

10. Pipeline Orchestration

10.1 Orchestration Approach

The end-to-end data and machine learning pipeline is orchestrated using **Prefect**, a Python-native workflow orchestration framework that runs seamlessly on Windows and Linux environments.

Prefect is used to:

- Automate pipeline execution
- Define task dependencies
- Provide task-level logging and retries
- Monitor pipeline execution status

This orchestration layer ensures reliable and repeatable execution of the complete recommendation system pipeline.

10.2 Pipeline Workflow

The pipeline is executed as a single flow with the following sequence of stages:

Data Ingestion

- Data Validation
- Data Preparation and EDA
- Feature Engineering
- Feature Store Access
- Model Training
- Model Evaluation

Each stage corresponds to an independent task implemented in earlier pipeline steps.

10.3 Orchestration Implementation

The pipeline is implemented as a Prefect flow with multiple tasks, where each task invokes an existing processing script.

Flow Definition

`src/orchestration/prefect_flow.py`

The flow coordinates execution of:

- Clickstream and product data ingestion
 - Data validation checks
 - Data preparation and exploratory analysis
 - Feature engineering and storage in PostgreSQL
 - Feature store access demonstration
 - Model training and evaluation
-

10.4 Logging and Monitoring

Prefect provides built-in logging and monitoring capabilities:

- Each task logs execution start and completion status
- Failures are captured and retried automatically
- Downstream tasks are blocked on upstream task failures
- Execution status can be monitored via the Prefect UI or terminal logs

This enables effective observability and failure handling across the pipeline.

10.5 Execution and Verification

The pipeline is executed using:

```
python -m src.orchestration.prefect_flow
```

Successful execution is verified through:

- Console logs indicating completion of each task
- Prefect UI showing successful flow runs and task statuses

Screenshots or logs from successful executions are included as evidence of orchestration.

Flows / dm4ml_recommendation_pipeline

Flow Runs

2 total

Task Runs

8

8 Completed 100%

Runs Deployments Details

2 Flow runs

Search by run name

All run states

Newest to oldest

dm4ml_recommendation_pipeline > aquamarine-albatross

Completed 2026/01/20 02:25:11 PM 0 Parameters 13s 8 Task runs

dm4ml_recommendation_pipeline > judicious-seal

Failed 2026/01/20 02:21:53 PM 0 Parameters 30s 1 Task run

Items per page 10

Ready to scale? Upgrade

Join the Community

Settings

This screenshot shows the 'Flows / dm4ml_recommendation_pipeline' page. It displays two flow runs: one completed successfully and one that failed. The completed run was executed on January 20, 2026, at 02:25:11 PM, took 13 seconds, and had 8 task runs. The failed run was executed on the same day at 02:21:53 PM, took 30 seconds, and had 1 task run. The interface includes filters for 'Flow runs', a search bar, and sorting options ('Newest to oldest'). A sidebar on the left provides links to various Prefect features like Dashboard, Runs, Flows, and Automations.

Dashboard

Flow Runs

2 total

Task Runs

8

8 Completed 100%

Hide subflows All tags Past day

Active Work Pools

There are no active work pools to show. Any work pools you do have are paused.

View all work pools

dm4ml_recommendation_pipeline

4m 6s ago

1 ✓

Ready to scale?

Webhooks, role and object-level security, and serverless push work pools on Prefect Cloud

Upgrade to Cloud

Ready to scale? Upgrade

Join the Community

Settings

This screenshot shows the main 'Dashboard' page. It highlights the 'Flow Runs' section with 2 total runs, 8 task runs, and 100% completion. Below this, there's a message stating that no active work pools are shown, and a link to view all work pools. The sidebar on the left is identical to the previous screenshot, showing navigation links for various Prefect components.

```

prefect_flow.py 2, U pipeline.log M
logs > pipeline.log
1 2026-01-20 14:35:39,049 | INFO | prefect | Starting temporary server on http://127.0.0.1:8417
2 See https://docs.prefect.io/v3/concepts/server/how-to-guides for more information on running a dedicated Prefect server.
3 2026-01-20 14:35:39,049 | INFO | prefect | Starting temporary server on http://127.0.0.1:8417
4 See https://docs.prefect.io/v3/concepts/server/how-to-guides for more information on running a dedicated Prefect server.
5 2026-01-20 14:35:38,383 | INFO | httpx | HTTP Request: GET http://127.0.0.1:8417/api/admin/version "HTTP/1.1 200 OK"
6 2026-01-20 14:35:38,529 | INFO | httpx | HTTP Request: GET http://127.0.0.1:8417/api/admin/health "HTTP/1.1 200 OK"
7 2026-01-20 14:35:38,536 | INFO | httpx | HTTP Request: GET http://127.0.0.1:8417/api/carf-token?client=f2f25860-732c-40aa-b64e-946ef9e02373 "HTTP/1.1 422 Unprocessable Entity"
8 2026-01-20 14:35:38,596 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/flows/ "HTTP/1.1 201 Created"
9 2026-01-20 14:35:38,604 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/flow\_runs/ "HTTP/1.1 201 Created"
10 2026-01-20 14:35:38,618 | INFO | prefect:flow_runs | Beginning flow run 'esoteric-squirrel' for flow 'dm4ml_recommendation_pipeline'
11 2026-01-20 14:35:38,618 | INFO | prefect:flow_runs | Beginning flow run 'esoteric-squirrel' for flow 'dm4ml_recommendation_pipeline'
12 2026-01-20 14:35:38,618 | INFO | prefect:task_runs | Clickstream ingestion completed
13 2026-01-20 14:35:38,618 | INFO | prefect:task_runs | Clickstream ingestion completed
14 2026-01-20 14:35:38,891 | INFO | prefect:task_runs | Clickstream ingestion completed
15 2026-01-20 14:35:38,891 | INFO | prefect:task_runs | Clickstream ingestion completed
16 2026-01-20 14:35:38,891 | INFO | prefect:task_runs | Clickstream ingestion completed
17 2026-01-20 14:35:38,891 | INFO | prefect:task_runs | Clickstream ingestion completed
18 2026-01-20 14:35:38,901 | INFO | httpx | HTTP Request: GET http://127.0.0.1:8417/api/flows/ca1a7bbf-283f-4d2d-ac67-20f92b7bda4f "HTTP/1.1 200 OK"
19 2026-01-20 14:35:39,542 | INFO | prefect:task_runs | Product ingestion completed
20 2026-01-20 14:35:39,542 | INFO | prefect:task_runs | Product ingestion completed
21 2026-01-20 14:35:39,545 | INFO | prefect:task_runs | Finished in state Completed()
22 2026-01-20 14:35:39,545 | INFO | prefect:task_runs | Finished in state Completed()
23 2026-01-20 14:35:40,133 | INFO | prefect:task_runs | Data validation completed
24 2026-01-20 14:35:40,133 | INFO | prefect:task_runs | Data validation completed
25 2026-01-20 14:35:40,195 | INFO | prefect:task_runs | Finished in state Completed()
26 2026-01-20 14:35:40,195 | INFO | prefect:task_runs | Finished in state Completed()
27 2026-01-20 14:35:40,762 | INFO | httpx | HTTP Request: GET http://127.0.0.1:8417/api/csrf-token?client=2db7f49b-ed4a-4514-b044-0124a101e04b "HTTP/1.1 422 Unprocessable Entity"
28 2026-01-20 14:35:40,767 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/logs/ "HTTP/1.1 201 Created"
29 2026-01-20 14:35:41,731 | INFO | prefect:task_runs | Data preparation completed
30 2026-01-20 14:35:41,731 | INFO | prefect:task_runs | Data preparation completed
31 2026-01-20 14:35:41,733 | INFO | prefect:task_runs | Data preparation completed
32 2026-01-20 14:35:41,733 | INFO | prefect:task_runs | Finished in state Completed()
33 2026-01-20 14:35:42,559 | INFO | prefect:task_runs | Feature engineering completed
34 2026-01-20 14:35:42,559 | INFO | prefect:task_runs | Feature engineering completed
35 2026-01-20 14:35:42,561 | INFO | prefect:task_runs | Finished in state Completed()
36 2026-01-20 14:35:42,561 | INFO | prefect:task_runs | Finished in state Completed()
37 2026-01-20 14:35:42,782 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/logs/ "HTTP/1.1 201 Created"
38 2026-01-20 14:35:43,216 | INFO | prefect:task_runs | Feature store retrieval demo completed
39 2026-01-20 14:35:43,216 | INFO | prefect:task_runs | Feature store retrieval demo completed
40 2026-01-20 14:35:43,218 | INFO | prefect:task_runs | Finished in state Completed()
41 2026-01-20 14:35:43,218 | INFO | prefect:task_runs | Finished in state Completed()
42 2026-01-20 14:35:44,793 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/logs/ "HTTP/1.1 201 Created"
43 2026-01-20 14:35:45,829 | INFO | prefect:task_runs | Model training completed
44 2026-01-20 14:35:45,829 | INFO | prefect:task_runs | Model training completed
45 2026-01-20 14:35:45,830 | INFO | prefect:task_runs | Finished in state Completed()
46 2026-01-20 14:35:45,830 | INFO | prefect:task_runs | Finished in state Completed()
47 2026-01-20 14:35:46,818 | INFO | httpx | HTTP Request: POST http://127.0.0.1:8417/api/logs/ "HTTP/1.1 201 Created"
48 2026-01-20 14:35:46,818 | INFO | prefect:task_runs | Model evaluation completed

```

10.6 Summary

The Prefect-based orchestration layer automates the entire data and machine learning workflow, ensuring structured execution, observability, and fault tolerance. This completes the end-to-end pipeline implementation for the recommendation system.