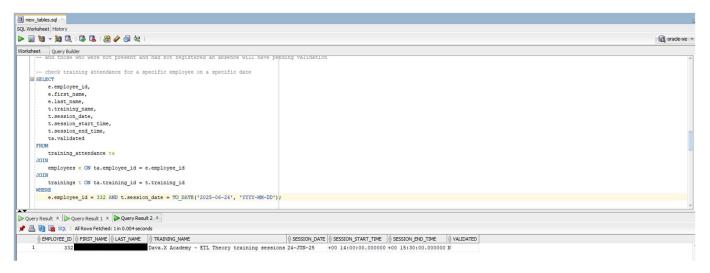# ETL Homework

Team X-tra Chill Cluj:

- Cosmin Gabriel Petriuc

- Adina Denisa Horj

- Daria Alexia Irimie

- Darius Fratila

- Ferdinand Alexandru Cabat

- Octavian Paul Ilies

- Tudor Gabriel Pascaru

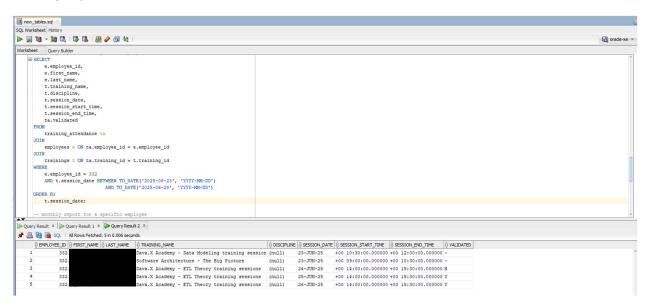- Daniel Kewin Chifa

- Raul Gabriel Chiorean

Steps:

1. Filled the employees table from the attendance lists at the three ETL training sessions (used the fill_employees.py script).
2. Filled the training sessions table from our calendar (used the fill_trainings.py script).
3. Filled the absences table from our [location](could be expanded to include other locations, but some of them had different formats for the table, which could not be exported). We had to subtract 3 hours from the times on Confluence since we believe that a different timezone was used (used the subtract_hours.py and fill_absences.py scripts).
4. Filled the training_attendance table as follows: it had as number of entries the number of employees multiplied by the number of training sessions. For each employee, if they had an absence at the same time a training was scheduled, we invalidated their presence. Validating their presence was possible only for the three ETL training sessions, if the employees were present in the meeting. For the other records, a placeholder char value '-' was chosen (used the fill_attendance.py script).

For running the script, create a database in Oracle's SQL Developer and make sure to modify the connection details in all Python scripts. Afterwards, create the tables in the new_tables.sql script, which also contains the queries highlighted below:

Employee on a specific date:


```sql
-- and those who were not present and had not registered an absence will have pending validation

-- check training attendance for a specific employee on a specific date
SELECT
    e.employee_id,
    e.first_name,
    e.last_name,
    t.training_name,
    t.session_date,
    t.session_start_time,
    t.session_end_time,
    ta.validated
FROM
    training_attendance ta
JOIN
    employees e ON ta.employee_id = e.employee_id
JOIN
    trainings t ON ta.training_id = t.training_id
WHERE
    e.employee_id = 332 AND t.session_date = TO_DATE('2025-06-24', 'YYYY-MM-DD');
```

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	TRAINING_NAME	SESSION_DATE	SESSION_START_TIME	SESSION_END_TIME	VALIDATED
332			Dava.X Academy - ETL Theory training sessions	24-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	N

Weekly report:


```sql
SELECT
    e.employee_id,
    e.first_name,
    e.last_name,
    t.training_name,
    t.discipline,
    t.session_date,
    t.session_start_time,
    t.session_end_time,
    ta.validated
FROM
    training_attendance ta
JOIN
    employees e ON ta.employee_id = e.employee_id
JOIN
    trainings t ON ta.training_id = t.training_id
WHERE
    e.employee_id = 332
    AND t.session_date BETWEEN TO_DATE('2025-06-23', 'YYYY-MM-DD')
                        AND TO_DATE('2025-06-29', 'YYYY-MM-DD')
ORDER BY
    t.session_date;

-- monthly report for a specific employee
```

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	TRAINING_NAME	DISCIPLINE	SESSION_DATE	SESSION_START_TIME	SESSION_END_TIME	VALIDATED
332			Dava.X Academy - Data Modeling training session	(null)	23-JUN-25	+00 10:30:00.000000	+00 12:00:00.000000	-
332			Software Architecture - The Big Picture	(null)	23-JUN-25	+00 09:00:00.000000	+00 10:30:00.000000	-
332			Dava.X Academy - ETL Theory training sessions	(null)	24-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	N
332			Dava.X Academy - ETL Theory training sessions	(null)	25-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	Y
332			Dava.X Academy - ETL Theory training sessions	(null)	26-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	Y

Monthly report:


```sql
-- monthly report for a specific employee
SELECT
    e.employee_id,
    e.first_name,
    e.last_name,
    t.training_name,
    t.discipline,
    t.session_date,
    t.session_start_time,
    t.session_end_time,
    ta.validated
FROM
    training_attendance ta
JOIN
    employees e ON ta.employee_id = e.employee_id
JOIN
    trainings t ON ta.training_id = t.training_id
WHERE
    e.employee_id = 332
    AND t.session_date BETWEEN TO_DATE('2025-06-01', 'YYYY-MM-DD')
                          AND TO_DATE('2025-06-30', 'YYYY-MM-DD')
ORDER BY
    t.session_date;
```

All Rows Fetched: 21 in 0.003 seconds

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	TRAINING_NAME	DISCIPLINE	SESSION_DATE	SESSION_START_TIME	SESSION_END_TIME	VALIDATED
8	332			Dava.X Academy - PL/SQL training session	(null)	10-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
9	332			Dava.X Academy - PL/SQL training session	(null)	11-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
10	332			Dava.X Academy - PL/SQL training session	(null)	12-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
11	332			Dava.X Academy - PL/SQL training session	(null)	13-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
12	332			Dava.X Academy - PL/SQL training session	(null)	17-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
13	332			Dava.X.Academy. Q&A Seesion	(null)	18-JUN-25	+00 16:00:00.000000	+00 17:00:00.000000	-
14	332			Dava.X Academy - PL/SQL training session	(null)	18-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
15	332			Dava.X Academy - PL/SQL training session	(null)	19-JUN-25	+00 09:00:00.000000	+00 10:00:00.000000	-
16	332			Dava.X Academy - Data Modeling training session	(null)	23-JUN-25	+00 10:30:00.000000	+00 12:00:00.000000	-
17	332			Software Architecture - The Big Picture	(null)	23-JUN-25	+00 09:00:00.000000	+00 10:30:00.000000	-
18	332			Dava.X Academy - ETL Theory training sessions	(null)	24-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	N
19	332			Dava.X Academy - ETL Theory training sessions	(null)	25-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	Y
20	332			Dava.X Academy - ETL Theory training sessions	(null)	26-JUN-25	+00 14:00:00.000000	+00 15:30:00.000000	Y
21	332			Git - Command Line	(null)	30-JUN-25	+00 10:30:00.000000	+00 11:30:00.000000	-