Ollama

Demonstration Environment

Development Board: Jetson Nano

SD (TF) Card: 64GB

```
Recommended for models with 4B parameters or less
```

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

Large Language Models (LLMs)

Large Language Models (LLMs) are a type of advanced text generation system based on artificial intelligence technology. Their main feature is their ability to learn and understand human language using large amounts of training data and generate natural and fluent text.

Ollama Installation

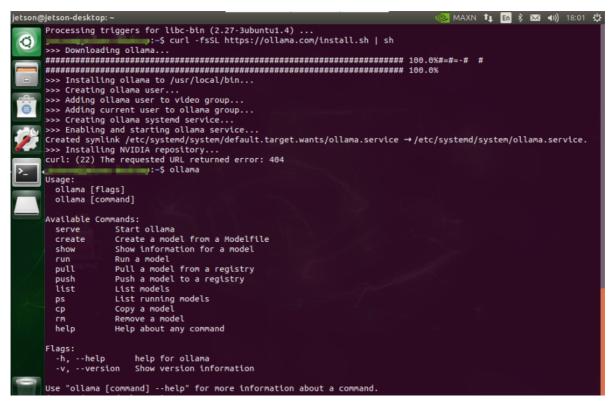This tutorial demonstrates how to install Ollama on a Jetson Nano using a script.

Script Installation

Note: Failure to pull the Ollama file is normal. You can try multiple times or use a proxy to pull the file. You will need to search online for proxy setup instructions.

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

```
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:
  gir1.2-gst-plugins-bad-1.0 libdc1394-dev libgdcm-dev libgphoto2-dev libraw1394-dev
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 71 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:~$
```

Using Ollama

Type "ollama" in the terminal to see the following prompt:


```
                                                    MAXN  En       18:01
jetson@jetson-desktop: ~
Processing triggers for libc-bin (2.27-3ubuntu1.4) ...
             :~$ 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
             :~$ 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.
```

Command	Function
ollama serve	Start Ollama
ollama create	Create a model from a model file
ollama show	Display model information
ollama run	Run a model
ollama pull	Pull a model from the registry
ollama push	Push a model to the registry
ollama list	List models
ollama ps	List running models
ollama cp	Copy a model
ollama rm	Remove a model
ollama help	Get help on any command

Ollama Uninstall

- Removing the service

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

- Removing files

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

- Removing the model and service user and group

```
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