Usage instructions:

1. Launch the product via 1-click from AWS Marketplace. **Wait** until the instance status changes to 'Running' and passes all health checks. Then, connect to your instance using your Amazon private key and the '**ubuntu**' user."

   To update software, use: **sudo apt update && sudo apt upgrade -y**

2. **Run first boot**: On the server run:

   sudo codecreator-weaviate-firstboot

3. The setup wizard will ask a few simple questions. For most users, these choices are a good start:

- Weaviate admin identity [admin]:
- Weaviate admin API key [auto-generate if blank]:
- HTTP port [8080]:
- gRPC port [50051]:
- Local Prometheus port [2112]:
- Local profiler port [6060]:
- Default query limit [25]:
- Log level [info]:
- Enable Prometheus monitoring [Y]:
- Enable async indexing for heavier imports [Y]:
- Disable telemetry [Y]:
- Enable local filesystem backups [Y]:
- Enable S3 backups [N]:

Weaviate's ready endpoint returns 200 when the server is ready and 503 while it is still starting, so it is normal for first boot to pause briefly while the service becomes ready.

4. **Read your saved connection details**

After first boot completes:

cat /opt/codecreator/weaviate/FIRST_LOGIN.txt

5. **Basic health checks**

Check readiness:

source /opt/codecreator/weaviate/.env

**curl -i -H "Authorization: Bearer ${WEAVIATE_API_KEY}" **
http://127.0.0.1:${WEAVIATE_HTTP_PORT}/v1/.well-known/ready

A healthy result looks like:

HTTP/1.1 200 OK

**curl -s -H "Authorization: Bearer ${WEAVIATE_API_KEY}" **
http://127.0.0.1:${WEAVIATE_HTTP_PORT}/v1/meta | jq

6. **Find your public IP from inside EC2, run**

curl -s http://169.254.169.254/latest/meta-data/public-ipv4

Then your Weaviate API URL will be:

http://YOUR_PUBLIC_IP:8080

7. **Test the API from another machine**

From your laptop or workstation:

export WEAVIATE_URL="http://YOUR_PUBLIC_IP:8080"
export WEAVIATE_API_KEY="PASTE_YOUR_API_KEY_HERE"

**curl -i -H "Authorization: Bearer ${WEAVIATE_API_KEY}" **
"${WEAVIATE_URL}/v1/.well-known/ready"

Then:

**curl -s -H "Authorization: Bearer ${WEAVIATE_API_KEY}" **
"${WEAVIATE_URL}/v1/meta" | jq

Other Helpful info: For additional help see: https://docs.weaviate.io/weaviate

Note: *This AMI provides the Weaviate server backend only. To use the server, customers should connect with their own application, API requests, or supported client libraries such as Python or JavaScript from a local machine, workstation, or another server. A built in web GUI is not included, and some use cases may also require the customer to supply their own embeddings workflow or vectorized data.*

Extra Information: **(Optional)**

Allocate Elastic IP

To ensure that your instance **keeps its IP during restarts** that might happen, configure an Elastic IP. From the EC2 console:

1. Select ELASTIC IPs.
2. Click on the ALLOCATE ELASTIC IP ADDRESS.
3. Select the default (Amazon pool of IPv4 addresses) and click on ALLOCATE.
4. From the ACTIONS pull down, select ASSOCIATE ELASTIC IP ADDRESS.
5. In the box that comes up, note down the Elastic IP Address, which will be needed when you configure your DNS.
6. In the search box under INSTANCE, click and find your INSTANCE ID and then click ASSOCIATE.
7. Your instance now has an elastic IP associated with it.
8. For additional help: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/elastic-ip-addresses-eip.html