

Cairo University
Faculty of Computers and Artificial Intelligence

Third Year
Data Warehouse Assignment
2023

In this assignment you should solve the below **three** questions using SSIS:

1. **[2 marks]** Consume any **REST API** and load the response to database. You don't need to load all the response fields, **3 or 4 is okay for me**.

For example, this is a response from an API that is used to search for universities:

```
1  [
2  {
3      "state-province": null,
4      "domains": [
5          "mga.edu"
6      ],
7      "name": "Middle Georgia State College",
8      "country": "United States",
9      "web_pages": [
10         "http://www.mga.edu/"
11     ],
12     "alpha_two_code": "US"
13 },
14 {
15     "state-province": null,
16     "domains": [
17         "meu.edu.jo"
18     ],
19     "name": "Middle East University",
20     "country": "Jordan",
21     "web_pages": [
22         "http://www.meu.edu.jo/"
23     ],
24     "alpha_two_code": "JO"
25 },
26 > { ...
27 > { ...
28 > { ...
```

You can create a table named “*UNIVERSITY*” with **three** columns: (name, country & alpha_two_code) and load these fields only.

2. [2 marks] Implement **SCD type 6** for the below “EMPLOYEE_Q2” table:

ID	Name	City	Email	Update_Date
1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023
1002	Alaa	Giza	alaa@mail.com	20-04-2023
1003	Samy	Cairo	samy@mail.com	20-04-2023

Notes:

1. The SCD fields are **City and Email**.
 2. Read source data using **Incremental Load**.
3. [2 marks] Load data to a table using **versioning** like below:

3.1 Source table “EMPLOYEE_Q3”

ID	Name	City	Email	Schedule_Date
1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023
1002	Alaa	Giza	alaa@mail.com	20-04-2023
1003	Samy	Cairo	samy@mail.com	20-04-2023

3.2 Target table after first run on the same day

Emp_Key	ID	Name	City	Email	Insert_Date	Active_Flag	Version_No
1	1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023	1	1
2	1002	Alaa	Giza	alaa@mail.com	20-04-2023	1	1
3	1003	Samy	Cairo	samy@mail.com	20-04-2023	1	1

3.3 Target table after second run on the same day

Emp_Key	ID	Name	City	Email	Insert_Date	Active_Flag	Version_No
1	1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023	0	1
2	1002	Alaa	Giza	alaa@mail.com	20-04-2023	0	1
3	1003	Samy	Cairo	samy@mail.com	20-04-2023	0	1
4	1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023	1	2
5	1002	Alaa	Giza	alaa@mail.com	20-04-2023	1	2
6	1003	Samy	Cairo	samy@mail.com	20-04-2023	1	2

3.4 Target table after first run on the next day (just change Schedule Date in the source data to simulate to next day)

Emp_Key	ID	Name	City	Email	Insert_Date	Active_Flag	Version_No
1	1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023	0	1
2	1002	Alaa	Giza	alaa@mail.com	20-04-2023	0	1
3	1003	Samy	Cairo	samy@mail.com	20-04-2023	0	1
4	1001	Ahmed	Cairo	ahmed@mail.com	20-04-2023	0	2
5	1002	Alaa	Giza	alaa@mail.com	20-04-2023	0	2
6	1003	Samy	Cairo	samy@mail.com	20-04-2023	0	2
7	1001	Ahmed	Cairo	ahmed@mail.com	21-04-2023	1	1
8	1002	Alaa	Giza	alaa@mail.com	21-04-2023	1	1
9	1003	Samy	Cairo	samy@mail.com	21-04-2023	1	1

Notes:

1. I have shown two runs only on the same day but during discussion I can run **n** times and see the behavior of your solution.
2. Add **new version** as long as you are running on the **same** day and **close** all old records.
3. Start from **version 1** again on the **next** day and so on.
4. **Don't** check for any change in the source data, load it as it is.

General Notes:

1. No late submission will be accepted for any reason.
2. Team should consist of **two** students only.
3. **All** team members must present on the discussion dates. **ZERO** points will be given to the absentees.

Where to submit?

1. Prepare a **zip** file that contains three folders, one for each problem, and name it as follows:

DWH_Assignment_TA_Name_ID1_ID2.zip

Ex: DWH_Assignment_Ahmed_Galal_20200011_20200022.zip

2. Upload the file to this Google Form: <https://forms.gle/4Zt8eMhJcaeWnPPm8>

The form will be closed on **Wednesday May 17th at 11:59 PM.**