

**软件工程综合课程设计**

--开发手册

**题 目** 政府采购网中标数据采集工具

**院 系**  计算机科学与技术学院

**专 业**  软件工程

**学生姓名**温志豪,高正东,柳家胜,屈有昊,曾君宇

**指导教师** 张德平

**二**〇**一五 年 十二 月 三十 日**

目录

[1. Python 2.7 1](#_Toc439541483)

[1.1. Python下载 1](#_Toc439541484)

[1.2. Python解压 1](#_Toc439541485)

[1.3. Python编译 1](#_Toc439541486)

[1.4. Python安装 2](#_Toc439541487)

[2. Scrapy 1.0.3 4](#_Toc439541488)

[2.1. 依赖软件 4](#_Toc439541489)

[2.2. 安装 4](#_Toc439541490)

[2.3. 验证 5](#_Toc439541491)

[3. Scrapyd 5](#_Toc439541492)

[3.1. 依赖软件 5](#_Toc439541493)

[3.2. 安装 5](#_Toc439541494)

[3.3. 验证 6](#_Toc439541495)

[4. Scrapyd-client 7](#_Toc439541496)

[4.1. 下载 7](#_Toc439541497)

[4.2. 解压 7](#_Toc439541498)

[4.3. 安装 7](#_Toc439541499)

[4.4. 配置 7](#_Toc439541500)

[5. Supervisord 8](#_Toc439541501)

[5.1. 安装 8](#_Toc439541502)

[5.2. 配置 9](#_Toc439541503)

[5.3. 启动 9](#_Toc439541504)

[5.4. 管理进程 9](#_Toc439541505)

[6. Mysql 10](#_Toc439541506)

[6.1. 安装 10](#_Toc439541507)

[6.2. 配置 12](#_Toc439541508)

[6.3. 管理软件 12](#_Toc439541509)

[7. Jdk 1.7 12](#_Toc439541510)

[7.1. 下载 12](#_Toc439541511)

[7.2. 解压 13](#_Toc439541512)

[7.3. 配置 13](#_Toc439541513)

[7.4. 验证 14](#_Toc439541514)

[8. Tomcat 7.0 14](#_Toc439541515)

[8.1. 下载 14](#_Toc439541516)

[8.2. 解压 14](#_Toc439541517)

[8.3. 安装 14](#_Toc439541518)

[8.4. 配置 14](#_Toc439541519)

[8.5. 验证 15](#_Toc439541520)

# Python 2.7

## Python下载

打开ubuntu下的shell终端，通过wget命令下载python源码包，如下图所示:



将python-2.7.3.tgz下载至/opt目录下。

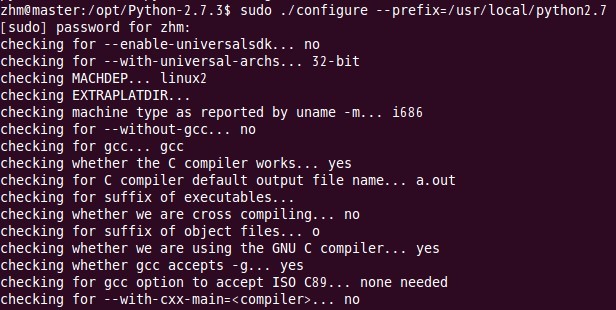
## Python解压

http://img.my.csdn.net/uploads/201303/16/1363433981_3703.jpg

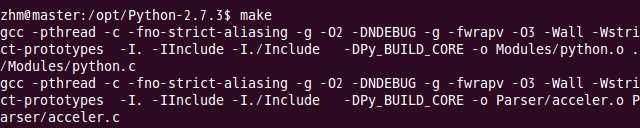
## Python编译

在对python进行编译之前，必须对它进行配置。在unix/linux平台上的安装过程中，配置和编译过程全部已经自动化了，我们所需要做的就是输入几条命令然后回车即可。

  通过执行脚本./configure开始编译， 这一过程可以决定系统的配置并创建相应的Makefile文件。你也可以传递一些参数选项来配置这个脚本。当这一步完成后，就可以运行make了。操作如下图所示：



通过添加--prefix用于指定python的安装路径于/usr/local/python2.7,配置完成后，我们就可以执行make操作了.

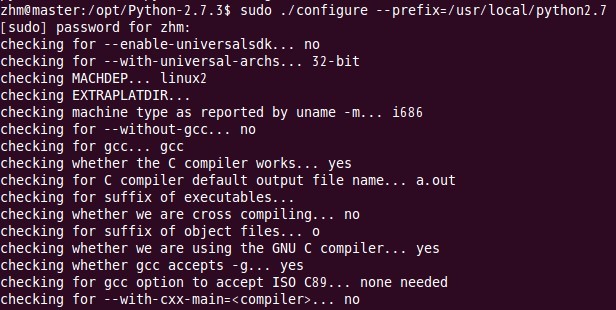


上面make编译的时间可能会比较长，大概5-6分钟，完成之后我们就可以进行安排操作了，输入指令make install(注意权限)， 如下:

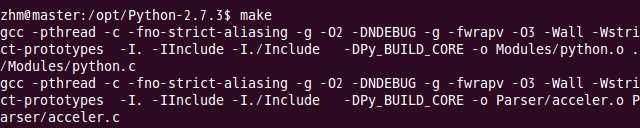
http://img.my.csdn.net/uploads/201303/18/1363604024_1697.jpg

## Python安装

对python进行编译之前，必须对它进行配置。在unix/linux平台上的安装过程中，配置和编译过程全部已经自动化了，所需要做的就是输入几条命令然后回车即可。通过执行脚本./configure开始编译， 这一过程可以决定系统的配置并创建相应的Makefile文件。你也可以传递一些参数选项来配置这个脚本。当这一步完成后，可以运行make。操作如下图所示：



通过添加--prefix用于指定python的安装路径于/usr/local/python2.7,配置完成后，就可以执行make操作了.



make编译的时间可能会比较长，大概5-6分钟，完成之后可以进行安排操作，输入指令make install(注意权限)， 如下:

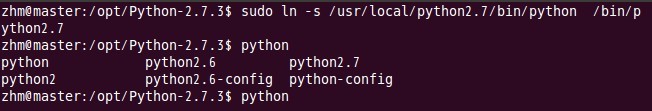
http://img.my.csdn.net/uploads/201303/18/1363604024_1697.jpg

安装后的python目录：

http://img.my.csdn.net/uploads/201303/18/1363604106_4807.jpg

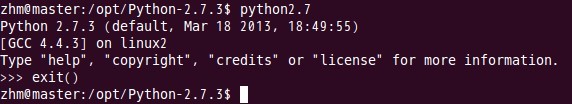
  上面在./configure的时候增加python的安装目录，所以用ls命令查看此目录的信息。其中bin目录下用于存放python的相关执行程序。

但是，由于手动将python安装于/usr/local/python2.7目录下，此时$PATH环境变量无法找到此目录下的python解释器，为此可以增加一个软链接，代码如下：



通过ln -s /usr/local/python2.7/bin/python /bin/python2.7命令可以在/bin路径下创建一软链接文件python2.7，当访问此链接文件，就可以间接地访问/usr/local/python2.7下的python程序了。这种方式与windows下的快捷方式类似。

下面可以通过运行python命令交互模式以查看刚刚安装的python版本情况：



如上所示，通过直接执行python2.7可以访问/usr/local/python2.7下的python解释器，其版本为Python 2.7.3，这个正是python版本

D

# Scrapy 1.0.3

## 依赖软件

* Python 2.7
* Python Package: pip and setuptools
* Lxml
* OpenSSL

## 安装

* 把Scrapy签名的GPG密钥添加到APT的钥匙环中:

sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 627220E7

* 执行如下命令，创建 /etc/apt/sources.list.d/scrapy.list 文件:

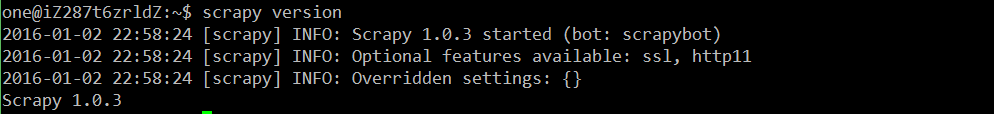
echo 'deb <http://archive.scrapy.org/ubuntu> scrapy main' | sudo tee /etc/apt/sources.list.d/scrapy.list

* 更新包列表并安装 scrapy:

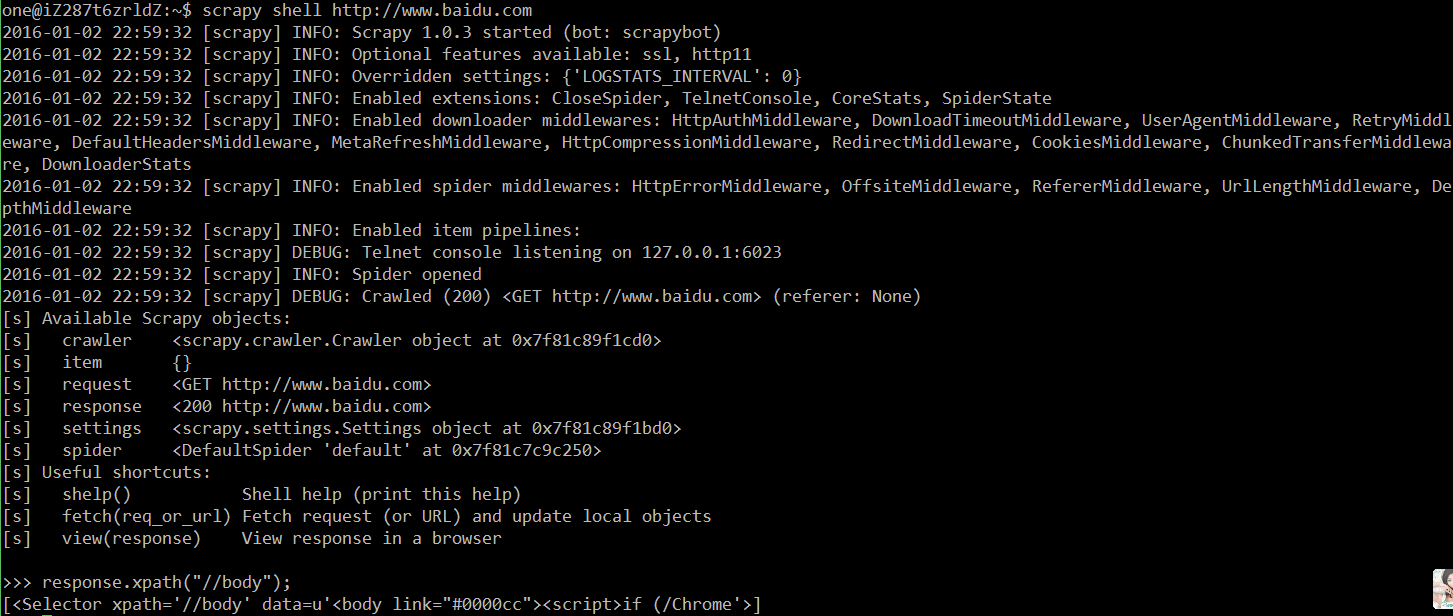
sudo apt-get update && sudo apt-get install scrapy-1.0.3

## 验证

* 运行scrapy命令，看是否出现版本号的信息



* 使用scrapy的Shell命令访问百度首页，看是否正常解析



证明安装成功而且能正常访问网页。

# Scrapyd

## 依赖软件

* Python 2.6 or above
* Twisted 8.0 or above
* Scrapy 0.17 or above

## 安装

安装scrapyd（一般方法）

pip install scrapyd

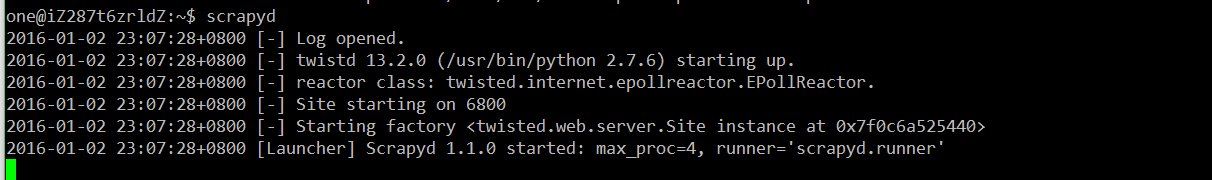
在Ubuntu上安装scrapyd

apt-get install scrapyd

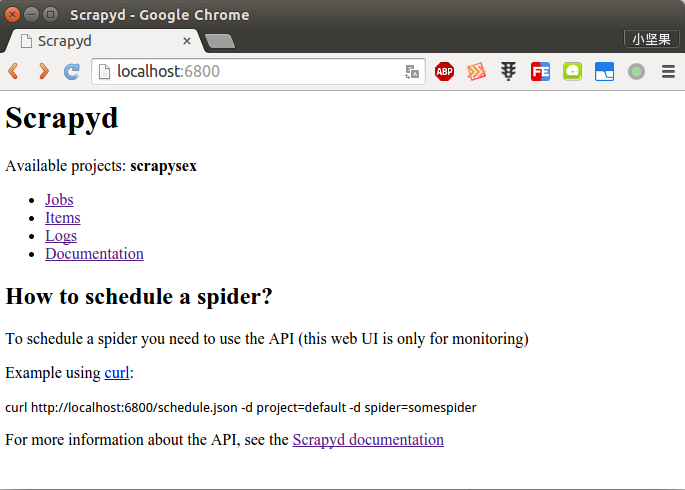
## 验证

运行命令如下

scrapyd



默认情况下scrapyd 监听 0.0.0.0:6800 端口，所有运行scrapyd 后在浏览器http://localhost:6800/ 即可查看到当前可以运行的项目：



# Scrapyd-client

## 下载

下载scrapyd-client

wget <https://github.com/scrapy/scrapyd-client/archive/master.zip>

## 解压

解压scrapy-client

tar –zxvf scrapyd-clinet-master

## 安装

进入scrapy-client目录

cd scrapyd-client-master

安装 scrapyd-client

python setup.py install

## 配置

部署scrapy 项目

修改工程目录下的 scrapy.cfg 文件

[deploy:scrapyd2] #默认情况下并没有scrapyd2，它只是一个名字，可以在配置文件中写多个名字不同的deploy

url = http://scrapyd.mydomain.com/api/scrapyd/ #要部署项目的服务器的地址

username = john #访问服务器所需的用户名和密码（如果不需要密码可以不写）

password = secret

其中的username 和 password 用于在部署时验证服务器的HTTP basic authentication，须要注意的是这里的用户密码并不表示访问该项目须要验证，而是登录服务器用的。如何对scrapyd 的服务器设置密码验证将在后文提到。

我们可以通过scrapyd-deploy -l 查看当前目录下的可以使用的部署方式（target）

scrapyd http://localhost:6800/ #这里显示两个target 即scrapy.cfg 中有两个deploy，名字分别是scrapyd 和scrapyd2

scrapyd2 <http://scrapyd.mydomain.com/api/scrapyd/>

scrapyd-deploy -L scrapyd 可以查看某一target 下的项目。

使用

scrapyd-deploy scrapyd2 -p project1 #scrapyd-deploy 是命令名，scrapyd2 是配置文件中的target 名

部署项目到服务器，如果成功将会看到如下的返回信息：

Deploying myproject-1287453519 to http://localhost:6800/addversion.json

Server response (200):

{"status": "ok", "spiders": ["spider1", "spider2"]}

爬虫的版本控制

默认情况下scrapyd-deploy 使用当前时间戳作为项目的版本号，服务器将运行版本号最新的项目。可以手动设置版本号：

scrapyd-deploy scrapyd -p project1 --version 54

或者使用git 等版本控制工具时：

scrapyd-deploy scrapyd -p project1 --version GIT

这时我们也可以在项目的设置文件中加上 version = GIT

[deploy]

url = http://scrapyd.mydomain.com/api/scrapyd/

username = john

password = secret

project = project1

version = GIT

由于设置中已经指明了project 且只有一个target（默认的target 名为default），部署的时候只需要以下命令即可。

scrapyd-deploy

使用JSON web service 控制爬虫

开启爬虫 schedule

curl http://localhost:6800/schedule.json -d project=myproject -d spider=somespider

停止 cancel

Curl <http://localhost:6800/cancel.json> -d project=myproject -d job=6487ec79947edab326d6db28a2d86511e8247444

其他更多API 可以参考官方文档 http://scrapyd.readthedocs.org/en/latest/api.html

# Supervisord

## 安装

安装supervisor

sudo pip install supervisor

## 配置

安装完成之后，就可以用"echo\_supervisord\_conf"命令来生成配置文件，例如

echo\_supervisord\_conf > /etc/supervisord.conf

echo\_supervisord\_conf > /path/to/supervisord.conf

## 启动

启动supervisord

supervisord #默认使用/etc/supervisord.conf的配置文件

supervisord -c /path/to/supervisord.conf

## 管理进程

通过supervisorctl就监控管理program

$ supervisorctl -c conf/app.conf status

node\_app RUNNING pid 6916, uptime 0:00:00

tornado\_app RUNNING pid 6917, uptime 0:00:00

$ supervisorctl -c conf/app.conf stop node\_app

node\_app: stopped

$ supervisorctl -c conf/app.conf stop tornado\_app

tornado\_app: stopped

$ supervisorctl -c conf/app.conf status

node\_app STOPPED Apr 04 02:34 AM

tornado\_app STOPPED Apr 04 02:35 AM

$ supervisorctl -c conf/app.conf start all

node\_app: started

tornado\_app: started

$ supervisorctl -c conf/app.conf status

node\_app RUNNING pid 8080, uptime 0:00:00

tornado\_app RUNNING pid 8079, uptime 0:00:00

通过web管理进程

$ grep -A 3 "inet\_http\_server" conf/app.conf

[inet\_http\_server] ; inet (TCP) server disabled by default

port=0.0.0.0:9001 ; (ip\_address:port specifier, \*:port for all iface)

#username=user ; (default is no username (open server))

#password=123 ; (default is no password (open server))

然后打开 http://localhost:9001 就可以访问了

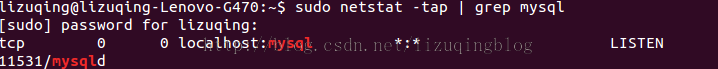
# Mysql

## 安装

1. 首先检查系统中是否已经安装了MySQL

在终端里面输入 sudo netstat -tap | grep mysql

若没有反映，没有显示已安装结果，则没有安装。若如下显示，则表示已经安装



1. 如果没有安装，则安装MySQL

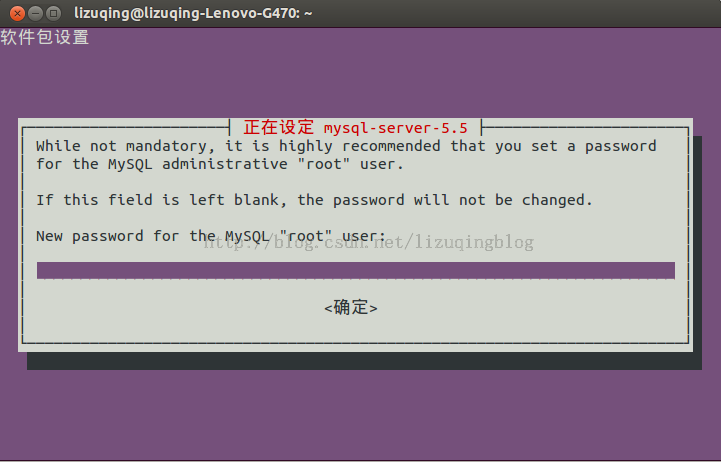
在终端输入

 sudo apt-get install mysql-server mysql-client

运行结果如下所示:



在此安装过程中会让要求输入root用户(管理MySQL数据库用户，非Linux系统用户)密码，按照要求输入即可。如下所示：

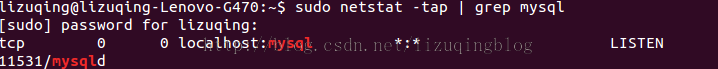


1. 测试安装是否成功：

在终端输入

sudo netstat -tap | grep mysql

出现如下结果则安装成功：

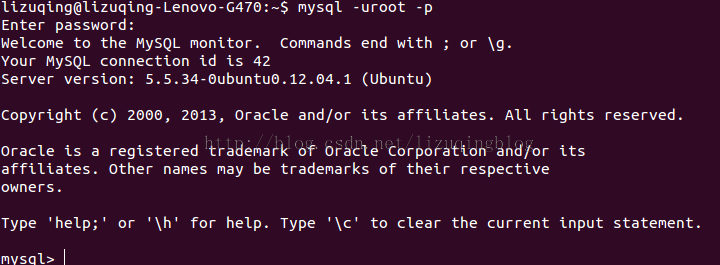


也可通过登录MySQL测试

在终端输入

mysql -uroot -p

接下来会提示输入密码，输入正确密码，即可进入。如下所示：



## 配置

启动MySQL服务：

sudo start mysql

停止MySQL服务：

sudo stop mysql

修改 MySQL 的管理员密码：

sudo mysqladmin -u root password newpassword

设置远程访问(正常情况下，mysql占用的3306端口只是在IP 127.0.0.1上监听，拒绝了其他IP的访问（通过netstat可以查看到）。取消本地监听需要修改 my.cnf 文件：)：

sudo vi /etc/mysql/my.cnf

bind-address = 127.0.0.1 //找到此内容并且注释

## 管理软件

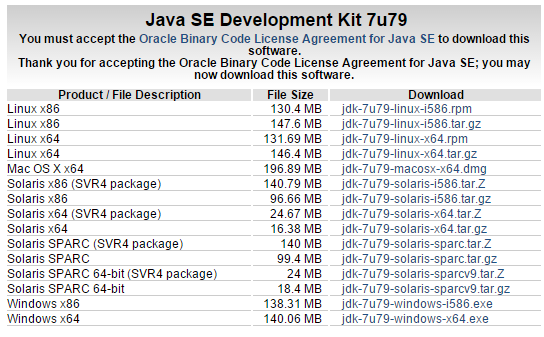
* 开源：MySQL Workbench
* 商业：Navicat

# Jdk 1.7

## 下载

下载jdk-7u45-linux-x64.tar.gz

  网址：http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html



## 解压

  进入JDK的下载目录



## 配置



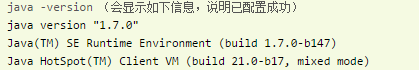
 打开profile文件输入



设置系统默认JDK



## 验证



# Tomcat 7.0

## 下载

获得tomcat的安装包



## 解压



## 安装



## 配置

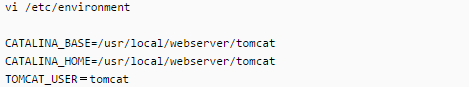
设置tomcat以独立的用户运行



设置用户tomcat对tomcat目录的访问权限

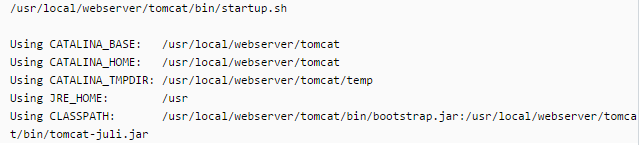


配置Tomcat环境变量



## 验证

启动tomcat



测试

访问http://localhost:8080/，如果看到Tomcat缺省界面就表示成功了。

