

Ollama

Ollama

- 1. Large Language Model (LLM)
- 2. Ollama Installation
 - Script Installation
- 3. Use of Ollama
- 4. Uninstall Ollama
- References

Demonstration environment

Development board: RDK X5 motherboard

SD card: 64G

Tutorial application scope

Mother board model	Support
rdk x5 4gb	√
rdk x5 8gb	√

Ollama is an open source tool that aims to simplify the deployment and operation of large language models, allowing users to use high-quality language models in local environments.

1. Large Language Model (LLM)

Large Language Models (LLM) are a type of advanced text generation system based on artificial intelligence technology. Its main feature is that it can learn and understand human language through large-scale training data and can generate natural and fluent text.

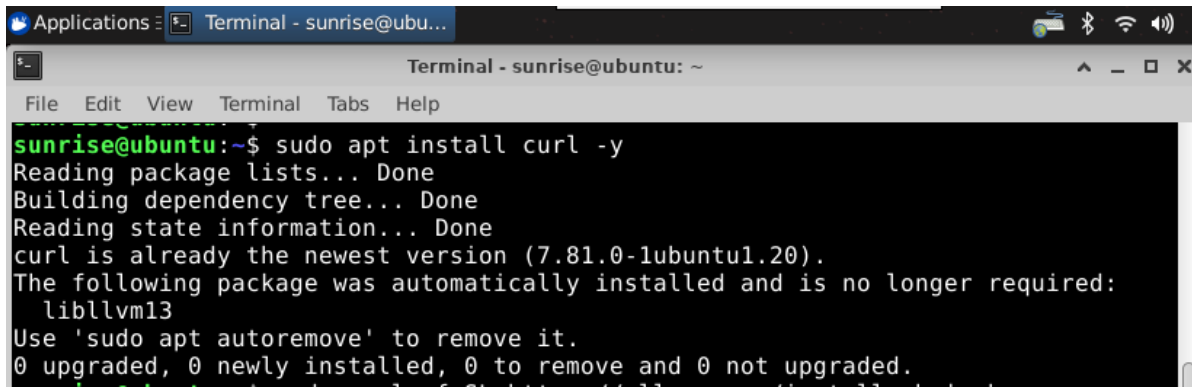
2. Ollama Installation

Note: It is recommended to use the image of the environment configured by Yabo Intelligent. It takes a long time to configure the environment personally, and there may be exceptions when installing different system versions.

The tutorial demonstrates the use of scripts to install Ollama on the rdk x5 motherboard. **(The image of the large language model provided by Yabo Intelligent has already configured the environment, no need to install it again)**

Script Installation

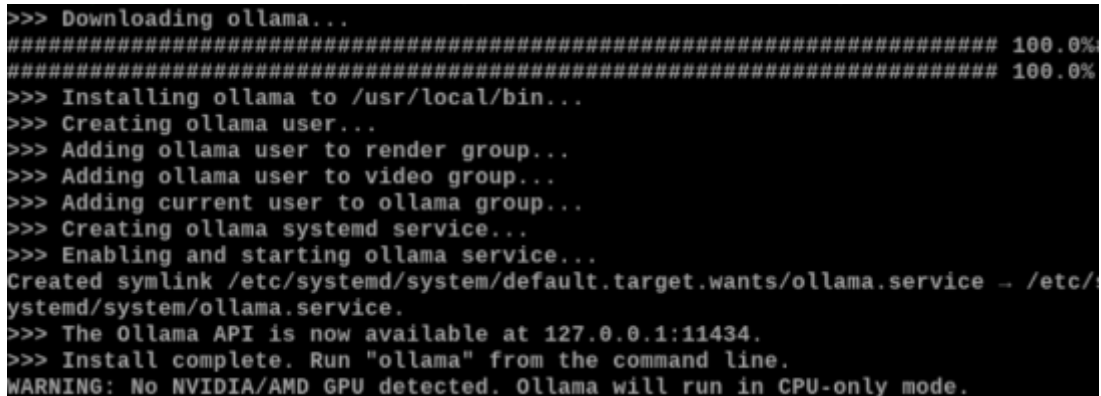
```
sudo apt install curl -y
sudo curl -fsSL https://ollama.com/install.sh | sh
```

A terminal window titled "Terminal - sunrise@ubu..." showing the command "sudo apt install curl -y" and its output. The output indicates that curl is already the newest version (7.81.0-1ubuntu1.20) and that libllvm13 was automatically installed and is no longer required. It also shows that 0 packages were upgraded, 0 newly installed, 0 to be removed, and 0 not upgraded.

```
sunrise@ubuntu:~$ sudo apt install curl -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
curl is already the newest version (7.81.0-1ubuntu1.20).
The following package was automatically installed and is no longer required:
  libllvm13
Use 'sudo apt autoremove' to remove it.
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
```

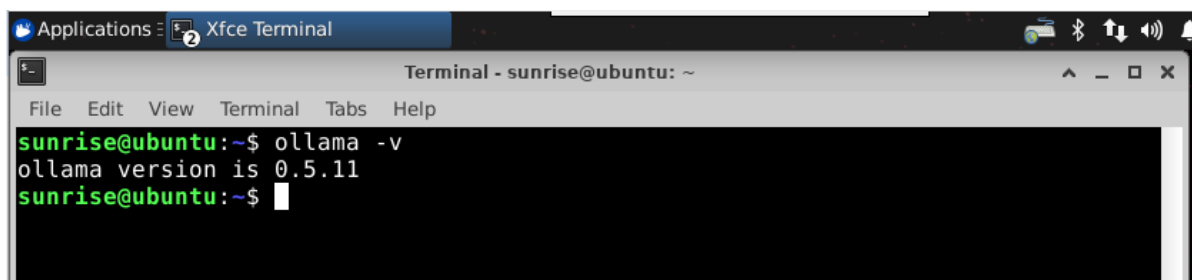
The entire installation process takes a long time, please wait patiently!

When the installation is complete, it will show.

A terminal window showing the installation of ollama. The output includes downloading ollama, installing it to /usr/local/bin, creating an ollama user, adding the user to the render and video groups, adding the current user to the ollama group, creating the ollama systemd service, enabling and starting the service, and creating a symlink. It also states that the Ollama API is now available at 127.0.0.1:11434 and that the installation is complete. A warning indicates that no NVIDIA/AMD GPU was detected, so ollama will run in CPU-only mode.

```
>>> Downloading ollama...
##### 100.0%
##### 100.0%
>>> Installing ollama to /usr/local/bin...
>>> Creating ollama user...
>>> Adding ollama user to render group...
>>> Adding ollama user to video group...
>>> Adding current user to ollama group...
>>> Creating ollama systemd service...
>>> Enabling and starting ollama service...
Created symlink /etc/systemd/system/default.target.wants/ollama.service -> /etc/systemd/system/ollama.service.
>>> The Ollama API is now available at 127.0.0.1:11434.
>>> Install complete. Run "ollama" from the command line.
WARNING: No NVIDIA/AMD GPU detected. Ollama will run in CPU-only mode.
```

Check the Ollama version number.

A terminal window titled "Xfce Terminal" showing the command "ollama -v" and its output. The output is "ollama version is 0.5.11".

```
sunrise@ubuntu:~$ ollama -v
ollama version is 0.5.11
sunrise@ubuntu:~$
```

3. Use of Ollama

Type ollama in the terminal and you will see the prompt.

```
Terminal - sunrise@ubuntu: ~
File Edit View Terminal Tabs Help
sunrise@ubuntu:~/Desktop$ cd
sunrise@ubuntu:~$ ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create     Create a model from a Modelfile
  show       Show information for a model
  run        Run a model
  stop       Stop a running model
  pull       Pull a model from a registry
  push       Push a model to a registry
  list       List models
  ps         List running models
  cp         Copy a model
  rm         Remove a model
  help       Help about any command

Flags:
  -h, --help      help for ollama
  -v, --version    Show version information
```

Command	Function
ollama serve	Start ollama
ollama create	Creating a model from a model file
ollama show	Display model information
ollama run	Run model
ollama pull	Pulling models from the registry
ollama push	Pushing the model to the registry
ollama list	List models
ollama ps	List models run
ollama cp	Copy model
ollama rm	Delete model
ollama help	Get help information about any command

4. Uninstall Ollama

- Delete service

```
sudo systemctl stop ollama
sudo systemctl disable ollama
sudo rm /etc/systemd/system/ollama.service
```

- Deleting files

```
sudo rm $(which ollama)
```

- Deleting model and service users and groups

```
sudo rm -r /usr/share/ollama  
sudo userdel ollama  
sudo groupdel ollama
```

References

Ollama

Website: <https://ollama.com/>

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