# BÁO CÁO WORKSHOP #1 – NEX INTERNSHIP 2025

**Người viết báo cáo:** Nguyễn Tô Thiên Bảo

## Mục tiêu buổi workshop

- Cung cấp góc nhìn doanh nghiệp về lĩnh vực Data và AI.

- Chia sẻ các kiến thức nền tảng quan trọng cho việc xây dựng hệ thống ứng dụng thực tế.

- Giới thiệu về Data Warehouse – một chủ đề trọng tâm trong nội dung thực tập, đặc biệt cần thiết với các bạn làm dự án “Agent trên data của doanh nghiệp”.

## Nội dung chia sẻ chính (Anh Đức Anh trình bày)

1. **Giới thiệu tổng quan về Data Warehouse (DW)**

- Data Warehouse (DW) là hệ thống lưu trữ dữ liệu tập trung, hỗ trợ phân tích dữ liệu doanh nghiệp.

- Khác với hệ thống OLTP (Online Transaction Processing):

* OLTP: dùng cho nghiệp vụ thường nhật, tốc độ truy xuất cao.
* DW: dùng cho phân tích, thống kê, hỗ trợ quyết định kinh doanh.

1. **Vì sao cần Data Warehouse?**

- Đảm bảo hệ thống phân tích không ảnh hưởng đến hệ thống giao dịch chính.

- Cho phép phân tích lịch sử dài hạn, xử lý dữ liệu lớn.

- Tối ưu hóa truy vấn tổng hợp, thống kê.

1. **Các thành phần kiến trúc của hệ thống DW**

- Data Sources: gồm databases, file excel, thiết bị IoT,...

- ETL/ELT: quy trình Extract – Transform – Load (hoặc Load trước rồi Transform).

- Data Warehouse Layer: nơi tập trung dữ liệu chuẩn hóa.

- Data Marts: phân tách dữ liệu theo bộ phận hoặc mục đích sử dụng.

- Visualization Layer: nơi cung cấp dữ liệu cho các công cụ dashboard.

1. **Các khái niệm và mô hình quan trọng**

- ETL vs ELT: tuỳ vào kiến trúc, doanh nghiệp sẽ lựa chọn phù hợp.

- Medallion Architecture (Bronze – Silver – Gold): mô hình lưu trữ theo tầng.

- So sánh hệ thống DW truyền thống vs hiện đại (cloud-native).

1. **Vai trò của các thành viên liên quan**

- Data Engineer: xây dựng ETL pipeline, xử lý và tích hợp dữ liệu.

- Business Analyst / Data Analyst: phân tích dữ liệu, mô hình hóa.

- AI Engineer: sử dụng dữ liệu từ DW để huấn luyện, triển khai mô hình AI.

## Chuyển giao sang phần AI (Anh Bằng trình bày)

### Xu hướng công việc AI hiện nay

- AI Engineer ngày nay cần hiểu cả mô hình lẫn hệ thống tích hợp.

- Biết làm việc với backend, frontend để triển khai ứng dụng AI.

### Kỹ năng cần thiết

- Biết Docker, Git, môi trường dev/test/prod.

- Hiểu sự khác biệt giữa Docker và VMware.

- Làm chủ việc xây dựng & triển khai hệ thống ML.