

Redhat Deployment Guide

erwin Data Quality V3.1.0

Prepared by



Introduction	1
Software Requirements (Auto Install)	2
PostgreSQL Installation (Only applicable for DB isolated Deployment)	2
Pre-Installation Setup	4
Configuration Setup	5
Details about the config file	5
Verification of Installation	7
Post-Installation Procedure	8

Introduction

This document provides a comprehensive, step-by-step process for installing DQLabs on RHEL 9. It is designed for personnel with technical knowledge of RHEL 9 and Linux Operating Systems. The guide walks you through the installation steps, ensuring that you have a clear understanding of the process and any prerequisites specific to RHEL. By following this guide, you will be able to successfully set up the DQLabs application and leverage its features in your RHEL environment.

System Requirements

This section provides the minimum system and mandatory requirements needed to install the DQLabs application in the Linux environment successfully.

Category	Recommended
Operating system	Red Hat Enterprise Linux 9.x(Supported on Azure, AWS, and on-prem VMs) If AWS, use Amazon Machine Image (AMI) (HVM) with SSD Volume Type.
Processor	64-bit processor
Disk Space	Minimum 100 GB, and it should be in the <code>/directory</code> or the DQLabs installation folder, and the install user home directory should have a minimum of 5GB

Package	Core and RAM Specifications
Bronze	4 Core 8 GB RAM
Silver	4 Core 8 GB RAM
Gold	8 Core 16 GB RAM
Platinum	16 Cores 32 GB RAM
Titanium	32 Cores 64 GB RAM

Postgres Server Prerequisites (Only in case of DB Isolated Deployments)

Operating System	Redhat 9.x
CPU Core	4 Cores or more
RAM	8GB or more

1. **Dedicated Server:** DQLabs needs a dedicated server for installation
2. **Internet Access:**

URL	Purpose	Required during
https://s3.amazonaws.com	This URL must be whitelisted to allow downloading binaries from the DQLabs repository.	Required only before installation. The file can be downloaded externally and moved to the server if needed
cdn.redhat.com subscription.rhsm.redhat.com	Official RHEL repositories are accessed to download and update the necessary packages Command	These repositories can be limited after updates and before installation

3. **License key:** A new DQ license key is required for activating the product upon installing the product
4. **Ports to be opened:** Ports Used for Internal Communication within the Application:

PostgreSQL	5432 (Mandatory)
Airflow	8080
HTTP	80
HTTPS	443

Software Requirements (Auto Install)

The requirements mentioned below are auto-installed with the script; the user should not manually install any of the software requirements in the DQServer

Services	Version
PostgreSQL	15.12
Python	3.11.0
Java	17.0.14
Airflow	2.8.1
Drivers	MSSQL, Oracle, PostgreSQL, MySQL ODBC/JDBC
Spark	3.5.1

Before proceeding with Single Server Deployments, ensure the following:

1. The server is free from any pre-installed applications.
2. PostgreSQL should not be manually installed for single-server deployments.
3. No third-party or external applications should be present on the server.

Pre-Installation Setup

Step 1: Install the wget command using the following command:

```
None
sudo yum install wget -y
```

Step 2: Download the Prerequisite File from S3 into the directory where the application should be installed

```
None
wget
https://s3.us-east-1.amazonaws.com/erwin-2.0/code/linux/application-code/3.1.0/On_premise/Packages_3.1.0/Erwin-Redhat-Packages/Erwin-prerequisites.tar
```

Step 3: Extract the Downloaded Tar File

```
None
tar -xvf Erwin-prerequisites.tar
```

Step 4: Execute the following commands:

```
None
#Navigate to the p7zip_file folder and Install the p7zip Plugin
cd prerequisites/p7zip_file

#Install the p7zip plugin:
sudo rpm -ivh *.rpm

#Remove the redundant tar file
sudo rm -rf ~/Erwin-prerequisites.tar

#Return to Prerequisites Folder
cd ~
```

Step 5: Download the Installation File by using the commands below in the directory where the application should be installed

```
None
wget
https://s3.us-east-1.amazonaws.com/erwin-2.0/code/linux/application-code/3.1.0/On_premise/Packages_3.1.0/Erwin-Redhat-Packages/Erwin-installer.7za
```

Step 6: Extract the Installation File by using the commands below:

```
None
sudo 7za x Erwin-installer.7za
```

Step 7: Remove the redundant zip file after extraction

```
None
sudo rm -rf Erwin-installer.7za
```

Configuration Setup

Step 1: Navigate to the Pre-requisites directory

None

```
#If the application is installed on home directory
cd prerequisites/

#If the application is installed on a custom directory
cd <custom_directory>/prerequisites/
```

Ensure that the `config.txt` and `Erwin-installer.sh` are in the same directory in the prerequisites folder.

```
dqlabs@ubuntuarun:~$ cd prerequisites/
dqlabs@ubuntuarun:~/prerequisites$ ls
Erwin-installer.sh  config.txt
```

Step 2: Edit the `config.txt` file using the command below:

None

```
sudo vi config.txt
```

Step 3: Update the `config.txt` file by clicking “**I**” to get into insert mode

Details about the config file

- 1. **LOCAL_REPO=yes**
- 2. **Source and Destination Locations:** Please verify that the **SOURCE_LOCATION** path specified in the `config.txt` is accurate and ensure the installation file is downloaded correctly to that location.
 - a. **SOURCE_LOCATION:** Define the source location path.
Example - (/home/ubuntu/Dqlabs-installer)
 - b. **DESTINATION_LOCATION:** Define the destination location path.
Example - (/usr/src/Dqlabs)
- 3. **DQLabs Access Information:**
 - a. **DNS_CONFIGURE:** yes or no
 - i. If yes, then update the following:
 - 1. **DNS_NAME:** e.g., <http://foo.subdomain.com>
 - 2. **SSL_PROTOCOL:** http or https
 - ii. If no, then update the following:
 - 1. **ACCESS_MODE:** public or private
 - 2. **SSL_PROTOCOL:** http or https
- 4. **Administrator Information:**
 - a. **ADMIN_EMAIL:** Administrator's email address (Ensure valid email format)
Example - (admin@dqlabs.ai)
 - b. **ADMIN_PASSWORD:** Administrator's password
Example - (Dql@b\$)

5. PostgreSQL Configuration:

- i. **PG_USERNAME=<username>**
- ii. **PG_PASSWORD=<password>**
- iii. **PG_DB_NAME=<dbname>**
- iv. **AIRFLOW_USERNAME=<username>**
- v. **AIRFLOW_PASSWORD=<password>**
- vi. **AIRFLOW_DB_NAME=<dbname>**

Do not use special characters and spaces in your username, dbname, and passwords

- 6. **Installation Options(Mandatory):**
 - a. **FULL_INSTALLATION:**
 - i. **yes** -> Install DQLabs Application along with Postgres database.
 - ii. **no** -> Install only the Postgres database.

- iii. **dqlabs** -> Install DQLabs without a Postgres database.

In case of Postgres Installation on the secondary server, provide the input as no

7. Dual server Deployment (Update only in case of db-separation deployment and if the input to the 9th variable is **dqlabs**)

- a. **PG_PORT_NO**=<port no>
- b. **PG_HOST**=<postgres server IP>

8. Installation Options:

- a. **INSTALL_SPARK_LIVY**= yes or no

```
##### Dqlabs #####

# Do you want to install packages from Local Repo (yes) or RHEL Repo (no)
LOCAL_REPO=yes

# Please provide the Dq-installer file path as source location (example: /home/ec2-user/Dq-installer)
SOURCE_LOCATION=

# Please provide the destination location (example: /usr/src/Dqlabs)
DESTINATION_LOCATION=

# DQLabs access Information

# DNS configuration:
# If a DNS name is available, set DNS_CONFIGURE to yes and provide the DNS value below.
# Otherwise, leave DNS_CONFIGURE as no.
DNS_CONFIGURE=no
DNS_NAME=

# Specify the access mode:
# For public IP, set ACCESS_MODE to 'public'.
# For private IP, set ACCESS_MODE to 'private'.
ACCESS_MODE=

# Specify the protocol to use (mandatory): http or https.
SSL_PROTOCOL=

# Login Inputs:
# Provide the administrator email and password for DQLabs access.
ADMIN_EMAIL=
ADMIN_PASSWORD=

# Postgres credentials to be created and used during DQLabs installation.
PG_USERNAME=
PG_PASSWORD=
PG_DB_NAME=

# Airflow credentials and database details.
AIRFLOW_USERNAME=
AIRFLOW_PASSWORD=
AIRFLOW_DB_NAME=

# Installation options (mandatory):
# Enter one of the following values:
# "yes" -> Install DQLabs Application along with Postgres database.
# "no" -> Install only the Postgres database.
# "dqlabs" -> Install DQLabs without Postgres database.
FULL_INSTALLATION=

### (Below variables are applicable if FULL_INSTALLATION is set to "dqlabs")
PG_PORT_NO=
PG_HOST=

### If you want to install spark and livy in this machine (yes) or (no)
INSTALL_SPARK_LIVY=

~
```

Step 4: Go to command mode by pressing the escape key. Save and exit the editor using **:wq!** and press Enter

Installation Process

Step 1: Set Permissions and execute the script using the commands below:

```
None

#Navigate to the prerequisites directory
cd prerequisites

#Grant permissions to the script
sudo chmod 777 Erwin-installer.sh

#Execute the setup script
./Erwin-installer.sh
```

Users without password-based authentication should not execute the script.

Verification of Installation

Once the installation process is complete, verifying that the DQLabs application and related services are running successfully on your Redhat machine is important. Perform the following verifications:

Step 1: Ensure the following output is shown at the end of the script execution:

```
airflow-webserver is running.
airflow-scheduler is running.
apache2 is running.
dqlabs.service is running.
livy.service is running.
true
*****
All services are running. DQ installation completed successfully!
*****
/home/dqlabs/DQ/dqlabsenv/lib/python3.11/site-packages/drf_yasg/___init___py:2: UserWarning: pkg_resources is deprecated as an API. See https://setuptools.pypa.io/en/latest/pkg_resources.html. The pkg_resour
ces package is slated for removal as early as 2025-11-30. Refrain from using this package or pin to Setuptools<81.
  from pkg_resources import DistributionNotFound, get_distribution
***** Airflow Token *** YWRtaW46YWRtaW4=
default
dqlabs@m2ubuntu310Freshinstall:~$
```

Note: The execution log will be stored on the server and can be accessed using the commands below:

```
None

#Navigate to prerequisites directory
cd prerequisites

#View the logs
vi script.log
```

Patch Process

Step 1: Download the Patch script into the DQ Server under the same user with which DQ was installed

```
None

wget
https://s3.us-east-1.amazonaws.com/erwin-2.0/code/linux/application-code/3.1.0/On_premise/erwin_v310_patchscript.sh
```

Step 2: Open the script file

```
None

sudo vi erwin_v310_patchscript.sh
```

Step 3: Update the variables

```
None

#Provide the directory path where the script is downloaded
PATCH_DIR="/home/dqlabs/"

#Provide the path to DQLabs-server folder
TARGET_DIR="/home/dqlabs/dq310/DQLabs-Server"

#Provide the destination directory of the application (Same as TARGET_DIR, exclude DQLabs-Server)
PARENT_DIR="/home/dqlabs/dq310"
```

Step 4: Save the file and exit

```
None

:wq!
```

Step 5: Grant permission to execute the script

```
None

sudo chmod 777 erwin_v310_patchscript.sh
```

Step 6: Execute the script

```
None

./erwin_v310_patchscript.sh
```

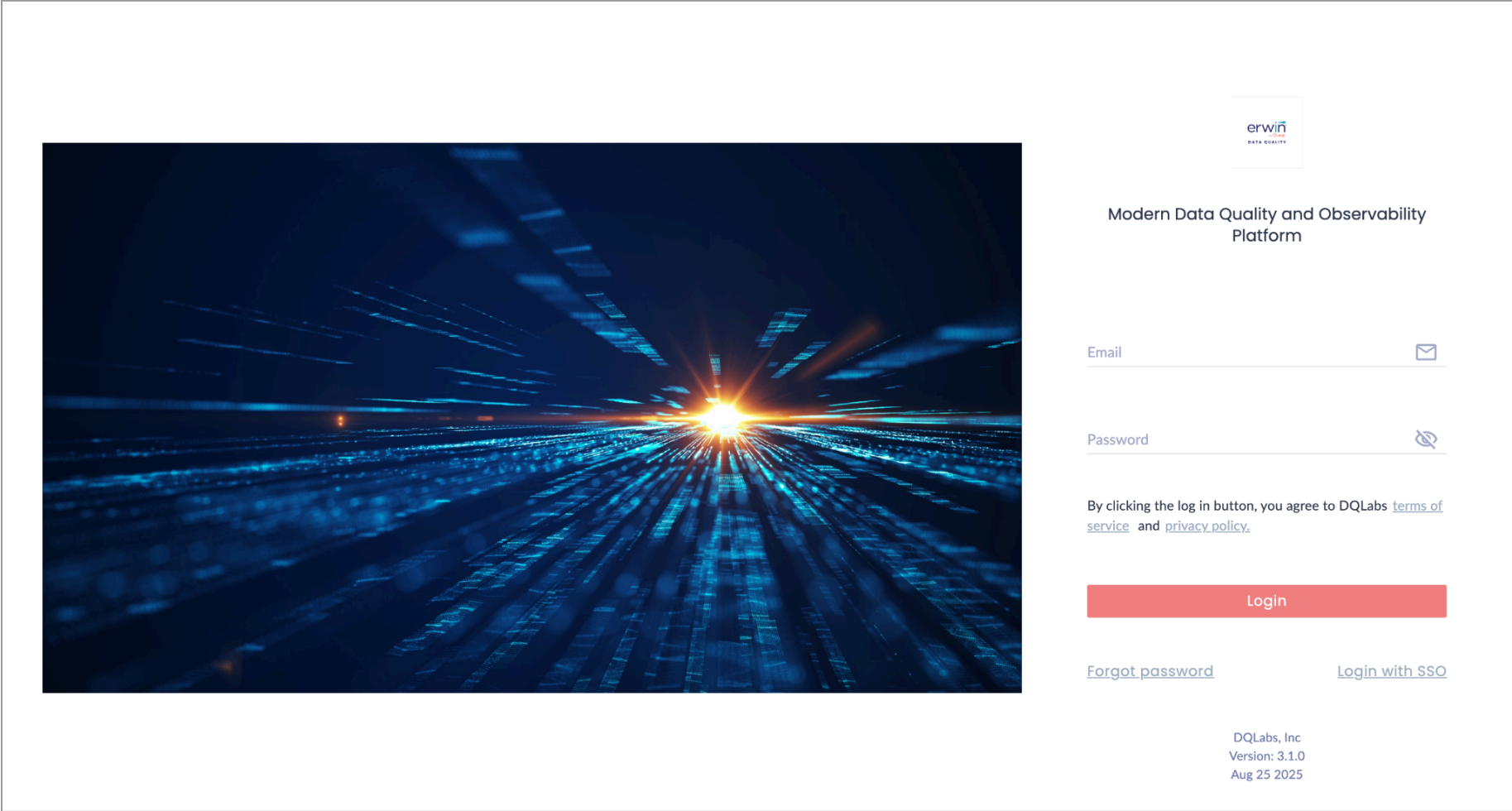

Post-Installation Procedure

After completing the installation and verifying the successful setup of DQLabs on your Redhat machine, you can now log in to the application and proceed with the final steps. Follow the instructions below:

Step 1: Launch any supported web browser on your machine

Step 2: In the address bar of the browser, enter the IP address or DNS name used during installation

Step 3: The browser will load the DQLabs application, and you will be presented with the login page



Step 4: After logging in to the application, you will be prompted to add the license. Once the license has been activated, the platform is ready to use.

