
# Data Warehouse & Business Intelligence Fundamentals

Todor Kichukov

todor.kichukov@bipartner.biz

https://www.facebook.com/groups/SUDWBI2022/

Faculty of Mathematics and Informatics

Sofia University

2022

# Data Warehouse & Business Intelligence Fundamentals

## Course Scope

- DW Concept
- DW Architecture
- DW Data Modeling
- Data Integration
- Gathering and Analyzing Requirements
- Business Intelligence
- Deployment, Support and Maintenance

# Data Warehouse
# Data Integration

- Data Integration Definition
- Levels of Data Integration
- Raw Data vs Business Rules
- ETL (Extract, Transform, Load) Process
- Data Materialization vs Data Virtualization
- Data Lineage
- Terminology
- Project Work

# Data Integration

- Data Integration is one of the defining characteristics of the Data Warehouse (*which were the others?*).
- **Data Integration** means gathering data from multiple sources and combining (merging) that data in a structured and meaningful way to form a whole.

# Levels of Data Integration

- **Co-locating data** - Placing individual copies of data, in source system table constructs, from various sources into a **common physical location** (physical database). One set of table constructs for each source.
- **Format Integration** - Placing individual copies of data, **converted to common table constructs**, from various sources into a common physical location. One set of table constructs for each source.
- **Structural Integration** - Loading data from various sources into **common shared tables** within a common physical location. One set of table constructs contains all data from all sources. Can have redundant records from multiple sources.
- **Soft Data Integration** - Loading data from various sources into common shared tables within a common physical location and **integrating instances of data** from the various sources **on business key level only**. One set of table constructs contains all data from all sources. Single instances of each record, keeping all data from all sources as received.
- **Full Data Integration** -Loading data from various sources into common shared tables within a common physical location and **integrating instances of data** from the various sources **for every single entity/attribute**. One set of table constructs contains all data from all sources. Single instances of each record, no redundancy.

# Raw Data *vs* Business Rules

- All Staging Areas have Raw Data – no Business Rules applied on data – it supports the quick offloading of the source systems.
- All Data Marts have Business Rules applied on data – it supports analyzing data in a standardized way.
- The Data Warehouse itself may/may not have Business Rules applied on data – it depends on the architecture, data model, the business requirements for raw data availability, etc.

# ETL (Extract, Transform, Load) Process

- Ensures data transportation through all the DW architecture – from the source systems to the data marts.
- Could be ETL or ELT, depending on the point of view and the tools used.
- Rule of thumb – use as much as possible the database capabilities to transform data.
- Usually the most “expensive” back-room process as a DW activity.
- Usually the longest task in a DW project.

# Data Materialization *vs* Data Virtualization

- Physical copy of data
- Needs time/resources for ETL
- Need time for implementation
- Data is transformed to the needed model before to use it
- Most/All calculations should be done in advance
- Better performance

- Logical data representation
- No ETL, reads data from wherever it is
- Quicker solution
- Data is transformed in the moment of usage (on-the-fly)
- All calculations take place in the moment of usage
- Worse performance

# Data Lineage

- Provides two-way information about:
  - How the data is calculated, which source data it uses
  - Which data is affected by specific source data change
- The Data Lineage tool should integrate well with the ETL tool and the BI tool.

# Terminology

- Data Integration
- ETL
- Data Materialization, Data Virtualization
- Data Lineage

# Project Work

- Today
  - Staging Area logical model
  - Justification of selected DWH building approach
  - Clarify all open questions
- Next steps
  - DWH model (draft)