

Kafka y Procesamiento en Tiempo Real

Tarea fin de módulo

Autor: Oscar Gómez

Actualizado Enero 2024

OBJETIVO.....	3
Kafka Connect.....	3
Arquitectura.....	4
Worker.....	4
Plugin.....	5
Connector.....	5
Conceptos Básicos.....	6
Particionado.....	6
Paralelismo y Escalabilidad.....	7
Connect API.....	10
Ecosistema de Conectores.....	10
Recomendación de Instalación de plugins.....	10
Kafka Streams.....	11
Arquitectura.....	11
Processing Topology.....	12
Modelo de Paralelismo.....	13
Estado.....	14
Semánticas de Entrega y State Stores.....	14
Conceptos Principales.....	15
La Dualidad Stream-Table.....	15
KStream.....	17
KTable.....	17
GlobalKTable.....	17
Agregaciones.....	18
Joins.....	18
Operaciones basadas en Ventana de tiempo.....	18
KSQL.....	19
Arquitectura.....	19
ksqlDB Engine.....	20
ksqlDB CLI.....	20
Conceptos.....	21
Data Definition Language (DDL) Statements.....	21
Data Manipulation Language (DML) Statements.....	21

OBJETIVO

El objetivo de la tarea es familiarizar al alumno con los procesos con las herramientas del Ecosistema Kafka y como estas, desde su orientación al streaming de datos pueden usarse tanto para casos de uso en tiempo real como “batch”

DESCRIPCIÓN

Partiendo de los datos del [Topic flights](#) resultado de la última semana de trabajo se pide:

1. Realizar las transformaciones necesarias (bien con KSQL o Kafka Streams) para responder las siguientes preguntas:
 - a. Cuantos vuelos llegaron a [New Orleans, LA](#) entre el 1 y el 20 de Enero de 2023
 - b.Cuál fue la ciudad de la que más vuelos partieron en Enero de 2023
2. Realizar las transformaciones necesarias para usando el [JDBC Sink Connector](#) escribamos en una tabla flights en Postgre SQL cumpliendo con el siguiente modelo:

Campo en Tabla	Campo en Topic
date	flightDate
flightNumber	Flight_Number_Reporting_Airline
originCity	OriginCityName
destinationCity	DestCityName

REALIZACIÓN Y ENTREGA

Como “infraestructura” base para la realización de la tarea se utilizará el apartado **TAREA** del repositorio de ejercicios del módulo.

El entregable será una descarga comprimida de dicho repositorio conteniendo las soluciones tal como se describen en el README del apartado.