

Ollama

Ollama

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

Demo Environment

Development board: Jetson Orin series motherboard

SSD: 128G

Tutorial Scope

Motherboard Model	Supported
Jetson Orin NX 16GB	√
Jetson Orin NX 8GB	√
Jetson Orin Nano 8GB	√
Jetson Orin Nano 4GB	√

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

The tutorial demonstrates the use of scripts to install Ollama on the Jetson Orin series motherboard.

Script Installation

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

```
Activities Terminal Jan 21 09:09 MAXN SUPER jetson@yahboom: ~

jetson@yahboom:~$ sudo apt install curl -y
[sudo] password for jetson:
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 packages were automatically installed and are no longer required:
  gdal-data libaec0 libarmadillo10 libarpack2 libavcodec-dev libavformat-dev libavutil-dev libblosc1 libbcfistio9 libcharls2
  libdc1394-dev libdeflate-dev libexif-dev libfreexl1 libfyba0 libgdal30 libgdcm-dev libgdcm3.0 libgeos-c1v5 libgeos3.10.2
  libgeotiff5 libgl2ps1.4 libglew2.2 libgphoto2-dev libhdf4-0-alt libhdf5-103-1 libhdf5-hl-100 libheif1 liblmbase-dev libjbig-dev
  libjpeg-dev libjpeg-turbo8-dev libjpeg8-dev libkmlbase1 libkmldev1 libkmlengine1 liblpt5 libminizip1 libmysqlclient21
  libnetcdf19 libodbc2 libodbcinst2 libogdi4.1 libopencv-calib3d4.5d libopencv-contrib4.5d libopencv-dnn4.5d
  libopencv-features2d4.5d libopencv-flann4.5d libopencv-highgui4.5d libopencv-imgcodecs4.5d libopencv-imgproc4.5d
  libopencv-ml4.5d libopencv-objdetect4.5d libopencv-photo4.5d libopencv-shape4.5d libopencv-stitching4.5d libopencv-superres4.5d
  libopencv-video4.5d libopencv-videoio4.5d libopencv-videostab4.5d libopencv-viz4.5d libopenexr-dev libpng-dev libpq5 libproj22
  libraw1394-dev librttopo1 libsocket++1 libspatialite7 libsuperlu5 libswresample-dev libswscale-dev libsz2 libtbb-dev
  libtesseract4 libtiff-dev libtiffxx5 liburiparser1 libvtk9.1 libxerces-c3.2 mysql-common proj-data unixodbc-common
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 64 not upgraded.
jetson@yahboom:~$ sudo curl -fsSL https://ollama.com/install.sh | sh
>>> Installing ollama to /usr/local
>>> Downloading Linux arm64 bundle
##### 100.0%
>>> Downloading JetPack 6 components
##### 100.0%
>>> 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.
>>> NVIDIA JetPack ready.
>>> The Ollama API is now available at 127.0.0.1:11434.
>>> Install complete. Run "ollama" from the command line.
jetson@yahboom:~$
```

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

3. Use Ollama

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

```
Activities Terminal Jan 21 09:11 MAXN SUPER jetson@yahboom: ~

>>> Installing ollama to /usr/local
>>> Downloading Linux arm64 bundle
##### 100.0%
>>> Downloading JetPack 6 components
##### 100.0%
>>> 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.
>>> NVIDIA JetPack ready.
>>> The Ollama API is now available at 127.0.0.1:11434.
>>> Install complete. Run "ollama" from the command line.
jetson@yahboom:~$ 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

Use "ollama [command] --help" for more information about a command.
jetson@yahboom:~$
```

Command	Purpose
ollama serve	Start ollama
ollama create	Create a model from a model file
ollama show	Show model information
ollama run	Run a model
ollama pull	Pull a model from a registry
ollama push	Push a model to a registry
ollama list	List models
ollama ps	List running models
ollama cp	Copy a model
ollama rm	Delete a model
ollama help	Get help information about any command

4. Ollama Uninstall

- Deleting a 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

Official website: <https://ollama.com/>

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