搭建PySpider爬虫服务

****1. 环境准备****

*前置环境部署*

在开始部署前，我们需要做一些前置准备 。

yum install epel-release

yum 更新

yum update -y

安装开发编译工具

yum install gcc gcc-c++ -y

安装依赖库

yum install python-pip python-devel python-distribute libxml2 libxml2-devel python-lxml libxslt libxslt-devel openssl openssl-devel -y

升级pip

pip install --upgrade pip

该步骤可选，但为了部署的稳定性，推荐执行

该步骤耗时可能较长（5~10min），请耐心等待

****2. 部署 mariadb****

由于 CentOS 7 中 MySQL 数据库已从默认的程序列表中移除，我们使用 mariadb 代替。

*安装 mariadb*

yum install mariadb-server mariadb -y

*启动 mariadb 服务*

*systemctl start mariadb*

*设置 root 密码*

默认的root用户密码为空，你可以使用以下命令来创建 root 用户的密码

*mysqladmin -u root password "123456"*

*检查是否安装成功*

现在你可以尝试通过以下命令来连接到 Mysql 服务器

*mysql -u root -p*

然后输入您刚才设置的密码 （ 默认：123456 ），如果一切正常，您应该可以在命令行看到以 MariaDB [(none)]> 或 mysql> 开头的提示了，说明连接成功。

此时输入 SHOW DATABASES; 并回车，应该可以看到类似下面这样的输出，说明一切正常。

MariaDB [(none)]> show databases;

+--------------------+

| Database |

+--------------------+

| information\_schema |

| mysql |

| performance\_schema |

| test |

+--------------------+

4 rows in set (0.00 sec)

完成后，可以通过快捷键 Ctrl+C 或命令行键入 exit 来退出，进入下一步。

****3. 部署 redis****

*下载、解压安装包*

下载安装包

wget <http://download.redis.io/redis-stable.tar.gz>

解压安装包

tar -xzvf redis-stable.tar.gz

移动解压包到 /usr/local 内

mv redis-stable /usr/local/redis

编译安装

cd /usr/local/redis

make

make install

*设置 redis 配置*

设置配置文件路径

mkdir -p /etc/redis

cp /usr/local/redis/redis.conf /etc/redis/redis.conf

修改 /etc/redis/redis.conf 文件的 daemonize 配置项为如下：

daemonize yes

*启动 redis 服务*

*/usr/local/bin/redis-server /etc/redis/redis.conf*

****4. 部署 pyspider****

*安装依赖*

*pip install --upgrade chardet*

*easy\_install mysql-connector==2.1.3*

*easy\_install redis*

*安装 pyspider*

*pip install pyspider*

然后 /etc/pyspider 目录下创建 pyspider.conf.json，参考下面的内容。

具体配置的说明文档请参考 [官方文档](http://docs.pyspider.org/en/latest/Deployment/" \l "configjson" \t "https://blog.csdn.net/qq_26954773/article/details/_blank)

{

"taskdb": "mysql+taskdb://root:123456@127.0.0.1:3306/taskdb",

"projectdb": "mysql+projectdb://root:123456@127.0.0.1:3306/projectdb",

"resultdb": "mysql+resultdb://root:123456@127.0.0.1:3306/resultdb",

"message\_queue": "redis://127.0.0.1:6379/db",

"webui": {

"username": "root",

"password": "123456",

"need-auth": true

}

}

其中 mysql 配置中的 root 为您 mysql 的用户名， root: 后面的 123456 为您刚设置的密码。

webui 配置中的 username 及 password 为您访问 WebUI 时候需要的用户名，你也可以不设置用户名密码，直接将 need-auth 设为 false 即可。

*启动服务*

*pyspider -c /etc/pyspider/pyspider.conf.json*

如果一切正常，现在访问 [http://yourip:5000](http://yourip:5000/" \t "https://blog.csdn.net/qq_26954773/article/details/_blank) 或 [http://yourdomain.com:5000](http://yourdomain.com:5000/" \t "https://blog.csdn.net/qq_26954773/article/details/_blank)，应该可以看到 pyspider dashboard 的首页了。

服务能够正常启动后，我们需要让它能够在后台运行，可以通过以下命令让服务在后台运行

*nohup pyspider -c /etc/pyspider/pyspider.conf.json &*

****5 部署完成****