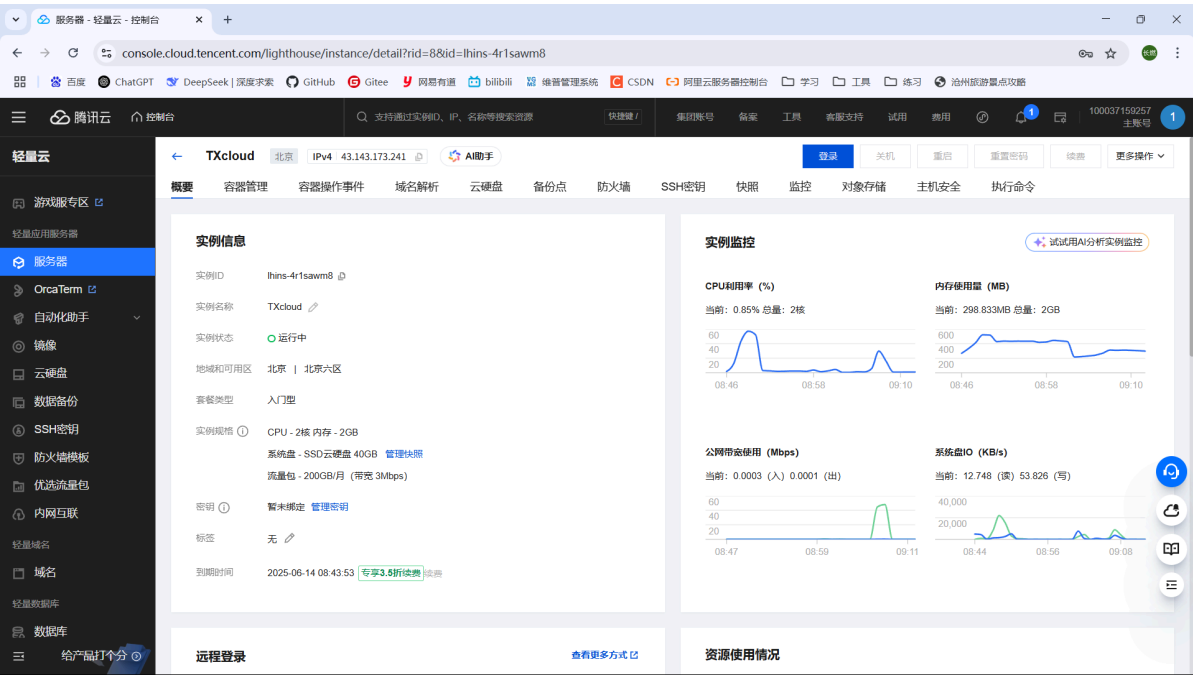
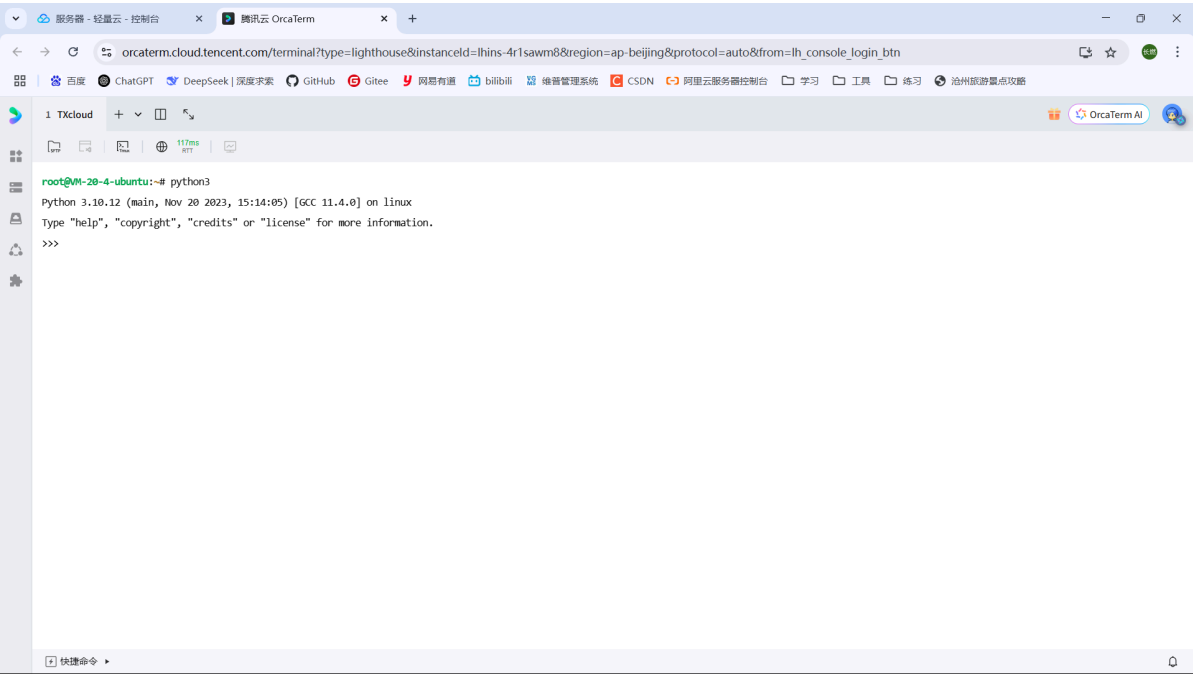


服务器部署分布式爬虫

1 服务器选择（腾讯云服务器）



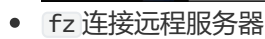
2 查看Python版本



3 fzf 连接远程服务器，上传MiniConda安装包并完成安装配置

Python版本为3.10.12，不知道能不能用，直接装 miniConda 最新版了

- 放行22端口



```
root@VM-28-4-ubuntu:~# telnet 43.143.173.241 21
Trying 43.143.173.241...
telnet: Unable to connect to remote host: Connection refused
root@VM-28-4-ubuntu:~# telnet 43.143.173.241 22
Trying 43.143.173.241...
Connected to 43.143.173.241.
Escape character is '^['.
SSH-2.0-OpenSSH_8.9p1 Ubuntu-3ubuntu0.6
Connection closed by foreign host.
root@VM-28-4-ubuntu:~# cd /home/ubuntu
root@VM-28-4-ubuntu:~/home/ubuntu# ls
Miniconda3-latest-Linux-x86_64.sh
root@VM-28-4-ubuntu:~/home/ubuntu# bash
.bash_history          .bashrc               Miniconda3-latest-Linux-x86_64.sh .profile               .ssh/
.bash_logout           .cache/               .pip/                  .pydistutils.cfg      .sudo_as_admin_successful
root@VM-28-4-ubuntu:~/home/ubuntu# bash Miniconda3-latest-Linux-x86_64.sh

Welcome to Miniconda3 py313_25.3.1-1

In order to continue the installation process, please review the license
agreement.
Please, press ENTER to continue
>>>
```

```
run the following command when conda is activated:

conda config --set auto_activate_base false

You can undo this by running `conda init --reverse $SHELL`? [yes/no]
[no] >>> yes
no change      /root/miniconda3/condabin/conda
no change      /root/miniconda3/bin/conda
no change      /root/miniconda3/bin/conda-erv
no change      /root/miniconda3/bin/activate
no change      /root/miniconda3/bin/deactivate
no change      /root/miniconda3/etc/profile.d/conda.sh
no change      /root/miniconda3/etc/fish/conf.d/conda.fish
no change      /root/miniconda3/shell/condabin/conda.ps1
no change      /root/miniconda3/shell/condabin/conda-hook.ps1
no change      /root/miniconda3/lib/python3.13/site-packages/xontrib/conda.xsh
no change      /root/miniconda3/etc/profile.d/conda.csh
modified       /root/.bashrc

==> For changes to take effect, close and re-open your current shell. <==

Thank you for installing Miniconda3!
root@VM-28-4-ubuntu:~/home/ubuntu#
```

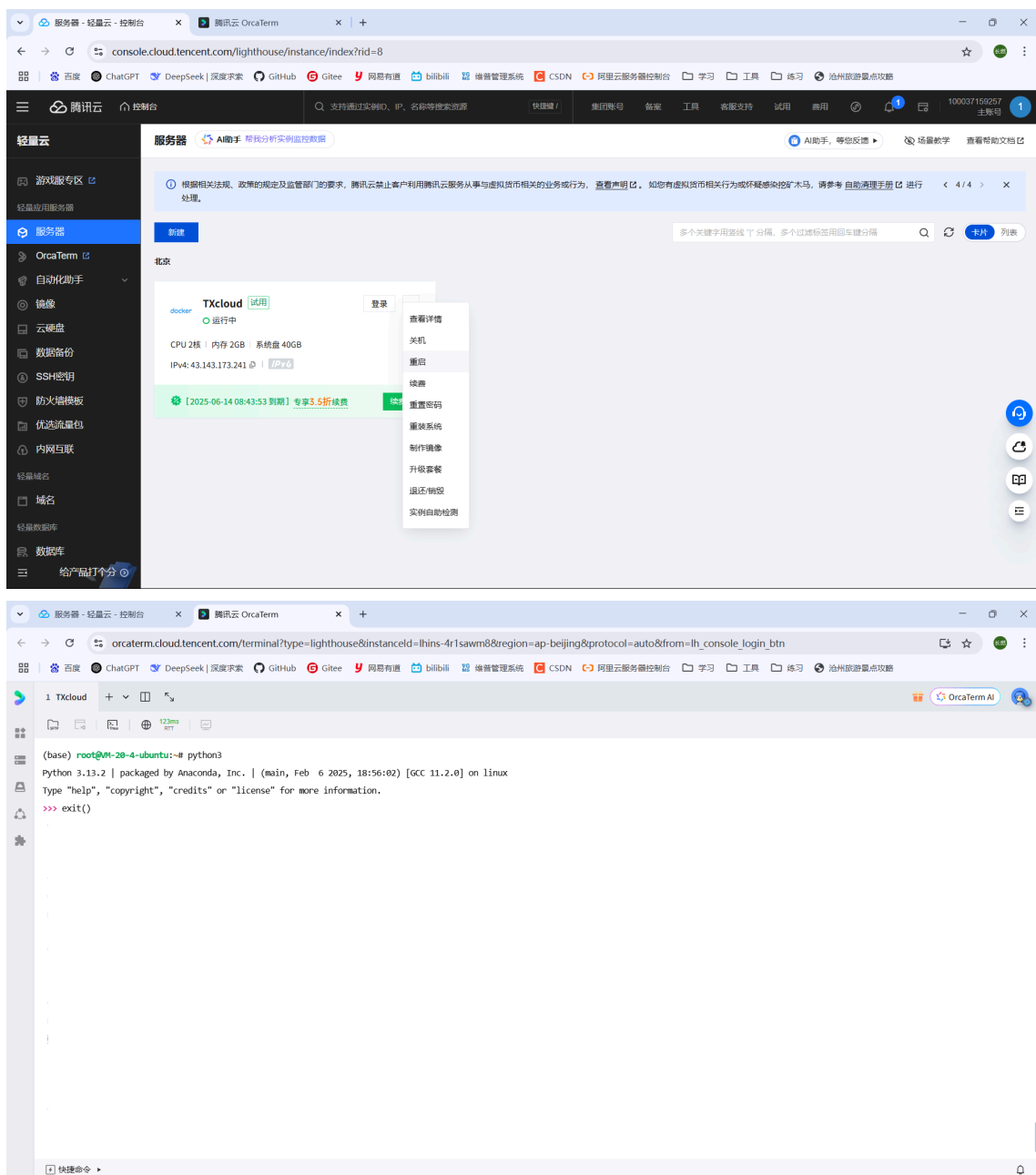
- 配置环境变量

```
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if [ -f /etc/bash_completion ] && ! shopt -oq posix; then
#   . /etc/bash_completion
#fi

#>>> conda initialize >>>
# !! Contents within this block are managed by 'conda init' !!
__conda_setup="$(/root/miniconda3/bin/conda 'shell.bash' 'hook' 2> /dev/null)"
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/root/miniconda3/etc/profile.d/conda.sh" ]; then
        . "/root/miniconda3/etc/profile.d/conda.sh"
    else
        export PATH="/root/miniconda3/bin:$PATH"
    fi
fi
unset __conda_setup
#<<< conda initialize <<<

:wq
```

- 重启服务器，验证安装配置



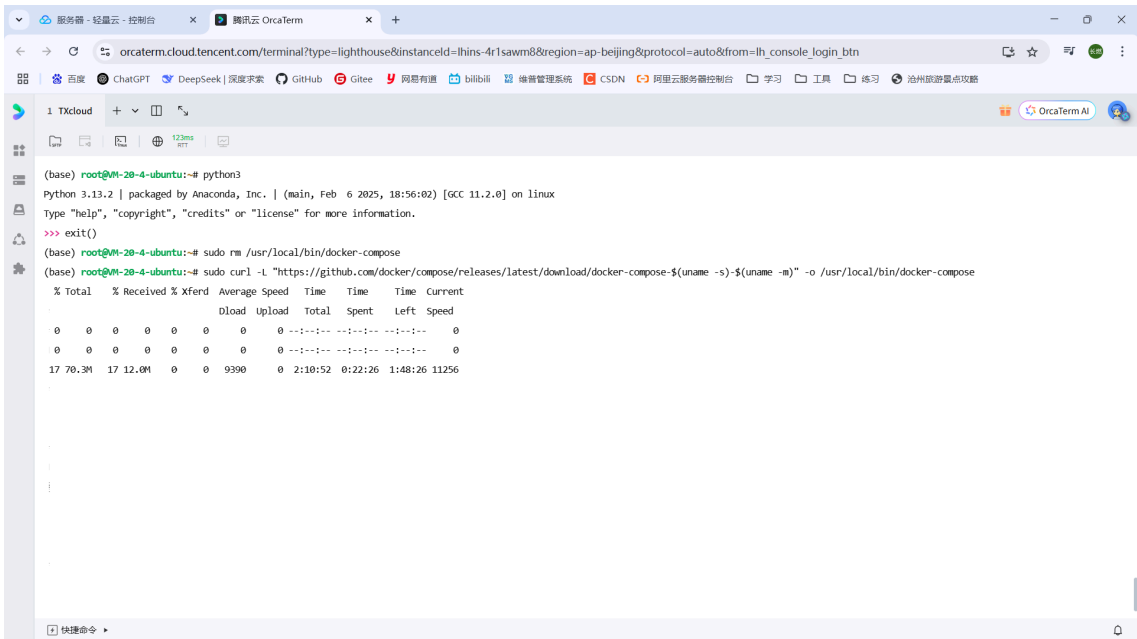
4 安装 docker-compose

- 清理之前的残余

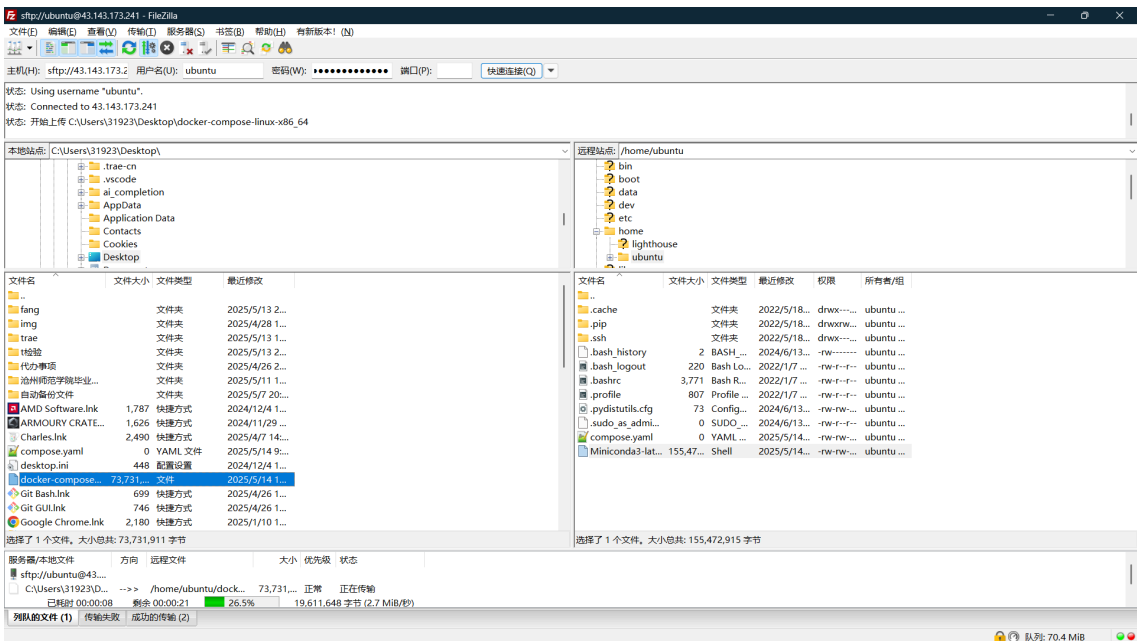
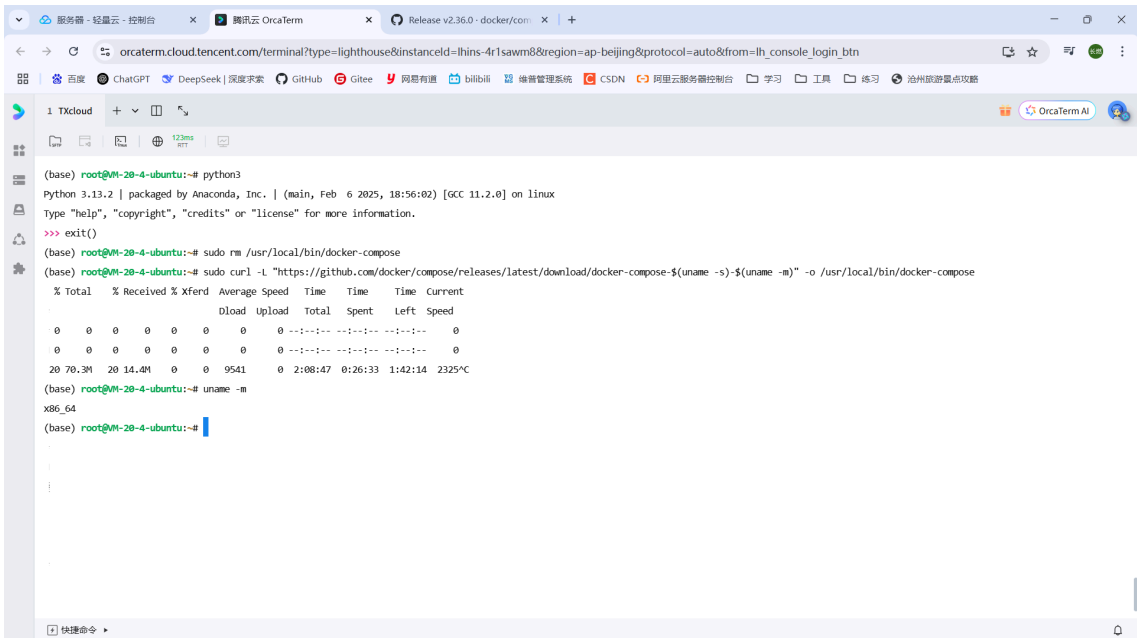
```
sudo rm /usr/local/bin/docker-compose
```

- 下载安装（由于下载太慢，这里我选择在本地windows上下载）

```
sudo curl -L
"https://github.com/docker/compose/releases/latest/download/docker-
compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose
```



- 访问<https://github.com/docker/compose/releases/latest>
- 查看远程服务器架构并下载对应版本

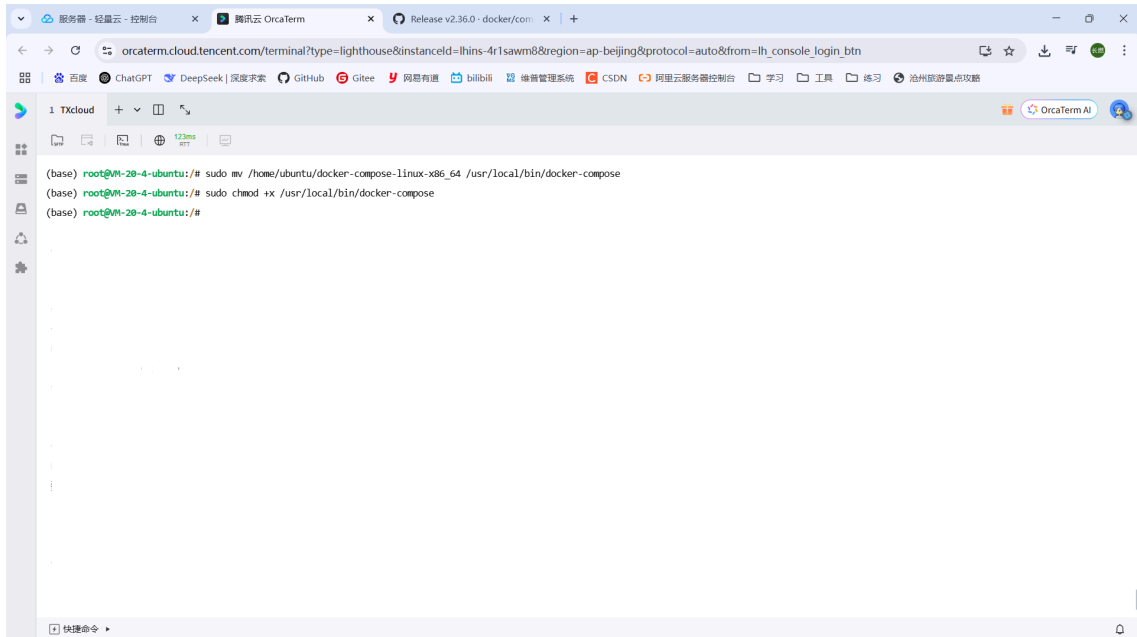


- 重命名文件

```
sudo mv /home/ubuntu/docker-compose-linux-x86_64 /usr/local/bin/docker-  
compose
```

- 添加权限

```
sudo chmod +x /usr/local/bin/docker-compose
```

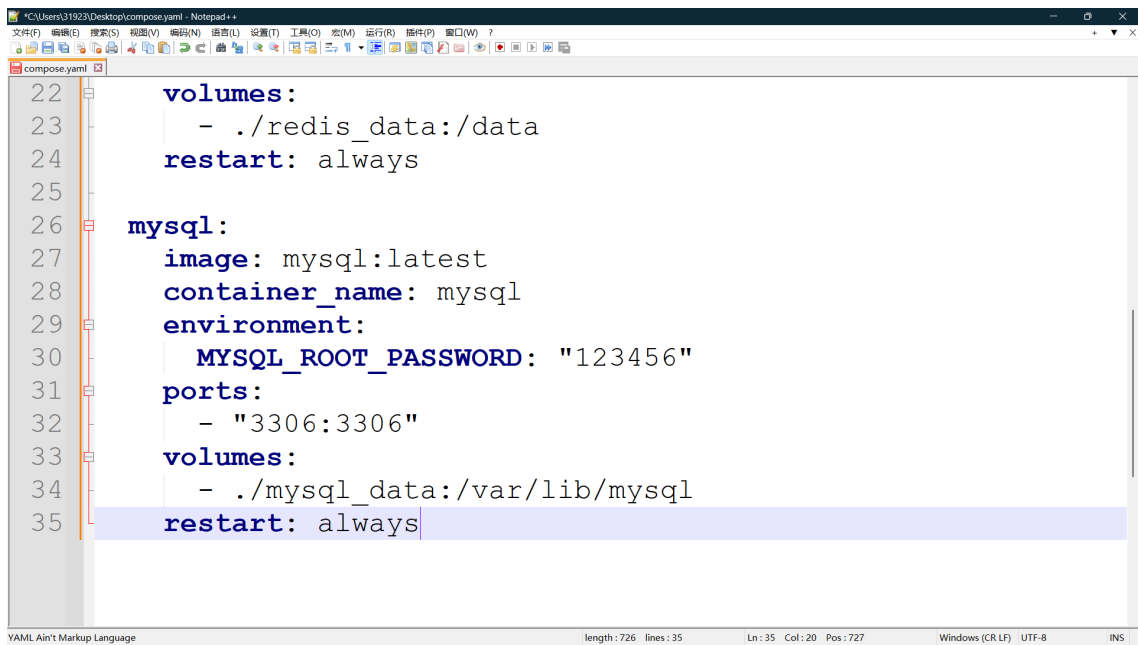


5 编写Compose 配置文件，上传到远程服务器上并启动

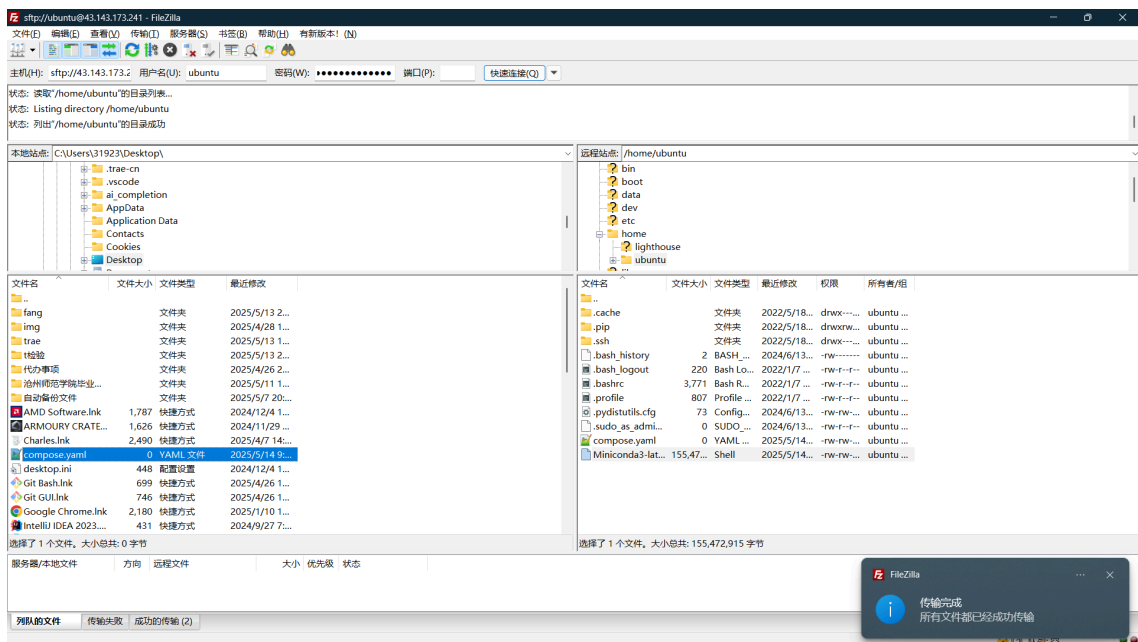
- 编写compose.yaml

```
version: '3.8'  
  
services:  
  mongo:  
    image: mongo:latest  
    container_name: mongo  
    environment:  
      MONGO_INITDB_ROOT_USERNAME: root  
      MONGO_INITDB_ROOT_PASSWORD: "123456"  
    ports:  
      - "27017:27017"  
    volumes:  
      - ./mongo_data:/data/db  
    restart: always  
  
  redis:  
    image: redis:latest  
    container_name: redis  
    command: redis-server --requirepass "123456"  
    ports:  
      - "6379:6379"  
    volumes:  
      - ./redis_data:/data  
    restart: always
```

```
mysql:
  image: mysql:latest
  container_name: mysql
  environment:
    MYSQL_ROOT_PASSWORD: "123456"
  ports:
    - "3306:3306"
  volumes:
    - ./mysql_data:/var/lib/mysql
  restart: always
```



- 通过 fz 上传到远程服务器上



- 在远程服务器上启动

```
docker-compose up
```

```
orcaterm.cloud.tencent.com/terminal?type=lighthouse&instanceId=ihins-4r1sawm8&region=ap-beijing&protocol=auto&from=lh_console_login_btn

(base) root@M-20-4-ubuntu:/# sudo mv /home/ubuntu/docker-compose-linux-x86_64 /usr/local/bin/docker-compose
(base) root@M-20-4-ubuntu:/# sudo chmod +x /usr/local/bin/docker-compose
(base) root@M-20-4-ubuntu:/# ls
bin boot data dev etc home lib lib32 lib64 libx32 lost+found media mnt opt proc root run sbin snap srv sys usr var
(base) root@M-20-4-ubuntu:/# cd /home/ubuntu/
(base) root@M-20-4-ubuntu:/home/ubuntu# ls
compose.yaml Miniconda3-latest-Linux-x86_64.sh
(base) root@M-20-4-ubuntu:/home/ubuntu# docker-compose up
empty compose file
(base) root@M-20-4-ubuntu:/home/ubuntu#
```

这里在本地编辑后忘保存了，又重新重复了下一步骤

```
orcaterm.cloud.tencent.com/terminal?type=lighthouse&instanceId=ihins-4r1sawm8&region=ap-beijing&protocol=auto&from=lh_console_login_btn

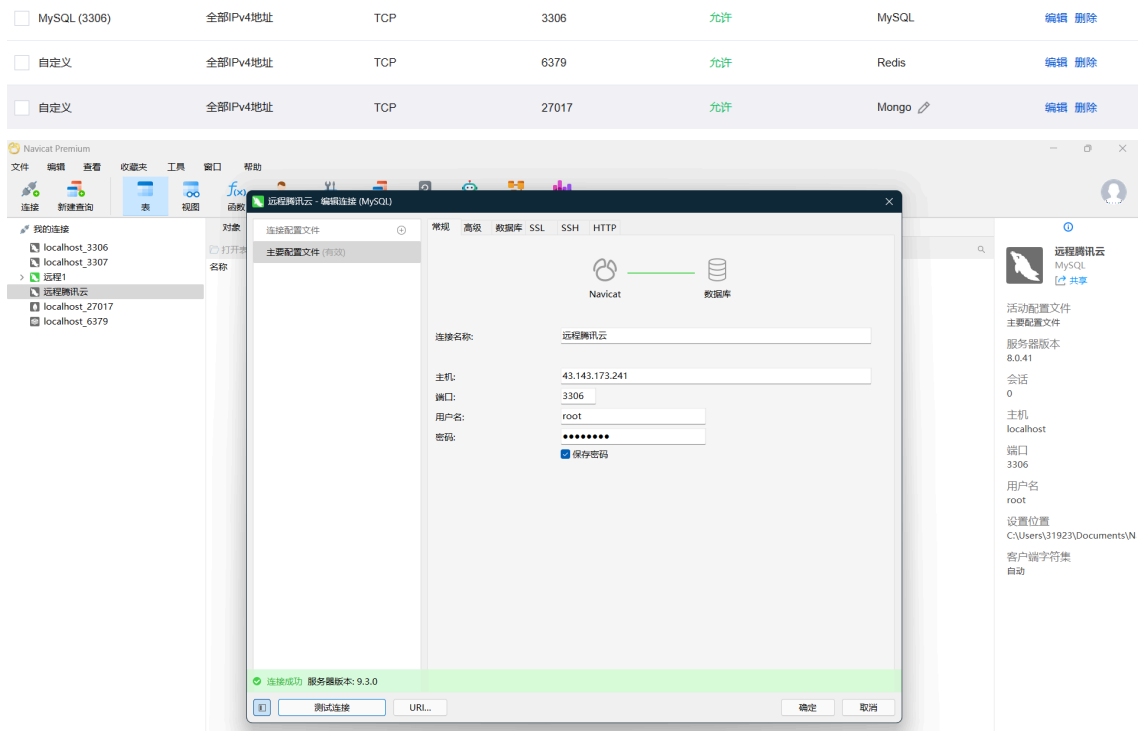
dash dash-grub-glue-efi nproc scriptreplay system-tty-ask-password-agent ypdomainname
date date-grub-kbdcomp nroff scsi_logging_level systemd-umount zcat
(base) root@M-20-4-ubuntu:/home/ubuntu# docker-compose up
WARN[0000] /home/ubuntu/compose.yaml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion
[+] Running 28/28
✓ mongo Pulled 16.2s
✓ redis Pulled 11.1s
✓ mysql Pulled 20.5s
```

```
orcaterm.cloud.tencent.com/terminal?type=lighthouse&instanceId=ihins-4r1sawm8&region=ap-beijing&protocol=auto&from=lh_console_login_btn

indexes for collections:"0 ms","Load cluster parameters from disk for a standalone:"0 ms","Build user and roles graph:"0 ms","Verify indexes for admin.system.users collection:"0 ms","Set up the background thread pool responsible for waiting for opTimes to be majority committed:"0 ms","Start up the replication coordinator:"2 ms","Ensure the change stream collections on startup contain consistent data:"0 ms","Write startup options to the audit log:"0 ms","Start transport layer:"1 ms","_initAndListen total elapsed time:"827 ms"}))}
mysql | 2025-05-14T02:42:18.394886Z 1 [System] [MY-013577] [InnoDB] InnoDB initialization has ended.
mysql | 2025-05-14T02:42:18.658582Z 0 [Warning] [MY-010068] [Server] CA certificate ca.pem is self signed.
mysql | 2025-05-14T02:42:18.658597Z 0 [System] [MY-013602] [Server] channel mysql_main configured to support TLS. Encrypted connections are now supported for this channel.
mysql | 2025-05-14T02:42:18.657292Z 0 [Warning] [MY-011810] [Server] Insecure configuration for --pid-file: location '/var/run/mysql/' in the path is accessible to all OS users. Consider choosing a different directory.
mysql | 2025-05-14T02:42:18.700736Z 0 [System] [MY-011323] [Server] X Plugin ready for connections. Bind-address: '::' port: 33060, socket: '/var/run/mysql/mysqlx.sock'
mysql | 2025-05-14T02:42:18.700823Z 0 [System] [MY-010931] [Server] /usr/sbin/mysqld: ready for connections. Version: '9.3.0' socket: '/var/run/mysql/mysql.sock' port: 3306 MySQL Community Server - 6.0.
mongo | {"t":{"$date":"2025-05-14T02:42:11.004+00:00"},"s":"M", "c":"CONTROL", "id":"636300", "ctx":"ftdc","msg":"Use of deprecated server parameter name","attr":{"deprecatedName":"internalQueryCacheSize","canonicalName":"internalQueryCacheWwEntriesPerCollection"}}
mongo | {"t":{"$date":"2025-05-14T02:42:11.004+00:00"},"s":"M", "c":"CONTROL", "id":"636300", "ctx":"ftdc","msg":"Use of deprecated server parameter name","attr":{"deprecatedName":"oplogSamplingIntervalSeconds","canonicalName":"collectionSamplingLogIntervalSeconds"}}
mongo | {"t":{"$date":"2025-05-14T02:42:11.005+00:00"},"s":"M", "c":"NETWORK", "id":"23803", "ctx":"ftdc","msg":"Use of deprecated server parameter 'sslMode', please use 'tlsMode' instead."}
mongo | {"t":{"$date":"2025-05-14T02:42:11.006+00:00"},"s":"M", "c":"CONTROL", "id":"636300", "ctx":"ftdc","msg":"Use of deprecated server parameter name","attr":{"deprecatedName":"wiredTigerConcurrentReadTransactions","canonicalName":"storageEngineConcurrentReadTransactions"}}
mongo | {"t":{"$date":"2025-05-14T02:42:11.006+00:00"},"s":"M", "c":"CONTROL", "id":"636300", "ctx":"ftdc","msg":"Use of deprecated server parameter name","attr":{"deprecatedName":"wiredTigerConcurrentWriteTransactions","canonicalName":"storageEngineConcurrentWriteTransactions"}}

w Enable Watch
```

- 远程连接mysql测试（第一次连接不上，看了下发现忘记放行端口了），放行端口



6 部署分布式爬虫

- 启动scrapyd服务

```
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

安装最新的 PowerShell, 了解新功能和改进! https://aka.ms/PSWindows

加载个人及系统配置文件用了 6413 毫秒。
(base) PS C:\Users\31923\Desktop\自动备份文件\3.学习\网络爬虫\作业\2025\05-13> scrapyd
2025-05-14T10:54:00+0800 [-] Loading C:\Users\31923\AppData\Roaming\Python\Python312\site-packages\scrapyd\txapp.py...
2025-05-14T10:54:01+0800 [scrapyd.basicauth#info] Basic authentication disabled as either 'username' or 'password' is un
set
2025-05-14T10:54:01+0800 [scrapyd.app#info] Scrapyd web console available at http://127.0.0.1:6800/
2025-05-14T10:54:01+0800 [-] Loaded.
2025-05-14T10:54:01+0800 [twisted.application.app.AppLogger#info] twisted 24.11.0 (D:\miniconda3\python.exe 3.12.9) start
ing up.
2025-05-14T10:54:01+0800 [twisted.application.app.AppLogger#info] reactor class: twisted.internet.selectreactor.SelectRe
actor.
2025-05-14T10:54:01+0800 [-] Site starting on 6800
2025-05-14T10:54:01+0800 [twisted.web.server.Site#info] Starting factory <twisted.web.server.Site object at 0x0000019AEE
0F86E0>
2025-05-14T10:54:01+0800 [Launcher] Scrapyd 1.5.0 started: max_proc=64, runner='scrapyd.runner'
```

- 启动gerapy可视化界面

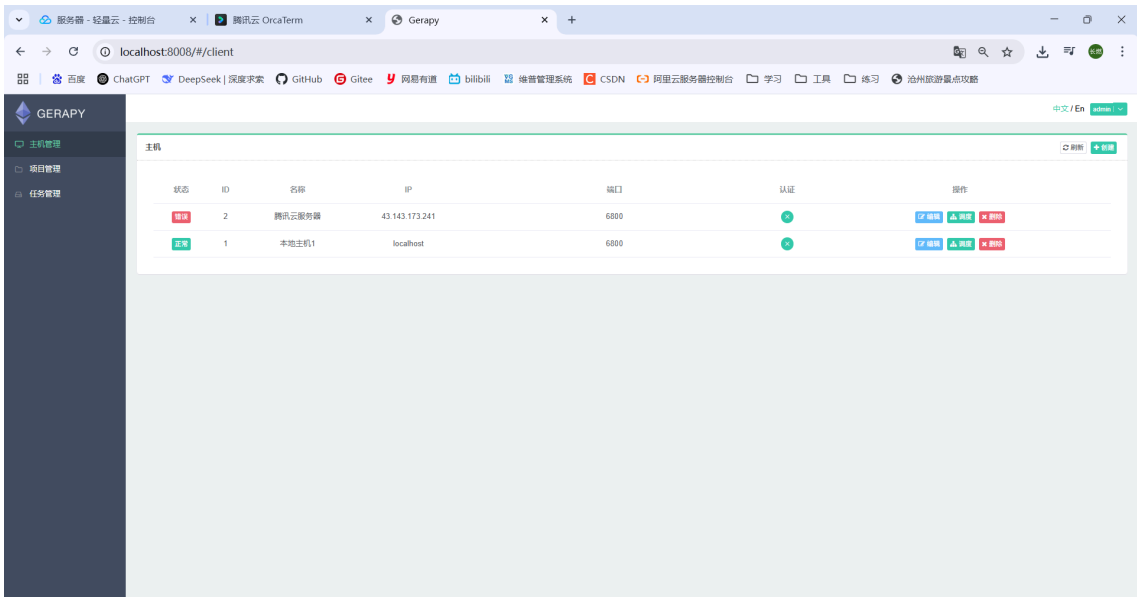
```
Windows PowerShell
版权所有 (C) Microsoft Corporation。保留所有权利。

安装最新的 PowerShell, 了解新功能和改进! https://aka.ms/PSWindows

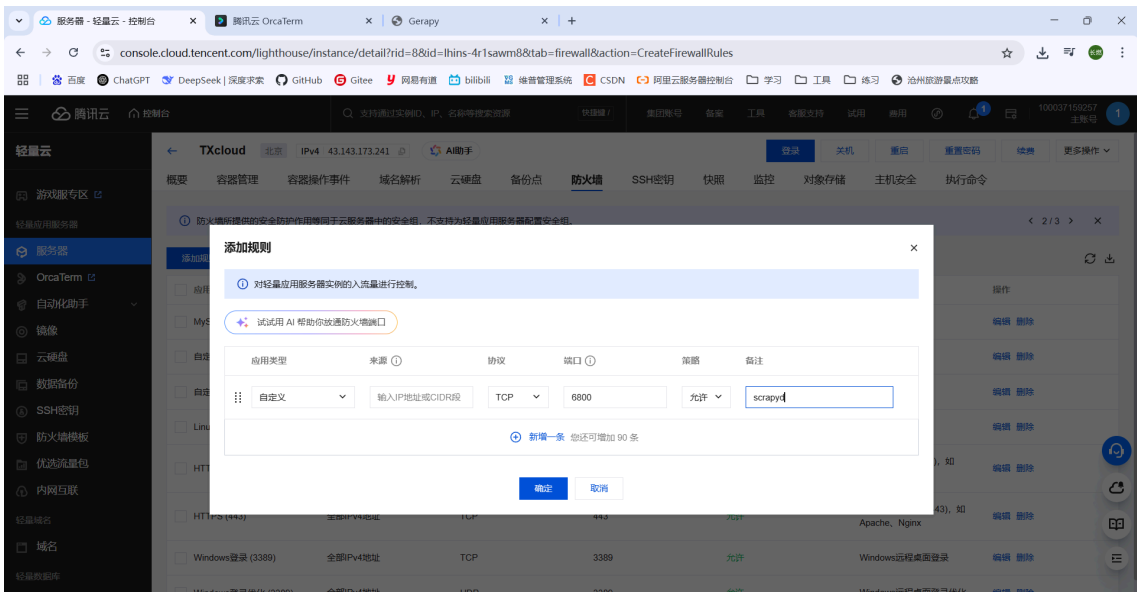
加载个人及系统配置文件用了 2969 毫秒。
(base) PS C:\Users\31923> d:
(base) PS D:\> cd .\Gerapy\
(base) PS D:\Gerapy> cd .\gerapy\
(base) PS D:\Gerapy\gerapy> gerapy runserver 127.0.0.1:8008
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).
INFO - 2025-05-14 10:56:17,628 - process: 14040 - scheduler.py - gerapy.server.core.scheduler - 105 - scheduler - succes
sfully synced task with jobs with force
May 14, 2025 - 10:56:17
Django version 2.2.28, using settings 'gerapy.server.server.settings'
Starting development server at http://127.0.0.1:8008/
Quit the server with CTRL-BREAK.
```

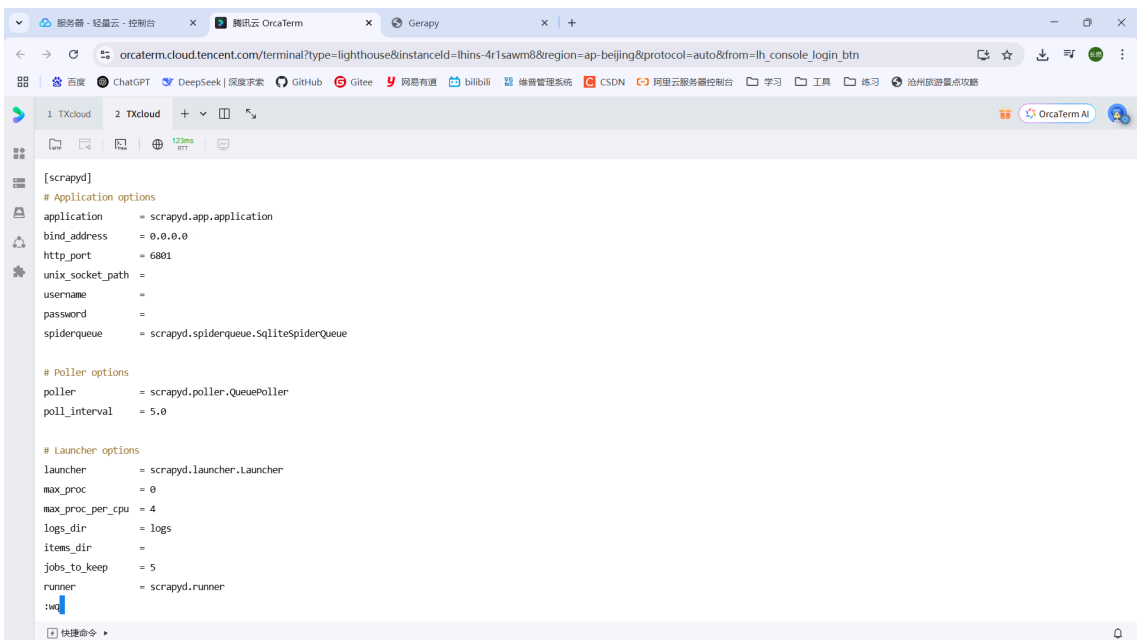
- 连不上，又忘记放行端口了。。。。不对，scrapyd 的配置文件也没改。。。。



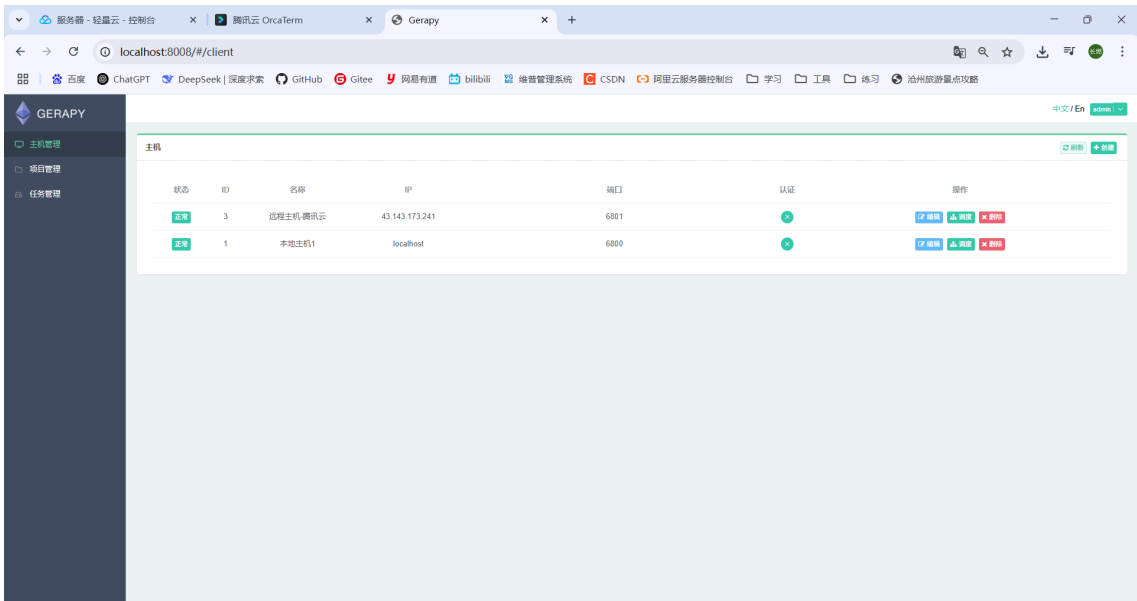
- 放行6800端口



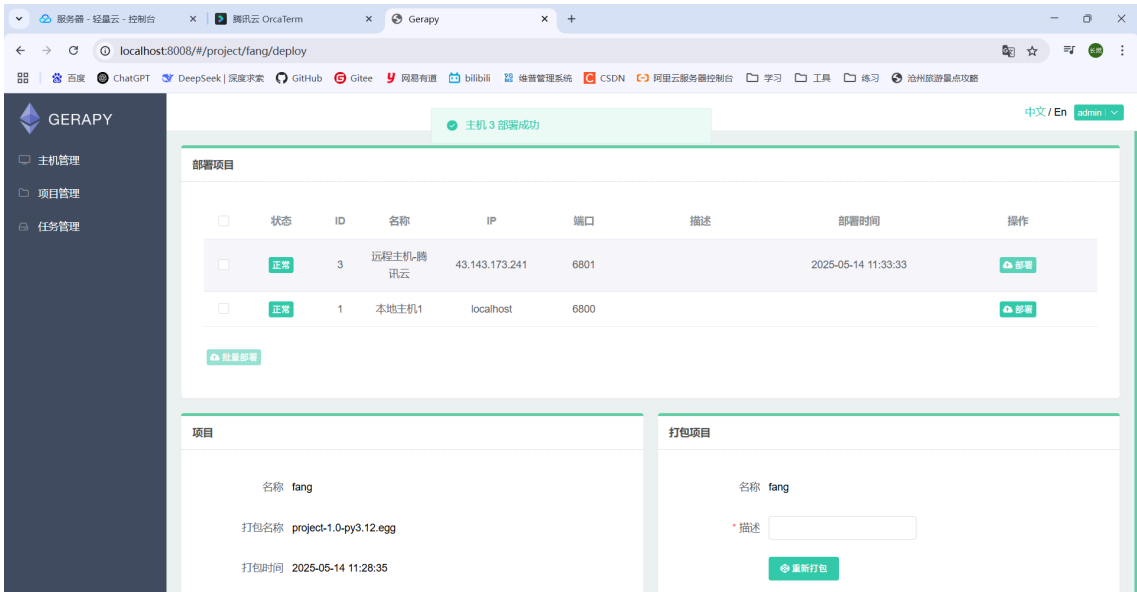
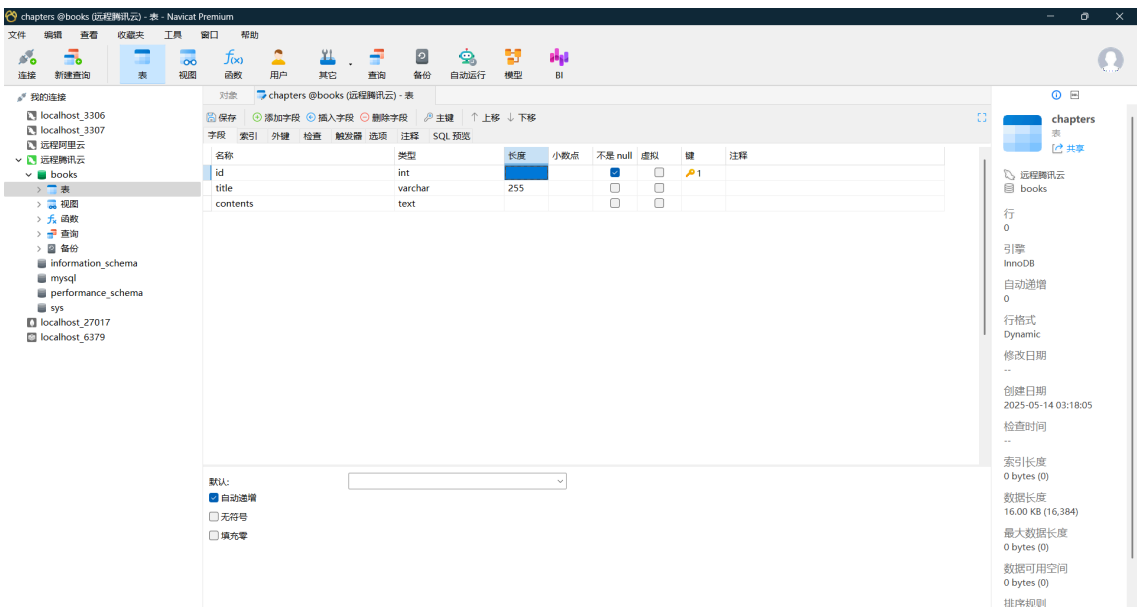
- 修改配置文件



- 重新连接，嗯。。。。忘了改端口了，放行端口6801，再重连



- 在腾讯云服务器上部署项目并运行，忘了建表了。。。。，先建表



- 没部署上，远程服务器包没装全，装包后运行项目
- 修改项目（修改mysql和redis的ip地址）
- 向redis中添加初始url

```
redis-cli -h 43.143.173.241 -p 6379 -a 123456 lpush baidu:start_urls  
https://www.xiaoshuopu.com/class_1/
```

- 运行结果
 - 远程mysql数据库

The screenshot shows the Navicat Premium interface. On the left, the 'Connections' pane shows a remote connection to 'localhost_6379' (remote MySQL cloud). The 'Database' pane shows the 'books' database selected. The 'Tables' pane shows the 'chapters' table selected. The main window displays the 'chapters' table structure and its contents.

id	title	contents
# int	# varchar(255)	# text
1	第1章 题名, 李奥	头疼, 剧烈的
2	第2章 却不收录	【姓名: 李奥
3	第3章 废神血	当章式浮动时
4	第4章 三六九等	在一个阳光明
5	第5章 至臻	一个时辰, 边
6	第6章 三圣地	此外, 随臂剑
7	第7章 肉身道	长春融盟, 盟
8	第8章 透力境	这一年, 燕北
9	第10章 万金属性	李奥位想, 如
10	第9章 太初	如两拳的力
11	第11章 李奥被剑	昨天还愈后还
12	第12章 死讯	林海海后还
13	第13章 和亲	看到提示, 李
14	第1章 世子	一夜起始战场
15	第2章 玉玺	这二人到底差
16	第15章 高我	燕北边境, 渊
17	第14章 最漂亮的地	什么情况? 干
18	第3章 幻相	皇子与世子们
19	第4章 陈点	明武阳正能大
20	第17章 剑圣	边如意真了。
21	第5章 始练	邹芳一弹远响
22	第一章 精魔人基里	白果园郊外的
23	第16章 心性	等女侍道下。
24	第18章 绝世面子	“如圣前辈, 且
25	第6章 太清	破关之诀在武
26	第二章 这位夫人。	黄昏时分, 白
27	第7章 出手	比翼的消息在

At the bottom, the SQL query is shown: `SELECT * FROM 'books'.chapters LIMIT 0,1000`. The status bar indicates '第1条记录 (共304条) 于第1页'.