目录

[1.whois本地化系统介绍 1](#_Toc534711878)

[2.1部署 1](#_Toc534711879)

[2.2运行方式 1](#_Toc534711880)

[2.3查询方式 1](#_Toc534711881)

[2.自动化批量处理 1](#_Toc534711882)

[3.1部署 1](#_Toc534711883)

[3.2运行 2](#_Toc534711884)

[3.3查询方式 2](#_Toc534711885)

[3.4数据库 2](#_Toc534711886)

[3.网络爬虫 2](#_Toc534711887)

[4.1部署 2](#_Toc534711888)

[4.2运行 2](#_Toc534711889)

[4.3查询方式 2](#_Toc534711890)

[5.mongodb数据库 2](#_Toc534711891)

[5.1部署 2](#_Toc534711892)

[6.web可视化及查询接口 3](#_Toc534711893)

[6.1web可视化 3](#_Toc534711894)

[6.2查询接口 3](#_Toc534711895)

1.whois本地化系统介绍

whois数据来源为全世界5大主要的RIR。

[Apnic](https://www.apnic.net/manage-ip/using-whois/guide/) [Arin](https://www.arin.net/resources/services/whois_guide.html) [Afrinic](https://www.afrinic.net/support)  [Lacnic](https://www.lacnic.net/en/web/lacnic/documentos-tecnicos-whois) [Ripe](https://www.ripe.net/manage-ips-and-asns/db/support/documentation/)

在Linux中，通过whois 命令可以从相应的组织中查询whois信息。由于这5个组织数据可获取的来源并不一致，因此whois本地化系统分为三大部分：分别是：

1. Whois镜像服务器
2. 自动化批量下载程序
3. 网络爬虫

接下来，分别介绍这三个部分的部署方式，使用方式，以及数据查询方式。

2.whois镜像服务

Whois镜像将RIPE以及AFRINIC两大组织的whois数据同步到本地。

2.1部署

目前服务运行于10.10.11.130，目录/home/hitnis/Document/whois-whois-1.87。部署方式，参考了这篇[博客](https://www.freebuf.com/articles/network/107372.html)以及[官方文档](https://github.com/RIPE-NCC/whois/wiki)。部署过程费时费力，没有特别需要，无需重新部署。另，目录下的properties为重要配置文件，重新部署时，直接替换原始文件即可。

2.2运行方式

只需进入该目录，使用命令 ./whois.init start即可启动。遇到服务器宕机，使用此命令再次启动即可。日志文件位于var/console.log。

2.3查询方式

Whois镜像服务部署完毕后，会将RIPE以及AFRINIC的数据定期同步到maridb数据库中。既然它是一个服务，其开放查询端口由我们在2.1中所说的properties文件中第6行port.query=8888，因此我们本地使用命令 whois –h 10.10.11.130 -p 8888 202.118.236.100 进行本地查询

3.自动化批量处理

3.1部署

由于Whois镜像服务无法将5大RIR的数据全部获取，在数据获取中发现APNIC和ARIN的数据可以批量获取，即可以一次性下载其全部数据。因此可以进行自动批量化下载、处理、入库。

程序位于

/home/hitnis/Document/whois-whois-1.87/dump/auto\_download\_raw\_whois/auto\_download\_inetnum\_and\_insert\_db.py

使其保持运行状态即可，为防止ssh终端而强行被终止，可以使用screen –S 命令创建新窗口后再运行该自动化脚本。另外，ARIN的数据并不直接公开，在脚本中可以发现下载链接后有”apikey=”，此key为本人去[Arin](https://www.arin.net/resources/services/whois_guide.html)官网自行申请获得。

3.2运行

进入该目录后，直接python auto\_download\_inetnum\_and\_insert\_db.py即可运行，目前脚本中设置数据更新周期为1个月。如有需要，自行更改脚本。

脚本执行过程为：批量下载数据，解压缩，创建新的数据表，将数据插入新表，完成后删除旧表，并重命名新表。

3.3查询方式

从3.2中的脚步程序中可以看出，数据入库为mysql数据库。存储在名为whois\_inetnum的数据库中，数据表分别是apnic、arin。为了加快查询速度，查询逻辑采用前缀匹配，因此多了两个表：apnic\_cidr、arin\_cidr。

因此，可以通过数据库命令进行查询。

前缀匹配的逻辑为：

3.4数据库

Mysql数据库部署在10.10.11.130，使用命令mysql登陆，无需用户名密码。

4.网络爬虫

最后，只剩下lacnic的数据，是以上两种方式均无法一次性获取的，因此，通过多线程爬虫脚本进行自动化爬取入库。

4.1部署

自动化爬取工具位于

/home/hitnis/Document/get\_whoises/ prefect\_only\_lacnic\_V\_0\_1.py

另外该程序依赖同目录下ip\_del.h文件。简单的说，该程序读取ip列表，可以控制开启多个线程同时爬取whois信息，通过ip\_del.h文件过滤，保证只爬取lacnic的数据，处理后，更新到mongodb数据库。数据库名：ly,数据表名：lacnic

4.2运行

进入该目录后，直接python prefect\_only\_lacnic\_V\_0\_1.py即可运行,保持长时间运行需注意防止ssh通信中断而导致程序终止的情况。上面说过解决方案。

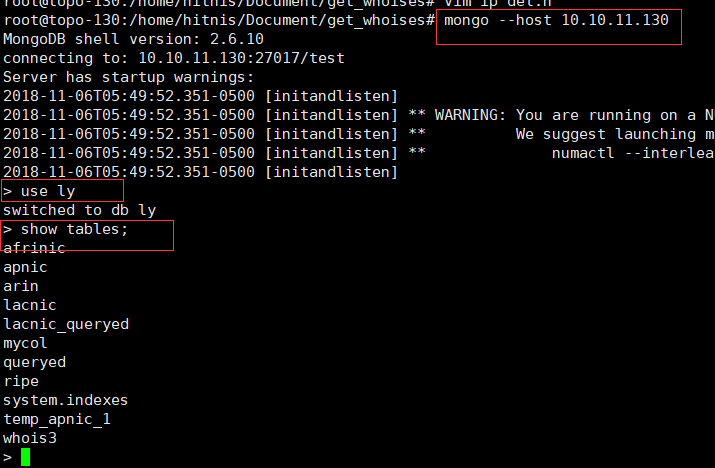
4.3查询方式

该数据也是存储的mongodb中，下一章介绍数据库登陆，查询方法。

5.mongodb数据库

5.1部署

Mongodb部署在10.10.11.130上，毕竟3，4两章的自动化脚本也在同一台机器。数据库登陆方式：mongo --host 10.10.11.130，详情见下方截图。需要注意的是只有lacnic数据保存在mongodb数据库中。



5.2迁移

如果数据库需要迁移到其他机器，请自行google，并注意更改3，4章节自动化脚本中数据库连接的主机IP。

6.web可视化及查询接口

6.1web可视化

Web程序使用基于PHP的Laravel框架，使用Lampp进行快速部署。Web源程序位于10.10.11.130机器的/var/www/html/test4c目录下。下面列出该目录下关键代码说明：

app/Http/Controllers/WhoisController.php #主要查询逻辑文件

routes/web.php #路由信息

app/\*.php #数据库连接模型

web访问地址为：

http://10.10.11.130/test4c/public/whois

6.2查询接口

虽然本地化数据有三大来源，但是统一的数据查询接口为：

<http://10.10.11.130/test4c/public/whois_api?ip=202.6.91.0>

其他：

Github相关源码：https://github.com/yiqiaoxihui/get\_whoises