Data Engineer - Technical Assessment

Background:
You have been hired by a fictional company called "DataCo" to build a real-time data pipeline that streams data from Kafka, processes the data in a data store, and indexes the processed data in Elasticsearch. The data pipeline will be used to ingest and analyze clickstream data from a web application.

Task:
You are required to build a data pipeline that performs the following steps:

Ingest clickstream data from Kafka.

Store the ingested data in data store of your choice, with the following schema:

Row key: Unique identifier for each click event.
Column families:
click_data: Contains columns for the user ID, timestamp, and URL of the clicked page.
geo_data: Contains columns for the user's country and city, as determined by their IP address.
user_agent_data: Contains columns for the user's browser, operating system, and device, as determined by their user agent string.
Periodically process the stored clickstream data in any data store by aggregating the data by URL and country, and calculating the number of clicks, unique users, and average time spent on each URL by users from each country.

Index the processed data in Elasticsearch.

Requirements:
You should use the following tools and technologies to build the data pipeline:

Apache Kafka for data ingestion.
Data storage and processing.
Elasticsearch for data indexing and searching.
Apache Spark for data processing and aggregation.
Deliverables:
You should provide the following deliverables:
A brief report that summarizes the approach taken and any assumptions made during the implementation of the data pipeline.

Evaluation Criteria:
The following criteria will be used to evaluate your solution:
Correctness of the implementation
Efficiency and scalability of the data pipeline
Readability and maintainability of the code
Clarity and completeness of the report

You have 48 hours to get back to us. Good luck with the evaluation!