Pentru a implementa ETL pentru tema propusă am creat 3 useri în baza de date:

* etl\_src\_user
  + Conține două tabele care vor avea datele preluate din fișierele csv într-o formă denormalizată, cu valori *null*.

A screenshot of a computer

AI-generated content may be incorrect.

* + Valorile din surse sunt prelucrate în python și inserate ulterior în tabele sub forma prezentată în .csv-uri.
  + Fișierele python care se ocupă cu inserarea datelor sunt in folderul Extract din cadrul proiectului.
* etl\_stg\_user
  + Conține 5 tabele care conțin datele normalizate în urma procesului de transformare din tabelele sursă.

A screenshot of a computer screen

AI-generated content may be incorrect.

* + Tabelele DIM\_STG\_ABSENCES si DIM\_STG\_ABSENCE\_CODES provin din tabela sursă DIM\_RAW\_ABSENCES
  + Tabelele DIM\_STG\_EMPLOYEES\_ON\_TRAINING si DIM\_STG\_TRAINING\_SESSION provin din tabela sursă DIM\_RAW\_TRAINING\_ATTENDANCE
  + Tabela DIM\_STG\_EMPLOYEE este populată pe baza datelor din tabela sursă DIM\_RAW\_ABSENCES de unde am preluat numele persoanelor din DU-urile Suceava și Brașov.
* etl\_tgt\_user
  + A screenshot of a computer screen

    AI-generated content may be incorrect.Conține 4 tabele de fapte și 5 tabele dimensiune.
  + Tabela FACT\_TGT\_EMPLOYEE și tabela DIM\_TGT\_EMPLOYEE\_HISTORY formează un SCD2 pentru a menține istoricul angajaților în situațiile de inserare / actualizare a datelor.
  + Tabela FACT\_TGT\_EMPLOYEES\_ON\_TRAINING va conține datele din userul anterior ( tabela DIM\_STG\_EMPLOYEES\_ON\_TRAINING ), la fel si DIM\_TGT\_TRAINING\_SESSION
  + Tabela DIM\_TGT\_ABSENCES va conține înregistrări precum tipurile de absențe ce se regăsesc în DIM\_STG\_DU\_ABSENCE\_CODES dar și alte valori precum R – Regular si OVT – overtime pentru a putea gestiona timpul în FACT\_TGT\_TIMESHEET.
  + Tabela DIM\_TGT\_PROJECT este folosită pentru a putea menține un flow corect al înregistrărilor din FACT\_TGT\_TIMESHEET
  + Tabela DIM\_TGT\_DATE are ca înregistrări toate zilele de la data de 1 mai pana la 31 august pentru a oferi funcționalități de analiză temporală mult mai bogate decât ar fi doar cu un câmp dateID numeric sau DATE în tabela de fapt. În procesul de încărcare cred că nu ar trebui să se axeze pe logică calendaristică (de exemplu calculul dacă o zi e weekend sau nu); pur și simplu o încărcare o singură dată în tabela de date este mult mai eficient.
  + Tabela FACT\_TGT\_TIMESHEET este tabelul care măsoară timpul alocat de un anumit angajat pentru proiectul la care este alocat. Practic, aici se află tot ce înseamă ore raportate în organizație pe baza legăturilor către dimensiuni.

Tot procesul ETL se bazează pe utilizarea de cod python care *extrage* datele din fișierele .csv sub forma lor brută, le *transformă* într-un mod normalizat astfel încât să nu existe valori nule, după care *încarcă* toate aceste date prelucrate în tabelele de dimensiune și de fapte.

Pentru a putea executa fișierele python din proiect trebuie executate comanzile***: pip install reportlab, pip install matplotlib, pip install "setuptools<81"*** pentru a putea funcționa toate fișierele în momentul execuției.

Folderele **etl** și **duAbsences** conțin fișierele .csv prelucrate puțin pentru un flow bun al temei față de fișierele „default”.

Folderul **\_\_\_Extract** conține codurile care încarcă datele din fișierele .csv în tabelele sursă.

Folderul **\_\_Transform** conține codurile care normalizează datele extrase și le inserează în tabelele corespunzătoare.

Folderul **\_Load** conține procesul final prin care fișierele incarcă datele în tabelele de dimensiune și de fapte.

În folderul **rapoarte** am generat deja câteva rapoarte .pdf bazându-mă pe informațiile din baza de date. Acesta conține atât codurile de execuție în python pentru a genera rapoartele, dar și rezultatele generate. Întrebările care au fost puse în vederea acestor rapoarte sunt:  
Question 1: Cine a lipsit de la trainingurile ETL și din ce motiv?

Answer 1: Vezi fișierele

* Who missed the Dava.X Academy - ETL Theory training sessions 1 and why
* Who missed the Dava.X Academy - ETL Theory training sessions 2 and why
* Who missed the Dava.X Academy - ETL Theory training sessions 3 and why

Question 2: Ce detalii pot avea despre unul din trainingurile ETL?

Answer 2: Vezi fișierele

* Details about Dava.X Academy - ETL Theory training sessions 1
* Details about Dava.X Academy - ETL Theory training sessions 2
* Details about Dava.X Academy - ETL Theory training sessions 3

Question 3: Câte persoane s-au învoit până acum din academia DAVA și care au fost motivele?

Answer 3: Vezi fișierul Raport\_Grafic\_Numar\_Absente.pdf

Question 4: Ce a făcut angajatul X în perioada 1 iunie 2025 – 31 august 2025 ?

Answer 4: M-am bazat pe existența datelor din tabela FACT\_TGT\_TIMESHEET, DIM\_TGT\_ABSENCES, DIM\_TGT\_DATE și DIM\_TGT\_TRAINING\_SESSION prin join-uri diverse astfel încât să afișez rezultatele din fișierul **Employee\_Activity\_June\_2025**

**SCD2 în SQLDeveloper**

Am creat un trigger care va insera automat în DIM\_TGT\_EMPLOYEE\_HISTORY și va modifica datele din același tabel astfel încât istoricul să fie conform cu realitatea.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.