≫ Extract, transform, Load (ETL)

← فرایندی >> data integration , data warehousing

extract data ——→ transforming it ——→ Loading it into
from various sources   into various format   a target database
                                                     or data warehouse

1. Extract

retriving from →P | DBs
data | flat files   (P.csv)
                    | web services (API)
                    | etc...

2. Transform

تغییرات { cleaning   تبدیلات ⇒ { data cleansing
          formating   normalization
          converting   aggregation   (تجمیع)
                                    deduplication (حذف تکراری)
                                    applying business rules
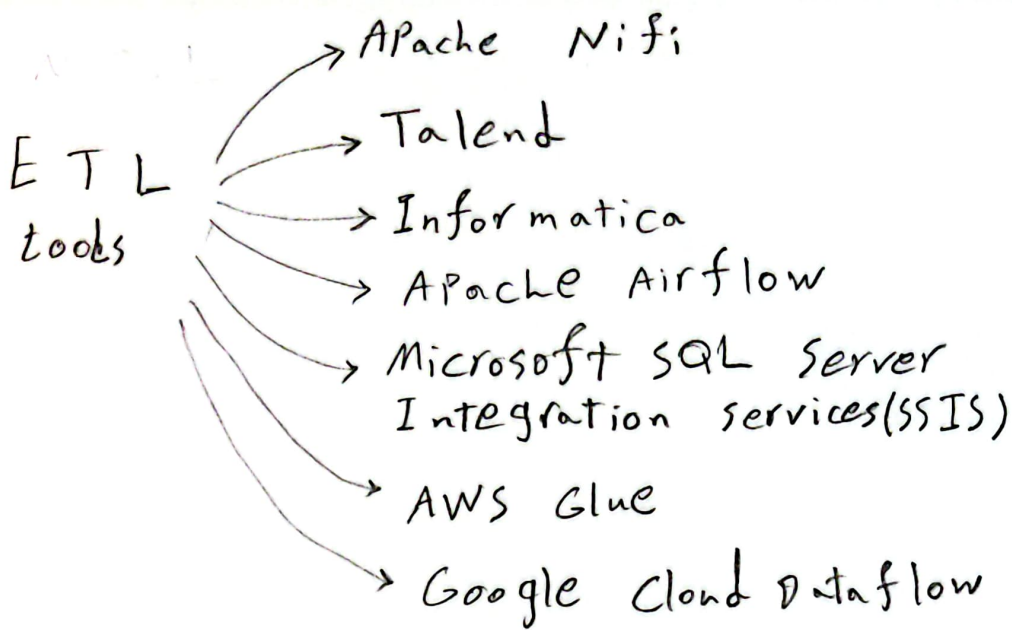
P | data cleaning —¹→ Handling missing values
  | Correcting errors
  |
  | Aggregation —²→ Summarizing data, calculating averages

3. Load

——→ inserting the transformed data into target { DB or warehouse

                                          →P { Amazon RedshA
                                              { Google BigQuery

ETL tools
- Apache Nifi
- Talend
- Informatica
- Apache Airflow
- Microsoft SQL Server Integration services (SSIS)
- AWS Glue
- Google Cloud Dataflow

>> Reporting vs Business Intelligence (BI) vs Analytics

* Reporting

→ { Collecting
    formatting   data in a structured manner
    presenting }

ویژگی‌ها { Periodic
          static
          snapshot of what has happened
          Focus on Past }

مثل { monthly sales reports
      annual financial statements. }

ابزارها
- Microsoft Excel
- Crystal Reports
- SQL server Reporting services (SSRS)
- Google Data studio

* Business Intelligence (BI)

اهداف
← ایجاد بینش برای business operations
← تصمیم گیری استراتژیک و تاکتیکی
← ارائه‌ی دید جامع از عملکرد کسب و کار

ویژگی های BI
- interactive & Dynamic
- Real-time or near-real-time data processing
- Focuses on both historical & current data
- Provides dashboards, score carts & ad-hoc querying capabilities.

Tools
- Tablean
- Power BI
- Qlik View
- Looker

* Analytics

اهداف
- پیش بینی trend های آینده
- تشخیص الگو و همبستگی در داده ها
- بهینه کردن فرایند های کسب و کار
- فراهم کردن پیشنهادات و آینده نگری

ویژگی ها
- تکنیک های پیشرفته پردازش داده (آنالیزهای آماری، یادگیری ماشین و...)
- تمرکز بر گذشته، حال و آینده داده ها
- فراهم کردن بینش، پیش بینی و توصیه ها
- شامل آنالیزهای اکتشافی داده ها و الگوریتم های پیچیده

ابزارها
- Python
- R
- SAS
- A Pache Spark