**Project Data Pipeline for Recruitment Start Up**

**Objectives : Tính toán tổng hợp số lượt click , số ứng viên ứng tuyển , số người đạt chuẩn , không đạt chuẩn của các job được chạy trên website.**

**Data Flow :**

**Diagram

Description automatically generated**

**Data Structure**

**In datalake :**

**Chứa dữ liệu thô ( logs data ) về 1 cái event vừa xảy ra**

**In data warehouse ( Event đã được tổng hợp )**

**( kiểu vào giờ đó , ngày đó , cái job này , có bao nhiêu người click vào , của campaign nào , nhà tuyển dụng nào , nhóm nào ,... )**

Diagram

Description automatically generated with medium confidence

**Input : Logs từ datalake , các bảng dimension trong Datawarehouse**

**Output : Events table trong Datawarehouse**

**Requirements : Chạy realtime hoặc near real time , và job được viết bằng Spark ( Spark Job )**