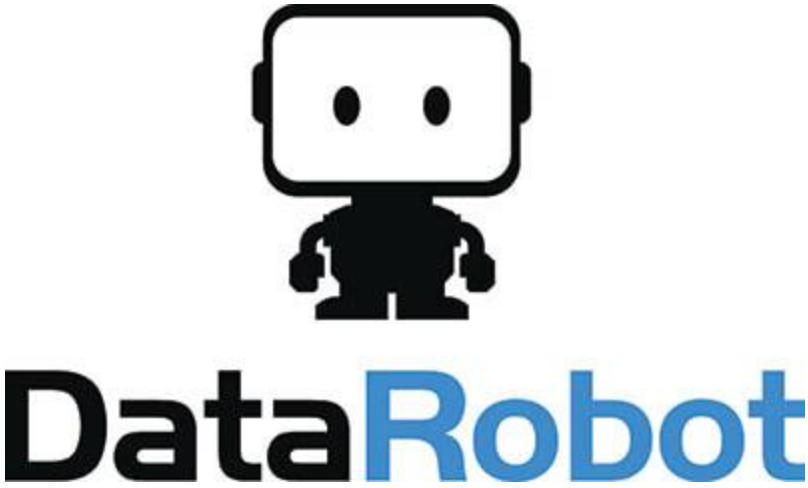

AWS Marketplace Single Node Provisioner for Platform Installation

---

Summary

DataRobot Platform 8 installation for AWS marketplace

License and Documentation.

This AWS Marketplace offering is a bring your own License. Installation and Configuration guide is provided with your license as part of Evidence of Delivery of the DataRobot subscription.

Infrastructure Planning

This datarobot offering is a single node configuration, the recommended configuration is to use the provided AMI and install the platform.

The quick start configuration of DataRobot is a single ec2 node platform that will support 4 modeler threads (workers) on r6i.4xlarge with S3 used for internal data storage with support for 5 GiB file ingest.

The cluster configurations of DataRobot use multiple ec2 nodes which balance the internal services for scalability and redundancy. The clusters start with 4 nodes: application, modeler, custom model, and predication. High availability data configurations can be added with additional 3 or 6 nodes. Additional modeler workers can be added in standing configuration or using AutoScaling in conjunction with metrics publishers. Please refer to the installation and configuration guide found in the home directory datarobot service account home directory.

Spin up Ec2 nodes

Locate the Datarobot offering in the AWS marketplace

Choose the “Launch through EC2” option

Select your localized EC2 options: tags, instance size, key pair, network settings.

Infrastructure Setup

1. BackEnd Storage
   Quickstart configuration uses an internal MinIO configuration.
   DataRobot can be configured to use a S3 bucket for backend storage, please refer to the installation and configuration guide for more information.
2. Security Group Rules
   Port 22 - will be used for admin access (during installation)
   Port 80/443 - will be used for end user access to the application

Installation of Platform

Login into the Application node with your ec2-user via ssh

1. Open an SSH client.

- Locate your private key file. The key used to launch this instance is me.pem
- Run this command, if necessary, to ensure your key is not publicly viewable.
- chmod 400 "me.pem"
- Connect to your instance using its Public DNS:
- ec2-hostname.compute-1.amazonaws.com
- Example:
  - ssh -i "me.pem" ec2-user@ec2-hostname.compute-1.amazonaws.com
  -

Switch to the service account ‘datarobot’

sudo su - datarobot

create ssh keys for the datarobot service account

2. Create ssh keys

```
[datarobot@ip-10-141-0-238 ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/datarobot/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/datarobot/.ssh/id_rsa.
Your public key has been saved in /home/datarobot/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:TPrDo8OvuDu0x+qnJfVl/oVXvRwx/jYWG6HuqZOyQo0
datarobot@ip-10-141-0-238
The key's randomart image is:
+---[RSA 3072]----+
|                 |
|                 |
|        .      + |
|       +      o =|
|      o S o  . =o|
|    .. E =  ....B|
|   ..+o * . o.o*o|
|    o+*o + +.oo..|
|   .BOo+o.o.+o   |
+----[SHA256]-----+
[datarobot@ip-10-141-0-238 ~]$ cd .ssh
[datarobot@ip-10-141-0-238 .ssh]$ ls
id_rsa  id_rsa.pub
[datarobot@ip-10-141-0-238 .ssh]$ cat id_rsa.pub >> authorized_keys
[datarobot@ip-10-141-0-238 .ssh]$ chmod 600 authorized_keys
[datarobot@ip-10-141-0-238 .ssh]$ ssh 10.141.0.238 date
The authenticity of host '10.141.0.238 (10.141.0.238)' can't be
established.
ECDSA key fingerprint is
SHA256:ke487w9aMGtj8CJnu+RW3X6jT4h6Ym/nbzH9IWYd7Is.
Are you sure you want to continue connecting (yes/no/[fingerprint])?
yes
Warning: Permanently added '10.141.0.238' (ECDSA) to the list of
known hosts.
Fri Aug 16 19:12:31 UTC 2024
```

```
[datarobot@ip-10-141-0-238 .ssh]$
```

. Modify the Config,yaml for ip address

Get your local system IP

hostname -i | awk -F ' ' '{print $2}'

```
Example

   [datarobot@ip-10-141-0-238 DataRobot-8.0.25]$ hostname -i | awk -F
' ' '{print $2}'

10.141.0.238
```

Update config.yaml replacing - 10.141.0.6 with local system IP

```
hosts:

- 10.141.0.238
```

3 . unpack your kit

```
   tar xvf DataRobot-RELEASE-8.0.*.tar.gz
```

4 . validate your config.yaml

./bin/datarobot validate

```
[datarobot@ip-10-141-0-238 DataRobot-8.0.25]$ ./bin/datarobot
validate

Validating configurations for DataRobot.

Validated config.yaml.

DataRobot config files are valid.
```

5 . Run setup dependencies

./bin/datarobot setup-dependencies

```
[datarobot@ip-10-141-0-238 DataRobot-8.0.25]$ ./bin/datarobot
setup-dependencies
```

6 . Run install

./bin/datarobot install

```
[datarobot@ip-10-141-0-238 DataRobot-8.0.25]$ ./bin/datarobot install
```

7 . Wait for the system to complete starting

```
[datarobot@ip-10-141-0-238 ~]$ docker ps | grep app
```

…

```
0d22cff3b272   10.141.0.238:5000/datarobot/datarobot-runtime:8.0.25r5
"/entrypoint datarob…"   18 hours ago   Up 5 minutes (healthy)
app
```

8 . create initial admin user

./bin/datarobot users reset-admin-credentials

```
[datarobot@ip-10-141-0-238 DataRobot-8.0.25]$ ./bin/datarobot  create
initial admin user
```

username is email style, i.e admin@datarobot.com , password 8 characters, 1 upper, 1 number

9. Log into the GUI with chrome browser.

10. Apply your license

Person icon (upper right corner) -> License -> <paste license string> -> validate -> apply

11. Test the system, create a new project with a sample dataset can be downloaded as follows:

https://s3.amazonaws.com/datarobot_public_datasets/10k_diabetes.xlsx (use "readmitted" as your target variable)

A copy of the 10k_diabetes.xlsx is located the home directory of datarobot