

1.Environments deployment

1.Environment deployment

[LLM](#)

[Install Ollama](#)

[Install scripts](#)

[Using Ollama](#)

[Uninstall Ollama](#)

[References](#)

1.Environment deployment

Development Boards: Jetson Nano

SD(TF)card: 64G

It is recommended to run the 4B and below parameter models

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.

LLM

Large Language Models (LLMs) are a type of advanced text generation system based on artificial intelligence technology. Their main features are the ability to learn and understand human language through large-scale training data and to generate natural and fluent text.

Install Ollama

This course mainly demonstrates how to install Ollama on Jetson nano using scripts.

Install scripts

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

```
jetson@jetson-desktop:~$ sudo apt install curl
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
apt-c-clone archdetect-deb bogl-bterm busybox-static cryptsetup-bin dpkg-repack gir1.2-timezonemap-1.0
gir1.2-xkl-1.0 grub-common kde-window-manager kinit kio kpackagetool5 kwayland-data kwin-common kwin-data
kwin-x11 libdebconf-indebconf libkdecorations2-5v5 libkdecorations2-private5v5 libkf5activities5
libkf5attica5 libkf5completion-data libkf5completion5 libkf5declarative-data libkf5declarative5
libkf5doctools5 libkf5globalaccel-data libkf5globalaccel5 libkf5globalaccel-private5 libkf5idletime5
libkf5jobwidgets-data libkf5jobwidgets5 libkf5kcmutils-data libkf5kcmutils5 libkf5kiocore5 libkf5kiontln5
libkf5skio-widgets5 libkf5newstuff-data libkf5newstuff5 libkf5newstuffcore5 libkf5package-data libkf5package5
libkf5plasma5 libkf5quickaddons5 libkf5solid5 libkf5solid5-data libkf5sonnet5-data libkf5sonnetcore5
libkf5sonnetui5 libkf5textwidgets-data libkf5textwidgets5 libkf5waylandclient5 libkf5waylandserver5
libkf5xmlgui-bin libkf5xmlgui-data libkf5xmlgui5 libkscreenlocker5 libkwin4-effect-builtins1
libkwin5effects11 libkwin5renderutils11 libkwin5renderutils-p1 libkwin5designer5 libkwin5help5
libkwin5multimedia5 libkwin5multimedia5-plugins libkwin5multimediaquick-p5 libkwin5multimediawidgets5 libkwin5opengl5
libkwin5quickwidgets5 libkwin5sql5 libkwin5test5 libxcb-composite0 libxcb-cursor0 libxcb-damage0 os-prober
python3-dbus.mainloop.pyqt5 python3-icu python3-pam python3-pyqt5 python3-pyqt5.qtsvg
python3-pyqt5.qtwebkit qml-module-org-kde-kquickcontrolsaddons qml-module-qtmultimedia qml-module-qtquick2
rdate tasksel tasksel-data
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
libcurl4
The following NEW packages will be installed:
curl
The following packages will be upgraded:
libcurl4
1 upgraded, 1 newly installed, 0 to remove and 346 not upgraded.
Need to get 337 kB of archives.
After this operation, 387 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://ports.ubuntu.com/ubuntu-ports bionic-updates/main arm64 libcurl4 arm64 7.58.0-2ubuntu3.24 [185 kB]
Get:2 http://ports.ubuntu.com/ubuntu-ports bionic-updates/main arm64 curl arm64 7.58.0-2ubuntu3.24 [152 kB]
Fetched 337 kB in 2s (162 kB/s)
debconf: delaying package configuration, since apt-utils is not installed
(Reading database ... 178835 files and directories currently installed.)
Preparing to unpack .../libcurl4_7.58.0-2ubuntu3.24_arm64.deb ...
Unpacking libcurl4:arm64 (7.58.0-2ubuntu3.24) over (7.58.0-2ubuntu3.16) ...
```

```
jetson@jetson-desktop:~$ sudo apt install curl
The following additional packages will be installed:
libcurl4
The following NEW packages will be installed:
curl
The following packages will be upgraded:
libcurl4
1 upgraded, 1 newly installed, 0 to remove and 346 not upgraded.
Need to get 337 kB of archives.
After this operation, 387 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://ports.ubuntu.com/ubuntu-ports bionic-updates/main arm64 libcurl4 arm64 7.58.0-2ubuntu3.24 [185 kB]
Get:2 http://ports.ubuntu.com/ubuntu-ports bionic-updates/main arm64 curl arm64 7.58.0-2ubuntu3.24 [152 kB]
Fetched 337 kB in 2s (162 kB/s)
debconf: delaying package configuration, since apt-utils is not installed
(Reading database ... 178835 files and directories currently installed.)
Preparing to unpack .../libcurl4_7.58.0-2ubuntu3.24_arm64.deb ...
Unpacking libcurl4:arm64 (7.58.0-2ubuntu3.24) over (7.58.0-2ubuntu3.16) ...
Selecting previously unselected package curl.
Preparing to unpack .../curl_7.58.0-2ubuntu3.24_arm64.deb ...
Unpacking curl (7.58.0-2ubuntu3.24) ...
Setting up libcurl4:arm64 (7.58.0-2ubuntu3.24) ...
Setting up curl (7.58.0-2ubuntu3.24) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
Processing triggers for libc-bin (2.27-3ubuntu1.4) ...
jetson@jetson-desktop:~$ curl -fsSL https://ollama.com/install.sh | sh
>>> Downloading ollama...
##### 100.0%#=#-# #
##### 100.0%
>>> Installing ollama to /usr/local/bin...
>>> Creating ollama user...
>>> 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.
>>> Installing NVIDIA repository...
curl: (22) The requested URL returned error: 404
jetson@jetson-desktop:~$
```

The entire installation process takes a long time, so be patient.

The message curl: (22) The requested URL returned error: 404 appears. You can ignore it!

Using Ollama

Enter ollama in the terminal to see the prompt:

```
jetson@jetson-desktop: ~  
Processing triggers for libc-bin (2.27-3ubuntu1.4) ...  
jetson@jetson-desktop:~$ curl -fsSL https://ollama.com/install.sh | sh  
>>> Downloading ollama...  
##### 100.0%#=#-# #  
##### 100.0%  
>>> Installing ollama to /usr/local/bin...  
>>> Creating ollama user...  
>>> 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.  
>>> Installing NVIDIA repository...  
curl: (22) The requested URL returned error: 404  
jetson@jetson-desktop:~$ 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  
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  
  
Use "ollama [command] --help" for more information about a command.  
jetson@jetson-desktop:~$
```

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

Uninstall Ollama

- Delete service

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

- Delete file

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

- Delete 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>