

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-LIFECYCLE

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.

ent writes a research paper or a company crea-

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 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 INFORMATION LIFECYCLE

s Use, where data is actively accessed, prepared, manipulated, communicated, and used for operations and decision-making. At this stage, relevant information is collected, organized, and communicated through reporting, analysis, or visualization.

teacher reads the paper to grade it, or employees make decisions based on the information. At this stage, relevant information is collected, organized, and communicated through reporting, analysis, or visualization.

STAGES OF THE INFORMATION LIFECYCLE

If it is no longer needed for daily operations, it either moves 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 long-term storage 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 when legally required. Secure disposal prevents unauthorized access to sensitive information. This may include shredding physical documents or physically destroying storage devices.

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

DATA RETENTION CHALLENGE

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

and reduce reliance on paper records.

Cloud Storage Lifecycle Management

- Automates moving files to cloud storage.
- Ensures integrity remains as it lowers costs.
- Offers scalability and organization.

MANAGING DATA THROUGH ITS LIFECYCLE

Document Management
Systems (DMS)

uses a DMS to store students' 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 been opened in over a year, it moves to a cheaper storage, but can still be accessed if needed.
- A photographer stores client photos in a central storage, but the old photos are stored in a lower-cost storage to save space.

MANAGING DATA THROUGH ITS LIFE CYCLE

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 specific 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 LIFE CYCLE

Encryption 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 company that stores patient medical records. If a cyberattack happens, they can restore the files and treat patients.
- A company stores business files on office computers. If a fire destroys the building, they can recover all data and continue operations.

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 different 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.

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 school's system and leaks student records, the school loses trust.

Compliance Risk

Failing to comply with regulations can lead to legal penalties, reputational damage, and loss of customer trust.

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

**THANK
YOU**



ent

Contact

PAQUAL

MANAGEMENT

Start

ent

Contact

Quality Management (DQM)

Management is the process of maintaining data that 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, phone number, and email) 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 all fields.

Data must be up-to-date and accurate.

ent

Contact

ality Frameworks and St

Organization for Standardization
standard for data quality.

data is accurate, complete, a

ent

Contact

ality 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

ent

Contact

Improving Data Quality

ools:

- Removing errors, duplicates
- Cleans student records to merge duplicates
- Converts data into a unified format
- Standardizes date formats (MM/DD/YY)

ent

Contact

Improving Data Quality

ools:

- Detecting and removing duplicates: Duplicate product listings can lead to inaccurate sales figures and confuse customers. Tools like DataPump help identify and remove these entries.
- Adding external data to enrich customer profiles: Integrating data from external sources such as social media or third-party databases can provide more comprehensive information about customers, helping businesses tailor their offerings.

ent

Contact

Governance Strategies

Policies - Rules for data handling and usage.

sets strict policies on user data access and retention.

Monitoring - Continuous checks on system activity and transaction data daily to detect anomalies.

ent

Contact

Governance Strategies

Structures – Assigning responsibilities

nt agencies, specific offices are managing databases.

Awareness – Educating staff on the importance of accuracy

instructing staff to input passenger data correctly

ent

Contact

ANK **Y**OU

3



MASTER DATA MANAGEMENT

&

REFERENCE DATA MANAGEMENT

GROUP 7



WHAT IS DATA

PROBLEM

DATA CAN BE DUPLICATED,
INCONSISTENT, OR
INCOMPLETE.

EDUCATIONAL DATA MANAGEMENT (EDM)

THE PRACTICE OF CREATING ONE TRUSTED VERSION OF THE

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

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

IMPORTANCE OF MDM

AVOIDS DUPLICATION.

IMPROVES REPORTING AND ANALYTICS.

INCREASES EFFICIENCY.

REFERENCE DATA MANAGEMENT

LISTS, CODES, AND CATEGORIES TO ENSURE ALL SYSTEMS ARE CONSISTENT.

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

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

ROLE OF RDM IN INTEGRATION

VALUES FROM ENTERING SYSTEMS DURING INTEGRATION.

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

INTEROPERABILITY

SIMILAR SYSTEMS CAN TALK TO EACH OTHER SMOOTHLY BECAUSE THEY USE THE SAME 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

INES

PHONES, CELL PHONES

3



ARCHITECTURE & PROCESSES



COMPONENTS OF MDM ARCHITECTURE & P

ES

NAL SYSTEMS WHERE DATA IS CREATED AND STOR

(ENTERPRISE RESOURCE PLANNING), CRM (CU

HR, FINANCE, SPREADSHEETS.

COMPONENTS OF MDM ARCHITECTURE & FUNCTIONALITY

EDITION LAYER

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

IMPORTANT: DIFFERENT SYSTEMS STORE DATA IN DIFFERENT FORMATS AND LANGUAGES.

EDITION LAYER CONVERTS THEM INTO A COMMON FORMAT AND LANGUAGE.

WE CALL THIS LAYER THE “TRANSLATOR” OR “COMMON LANGUAGE LAYER” BECAUSE IT IS A TRANSLATOR THAT MAKES SURE ALL “LANGUAGES” CAN BE UNDERSTOOD.

CRM CAN BE UNDERSTOOD IN ONE PLACE.

COMPONENTS OF MDM ARCHITECTURE & THEIR FUNCTIONS

DATA HUB

A **CENTRAL DATABASE (THE "BRAIN") THAT STORES**
OPEN RECORDS.

IMPORTANT: IT'S THE SINGLE SOURCE OF TRUTH FOR
EVERYTHING 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

• FIX SPELLING MISTAKES, REMOVE DUPLICATES,
• (e.g., "PHILIPPIES" vs. "PHILIPPINES")

• GOLDEN RECORD: ENSURES THE GOLDEN RECORD IS ACCURATE

COMPONENTS OF MDM ARCHITECTURE & THEIR FUNCTIONS

RS

S, APPLICATIONS, AND REPORTS THAT USE THE DATA FROM MDM.

BUSINESS APPLICATIONS (ERP, CRM), BI DASHBOARDS,

FINANCIAL REPORTS.

MANUFACTURING: THEY RELY ON MDM TO GET CONSISTENT DATA FOR PRODUCTION PLANNING.

WORKING.

MDM PROCESSES

INTEGRATION

INTEGRATION: BRINGING DATA FROM MULTIPLE SOURCE SYSTEMS.

DATA VALIDATION

VALIDATION: IDENTIFYING INACCURATE RECORDS, RESOLVING CONFLICTS.

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

MDM PROCESSES

E & VALIDATION

BUSINESS RULES AND CHECKING ACCURACY.

DATE OF BIRTH MUST BE VALID, EMAIL MUST FOLLOW

MDM PROCESSES

INTEGRATION

MANAGE AND MAINTAIN GOLDEN RECORDS TO OTHER SYSTEMS.

CUSTOMER INFO SENT TO SALES, FINANCE, AND MARKETING

FROM A SINGLE SOURCE, AVOIDING DUPLICATES AND UPDATES TO THE SAME RECORD.

DATA ENRICHMENT & MAINTENANCE

REGULARLY CHECKING, UPDATING, AND IMPROVING DATA QUALITY

ENSURING THAT THE MDM DOESN'T BECOME OUTDATED OR INCONSISTENT

DATA QUALITY TOOLS

R TOOLS:

DATA MDM

ATMOSPHERE

DG

MDM

DATA QUALITY TOOLS

TOOLS:

USED MDM:

PLACEMENT MDM IN THE CLOUD INSTEAD OF ON-PREMISES IS PREFERABLE, SCALABLE, COST-EFFICIENT.

INSTEAD 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

? LEADERS (LIKE DEPARTMENT HEADS).

WHAT THE DATA MEANS AND WHO CAN ACCESS IT.

D DECIDES HOW "EMPLOYEE RECORDS" ARE STRUCTURED.

STEWARDS

DAY CARETAKERS OF DATA.

ENSURE ACCURACY, FIX ERRORS, AND MONITOR DATA QUALITY.

INCE DATA STEWARD ENSURES VENDOR BANK ACCOUNTS

ROLES IN GOVERNANCE & STEWARDSHIP

ADMINISTRATORS

E 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

CONTROL (RBAC – ROLE-BASED ACCESS CONTROL)

ROLE ONLY GETS THE DATA IT NEEDS.

AN SEE VENDOR BANK INFO NOT EMPLOYEE MEDICAL RECORDS

GOVERNANCE PROCESSES IN MDM

QUALITY MONITORING

CHECKS FOR DUPLICATES, MISSING FIELDS, OR IN-

' IS ENTERED AS CURRENCY, RDM + GOVERNANCE

COMPLIANCE

ASSESSED WHAT, WHEN, AND HOW.

COMPLIANCE WITH LAWS LIKE GDPR, HIPAA, OR LO

'GOVERNANCE & STEWARDSHIP MAKES IT POSSIBLE TO

CHAOS → WITHOUT RULES, EVERY SYSTEM IS A CHAOS.

T AS "CUSTOMER".

ACCURATE DATA → PREVENTS FINANCE FROM SAVING MONEY.

RY RECORDS.

TRANSPARENCY → PEOPLE ONLY TRUST DASHBOARDS AND REPORTS WHEN THE DATA IS ACCURATE AND CONSISTENT.

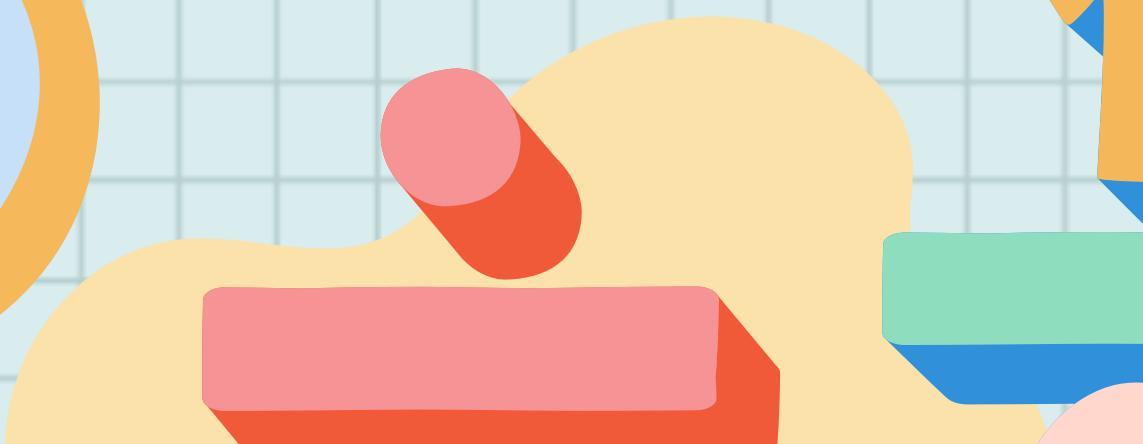
COMPLIANCE → AVOIDS LEGAL ISSUES (E.G.,

3



THE END

anks for listening



TA INTEGRATI

&

TA PROVENANT

GROUP 8

DATA INTEGRA

tion involves combining data from different sources for a common view for analysis, reporting, and other purposes

Source(Recipe , Procedure, Equipment, Lobo)

EXAMPLES OF DATA INTEGRATION

The diagram features three overlapping circles. The top-left circle is light blue and contains the text 'DATA GOVERNANCE'. The bottom-center circle is dark blue and contains the text 'SCALABILITY AND PERFORMANCE'. The right circle is light blue and contains the text 'FLEXIBILITY AND ADAPTABILITY'. The circles overlap in a way that suggests they are interconnected concepts.

DATA
GOVERNANCE

SCALABILITY
AND
PERFORMANCE

FLEXIBILITY
AND
ADAPTABILITY

ALITY AND CON-

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

Students Record : Name (John Doe)

Record : Name (Jane, Age : 90)

DATA GOVERNANCE

U policies and procedures for managing data assets, including data 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

Data is cleansed, standar

lized into a suitable format

before being loaded.

D OF DATA INTE

transform, Load):
processed data is loaded
typically 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

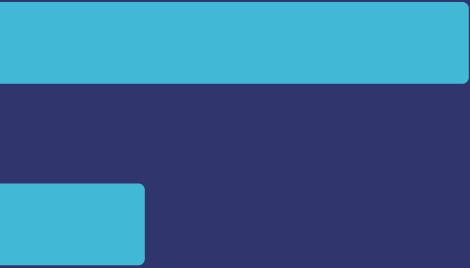
acilitate continuous data flow supporting various analytical ar

for managing the entire de
mption.

D OF DATA INTE

elines

ng Pipelines: Process large volumes of data over long intervals, ideal for tasks like monthly reports(Daily, Monthly, Yearly) Reporting Pipelines: Handle real-time data, such as user interactions or sentiment analysis (Like, Comment, Share) Real-time



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 CHALLENGE

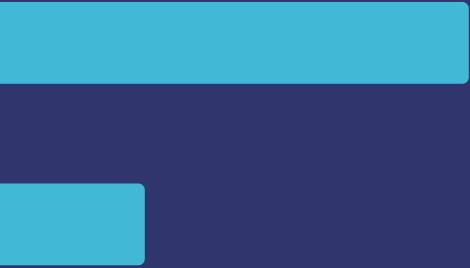
Warehouses provide centralized storage and distribution capabilities, but they also present several challenges. Existing systems and processes may not be designed to support the complex requirements of a warehouse environment. This can lead to inefficiencies, errors, and increased costs. Some common challenges include:

- Inventory management: Managing structured, semi-structured, and unstructured data types simultaneously.
- Order fulfillment: Handling large volumes of orders efficiently while ensuring accuracy and timely delivery.
- Supplier integration: Managing relationships with multiple suppliers to ensure timely delivery of raw materials.
- Customer service: Providing timely and accurate information to customers about product availability, shipping status, and delivery times.
- Compliance: Ensuring adherence to industry regulations and standards related to food safety, quality control, and environmental sustainability.

WAREHOUSING INTEGRATION CHALLENGES

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

Addressing these challenges requires advanced quality assurance mechanisms.



DATA PROVENANCE

DATI PROVENANTI

ance, also known as
cking the history and f

DATA PROVENANCE

include:

Data Flow: Identifying where data moves across systems.

Transformations: Recording when data is aggregated, or transformed.

Accountability: Providing the source of evidence, 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
from multiple sources in order
to find useful business information.

ANALYTICAL PROCESSING

layer that uses the multi-
model to enable users to
query 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
transformations, data cleaning procedures,
and preparation to be loaded into the data
warehouse.

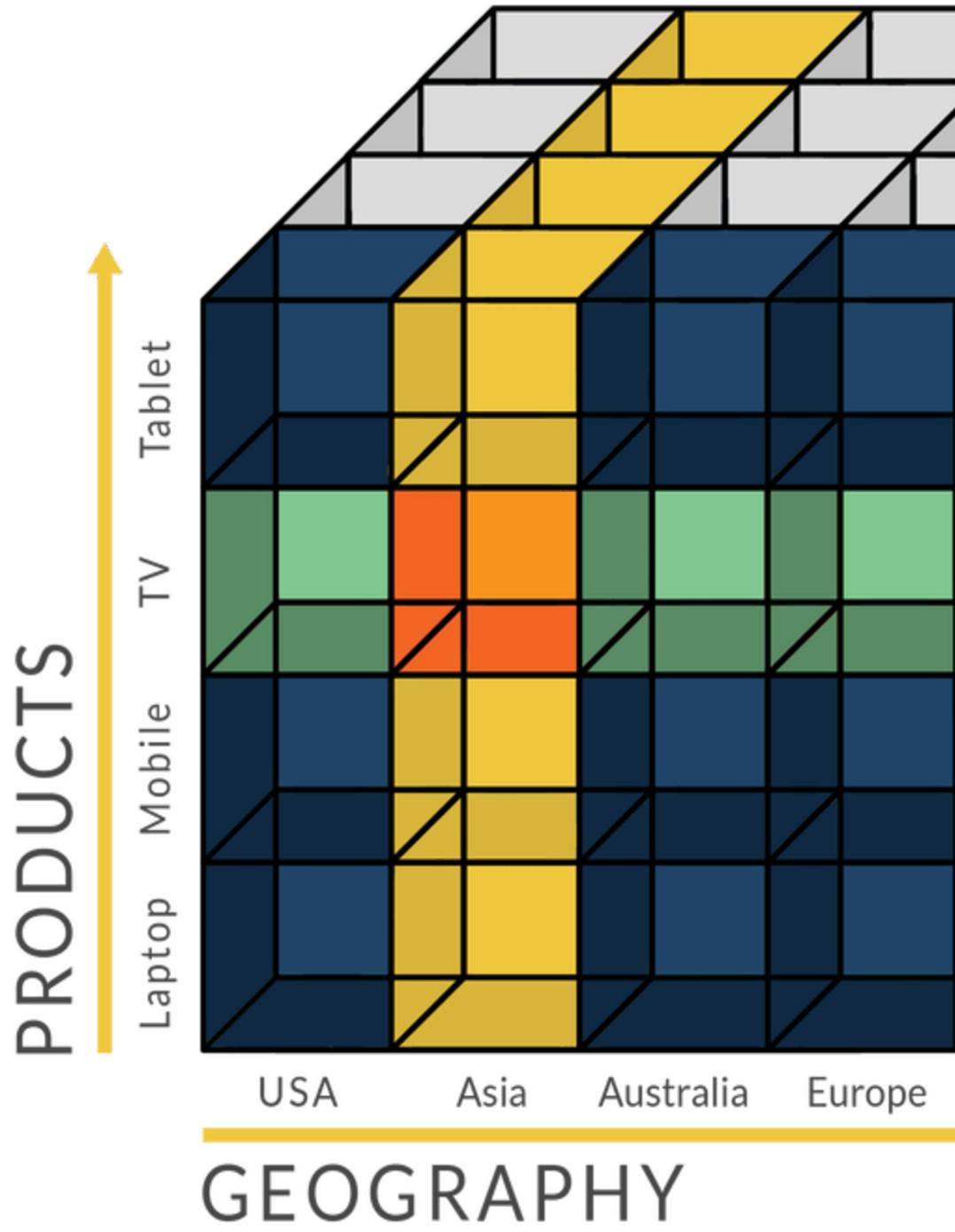
Change this process:

• Pulls raw data from sources.
• Standardizes and formats the data.
• Loads the data into the data warehouse

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

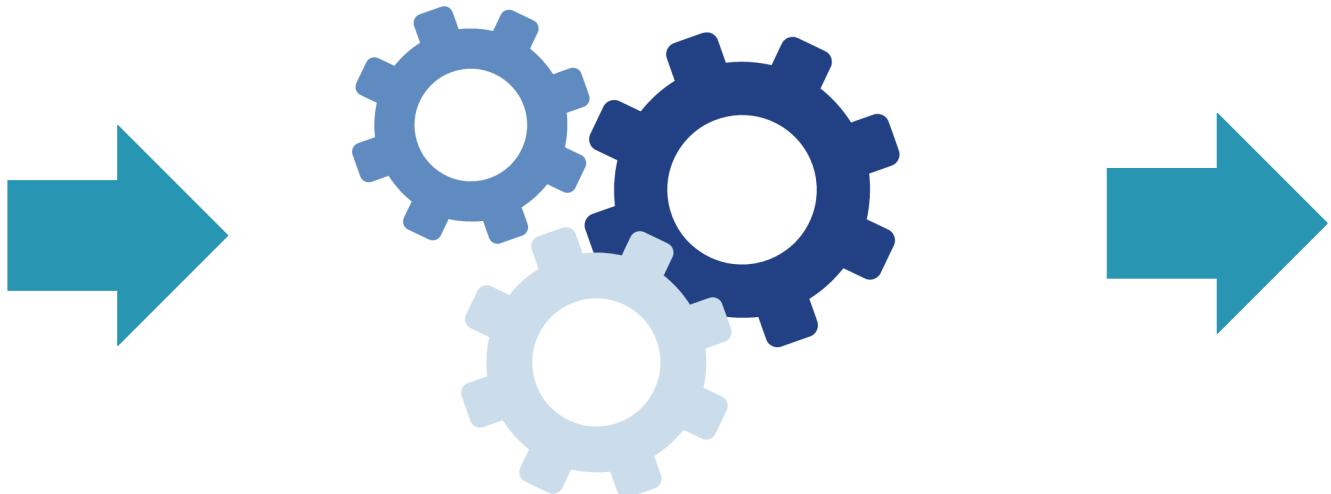
Ex. Sales, Marketing

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,



RACT, **T**RANSFORM, **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.



Collection: Data is collected from different sources, databases, spreadsheets, CRM systems, logs, and other raw data, no matter the format (structured or unstructured).

Integration/Tools: SQL, APIs, data connectors, Fivetran, Talend, Informatica, Alteryx.

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 ready for analysis.

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

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

Technologies/tools: ETL pipelines, batch/stream processing engines (Apache Flink, Apache Spark), data integration tools (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)

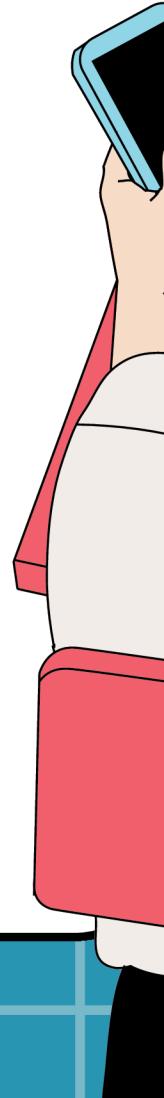
Processing approach emphasizing real-time. The majority of OLTP systems are made of short atomic operations that keep data consistent.

OLAP VS. OLTP

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

WHAT IS A STAR SCHEMA FOR BUSINESS INTELLIGENCE

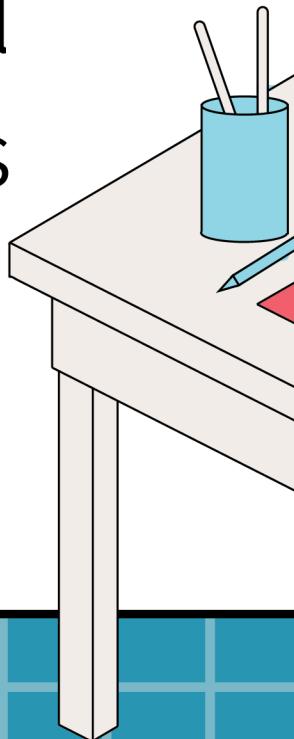
is a type of multidimensional structures information in a way simple to interpret and analyze. A star schema is important in Business Intelligence (BI) because it provides the ability to store, retrieve, monitor and organize data in a way that supports reporting and analysis fast and efficiently.

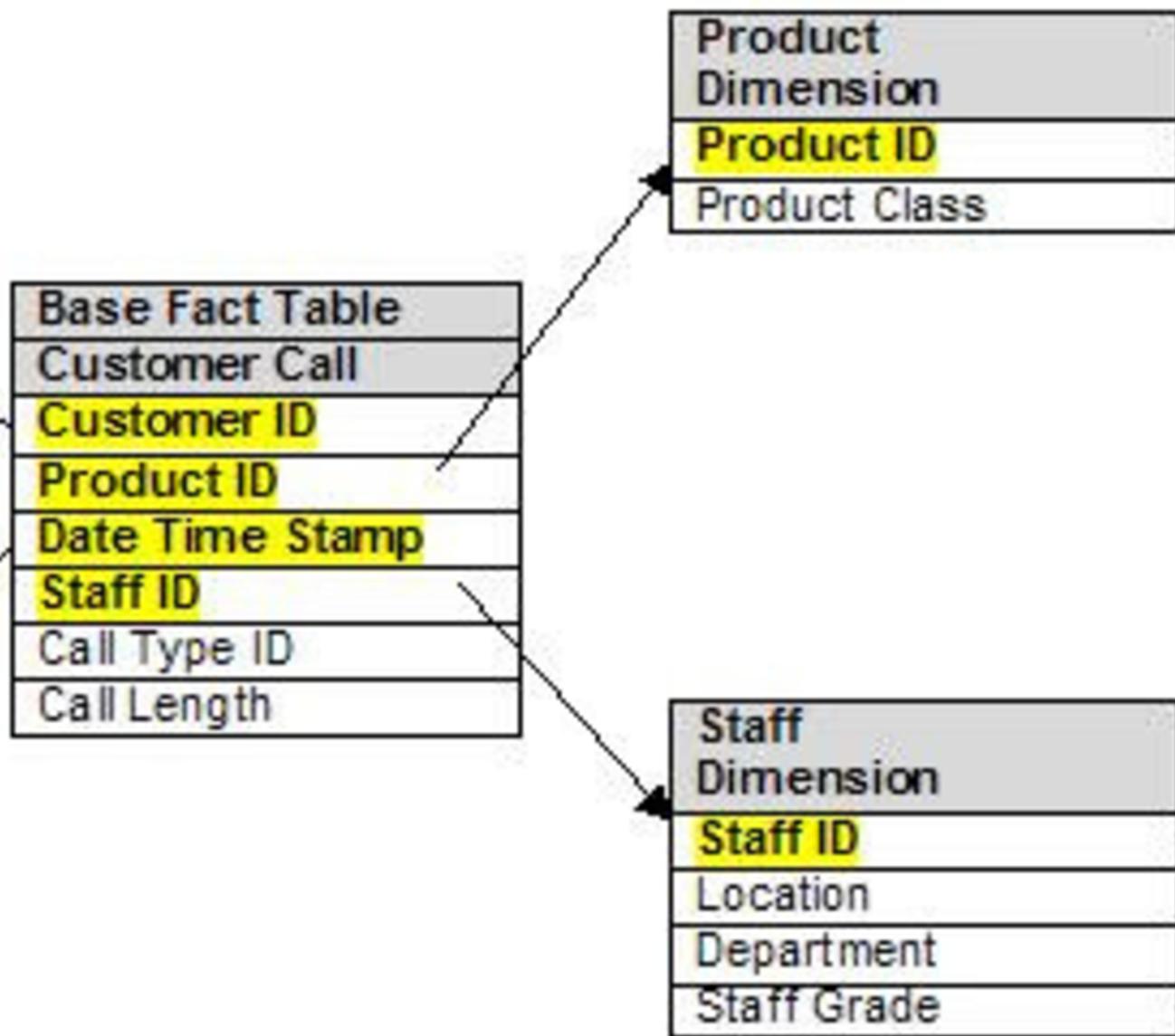


olds the key business metrics
in amounts and quantities.

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

[databricks.com/overview-of-data-warehouses](https://www.databricks.com/overview-of-data-warehouses)

geeks.org/dbms/data-warehouse-area

<https://www.software-testing.com/what-is-a-data-structure>

What is ETL?

geeks.org/dbms/difference-between

<https://www.semanticscience.org/glossary/star-schema>