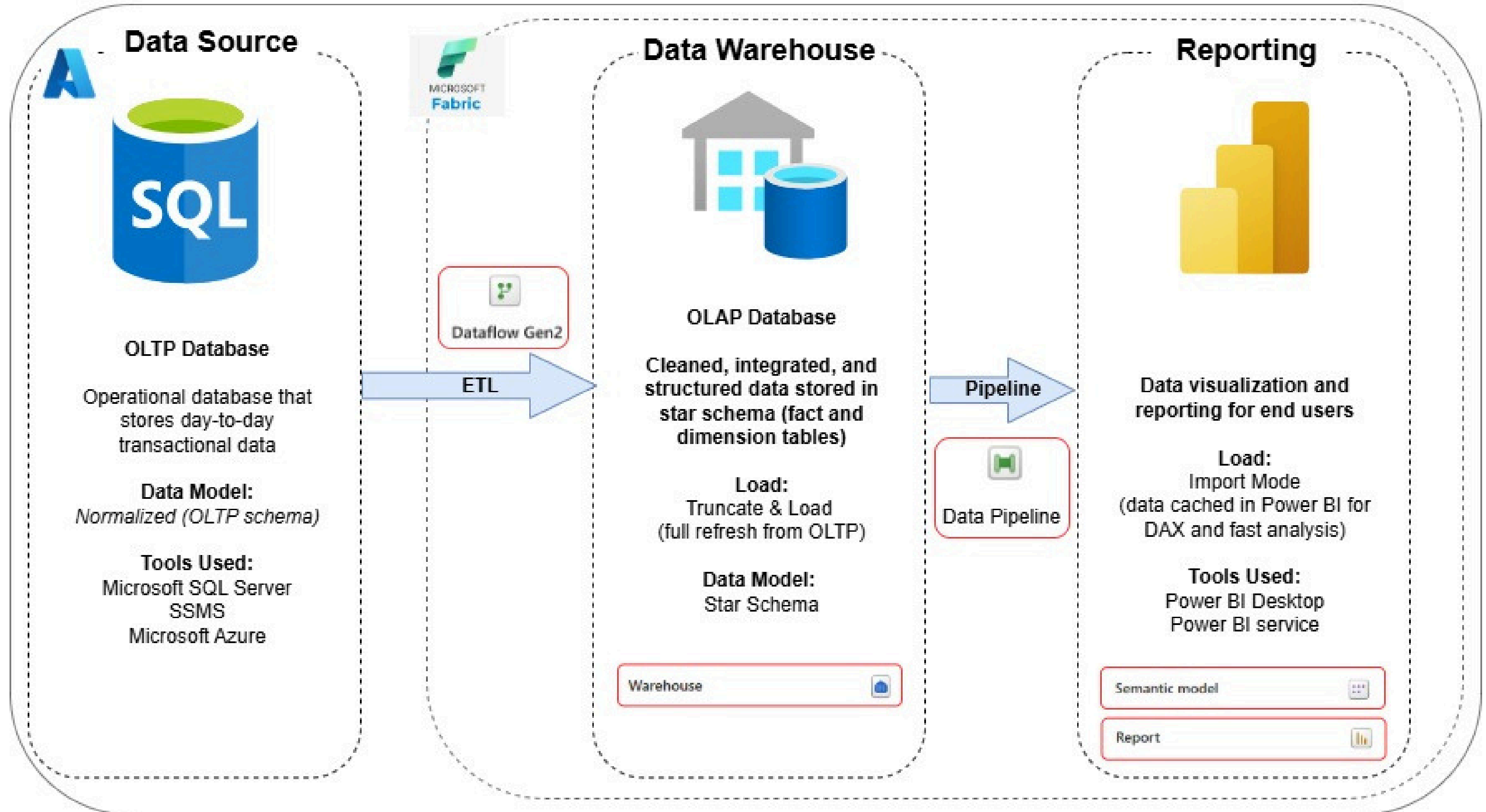


High Level Architecture





Microsoft Fabric

06

Seemlessly, the data is connected to Microsoft Fabric. This integration allows you to use Fabric for data transformation, modelling, reporting, and advanced insights.

Fabric

ITI_Lakehouse

ITI_Lakehouse

Dataflow_ETL

ITI_DWH

Search

Trial activated: 29 days left

Home

Workspaces

Copilot

OneLake catalog

Monitor

Real-Time

Workloads

ITI_Examination_system

My workspace

ITI_Examination_system

Create deployment pipeline

Create app

Manage access

Workspaces

Filter by keyword

Filter

List view

Name	Status	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
Dataflow_ETL		Dataflow Gen2 (C...	—	Mohamed Fayez ...	12/1/2025, 9:22:28 PM	—	—	—	
Final Project		Report	—	ITI_Examination_s...	12/11/2025, 1:47:34 P...	—	—	—	No
Final Project		Semantic model	—	ITI_Examination_s...	12/11/2025, 1:47:34 ...	N/A	—	—	
ITI_DWH		Warehouse	—	Mohamed Fayez ...	—	—	—	—	
ITI_Lakehouse		Lakehouse	—	Mohamed Fayez ...	—	—	—	—	
ITI_Lakehouse		SQL analytics en...	—	Mohamed Fayez ...	—	—	—	—	
ITI_Pipeline		Pipeline	—	Mohamed Fayez ...	—	—	—	—	
Linkedin Dataflow		Dataflow Gen2 (C...	—	Mohamed Fayez ...	11/24/2025, 8:43:28 P...	—	—	—	
Notebook_ITI		Notebook	—	Mohamed Fayez ...	—	—	—	—	

Dataflow Gen2

Prep, clean, and transform data.



Warehouse

Provide strategic insights from multiple sources into your entire business.



Lakehouse

Store big data for cleaning, querying, reporting, and sharing.



Notebook

Explore, analyze, and visualize data and build ML models. Supports Apache Spark, Python, T-SQL, and more.



Pipeline

Ingest data at scale and schedule data workflows.



ITI_Examination_system

+ New item New folder Import Micro

Create deployment pipeline Create app Manage access Workspaces List view

Filter by keyword Filter

Name	Status	Type	Task	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Included in app
Dataflow_ETL		Dataflow Gen2 (C...	—	Mohamed Fayez ...	12/1/2025, 9:22:28 PM	—	—	—	
Final Project		Report	—	ITI_Examination_s...	12/11/2025, 1:47:34 P...	—	—	—	<input type="checkbox"/> No
Final Project		Semantic model	—	ITI_Examination_s...	12/11/2025, 1:47:34 ...	N/A	—	—	
ITI_DWH		Warehouse	—	Mohamed Fayez ...	—	—	—	—	
ITI_Lakehouse		Lakehouse	—	Mohamed Fayez ...	—	—	—	—	
ITI_Lakehouse		SQL analytics en...	—	Mohamed Fayez ...	—	—	—	—	
ITI_Pipeline		Pipeline	—	Mohamed Fayez ...	—	—	—	—	
Linkedin Dataflow		Dataflow Gen2 (C...	—	Mohamed Fayez ...	11/24/2025, 8:43:28 P...	—	—	—	
Notebook_ITI		Notebook	—	Mohamed Fayez ...	—	—	—	—	

Completed (2.49 s) Columns: 3 Rows: 10 Column profiling based on top 1,000 rows [Add default destination...](#) ✓ Last run on 12/1/25 at 9:22 PM Step

Fabric

Dataflow_ETL

Search

Trials activated: 7 days left

Power Query

Home

Transform

Add column

View

Help

Queries [11]

DimBranch

DimCourse

DimDepartment

DimExam

DimInstructor

DimIntake

DimStudent

FactStudent_Grades

DimTrack

DimDate

Dimlinkedprofiles

Table.Select

123 Bran_ID

10 distinct, 10 u...

10 distinct

1	Cairo
2	Alexandria
3	Giza
4	Mansoura
5	Damanhur
6	Tanta
7	Asyut
8	Ismailia
9	Port Said
10	Minya

Completed (2.49 s)

Columns: 3

Rows: 10

Column profiling

SQL Server database

Server * ⓘ

iti-exam-sqlsrv.database.windows.net

Database

ITI_Examination_System

> Advanced options

OK

Cancel

> Data destination

X

Warehouse

Step

Query settings

> Properties

Name

DimBranch

> Applied steps

Source

Navigation 1

X Choose col...

> Data destination

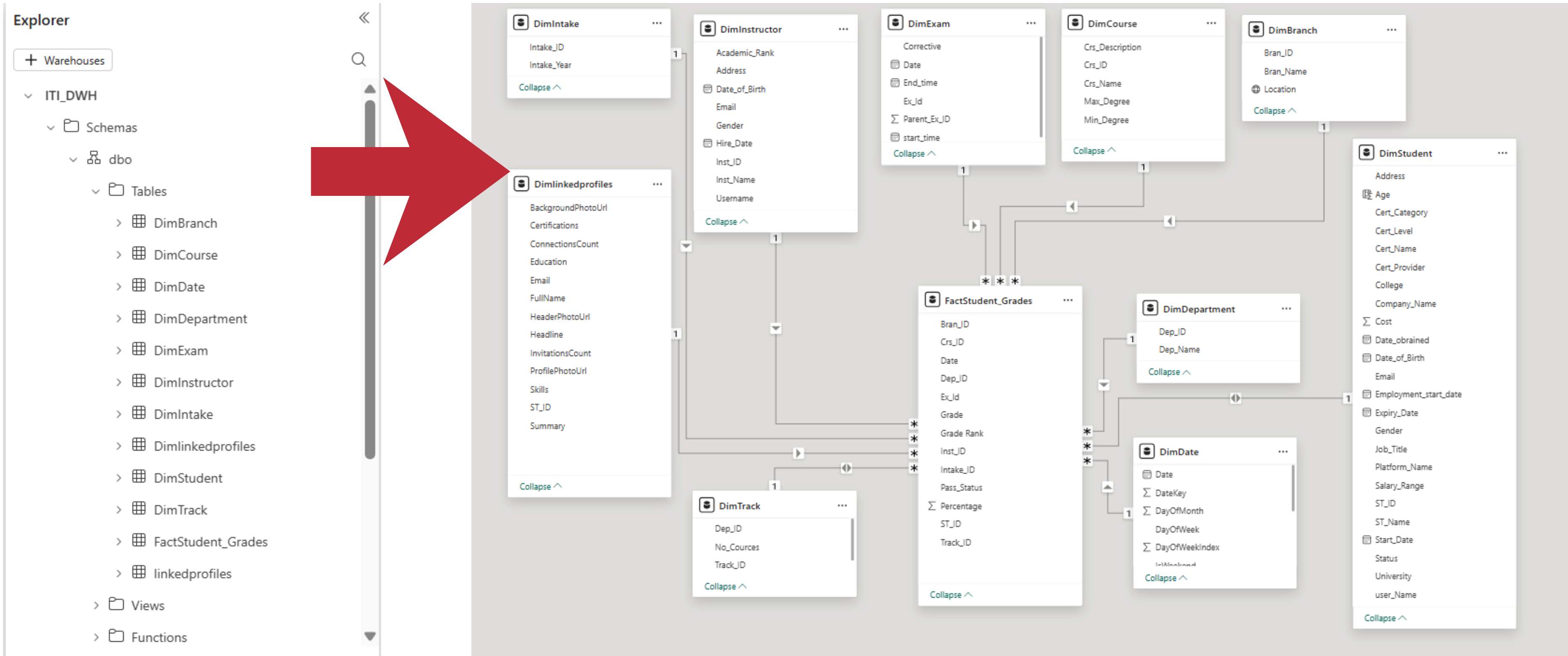
X Warehouse

Step



Data Modeling

Convert tables in the data warehouse (OLAP) into Fact and dimensions tables (Star Schema). It will help to improve performance and be clear to understand.



...

Fabric

ITI_Lakehouse

Notebook_ITI

Saved

Search

Trial activated: 7 days left

...

...

...

...

...

...

Home

Edit

AI tools

Run

View

Run all

Connect

PySpark (Python)

Environment

Workspace default

Data Wrangler

Copilot

Explorer

Data items

Resources

Add data items

Items

ITI_Lakehouse

Tables

linkedprofiles

Files

profiles

Ahmed

Aya

Manar

Mennatullah

Mohamed Elb...

Files

Ahmed

Search files

Certifications.csv

Connections.csv

Education.csv

Email Addresses.csv

Invitations.csv

Learning.csv

Profile.csv

Rich_Media.csv

Skills.csv

1

2

PART 1: Setup & Test Reading a CSV File

3

4

5

from pyspark.sql import SparkSession

6

from pyspark.sql import functions as F

7

from pyspark.sql import types as T

8

9

spark = SparkSession.builder.getOrCreate()

10

11

print("Spark session started")

12

[26]

✓

- Command executed in 267 ms by Mohamed Faye Abd Ellatif Elbasuoney on 2:35:21 PM, 11/24/25

PySpark (Python)

...

Spark session started

1

=====

2

PART 2 - Student folders + Safe CSV loader + Record template

3

=====

4

5

from pyspark.sql.utils import AnalysisException

6

7

8

1. List of student folders (based on your Lakehouse)

9

10

students = [

11

"Ahmed",

12

"Aya",

13

"Manar",

14

"Mennatullah",

15

"Mohamed Elbasuoney"

16

]

17

18

print("Students detected:", students)

Not connected

AutoSave: On

Selected Cell 1 of 5 cells

Data Flow & Pipeline

