

Avoid Training & Predicting Data Conflict using Feast

Nhóm TimeSeries - GRID070

Ngày 9 tháng 10 năm 2025

Nội dung sẽ chia thành 5 phần chính:

- **Revision: Data Versioning và ETL pipeline**
- **Overview: Giới thiệu về các 1 dòng Hết Mòn**
- **Feature Store và Feast là gì ?**
- **Case Study**
- **Final Revision: Ứng dụng thực tiễn trong bối cảnh team Time Series**

Intro: Từ ETL, Processing, Data Versioning và Feature Store

Hình thức các dữ liệu bất đồng bộ quy trình ETL (Extract, Transform, Load) thu thập, chuyển hóa dữ liệu thô từ nhiều nguồn khác nhau về một nơi, biến đổi chúng rồi lưu trữ vào 1 trung tâm lưu trữ tập trung như Data Warehouse.

Tuy nhiên, thực tế các dữ liệu "ng" có tính liên tục (Streaming data như tin nhắn, nhật ký) và dữ liệu "tĩnh" có thể thay đổi (static như sách, CCCD), tạo ra nhu cầu cho các phương pháp xử lý linh hoạt hơn như Apache Spark cho phép tải và xử lý dữ liệu phức tạp một cách linh hoạt. Ngoài ra còn có các công cụ truy vấn (SparkSQL) và học máy (MLlib).

Chờn s

1.1 ML/AI Project Lifecycle - Giới thiệu về các 1 dòng Hết Mòn

Data Engineering Modeling Deployment Business analysis AI Infrastructure - Các bước trong các 1 dòng AI 6 vai trò trong nhóm phát triển AI

Feature Store và Feast là gì ?

Vấn đề trong ML Project life cycle. Vấn đề Feature Store giải quyết -> Các thành phần chính trong Feature Store pipeline -> Lịch sử Feature Store trong bối cảnh X. (liệt kê các lịch sử).

Feast hỗ trợ là gì trong Feature ? (trình bày về Feast) -> Feast Pipeline -> Các thành phần của Feast.

Case Study

Final Revision: Ứng dụng thực tiễn trong bối cảnh team Time Series