Huawei Cloud - DeepSeek & Ollama Setup (Easy Guide)

This guide will help to deploy **DeepSeek 1.5B** on a Huawei Cloud ECS using **Ollama**, a tool to run language models easily.

Step 1: Set Up Your ECS

1.1 Create an ECS

1. Log in to [Huawei Cloud Console].
2. Go to **Elastic Cloud Server** and click **Buy ECS**.
3. Choose **Custom Config**:
    - **Billing**: Pay-per-use
    - **Region**: AP-Singapore
    - **Architecture**: x86
    - **Flavor**: c6.xlarge.2 (4 vCPUs, 8 GB RAM)
    - **Image**: CentOS 8.2 (40 GB)
    - **EIP**: Auto assign, 100 Mbps
    - **ECS Name**: `ecs-cent`
    - **Password**: `Huawei1234%` (Don't forget it)
4. Click **Submit**, then **Agree and Submit**.

Step 2: Log in to Your ECS

1. Use **CloudShell** from the web console.
2. Enter the ECS password: `Huawei1234%`

Step 3: Install Ollama

Ollama lets to run large language models easily.

Option 1: Install from GitHub

```
curl -fsSL https://ollama.com/install.sh | sh
ollama -v  # Verify installation
```

If download fails, try again or use Option 2.

Option 2: Install from OBS (Faster)

```
wget https://sandbox-experiment-files.obs.cn-north-
4.myhuaweicloud.com/deepseek/ollama-linux-amd64.tgz
sudo tar -C /usr -xzf ollama-linux-amd64.tgz
sudo chmod +x /usr/bin/ollama
```

Create a user and service

```
sudo useradd -r -s /bin/false -m -d /usr/share/ollama ollama
```

Create service file:

```
cat << EOF | sudo tee /etc/systemd/system/ollama.service
[Unit]
Description=Ollama Service
After=network-online.target

[Service]
Environment="OLLAMA_HOST=0.0.0.0:11434"
ExecStart=/usr/bin/ollama serve
User=ollama
Group=ollama
Restart=always
RestartSec=3

[Install]
WantedBy=default.target
EOF
```

Enable and start Ollama:

```
sudo systemctl daemon-reload
sudo systemctl enable ollama
sudo systemctl start ollama
ollama -v  # Check it works
```

Step 4: Install DeepSeek-R1 1.5B Model

Option 1: Pull from Ollama Hub

```
ollama pull deepseek-r1:1.5b
ollama run deepseek-r1:1.5b
```

If download is slow, press Ctrl+C and retry.

Option 2: Download from OBS (Faster)

```
wget https://sandbox-experiment-files.obs.cn-north-
4.myhuaweicloud.com/deepseek/ollama_deepseek_r1_1.5b.tar.gz
sudo tar -C /usr/share/ollama/.ollama/models -xzf ollama_deepseek_r1_1.5b.tar.gz
cd /usr/share/ollama/.ollama/models
mv ./deepseek/sha256* ./blobs
mkdir -p ./manifests/registry.ollama.ai/library/deepseek-r1
mv ./deepseek/1.5b ./manifests/registry.ollama.ai/library/deepseek-r1
rm -rf deepseek/
```

Run the Model

```
ollama list  # Check model exists
ollama run deepseek-r1:1.5b
```

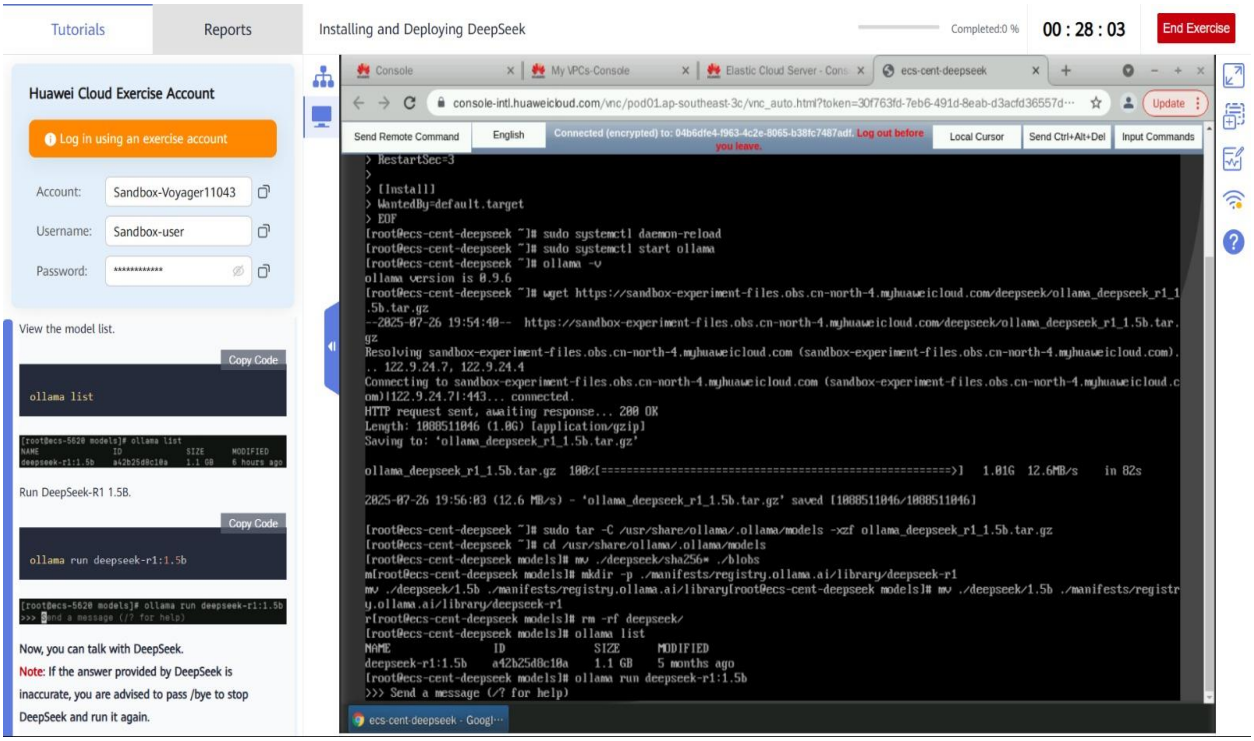
All Set!

now chat with DeepSeek right from ECS.

If responses are off, run `/bye` to stop and restart the model.

Enjoy using DeepSeek with Huawei Cloud!

Results:



The End