OLLAMA部署

VPN环境

1. 节点来源：https://919101.cc/；

2. Ubuntu安装**Clash Verge**：**https://github.com/zzzgydi/clash-verge/releases**；

3. 配置-配置连接输入订阅地址，设置-系统代理打开。

OLLAMA

1. 下载OLLAMA：sudo wget https://ollama.com/download/ollama-linux-amd64 -O /tmp/ollama；

2. 配置ollama.service：

   a. sudo nano /etc/systemd/system/ollama.service

   [Unit]

   Description=Ollama Service

   After=network.target

   [Service]

   ExecStart=/usr/local/bin/ollama serve

   Restart=always

   RestartSec=3

   User=libra34567

   Group=libra34567

   Environment=PATH=/usr/bin:/usr/local/bin

   Environment=NODE_ENV=production

   Environment=OLLAMA_MODELS=/root/.ollama/models

   Environment=OLLAMA_HOST=0.0.0.0:11434

   Environment=OLLAMA_NUM_PARALLEL=3

   b. sudo systemctl daemon-reload；

3. 启动OLLAMA：

   a. sudo systemctl stop ollama；

  b. sudo systemctl start ollama;

  c. sudo systemctl status ollama;

4. 下载LLAMA3：ollama run llama3:70b；

5. 在ubuntu中通过Software&Update/Settings/Driver/ 更新或安装GPU的驱动

6. 设置并行数：

  a. sudo nano /etc/environment

  b. OLLAMA_NUM_PARALLEL=3;

7. 测试：curl -X POST http://localhost:11434/api/chat -d '{"model":"llama3:70b","messages":[{"role":"user","content":"hello"}],"stream":true}'

1. Mono x64

2. Go to server/MultiAgents

  a. dotnet publish -c Release -r linux-x64 --self-contaed true

3. Zip linux-x64

4. Run MutliAgentsServer_Release

  a. sudo chmod +x ./MultiAgentsServer_Release

  b. sudo ./MultiAgentsServer_Release

RAW：

[Unit]

Description=Ollama Service

After=network-online.target

[Service]

ExecStart=/usr/local/bin/ollama serve

User=ollama

Group=ollama

Restart=always

RestartSec=3

Environment="PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/libra34567/.dotnet:"

[Install]

WantedBy=default.target

只需要添加：Environment=OLLAMA_NUM_PARALLEL=3