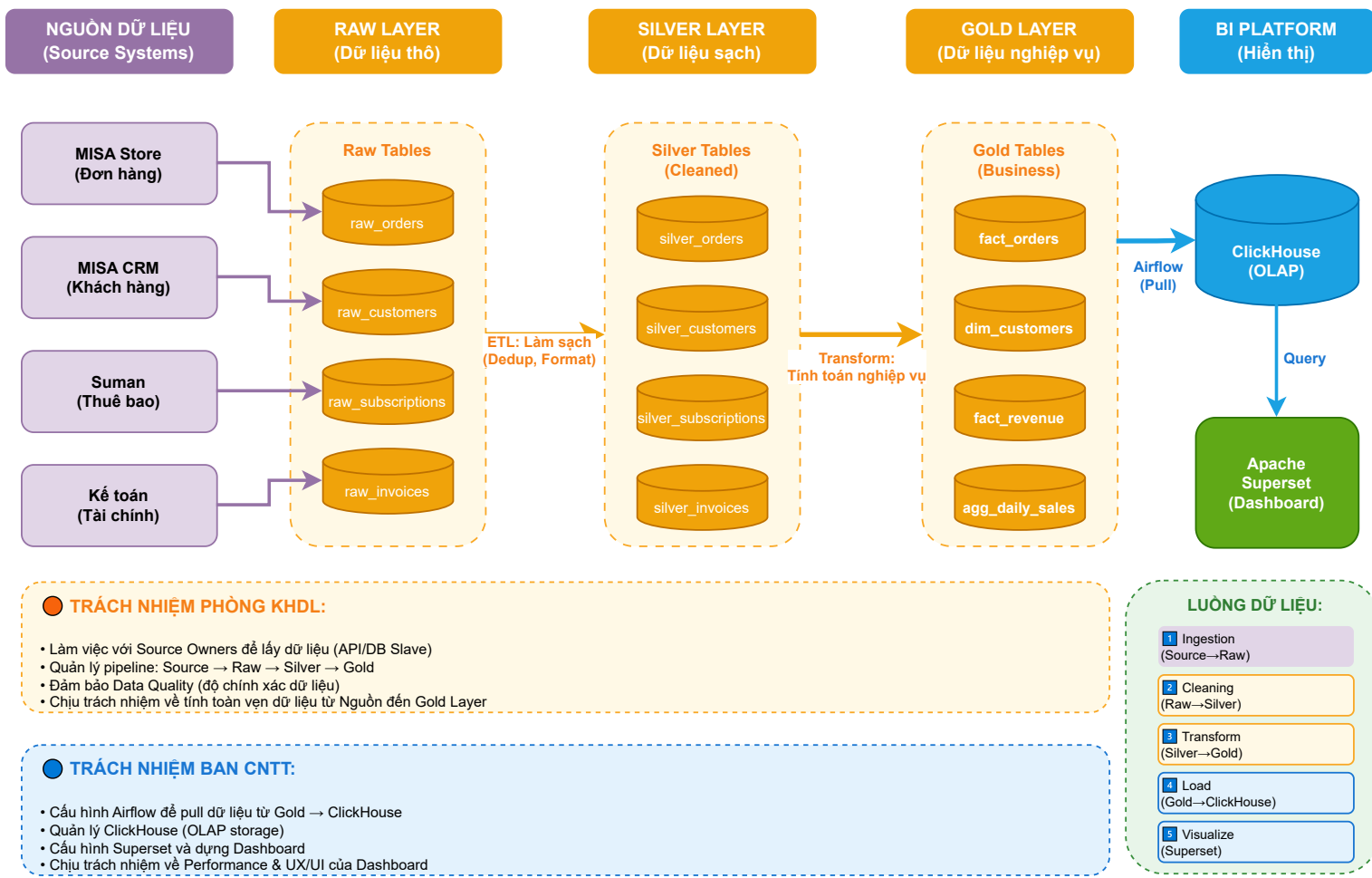


## KIẾN TRÚC DỮ LIỆU & LƯỒNG ETL (Data Architecture & ETL Flow)



## MA TRẬN TRÁCH NHIỆM (RACI MATRIX) Ban CNTT vs Phòng KHDL

### GIẢI THÍCH RACI:

R - Responsible (Người thực hiện)

A - Accountable (Chịu trách nhiệm)

C - Consulted (Được tham vấn)

I - Informed (Được thông báo)

Lưu ý: Mỗi công việc chỉ có 1 người Accountable (A), nhưng có thể có nhiều người Responsible (R)

HẠNG MỤC CÔNG VIỆC	BAN CNTT (MCP BI Team)	PHÒNG KHDL (Data Foundation)	ĐỐI TÁC NGUỒN (3rd Party)
1. Tiếp nhận yêu cầu Biz	R	C	-
2. Làm việc với Source Team lấy API	C	R	I
3. Pipeline: Source → Raw → Gold	-	R	-
4. Data Quality (Độ đúng đắn dữ liệu)	I	A	-
5. Pipeline: Gold → ClickHouse	R	C	-
6. Performance Dashboard (Tốc độ)	A	-	-
7. UX/UI Báo cáo	R	-	-

### TÓM TẮT PHÂN CÔNG:

#### BAN CNTT (MCP BI Team):

- Tiếp nhận yêu cầu từ Người dùng
- Quản lý lớp Hiển thị (Superset)
- Quản lý Xử lý trung gian (Airflow/ClickHouse)
- Chịu trách nhiệm về Performance & UX/UI

#### PHÒNG KHDL (Data Foundation Team):

- Quản lý lớp Dữ liệu nền (Raw/Silver/Gold)
- Làm việc với các team sản phẩm (Source Owners)
- Chịu trách nhiệm về Data Quality (Độ chính xác)
- Thực hiện ETL & Data Quality Assurance

## QUY TRÌNH PHỐI HỢP: BAN CNTT & PHÒNG KHDL (End-to-End Flow: Từ Yêu Cầu Nghiệp Vụ đến Báo Cáo BI)

### CÁC BÊN THAM GIA (ACTORS):

- Người dùng
- Ban CNTT
- Phòng KHDL
- Đối tác nguồn

#### GIẢI ĐOẠN 1: TIẾP NHẬN & PHÂN TÍCH (Ban CNTT chủ trì)

##### Bước 1: Tiếp nhận yêu cầu

Actor: Ban CNTT (BA/PO)

- Nhận Ticket/Email từ Người dùng
- Làm rõ: Metric? Dimension? Tần suất?

##### Bước 2: Kiểm tra kho dữ liệu

Actor: Ban CNTT

- Truy vấn thử Gold Layer
- Kiểm tra Data Availability

CHƯA CÓ

ĐÃ CÓ

#### GIẢI ĐOẠN 2: YÊU CẦU DỮ LIỆU & XỬ LÝ NỀN (Phòng KHDL chủ trì)

##### Bước 3: Chuyển yêu cầu

- Ban CNTT tạo Ticket Jira
- Gán cho KHDL
- Mô tả nghiệp vụ + Output

##### Bước 4: Phân tích nguồn

- KHDL kiểm tra Raw/Silver
- Làm việc với Source Owners
- Yêu cầu API/DB Slave

##### Bước 5: ETL & DQA

- Raw→Silver: Làm sạch
- Silver→Gold: Tính toán
- Output: Bảng Gold ready

#### GIẢI ĐOẠN 3: TÍCH HỢP & HIỂN THỊ (Ban CNTT chủ trì)

##### Bước 6: Kéo dữ liệu về BI Platform

- Data Engineer cấu hình Airflow
- Pull Gold→ClickHouse
- Transform nhẹ (mapping tên)

##### Bước 7: Dựng Dashboard

- Cấu hình Superset→ClickHouse
- Vẽ biểu đồ, thiết lập filter
- Theo yêu cầu ban đầu

#### GIẢI ĐOẠN 4: UAT & NGHIỆM THU

##### Bước 8: Kiểm thử nội bộ (Internal UAT)

- Data Check: Gold vs Dashboard
- Performance Check: Tốc độ load
- KHDL + Ban CNTT đối chiếu

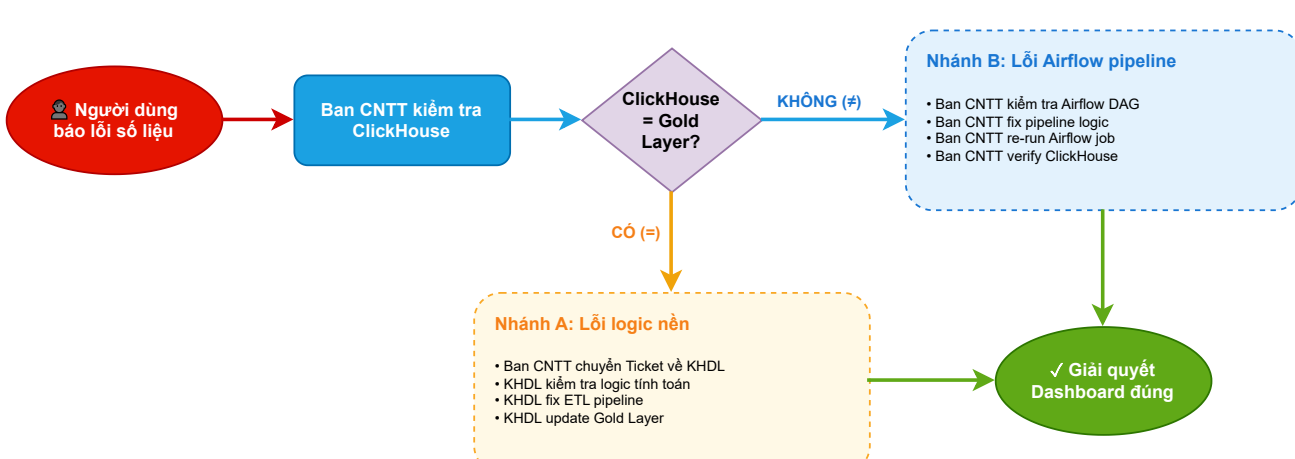
##### Bước 9: Bàn giao người dùng (Client UAT)

- Ban CNTT gửi link Dashboard
- Người dùng xác nhận nghiệm thu
- Đóng Ticket → Live!

✓ HOÀN THÀNH Dashboard Live

## CƠ CHẾ XỬ LÝ LỖI & PHÁT SINH Quy trình khắc phục lỗi và xử lý vấn đề phát sinh từ nguồn

### LỖI: LỆCH SỐ LIỆU (Data Discrepancy)



### PHÁT SINH: THAY ĐỔI CẤU TRÚC NGUỒN (Schema Drift)

#### QUY TRÌNH XỬ LÝ SCHEMA DRIFT (Timeline):

T-3 ngày

Source Team báo cho KHDL về thay đổi cấu trúc DB

T-2 ngày

KHDL phân tích impact KHDL update ETL pipeline (Raw→Silver→Gold)

T-1 ngày

KHDL báo cho Ban CNTT Ban CNTT update Airflow (nếu cần)

T-0 (Deploy)

Source Team deploy thay đổi cấu trúc Pipeline hoạt động bình thường

⚠ LƯU Ý QUAN TRỌNG: Source Team PHẢI báo trước 3 ngày để KHDL có thời gian chuẩn bị!

#### NGUYÊN TẮC XỬ LÝ:

##### 1. Xác định ranh giới trách nhiệm rõ ràng:

- KHDL chịu trách nhiệm về độ chính xác dữ liệu (Gold Layer)
- Ban CNTT chịu trách nhiệm về hiệu năng & hiển thị (ClickHouse/Superset)

##### 2. Giao tiếp chủ động & kịp thời:

- Source Team phải báo trước 3 ngày khi thay đổi cấu trúc
- KHDL phải thông báo Ban CNTT khi update Gold Layer

##### 3. Kiểm tra đối chiếu (Data Reconciliation):

- Luôn so sánh: Gold Layer ↔ ClickHouse ↔ Dashboard