WEEK 5

NFORMATION LIFECYCLE MANAGEMEN

RODUCTION

at 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

TION

STO

USE

IVAL

DI

ES OF THE
TION LIFECYCLE

tially generated from various sources such as
stems. This can include structured data
structured data like emails, images, and docum

ent writes a research paper or a company crea

ES OF THE
TION LIFECYCLE

e data moves into the Storage stage, where it
torage may be physical, such as paper files
atabases, servers, or cloud systems. Proper st
zed, and available when needed.

the research paper on a computer hard drive,

ES OF THE
TION LIFECYCLE

s Use, where data is actively accessed, pr
rations and decision-making. At this stage, r
information through reporting, analysis, o

eacher reads the paper to grade it, or empl
ion-making.

ES OF THE
TION LIFECYCLE

l is no longer needed for daily operations, it e
to long-term storage, where it remains availabl
es but does not burden active systems. Archive
solutions optimized for long-term retention.

the semester, the research paper is stored in
or the company files the report for auditing purp

ES OF THE
TION LIFECYCLE

n reaches the Disposal stage, where it is securel
gally required. Secure disposal prevents unauthoriz
ations. This may include shredding physical docum
ysically destroying storage devices.

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

A RETENTION A
CHING POLIC

s

ntion

r 7

ears

ay

s in

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.

Kee
rev
and

•

•

GING DATA
UGH ITS LIFECYCLE

nent Management
ms (DMS)

anize, track, and
age documents in
al form.
vide version control
easy retrieval of
.

uce reliance on paper
rds.

Cloud Stor
Lifecycle
Manageme

- Automati
  files to cl
  storage.
- Ensures i
  remains a
  lowering
- Offers sc
  flexibility
  organizat

GING DATA
UGH ITS LIFECYCLE

nent Management
ms (DMS)

uses a DMS to store
' grades, assignments,
ndance records digitally
of keeping piles of
lders.
es in a law firm can
search and retrieve
s without digging
filing cabinets.

Cloud Storage w
Lifecycle Manag
Rules

EXAMPLE:

- A company s
files in Goog
not opened i
move to a ch
but can still
needed.
- A photograph
client photos
storage, but
photos are s
storage to s

AGING DATA UGH ITS LIFECYCLE

ion and Access

t sensitive information
nauthorized access.
otion ensures data
s unreadable without
rrect key.
s controls (like
ords, roles, and multi-
authentication) allow
thorized staff to view
data.

Backup and D
Recovery Sys

- Create regu
  critical dat
- Ensure quic
  system fail
  or natural c
- Provide bus
  reduce dow

AGING DATA
UGH ITS LIFECYCLE

n and Access

: encrypts customer
t numbers, so even if
s steal the database,
ormation looks like
n symbols without the
tion key.
chool, only teachers with
word and 2-step
ation can access
ts' grades.

Backup and Disa
Recovery Systen

EXAMPLE:

- A hospital c
patient medi
cyberattack
happens, the
the files and
patients.
- A company s
business file
office compu
destroyed b
recover all c
continue ope

CHALLENGES IN M
INFORMATION

xity

nd variety of data are
rapidly. Multiple data
and formats increase
ement difficulties.

EXAMPLE:

t has photos, videos,
l assignments saved on
ne, laptop, and Google
Since the files are in
olaces and formats, it's
p everything organized.

Cost

Managing the
requires sign
storage infra
data managem
software
pro

E

The studen
Google Drive s
paying for e
month. That's
manag

CHALLENGES IN M
INFORMATION

Concerns

nstantly under
yberattacks such
e, phishing, and
mpts, which can
ere financial and
nal damage.

someone hacks a
stem and leaks
rds, the school
s trust.

Compliance Risk

Failing to comply
in legal pena
reputational dam
loss of customer

Example: If a
doesn't keep p
records properly
face big legal p


THANK
YOU

ent

Contact

TA QUAL

MANAGEMENT

Start

ent

Contact

Quality Management (DQM)

nagement is the process of m
, complete, consistent, and up
l used effectively. It focu
iable by setting rules, check
ta over time.

DQM to ensure customer account de
actions) are accurate and up to date.

ent

Contact

IMENSIONS OF DATA QUALI

should correctly represent rea

o missing or incomplete values

Data should be uniform acro

ata must be up-to-date and

ality Frameworks and St

Organization for Standardizat

ndard for data quality.

data is accurate, complete, a

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

ecks – Ensure inputs meet rule

ing – Analyzing dataset

es. Audits – Independent re

ent

Contact

Improving Data Quality

ols:

- Removing errors, duplica

cleans student records to merge dupl

n – Converting data into a uni

standardizes date formats (MM/DD/

ent

Contact

Improving Data Quality

ools:

– Detecting and removing dupli

ges duplicate product listings to av

t – Adding external data to im

ch customer profiles with updated c

Governance Strategies

es – Rules for data handling ar

sets strict policies on user data acce

toring – Continuous checks on c

tor transaction data daily to detect

ent

Contact

Governance Strategies

Structures – Assigning re

nt agencies, specific offices are m
tabases.

reness – Educating staff on t

n staff to input passenger data correc

ent

Contact

ANK YO

MASTER DATA MANAGEMEN

&

FERENCE DATA MANAGEM

Group 7

WHAT IS DATA

ce it

d

Problem

Data can be duplicated, inconsistent, or incomplete.

ER DATA MANAGEMENT (

ICE OF CREATING ONE TRUSTED VERSION OF THE

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

GLE, CORRECT VERSION OF THAT STUDENT RECORD ACROSS ALL SYSTEMS

IMPORTANCE OF MDM

VES DUPLICATION.

VES REPORTING AND ANALYTICS.

S EFFICIENCY.

ENCE DATA MANAGEMENT

ISTS, CODES, AND CATEGORIES TO ENSURE ALL SYSTEM

INTO PHP (PHILIPPINE PESO).

OLE OF RDM IN INTEGRAT

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

OLE OF RDM IN INTEGRAT

ATION IN INTEGRATION

VALUES FROM ENTERING SYSTEMS DURING INTEGRATION.

W VENDOR HAS CURRENCY = "PHL," RDM REJECTS IT SI

TEROPERABILITY

ATIONS CAN TALK TO EACH OTHER SMOOTHLY BECAUSE THEY

ROLE OF RDM IN REPORTI

E AGGREGATION

TA FROM MULTIPLE SOURCES WITHOUT DUPLICATION

S," "USA," AND "UNITED STATES" ARE ALL REPO

FD DECISION-MAKING

ELY ON TRUSTED, STANDARDIZED DATA → NOT CONI

ROLE OF RDM IN REPORTI

CONSISTENCY

GORIES CHANGE (LIKE PRODUCT LINES), RDM MAINTAI

ONSISTENT.

1

NES

PHONES, CELL PHONES

M ARCHITECTURE & PROCES

PONENTS OF MDM ARCHITECTURE & F

ES

NAL SYSTEMS WHERE DATA IS CREATED AND STOR

(ENTERPRISE RESOURCE PLANNING), CRM (CU

HR, FINANCE, SPREADSHEETS.

ponents of MDM architecture & F

on layer

and processes (like ETL – Extract, Trans

ource systems to the MDM hub.

tant: Different systems store data in dif

layer converts them into a common form

e a translator that makes sure all "lan

CRM can be understood in one place.

ponents of MDM Architecture & F

data hub

ntral database (the "brain") that stores

den Records.

tant: It's the single source of truth for

e the registrar's office in a school that

—no duplicates, no errors.

PONENTS OF MDM ARCHITECTURE & F

LITY TOOLS

ATED TOOLS THAT CLEAN, VALIDATE, AND STAN

IX SPELLING MISTAKES, REMOVE DUPLICATES,

VS. "PHILIPPINES")

ANT: ENSURES THE GOLDEN RECORD IS ACCUR

PONENTS OF MDM ARCHITECTURE & F

RS

S, APPLICATIONS, AND REPORTS THAT USE THE

USINESS APPLICATIONS (ERP, CRM), BI DAS

NANCIAL REPORTS.

ANT: THEY RELY ON MDM TO GET CONSISTEN

KING.

MDM PROCESSES

TION

W DATA FROM MULTIPLE SOURCE SYSTEMS.

IDATION

ICATE RECORDS, RESOLVING CONFLICTS.

CRUZ" + "J PAUL CRUZ" → JOHN PAUL

MDM PROCESSES

E & VALIDATION

SINESS RULES AND CHECKING ACCURACY.

TE OF BIRTH MUST BE VALID, EMAIL MUST FO

MDM PROCESSES

BUTION

AN, GOLDEN RECORDS TO OTHER SYSTEMS.

STOMER INFO SENT TO SALES, FINANCE, AND

THE SAME RECORD.

G & MAINTENANCE

CHECKING, UPDATING, AND IMPROVING DATA

M DOESN'T BECOME OUTDATED OR INCONSIS

DATA QUALITY TOOLS

R TOOLS:

A TICA MDM

FOSPHERE

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.

GOVERNANCE & STEWARD

MDM GOVERNANCE & STEWARDSHIP

GOVERNANCE

E POLICIES, RULES, AND DECISION-MAKING FRAME

N IN MDM.

TEWARDSHIP

E PEOPLE AND ROLES RESPONSIBLE FOR MAINTAIN

COMPLIANCE.

ROLES IN GOVERNANCE & STEWARD

NERS

R LEADERS (LIKE DEPARTMENT HEADS).

WHAT THE DATA MEANS AND WHO CAN ACCESS IT.

D DECIDES HOW "EMPLOYEE RECORDS" ARE STRUCTURE

WARDS

-DAY CARETAKERS OF DATA.

E ACCURACY, FIX ERRORS, AND MONITOR DATA QUALITY.

NCE DATA STEWARD ENSURES VENDOR BANK ACCOUNTS

ROLES IN GOVERNANCE & STEWARD

MINISTRATORS

E SYSTEM RULES, PERMISSIONS, AND INTE

KE SURE HR ONLY GETS EMPLOYEE DATA,

GOVERNANCE PROCESSES IN MDM

CIES

W DATA IS CREATED, STORED, SHARED, AND SEC

TOMER RECORDS MUST INCLUDE EMAIL + PHON

ITROL (RBAC – ROLE-BASED ACCES

OLE ONLY GETS THE DATA IT NEEDS.

N SEE VENDOR BANK INFO NOT EMPLOYEE MED

GOVERNANCE PROCESSES IN MDM

LITY MONITORING

CHECKS FOR DUPLICATES, MISSING FIELDS, OR IN

' IS ENTERED AS CURRENCY, RDM + GOVERNA

MPLIANCE

SSED WHAT, WHEN, AND HOW.

IANCE WITH LAWS LIKE GDPR, HIPAA, OR LO

GOVERNANCE & STEWARDSHIP MA

CHAOS → WITHOUT RULES, EVERY SYSTEM

T AS "CUSTOMER".

TIVE DATA → PREVENTS FINANCE FROM S

RY RECORDS.

T → PEOPLE ONLY TRUST DASHBOARDS A

E DATA IS ACCURATE AND CONSISTENT.

MPLIANCE → AVOIDS LEGAL ISSUES (E.G.,

THE END

anks for listenin

TA INTEGRATI

&

TA PROVENAN

GROUP 8

DATA INTEGRA

n involves combining data from di
view for analysis, reporting, and
poses

Source(Recipe , Procedure, Equ
lobo)

LES OF DATA GRATION

DATA GOVERNANCE

SCALABILITY AND PERFORMANCE

FL

AD

ALITY AND CONS

curacy, completeness, ar
ss all sources and targets

tudents Record : Name (
d : Name (Jam, Age : 90)

ATA GOVERNAN

g policies and proce
data assets, includin
d compliance.

ata Privacy (G-cash)

LITY AND PERFO

ntegration solutions th
lata volumes and proce

der Monday (100) - Orde

LITY AND ADAP

ems that can easily integ
adapt to evolving busines

dent Record

DATA MANAGE

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

lethods of Data Int
(ETL,ETL, DATA PIP

D OF DATA INTE

ansform, Load):

a is extracted from sourc
Data is cleansed, stand
into a suitable format
being loaded.

D OF DATA INTE

ransform, Load):

processed data is loade

ically a data warehouse.

al Different source:(Patient

emove duplicate name,

r, Snowflake

D OF DATA INTE

Load, Transform):

data is extracted from so

w data is loaded directly

cloud data warehouse or

D OF DATA INTE

oad, Transform):

nsformations are perfor

leveraging its processing

D OF DATA INTE

a series of automated process
m its source to a destination.
s various integration methods
g, depending on the specific re

D OF DATA INTE

acilitate continuous data flo
porting various analytical a

for managing the entire d
mption.

D OF DATA INTE

elines

ng Pipelines: Process large vo
rvals, ideal for tasks like month
s(Daily, Monthly, Yearly) Repor
elines: Handle real-time data,
such as user interactions or sen
k (Like, Comment, Share) Real-

Data Warehous and Integration Challe

WAREHOUSING
GRATION CHALLE

g is the process of collect
amounts of data from differe
ed a data warehouse) for anal

WAREHOUSING
GRATION CHALLE

ehouses provide centralized
several challenges exist
urces:
y: Managing structured, sem

WAREHOUSING
GRATION CHALLE

nsuring accuracy, completeness
dling large and growing volume
cing real-time and batch data in

e challenges requires adva
ality assurance mechanisms.

DATA PROVENA

TA PROVENANC

ance, also known as
cking the history and f

TA PROVENANC

lude:

ta Flow: Identifying where
es across systems.
Transformations: Recordi
gated, or transformed.
ccountability: Providing t
iance, and troubleshooting.

TA PROVENANC

ning lineage, organ
ust in their data
ompliance.

THANK YOU

TA WAREHOUSIN

NLINE ANALYTIC

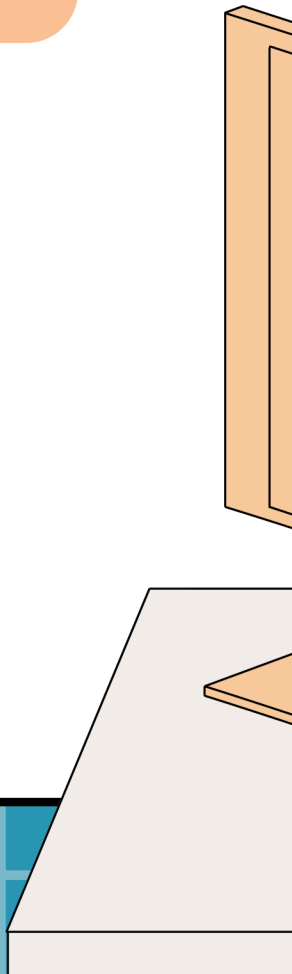
ROCESSING (OLA

HOUSING

gathering, keeping, and
m multiple sources in order
ful business information.

LYTICAL PROCESSING

that uses the multi-
model to enable users to
n several database systems



AREHOUSING ARCHITEC

that helps businesses make bett
y combining data from variou
d organizing it under a sing
It simplifies data managemer
and storage, increasing th
s of analysis.

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

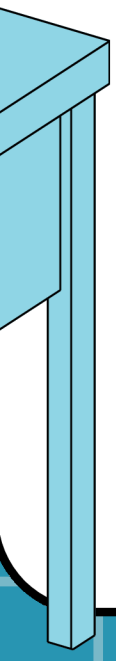
ge this process:

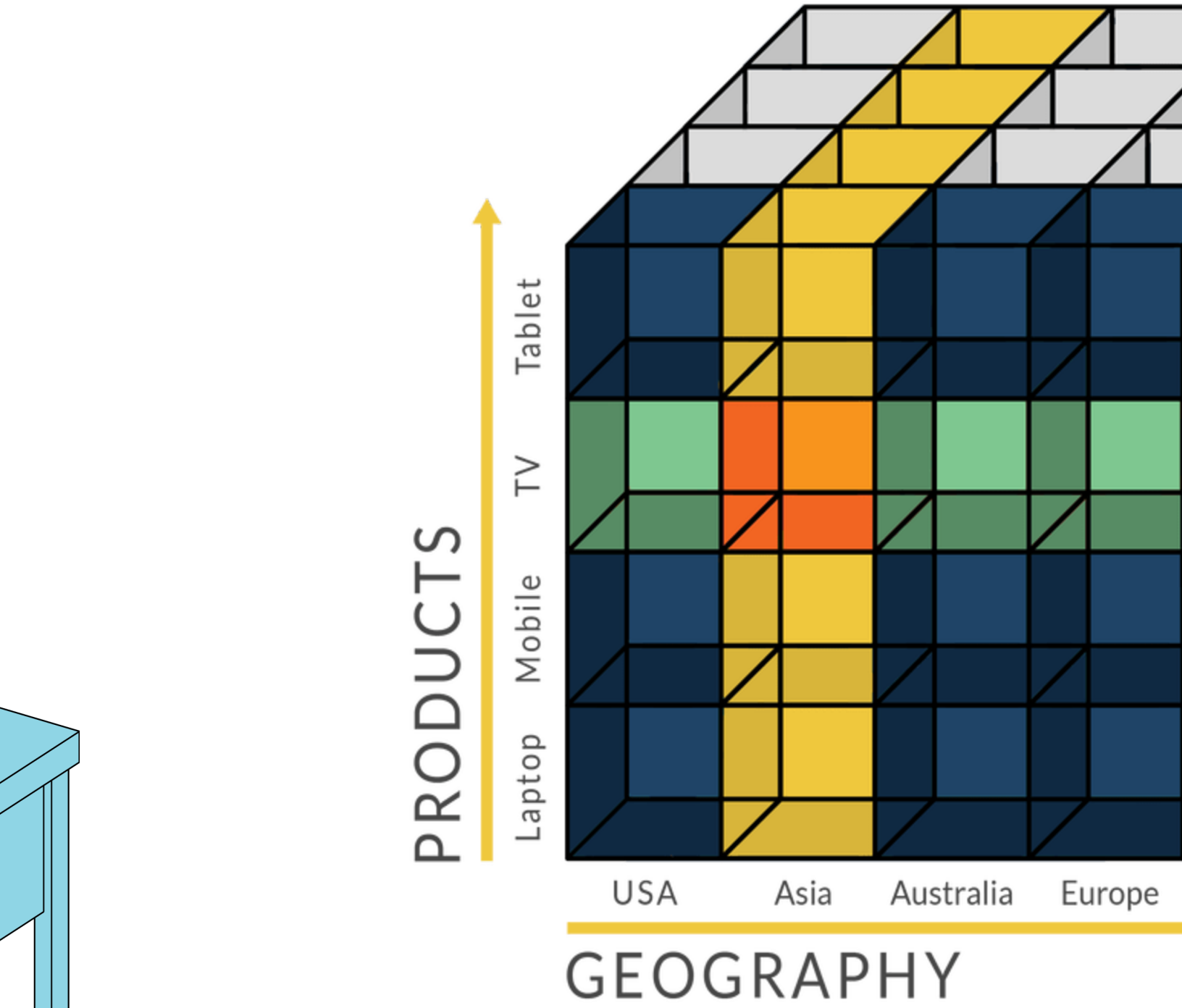
w data from sources.
dardizes and formats the data.
e data into the data warehouse

Data Marts is the smaller, speci
the data warehouse that serve
giving them direct access to the d

Ex. Sales, Marketing

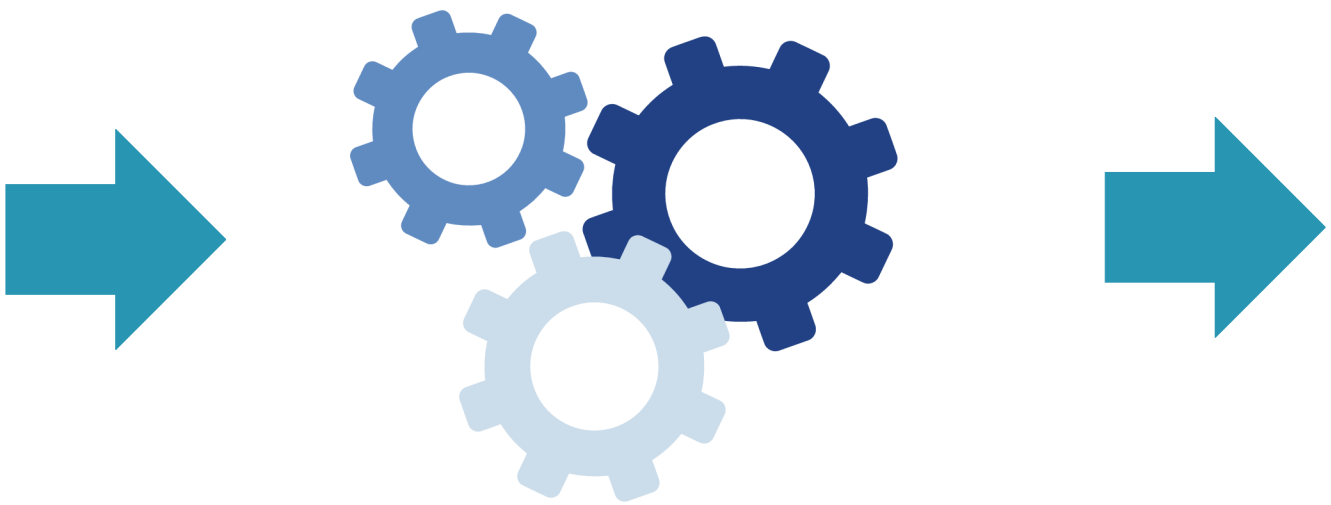
An **OLAP cube** is a special way
multi-dimensional structure that
faster to analyze. Instead of loo
tables,




PRODUCTS
Laptop
Mobile
TV
Tablet
USA
Asia
Australia
Europe
GEOGRAPHY

RACT, TRANSFORM, L

ocess of gathering data from differen
to a central storage system called a d
cess, raw data is cleaned, structured,
rules so that it's ready for storage, analy



pens: Data is collected from different
ses, spreadsheets, CRM systems, lo

er raw data, no matter the format (str
unstructured).
gies/tools: SQL, APIs, data connect
oop, Talend, Informatica, Fivetran.

pens: The extracted data is cleaned,
d according to business rules. This
duplicates, correcting errors, standar
time), and sometimes enriching the d
vert messy raw data into consisten
n ready for analysis.
gies/tools: Apache Spark, AWS G
Talend.

pens: The transformed data is store
e (like Amazon Redshift, Goog
, or Microsoft SQL Server).
e the data accessible for analytics
and machine learning.
gies/tools: ETL pipelines, batch/strea
ces (Airbyte, Informatica, Azure Data

OLAP VS. OLTP

l Processing (OLAP)

tware tools used for the analysis of
king processes. It generally allow use
m various perspectives.

ion Processing (OLTP)

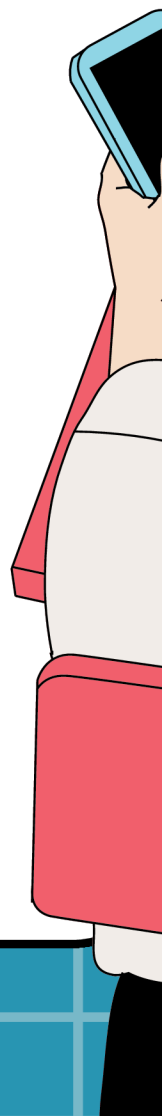
cessing approach emphasizing real-ti
The majority of OLTP systems are m
hort atomic operations that keep datal

OLAP VS. OLTP

	OLAP (Online Analytical Processing)	OLTP (
	Used for **analysis** and **decision-making**	Use
	Forecasting, business reporting, dashboards, market trend analysis	Ban e-com airlin
ols	Microsoft SSAS, Oracle OLAP, SAP BW, Tableau	MySQ Se

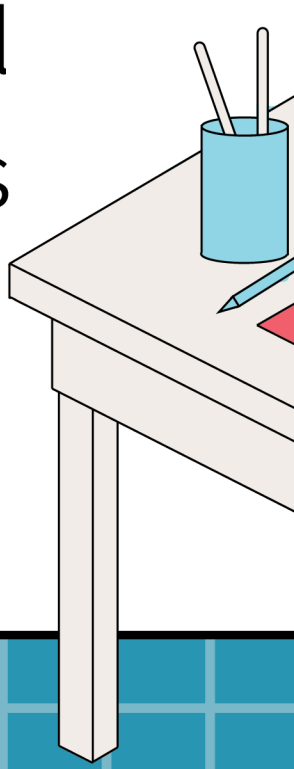
G A STAR SCHEMA FOR ESS INTELLIGENCE

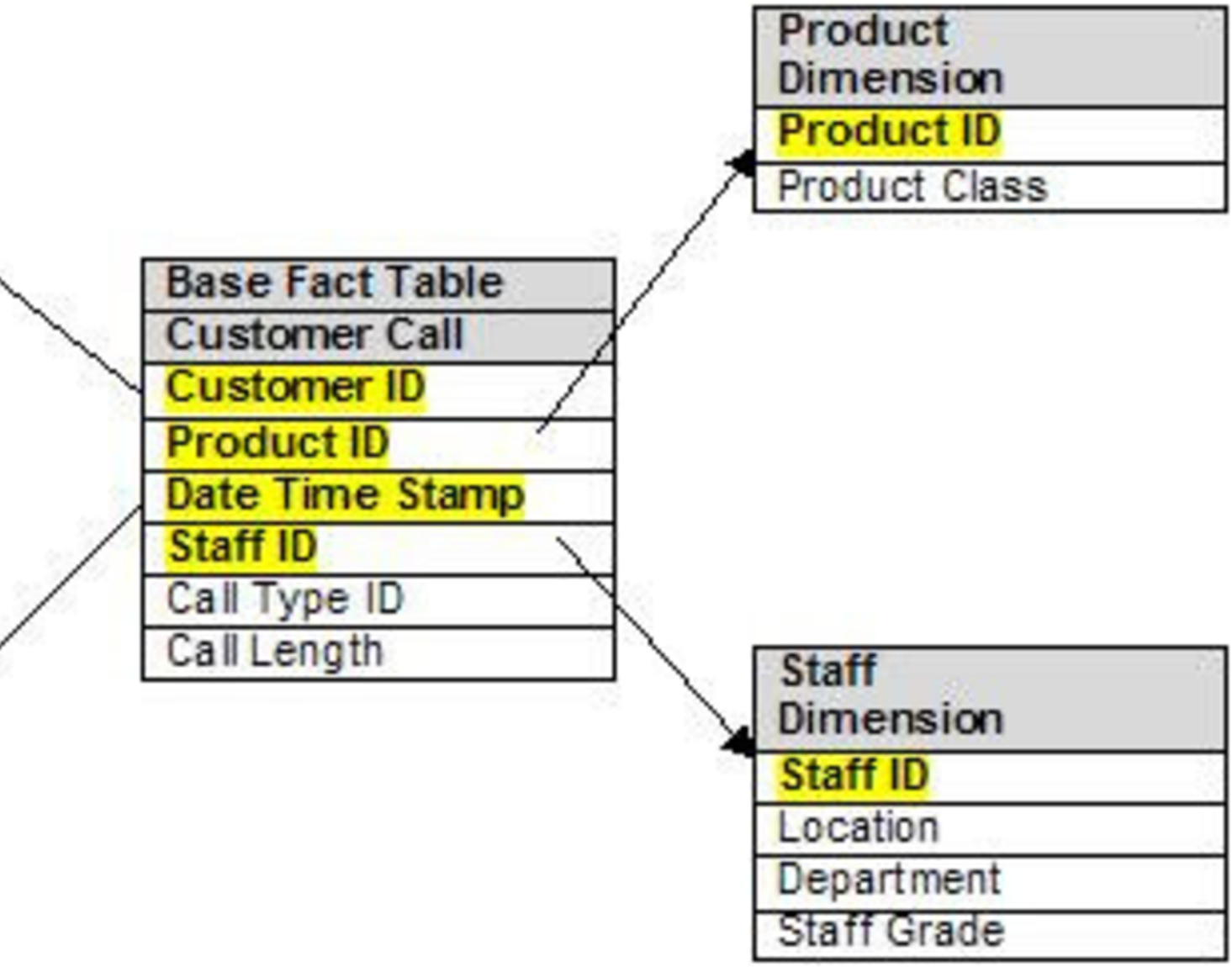
is a type of multidimensional structures information in a way nple to interpret and analyze. A is important in Business) because it provides the toring and organizing data in a reporting and analysis fast and



lds the key business metrics
in amounts and quantities.
es: provide the context like
t

is the center (measures) and
tables are the sides
and they are connected
key–foreign key relationships
eaningful.




Product Dimension
Product ID
Product Class
Base Fact Table
Customer Call
Customer ID
Product ID
Date Time Stamp
Staff ID
Call Type ID
Call Length
Staff Dimension
Staff ID
Location
Department
Staff Grade

THANK YOU

oint.com/overview-of-data-warehous

geeks.org/dbms/data-warehouse-ar

org/software-testing/what-is-a-data-

om/what-is/etl/

geeks.org/dbms/difference-betwee

s.com/glossary/star-schema