

# 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 đơn vị Hc Mòy**
- **Feature Store và Feast là gì ?**
- **Case Study**
- **Final Revision: Tổng kết và bài học trong bối cảnh team Time Series**

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

Hình thức của dữ liệu bất kỳ quy trình ETL (Extract, Transform, Load) thu thập, chuyển đổi dữ liệu từ nhiều nguồn khác nhau về định dạng, biến đổi dữ liệu và lưu trữ vào 1 trung tâm lưu trữ tập trung như Data Warehouse.

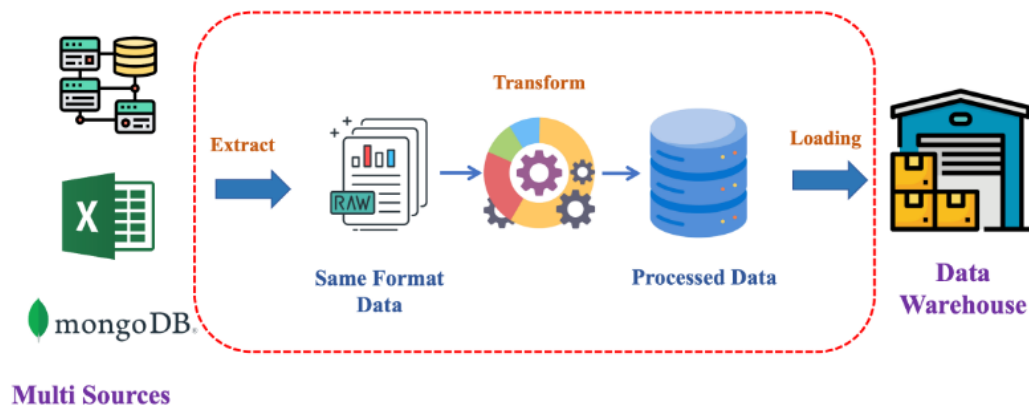


Figure 1: ETL pipeline

Tuy nhiên, trong thực tế xử lý 1 lượng dữ liệu liên tục "ng" có các nhập liệu liên tục (Streaming data như tin nhắn, hệ thống shopee) và dữ liệu "tĩnh" ở trạng thái sau khi lưu (static như sách, thông tin cơ bản) tạo nhu cầu cho 1 phương pháp xử lý linh hoạt hơn như Apache Spark cho phép ta xử lý cả lượng dữ liệu phức tạp và tĩnh. Khi đó hệ thống còn hỗ trợ truy vấn (SparkSQL) và học máy (MLlib).

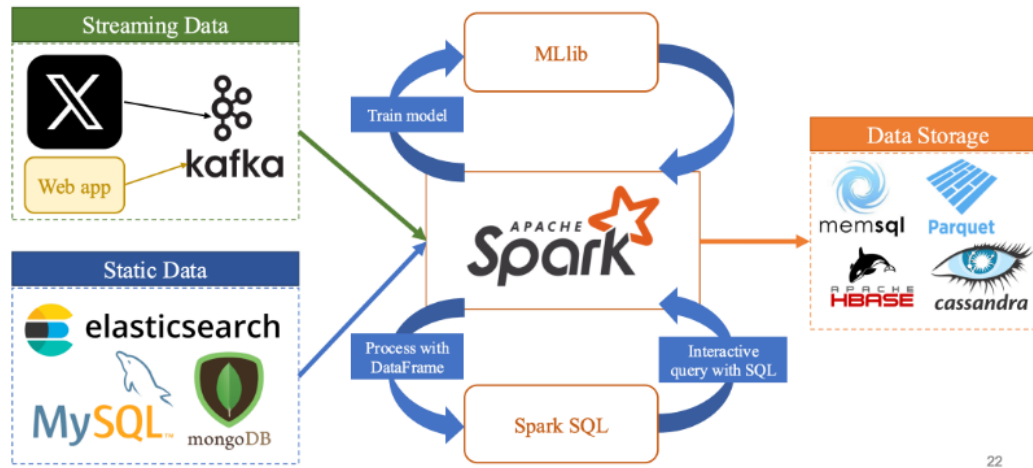


Figure 2: Process pipeline

Chờn vớ s phòt trin cà còc quy trỡnh hc mỳ ỏ to ra thỏch thc qun lý mi phiỏn bn d liu/mũ hỡnh bao gm qun lý vũ s artifacts nh tỳ tham s, phiỏn bn mũ hỡnh vớ b d liu mi c to ra. Vớ Data Versioning, nú m bo kh nng theo dừi, kim soỏt vớ tòi th nghim tng t nh cỏch Git qun lý mỗ ngun.

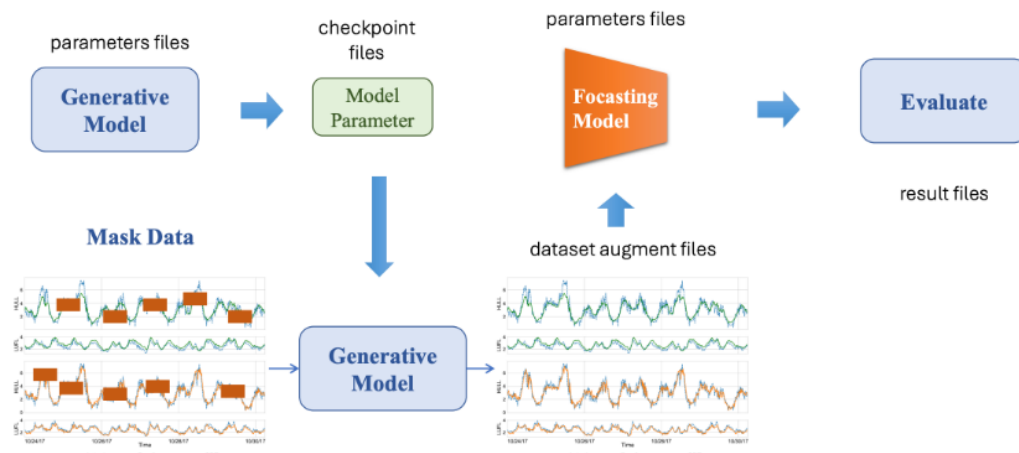


Figure 3: Data versioning Pipeline

Cui cớng sau khi ỏ lu d liu vớo trung tóo lu tr data warehouse, x lý d liu lng ln c tnh vớ ng nh Apache Spark sau ú qun lý mi phiỏn bn d liu/mũ hỡnh khi hun lỳn vớ data versioning, 1 cớu hi c t ra lỳ lỳm th nộ kt hp c 3 quy trỡnh chun húa, lu tr ri phc v còc c trỡng ỏ x lý 1 cỏch nht quỏn cho c vic hun lỳn mũ hỡnh (training) vớ d oỏn (inference). Tt c cỏc bc trỏn hi t v mt gii phỏp trung tóo vớ hii qu: Feature Store.

## ML/AI Project Lifecycle - Gii thiu vùng i ca 1 d òn Hc Mỳ

Data Engineering Modeling Deployment Business analysis AI Infrastructure - C s h tng ca 1 d òn AI 6 vai trũ trong nhúm phòt trin AI

## **Feature Store và Feast là gì ?**

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

Feast hỗ trợ là gì trong Feature ? (trị lý của Feast) -> Feast Pipeline -> Các thành phần của Feast.

## **Case Study**

**Final Revision:** ôn lại bài khởi nghiệp trong bối cảnh team Time Series