071 使用crontab定时任务自动执行爬虫程序

* 核心技术

crontab命令被用来提交和管理用户的需要周期性执行的任务，与windows下的计划任务类似，当安装完成操作系统后，默认会安装此服务工具，并且会自动启动crond进程，crond进程每分钟会定期检查是否有要执行的任务，如果有要执行的任务，则自动执行该任务。

（1）crontab语法格式。crontab语法格式如下：

crontab(选项)(参数)

下面分别介绍一下语法中的选项和参数。

crontab有如下选项：

* -e：编辑该用户的计时器设置；
* -l：列出该用户的计时器设置；
* -r：删除该用户的计时器设置；
* -u<用户名称>：指定要设定计时器的用户名称。

crontab的参数，是一个crontab文件，指定包含待执行任务的crontab文件。

（2）crontab文件含义。用户所建立的crontab文件中，每一行都代表一项任务，每行的每个字段代表一项设置，它的格式共分为六个字段，前五段是时间设定段，第六段是要执行的命令段，格式如下：

minute hour day month week [command](http://man.linuxde.net/command) 顺序：分 时 日 月 周

其中含义如下：

* minute： 表示分钟，可以是从0到59之间的任何整数。
* hour：表示小时，可以是从0到23之间的任何整数。
* day：表示日期，可以是从1到31之间的任何整数。
* month：表示月份，可以是从1到12之间的任何整数。
* week：表示星期几，可以是从0到7之间的任何整数，这里的0或7代表星期日。
* command：要执行的命令，可以是系统命令，也可以是自己编写的脚本文件。

在以上各个字段中，还可以使用以下特殊字符：

* 星号（\*）：代表所有可能的值，例如month字段如果是星号，则表示在满足其它字段的制约条件后每月都执行该命令操作。
* 逗号（,）：可以用逗号隔开的值指定一个列表范围，例如，“1,2,5,7,8,9”
* 中杠（-）：可以用整数之间的中杠表示一个整数范围，例如“2-6”表示“2,3,4,5,6”
* 正斜线（/）：可以用正斜线指定时间的间隔频率，例如“0-23/2”表示每两小时执行一次。同时正斜线可以和星号一起使用，例如\*/10，如果用在minute字段，表示每十分钟执行一次。

（3）定时任务实例。

每1分钟执行一次command，示例如下：

\* \* \* \* \* command

每小时的第3和第15分钟执行，示例如下：

3,15 \* \* \* \* command

在上午8点到11点的第3和第15分钟执行，示例如下：

3,15 8-11 \* \* \* command

每隔两天的上午8点到11点的第3和第15分钟执行，示例如下：

3,15 8-11 \*/2 \* \* command

每个星期一的上午8点到11点的第3和第15分钟执行，示例如下：

3,15 8-11 \* \* 1 command

每晚的21:30重启smb ，示例如下：

30 21 \* \* \* /etc/[init](http://man.linuxde.net/init" \t "_blank" \o "init命令).d/smb restart

每月1、10、22日的4:45重启smb，示例如下：

45 4 1,10,22 \* \* /etc/init.d/smb restart

每周六、周日的1:10重启smb，示例如下：

10 1 \* \* 6,0 /etc/init.d/smb restart

每天18:00至23:00之间每隔30分钟重启smb，示例如下：

0,30 18-23 \* \* \* /etc/init.d/smb restart

每星期六的晚上11:00 pm重启smb，示例如下：

0 23 \* \* 6 /etc/init.d/smb restart

每一小时重启smb，示例如下：

\* \*/1 \* \* \* /etc/init.d/smb restart

晚上11点到早上7点之间，每隔一小时重启smb，示例如下：

\* 23-7/1 \* \* \* /etc/init.d/smb restart

每月的4号与每周一到周三的11点重启smb，示例如下：

0 11 4 \* mon-wed /etc/init.d/smb restart

一月一号的4点重启smb，示例如下：

0 4 1 jan \* /etc/init.d/smb restart

每小时执行/etc/cron.hourly目录内的脚本，示例如下：

01 \* \* \* \* root run-parts /etc/cron.hourly