# 法官AI助手安装文档

# 1 AI服务器安装

AI服务器安装主要参照如下文档



## 前置准备

1. 准备netstat,dos2uix,curl这样的常用离线安装包
2. 安装netstat,dos2uix,curl
3. 在ubuntu机子上创建root权限

https://blog.csdn.net/janeqi1987/article/details/106173711

## 大模型安装

### 安装大模型

大模型安装包地址：

http://192.168.1.9/svn/product\_manager/Version/45.法官AI助手/JAI\_V100R001C01/04 开发/转测归档/大模型安装

由于大模型安装包过大， 而且不会频繁改动， 不再在每一轮转测中归档， 建一个独立的目录。

1. 解压安装包：

tar -xzvf llm\_install.tar.gz

cd llm\_install

由于llm\_install.tar.gz文件过大，重新打包时间过长， 在安装之前请将SP1的内容拷贝至相关目录下：

$ cp sp1/docker-compose.yml llm\_install/

$ cp -r sp1/large\_doc\_summarize llm\_install/Langchain-Chatchat/

$ chmod a+x Langchain-Chatchat/large\_doc\_summarize/bin/start.sh

SP1地址：

http://192.168.1.9/svn/product\_manager/Version/45.法官AI助手/JAI\_V100R001C01/04 开发/转测归档/大模型安装/SP1

1. 安装 Docker：

sudo dpkg -i ./docker/\*.deb

sudo apt-get install -f

3. 将当前用户添加到 docker 组：

sudo usermod -aG docker $USER

注意：执行此命令后需要重新登录以使更改生效。

4. 安装 NVIDIA Container Toolkit：

sudo dpkg -i ./nvidia-container-toolkit/\*.deb

sudo apt-get install -f

sudo nvidia-ctk runtime configure --runtime=docker

sudo systemctl restart docker

5. 导入镜像：

docker load -i vllm-openai-0.3.3.tar

docker load -i chatchat-as-service-0.2.10.tar

6. 验证安装：

docker run --runtime nvidia --gpus all vllm/vllm-openai:0.3.3 nvidia-smi

docker run --runtime nvidia --gpus all --entrypoint nvidia-smi vllm/vllm-openai:0.3.3

### 1.2.2大模型启动

1. 快速启动

cd llm\_install

docker compose up -d

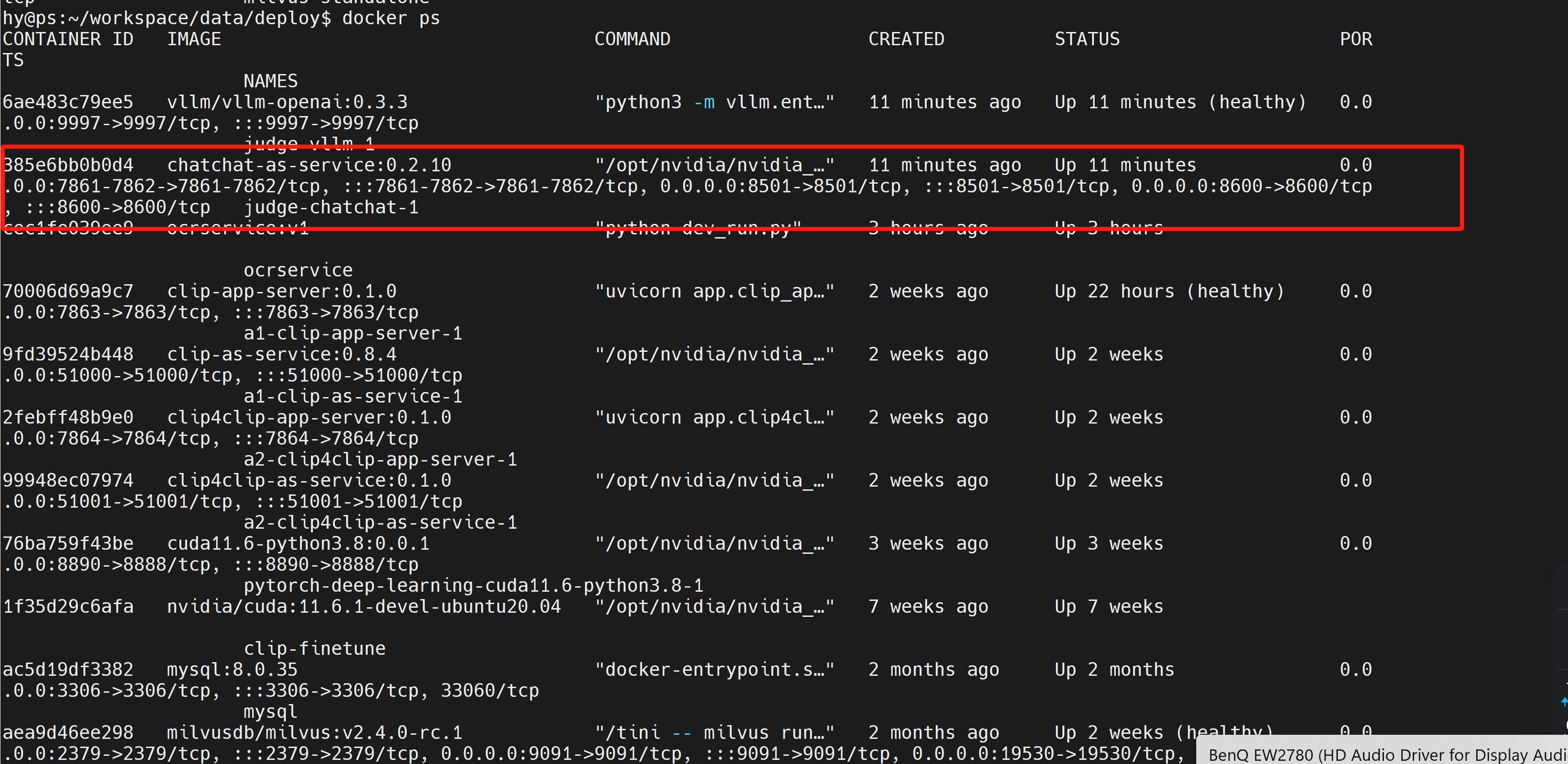
1. 停止服务

cd llm\_install

docker compose down

### 1.2.3 启动large\_doc\_summarize

查看chatchat-as-service的containerId



进入容器：

$ docker exec -it [containerId] bash

在容器中执行

# cd large\_doc\_summarize/bin

# nohup ./start.sh &

### 1.2.4安装OcrService

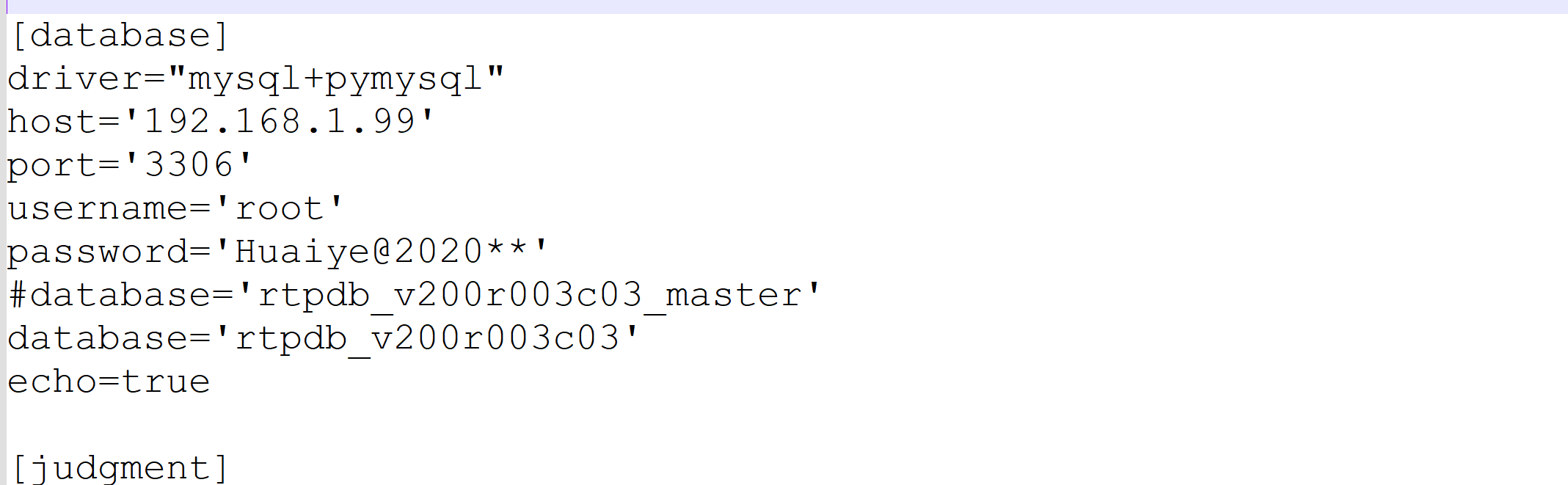
$ tar -zxvf ocr\_install.tar.gz

$ cd ocr\_install

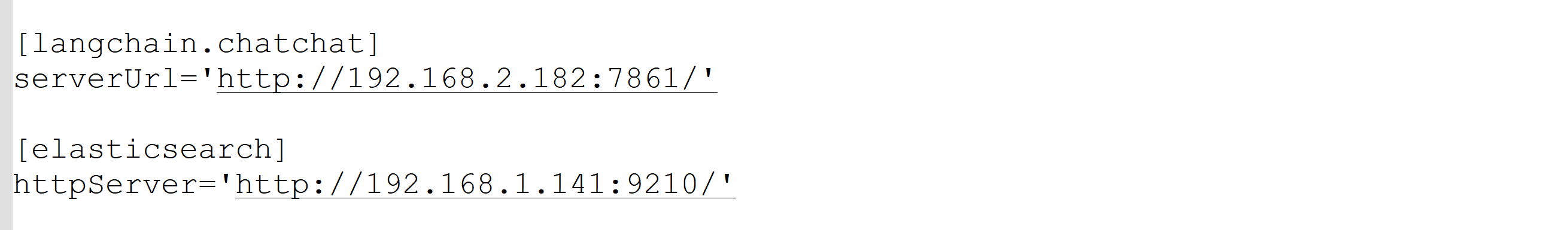
修改配置 settings.toml

主要修改：

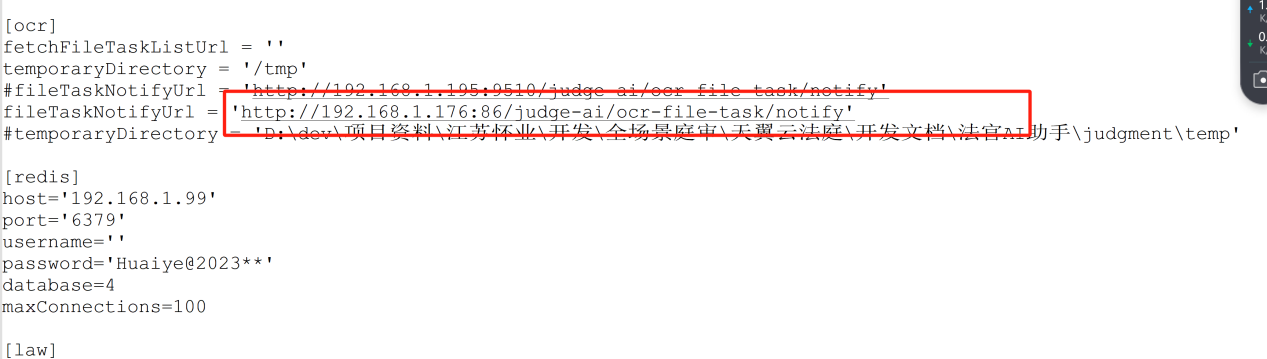
数据库配置



大模型与ES配置



redis配置与法官AI助手的回调地址



导入容器：

$ docker load -i ocrservice.tar

启动服务

docker compose up -d

停止服务：

$ cd ocr\_install

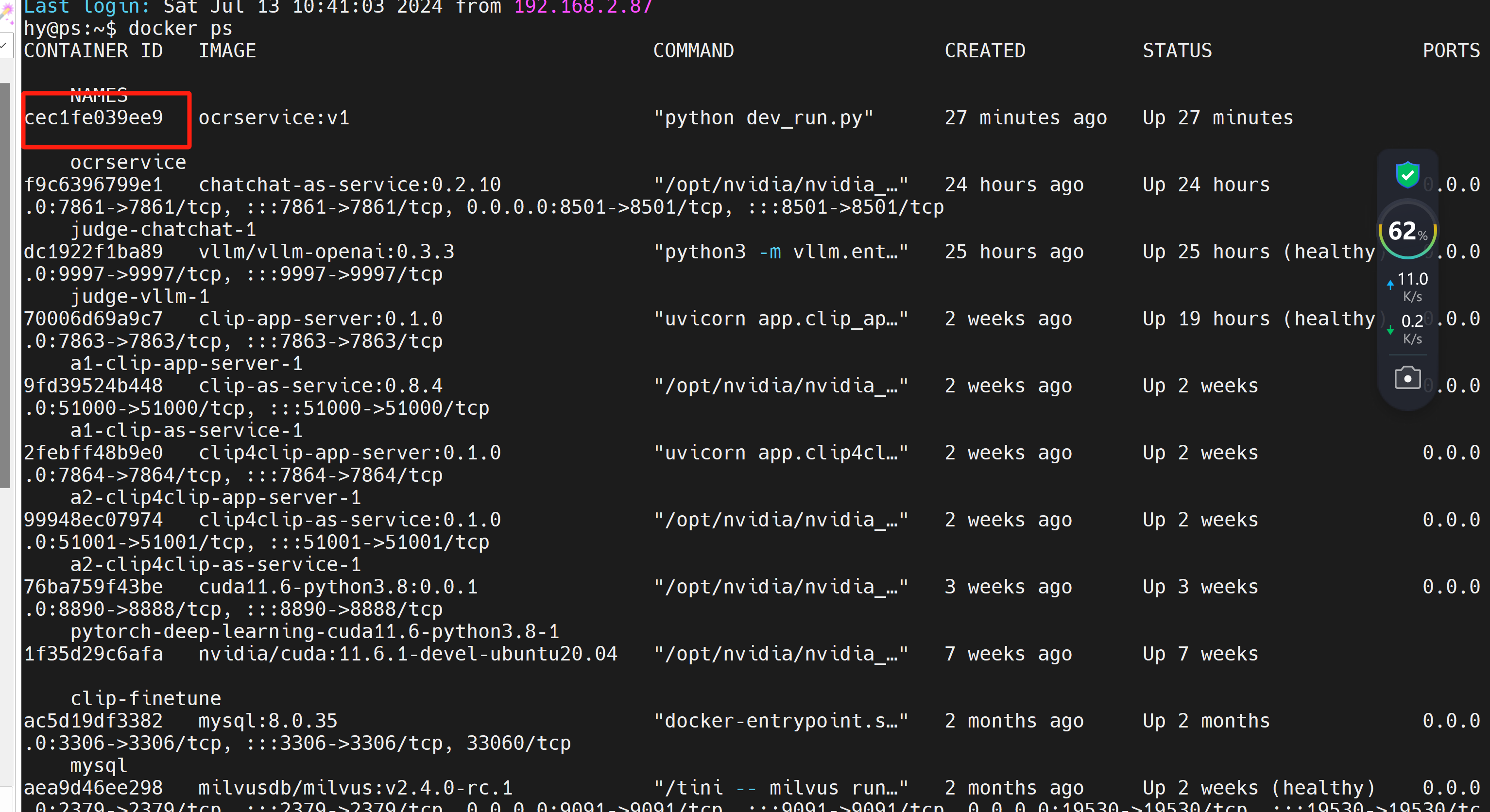
$docker compose down

常用操作

进入容器：

$ docker ps

查看containerId



$ docker exec -it [containerId] bash

查看日志：

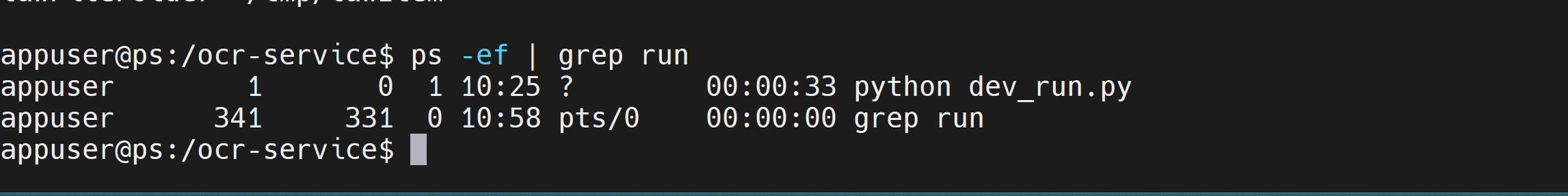
进入容器后

# tail -fn 1000 logs/2024-07-13.0001.ocr\_service.log

重启服务

进入容器后

# ps -ef | grep run



# kill -9 [pid]

# python dev\_run &

### 1.2.5安装Ocr集群

参考如下文档安装OCR集群， 1.2.4可以不用执行



### 1.2.6 MiniCPM 安装

1.解压安装包：

tar -xzvf minicpm.tar.gz

cd minicpm

2.确保已经安装了Docker、NVIDIA Container Toolkit（若未安装，移步至llm大模型《安装说明.MD》安装）

3.导入镜像

docker load -i ./minicpm-image.tar

4.创建容器

docker run -it --gpus all -v /home/hy/workspace/MiniCPM-V-2\_6:/mnt/MiniCPM-V-2\_6 --name minicpm -p 8050:8050 -e API\_KEY=asd123 -e GPU\_MEMORY\_UTILIZATION=1 -e PORT=8050 -e MAX\_MODEL\_LEN=2048 --ipc=host minicpm-image

注意：/home/hy/workspace/MiniCPM-V-2\_6为解压后模型路径，根据安装目录调整,命令输入后模型自动运行，关闭窗口即可

停止minicpm命令:docker stop minicpm

运行minicpm命令:docker start minicpm

### 1.2.7 大模型升级配置

将prompt\_config.py拷贝至llm\_install

/Langchain-Chatchat/configs目录下，最好先备份一下现网的prompt\_config.py

然后重启大模型

$ cd llm\_install

$ docker compose stop

$ docker compose start

待容器启动成功后， 进入容器启动large\_doc\_summarize

进入容器：

$ docker exec -it [containerId] bash

在容器中执行

# cd large\_doc\_summarize/bin

# nohup ./start.sh &

# 2 ES安装

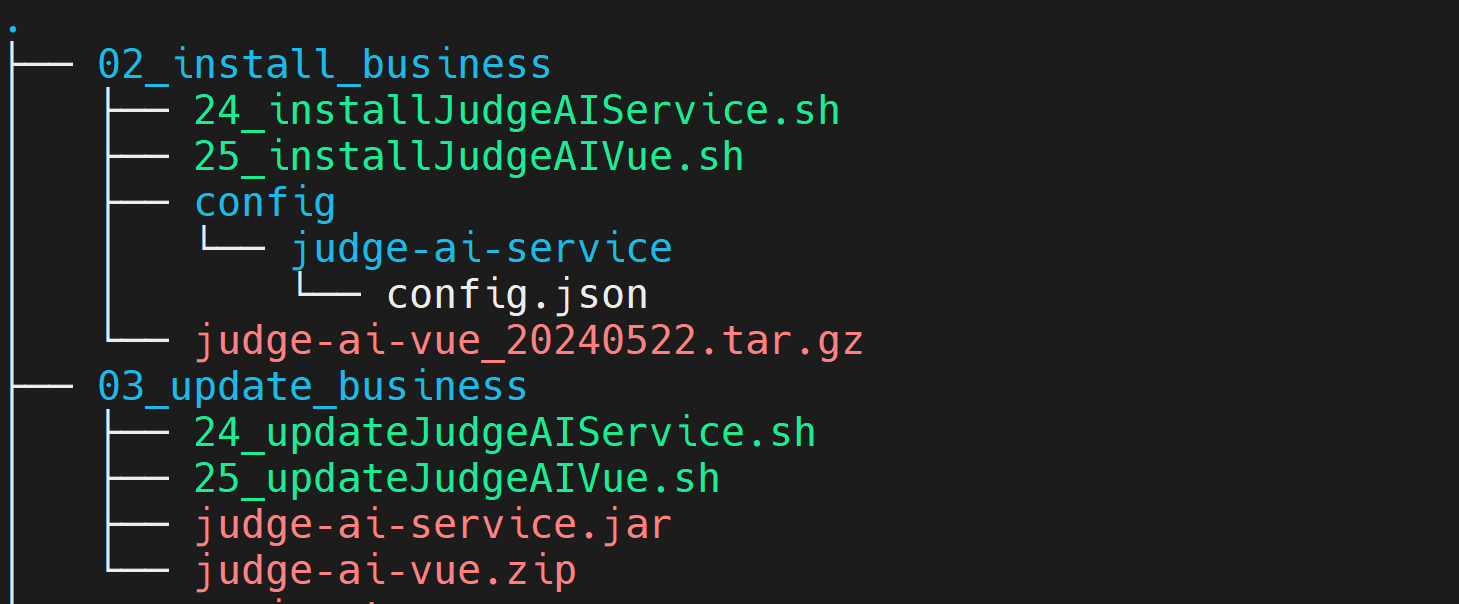
参考以下文档的2.1.2 章节进行Elasticsearch 7.x的安装



注意：如果现网的业务服务器配置较低， 可以将Elasticsearch安装到AI服务器上， 将setup\_in目录拷贝到AI服务器上， 按照上述文档进行按照即可。

# 3 业务安装

目前法官AI助手的业务包暂时未打包至业务全量包， 需要单独将安装脚本和安装包拷贝到相应的目录下



最新的包中， 已经将法官AI助手的业务包打包进setup\_in全量包中

## 3.1 安装

由于法官AI助手不是庭审标配产品， 不支持一键安装。

$ cp /home/setup\_in/02\_install\_business/package/judge-ai-service\_\* /home/setup\_in/02\_install\_business

$cp /home/setup\_in/02\_install\_business/package/judge-ai-vue\_\* /home/setup\_in/02\_install\_business

$./24\_installJudgeAIService.sh

$./25\_installJudgeAIVue.sh

## 3.2 更新

$./24\_updateJudgeAIService.sh

$./25\_updateJudgeAIVue.sh

## 3.3 修改nginx配置

将下面nginx.conf的配置追加至现网的nginx配置中，新增86端口



切换到rtpuser用户， 在用户目录下创建 如下目录

judge-ai-files/images

judge-ai-soft

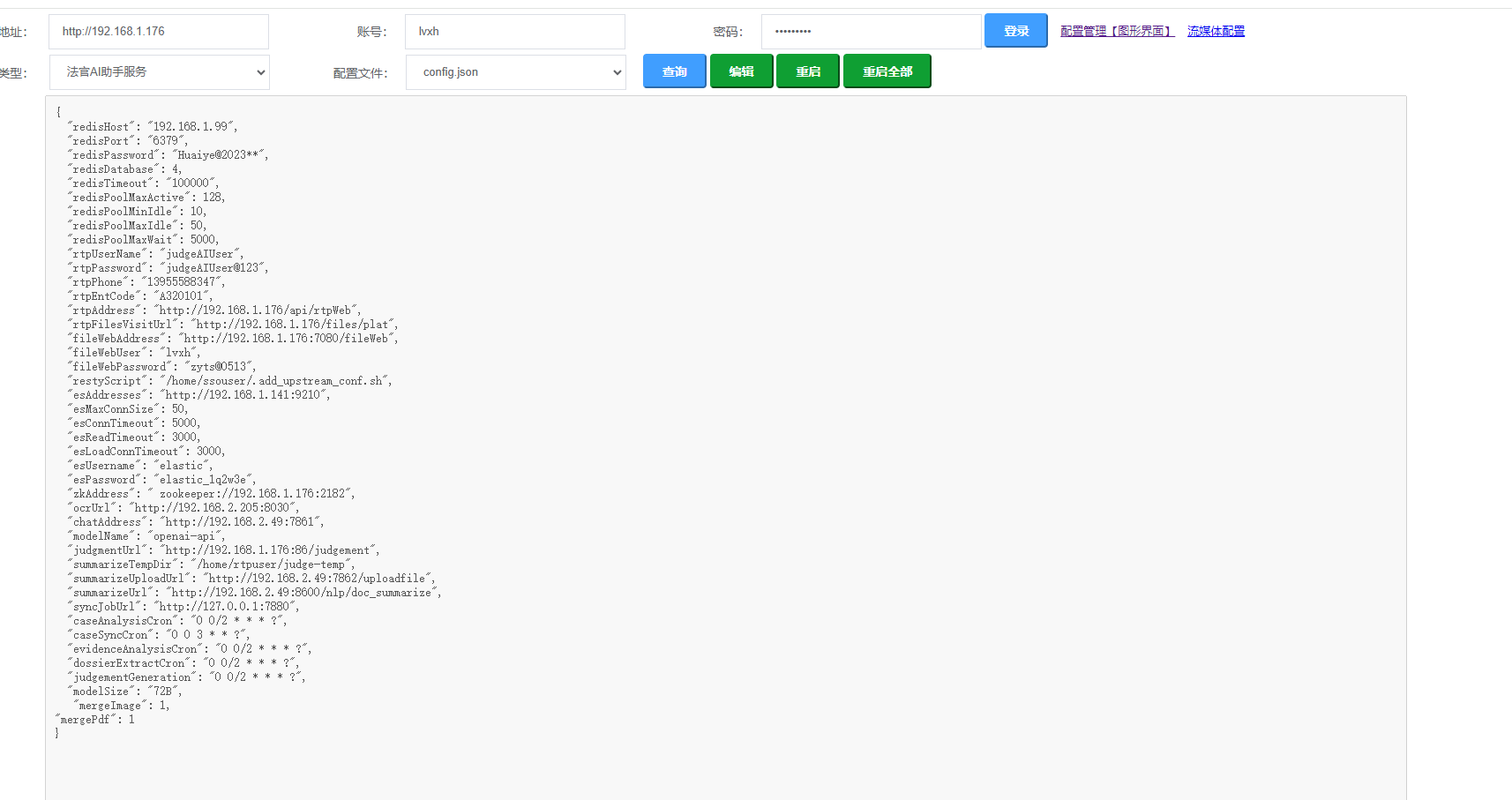
$ su - rtpuser

$ mkdir -p judge-ai-files/images

$ mkdir -p judge-ai-soft

## 3.4 业务配置

访问hymonitor页面， 对法官AI服务进行配置



主要配置说明如下：

redisHost：redis的IP

redisPort：redis的端口

redisPassword：redis密码

rtpAddress：庭审系统接口地址

rtpFilesVisitUrl：文件访问地址前缀

fileWebAddress：文件服务地址

fileWebUser：用户

fileWebPassword：密码

esAddresses：Elasticsearch访问地址

zkAddress：zookeeper地址

ocrUrl：OCR地址

chatAddress：大模型访问地址

judgmentUrl：审判文书访问地址， 和业务地址的IP和端口相同

summarizeUploadUrl

summarizeUrl

上述两个配置的IP与大模型访问地址的IP相同。

syncJobUrl：syncDataJob的服务地址， 用于现网的卷宗同步

caseSyncCron： 案件同步时间， 会同步庭审系统中的案件

mergeImage：是否合并图片，默认为0， 如需开启， 设置成1

mergePdf：是否合并PDF，默认为0， 如需开启， 设置成1

参考的配置文件如下：



## 3.5 验证

访问http://Ip:86

如：<http://192.168.1.176:86/>

使用庭审系统的账号即可登录

# 4导入数据

访问http://Ip:86

## 4.1 导入判决书

数据源：

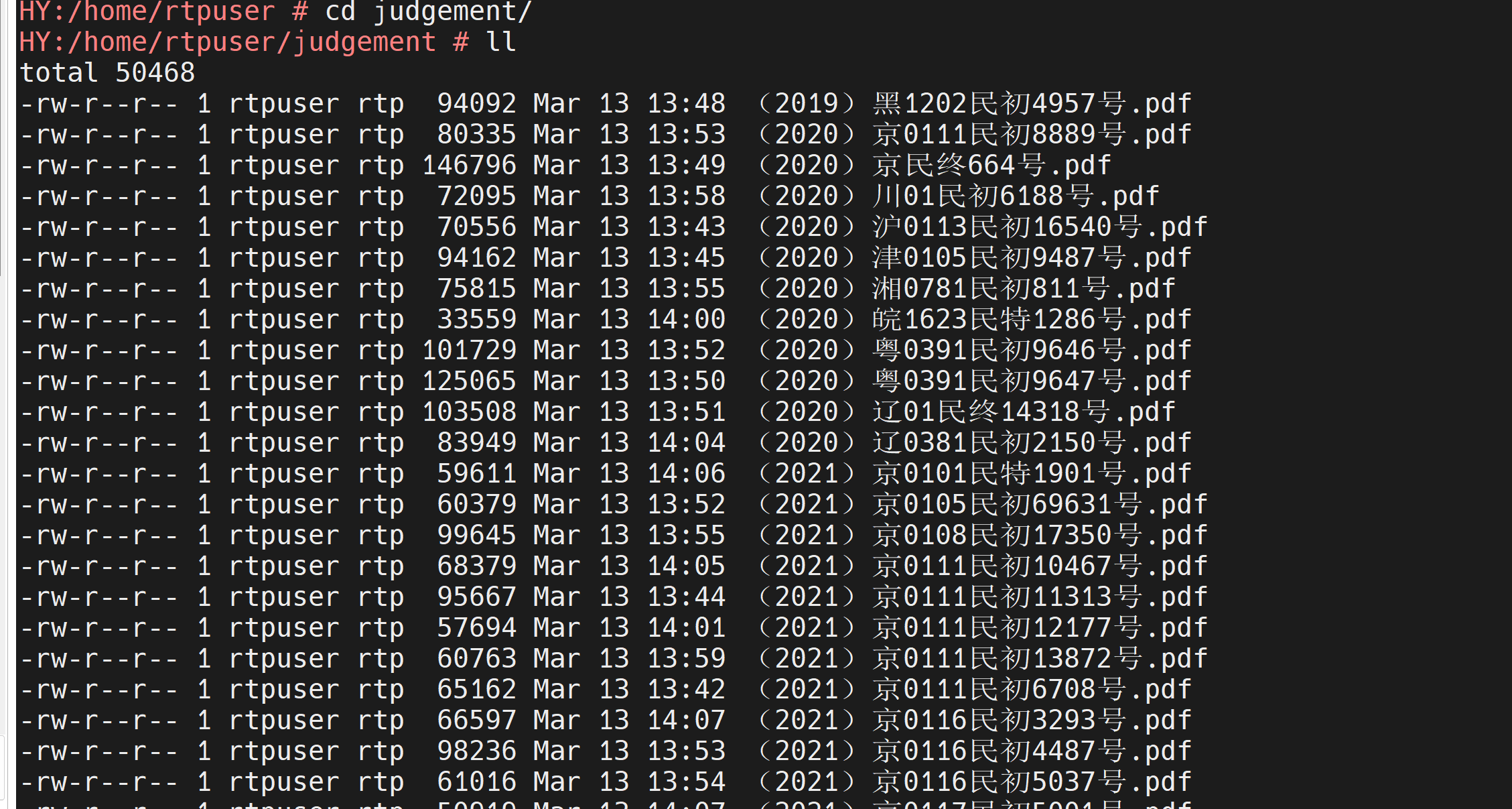
http://192.168.1.9/svn/product\_manager/Version/45.法官AI助手/JAI\_V100R001C01/04 开发/转测归档/B01/业务/初始化数据

在业务服务上

$ su - rtpuser

$ mkdir judgement

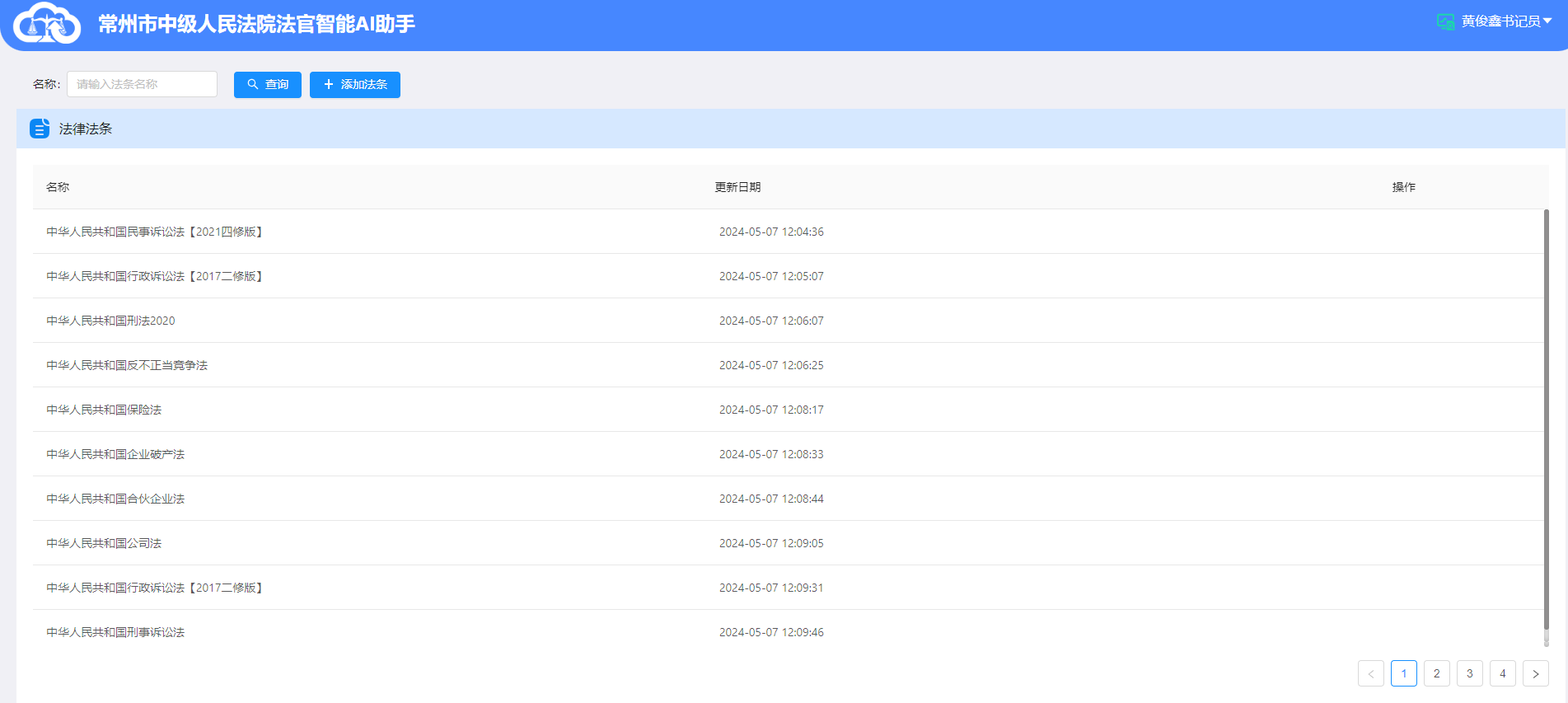
将judgement.zip解压至/home/rtpuser/judgement



## 4.2 法条维护

通过AI法官助手的系统， 将PDF法条上传至系统





## 4.3 法条导入知识库

如果新安装的大模型，需要将法律法条导入知识库

curl -X 'POST' \ 'http://192.168.2.49:8030/v1/law/upload-law-knowledge' \ -H 'accept: application/json' \ -d ''

导入法条时间比较久， 使用下述接口可以查询到27个文件就说明导入成功了

curl -X 'GET' \ 'http://192.168.2.182:7861/knowledge\_base/list\_files?knowledge\_base\_name=lawItem' \ -H 'accept: application/json'