

**WEEK 5**

# **INFORMATION LIFECYCLE MANAGEMENT**

# Introduction

## What is Information Lifecycle Management?

Information Lifecycle Management is a comprehensive approach to managing data and information throughout its entire existence - from creation to final disposal. It's a strategic methodology that ensures information is properly handled, stored, protected, and disposed of according to its value and regulatory requirements.

# S OF THE ON-LIFE CYCLE

ITION

STO

USE

IVAL

DIS

# STAGES OF THE INFORMATION LIFECYCLE

tially generated from various sources such as systems. This can include structured data like emails, images, and documents. A student writes a research paper or a company creates a presentation.

# STAGES OF THE INFORMATION LIFECYCLE

The data moves into the Storage stage, where it is stored. Storage may be physical, such as paper files or electronic databases, servers, or cloud systems. Proper storage ensures the data is organized, and available when needed.

For example, consider the research paper on a computer hard drive, which is stored in digital form.

# STAGES OF THE EDUCATION LIFECYCLE

Stage 3: Use, where data is actively accessed, processed, evaluated, communicated, and used for operations and decision-making. At this stage, relevant information is collected and communicated through reporting, analysis, or communication.

Stage 4: Feedback, where data is used to evaluate and improve operations and decision-making. This stage involves a teacher reading the paper to grade it, or employees providing feedback to management.

# STAGES OF THE INFORMATION LIFECYCLE

If it is no longer needed for daily operations, it either goes to long-term storage, where it remains available but does not burden active systems. Archives are solutions optimized for long-term retention.

At the end of the semester, the research paper is stored in the company files for auditing purposes or the company files the report for auditing purposes.

# STAGES OF THE INFORMATION LIFECYCLE

When it reaches the Disposal stage, where it is securely destroyed as legally required. Secure disposal prevents unauthorized access to documents. This may include shredding physical documents or physically destroying storage devices.

Old drafts from the computer, shredding out old digital files no longer needed.

# DATA RETENTION CHALLENGES

## Regulatory Requirements

- Healthcare: HIPAA requires patient records to be retained for a defined period.
- Finance: SEC mandates retention of trading records.

Pharmaceutical Industry (FDA, U.S.): Clinical trial records must be retained for at least 2 years after a drug is approved.

# MANAGING DATA THROUGH ITS LIFECYCLE

## Document Management Systems (DMS)

Organize, track, and manage documents in digital form.

Provide version control for easy retrieval of historical data.

Reduce reliance on paper records.

## Cloud Storage Lifecycle Management

- Automates the migration of files to cloud storage.
- Ensures integrity of data remains a priority by lowering costs.
- Offers scalability and flexibility for organizations.

# MOVING DATA THROUGH ITS LIFECYCLE

## Document Management Systems (DMS)

uses a DMS to store student's grades, assignments, attendance records digitally instead of keeping piles of papers.

Attorneys in a law firm can search and retrieve cases without digging through filing cabinets.

Cloud Storage with Lifecycle Management Rules

### EXAMPLE:

- A company stores files in Google Drive. If a file has not opened in 90 days, it moves to a cheaper storage but can still be accessed if needed.
- A photography studio stores client photos in cloud storage, but the old photos are stored in local storage to save costs.

# MANAGING DATA THROUGH ITS LIFECYCLE

## Protection and Access

Protect sensitive information from unauthorized access.

Encryption ensures data is unreadable without the correct key.

Access controls (like passwords, roles, and multi-factor authentication) allow authorized staff to view data.

## Backup and Disaster Recovery Systems

- Create regular backups of critical data
- Ensure quick recovery after system failure or natural disaster
- Provide business continuity to reduce downtime

# MANAGING DATA THROUGH ITS LIFECYCLE

## Storage and Access

• encrypts customer credit card numbers, so even if someone steals the database, the information looks like random symbols without the decryption key.

• In school, only teachers with a password and 2-step verification can access students' grades.

## Backup and Disaster Recovery System

### EXAMPLE:

- A hospital computer system stores patient medical records. If a cyberattack occurs and the data is destroyed, the backup system restores the files and protects the patients.
- A company stores its business files on office computers. If a fire destroys the building, the backup system recovers all data and allows the company to continue operating.

# CHALLENGES IN INFORMATION

## Complexity

and variety of data are increasing rapidly. Multiple data sources and formats increase management difficulties.

## EXAMPLE:

A student has photos, videos, school assignments saved on their phone, laptop, and Google Drive. Since the files are in many places and formats, it's hard to keep everything organized.

## Cost

Managing the increased data requires significant storage infrastructure, data management software, and professional services.

The student uses Google Drive since it's free, but they are paying for extra storage space each month. That's a significant cost for managing their information.

# CHALLENGES IN INFORMATION

## Concerns

constantly under cyberattacks such as malware, phishing, and ransomware attempts, which can cause severe financial and reputational damage.

If someone hacks a system and leaks student records, the school loses trust.

## Compliance Risk

Failing to comply with legal requirements can lead to reputational damage and loss of customers.

Example: If a company doesn't keep proper records properly, it may face big legal problems.

**THANK  
YOU**



ent

Contact

AQUAL

M A N A G E M E N T

Start

ent

## Contact

### Quality Management (DQM)

Management is the process of making sure data is accurate, complete, consistent, and up-to-date. It is used effectively. It focuses on making data reliable by setting rules, checking data over time.

DQM is used to ensure customer account details (such as address and phone number) are accurate and up to date.

ent

Contact

## IMENSIONS OF DATA QUALITY

Data should correctly represent reality

No missing or incomplete values

Data should be uniform across

Data must be up-to-date and

ent

Contact

# ality Frameworks and St

Organization for Standardization  
standard for data quality.

data is accurate, complete, a

ent

Contact

# ability Frameworks and St

Management Association Interna  
ent Body of Knowledge.

organizations set governance  
ure secure and consistent data

ent

Contact

# ing and Measuring Data

- checks – Ensure inputs meet rules
- ing – Analyzing database
- es.      Audits – Independent re

# Improving Data Quality

ools:

- Removing errors, duplicates

cleans student records to merge duplicates

on - Converting data into a uniform standardizes date formats (MM/DD/YY)

ent

Contact

# Improving Data Quality

ools:

- Detecting and removing duplicates - Identifying and fixing errors in product data

nt - Adding external data to improve customer profiles with updated contact information

ent

Contact

# Governance Strategies

**Policies** – Rules for data handling and access. The company has strict policies on user data access and usage.

**Monitoring** – Continuous checks on transaction data daily to detect anomalies or suspicious activity.

ent

Contact

# Governance Strategies

**Structures** – Assigning responsibilities

Government agencies, specific offices are responsible for maintaining databases.

**Awareness** – Educating staff on the importance of accuracy

Training staff to input passenger data correctly

ent

Contact

ANK YOU

3



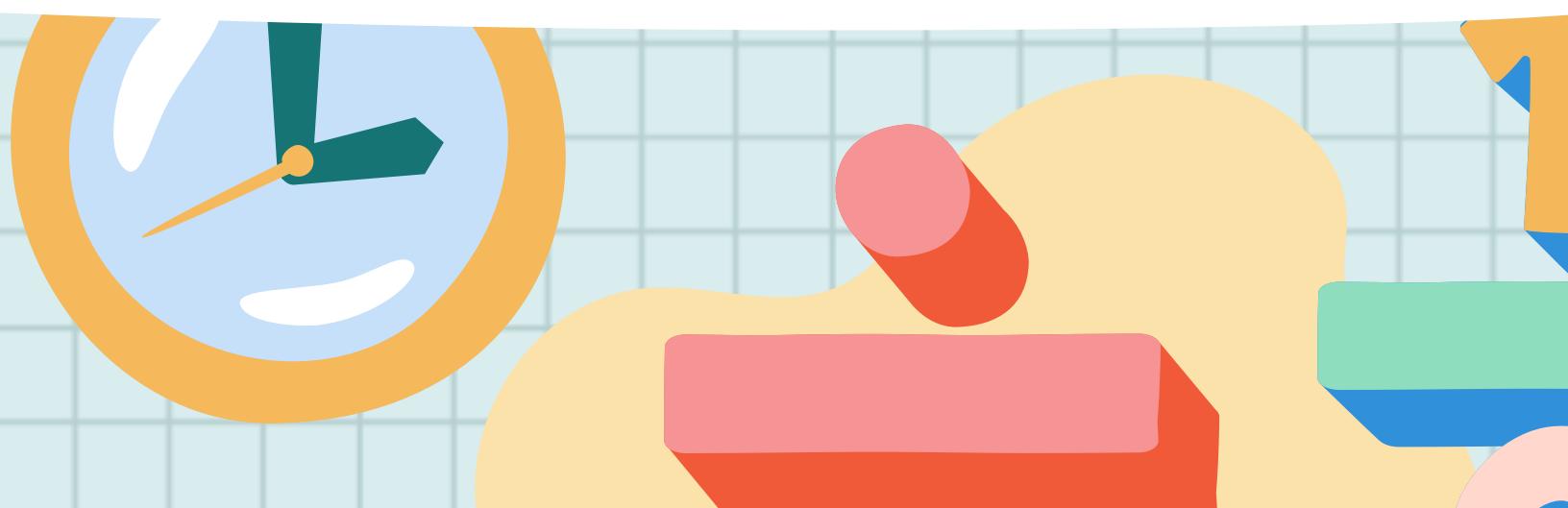
MASTER DATA MANAGEMENT

&

REFERENCE DATA MANAGEMENT

GROUP 7

---



# WHAT IS DATA

---

## PROBLEM

DATA CAN BE DUPLICATED,  
INCONSISTENT, OR  
INCOMPLETE.

# **CENTER DATA MANAGEMENT (CDM)**

*THE PRACTICE OF CREATING ONE TRUSTED VERSION OF THE*

*NAME APPEARS DIFFERENTLY IN EACH SYSTEM (E.G., JOHN PAUL CRUZ*

*IS TRANSFORMED INTO A SINGLY, CORRECT VERSION OF THAT STUDENT RECORD ACROSS ALL SYSTEMS)*

# **IMPORTANCE OF MDM**

**REDUCES DUPLICATION.**

**IMPROVES REPORTING AND ANALYTICS.**

**INCREASES EFFICIENCY.**

# **ENCE DATA MANAGEMENT**

*ISTS, CODES, AND CATEGORIES TO ENSURE ALL SYSTEM*

*INTO PHP (PHILIPPINE PESO).*

# **ROLE OF RDM IN INTEGRATION**

*WHEN SYSTEMS EXCHANGE DATA, THEY ARE ALIGNED*

**ATION ACROSS SYSTEMS**

*ES, CATEGORIES, AND VALUES MEAN THE SAME THING*

*MALE, "F" = FEMALE ACROSS HR, CRM, AND FIN*

# **ROLE OF RDM IN INTEGRATION**

**VALUES FROM ENTERING SYSTEMS DURING INTEGRATION.**

**W VENDOR HAS CURRENCY = "PHL," RDM REJECTS IT SINCE IT ISN'T IN THE LIST.**

## **INTEROPERABILITY**

**SYSTEMS CAN TALK TO EACH OTHER SMOOTHLY BECAUSE THEY USE A COMMON LANGUAGE.**

# **ROLE OF RDM IN REPORTING**

## **E AGGREGATION**

*DATA FROM MULTIPLE SOURCES WITHOUT DUPLICATION*  
"COUNTRIES," "USA," AND "UNITED STATES" ARE ALL REPO

## **ED DECISION-MAKING**

*RELIABLY ON TRUSTED, STANDARDIZED DATA → NOT CONF*

# ROLE OF RDM IN REPORTING

## CONSISTENCY

CATEGORIES CHANGE (LIKE PRODUCT LINES), RDM MAINTAINS CONSISTENT.

1

LINES

PHONES, CELL PHONES

3



# ARCHITECTURE & PROCESSES

---



# COMPONENTS OF MDM ARCHITECTURE & FLOW

ES

NAL SYSTEMS WHERE DATA IS CREATED AND STORED

(ENTERPRISE RESOURCE PLANNING), CRM (CUSTOMER RELATIONSHIP MANAGEMENT)

HR, FINANCE, SPREADSHEETS.

# COMPONENTS OF MDM ARCHITECTURE & FUNCTIONAL LAYER

## EDITION LAYER

AND PROCESSES (LIKE ETL - EXTRACT, TRANSFORM, LOAD) THAT MOVE DATA FROM SOURCE SYSTEMS TO THE MDM HUB.

IMPORTANT: DIFFERENT SYSTEMS STORE DATA IN DIFFERENT FORMATS.

LAYER CONVERTS THEM INTO A COMMON FORMAT.

WE HAVE A TRANSLATOR THAT MAKES SURE ALL "LANGUAGES" CAN BE UNDERSTOOD IN ONE PLACE.

CRM CAN BE UNDERSTOOD IN ONE PLACE.

# COMPONENTS OF MDM ARCHITECTURE & FUNCTIONALITIES

## DATA HUB

A CENTRAL DATABASE (THE "BRAIN") THAT STORES

OPEN RECORDS.

IMPORTANT: IT'S THE SINGLE SOURCE OF TRUTH FOR

E THE REGISTRAR'S OFFICE IN A SCHOOL THAT

—NO DUPLICATES, NO ERRORS.

**COMPONENTS OF MDM ARCHITECTURE & FUNCTIONALITY**

## **DATA QUALITY TOOLS**

*INTEGRATED TOOLS THAT CLEAN, VALIDATE, AND STANDARDIZE DATA*

*(AUTOMATICALLY FIX SPELLING MISTAKES, REMOVE DUPLICATES, ETC.)*

*(VS. "PHILIPPINES")*

*DATA CONSISTENCY: ENSURES THE GOLDEN RECORD IS ACCURATE*

# COMPONENTS OF MDM ARCHITECTURE & PARTNERS

S, APPLICATIONS, AND REPORTS THAT USE THE

BUSINESS APPLICATIONS (ERP, CRM), BI DAS

FINANCIAL REPORTS.

ANT: THEY RELY ON MDM TO GET CONSISTEN

KING.

# **MDM PROCESSES**

**INTEGRATION**

*MERGE AND INTEGRATE DATA FROM MULTIPLE SOURCE SYSTEMS.*

**VALIDATION**

*VERIFY AND VALIDATE RECORDS, RESOLVING CONFLICTS.*

*"JOHN CRUZ" + "J PAUL CRUZ" → JOHN PAUL CRUZ*

# **MDM PROCESSES**

## **E & VALIDATION**

**BUSINESS RULES AND CHECKING ACCURACY.**

**DATE OF BIRTH MUST BE VALID, EMAIL MUST FOR**

# **MDM PROCESSES**

## **INTEGRATION**

*INTEGRATION, GOLDEN RECORDS TO OTHER SYSTEMS.*

*CUSTOMER INFO SENT TO SALES, FINANCE, AND MARKETING*

*FROM A SINGLE SOURCE, THE SAME RECORD.*

## **DATA ENRICHMENT & MAINTENANCE**

*CHECKING, UPDATING, AND IMPROVING DATA QUALITY*

*ENSURING THAT YOUR DATA IN YOUR MDM DOESN'T BECOME OUTDATED OR INCONSISTENT*

# **DATA QUALITY TOOLS**

**R TOOLS:**

**DATA MDM**

**ATOS SPHERE**

**DG**

**MDM**

# **DATA QUALITY TOOLS**

**TOOLS:**

**SED MDM:**

**STING MDM IN THE CLOUD INSTEAD OF ON-P  
BLE, SCALABLE, COST-EFFICIENT.**

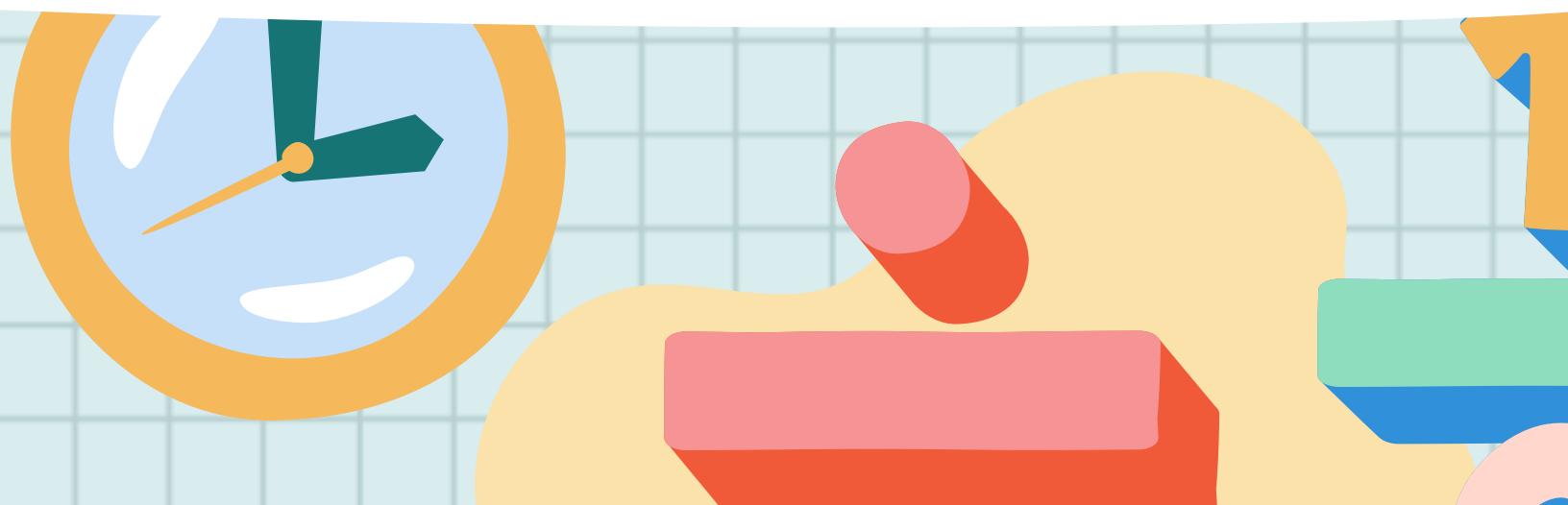
**AD OF BUYING SERVERS, COMPANIES CAN USE  
THEIR DATA GROWS.**

3



# **GOVERNANCE & STEWARDSHIP**

---



# **MDM GOVERNANCE & STEWARDSHIP**

## **GOVERNANCE**

*E POLICIES, RULES, AND DECISION-MAKING FRAMEWORKS IN MDM.*

## **STEWARDSHIP**

*E PEOPLE AND ROLES RESPONSIBLE FOR MAINTAINING COMPLIANCE.*

# ROLES IN GOVERNANCE & STEWARDSHIP

## OWNERS

DATA OWNERS (LIKE DEPARTMENT HEADS).

WHAT THE DATA MEANS AND WHO CAN ACCESS IT.

DATA DECIDES HOW "EMPLOYEE RECORDS" ARE STRUCTURED.

## STEWARDS

DAY-DAY CARETAKERS OF DATA.

ENSURE DATA ACCURACY, FIX ERRORS, AND MONITOR DATA QUALITY.

DATA STEWARD ENSURES VENDOR BANK ACCOUNTS

# **ROLES IN GOVERNANCE & STEWARDSHIP**

## **ADMINISTRATORS**

*SYSTEM RULES, PERMISSIONS, AND INTEGRITY*

*MAKE SURE HR ONLY GETS EMPLOYEE DATA,*

# **GOVERNANCE PROCESSES IN MDM**

**CIES**

**HOW DATA IS CREATED, STORED, SHARED, AND SECURED.**

**CUSTOMER RECORDS MUST INCLUDE EMAIL + PHONE NUMBER.**

**ROLE-BASED ACCESS CONTROL (RBAC) – ROLE-BASED ACCESS CONTROL**

**ROLE ONLY GETS THE DATA IT NEEDS.**

**AN ANALYST CAN SEE VENDOR BANK INFO NOT EMPLOYEE MEDICAL RECORDS.**

# **GOVERNANCE PROCESSES IN MDM**

## **LITY MONITORING**

*CHECKS FOR DUPLICATES, MISSING FIELDS, OR INACCURACIES. FOR EXAMPLE, IF A PRICE IS ENTERED AS CURRENCY, RDM + GOVERNANCE*

## **COMPLIANCE**

*ASSSED WHAT, WHEN, AND HOW.*

*IANCANCE WITH LAWS LIKE GDPR, HIPAA, OR LOCAL REGULATIONS.*

*/ GOVERNANCE & STEWARDSHIP MA*

CHAOS → WITHOUT RULES, EVERY SYSTEM

T AS "CUSTOMER".

ACTIVE DATA → PREVENTS FINANCE FROM S

RY RECORDS.

UI → PEOPLE ONLY TRUST DASHBOARDS A

E DATA IS ACCURATE AND CONSISTENT.

COMPLIANCE → AVOIDS LEGAL ISSUES (E.G.,

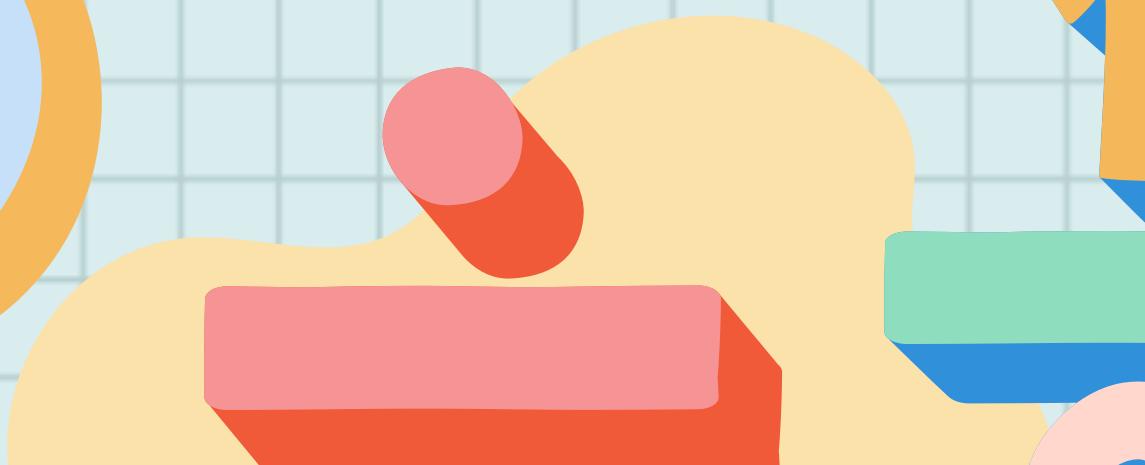
3



# THE END

anks for listening

---



TA INTEGRATI

&

TA PROVENANT

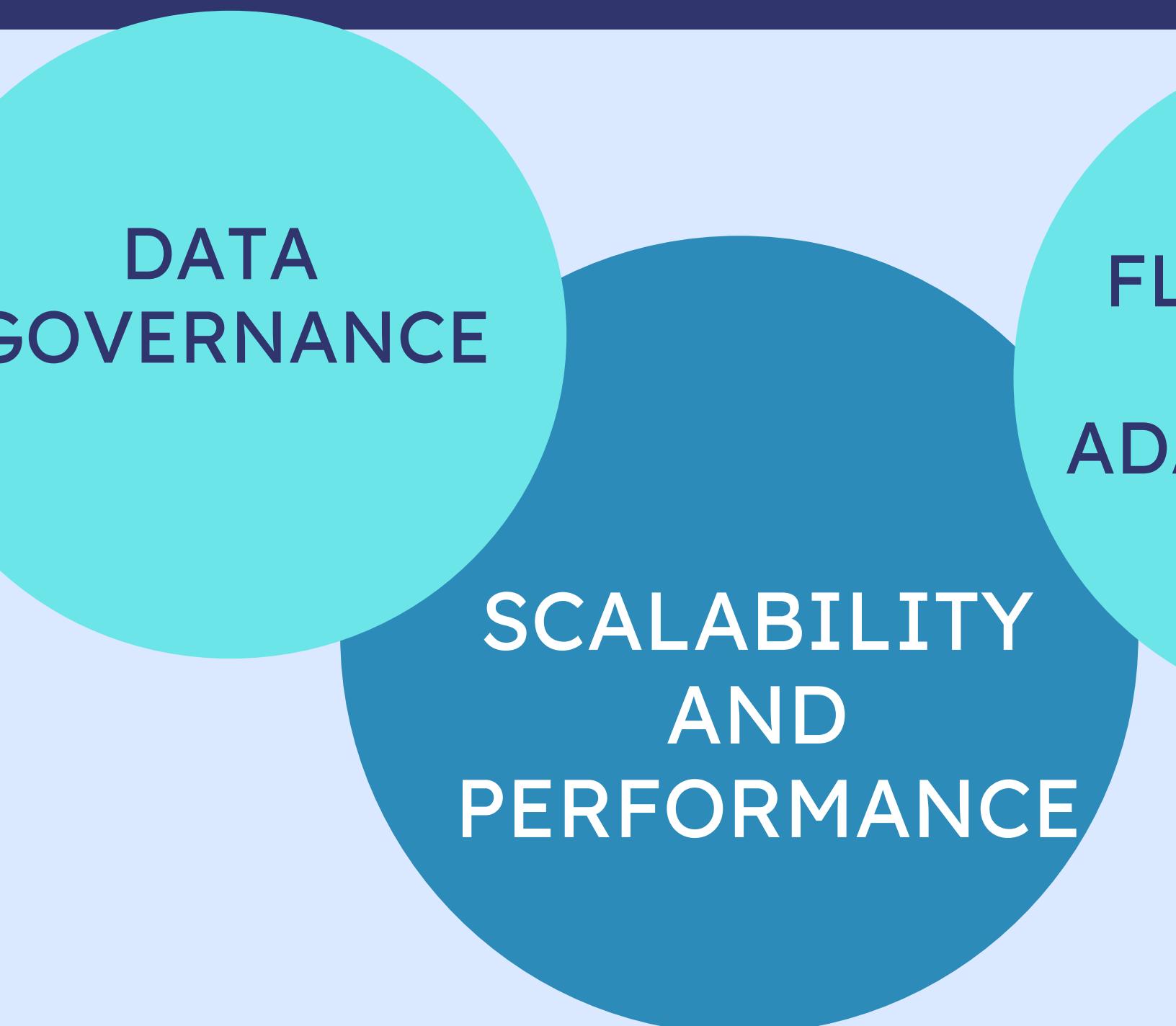
GROUP 8

# DATA INTEGRA

tion involves combining data from different sources for review for analysis, reporting, and other purposes

Source( Recipe , Procedure, Equipment, Lobo)

# EXAMPLES OF DATA INTEGRATION



# ALITY AND con

accuracy, completeness, and consistency across all sources and targets.

students Record : Name (John Smith : Name (Jam, Age : 90)

# DATA GOVERNANCE

U policies and procedures for managing data assets, including data privacy and compliance.

Data Privacy (G-cash)

# LITY AND PERFO

ntegration solutions the  
data volumes and proce

der Monday (100) - Order

# LITY AND ADAP

ems that can easily integ  
adapt to evolving busines

dent Record

# DATA manage

management refers to  
and using metadata to  
and quality of an organiz

# Methods of Data Integration (ETL, ETL, DATA PIPELINES)

# D OF DATA INTE

(Extract, Transform, Load):

Data is extracted from source systems.  
Data is cleansed, standardised  
and transformed into a suitable format  
before being loaded.

# D OF DATA INTE

transform, Load):

processed data is loaded

ically a data warehouse.

ral Different source:(Patient

remove duplicate name,

r, Snowflake

# D OF DATA INTE

Load, Transform):

data is extracted from sources, new data is loaded directly into a cloud data warehouse or

# D OF DATA INTE

oad, Transform):

nsformations are performed leveraging its processing

# D OF DATA INTE

a series of automated processes from its source to a destination. It uses various integration methods, depending on the specific require

# D OF DATA INTE

acilitate continuous data flow supporting various analytical and

for managing the entire deployment.

# D OF DATA INTE

## elines

ing Pipelines: Process large volumes of data at regular intervals, ideal for tasks like monthly reports(Daily, Monthly, Yearly) Real-time Pipelines: Handle real-time data, such as user interactions or social media posts (Like, Comment, Share) Real-



# Data Warehousing and Integration Challenges

# WAREHOUSING INTEGRATION CHALLENGE

ing is the process of collecting large amounts of data from different sources (called a data warehouse) for analysis.

# WAREHOUSING INTEGRATION CHALLENGES

Warehouses provide centralized storage, but they also present several challenges. Existing systems and data sources can be incompatible, leading to difficulties in managing structured, semi-structured, and unstructured data. Additionally, the physical infrastructure of warehouses, such as shelving and picking paths, can limit the efficiency of data collection and processing. Addressing these challenges requires a comprehensive approach that considers both technical and operational factors.

# WAREHOUSING INTEGRATION CHALLENGES

Ensuring accuracy, completeness, and consistency across multiple systems while handling large and growing volumes of data, managing real-time and batch data integration, and maintaining data quality.

Addressing these challenges requires advanced data management and quality assurance mechanisms.

# DATA PROVENCE

# DATA PROVENANCE

ance, also known as  
cking the history and f

# DATA PROVENANCE

include:

Data Flow: Identifying where data moves across systems.

Transformations: Recording what data is aggregated, or transformed.

Accountability: Providing the source, lineage, and troubleshooting.

# DATA PROVENANCE

Ensuring lineage, origin and context in their data for compliance.

# **THANK YOU**

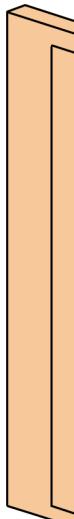
# **DATA WAREHOUSING ONLINE ANALYTIC PROCESSING (OLAP)**

# **HOUSING**

gathering, keeping, and combining multiple sources in order to produce useful business information.

## **ANALYTICAL PROCESSING**

System that uses the multidimensional model to enable users to analyze several database systems



# DATA WAREHOUSING ARCHITECTURE

that helps businesses make better decisions by combining data from various sources and organizing it under a single schema. It simplifies data management and storage, increasing the speed and efficiency of analysis.

a process in which data goes through  
mations, data cleaning procedures,  
on to be loaded into the data  
ty.

ge this process:

ew data from sources.

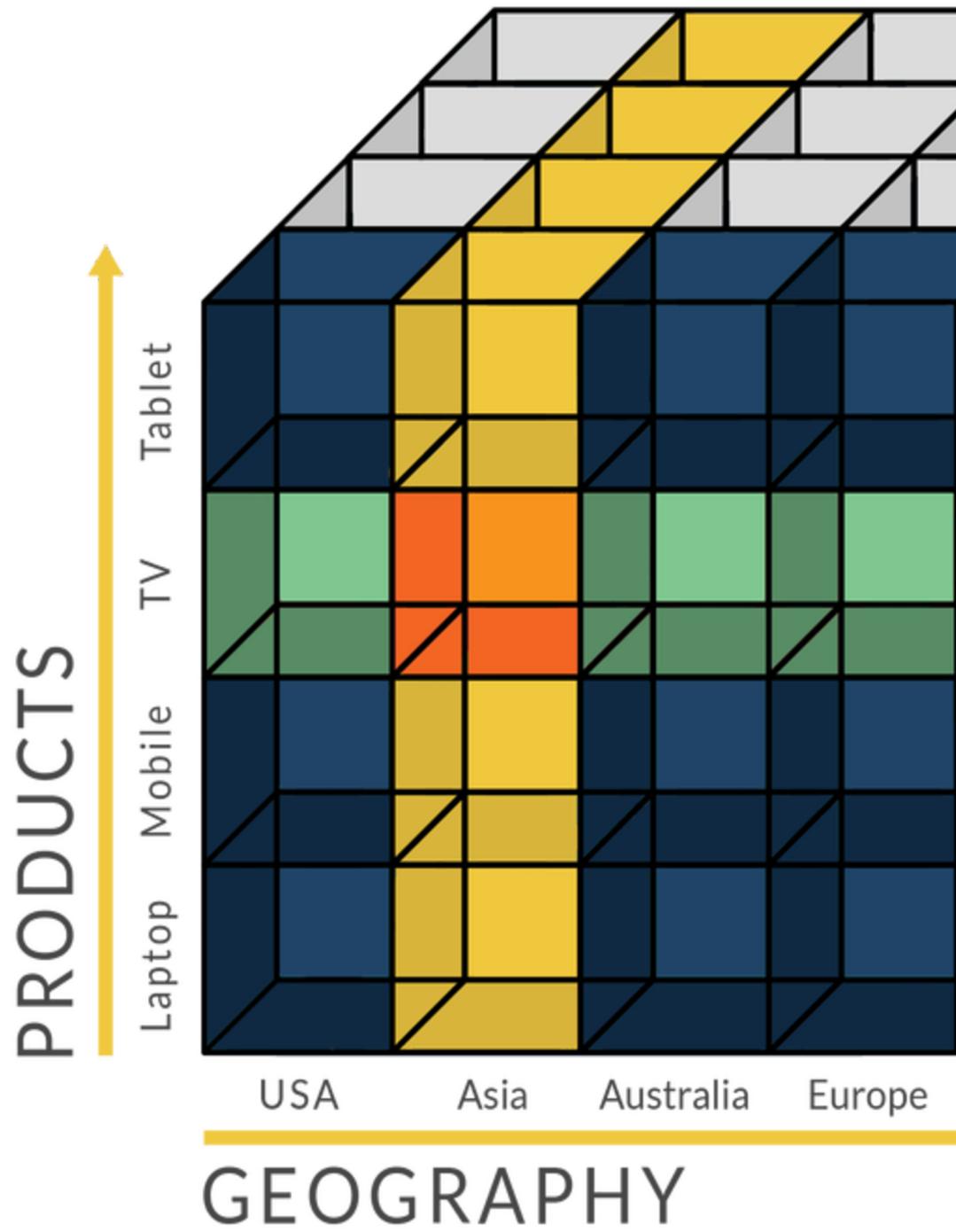
dardizes and formats the data.

e data into the data warehouse

**Data Marts** is the smaller, specialized data warehouses that serve giving them direct access to the data.

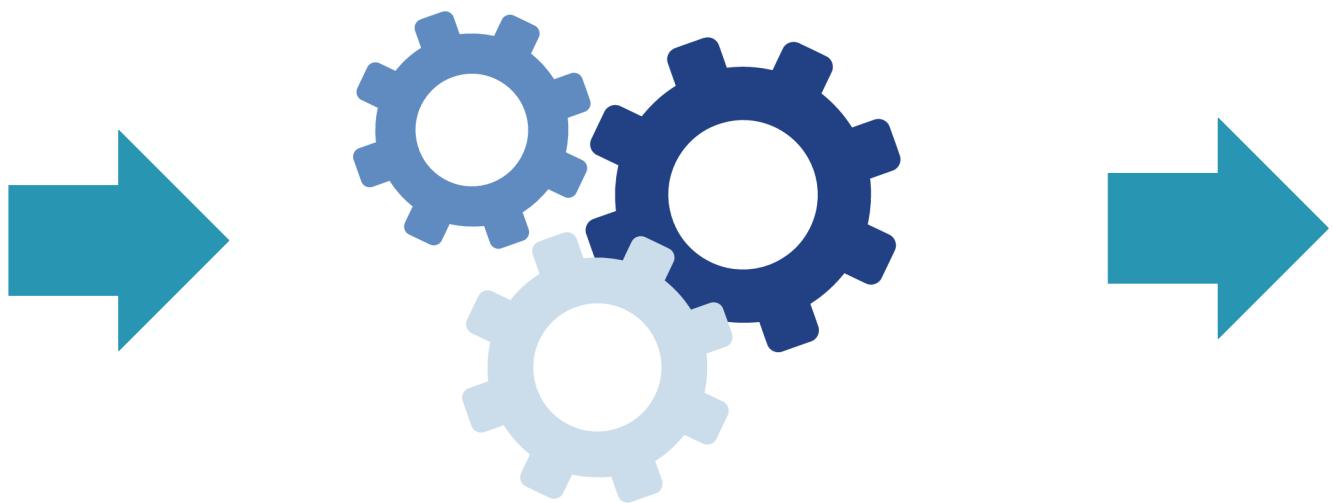
Ex. Sales, Marketing

An **OLAP cube** is a special way of organizing data into a multi-dimensional structure that makes it faster to analyze. Instead of looking at tables,



# TRACT, TRANSFORM, L

process of gathering data from different sources and moving it into a central storage system called a data warehouse. In this process, raw data is cleaned, structured, and transformed according to specific rules so that it's ready for storage, analysis, and reporting.



**Imports:** Data is collected from different databases, spreadsheets, CRM systems, log files, and other sources.

er raw data, no matter the format (structured or unstructured).

**Technologies/tools:** SQL, APIs, data connectors, Fivetran, Talend, Informatica, Alteryx, Looker, Tableau.

**cleans:** The extracted data is cleaned, and according to business rules. This includes removing duplicates, correcting errors, standardizing formats (if time), and sometimes enriching the data by adding new fields. This step is crucial to convert messy raw data into consistent and accurate data sets ready for analysis.

**Analytics/tools:** Apache Spark, AWS Glue, Talend.

**pens:** The transformed data is stored in a database (like Amazon Redshift, Google BigQuery or Microsoft SQL Server). This makes the data accessible for analytics and machine learning.

**Strategies/tools:** ETL pipelines, batch/stream processing engines (Airbyte, Informatica, Azure Data Factory).

# **OLAP VS. OLTP**

## **Online Analytical Processing (OLAP)**

Software tools used for the analysis of existing processes. It generally allow users from various perspectives.

## **Online Transaction Processing (OLTP)**

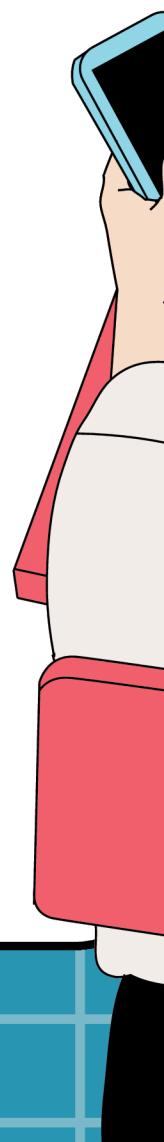
Processing approach emphasizing real-time. The majority of OLTP systems are made of short atomic operations that keep data consistent.

# OLAP VS. OLTP

	<b>OLAP (Online Analytical Processing)</b>	<b>OLTP (Online Transaction Processing)</b>
	Used for <b>analysis</b> and <b>decision-making</b>	Used for <b>transaction processing</b>
	Forecasting, business reporting, dashboards, market trend analysis	Banking, e-commerce, airlines
Tools	Microsoft SSAS, Oracle OLAP, SAP BW, Tableau	MySQL, PostgreSQL, Oracle Database

# **G A STAR SCHEMA FOR BUSINESS INTELLIGENCE**

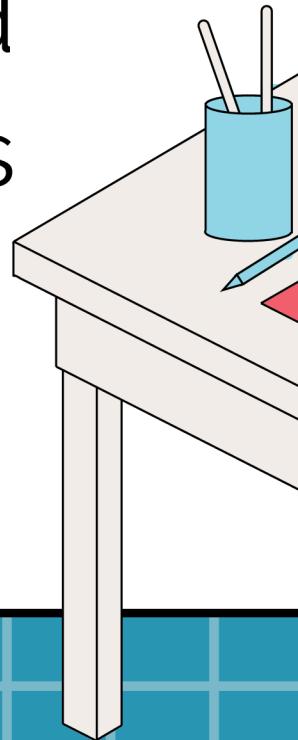
is a type of multidimensional structures information in a way simple to interpret and analyze. A star schema is important in Business Intelligence (BI) because it provides the ability to store, retrieve, monitor and organize data in a way that supports reporting and analysis fast and efficiently.

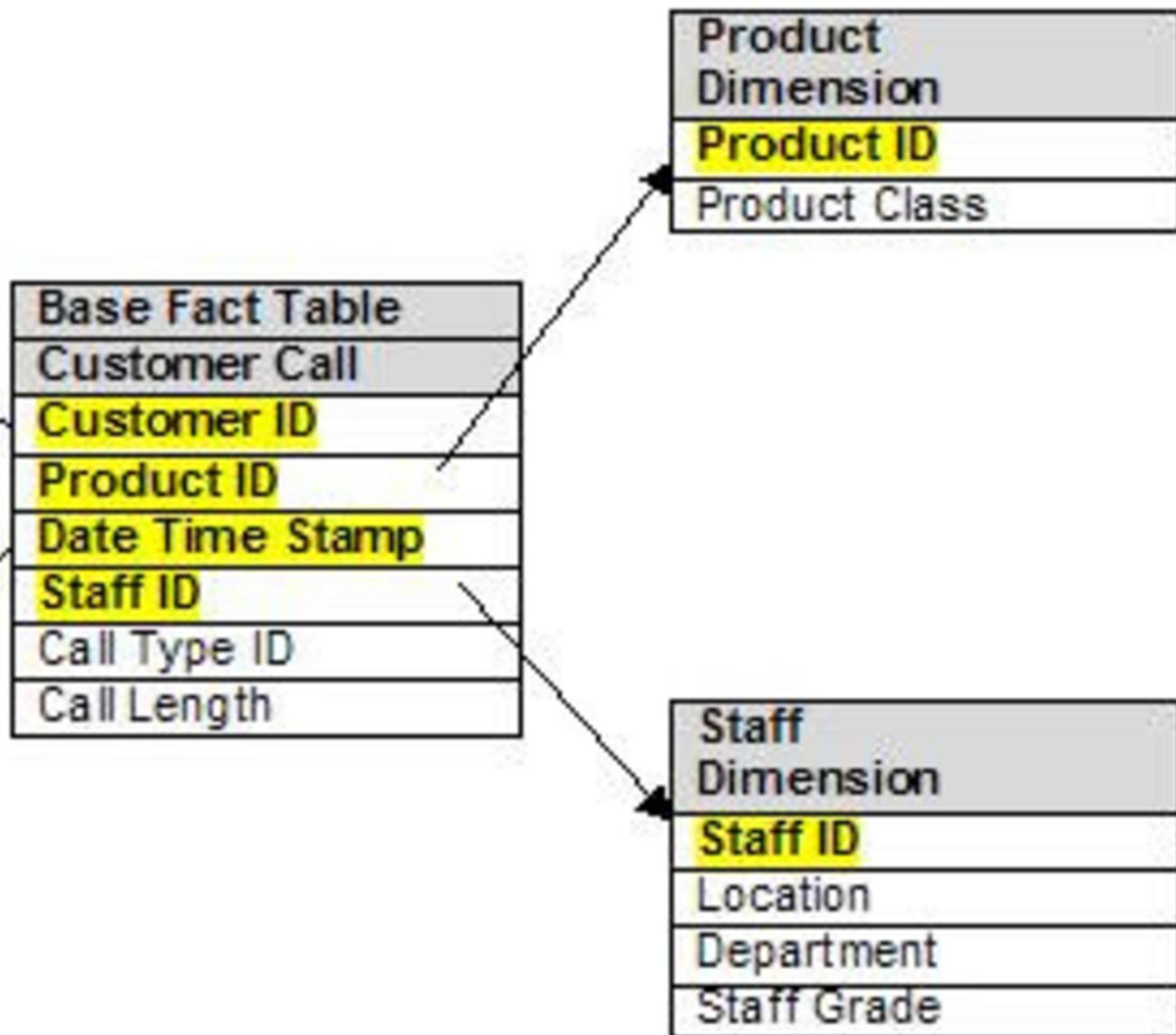


olds the key business metrics  
in amounts and quantities.

**Dimensions:** provide the context like  
product

is the center (measures) and  
**dimensions** and **tables** are the sides  
and they are connected  
by key-foreign key relationships  
making it meaningful.





**THANK YOU**

[point.com/overview-of-data-warehouse](https://www.geeksforgeeks.org/software-testing/what-is-a-data-warehouse/)  
[geeks.org/dbms/data-warehouse-and-dwml](https://www.geeks.org/dbms/data-warehouse-and-dwml/)  
[org/software-testing/what-is-a-data-warehouse](https://www.geeks.org/dbms/difference-between-star-and-snowflake-schemas/)  
[om/what-is/etl/](https://www.geeks.org/dbms/difference-between-star-and-snowflake-schemas/)  
[geeks.org/dbms/difference-between-star-and-snowflake-schemas/](https://www.geeks.org/dbms/difference-between-star-and-snowflake-schemas/)  
[s.com/glossary/star-schema](https://www.geeks.com/glossary/star-schema)