PurpleCube AI®

30-DAY FREE TRIAL

Your Step-to-Step Guide to Downloading PurpleCube AI's Free Trial On-Premise

Installation Guide

Table of Contents

1. Introduction

1.1. Purpose of the Document

This document serves as a comprehensive guide to installing PurpleCube AI's 30-day Free Trial.

1.2. End Users

This document is designed for data scientists, data engineers, data architects, data executives, and organizations interested in installing PurpleCube AI's free trial on-premises.

2. About PurpleCube AI

2.1. Introduction

PurpleCube AI is a unified data orchestration platform on a mission to revolutionize data engineering with the power of Generative AI. This unique approach enables us to automate complex data pipelines, optimize data flows, and generate valuable insights cost-effectively and with efficiency and accuracy.

PurpleCube AI's unified data orchestration platform is your key to:

- **Unify** all data and data engineering functions on a single platform with real-time Gen AI assistance.
- **Automate** complex data pipelines by provisioning data sets with comprehensive metadata and governance for optimal business use.
- **Activate** all kinds of analytics, including English Language Queries and Exploratory Data Analytics.

Beyond traditional data lake and warehouse automation, PurpleCube AI leverages the power of language models to unlock a variety of innovative use cases. This includes conducting exploratory data analysis and natural language queries, automating metadata generation and enrichment, enhancing data quality assessment, and optimizing data governance through relationship modeling.

2.2. Unified Data Orchestration Platform Features

Today, multiple platforms are required to take care of a variety of data movement and transformation activities, creating wasted time, money, and resources. Every organization is doing data replication, data integration, API integration, big data integration, cloud data integration, streaming data management, data pipeline management, data orchestration, and data preparation.

Below are some of the capabilities that make PurpleCube AI's unified data orchestration platform a perfect choice for data engineers, data scientists, data architects, and data executives:

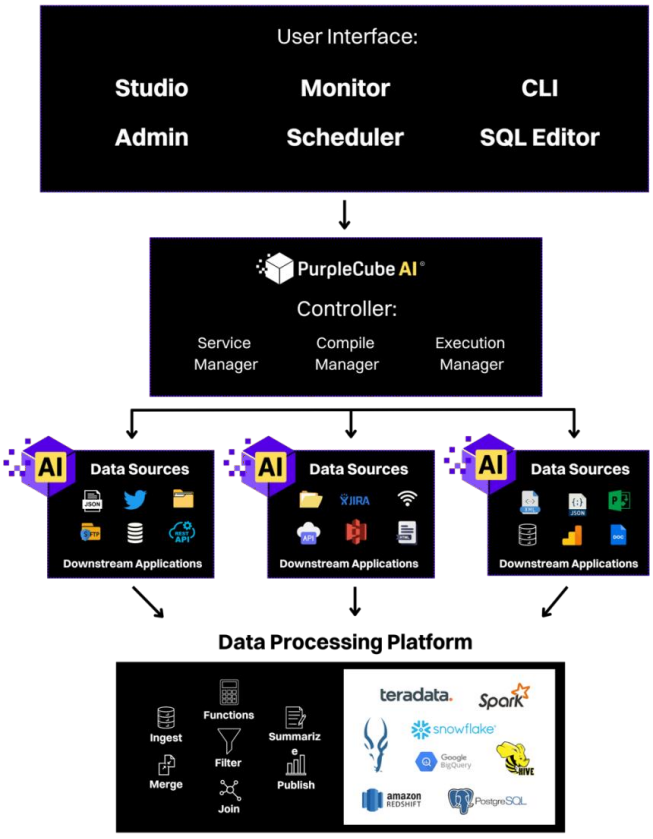
- Maximize Data Engineering Asset Reusability
- Automate End-to-End Data Pipelines
- Deploy AI with Confidence and Scale
- Unlock Exponential Productivity with GenAI
- Gain Complete Visibility into Data Governance and Security
- Deliver Consistently Trustworthy Data Across Your Organization
- Rapidly Build and Deploy End-to-End Data Pipelines
- **Supercharge Data Engineering Productivity**

2.3. Architecture

Leveraging PurpleCube AI's Generative Artificial Intelligence (GenAI) for querying, data professionals are empowered to uncover subtle patterns from vast datasets, refining exploration methodologies for contextually relevant insights and positioning themselves at the forefront of data-driven innovation.

This dynamic interaction, underpinned by advanced algorithms, bridges the gap between raw data and actionable intelligence, ensuring optimized decision-making and a competitive edge in a data-centric landscape.

The technical architecture of our solutions is designed to be robust, scalable, and secure.



3. Pre-Installation Requirements

To install PurpleCube AI's free trial, servers must meet the minimum recommended requirements. These include specifications for server hardware, ports, user, repository database, and environment variables.

3.1. System Requirements

- ### Minimum Hardware Requirement for Controller

HW Specification	***Recommended***
CPU	2 x Intel(R) Xeon(R) CPU E5-2620 or Higher @2.00GHz
Cores	Quad or Higher
Memory	32 GB or Higher
Disk Space	Minimum 100 GB
Operating System	RHEL 9, Ubuntu 20.4 or higher
Software	Java 1.8x
Repository	PostgreSQL 15.x. (Ignore this requirement, if the installation is using embedded repository option)
DB Space for Repository	20 GB or higher. (Ignore this requirement, if the installation is using embedded repository option)

Port Requirement for Controller

Default Port No.	***Description***	***Network Firewall requirement***
9000	Controller Startup Port – The port on which the Controller service is running and accepts connections from all client modules.	This port should be open in network to accept connections from intranet or internet
9001	Controller Shutdown Port – The port through which Controller service shuts down.	This port need not be open as it is used internally

9002	Controller Messaging port/Controller broker startup Port – Port on which controller broker service will be running. This port is used to communicate between Agent and Controller.	This port should be opened in the network to accept connections from Agents.
9003	Embedded PostgreSQL Database port - Port on which the metadata database is running	This port need not be opened over the network.
1098	Broker Connector Port - Port used to connect to broker active MQ jmx	This port need not be opened over the network.
9005	Default Agent service Start Port - The port on which agent service is running.	This port need not be opened over the network.
9007	Data receiver port for Default Agent	This port needs to be opened only if the Default Agent needs to receive data from another agent.
9010	Default Agent service Stop Port - The port through which agent service shuts down.	This port need not be opened over the network.

▪ Minimum Hardware Requirement for Agent

HW Specification	***Recommended***
CPU	2 x Intel(R) Xeon(R) CPU E5-2620 or Higher @2.00GHz
Cores	Quad or Higher
Memory	32 GB or Higher
Disk Space	Minimum 100 GB *(Note: It should be twice the data being extracted for each batch)*
Operating System	RHEL 9, Ubuntu 14 or higher, Windows (64-bit)
Software	Java 1.8x

Port Requirement for Agent

Default Port Number	***Network Firewall requirement***	***Description***
9005	This port need not be opened over the network.	Agent service start port - The port on which agent service is running.
9007	This port needs to be opened only if the optional Agent needs to receive data from another agent.	Data receiver port
9010	This port need not be opened over the network.	Agent service stop port - The port through which agent service shuts down.

▪ Minimum Hardware Requirement for PostgreSQL Extension

HW Specification	***Recommended***
CPU	2 x Intel(R) Xeon(R) CPU E5-2620 or Higher @2.00GHz
Cores	Dual or Higher
Memory	8 GB or Higher
Disk Space	Minimum 20 GB
Operating System	RHEL 9, Ubuntu 20.4 or higher
Software	Java 1.8x
Version	PostgreSQL 15.x

Port Requirement for PostgreSQL Extension

Default Port Number	**Description**	**Network Firewall requirement**
5432	PostgreSQL Database connectivity port	This should be open in network to accept connections from controller

3.2. User Setup

The PurpleCube AI user environment should be set-up first. This article explains how to set up Linux environments for PurpleCube AI installation.

On Linux systems, PurpleCube AI components should be installed by creating a dedicated user account. The dedicated user account should be used for installing, running configurations, and maintaining updates of PurpleCube AI components.

1. Create a directory "app" under root directory.

- *mkdir /app*

2. Create the group and add user to the group.

- *groupadd purplecubev*
- *useradd -d /app/dictrl -g purplecube dictrl*

3. Set the password for username.

- *passwd dictrl*

Note: PurpleCube AI requires a dedicated Linux user account from where server components can be installed and managed.

Reference: User Setup

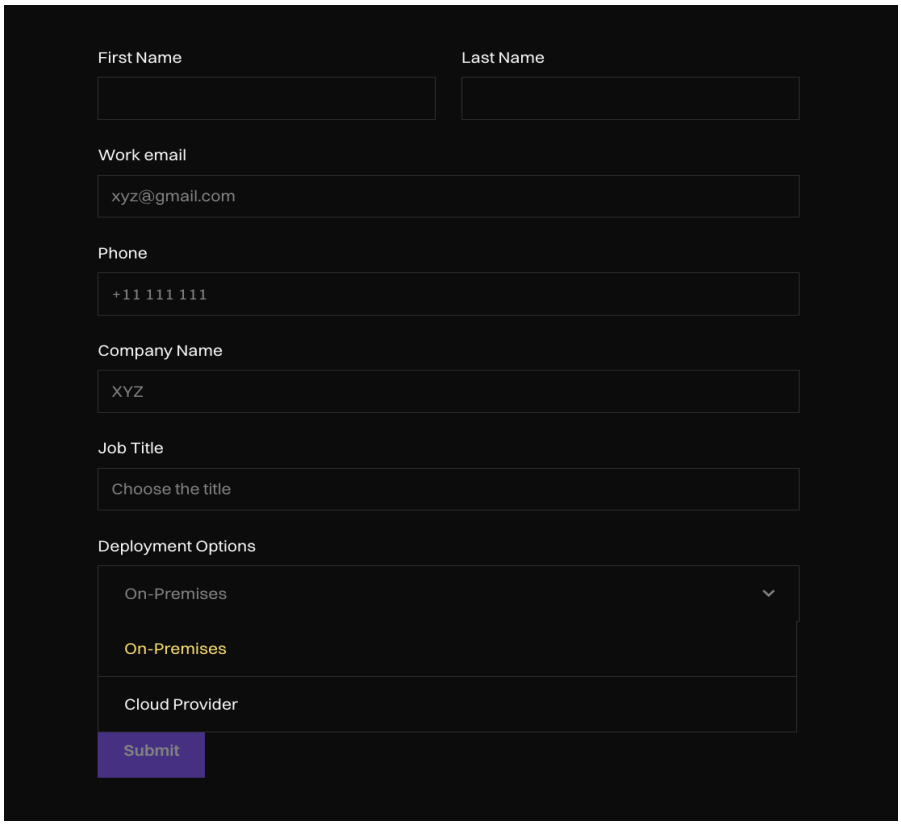
4. Installing PurpleCube AI Platform Free Trial

This section will walk you through the steps to install PurpleCube AI's free trial. All commands are executed on a Linux system for the Controller installation and upgrade. The Agent installation and upgrade process may vary slightly depending on the operating system.

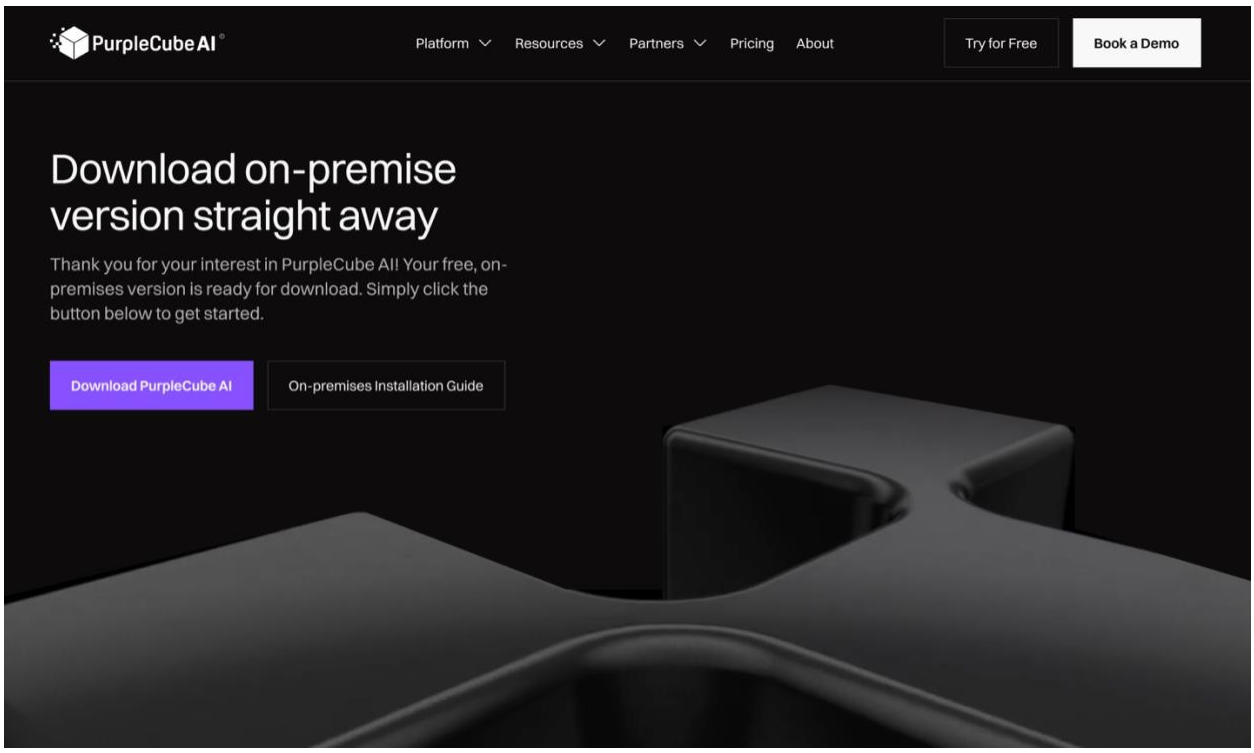
4.1. Downloading PurpleCube AI's Free Trial

To download PurpleCube AI Free Trial, follow the simple instructions below:

- Open https://www.purplecube.ai in your preferred browser
- Click the 'Try for Free' button
- The following form will appear. Enter your details and submit the form:



- While entering the details, you will have to select your Deployment Preference, either Cloud Provider or On-Premises.
- If you select On-Premises, the following window will appear:



- Click the Download PurpleCube AI button, and the installation will begin straightaway on- premises.
- Click the on-premises Installation Guide button to go through the installation steps in detail.

4.2. Interactive Installation of Controller with Default Agent

Login to the Server with Installer User ID

Download the Installer with wget in Linux

Controller Installer Download (It comes with Default Agent):

- *wget https://downloads.purplecube.ai/Controllers/5.8.0.0000.000/Purplecube_Controller_5.8.0.0000.000_linux-x64_installer.run --no-check-certificate*

Provide execute permission to the installer file

- *chmod 775 Purplecube_Controller_5.8.0.0000.000_linux-x64_installer.run*

Run Installer

- *./Purplecube_Controller_5.8.0.0000.000_linux-x64_installer.run*

Provide necessary details for Purplecube home path, components to be installed with respective ports, and database details.

Note: Open all mentioned ports as inbound/outbound. It is recommended to open a range of **9000-9040** as per configuration in Controller and Default Agent.

Reference: Interactive Installation of Controller

Verify PURPLCUBE HOME variables path and run Profile

Check variables:

- *cat ~/.bashrc*

Check variable paths:

- *echo $DIYOTTA_HOME and $DI_HOME*

Run bash profile:

- *source ~/.bashrc*

Start Controller:

Navigate to Controller path:

- *cd <PURPLCUBE_HOME>/controller/bin/*
- *./serverstartup.sh*

Verify if Controller and Postgres Database (Metadata) processes are running:

- *ps -ef | grep controller*
- *ps -ef | grep postgres*

Reference: Starting Purplecube Controller

Start Agent:

Navigate to Agent path:

- *cd <PURPLCUBE_HOME>/agent/bin/*
- *./agentstartup.sh*

Verify if Agent process is running:

- *ps -ef | grep agent*

Note: Before starting Agent, you need to apply for the license first.

Reference: Starting Purplecube Agent

Optional Steps:

Check Local Host Metadata Database Login

Navigate to:

- *cd <PURPLCUBE_HOME>/diyotta/controller-repo/bin*
- *./psql -h localhost -p 9003 -U direpusr -d direpdb*

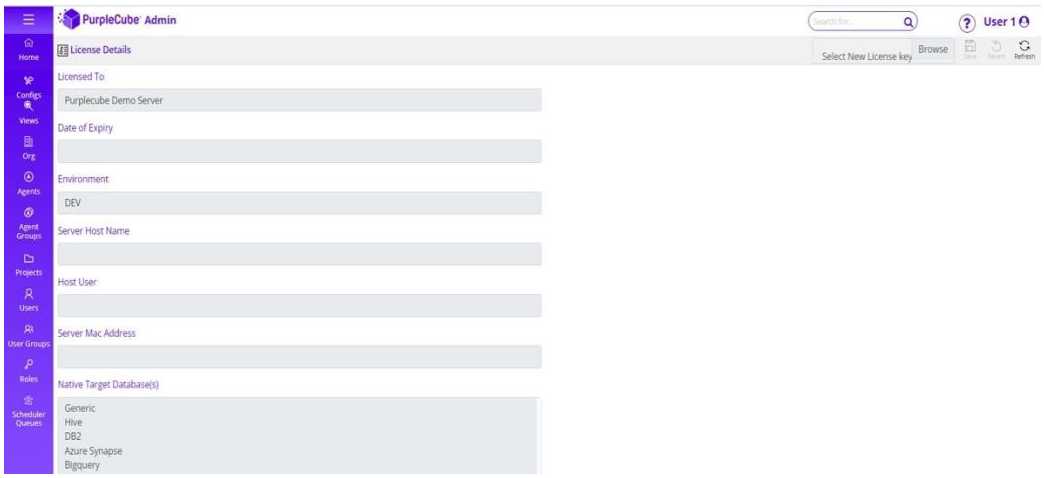
Note: If you face an error, run the following command and retry database login:

- *export LD_LIBRARY_PATH=/home/sysadmin/diyotta/controller-repo/lib:$LD_LIBRARY_PATH*

5. Post-Installation

5.1. Apply for License

- **Generate License:** After installation, we must generate a license to apply. To do so, you have to contact our experts at support@purplecube.ai with specific details about the connector and system.
- **Apply License:** After getting the license file, you must apply using the created URL.



5.2. Testing the Installation

- Check if the URL https://hostname:port/#/session/login is working.
  Ex: https://demo.purplecube.ai/#/session/login
- Contact to support@purplecube.ai if any error arises.

5.3. Common Issues and Troubleshooting

- **Issue:** The URL and Services are not accessible or running because ports have not been opened.
- **Resolution:** You should open all ports mentioned in the Controller and Agent configuration.
- **Issue:** Not able to login after successfully installing and service running fine due to Postgres
- **Resolution:** Check if the DB Jars are up to date.
- **Issue:** Service up failed
- **Resolution:** Check metadata file permissions.

If you have any other errors, please contact us support@purplecube.ai.

6. Appendices

6.1. Glossary of Terms

- **Data Orchestration**: The process of managing and coordinating the flow of data across various systems and applications.
- **Revolutionize**: To fundamentally change something in a way that significantly improves or transforms it.
- **Data Engineering**: The practice of designing and building systems for collecting, storing, and analyzing data.
- **Data Flows**: The movement of data between different systems, processes, or components.
- **Data Pipelines**: Automated systems that transport data from one place to another, often for processing or analysis.
- **Data Lakes**: Large storage repositories that hold vast amounts of raw data in their native format.
- **Data Replication**: The process of copying data from one location to another to ensure consistency and reliability.
- **Data Integration**: Combining data from different sources to provide a unified view or dataset.
- **Controller**: In computing, a controller is a hardware device or software program that manages or directs the flow of data between other devices or systems.
- **Agent:** A lightweight Java-based component that executes Controller instructions on target systems and returns results securely.
- **Broker:** A Java-based bridge that manages encrypted message exchange between the Controller and Agents through queues.
- **Metadata Repository:** A relational database that stores process definitions, configurations, and runtime metadata for PurpleCube AI.