

Task for home preparation

The tasks below aim to give us a perspective over your high-level competence in data management, message brokers & python literacy. They are not designed to validate your seniority but to get a perspective over your preferred work approach.

Please, send us the result in a format convenient for you.

There is no right/wrong answer, we want to understand your preferred work approach. Should you be invited to our technical interview, we will revise the tasks together and likely dive into more details.

A GDPR compliant company has tasked you with the creation of an ecommerce.db consisting of the following tables:

orders (order_id,customer_id,order_date,total_amount,status,shipping_address,payment_method,currency)
customers (customer_id,first_name,last_name,age,email,country,city,postal_code,phone_number,registration_date,last_login_date)
products (product_id,name,category,subcategory,price,cost,supplier_id,stock_quantity,weight,dimensions)
order_items (order_id,product_id,quantity,unit_price,discount_amount,total_price)
suppliers (supplier_id,company_name,contact_name,email,phone,country,lead_time)
product_reviews (review_id,product_id,customer_id,rating,review_text,review_date)

Implementation requirements

Database Creation:

- Develop SQL scripts to create the above-mentioned tables with appropriate data types, constraints, and indexes.
- Implement necessary foreign key relationships and ensure referential integrity.

Data Generation:

- Create a Python script to generate realistic sample data for all tables.
- Ensure data consistency across related tables and implement logic for realistic order histories, product inventories, and customer behaviors.

Data Streaming Infrastructure:

- Implement Kafka producer in Python that simulates real-time data generation for "customers" table and writes the data to "customers" Kafka topic.
- Develop a Kafka consumer that processes the streaming data and insert it into the corresponding table.
- Implement error handling and logging for both producers and consumers.

Data Validation and Transformation:

- Design and implement data validation checks in the Kafka consumers to ensure data quality before insertion into the database.
- Create transformation logic to handle any necessary data type conversions or derived fields.