System architecture of HPC-Cloud platform
图 3. HPC-Cloud 平台体系架构

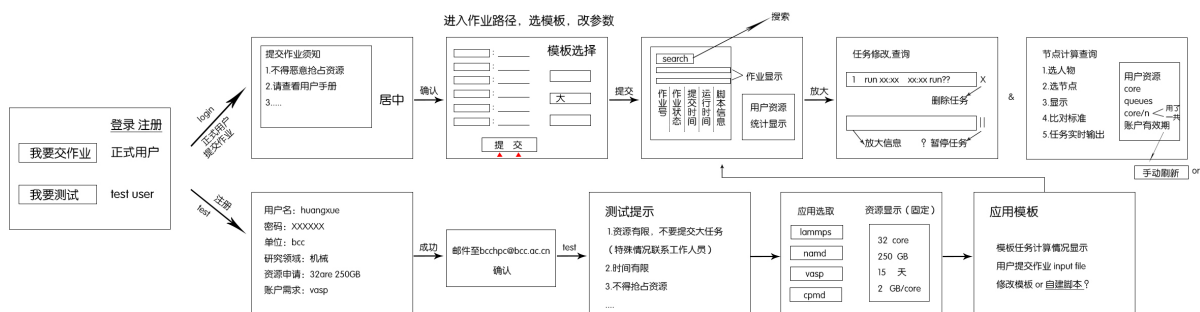


Figure 4. Business process of HPC-Cloud
图 4. HPC-Cloud 业务流程

规性需求的同时，还满足了用户对于灵活计算环境的要求。在不影响付费用户计算的同时，还能满足试用用户查看集群计算效率的要求。

图 5 是平台内的作业模板，在常规脚本参数设置的同时，考虑用户对于重要作业状态实时关注的需求，添加了邮件/短信提醒的功能。

5. HPC 用户群体分析的重要意义

如今以高校主导的用户群体主要以计算化学、分子动力学、材料仿真、生命科学等几大主要研究领域。每个高校的化学、生物、材料、机械等学院都是潜在用户，在紧张的课题项目压力下，用户没有额外的时间学习 HPC 集群是专业知识。同时，因为没有资金、或者没有专业机房知识不能搭建自己的机房，只有租用计算资源。全国可查的高校有 849 所，其中 100 多所高校在使用 HPC 业务，有 87% 的市场是未开发的。所以，HPC 业务市场拥有可观的市场前景。用户群体分析是进行产品定义的前提，是制定产品需求的立足点，是占领市场的核心竞争力。用户分析包括用户特征、使用习惯、行为习惯、行业特征等等，因为使用 HPC 业务的行业很多，每种行业都有自身的特点，只提取大众特点设计出的产品，受众面反而是最小的，如 IBM 网页版的 PAC (IBM Platform Application Center) [7] 就是典型的例子。因为 PAC 专为业务单一的用户准备，

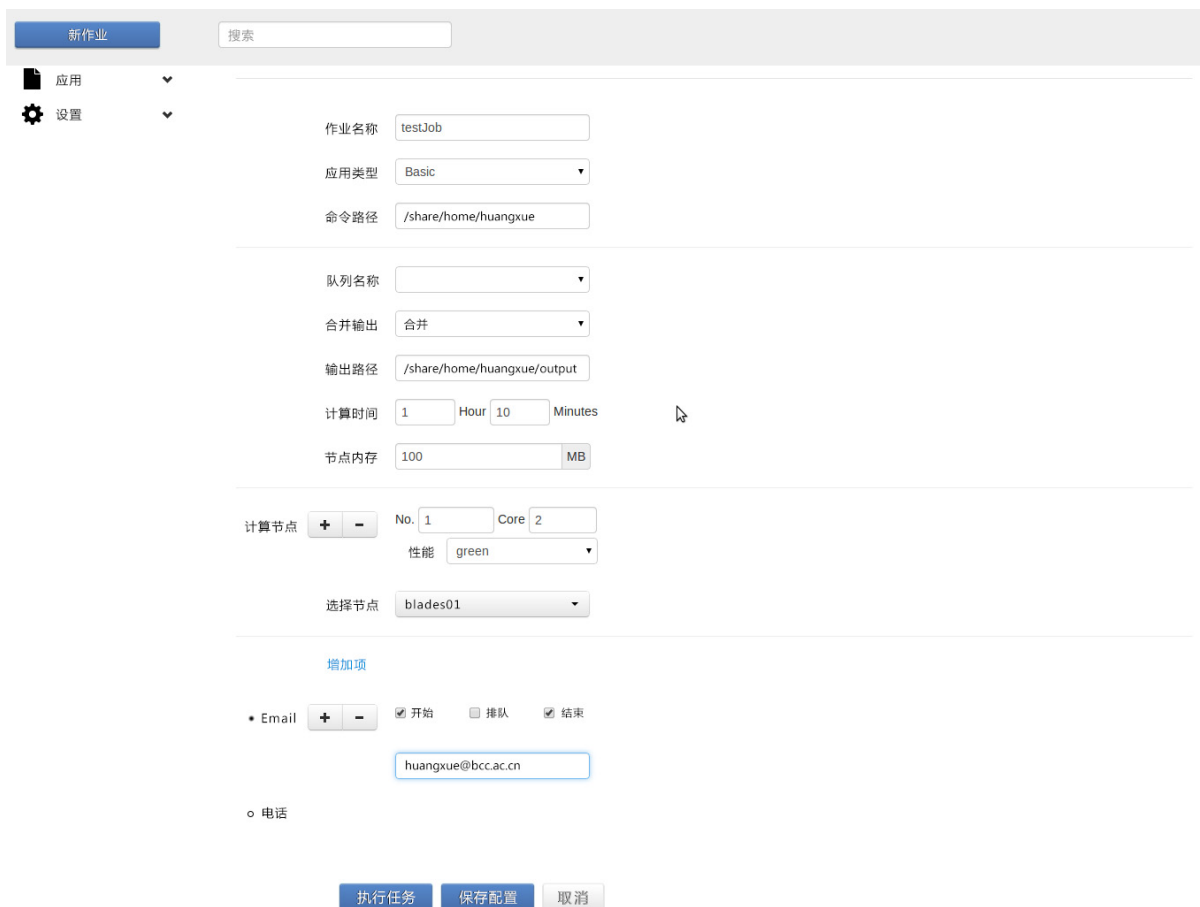


Figure 5. Job template of HPC-Cloud
图 5. HPC-Cloud 作业模板

客户计算类型一致，计算规模相仿。鉴于这些前提要求，虽然 PAC 软件质量高，却没有被大范围的应用，甚至很少用户知道。由此可见，产品前期用户群体分析要全面、深入，紧密贴近用户业务需求出发。

本文针对基于 Web 的高性能计算解决方案一直不被广泛应用现象展开讨论，通过用户调研得出问题原因。分析原因并以 HPC-Cloud 平台为例展示用户关注的业务点，阐明用户群体以及业务特征分析对于产品分析、定义阶段的重要性，为其他正在研究、开发基于 Web 的 HPC 平台具有一定的借鉴意义。

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] 远帆尔. 高性能计算的发展现状及趋势. http://blog.163.com/zjh_upc/blog/static/12259200820111212537239/
- [2] 360 百科. 高性能计算. <http://baike.so.com/doc/6189185.html>
- [3] Inspur 浪潮. 2013 年高性能计算用户大会. <http://www.it168.com/redian/inspur2013>
- [4] 范秋生 (2011) 浅谈集群管理. *电子商务*, 12.
- [5] 张帅. 浅谈云计算如何落地? IT 专家网. <http://security.ctocio.com.cn/securitycomment/388/8873388.shtml>, 2009.5.26

- [6] 琚攀攀 (2010) 基于 LSF 的集群管理研究——基于 Web 的高性能计算集群管理系统的设计与实现. 西安建筑科技大学, 西安.
- [7] IBM Platform Application Center.
<http://www03.ibm.com/systems/platformcomputing/products/lsf/applicationcenter.html>