Bootcathon 2024

Get an immersive learning experience on technology
and solve real energy business challenges

Analytics Track

Safety brief

Bootcathon Recap

Speakers

Suttipong (Best) Anuchitolan
System Architect
(Chemical Snowflake)
Corporate IT / EMIT

Narat (France) Maraprasertsak
BI Developer
(FVC Profitability)
SCCOT IT / EMIT

Doldej (Jet) Schuett
System Architect
(Financial Data Products)
Corporate IT / EMIT

How Success in adopting in D&A look like?

Exxon Mobil

What you will learn/do today?

Action Needed

Raise your voice and Share your thought as much as you can!!!

Exercise 1
What do you see?

What are insights from the dashboard?

What are insights from the dashboard?

- The quantity and totalprice seem to be positively correlated. An example wound be the line chart for TOTALPRICE and QUANTITY both show similar pattern.
- The TOTALPRICE contribution mostly come from order with more LINEITEM. Therefore, number of lineitem in each order seem to related with TOTALAMOUNT.

What did you learn from this exercise?

Exercise 2
Code-GPT

Instruction

Background

Mr.A has run his business for a while. He currently does not know on how well his company has been running and he also unable to answer who/where is his key customers. Therefore, Mr.A hire you as an analyst who will help him finding the answer and develop the product that change the data into insight. Currently, there is a new AI tool so called "Code-GPT". The tool can help you manage Mr.A company data by just telling it on what you want to do with the tables that contain Mr.A company data.

Objective: Provide insight that can help Mr.A to understand how well his company has been run.

Instruction:

1. Please review the provided data & Diagram
2. Develop the product by
    1. Starting from table 1
    2. Tell what you want Code-GPT to do
    3. See the result on the screen
    4. Move to next table
    5. Goto 1, Stop when Last table finish and all tables agree that the result acceptable

Instruction

Basic Data Command

1. SELECT <Column> From <Table>
2. Join Operation (LEFT/RIGHT/FULL OUTER)
3. Aggregated Function (Group By)
    1. SUM
    2. MIN
    3. MAX
    4. AVG
    5. COUNT
4. Date Manipulation Function
    1. MONTH
    2. YEAR
    3. DATE
5. String Manipulation Function
    1. SUBSTRING
    2. LEFT
    3. RIGHT

Basic Viz Generate

1. Display result in
    1. Line
    2. Dot
    3. Bar

Or "Try Speak out the Command and the GPT will try"

What did you learn from this exercise?

Let’s Guess

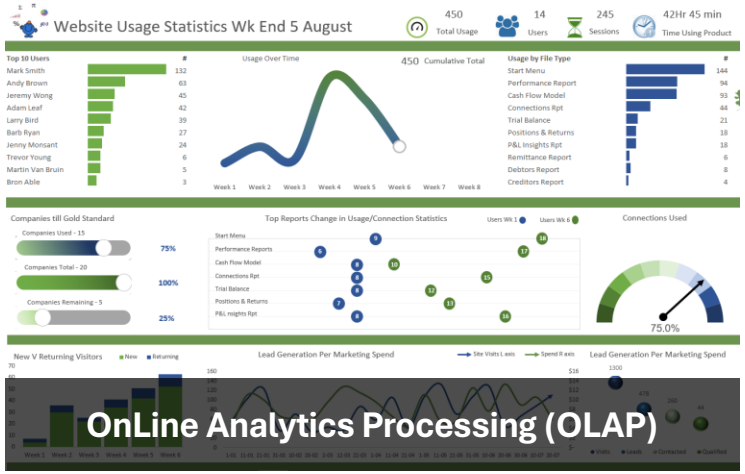
- How many transactions in 7-11 Retail System per day in Total?

What are the challenges of having very huge datasets?

Security
Storage
Integration
Processing
Speed

Big Data Challenges: OLTP vs OLAP

OnLine Transaction Processing (OLTP)



OnLine Analytics Processing (OLAP)

Desired properties

- Prioritize insert/update speed
- Operations performed on single- to few rows at a time
- Requires strong data consistency and integrity

Typical solutions/architecture

- Row storage
- Database indexes, constraints, and ACIDity

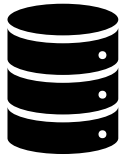
- Prioritize read/select speed
- Operations performed very large number of rows at a time
- Data consistency and integrity is assumed at ingestion time

- Columnar storage
- Distributed processing

Big Data Challenges: Data Warehouse and Data Lake

- Data lake is an *unstructured* data repository to centralize the storage of an enterprise's data from multiple source systems
- Because the data is unstructured and heterogenous, it is not easily directly usable

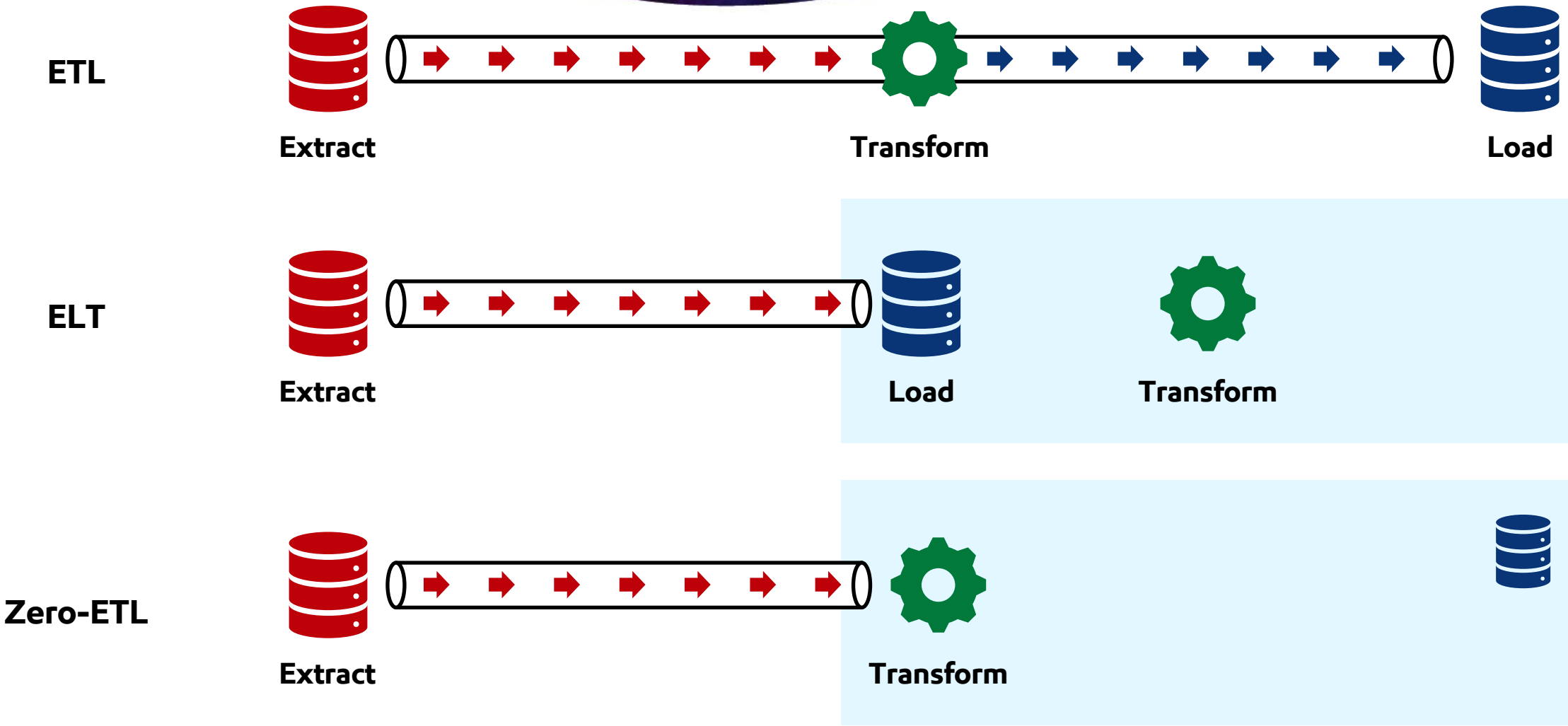
If data is not directly usable, why is a data lake useful?



- Data warehouse is a structured repository of mostly homogenous data
- It is (usually) optimized for processing large amounts of data
- Data is readily usable for further analysis

Why not ingest data directly into Data warehouse?

Big Data Challenges: ELT, ELT, and Zero-ETL



Importance of Speed to Market: The Case of Minecraft vs Lego Universe

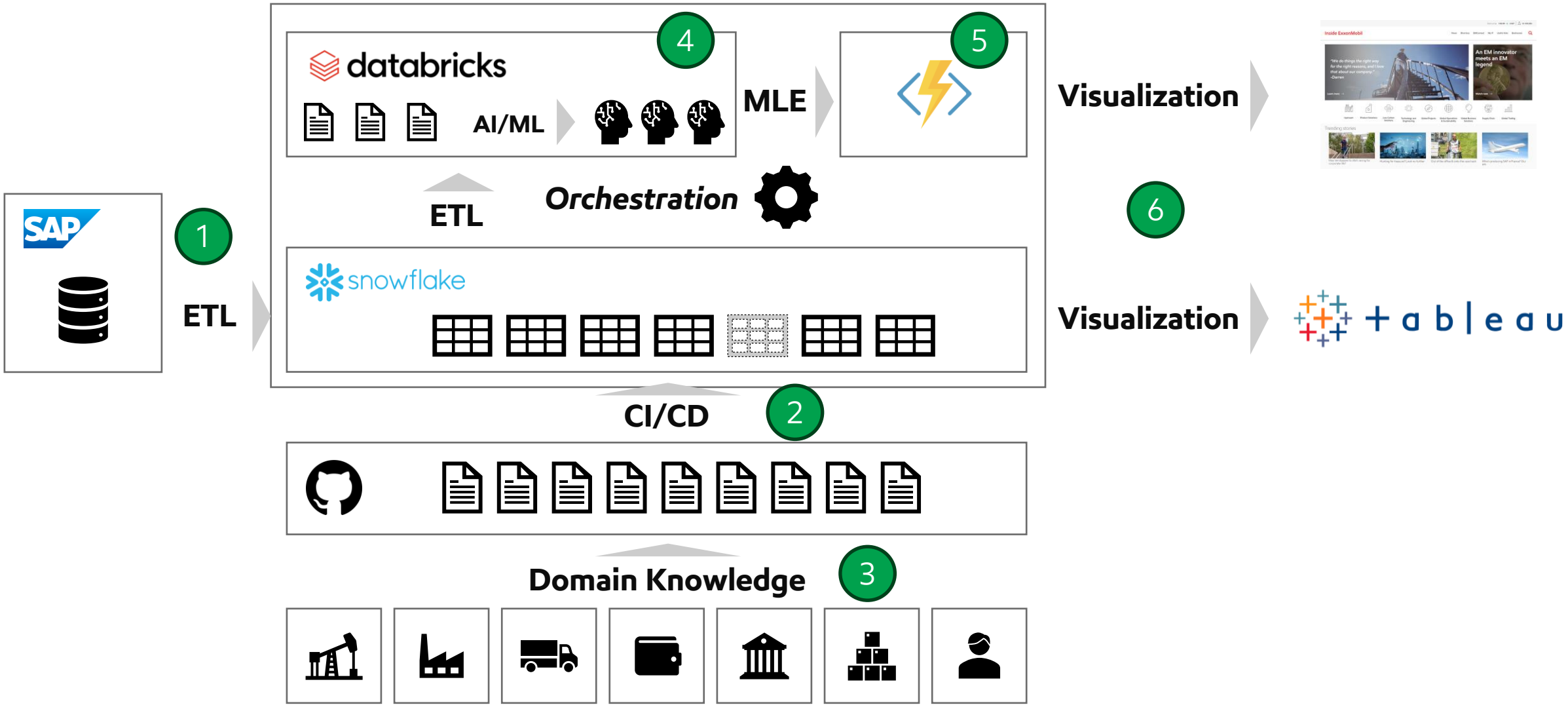
>100m monthly users | $300m+ annual revenue | acquired for $2.5B by Microsoft

Game was unprofitable and shutdown within 2 years

Analytics End-To-End

Role and Responsibility

Data "actors" need to know each others' scripts to produce an analytic solution



ExxonMobil

Lab 0

Prerequisite for Lab (Analytics)

- OS: Windows 8/10/11+
- IDE:
  - Visual Studio Code Extension (Preferred)
    - Python
    - Jupyter
  - Jupyter Notebook (Optional)
    pip3 install pandas
- Viz:
  - Power BI Desktop
- Download “package.zip” File from link below
  - t.ly/mTJ-C
- Running Setup.bat
- Download “Lab2_Starter.pbix” and “Lab2 Deck.pptx” from the link below
  - t.ly/qfj3r
  - t.ly/EIRIE
- Explore files in data folder

.bat completed installation screen