集群进程及任务管理脚本分享

植物研究部-生物信息工程师

陈 军

2019

目录

[1 写在前面 2](#_Toc13745834)

[2 集群进程及任务管理脚本分享 2](#_Toc13745835)

[2.1 ks：进程管理器 2](#_Toc13745836)

[2.2 qd：投递任务管理器 3](#_Toc13745837)

[2.3 qjk：集群job监控工具 4](#_Toc13745838)

[3 数据操作 4](#_Toc13745839)

[3.1 oss：集群数据互传工具 4](#_Toc13745840)

[3.2 md5/md5q：在当前目录计算所有文件md5和大小 5](#_Toc13745841)

[4 其他命令 6](#_Toc13745842)

[5 下期分享 6](#_Toc13745843)

# 写在前面

* 简介

快捷使用命令的一些操作

包含dfa, pp, p, ks, qd, qgjd, qjk, qjk, oss, md5, md5q, send, view, rep, fas, ccut, ca, cah, f, asum, getsize等10余个命令

* 使用方法

以下章节介绍的这些命令的使用，都只需要source一条命令后即可直接使用（可以将以下命令加入到.bashrc中，尽量加到.bashrc的末尾）

source的命令如下：

（注：mymail=xxx@xxx的邮箱设置自己的，并将wan230114@126.com加入收信人白名单）

天津：

#################### 导 入 工 具 ########################

**mymail="xxxx@xxx.com"**

**source /ifs/TJPROJ3/Plant/chenjun/mytools/Tools\_bashrc.sh**

###########################################################

南京：

#################### 导 入 工 具 ########################

**mymail="xxxx@xxx.com"**

**source /NJPROJ2/Plant/chenjun/mytools/Tools\_bashrc.sh**

###########################################################

# 集群进程及任务管理脚本分享

## ks：进程管理器

* 简介

自动找出给定关键词的子父进程，待用户确认后一键快捷杀复杂进程。

* 应用场景：

1、当父进程被kill时，子进程仍然运行，该进程管理器可找出子父进程一起杀掉。

2、当多个主进程运行多个子进程时，每个主进程包含有grep的关键词，而子进程则无，想全部终止包含这个关键词的主进程和子进程时用普通方法grep不好管理，该进程管理器也可以轻松解决该问题。

（如注释流程提交后，会在本地目录运行特别多的子进程，如果一开始配置文件出错，想全部杀掉重新开始会是一件很麻烦的事，使用本命令ks和下一个qd命令结合会使操作快捷很多倍）

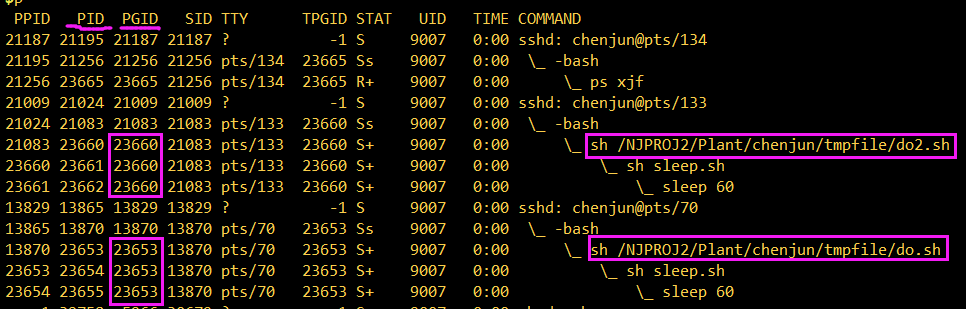
* 使用方法

ks keyword

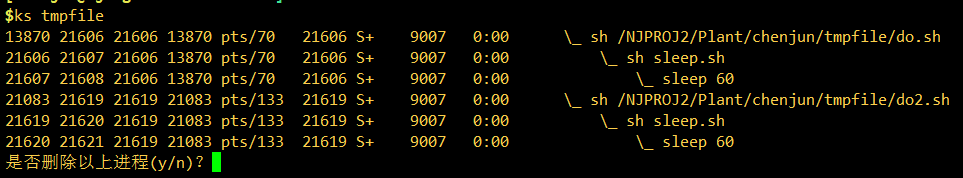
* 实例：

想终止/NJPROJ2/Plant/chenjun/tmpfile目录下运行的两个进程

1. 通过命令`ps xjf`查看进程情况，



1. 通过命令`ks tmpfile`找出欲杀掉的子父进程，并确认是否杀掉



## qd：投递任务管理器

* 应用场景：

每次批量杀任务都是用这样的命令：

vjob|grep P101SC16120962|cut -d ' ' -f 1|xargs qdel

qd封装了以上命令：

直接

qd P101SC16120962

即可完成相同操作

* 简介

给定删除任务关键词，可以批量qdel掉任务

* 使用方法

qd keyword

其中，keyword为vjob后能grep到的行的关键词

## qjk：集群job监控工具

* 简介

监控投递任务中能查找到的关键词，当关键词无法搜索到时即结束监控发送邮件。

在监控的过程中会生成日志文件，可用于监控资源使用情况。

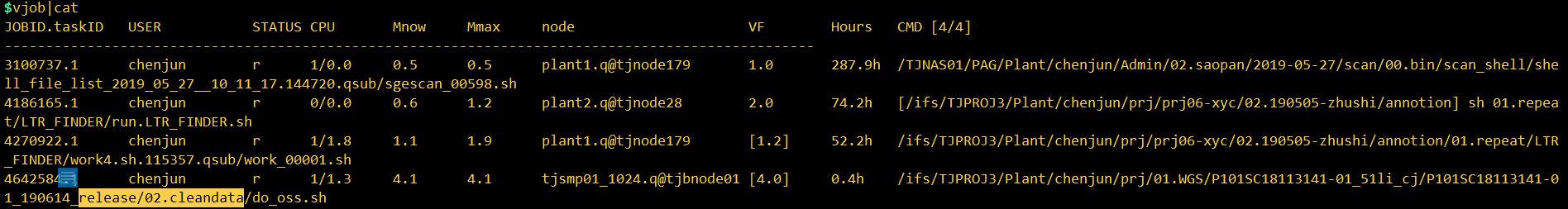
* 使用方法：

qjk keyword # keyword是指vjob里面能用关键词（keyword）grep到的行

* 示例：

1. 首先使用vjob命令查看投递的job，取出在里面能grep出的唯一关键词：

/home/leiyang/local/bin/vjob



1. 然后运行命令开始监控：

qjk release/02.cleandata



# 数据操作

## oss：集群数据互传工具

* 开发背景：

原始使用命令较为复杂：

天津查看同步到南京上传的：

ossutil ls oss://novo-tj-upload/novo-plant/chenjun/ -e oss-cn-beijing-internal.aliyuncs.com -i LTAIqS6yZi9YY9IQ -k gthbZClPZPQSa8S6b53m5X62edTiyH

天津查看从南京云同步过来的：

ossutil ls oss://novo-tj-down/novo-plant/chenjun/ -e oss-cn-beijing-internal.aliyuncs.com -i LTAIqS6yZi9YY9IQ -k gthbZClPZPQSa8S6b53m5X62edTiyH

南京查看从天津同步过来的：

ossutil ls oss://novo-nj-down/novo-plant/chenjun/ -e oss-cn-hangzhou-internal.aliyuncs.com -i LTAIqS6yZi9YY9IQ -k gthbZClPZPQSa8S6b53m5X62edTiyH

南京查看同步至天津的：

ossutil ls oss://novo-nj-upload/novo-plant/chenjun/ -e oss-cn-hangzhou-internal.aliyuncs.com -i LTAIqS6yZi9YY9IQ -k gthbZClPZPQSa8S6b53m5X62edTiyH

* 简介

无需配置脚本，应用南京和天津端的脚本自动配置好使用命令。

可将命令写入文件。

* 使用方法

（命令助记：命令中f含义是file，u含义是上传，d含义是下载）

oss fu file1 path/

oss fd path/ path2/

oss lsu [file1]

oss lsd [file1]

oss rmu [file1]

oss rmd [file1]

* 示例：

示例: 南京u-->天津d

1) 上传：将文件(夹)上传至南京云，南京云会自动同步到天津云

南京客户端: oss fu file(或dir/) dict1/

2) 查看：查看天津云（两者应该相同，否则请等待后台传输完毕(数据量大有延时)）

南京客户端: oss lsu

天津客户端: oss lsd

3) 下载：从天津云下载

1.下载所有文件到当前文件夹

天津客户端: oss fd

2.或下载指定文件(夹)到当前文件夹

天津客户端: oss fd dict1/ ./

4) 删除：清理云端数据（转移完毕去上传端删除数据）

1.删除所有的上传数据

南京客户端: oss rmu

2.或删除指定文件(夹)

南京客户端: oss rmu dict1/

（一般来说`oss rmd`不会用到，原则是在上传端去删除，这样删除操作才会同步至另一个中心，若在下载端rmd则会需要在上传端去再删除1次）

## md5/md5q：在当前目录计算所有文件md5和大小

* 简介

md5本地计算当前目录md5和checksize

md5q自动投递集群计算当前目录md5和checksize

(文件夹在2G以下时可使用md5, 2G以上尽量使用md5q，可以du -hLs查看文件夹大小)

* 使用方法：

md5

md5 num # num为数字，默认为2，是指使用多少进程计算

md5q

# 其他命令

send 发送邮件，具体用法见`send -h`

view 可视化文件夹，打包指定格式为html(svg,png,pdf)，具体用法见`view -h`

rep 给定一个替换list，对某一列批量替换，具体用法见`rep -h`

fas 快速统计碱基序列，目前详情统计模式算法待优化，具体用法见`fas -h`

ccut 补全cut不能去除末尾倒数多少列的问题，目前功能有限尚待优化

ca 带颜色输出，用法同cat

cah 将脚本分行处理输出

f 返回文件的绝对路径，用法同 `pwd`/file

...

# 下期分享

* 根据相同的某列合并两个文件的脚本（目前应用于集群管理下机单查找）