

- 一 . PBS脚本编写使用
- 二 . 曙光Gridview作业调度中间件

PBS作业提交步骤

1. **准备**：编写描述作业的脚本，包括作业名，需要的资源等。
2. **提交**：使用qsub命令将该作业提交给PBS服务器
3. **排队**：服务器将该任务排入适当的队列
4. **调度**：服务器检查各工作节点的状态是否符合该作业的要求，并进行调度。
5. **执行**：当条件满足时，作业被发给相应的执行服务器执行。程序运行时执行服务器会收集程序的标准输出和标准错误流，等程序结束时，将这些信息返回给用户。
6. **查询和调整**：当作业在运行时，用户可以使用qstat进行状态查询。用户发现作业提交错误时，可以使用qdel删除正在运行的作业。
7. **查看结果**：使用文本编辑软件vi或者系统命令cat, less等查看输出及错误信息显示。

□ 在PBS系统中，用户使用 qsub 命令提交用户程序。用户运行程序的命令及PBS环境变量设置组成PBS作业脚本，作业脚本使用如下格式提交到PBS系统运行：

```
qsub <PBS作业脚本>
```

```
qsub -N test.vasp -l nodes=4:ppn=2 -q defaults <PBS作业脚本>
```

- 本质是一个SHELL脚本
- 注释以 “#” 开头
- PBS运行参数，以 “#PBS” 开头
- 可以直接调用SHELL命令和系统命令

```
#PBS -N vasp
#PBS -l nodes=1:ppn=1
#PBS -l walltime=12:00:00
#PBS -q high

cd /home/test/work
./test.exe
```

在 PBS 脚本和 qsub 命令行中均有效，qsub命令行参数的优先级更高

运行参数	说 明
-a <作业开始运行的时间>	向PBS系统指定作业运行的开始时间。 作业运行时间格式为： [[[[CC]YY]MM]DD]hhmm[.SS]
-A <用户名>	使用不同的用户来提交作业，缺省使用当前用户名
-o <标准输出文件的路径> -e <标准错误输出的路径>	该参数指定标准错误输出的位置，缺省的情况下，PBS系统把标准输出和标准错误输出放在用户qsub命令提交作业的目录下。 标准错误输出：<作业名>.o<作业号> 标准错误输出：<作业名>.e<作业号> 路径使用如下格式标准： [<节点名>:]<路径名>
-N <作业名>	指定提交的作业名
-q <目标队列>	指定作业提交的目标队列，其中目标队列可以是目标队列、目标节点名或者是目标节点上的队列。如果目标队列是一个路由队列，那么服务器可能把作业路由到新的队列中。如果该参数没有指定，命令qsub会把作业脚本提交到缺省的队列中。
-l <申请资源列表>	该参数指定作业脚本申请的PBS系统资源列表。 申请资源列表使用如下格式： <资源名>[=[<数量>]][,资源名[=[<数量>]],] 例如作业希望申请在双路节点上申请5个CPU资源的情况，则可以在脚本中如下： #PBS -l nodes=2:ppn=2+1:ppn=1

PBS的环境变量

变量名	说明
登陆SHELL继承来的变量	包括\$HOME, \$LANG, \$LOGNAME, \$PATH, \$MAIL, \$SHELL和\$TZ。
\$PBS_O_HOST	qsub提交的节点名称
\$PBS_O_QUEUE	qsub提交的作业的最初队列名称
\$PBS_O_WORKDIR	qsub提交的作业的绝对路径
\$PBS_JOBID	作业被PBS系统指定的作业号
\$PBS_JOBNAME	用户指定的作业名，可以在作业提交的时候用qsub -N <作业名> 指定，或者在PBS脚本中加入#PBS -N <作业名>。
\$PBS_NODEFILE	<p>PBS系统指定的作业运行的节点名。该变量在并行机和机群中使用。当在PBS脚本中用#PBS -l nodes=2:ppn=2指定程序运行的节点数时，可以使用\$PBS_NODEFILE在脚本中引用PBS系统指定的作业运行的节点名。比如：</p> <pre>#PBS -l nodes=2:ppn=2 mpirun -np 4 -machinefile \$PBS_NODEFILE <程序名></pre>
\$PBS_QUEUE	PBS脚本在执行时的队列名

PBS脚本举例

```
# 这是一个串行作业脚本的例子
#PBS -N test
#PBS -l nodes=1:ppn=1

cd $HOME/test/
./a.out > $HOME/result/a.result
```

```
# 这是一个并行作业脚本的例子
#PBS -N vasp_job
#PBS -l nodes=2:ppn=8
#PBS -q low

echo This jobs is $PBS_JOBID@$PBS_QUEUE
cd $PBS_O_WORKDIR
mpirun -np 16 -machinefile $PBS_NODEFILE ./vasp
```

PBS脚本举例（续1）

□ 有时在PBS脚本中，需要对PBS环境变量的内容进行改造

□ 比如，\$PBS_NODEFILE，该文件内容格式为：

node1

node1

node2

node2

□ 对于一般MPI程序，可直接将 \$PBS_NODEFILE 作为 MPI 的
“-machinefile” 参数，如上例所示

□ 而一些软件有特殊的节点指定格式，比如ANSYS的命令行参数格式为：

ansys121 -dis -machines node1:2:node2:2 -i test.inp -o test.log

□ 这时我们可以对 \$PBS_NODEFILE 进行字符处理，得到需要的格式

PBS脚本举例（续2）

这是一个ANSYS并行作业的例子

```
#PBS -N ansys_job
```

```
#PBS -l nodes=2:ppn=8
```

```
#PBS -q low
```

```
INPUTFILE=test.inp
```

```
OUTPUTFILE=test.log
```

```
hosts=`cat $PBS_NODEFILE | uniq -c | awk '{print $2":"$1}' | tr '\n' ':' | sed 's/:$//'`
```

```
cd $PBS_O_WORKDIR
```

```
ansys121 -dis -machines $hosts -i $INPUTFILE -o $OUTPUTFILE
```

一个复杂的PBS脚本

```
#!/bin/bash
#PBS -N jobname
#PBS -l nodes=2:ppn=8

cat `echo $PBS_NODEFILE` > $HOME/$PBS_JOBID.nodes
for node in `cat $HOME/$PBS_JOBID.nodes`
do
    ssh $node mkdir /tmp/$PBS_JOBID
    ssh $node cp -rf $PBS_O_WORKDIR/* /tmp/$PBS_JOBID/
done

cd /tmp/$PBS_JOBID
mpirun -np 16 -machinefile $PBS_NODEFILE $HOME/bin/vasp

cp -rf /tmp/$PBS_JOBID/* $PBS_O_WORKDIR/
for node in `cat $HOME/$PBS_JOBID.nodes`
do
    ssh $node rm -rf /tmp/$PBS_JOBID
done
rm $HOME/$PBS_JOBID.nodes
```

指定节点特性

```
# 这是一个并行作业脚本的例子
#PBS -N vasp.Hg
#PBS -j oe
#PBS -l nodes=2:ppn=12:amd
#PBS -q low

echo "This jobs is "$PBS_JOBID@"$PBS_QUEUE
NP=`cat $PBS_NODEFILE | wc -l`
cd $PBS_O_WORKDIR
mpirun -np $NP -machinefile $PBS_NODEFILE ./vasp
```

```
#PBS -l nodes=2:ppn=4:amd:chem
```

```
#PBS -l nodes=1:ppn=8:chem+1:ppn=8:bio
```

```
#PBS -l nodes=node1:ppn=8+node2:ppn=8
```

作业提交后，会生成一个作业号，如：

```
[dawning@node1 ~]$ qsub test.pbs
```

93.node1

查看集群作业运行状态：

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
-----	-----	-----	-----	-	-----
93.node1	test.pbs	test	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default

作业状态说明：

E：退出

Q：排队

H：挂起

R：运行

C：结束

查询作业状态（续）

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default

查询某个作业运行状态：

qstat 93.node1 (或者 qstat 93)

显示作业运行在哪些节点上：

qstat -n 93.node1

显示作业运行详细信息：

qstat -f 93.node1

取消删除作业

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default

```
qdel 93.node1
```

注：用户只能删除自己的作业，管理员可以删除所有用户作业

作业挂起及取消

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default

挂起作业：

```
qhold 111.node1
```

取消作业挂起

```
qrls 111.node1
```

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default

更改作业运行队列：

```
qmove high 111.node1
```

更改作业资源属性：

```
qalter -l walltime=10:00:00 111.node1
```


交换作业顺序

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
-----	-----	-----	-----	-	-----
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
111.node1	structure	amber	0	Q	default
112.node1	gaussian	gauss	0	Q	default

交换两个作业的排队顺序：
qorder 111.node1 112.node1

```
[dawning@node1 ~]$ qstat
```

Job id	Name	User	Time Use	S	Queue
-----	-----	-----	-----	-	-----
93.node1	test.pbs	zhaocs	0	R	default
95.node1	vasp.Hg	vasp	0	E	default
112.node1	gaussian	gauss	0	Q	default
111.node1	structure	amber	0	Q	default

- PBS脚本中可以指定多个作业之间的依赖关系，比如作业提交前另一个作业必须完成，否则处于排队状态

```
#PBS -N step2
#PBS -l nodes=4:ppn=4
#PBS -q high
#PBS -W depend=after:<JOB_ID>
...
```

- 当指定作业非正常结束，作业才能提交

```
#PBS -N job_rerun
#PBS -l nodes=4:ppn=4
#PBS -q high
#PBS -W depend=afternotok:<JOB_ID>
...
```

应用场景

使用相同的PBS脚本，提交多个作业，每个作业运行环境基本一致，除了个别运行参数有不同。

可用 #PBS -t 或 qsub -t 的方式提交这样的Job Array，每个作业用环境变量 \$PBS_ARRAYID 进行区分。

简单的例子array.pbs：

```
#PBS -N array
#PBS -l nodes=1:ppn=8
#PBS -j oe
#PBS -t 1-100

cd $PBS_O_WORKDIR
mpirun -np 8 ./myprogram -parameter $PBS_ARRAYID
```

pbsnodes命令的主要参数

- a 列出所有结点及其属性，属性包括 “state” 和 “properties”
- o 将指定结点的状态标记为 “offline”。这将帮助管理员暂时停止某些结点的服务
- l 以行的方式列出被标记的结点的状态，如 -l free，-l offline
- c 清除结点列表中的 “offline” 或 “down” 状态设置，使结点可以被分配给作业
- r 清除指定结点的 “offline” 状态

一 . PBS脚本编写使用

二 . 曙光Gridview作业调度中间件

作业提交

该作业调度系统主要通过web页面将作业提交给调度器。系统中有三种提交方式：

- (1)模板方式提交
- (2)脚本方式提交
- (3)命令行方式提交

每次只能选择一种方式提交，在打开提交窗口时，默认是第一种提交方式。

作业提交的界面

×

作业提交

基本参数

高级参数

应用模板

集群名称:

请选择目的集群...

▼

作业名称:

请输入作业名字...

队列:

请选择队列...

▼

工作路径:

请选择工作路径...

🔍

提交方式

☒ 模板方式

请选择应用类型...

▼

☐ 脚本方式

请选择作业路径...

🔍

☐ 命令行方式

请输入命令...

上传文件

提交

清除

取消

作业提交

1.模板方式提交

2.脚本方式提交

3.命令行方式提交

4.其他参数设置

5.文件上传

1 模板方式提交

基本参数

高级参数

集群名称:

作业名称:

提交方式

☒ 模板方式

☐ 脚本方式

☐ 命令行方式

作业提交

基本参数

高级参数

应用模板

模板详细内容如下所示:

```
#!/bin/bash
#PBS -N wrf.30KM
#PBS -l nodes=18:ppn=8
#PBS -l walltime=01:40:00
#PBS -j oe
#PBS -q @node24

echo "This jobs is "$PBS_JOBID@"$PBS_QUEUE

cd $PBS_O_WORKDIR

date
time -p /PATH/mpirun -rsh -hostfile $PBS_NODEFILE -np 144
./wrf.exe
date

exit 0
```

上传文件

提交

清除

取消

点击选择集群

点击选择模板



作业提交

1.模板方式提交

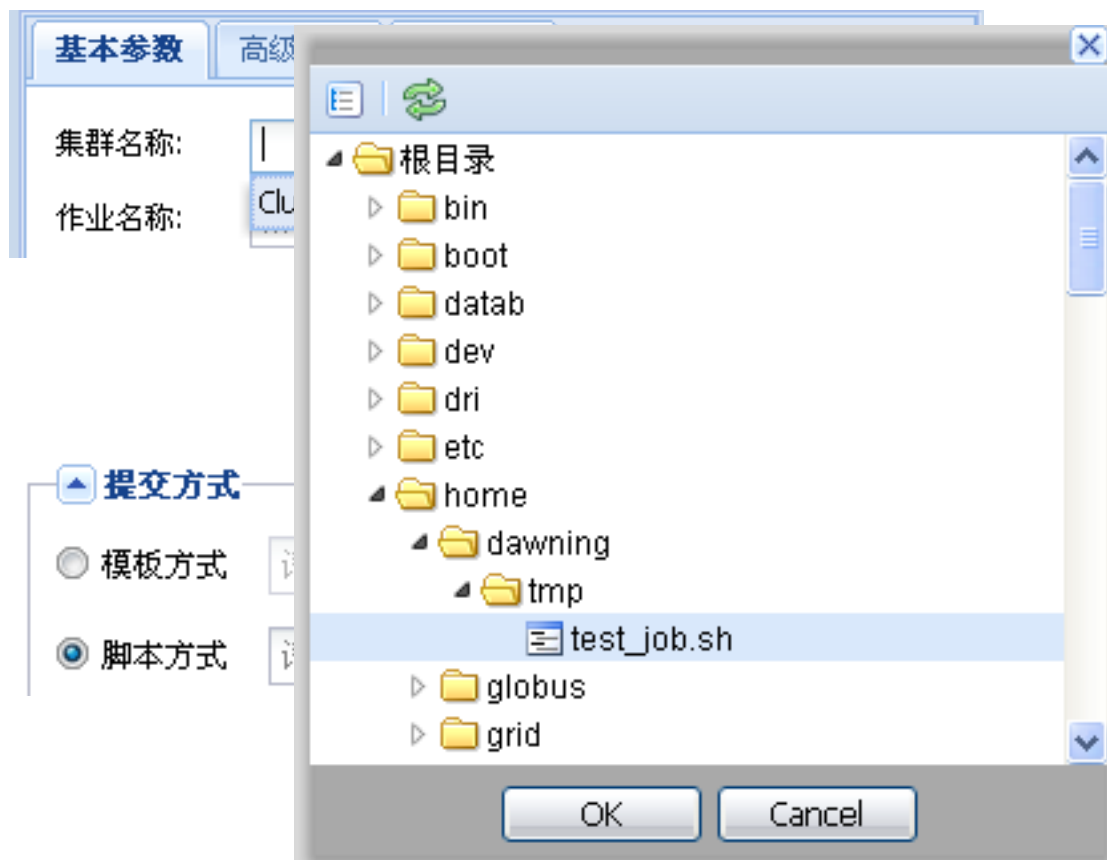
2.脚本方式提交

3.命令行方式提交

4.其他参数设置

5.文件上传

2 脚本方式提交



点击选择集群

点击选择脚本

作业提交

1.模板方式提交

2.脚本方式提交

3.命令行方式提交

4.其他参数设置

5.文件上传

3 命令行方式提交

基本参数

高级参数

应用模板

集群名称:

作业名称:

Cluster_dawning1(PBS)

点击选择集群



☒ 命令行方式

输入命令

作业提交

1.模板方式提交

2.脚本方式提交

3.命令行方式提交

4.其他参数设置

3.文件上传

4 其他可选参数

- 1.作业名称
- 2.队列
- 3.工作路径
- 4.节点列表
- 5.节点数
- 6.每个节点处理器数
- 7.标准输出
- 8.错误输出
- 9.批量提交
- 10.所需执行时间

作业提交

1.模板方式提交

2.脚本方式提交

3.命令行方式提交

4.其他参数设置

5.文件上传

5 文件上传

作业提交

基本参数

高级参数

应用模板

上传文件

上传文件选择:

浏览...

目的集群:

请选择目的集群...

▼

目的路径:

请选择上传的目的路径 ...

🔍

(*仅支持小文件(50M以内)上传)

上传

取消

上传文件

提交

清除

取消

上传文件选择: C:\Documents and Sett 浏览...

点击“浏览”

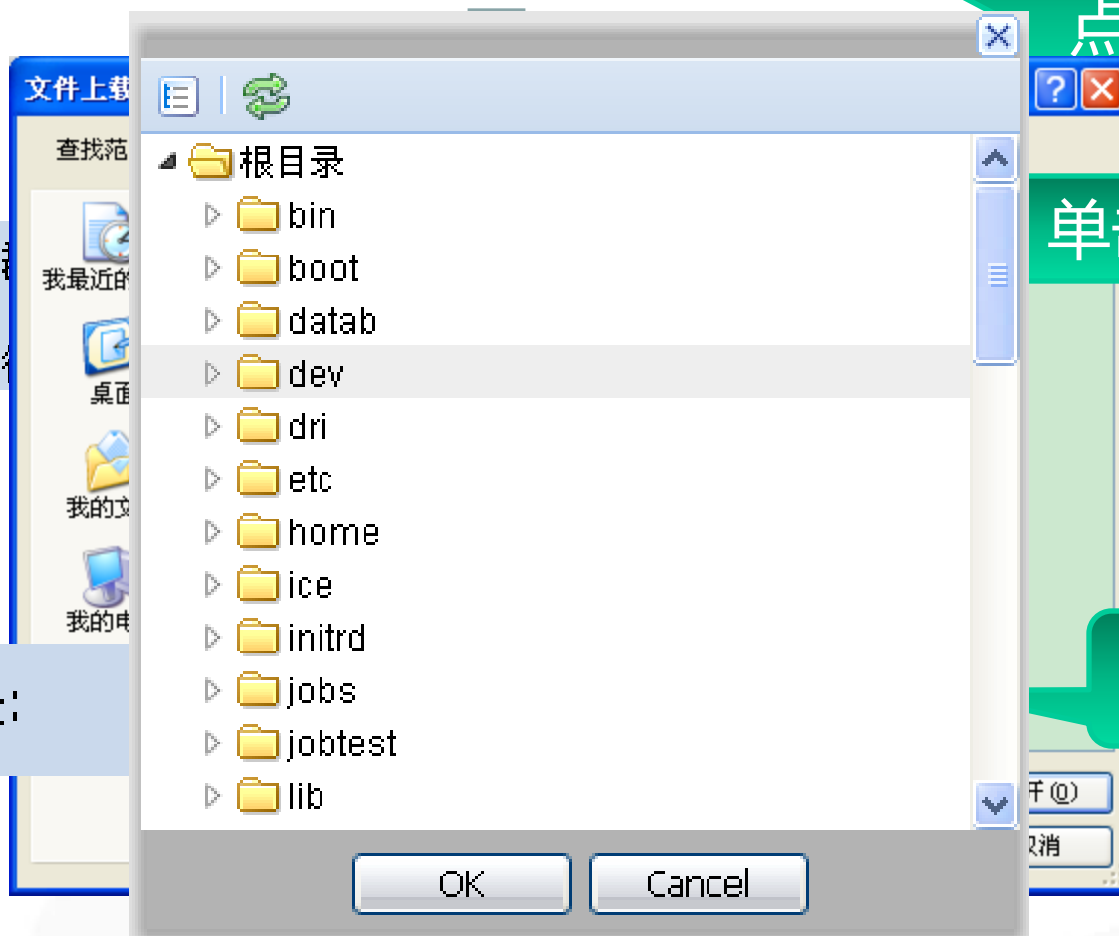
单击选择集群

选择目的路径

目的集群

目的路径

目的路径:



新增功能：基于应用的提交模版 Gridview Web Portal

vasp

集群类型:

PBS

VASP

VASP Portal Version 2.0 (with Checkpoint/Restart support)

Asterisk (*) = required field.

Job Name:

vasp_job

* Working Directory:

/home/libin/vasp.Hg

Browse_folder...

VASP version:

vasp.4.6.28

Network Type:

ethernet

Enable Checkpoint/Restart?:

yes

Checkpoint frequency (hour):

5

NO. of Nodes:

1

Cores Per Node:

4

Memory Maxsize(MB):

1024

CPU Time Maxsize(hh:mm:ss):

01:00:00

Queue:

high

Submit

Clear

作业管理

作业管理主要包含对作业的查询、控制和删除等操作。对作业的控制包括保留、释放、挂起、恢复、重新运行等操作。

作业管理界面

GRIDVIEW 曙光集群综合管理系统

作业提交 作业管理 节点状态监控 应用模板管理 调度系统管理

注册

作业管理

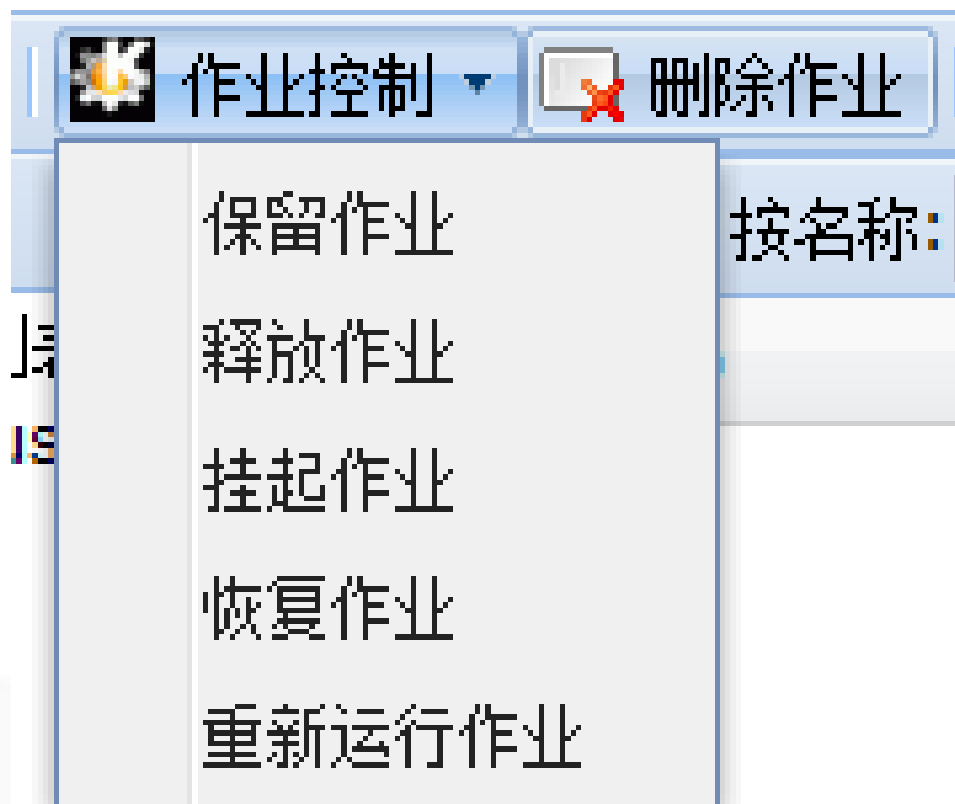
刷新作业 作业控制 删除作业 刷新设置 自动刷新

查询条件: 按名称: 请输入用户名 按状态: 请选择状态 按所有者: 请输入所有者 按创建时间: 开始查询

作业ID	作业名	集群名称	作业所有者	作业队列	创建时间	作业状态	执行主机	运行时间
29	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node92/2	00:09:16
28	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node92/1	00:09:16
27	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node92/0	00:09:16
26	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/7	00:09:14
25	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/6	00:09:14
24	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/5	00:09:14
23	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/4	00:09:14
22	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/3	00:09:14
21	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/2	00:09:14
20	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/1	00:09:14
19	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node91/0	00:09:14
18	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/7	00:09:04
17	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/6	00:09:04
16	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/5	00:09:04
15	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/4	00:09:04
14	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/3	00:09:04
13	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/2	00:09:04
12	STDIN	Cluster_node209	root	low	2009-04-16 17:26:55	Running	node90/1	00:09:04
11	STDIN	Cluster_node209	root	low	2009-04-16 17:26:23	Running	node90/0	00:09:36

作业控制

作业控制包括保留作业、释放作业、挂起作业、恢复作业和重新运行作业五种操作。



删除作业

在作业管理主界面的表格中选择要删除的作业，点击工具栏中的“删除作业”按钮即可实现删除操作。

管理员可以删除任何作业，普通用户只能删除自己的作业。

用户可以选择多个作业同时删除。