

EasyCluster V1.6 简明管理员操作手册

Easycluster V1.6 是由南京君富科技有限公司开发的用于 High Performace Computing(简称 HPC) Cluser 管理的平台级软件工具，包含任务调度/管理、系统监控、增值功能几个功能模块。

登录/退出

Easycluster V1.6 采用 Web 界面，因此，用户的操作多数是通过浏览器界面提交的，因数据是压缩传输的，因此速率和保密性均有增加。已经测试过的浏览器有 IE6.x 和 Firefox 1.5。由于 IBM 和 SUN 公司的 java 仅提供了 32-bit 的 plug-in，因此，64-bit 客户端系界面在使用上有限制。

首次登录或使用 Easycluster V1.6 系统，用户要检查自己的系统中是否安装了 java 运行环境，最好是在 Easycluster 的登录界面下方链接处，由 Easycluster 的管理节点下载 JRE 环境并安装。在 Windows 环境下安装 JRE 非常简单，安装完毕，请在“控制面板”的 java 项目中，将自动更新项目设置为无效。若客户端是 Linux 环境，则安装 JRE 环境后，需要在 firefox 的 plug-in 目录中手工增加一个控件的链接文件。

在登录界面上，输入用户名/口令，鼠标点击“进入系统”按钮，即可登录系统并提交任务了。若打算退出系统，请鼠标点击界面右上角的“退出”菜单，而不要直接点击窗口右上角的“X”，这样，窗口关闭了，可是服务器端认为链路依然存在，所以占用的资源暂时不会释放掉。

提交普通任务

该功能仅针对普通用户，管理员不能提交任务。提交任务的功能比较多，这里仅介绍普通任务的提交。如图 1 和图 2 所示，

EasyCluster 首页 > 提交任务

基本信息

任务名称: _____

任务类型: 串行 单机 (SMP) 并行 多机 (分布式) 并行

执行程序: _____ 浏览...

工作目录: _____ 浏览...

输出目录: _____ 浏览...

任务队列: ----- 选择任务队列 ----- ▾

并行环境 (PE): ----- 请先选择一个任务队列 ----- ▾

进程数: 最小: _____ 最大: _____

高级选项

提交 重置

图 1. 普通任务提交界面

高级选项

合并输出:

邮件提醒: EMail地址: _____
 任务启动 任务完成 任务挂起 任务恢复

启动时间: _____ 浏览...

命令行选项: _____

程序参数: _____

环境变量: _____ 浏览... 重置

资源: _____ 浏览... 重置

License: _____ 浏览...

隶属项目: ----- 请先选择一个任务队列 ----- ▾

任务依赖: _____ 浏览...

主任务队列: ----- 请先选择一个任务队列 ----- ▾

优先级: _____

提交 重置

图 2. 提交任务的高级选项

普通任务提交界面有下列项目，

任务名称

该项目要求以英文字母、数字填写，最多 8 个西文字母，且首字母必须为英文字母，不能在其中夹杂空格或下划线等字符。任务名称主要方便用户识别自己提交运行的众多任务，任务调度系统管理任务借助于 **jobID** 来管理任务，**jobID** 是全局唯一的，当任务提交成功后，无论任务能否正确或成功执行，系统均会创建提供一个唯一的 **jobID**，并以此为依据来管理、查询任务的状态。

任务类型

任务类型分为串行、单机(SMP)并行、多机(分布式)并行三种类型，其中串行是指单机单进程单线程的任务；单机(SMP)并行是指在一台机器上运行的多线程

程序，即用户的程序采用了 **multi-thread** 技术，只能在一台机器上运行的并行程序；多机(分布式)并行是指采用 **MPI** 技术，可以将任务分布在多台机器上计算的并行程序；只有选择了正确的任务类型，系统才能正确提交任务。

执行程序

该项目必须选择具有执行权限的程序，用户用鼠标点击右边的“浏览”按钮，选择正确的执行程序。

工作目录

该项目是指数据文件所在的目录，若不需要数据文件，则也需要指定一个路径，通常系统会自动在用户选择了执行程序后自动将该路径填写在项目中，用户可以用鼠标点击右边的“浏览”按钮，选择正确的路径。

输出目录

该项目是指将来存放调度系统创建的任务的 **stdout** 和 **stderr** 信息，文件名为 **<任务名>.o<jobID>** 和 **<任务名>.e<jobID>**。若在高级选项中使用“合并输出”，则系统仅创建一个输出文件 **<任务名>.o<jobID>**，**stderr** 的信息也会写入 **<任务名>.o<jobID>** 文件。

任务队列

该项目是指用户目前提交的任务在那里排队、调度、运行。任务队列是一个用户授权和运算资源的整体，用户必须要选择一个队列。若该项目的下拉项目内容是空的，请联络管理员。

并行环境(PE)

该项目仅适用于并行任务，串行任务无需并行环境。若你提交的任务是单机多线程的就请在“任务类型”项目选择“单机(SMP)并行”，否则选“多机(分布式)并行”。系统会自动根据“任务类型”、“任务队列”为你提供相应的选项列表，用户再根据自己的需要，进一步确定一个并行环境。若没有“并行环境”，请联系管理员。

进程数

该项目是一个区间，仅针对并行任务，用户可以根据自己的需要、**Cluster** 的负载情况来决定。例如，当程序限定 4 个进程做计算，则上下区间均填写 4；当任务的进程无限制时，可填写 4—20，表示有只要有四个 CPU 核的资源空闲，任务就可实施计算，最多 20 个 CPU 核，系统根据实际的空闲资源自动设定任务的实际进程数(CPU 核数目)。管理员会为每个用户设定一个任务允许的进程数

目限制，用户填写的上限必须小于等于该数字。

合并输出

任务调度系统会收集每个任务的屏幕输出信息(由 `stdout`、`stderr` 产生)到指定的文件中，该文件存放于用户指定的输出目录，文件名为 `<任务名>.o<jobID>` 和 `<任务名>.e<jobID>`。当该项目有效，则仅创建一个名为 `<任务名>.o<jobID>` 的文件，`stderr` 的信息也写入该文件中；否则，分别写入两个文件。

程序参数

该项目仅针对命令行运行程序时需要有输入参数的任务。例如，运行程序需要增加的输入数据文件、输出结果文件、运行参数/选项等。

其它选项涉及到比较复杂的调度策略，在高级策略相关的科目中描述，这里就省去了。在正确填写这些必要的项目后，鼠标点击页面右下脚的“提交任务”按钮，任务就会提交调度系统，系统返回一个任务的 `jobID` 号，表示该任务已经进入系统排队。

管理任务

任务提交系统排队后，调度系统就有了该任务的信息，普通用户和管理员均可以通过右边的菜单栏的“任务管理”-->“查询未完成任务”浏览目前的未完成任务列表。



<input type="checkbox"/> 选定	任务ID	优先级	任务名称	所属用户	状态	任务队列	进程数	展开
<input type="checkbox"/>	986	0.605000	FeCrS	qili	dr	all.q	6	⌵
<input type="checkbox"/>	1057	0.505000	d369	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1051	0.505000	d363	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1050	0.505000	d362	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1053	0.505000	d365	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1056	0.505000	d368	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1049	0.505000	d361	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1054	0.505000	d366	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1055	0.505000	d367	fwang	r	all.q	1	⌵
<input type="checkbox"/>	1096	0.605000	stat	qili	qw		6	⌵

图 3. 查询未完成任务列表(管理员界面)

管理员的界面和普通用户的不同，普通用户仅能够看见自己的任务，管理员则能够查看所有用户任务列表。

管理员能够“保持/重排”（正在排队的任务）、“挂起/恢复”（运行的任务）、“删除”、“加票点”（正在排队的任务）、“清除错误状态”等；普通用户只能够删除自己的任务。其中，“保持/重排”是指使排队的任务停止或重新排队；“挂起/恢复”是指运行的任务停止或继续运行，挂起的任务虽然不运行，但是并未退出系统，因此占用的资源也不会释放；“加票点”是使某些排队的任务往前排队的几种方法，具体的效果还要配合其它的高级策略一起实施，这里就不再多描述了；“清除错误状态”，当调度系统的执行模块遭遇异常或任务本身有问题，会导致错误状态的出现，任务一旦被认定是处于错误状态，“状态”列就会出现“E”，表示该任务有问题了，需要联系管理员，管理员使用该功能来清楚任务的E状态，使其继续运行。

每项任务列表的右边有一个向上的双箭头，鼠标点击双箭头(如图4所示)，任务的详细信息就展开显示，这对于查看多机并行任务很有帮助。特别是并行程序的任务，用户能够观察到并行任务在各个节点上的分布状态。状态栏会显示任务当前的状态 —— d(letion 删除) E(rror 出现错误) h(old 保持，即排队的任务停止排队) r(unning 正在运行) R(estarted 重新运行) s(uspended 挂起) t(ransfering 正在发布) qw(排队等候)。



图 4. 任务的详细信息

任务状态可以是组合的，例如，dr——表示该任务正在被删除。若用户的任务长期处于 dr 状态，说明系统不能删除该任务，这可能是因运行该任务的节点意外关机导致的，此时管理员需启动该节点，待节点启动后，在 Easycluster 管理员界面，再次删除此任务，若仍不能删除该任务，则在命令行状态以 root 身份执行命令：`qdel -f <jobID>`

如何在管理节点手工检查 Easycluster 是否启动成功?

Easycluster 系统的模块很多，启动的步骤也很复杂，在管理节点重新启动时，因某些模块的运行条件未满足，而导致启动失败的情况是有可能发生的。检查 Easycluster 是否启动成功的办法很简单，只要在浏览器的地址栏输入 Easycluster 服务器的地址，例如，<http://99.0.0.80/easycluster/>，浏览器能够让用户登录进入系统，则说明系统各模块启动成功。若浏览器中无法出现登录界面，则说明 Easycluster 的前台接口模块未启动成功，若有界面出现，但是不能登录系统则说明，后台模块未启动。

前台界面的检查和启动

`#ps aux | grep java` --- 该命令检查结果应发现两个 java 的进程，没有两个说明，前台模块未成功启动

`#/opt/easycluster/scripts/stop_interface.sh` --- 该命令是终止前台的界面模块的运行

`#/opt/easycluster/scripts/start_interface.sh` --- 该命令是启动前台的界面

后台界面的检查和启动

`#ps aux | grep easy_` --- 该命令检查结果应发现有 easy_s 进程，否则说明后台模块未启动

`#/opt/easycluster/background/easy_s` --- 该命令启动 easy_s 进程

如何删除系统帐户

删除系统帐户的功能比较复杂，Linux 在删除用户帐户后，会保留其目录和文件，目的是担心用户还需要自己的文件，给用户一个备份文件的机会。但是，当管理员将来创建一个同名的帐户就会遇到麻烦，因为用户主目录已经存在，但是用户确没有操作权限。因此，管理员通过 Easycluster 界面删除用户帐户后，首先，要确认将用户数据文件备份至某个特殊的目录，然后，手动删除该用户的主目录，这样，将来创建用户就不会出问题了。

用户/组的管理

管理员可以通过 Easycluster 来管理 Cluster 系统的用户和组，若用户手工在命令行环境添加了新用户或删除了某个帐户，应在 Easycluster 管理员界面执行“帐户管理”→“帐户同步”步骤同步 Easycluster 和 Linux 系统间的帐户信息。通常用户在 Easycluster 界面上修改用户帐户信息后，Easycluster 会自动将修改信息同步至整个 Cluster，使其生效。因此，管理员因优先采用 Easycluster 来管理用户的帐户。

通过 Easycluster 添加用户，会要求输入姓名和所属的组，这样做使系统的帐户数据看上去更简洁，缺省状态下，Redhat AS4 会自动为用户创建一个同名的组，导致用户组很多；SuSE Enterprise10 会把用户添加至 users 组。因此，通过 Easycluster 创建用户前，应先做简单规划，创建一些组，接下来再创建新用户帐户。

管理员在创建了新帐户后，必须为其再做三个设置，以便该用户能够通过 Easycluster 界面提交/管理任务 —— 1、“计费记账”→“用户充值” 2、“计费记账”→“用户计费标准设定”，“CPU 时”指实际的 CPU 资源占用时间，“任务运行时”指 3、“调度策略配置”→“设定用户提交任务限制”。

任务队列

任务队列是计算资源和用户授权的组合体，队列包含了节点、并行环境(PE)、允许/拒绝的用户集等项目。节点列表是指发布至该队列上的任务能够运行在那些节点上；并行环境仅针对并行任务，即该队列上允许的并行任务类型和策略列表，并行环境定义任务调度系统应如何启动并行任务、任务终止运行后如何处理、如何发布并行任务等；允许/拒绝用户集列表中，以拒绝的优先级更高，即某个用户即被拒绝又被允许，则该用户仍被拒绝，若未定义这两项列表，则任何用户都能够发布任务至该队列，不过，PE 也能够限制用户，这里是指在队列这个层次定义用户操作权限。其它项目涉及其它的高级策略配置，因此，这里就不描述来了。

创建任务队列时，管理员填写的队列名称也要求 8 个字母的长度限制，且首字母是英文字母，创建成功后，队列名称会自动被系统增加 .q 的后缀，因此用户自己不要在名称中添加 .q 了。

并行环境

并行环境定义了如何发布并行任务、用户授权列表、最多能够在该 PE 排队或运行的进程(或线程)数目等。对于刚刚接触 Easycluster 的管理员来说, 在创建普通任务 PE 时, 有三个项目比较难理解 ——

- 1、分配策略: 填充或循环, 填充是指发布任务时尽量填满一个节点的资源, 然后再寻找下一个有空闲资源的节点; 循环是指找到所有的有空闲资源的节点, 每个节点填一个进程, 若还有进程, 则再依此策略做一遍, 直到分配完任务的进程
- 2、PE 类型: 单机(SMP)并行或多机(分布式)并行, 多机(分布式)并行是指采用 MPI 的并行(特殊的应用程序除外, 例如, Linda), 单机(SMP)并行是指采用多线程并行的程序, 该任务只能在一个节点上运行
- 3、进程数: 是指使用该 PE 的所有任务的进程数或线程数的总和, 无论排队或运行的任务, 因此, 该项目不能够设置的过小, 否则使用该 PE 提交任务会遭遇没有资源的提示, 并无法排队的事情。

用户集

用户集是任务调度系统的概念, 这里不是指 Linux 系统的用户组, 一个用户可以同时隶属于多个用户集, 用户集可以被定义在 PE、任务队列的允许/拒绝的列表中, 以达到限制用户的目的。在创建用户集的时候, 在“类型”项目, 请选择“ACL”, 表示这是一个单纯的用户集, “DEPT”是定义 share-tree 策略需要的, 这里暂且省略。