PROJECT for DATA ENGINEERING Dept. at CSI’24 (May – July):

*Given tasks –*

***1****. Create a pipeline to fetch the 5 countries (india,us,uk,china,russia) data from Rest API*

*(https://restcountries.com/v3.1/name/{name} here replace the {name} with Country name like*

*https://restcountries.com/v3.1/name/us) and save it in separate file as JSON with File name*

*equal to Country name.*

***2.*** *Add the trigger to above pipeline in such a way that it will automatically run two times in a*

*day ( 12:00 AM and 12:00 PM IST)*

***3.*** *Create a pipeline to copy customer data from db to adls only if record count is more than 500.*

*Once a data get copy it should call a child pipeline (which will copy the product data from table if*

*customer record count is &gt; 600).*

***4****. Design the pipeline in such a manner that it will pass the Customer pipeline pass the*

*customer count to the child product pipeline via Pipeline parameter.*

---------------------------------------------------------------------------------------------------------------------------------------

**TASK 1:**

Use the pyhton file named “fetch\_country\_data.py” , it will execute a folder named “country\_data” according to task1.

**TASK 2:**

We will add a time-based trigger to the pipeline that you created in Task 1. This trigger will ensure the pipeline runs automatically at 12:00 AM and 12:00 PM IST.

So, go through the processes mentioned below step by step –

1. ***Create a Trigger*:**

* Go to the **Manage** tab in Azure Data Factory.
* Select **Triggers** and click on **New** to create a new trigger.
* Name the trigger (e.g., "TwiceDailyTrigger").
* Set the type to **Schedule**.
* Define the start time (e.g., the current date).
* Set the recurrence to **12 hours** to make it run at 12:00 AM and 12:00 PM IST.
* Ensure the time zone is set to **India Standard Time (IST)**.

1. ***Attach the Trigger to the Pipeline*:**

* Go to the **Author** tab.
* Open the pipeline you created in Task 1.
* Click on the **Add Trigger** button and select **New/Edit**.
* Select the trigger you created and activate it.

Use “TwiceDailyTrigger.json” file as trigger in upper mention processs.

**TASK 3:**

To copy customer data using pipeline we will follow the process –

**Step 1: *Check Record Count***

1. ***Create a Lookup Activity:***
   * In your new pipeline, add a **Lookup** activity to query the record count from the customer table in your database.
   * Configure the **Lookup** activity to connect to your database and execute a SQL query like SELECT COUNT(\*) AS recordCount FROM Customer.

(Use “lookup.json” file)

1. ***Add a Conditional Activity*:**
   * Add an **If Condition** activity to check if the record count is greater than 500.
   * Set the condition expression to @greater(activity('LookupActivity').output.firstRow.recordCount, 500).

(Use “condition.json” file)

**Step 2: *Copy Data if Condition is Met***

1. ***Inside the If Condition Activity:***
   * If the condition is true, add a **Copy Data** activity to copy the customer data fromyour database to Azure Data Lake Storage (ADLS).
   * Configure the source dataset (Customer table) and the sink dataset (ADLS).

**TASK 4:**

*1.* ***Add a Execute Pipeline Activity:***

* Still inside the **If Condition** activity, add an **Execute Pipeline** activity to call the child pipeline (Product pipeline).
* Pass the customer record count as a parameter to the child pipeline.
* Ensure that the child pipeline has a parameter defined to receive this value.

(Use “productpipeline.json” file)

*2.* ***Child Pipeline to Copy Product Data*:**

* Create a new pipeline for copying product data.
* Define a parameter (e.g., customerRecordCount).
* Add an **If Condition** activity to check if customerRecordCount is greater than 600.
* Inside the **If Condition** activity, add a **Copy Data** activity to copy the product data from your database to ADLS if the condition is true.