**Coding Assessment:**

1.Implement Processing JSON and CSV data with PySpark

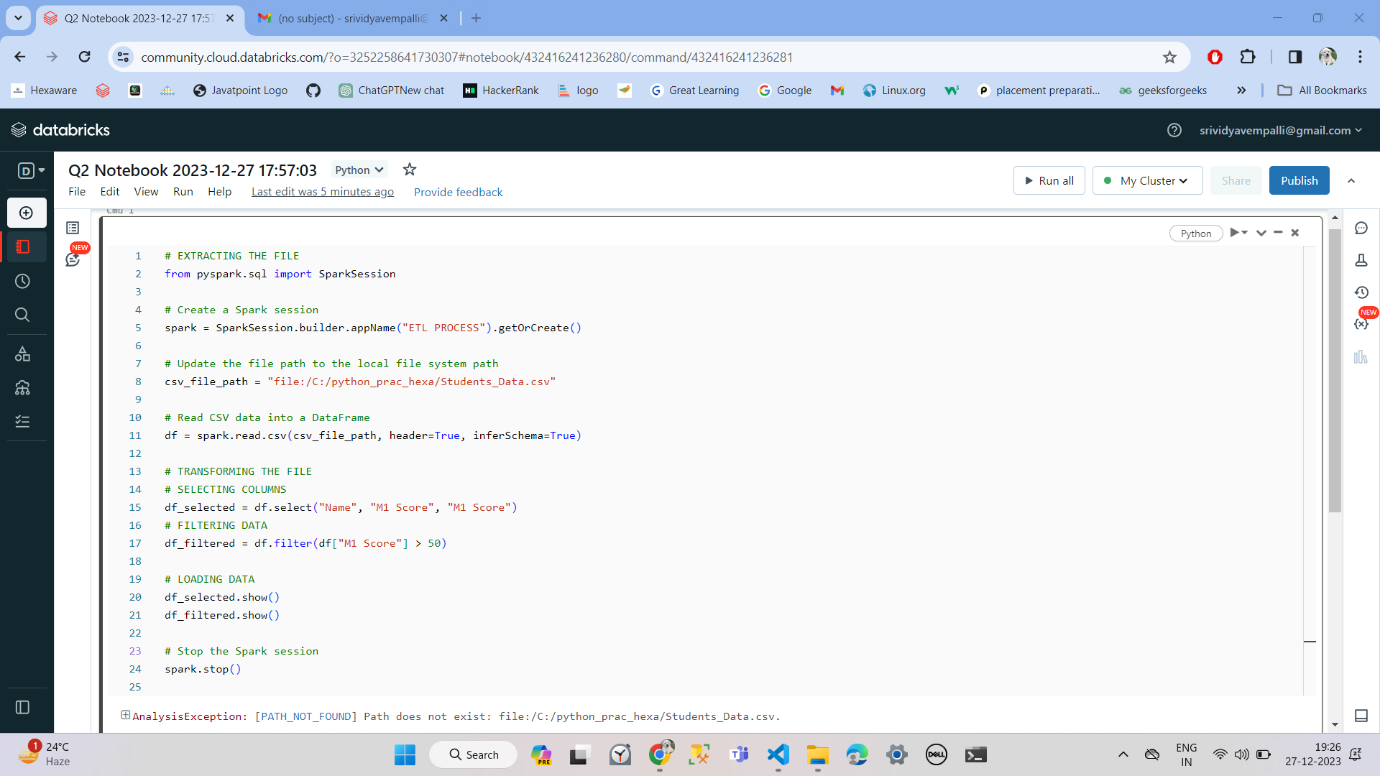
2. Explain ETL (Extract, Transform, Load) with PySpark

3.Using Spark SQL - Creating databases, tables

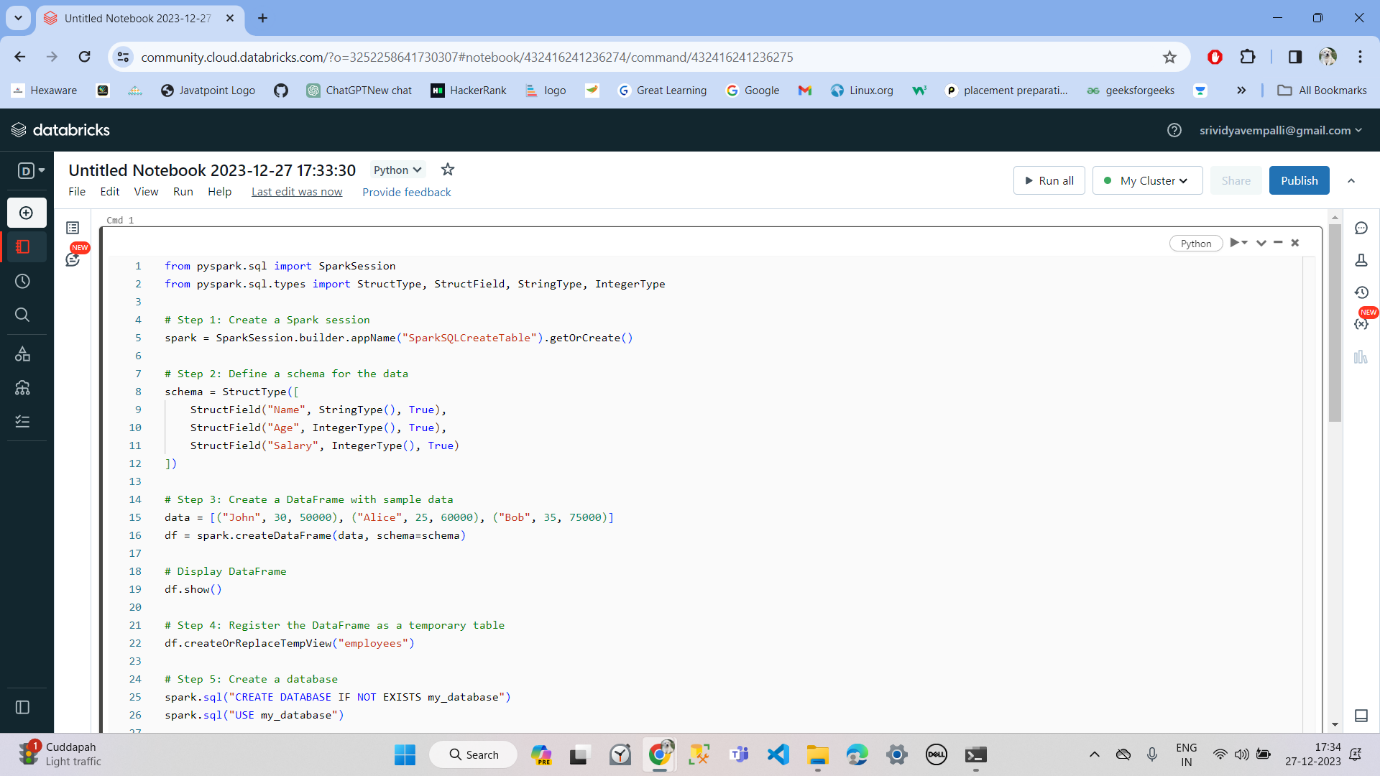
4. Using Spark SQL - Transformations such as Filter, Join, Simple Aggregations, Group By.

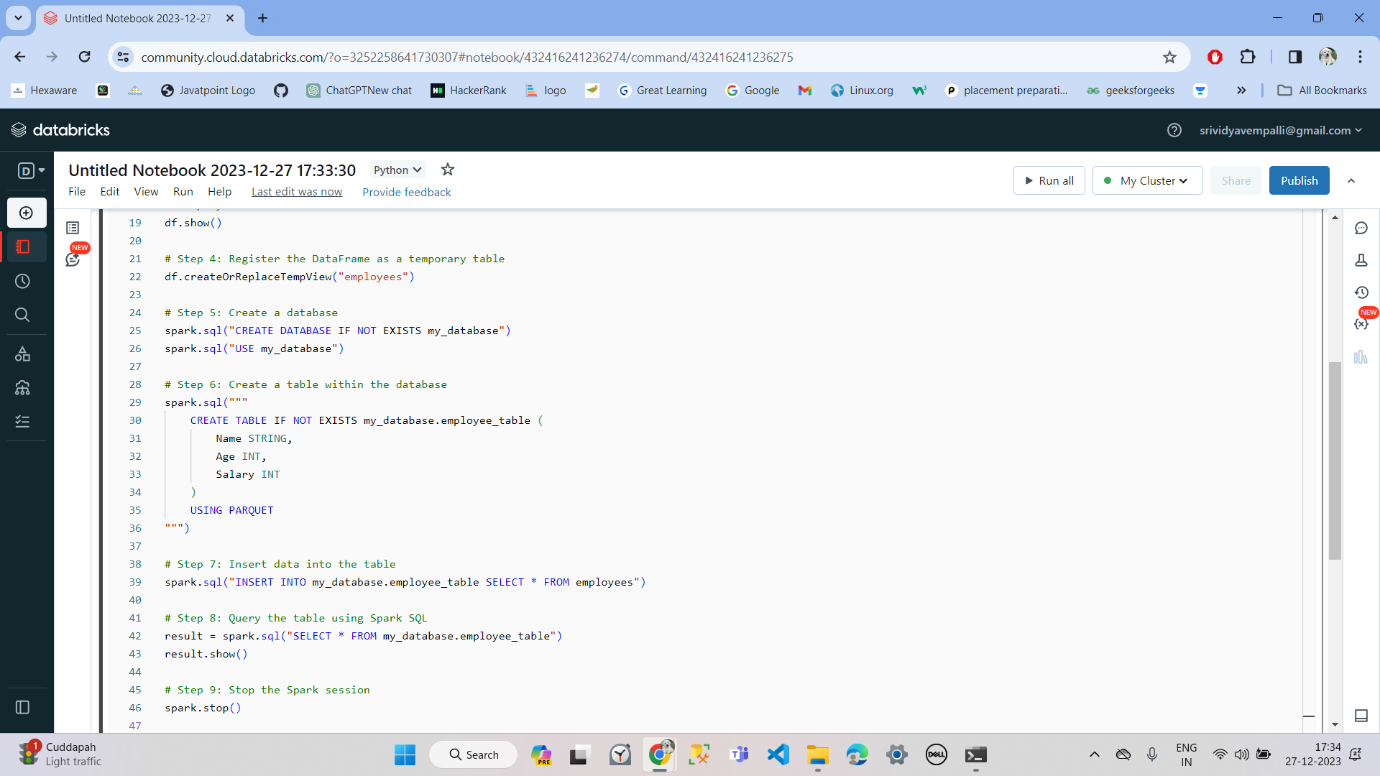
2. ETL is a process commonly used in data warehousing and data integration, and PySpark is a Python library for Apache Spark, a powerful distributed data processing framework. PySpark simplifies the implementation of ETL processes by providing a high-level API for distributed data processing.

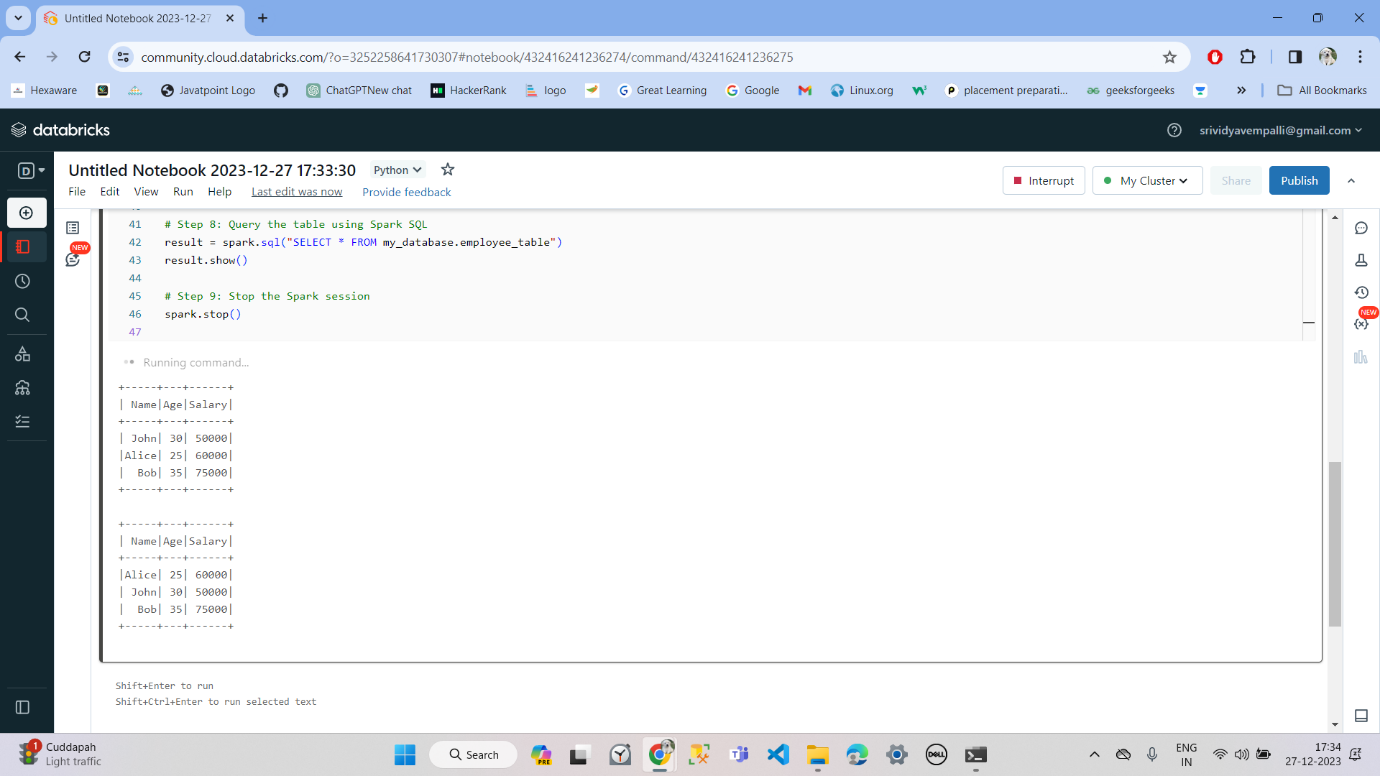
Here's an example of an ETL process using PySpark with a CSV file:



3.







4.

