**Tài Liệu Dự Án MLOps Platform**

[I. Giới thiệu 2](#_Toc2065766598)

[1. Tổng quan 3](#_Toc630757857)

[2. Mục tiêu 3](#_Toc831219756)

[3. Skateholder 3](#_Toc1257630740)

[II. Timeline 4](#_Toc1433490476)

[1. Thiết kế kiến trúc và thử nghiệm các nền tảng (12/7 - 31/9): 4](#_Toc138368940)

[2. Xây dựng platform trên hệ thống OpenShift (1/10 - 31/11): 4](#_Toc1476943796)

[3. Xây dựng công cụ cho ETL dữ liệu từ các nguồn vào Feature Store (1/12 - 31/12): 5](#_Toc729594963)

[III. Chức năng 5](#_Toc1220457544)

[IV. Kiến trúc hệ thống 5](#_Toc2132322451)

[1. Kiến trúc High-Level 5](#_Toc1757901918)

[2. Kiến trúc chi tiết 6](#_Toc1010964672)

[V. MLOps Platform 7](#_Toc1964553200)

[1. Dashboard MLOps Platform 7](#_Toc385597837)

[2. Feature Store 7](#_Toc1782952168)

[3. Giám sát quá trình training mô hình 8](#_Toc1748640419)

[4. Quản lý API endpoints 9](#_Toc878096215)

[5. Pipeline và theo dõi kết quả 10](#_Toc1273862260)

[6. Tinh chỉnh resource cho API và các job bằng giao diện và tự động 11](#_Toc1059494973)

[7. Quản lý job ETL 11](#_Toc248796744)

[8. Thực thi các job ETL thông qua Apache Spark 12](#_Toc219445872)

[9. Giám sát mô hình theo thời gian thực 13](#_Toc1191756564)

# **Giới thiệu**

## **Tổng quan**

MLOps, viết tắt của "Machine Learning Operations," là một hệ thống tổ chức và quản lý quy trình phát triển, triển khai, và duy trì mô hình máy học. Được xem là sự mở rộng của DevOps (Development Operations) vào lĩnh vực machine learning, MLOps giúp tối ưu hóa quy trình làm việc từ việc xây dựng mô hình đến triển khai và theo dõi chúng trong môi trường sản xuất.

Mục tiêu của MLOps là cải thiện tính ổn định, độ tin cậy và hiệu suất của các hệ thống máy học, đồng thời giảm thiểu sự phụ thuộc vào các quy trình thủ công, giảm thiểu rủi ro và tăng cường khả năng mở rộng của dự án machine learning.

Với ngành ngân hàng, MLOps không chỉ là một công cụ quản lý mô hình mà còn là một phương tiện để tối ưu hóa các dự án liên quan đến dự đoán tài chính, chống gian lận, và quản lý rủi ro.

## **Mục tiêu**

Mục tiêu cụ thể của nền tảng MLOps là tạo ra một môi trường làm việc mạnh mẽ và linh hoạt để quản lý và phát triển các dự án machine learning. Để đạt được những mục tiêu này, nền tảng MLOps đặt ra các ưu tiên sau đây:

* Cải thiện Hiệu suất: MLOps hướng đến việc tối ưu hóa hiệu suất của mô hình máy học. Điều này bao gồm việc tối ưu hóa các thuật toán, quy trình làm việc và tối ưu hóa tài nguyên tính toán để đảm bảo mô hình hoạt động hiệu quả trong môi trường sản xuất.
* Tự động hóa Quy trình: MLOps đặt mục tiêu tự động hóa quy trình phát triển, triển khai và duy trì mô hình. Tự động hóa giúp giảm thiểu sự phụ thuộc vào các quy trình thủ công, giảm lỗi và tăng tốc độ triển khai.
* Tăng tính Tái sản xuất: MLOps cung cấp khả năng tái sử dụng mô hình một cách linh hoạt và hiệu quả. Điều này giúp giảm thời gian và công sức cần thiết cho việc phát triển các mô hình mới bằng cách sử dụng lại các thành phần đã kiểm định và chứng minh hiệu suất.
* Quản lý Đôi mới và Tính nhất quán: MLOps hỗ trợ quản lý đối mới và đảm bảo tính nhất quán giữa các phiên bản khác nhau của mô hình. Điều này làm cho quá trình triển khai và cập nhật trở nên dễ dàng và an toàn hơn.
* Giảm Thiểu Rủi ro và Bảo mật: MLOps đặt mục tiêu trong việc giảm thiểu rủi ro liên quan đến triển khai mô hình và đảm bảo an toàn về mặt bảo mật của dữ liệu và mô hình máy học.
* Tối ưu hóa Chi phí: MLOps hướng đến việc giảm thiểu chi phí liên quan đến việc triển khai và duy trì mô hình bằng cách tối ưu hóa sử dụng tài nguyên và quy trình làm việc.

## **Skateholder**

Nền tảng MLOps đóng vai trò quan trọng trong việc kết nối và phối hợp giữa các bên liên quan, được biết đến là các bên liên quan hay "stakeholders." Các skateholder đóng góp vào quá trình phát triển và duy trì mô hình máy học theo các cách khác nhau, với mỗi bên đều mang theo vai trò và nhiệm vụ đặc biệt.

1. Nhóm Phát triển Mô hình (Data Scientists và Machine Learning Engineers):

* Vai trò: Tạo ra và huấn luyện mô hình máy học.
* Nhiệm vụ: Phát triển và cải thiện các mô hình dựa trên dữ liệu, đảm bảo tính chính xác và hiệu suất của chúng. Họ tham gia vào quá trình xây dựng và kiểm định mô hình, đồng thời cần liên tục cập nhật để đối mặt với thay đổi trong dữ liệu và yêu cầu kinh doanh.

1. Nhóm IT và DevOps:

* Vai trò: Triển khai và duy trì hạ tầng.
* Nhiệm vụ: Bảo đảm rằng hạ tầng hỗ trợ triển khai mô hình một cách hiệu quả và an toàn. Họ đảm bảo tích hợp liền mạch giữa các phần mềm và hệ thống tổ chức, đồng thời duy trì tính ổn định và đảm bảo an toàn thông tin.

1. Nhóm Kinh doanh và Quản lý Dự án:

* Vai trò: Xác định yêu cầu kinh doanh và đưa ra chiến lược.
* Nhiệm vụ: Giao tiếp giữa các đội ngũ kỹ thuật và bên kinh doanh để đảm bảo rằng mô hình máy học đáp ứng đúng các mục tiêu kinh doanh. Họ cần định rõ yêu cầu và ưu tiên, cũng như đảm bảo rằng các giải pháp máy học đóng góp vào chiến lược tổng thể của tổ chức.

1. Nhóm An toàn Thông tin và Tuân thủ:

* Vai trò: Bảo vệ dữ liệu và đảm bảo tuân thủ các quy định.
* Nhiệm vụ: Đảm bảo rằng triển khai và duy trì mô hình tuân thủ các quy định an toàn thông tin và quy định pháp luật liên quan. Họ cũng giữ vai trò quan trọng trong việc đánh giá và quản lý rủi ro liên quan đến mô hình máy học.

1. Người Dùng Cuối và Bộ Phận Kinh doanh:

* Vai trò: Sử dụng và đánh giá mô hình máy học trong môi trường thực tế.
* Nhiệm vụ: Cung cấp phản hồi và thông tin chi tiết về hiệu suất của mô hình từ quan điểm người dùng cuối. Họ cũng chịu trách nhiệm định rõ yêu cầu kỹ thuật từ quan điểm người dùng và giúp cải thiện tính ứng dụng của mô hình trong các tình huống thực tế.

# **Timeline**

## **Thiết kế kiến trúc và thử nghiệm các nền tảng (12/7 - 31/9):**

* Mục tiêu: Xác định phạm vi, yêu cầu và mục tiêu chiến lược cho dự án MLOps.
* Các kết quả đạt được:
* Kiến trúc tổng quan của hệ thống MLOps qua các giai đoạn.
* Kiến trúc hạ tầng triển khai trên hệ thống của Sacombank.
* Bảng đánh giá giữa các công cụ triển khai MLOps.
* Kế hoạch dự kiến cho nhu cầu resource để xây dựng platform trên môi trường DEV.

## **Xây dựng platform trên hệ thống OpenShift (1/10 - 31/11):**

* Mục tiêu: Phát triển và triển khai MLOps platform trên OpenShift.
* Các kết quả đạt được:
* Hệ thống xây dựng và thí nghiệm mô hình machine learning theo kiến trúc microservice.
* Hệ thống Pipeline để tự động việc xây dựng mô hình machine learning và triển khai mô hình ra các API.
* Hệ thống giám sát vận hành của mô hình theo thời gian thực.
* Hệ thống tự động đóng gói mô hình thành các Container, triển khai mô hình và tự động phân bổ tài nguyên.real-time.
* Dashboard thống nhất các hệ thống để quản lý tập trung các mô hình và các dự án machine learning.
* Xây dựng hạ tầng Feature Store với 2 lớp offline và online phục vụ cho nhu cầu dữ liệu cho trainning mô hình và dự đoán kết quả từ batch job hoặc on-demand.

## **Xây dựng công cụ cho ETL dữ liệu từ các nguồn vào Feature Store (1/12 - 31/12):**

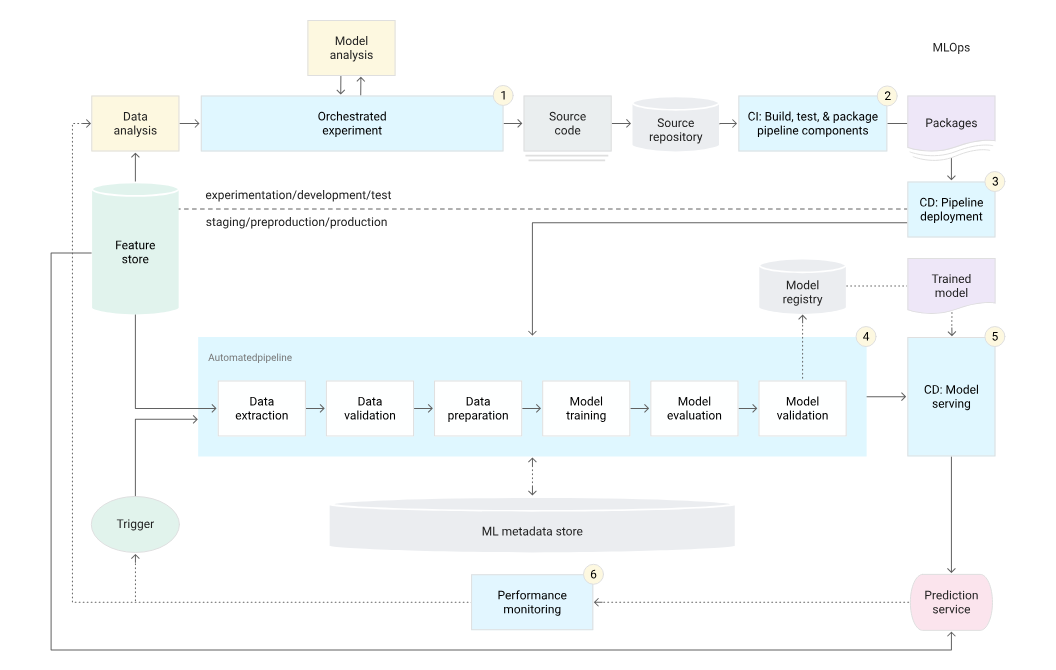
* Mục tiêu: Xây dựng công cụ phục vụ cho việc ETL dữ liệu vào feature store cho các bài toán ML
* Kết quả đạt được:
* Công cụ thiết lập lịch thực thi cho các job ETL
* Công cụ Apache Spark trên OpenShift để thực thi, tối ưu và đảm bảo các job ETL được thực thi một cách hiệu quả.
* Công cụ Apache Kafka để đón nhận dữ liệu và ETL real-time vào feature store.

# **Chức năng**

* Quản lý phiên bản mô hình và API: theo dõi và quản lý phiên bản của mô hình máy học và mã nguồn, đảm bảo tính nhất quán và khả năng tái sử dụng.
* Tự động hóa việc kiểm thử và triển khai mô hình: Tự động hóa các bước trong quy trình phát triển và triển khai, giảm thiểu lỗi, tăng tốc độ phát triển và quá trình triển khai mô hình máy học từ môi trường phát triển đến môi trường sản xuất.
* Quản lý và theo dõi hiệu suất của mô hình: Quản lý vòng đời của mô hình, bao gồm việc theo dõi hiệu suất, cập nhật mô hình và quản lý các phiên bản khác nhau. Theo dõi hoạt động của mô hình trong thời gian thực, cung cấp thông tin về hiệu suất và báo cáo tổng quan.
* Tôi ưu tài nguyên: Tối ưu hóa việc sử dụng tài nguyên máy tính để giảm thiểu chi phí và tăng cường hiệu suất.
* Quản lý các đặc trưng cho mô hình ML: Lưu trữ và quản lý các đặc trưng (features) được sử dụng trong quá trình huấn luyện và triển khai mô hình, đảm bảo tính nhất quán và sẵn sàng tái sử dụng. Tích hợp các đặc trưng từ nhiều nguồn khác nhau và chuẩn hóa chúng để có sự thống nhất trong quá trình huấn luyện và triển khai mô hình.
* Chuyển đổi và đồng bộ dữ liệu từ các nguồn về Feature Store: Trích xuất dữ liệu từ nhiều nguồn khác nhau, bao gồm cả cơ sở dữ liệu, kho dữ liệu, và các hệ thống bên ngoài. Thực hiện các bước xử lý để chuẩn hóa, làm sạch và biến đổi dữ liệu thành định dạng phù hợp để huấn luyện mô hình. Tích hợp cơ chế đồng bộ và lên lịch để tự động hóa quy trình ETL theo định kỳ hoặc dựa trên sự kiện. Theo dõi hiệu suất của quy trình ETL, đảm bảo tính ổn định và hiệu suất của quá trình xử lý dữ liệu.

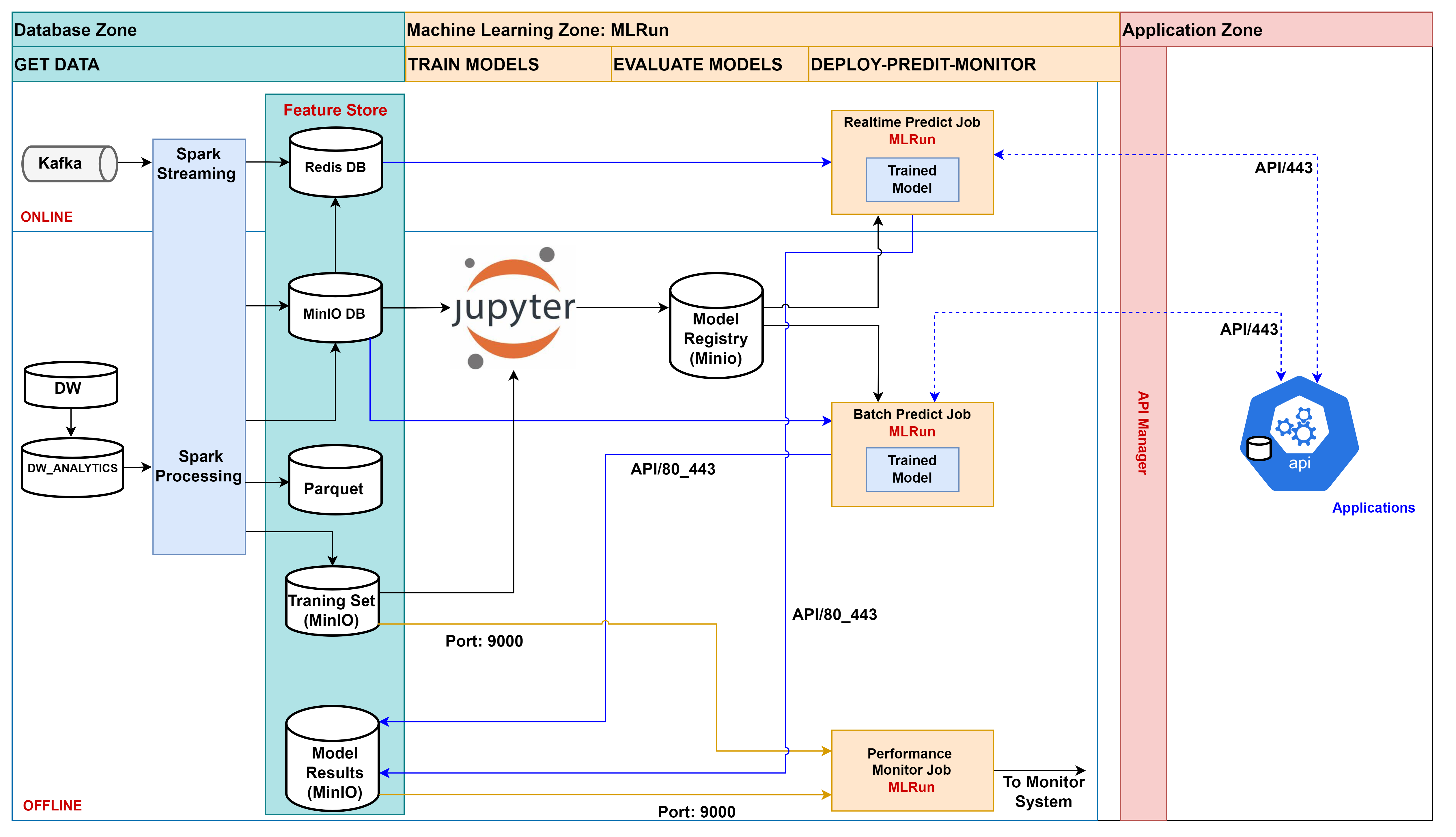
# **Kiến trúc hệ thống**

## **Kiến trúc High-Level**



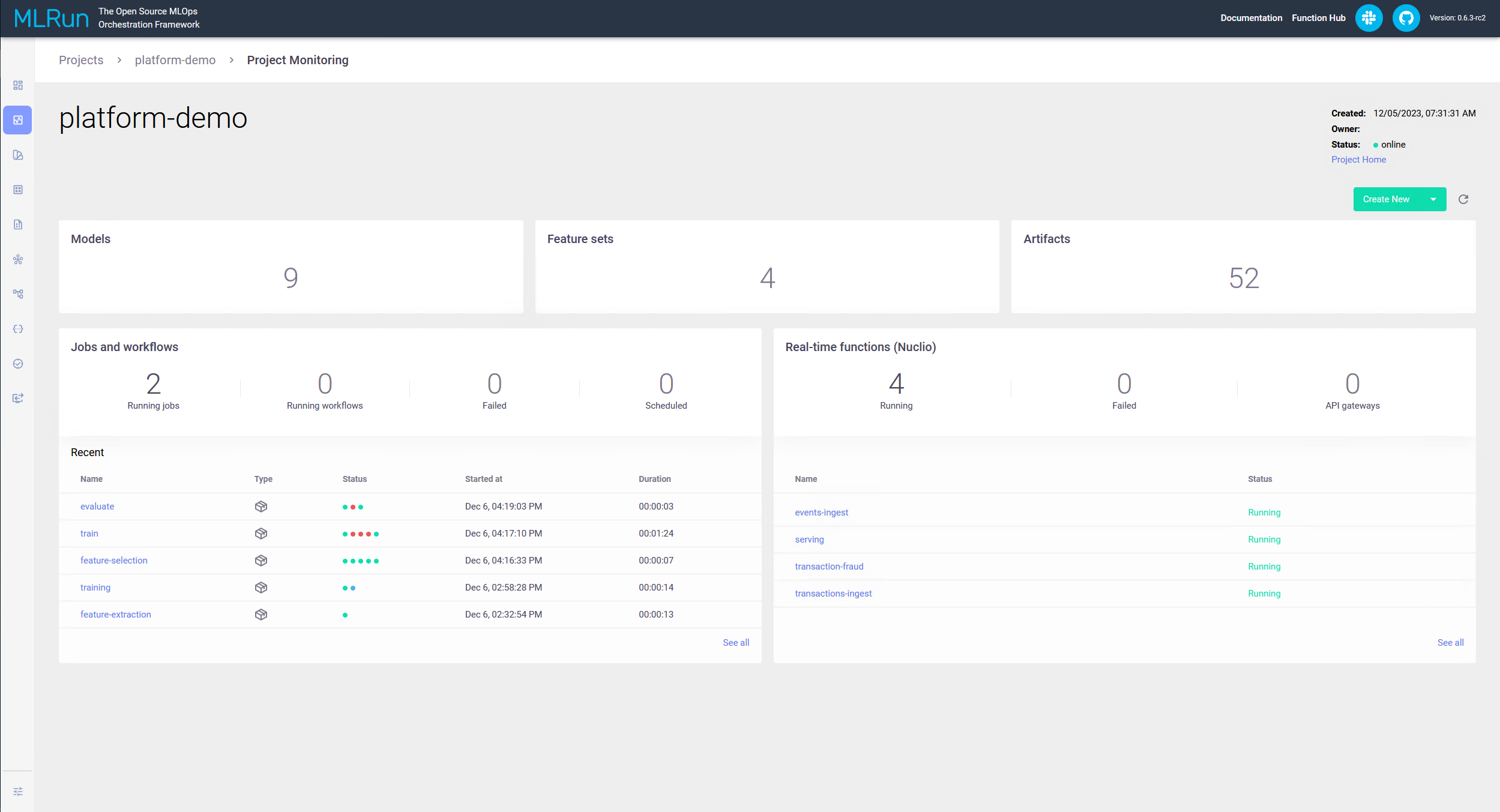
1. Phát triển và thử nghiệm: Thử các thuật toán ML mới và phát triển mô hình mới. Đầu ra của giai đoạn này là mã nguồn của các bước trong quy trình ML, sau đó được đẩy đến kho lưu trữ nguồn.
2. Kiểm thử và tạo các thành phần của Pipeline: Thực thi mã nguồn của mô hình và thực hiện kiểm thử. Kết quả của giai đoạn này là các thành phần của một pipeline ML phục vụ cho việc chạy tự động cho giai đoạn tiếp theo.
3. Thực hiện triển khai Pipeline: Triển khai Pipeline ML tự động lên các môi trường khác nhau nhầm phục vụ cho các giai đoạn của phát triển phần mềm (DEV, TEST, PRODUCTION)
4. Kích hoạt Pipeline ML: Pipeline xây dựng mô hình ML có thể được kích hoạt thủ công hoặc tự động theo lịch trình mà không cần phải chỉnh sửa.
5. Triển khai mô hình: Tiếp nhận mô hình được sinh ra ở Pipeline và tự động triển khai lên thành các microservice (API).
6. Giám sát và theo dõi: Có thể theo dõi và giám sát các số liệu thống kê về dữ liệu được đưa vào mô hình cũng như hiệu suất của mô hình trên dữ liệu được đưa vào theo thời gian thực.

## **Kiến trúc chi tiết**

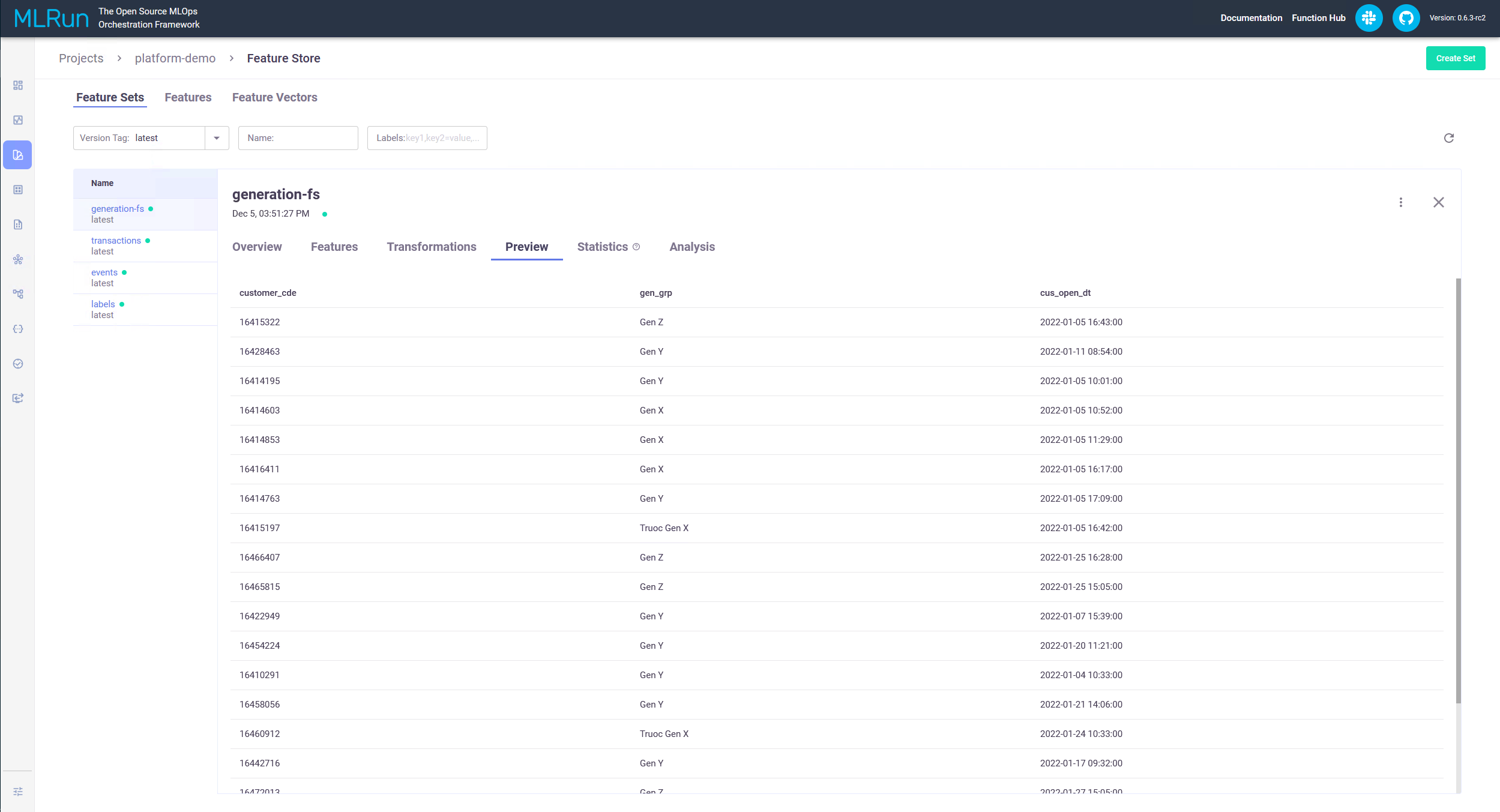


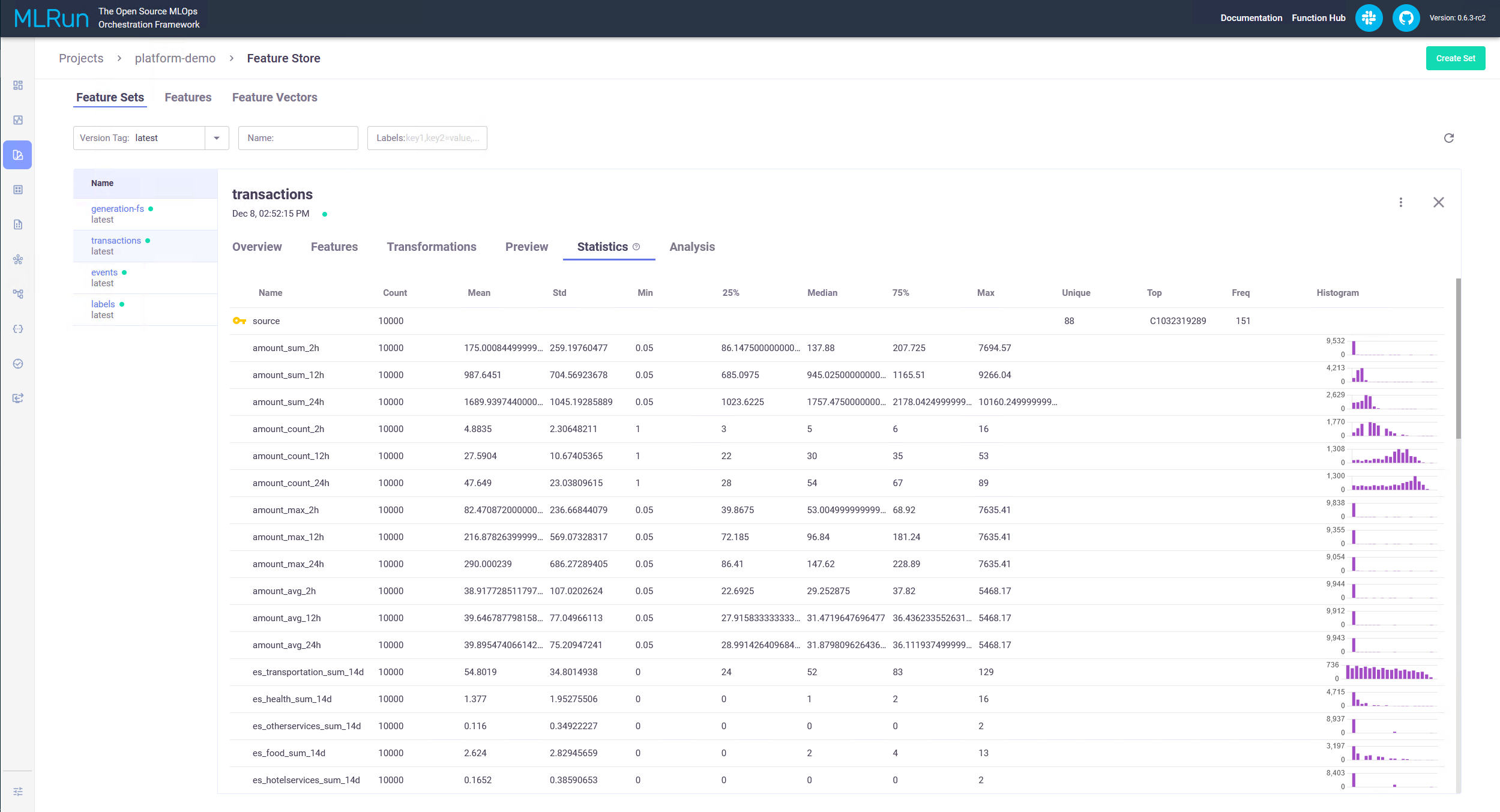
# **V. MLOps Platform**

## **Dashboard MLOps Platform**

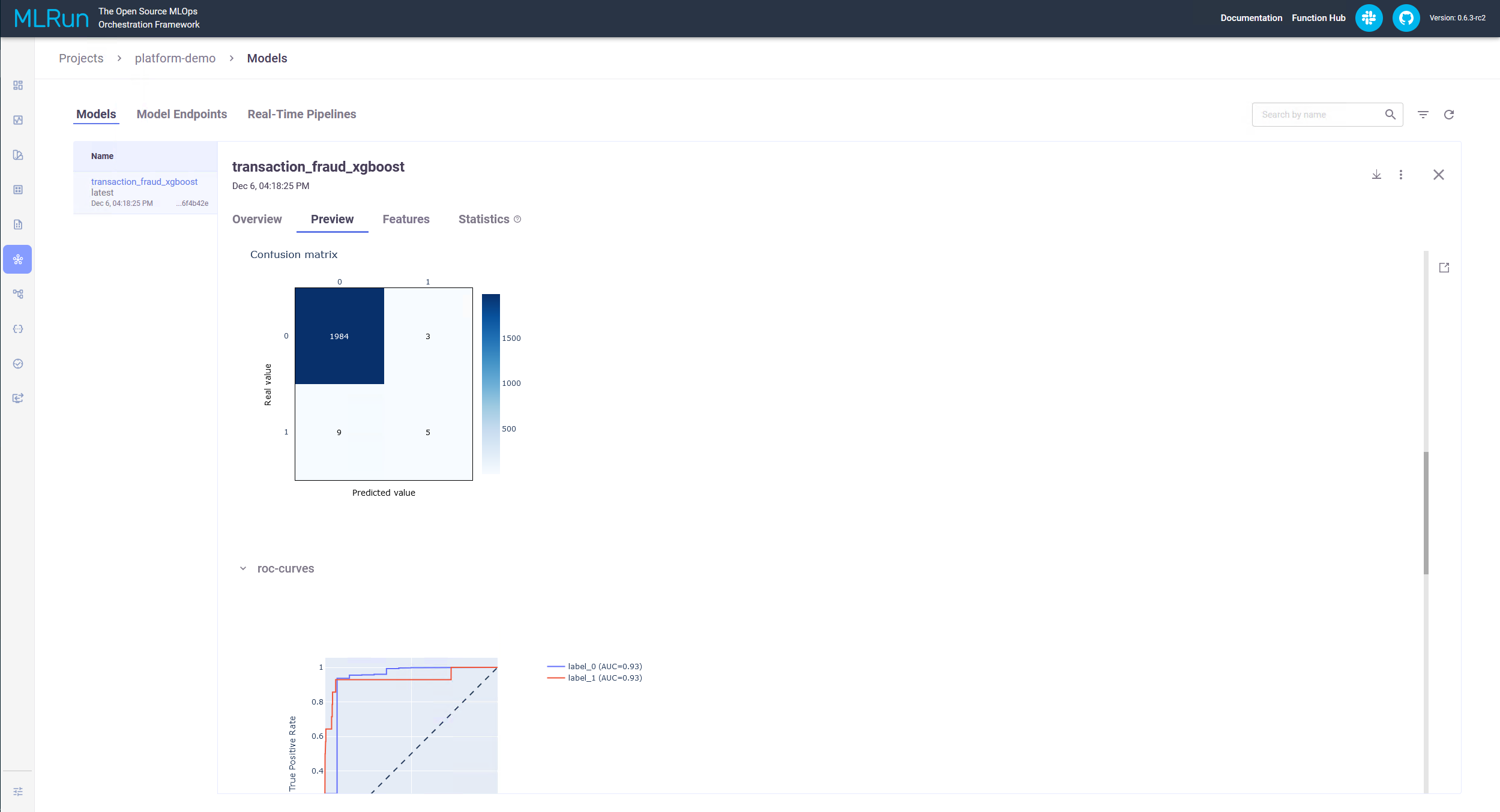


## **Feature Store**

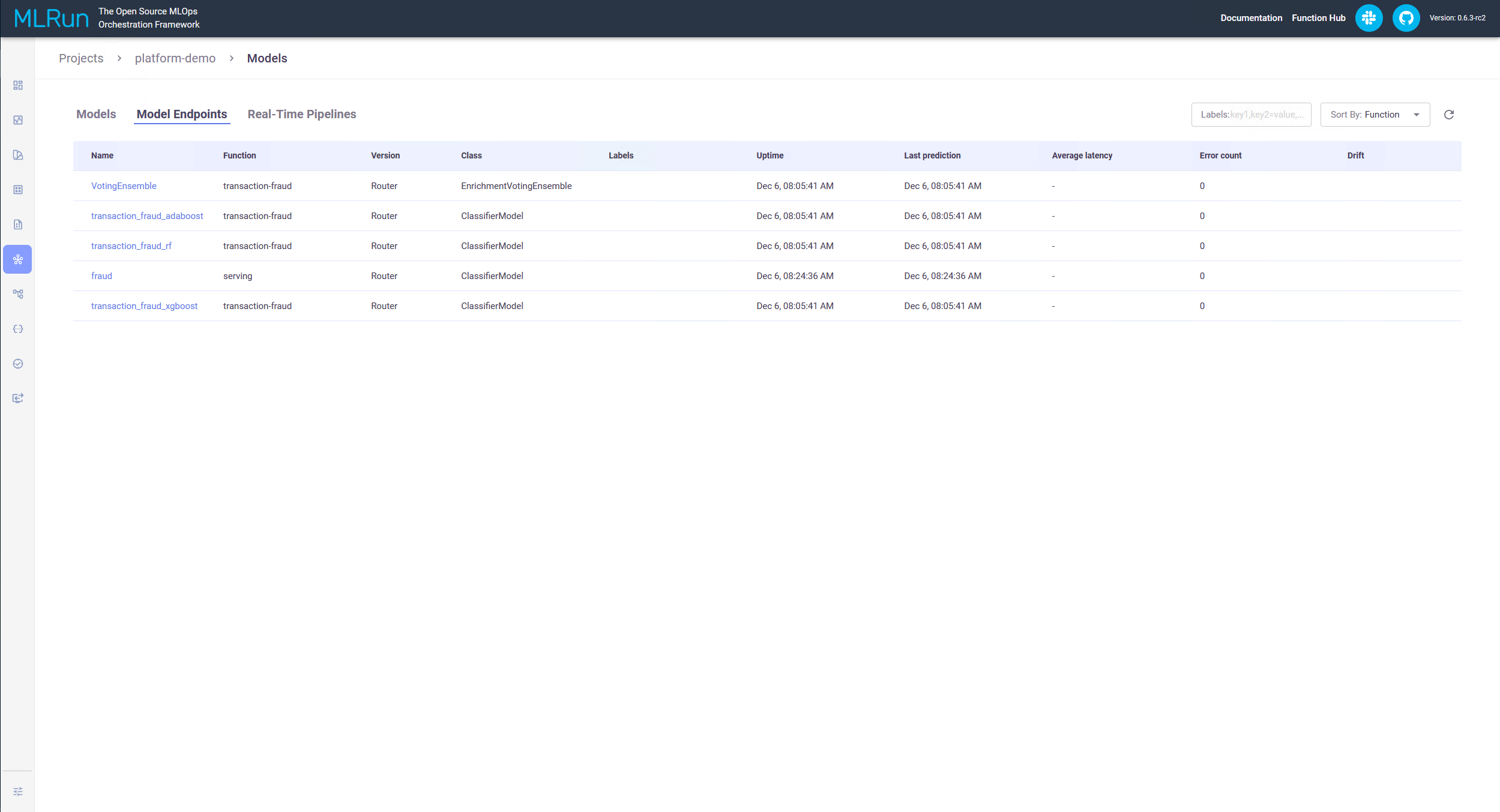


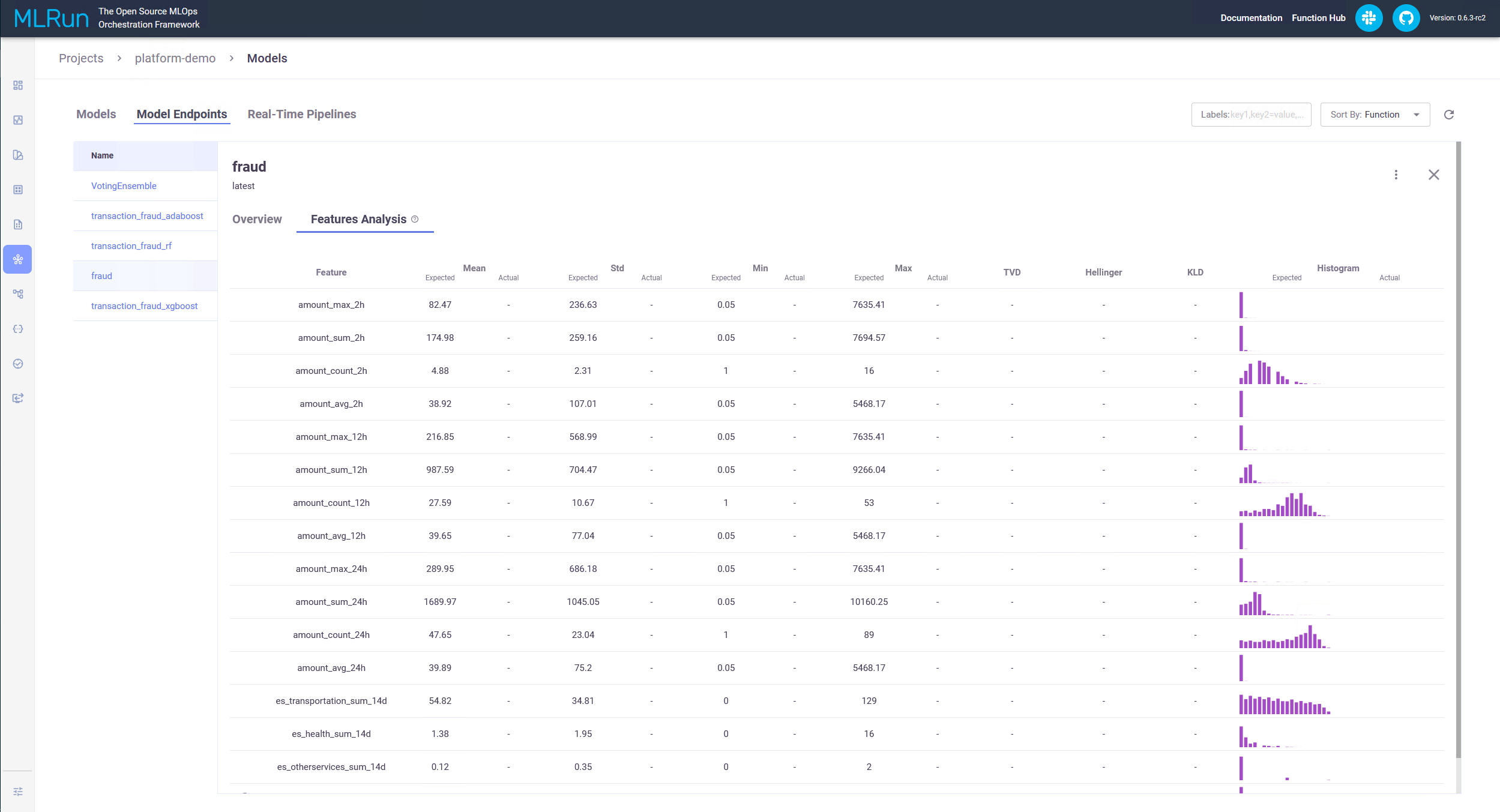


## **Giám sát quá trình training mô hình**

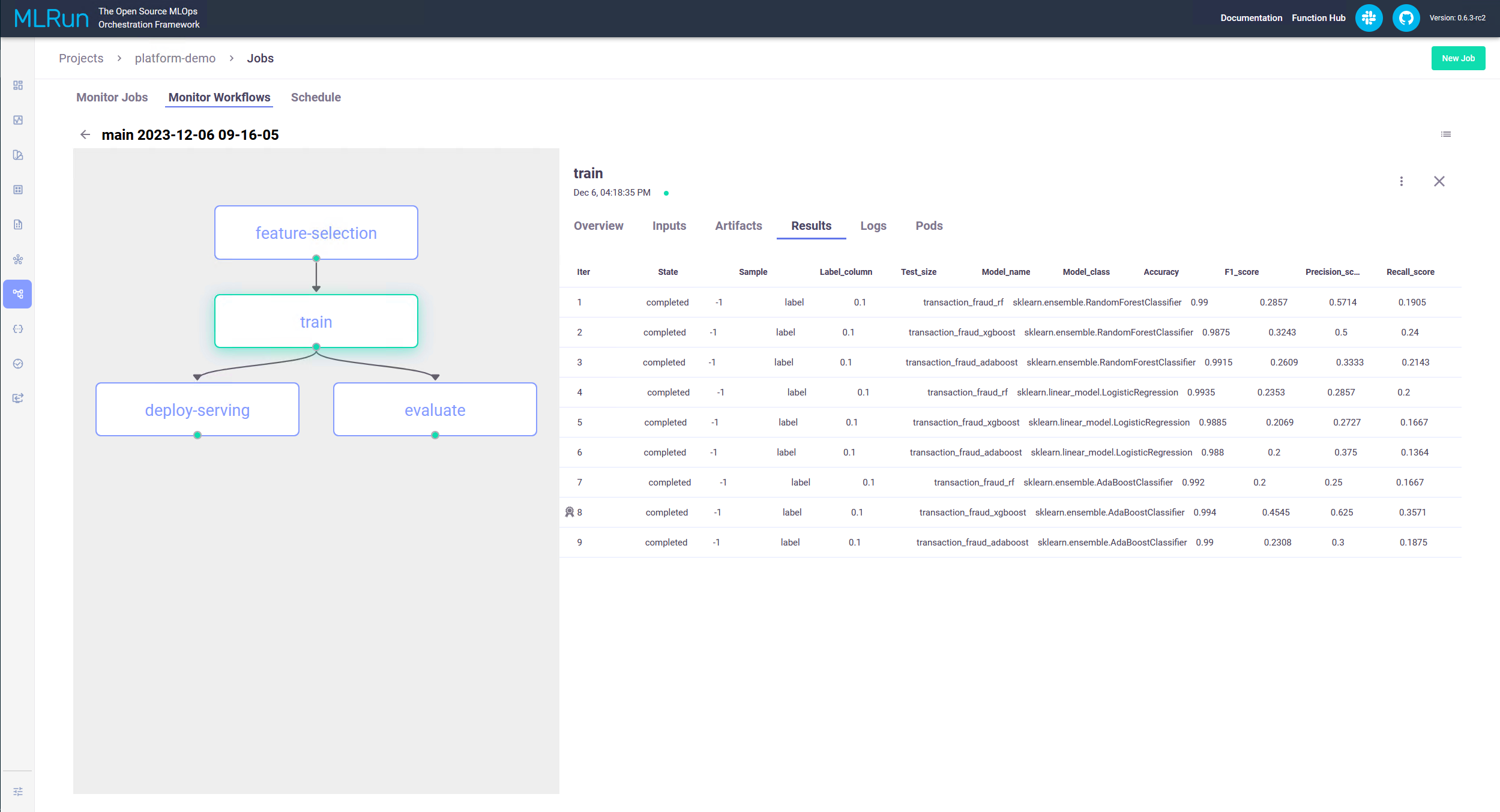


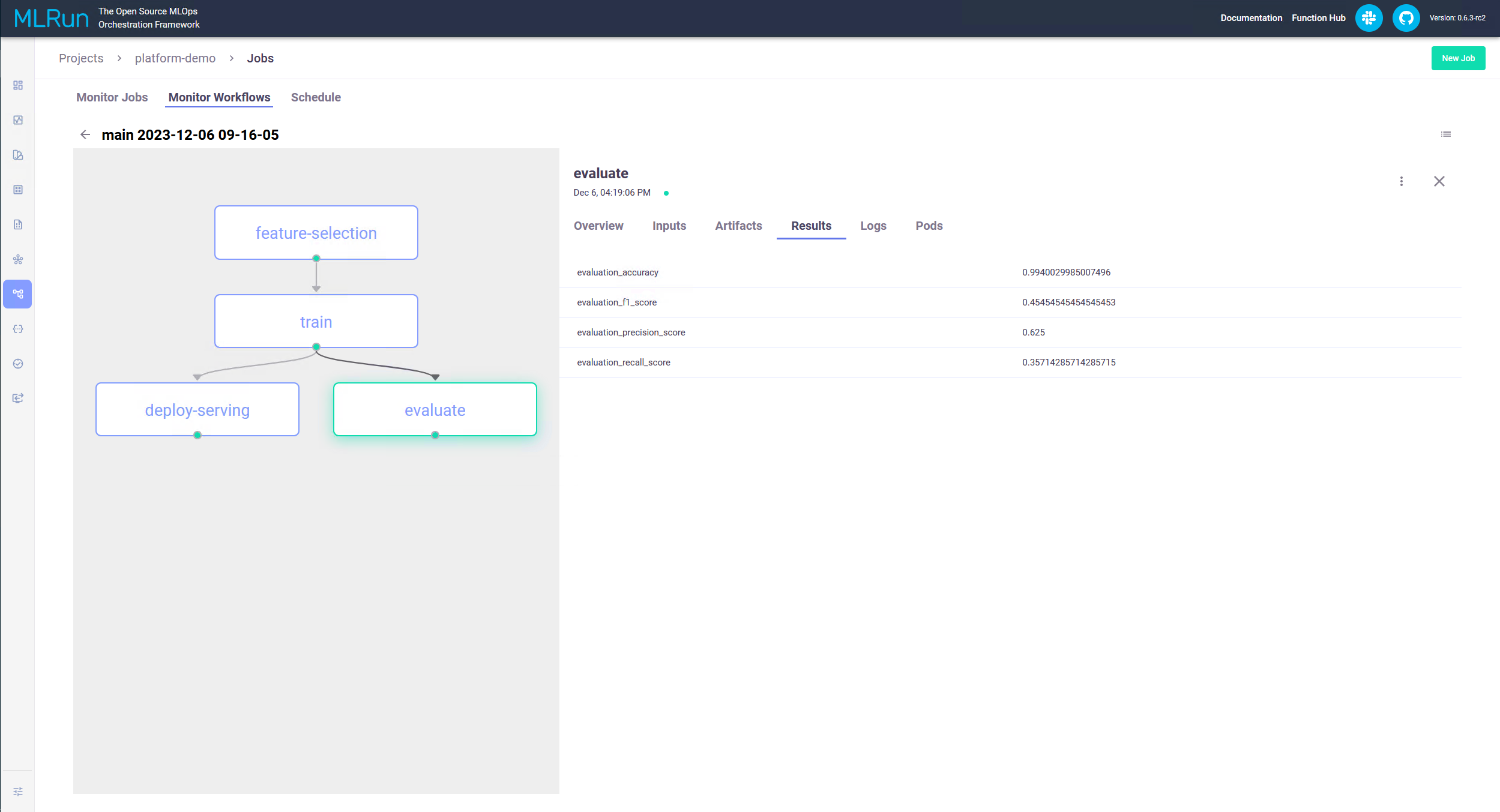
## **Quản lý API endpoints**



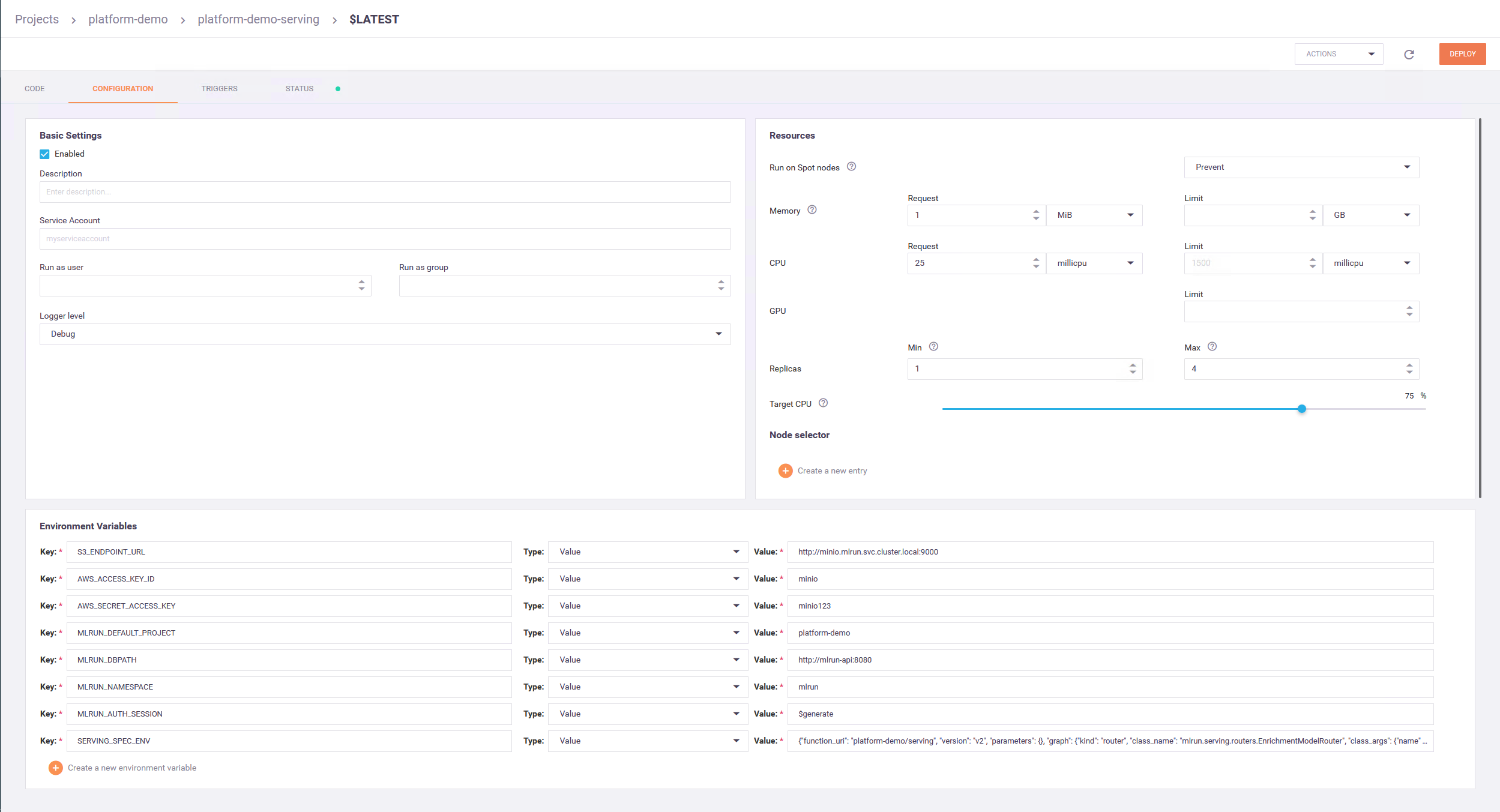


## **Pipeline và theo dõi kết quả**

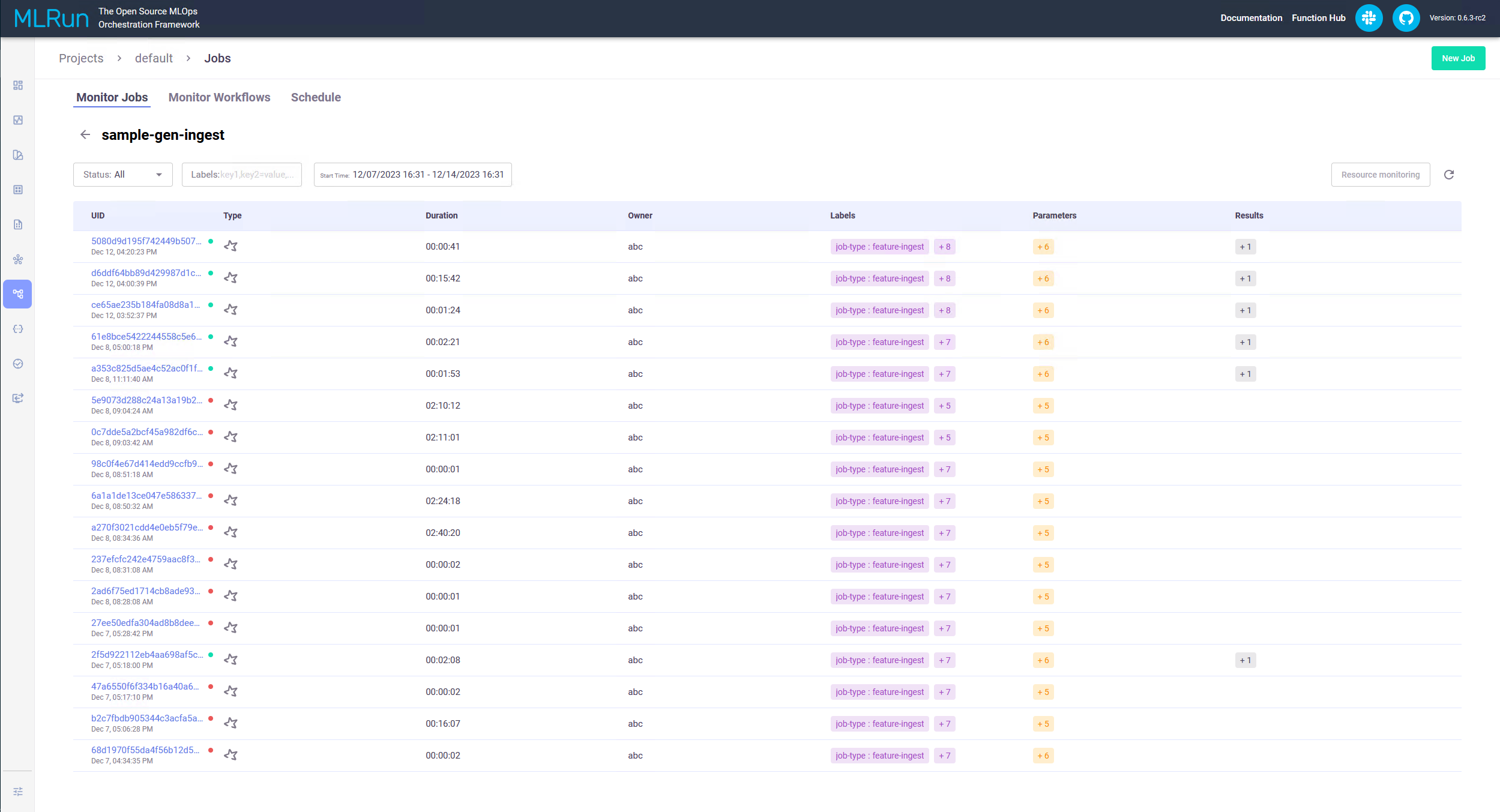




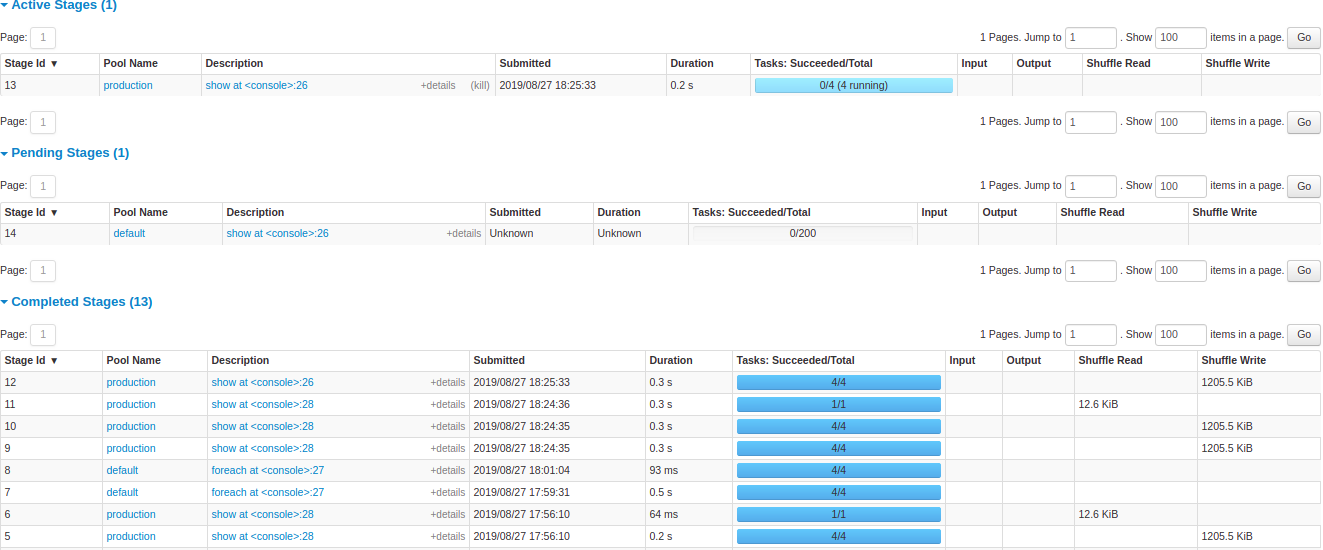
## **Tinh chỉnh resource cho API và các job bằng giao diện và tự động**

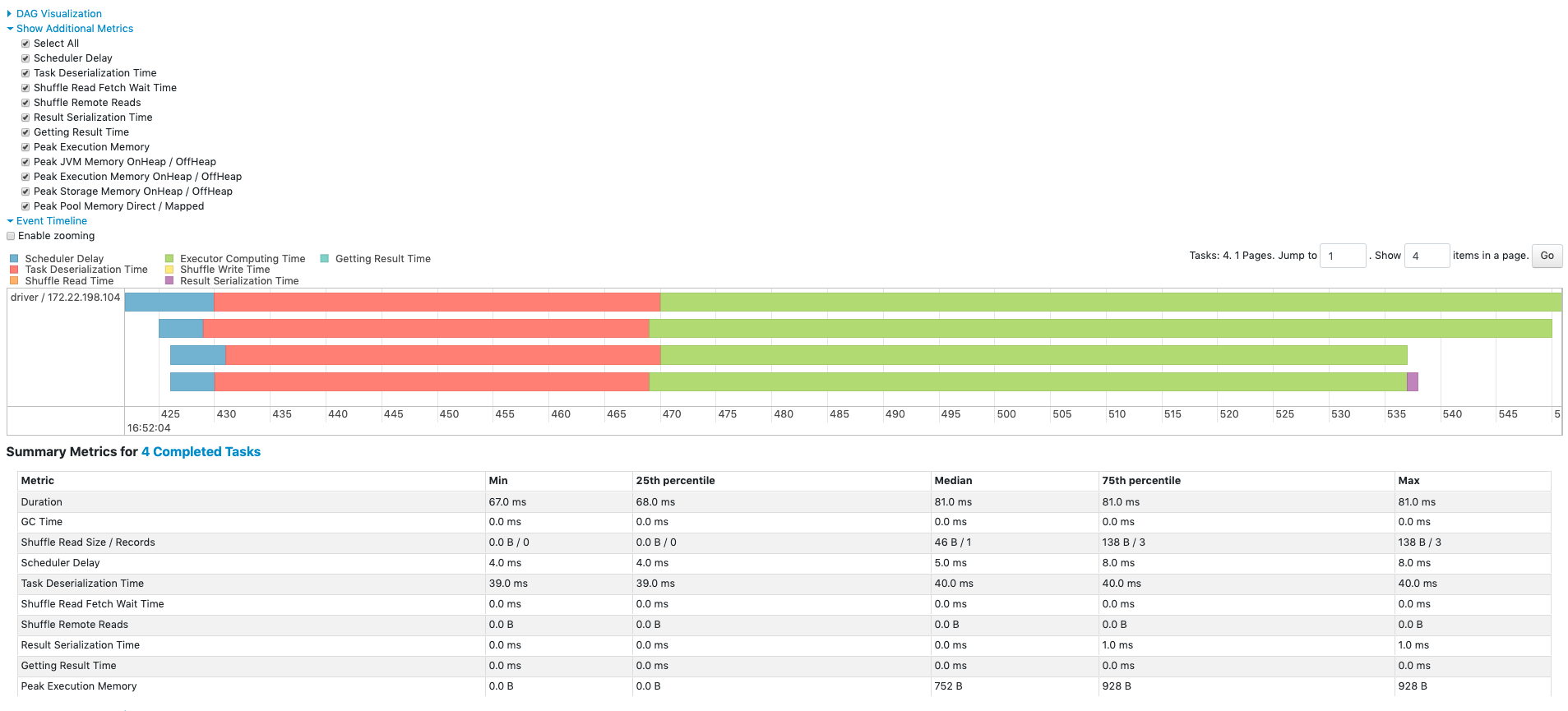


## **Quản lý job ETL**



## **Thực thi các job ETL thông qua Apache Spark**





## **Giám sát mô hình theo thời gian thực**

