INNOMATICS®
RESEARCH LABS
INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Workflow Orchestration using Prefect

Objective of the project

The key aims of this project include:

1. Data Preparation and Management
2. Model Training
3. Experimentation and Evaluation
4. Model Deployment
5. Monitor and Management
6. Automation of repetitive tasks

Introduction

Orchestration refers to the coordination and management of various tasks, resources, and processes involved in the end-to-end machine learning lifecycle.
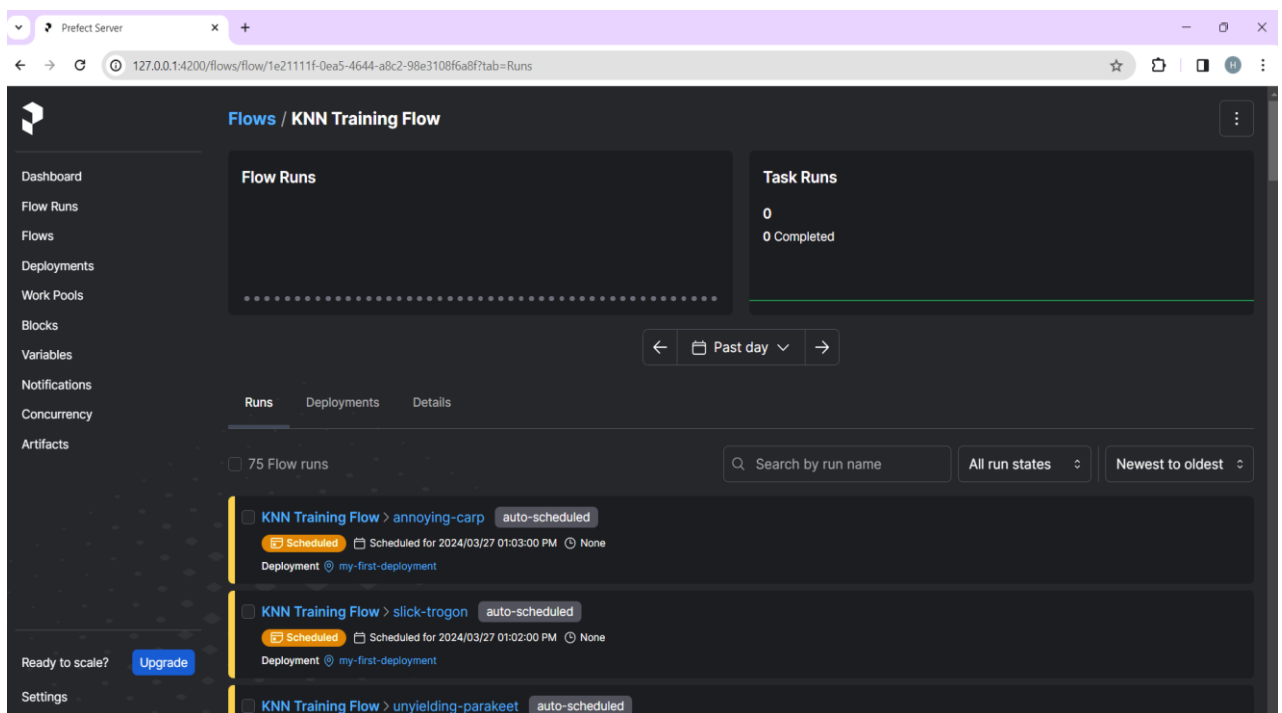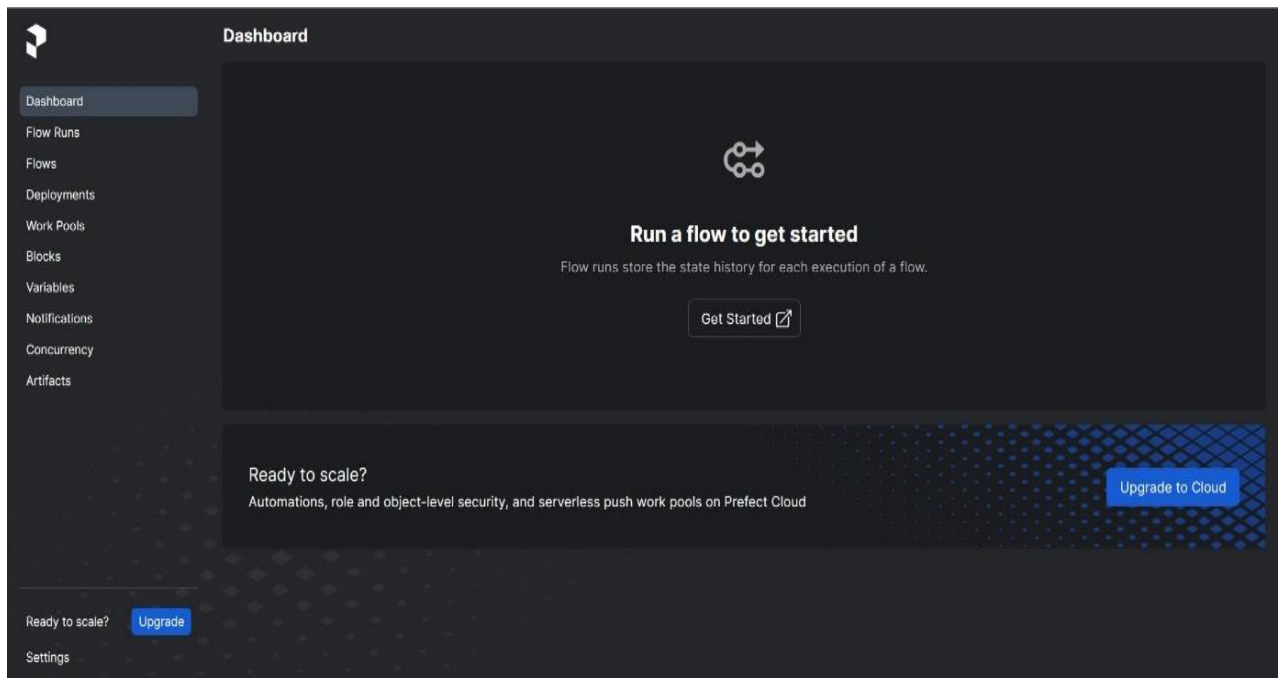
Prefect is an open-source orchestration and observability platform that empowers developers to build and scale resilient code quickly, turning their Python scripts into resilient, recurring workflows.

Prefect streamlines the orchestration of machine learning workflows by providing a flexible, scalable, and reliable framework for building, deploying, and managing complex data pipelines with ease. It empowers data scientists and engineers to focus on building machine learning models and solving business problems while abstracting away the complexities of workflow management and execution.

Building a Prefect Workflow

Run below commands to install prefect on your system:

1. Create Virtual Environment: **python -m venv .env**
2. Activate Virtual Environment: **.env\Scripts\activate**
3. Install Prefect: **pip install prefect**
4. Install Setuptools (Optional):   **pip install setuptools**
5. To open Prefect Dashboard: **prefect server start**

Dashboard

Run a flow to get started

Flow runs store the state history for each execution of a flow.

Get Started

Ready to scale?
Automations, role and object-level security, and serverless push work pools on Prefect Cloud

Upgrade to Cloud

- Dashboard
- Flow Runs
- Flows
- Deployments
- Work Pools
- Blocks
- Variables
- Notifications
- Concurrency
- Artifacts

Ready to scale? Upgrade

Settings

Prefect Server

127.0.0.1:4200/flows/flow/1e21111f-0ea5-4644-a8c2-98e3108f6a8f?tab=Runs

Flows / KNN Training Flow

Flow Runs

Task Runs

0
0 Completed

← Past day ∨ →

Runs   Deployments   Details

☐ 75 Flow runs   Search by run name   All run states ⇅   Newest to oldest ⇅

☐ **KNN Training Flow** > annoying-carp   auto-scheduled
⏱ Scheduled   🗓 Scheduled for 2024/03/27 01:03:00 PM   🕐 None
Deployment ⊚ my-first-deployment

☐ **KNN Training Flow** > slick-trogon   auto-scheduled
⏱ Scheduled   🗓 Scheduled for 2024/03/27 01:02:00 PM   🕐 None
Deployment ⊚ my-first-deployment

☐ **KNN Training Flow** > unyielding-parakeet   auto-scheduled

- Dashboard
- Flow Runs
- Flows
- Deployments
- Work Pools
- Blocks
- Variables
- Notifications
- Concurrency
- Artifacts

Ready to scale? Upgrade

Settings