

Orca Mini

Orca Mini

- 1、模型规模
- 2、拉取Orca Mini
- 3、使用Orca Mini
 - 3.1、运行Orca Mini
 - 3.2、进行对话
 - 3.3、结束对话
- 参考资料

演示环境

开发板：rdk x5系列主板

SD卡：64G

教程适用范围：主板是否可以运行，和系统的可使用内存相关，用户自己搭建环境和后台运行的程序都有可能

导致模型无法运行

1、模型规模

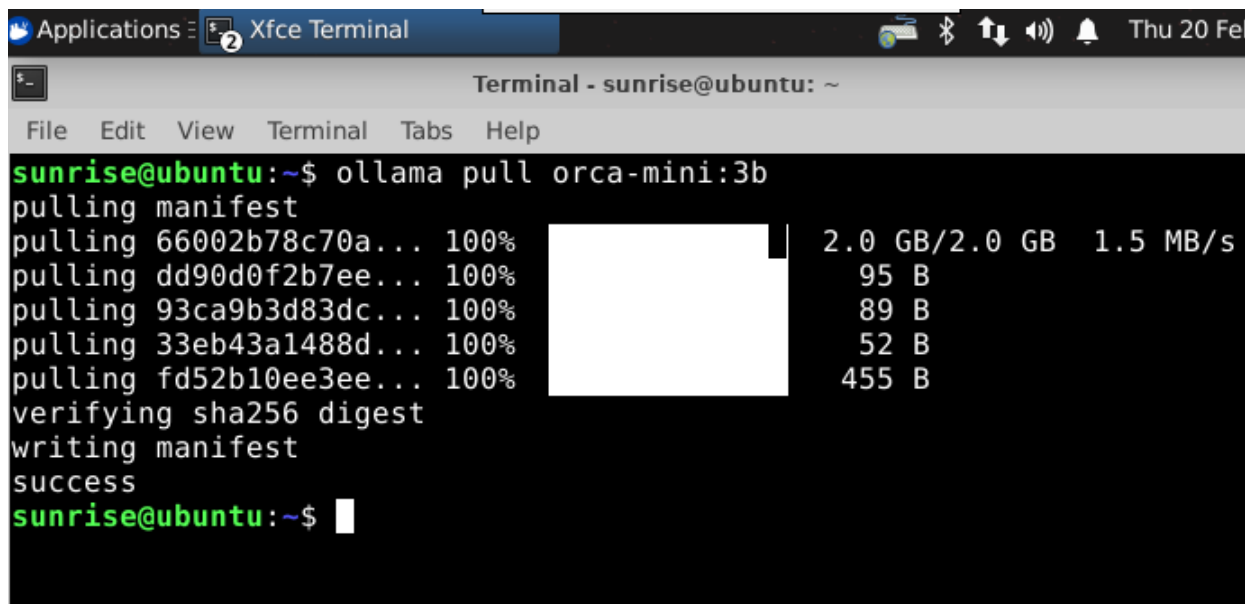
模型	参数
Orca Mini	3B

2、拉取Orca Mini

使用pull命令会自动拉取Ollama模型库的模型：

RDK X5 8Gb主板可以拉取此参数模型：

```
ollama pull orca-mini:3b
```



```
Applications ▢ Xfce Terminal
Terminal - sunrise@ubuntu: ~
File Edit View Terminal Tabs Help
sunrise@ubuntu:~$ ollama pull orca-mini:3b
pulling manifest
pulling 66002b78c70a... 100% ██████████ 2.0 GB/2.0 GB 1.5 MB/s
pulling dd90d0f2b7ee... 100% ██████████ 95 B
pulling 93ca9b3d83dc... 100% ██████████ 89 B
pulling 33eb43a1488d... 100% ██████████ 52 B
pulling fd52b10ee3ee... 100% ██████████ 455 B
verifying sha256 digest
writing manifest
success
sunrise@ubuntu:~$
```

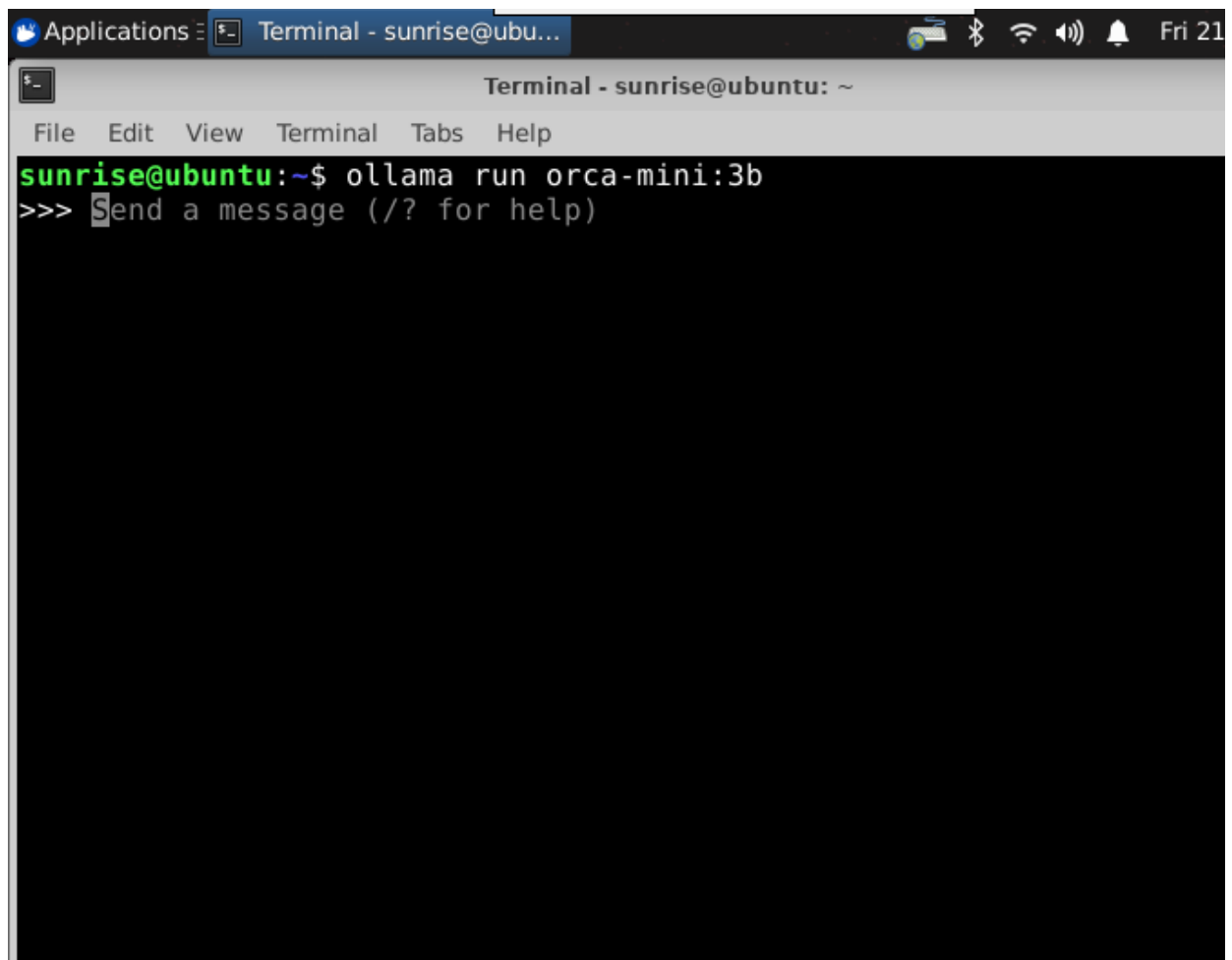
3、使用Orca Mini

3.1、运行Orca Mini

RDk X5 8Gb主板运行此参数模型

如果系统没有运行的模型，系统会自动拉取Orca Mini 1.1B模型并运行：

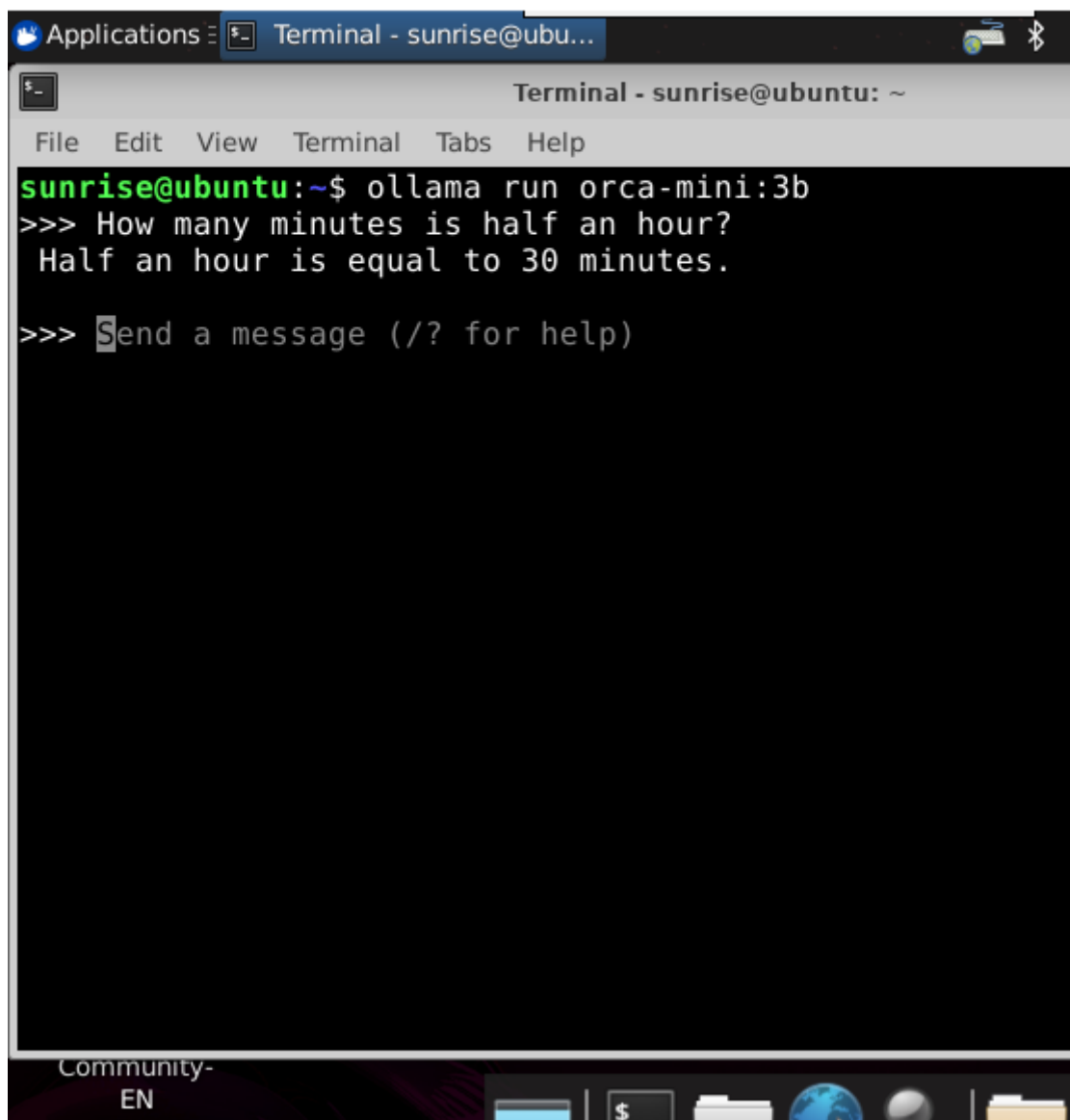
```
ollama run orca-mini:3b
```



3.2、进行对话

How many minutes is half an hour?

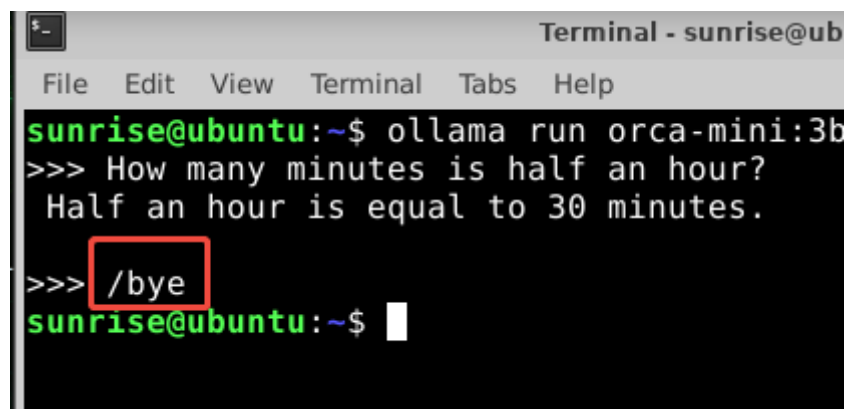
回复问题时间和硬件配置有关，耐心等待！



```
Applications ▢ Terminal - sunrise@ubu...  
Terminal - sunrise@ubuntu: ~  
File Edit View Terminal Tabs Help  
sunrise@ubuntu:~$ ollama run orca-mini:3b  
>>> How many minutes is half an hour?  
Half an hour is equal to 30 minutes.  
>>> Send a message (/? for help)
```

3.3、结束对话

使用 Ctrl+d 快捷键或者 /bye 可以结束对话！



```
Terminal - sunrise@ub  
File Edit View Terminal Tabs Help  
sunrise@ubuntu:~$ ollama run orca-mini:3b  
>>> How many minutes is half an hour?  
Half an hour is equal to 30 minutes.  
>>> /bye  
sunrise@ubuntu:~$
```

参考资料

Ollama

官网: <https://ollama.com/>

GitHub: <https://github.com/ollama/ollama>

Orca Mini

Ollama对应模型: <https://ollama.com/library/orca-mini>