

Real-Time Engagement Pipeline

thmanyah-task

Candidate: **Sara Bajaba**

Date: 4 August 2025

1. Executive Summary / الملخص التنفيذي

This project demonstrates a production-grade streaming pipeline that enriches engagement events in real time. Raw events are published by a Python producer to a Kafka topic (`engagements`). Apache Flink 1.19 (SQL API) consumes the stream, joins each event with content metadata stored in PostgreSQL, and writes the enriched records to `processed_engagements`. The design showcases exactly-once semantics, ANSI-SQL code readability, and an architecture that can scale horizontally.

2. Quick-Start / طريقة التشغيل السريعة

Prerequisites: Docker Desktop \geq 24, 8 GB RAM.

```
```bash
git clone https://github.com/Saraalii/thmanyah-task.git
cd thmanyah-task
docker compose up -d
open http://localhost:8081 # Flink UI
```
```