

THIẾT KẾ HỆ THỐNG DATA LAKE



DỮ LIỆU VIETTEL TELECOM



TĂNG TRƯỞNG DOANH
THU THOẠI SO VỚI NĂM TRƯỚC?

TỈNH NÀO CÓ NHIỀU
THUÊ BAO NHẤT?

DOANH THU GÓI V90?

HÔM NAY THU ĐƯỢC
BAO NHIÊU TIỀN?

THUÊ BAO NÀO CÓ
NGUY CƠ RỜI MẠNG?

GÓI KHUYẾN MẠI A PHÙ HỢP
VỚI THUÊ BAO X HAY Y?

TỈ LỆ ĐĂNG KÝ DỊCH VỤ KHI
TRUYỀN THÔNG GÓI KHUYẾN MẠI?



100+ triệu

thuê bao tại Việt Nam
và 10 thị trường nước ngoài



60+ tỷ/ ngày

bản ghi transaction: Voice,
SMS, Vas, MSC, URL...



24/7

dữ liệu real-time 24/7, 365
ngày 1 năm.

MỤC LỤC

- 1 TỔNG QUAN
- 2 CÁC KHÁI NIỆM CƠ BẢN
- 3 ỨNG DỤNG & GIÁ TRỊ MANG LẠI DEMO
- 4 QUY TRÌNH XÂY DỰNG
- 5 THIẾT KẾ

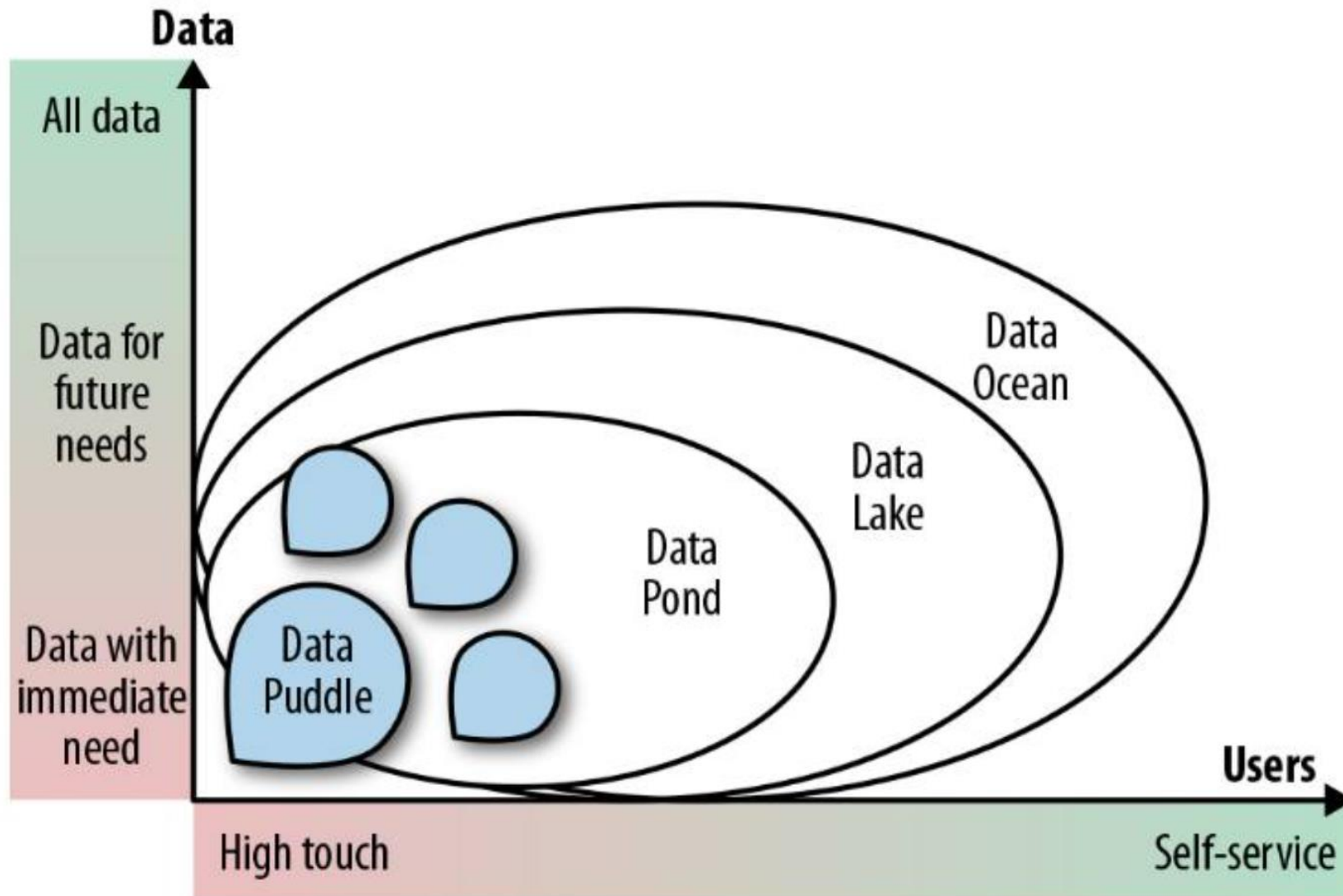
TỔNG QUAN

Data Lake là gì ?

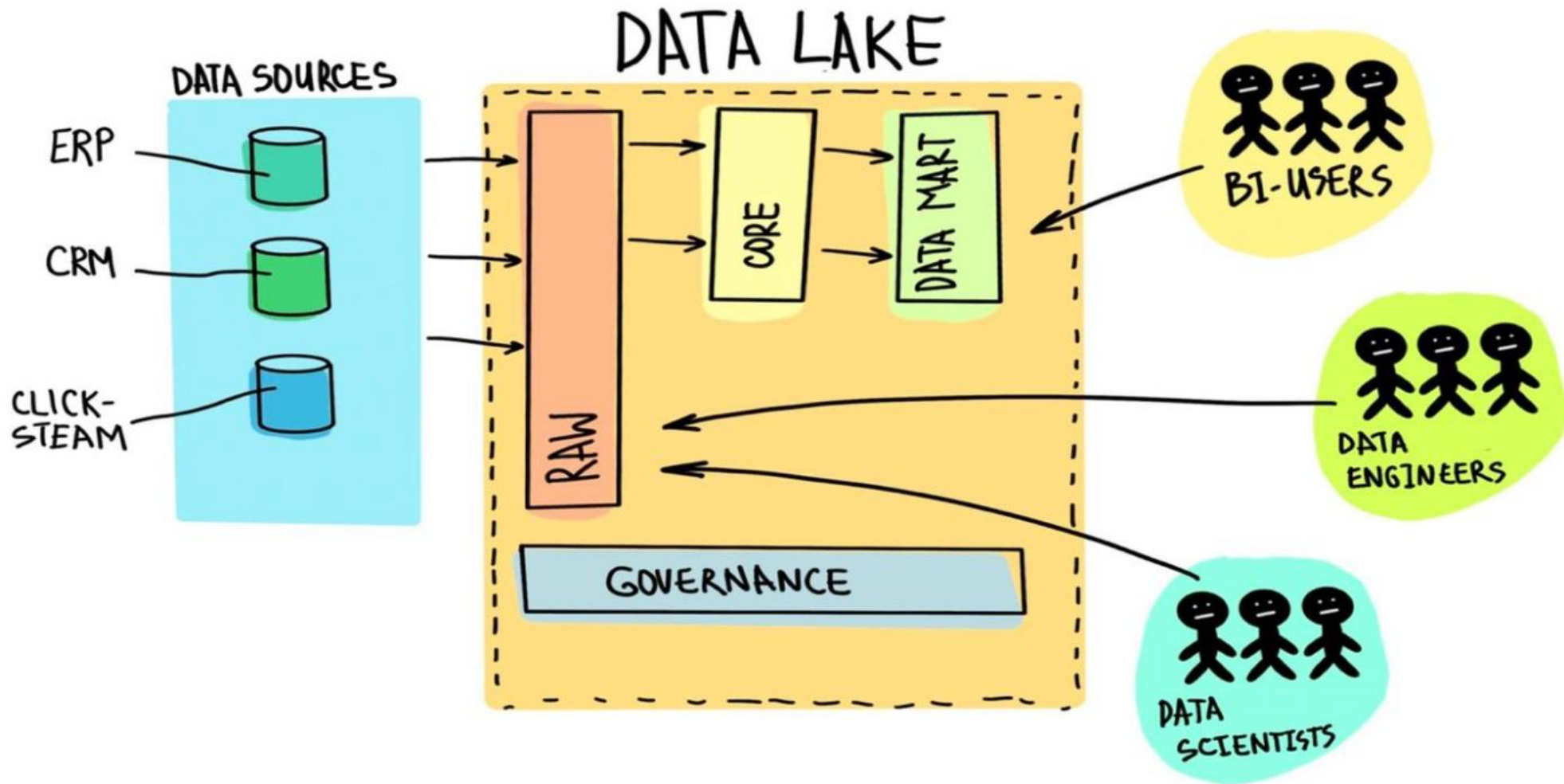
Data Lake là một siêu Kho lưu trữ mà bạn có thể lưu trữ khối lượng rất lớn dữ liệu không cấu trúc, bán cấu trúc và cấu trúc. Nó được tạo ra để lưu mọi loại dữ liệu với định dạng gốc mà không giới hạn dung lượng, bản ghi hay số file. Nó lưu trữ một lượng dữ liệu lớn để tăng khả năng phân tích và tích hợp đa nền tảng.



CÁC KHÁI NIỆM CƠ BẢN TRONG DATA LAKE



CÁC KHÁI NIỆM CƠ BẢN TRONG DATA LAKE



CÁC KHÁI NIỆM CƠ BẢN TRONG DATA LAKE

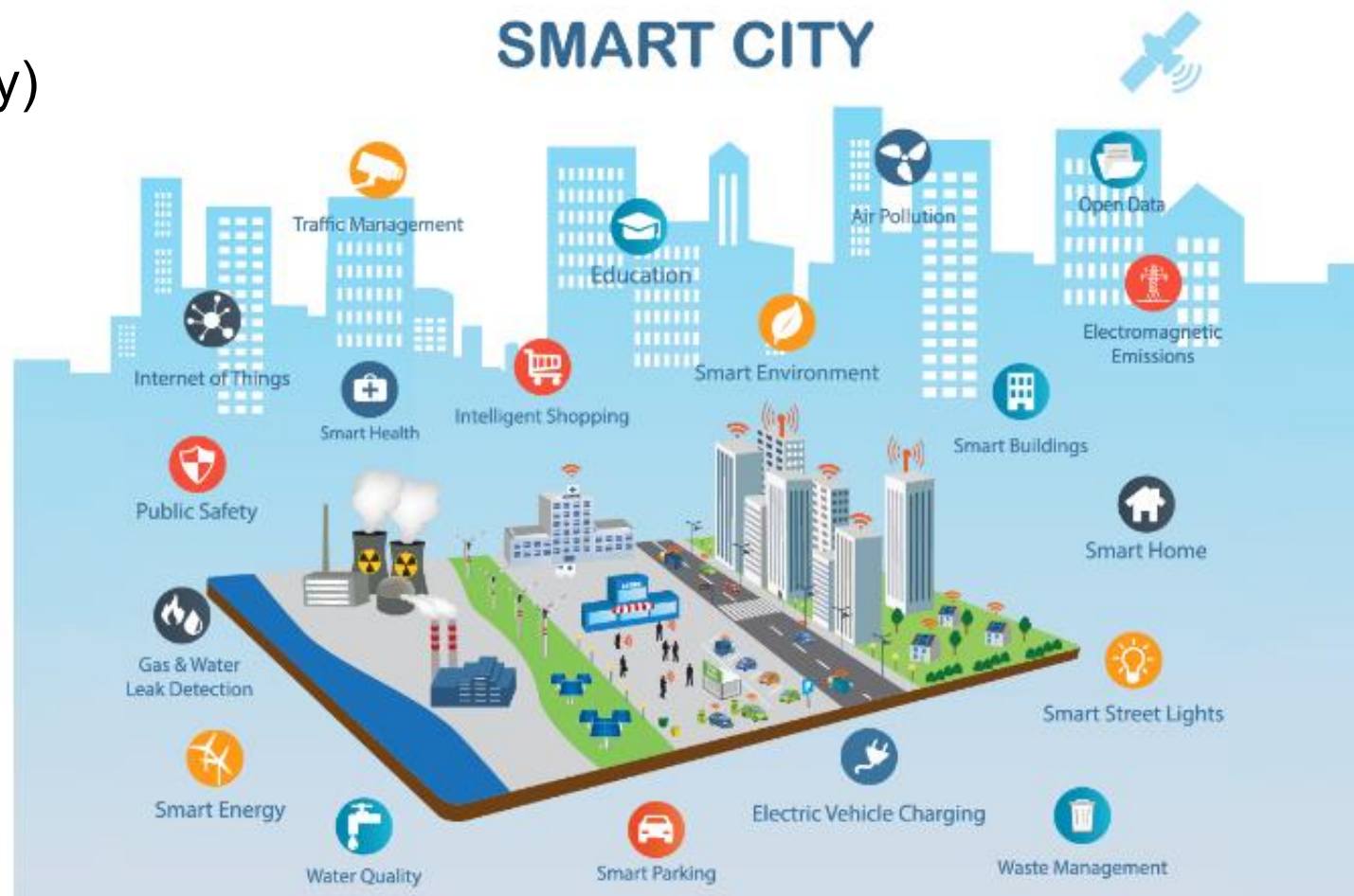


ỨNG DỤNG & GIÁ TRỊ MANG LẠI

Các công ty lớn đang sử dụng Data Lake như thế nào

Hiện có 6 nhóm ứng dụng phổ biến của Data Lake cho các doanh nghiệp:

1. Thành phố thông minh (Smart City)
2. Internet vạn vật (IOT)
3. Khoa học và đời sống
4. An ninh mạng và Bảo mật
5. Khách hàng và Marketing
6. T ư vấn và hỗ trợ



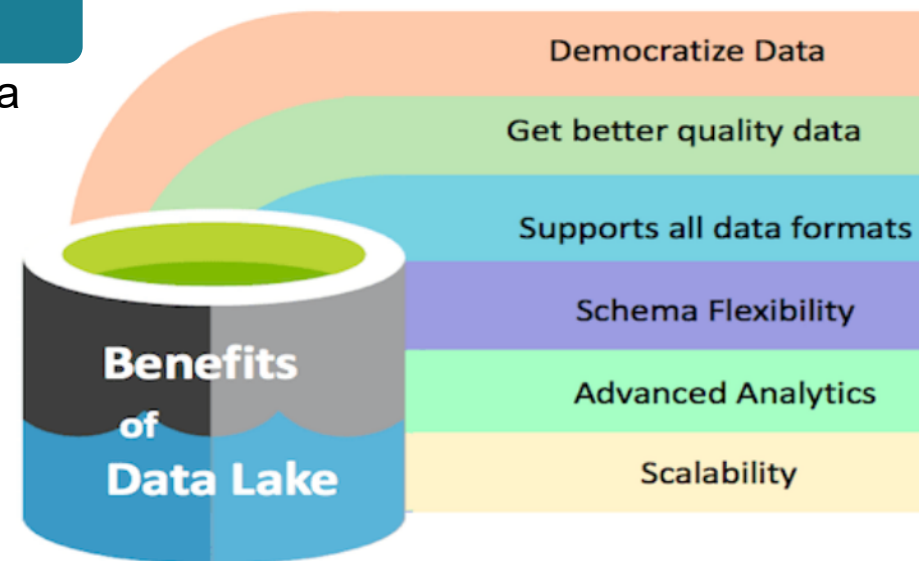
Vậy Data Lake có những lợi ích và bất cập gì ?

Lợi ích

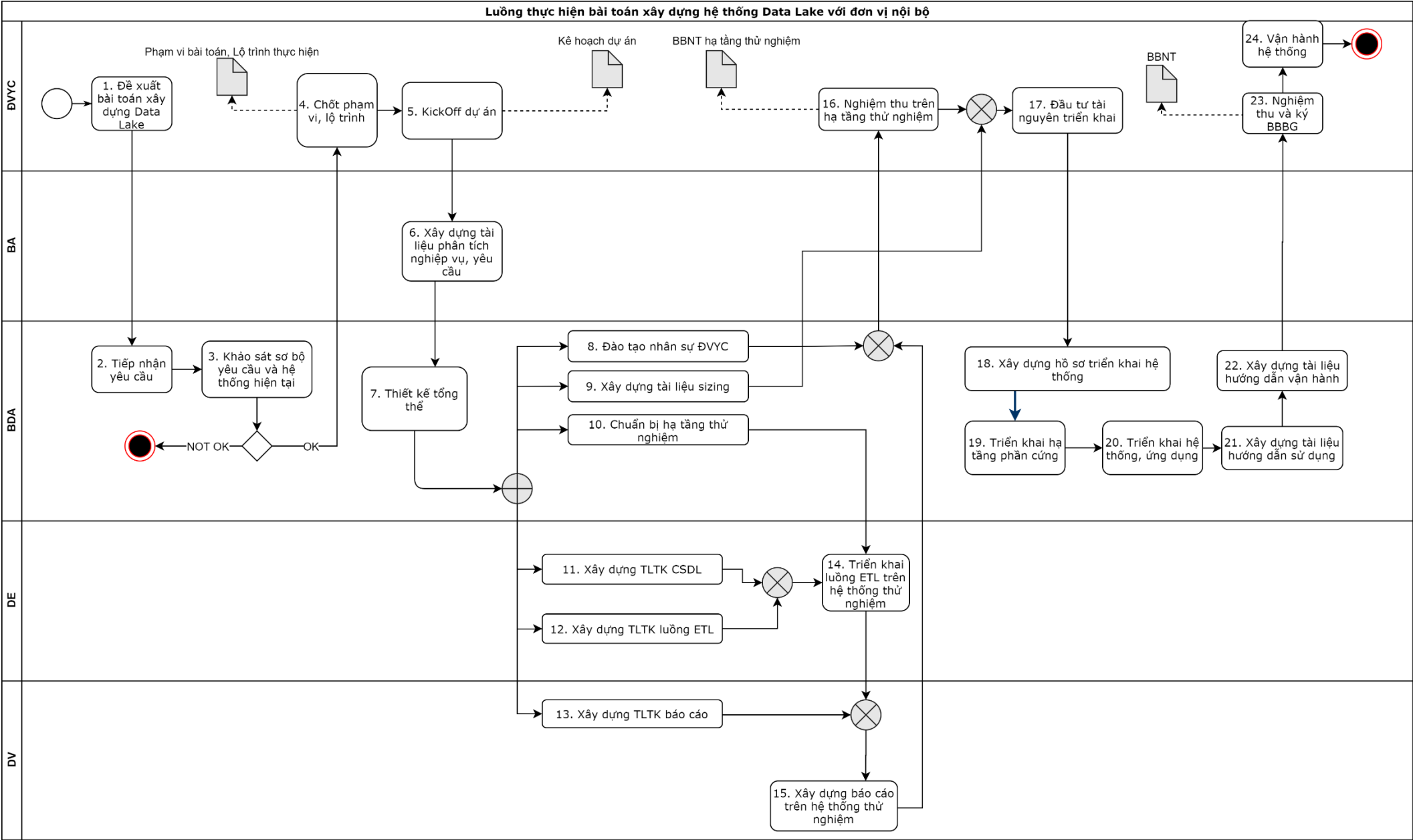
- Hệ thống hóa toàn bộ quy trình xử lý, khai thác và phân tích dữ liệu
- Cung cấp khả năng mở rộng linh hoạt và tối ưu chi phí
- Cung cấp khả năng không giới hạn về định dạng và lưu trữ dữ liệu
- Giảm chi phí sở hữu, lưu trữ dữ liệu lâu dài
- Cho phép lưu trữ hợp lý về chi phí
- Thích nghi nhanh với việc thay đổi
- Tập trung được toàn bộ các nguồn dữ liệu từ nhiều nguồn khác nhau
- Người dùng từ mọi nơi trên toàn cầu có thể dễ dàng truy cập và khai thác

Rủi ro:

- Sau một thời gian Data Lake có thể mất đi sự liên kết và ràng buộc giữa các thành phần
- Rủi ro lớn trong thiết kế do kiến trúc gồm rất nhiều thành phần và phân hệ phức tạp
- Việc lưu trữ lớn dữ liệu không cấu trúc có thể dẫn tới rối loạn dữ liệu, phức tạp và không thể khai thác được
- Do lưu trữ lâu và nhiều hơn lên cũng làm tăng chi phí
- Khó kiểm soát bảo mật và truy cập



QUY TRÌNH XÂY DỰNG DATA LAKE



MÔ HÌNH DỰ ÁN DATA LAKE

| Vai trò \ Đơn vị | Ban QTDL | Đơn vị |
|--|----------|--------|
| Tư vấn (Consultant) | ★ | |
| Quản trị dự án (Project Manager) | ★ | ★ |
| Thiết kế kiến trúc (Big Data Architecture) | ★ | |
| Khoa học dữ liệu (Data Scientist) | ★ | ★ |
| Phân tích nghiệp vụ (Business Analyst) | ★ | ★ |
| Xây dựng luồng dữ liệu (Data Engineer) | ★ | ★ |
| Phân tích dữ liệu (Data Analyst) | ★ | ★ |
| Kiểm thử (Tester) | | ★ |

THÀNH PHẦN CHÍNH CỦA DATA LAKE



BUSINESS ANALYST

PHÂN TÍCH KINH DOANH

Các nhà lãnh đạo Doanh nghiệp đưa ra những quyết định chiến lược đối với hoạt động kinh doanh của doanh nghiệp.



DATA MINING

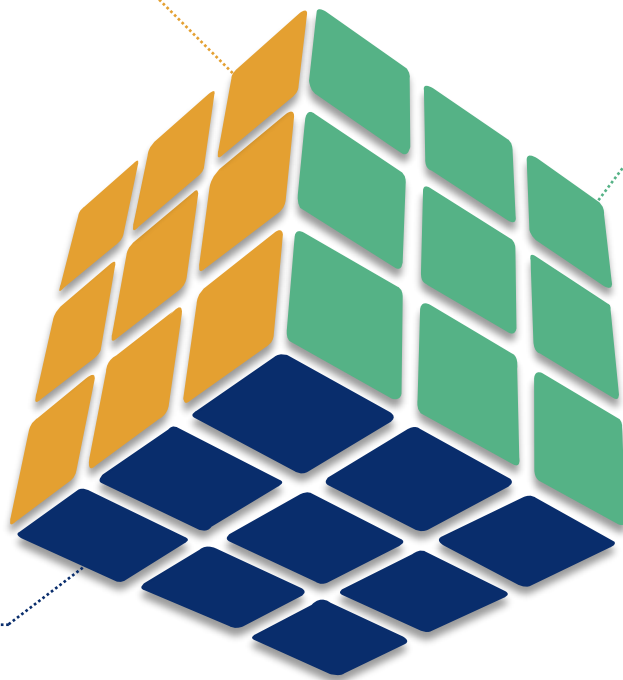
KHAÍ PHÁ DỮ LIỆU

Các kỹ thuật dùng để khai phá dữ liệu và phát hiện tri thức như phân loại (*Classification*), phân nhóm (*Clustering*), phát hiện luật kết hợp (*Association Rule*), Dự đoán (*Predcition*)

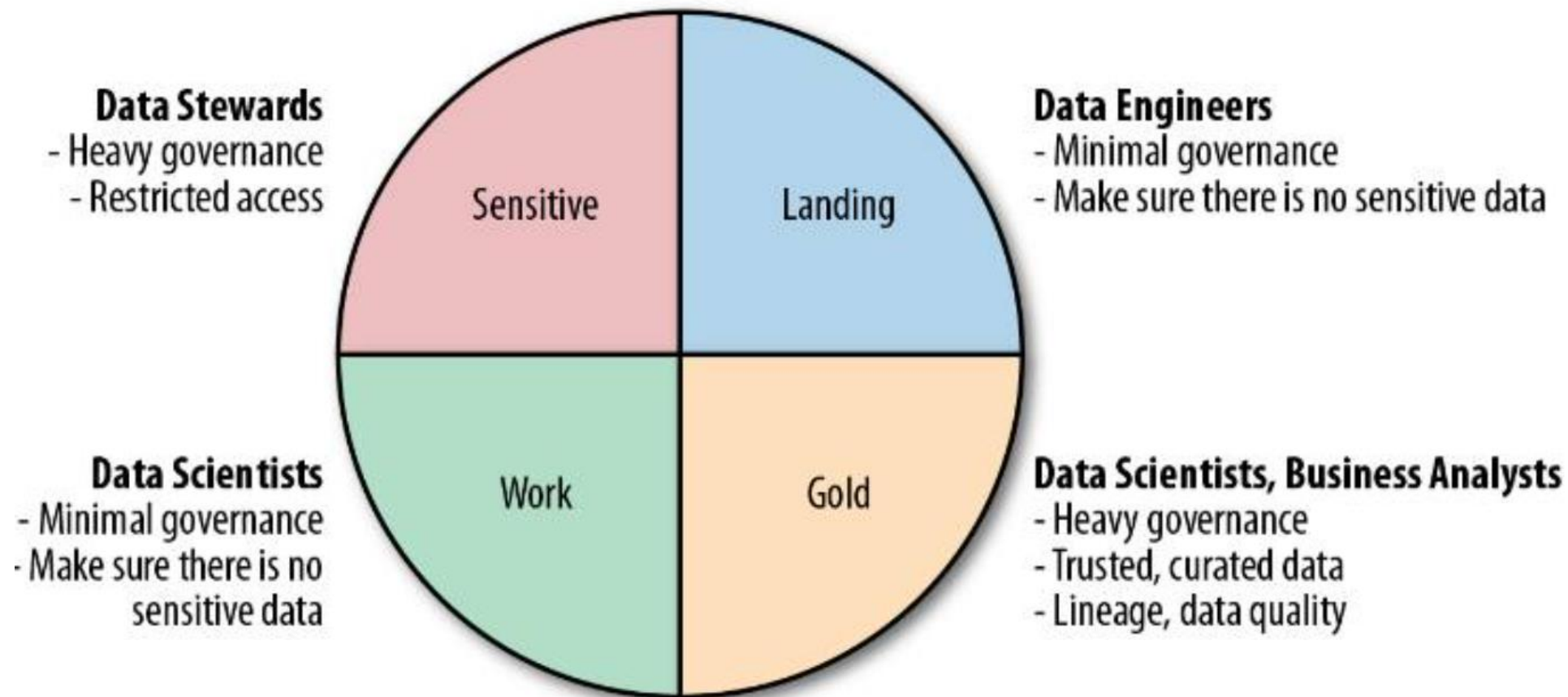
DATA WAREHOUSE

KHO DỮ LIỆU

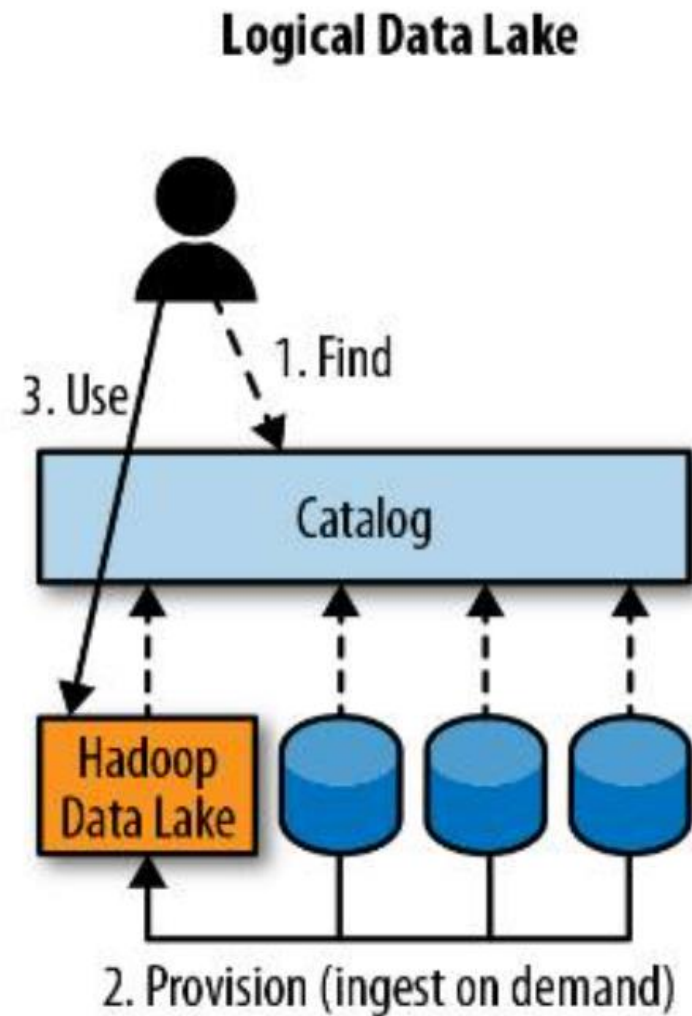
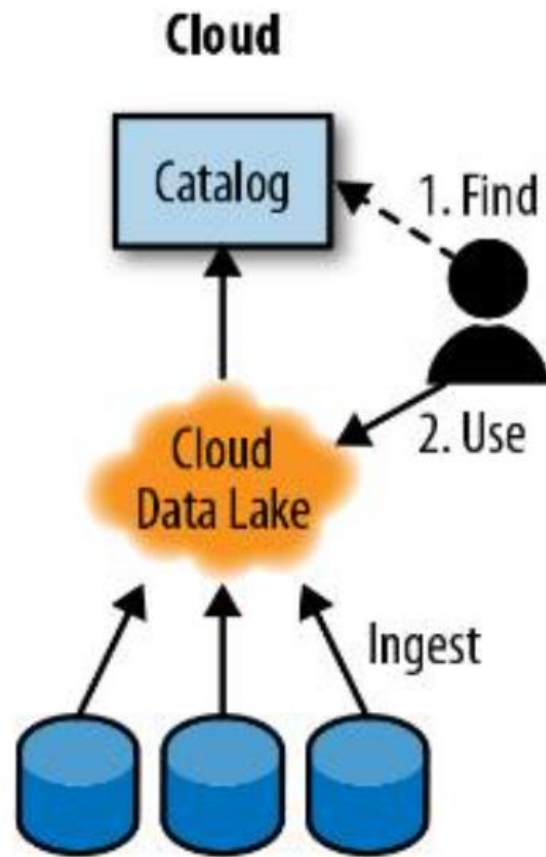
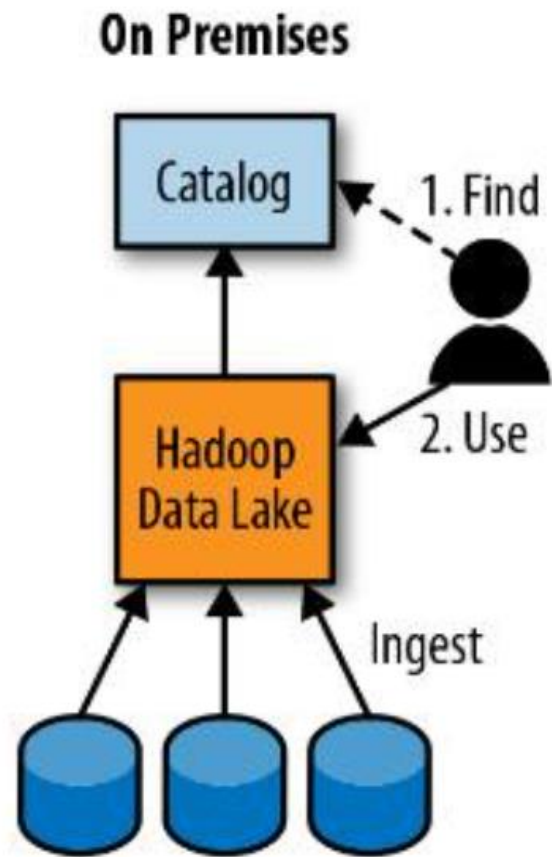
Chứa dữ liệu tổng hợp của doanh nghiệp



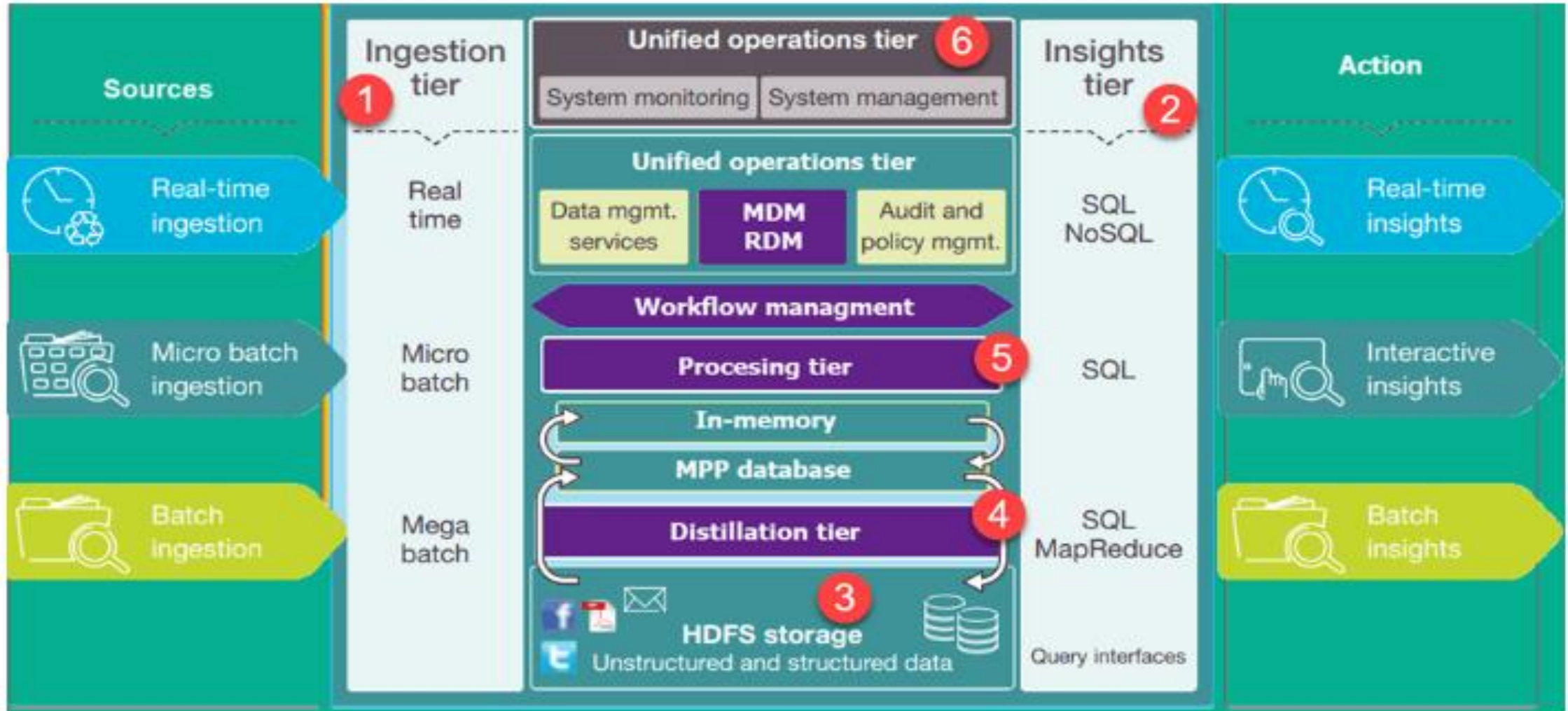
KIẾN TRÚC CƠ BẢN CỦA MỘT DATA LAKE



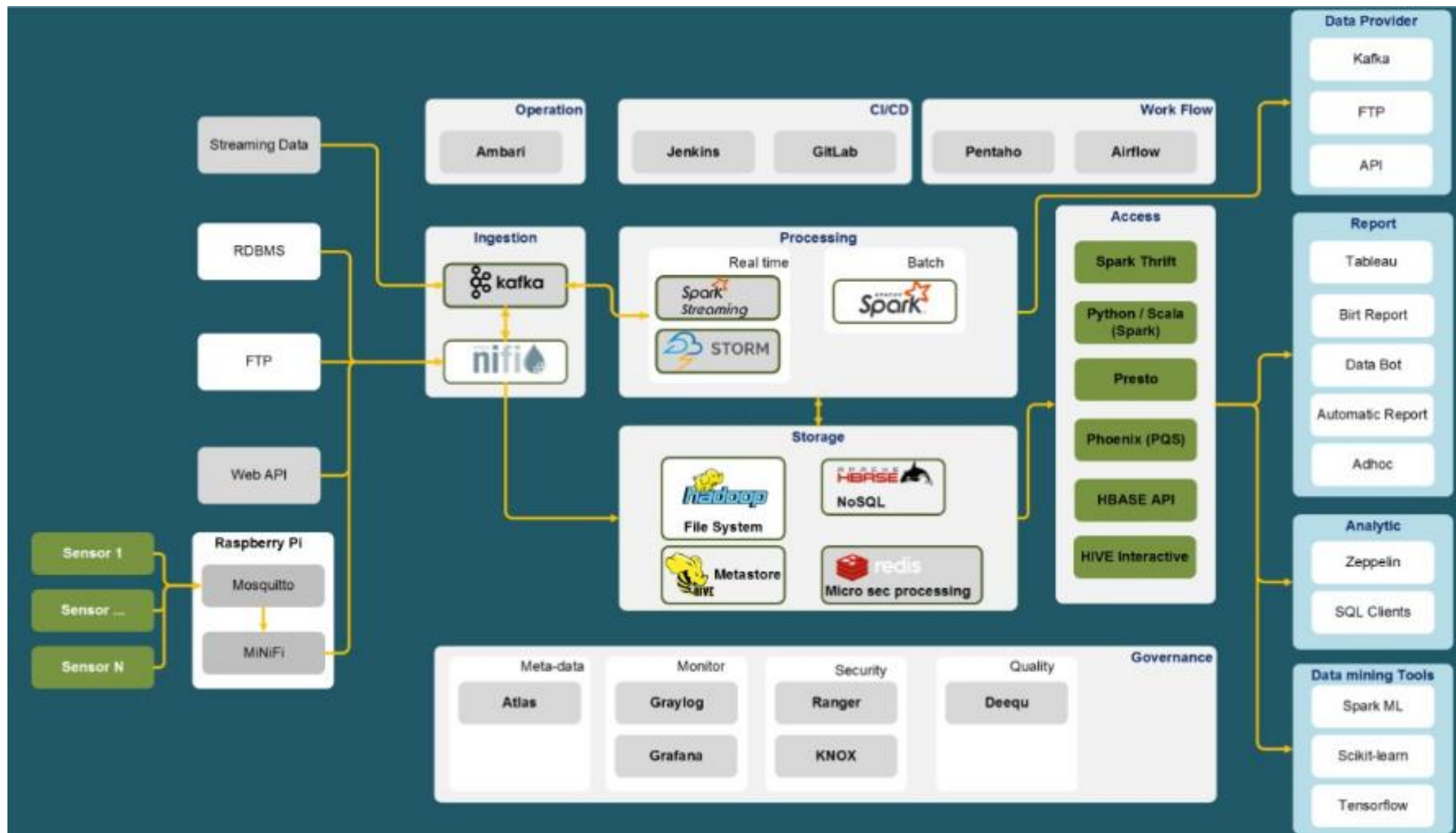
KIẾN TRÚC CƠ BẢN CỦA MỘT DATA LAKE



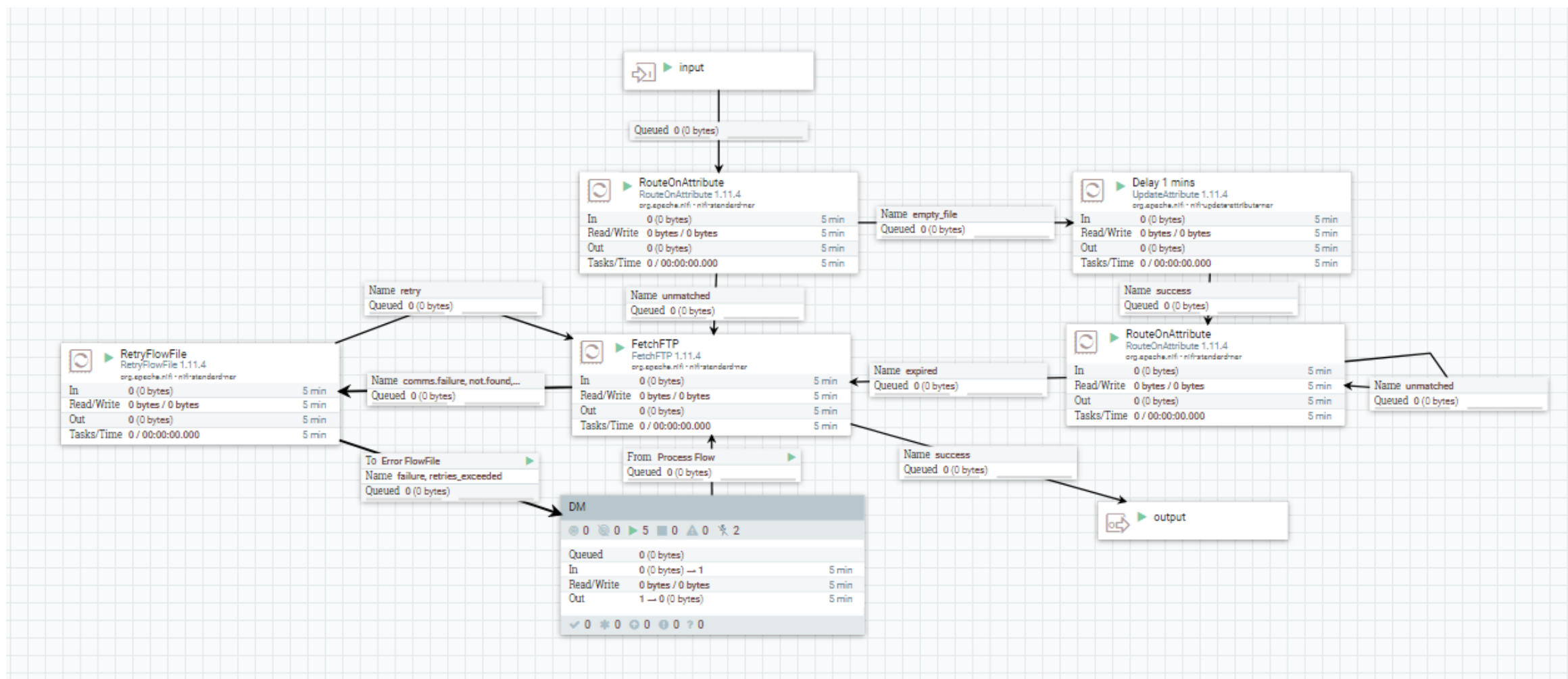
KIẾN TRÚC CƠ BẢN CỦA MỘT DATA LAKE



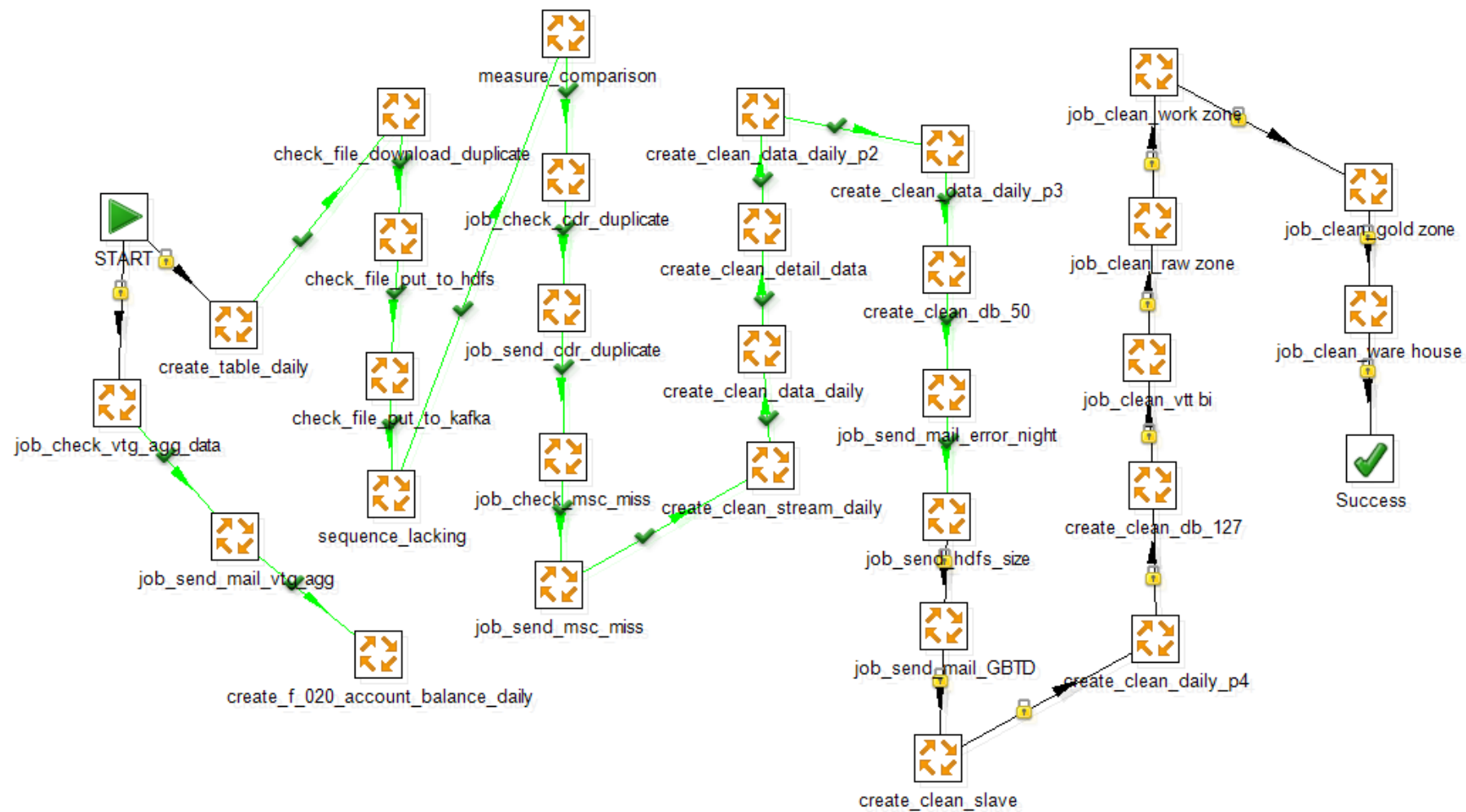
KIẾN TRÚC CƠ BẢN CỦA MỘT DATA LAKE



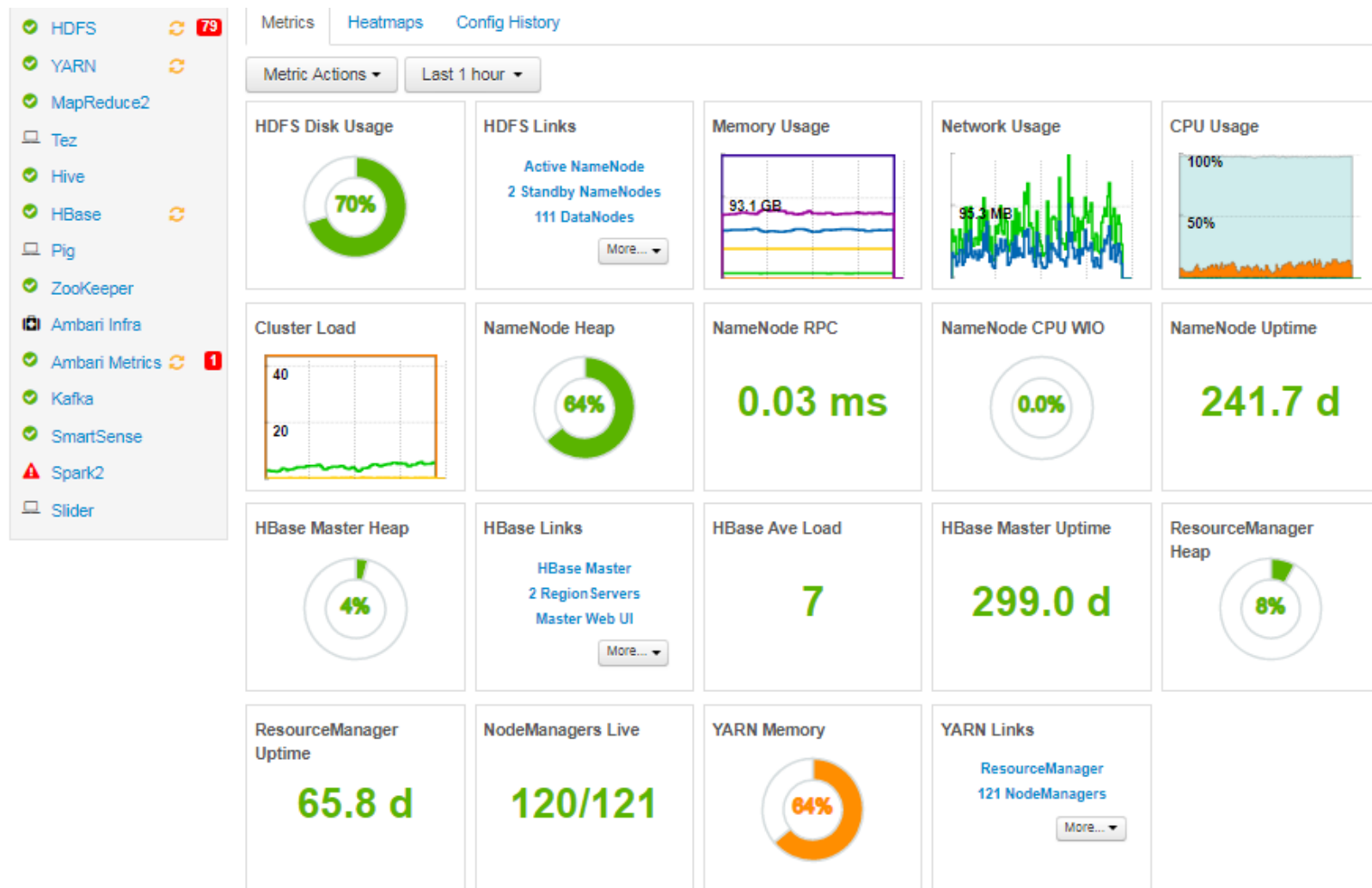
ỨNG DỤNG TRÊN DATA LAKE



ỨNG DỤNG TRÊN DATA LAKE



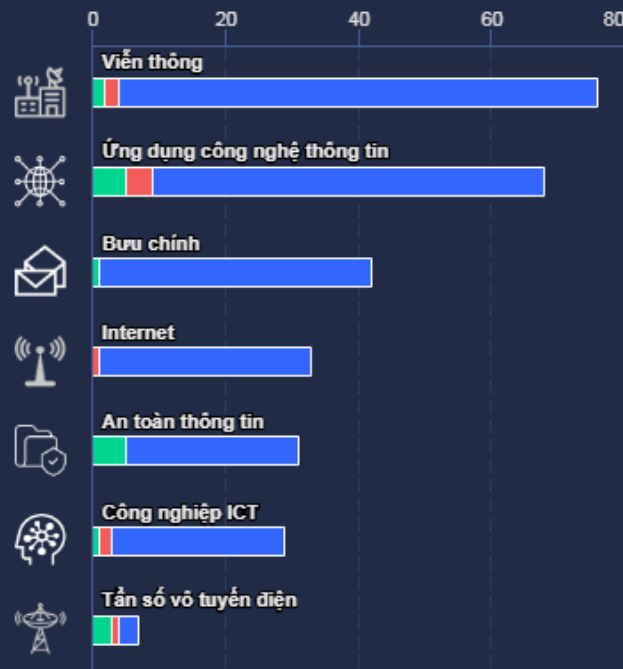
ỨNG DỤNG TRÊN DATA LAKE



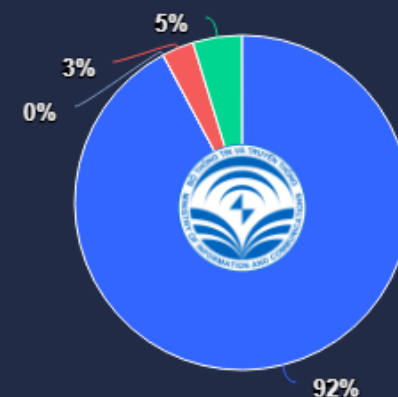
ỨNG DỤNG TRÊN DATA LAKE

Trang chủ > Tổng quan

Tổng quan tháng 08/2020

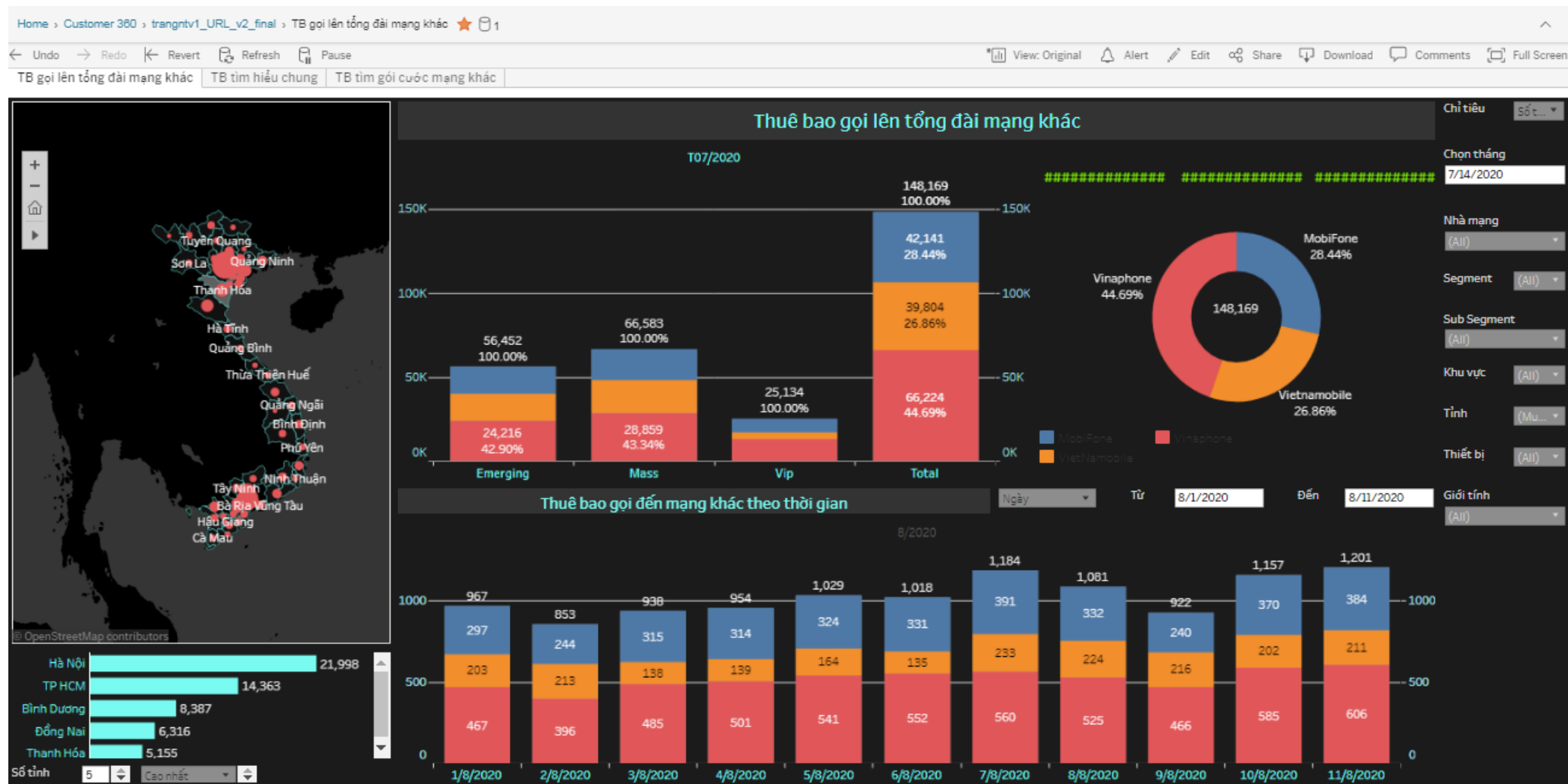


Mức độ hoàn thành

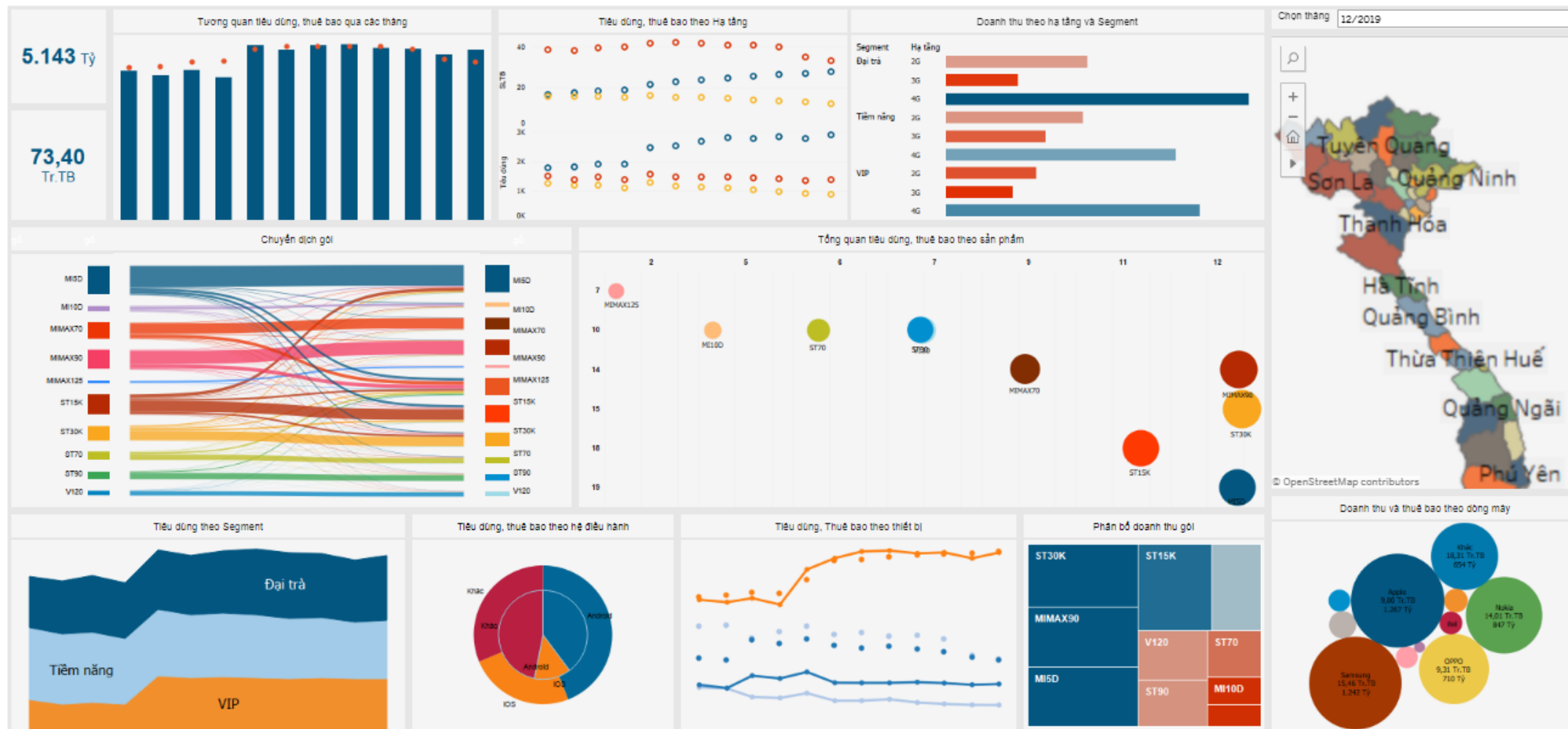


486 Tổng số chỉ tiêu
448 Chỉ tiêu chưa xác định mục tiêu
0 Chỉ tiêu chưa được cập nhật số liệu
15 Chỉ tiêu chưa đạt mục tiêu
23 Chỉ tiêu đạt, vượt mục tiêu

ỨNG DỤNG TRÊN DATA LAKE



ỨNG DỤNG TRÊN DATA LAKE



Trân trọng cảm ơn!