Learning Progress Review

Week 2

Introduction to Data and Database

Basic SQL

Intermediate SQL

DigitalSkola Batch 11

Kelompok 5 – Anak Digital

Eliza Harumanti

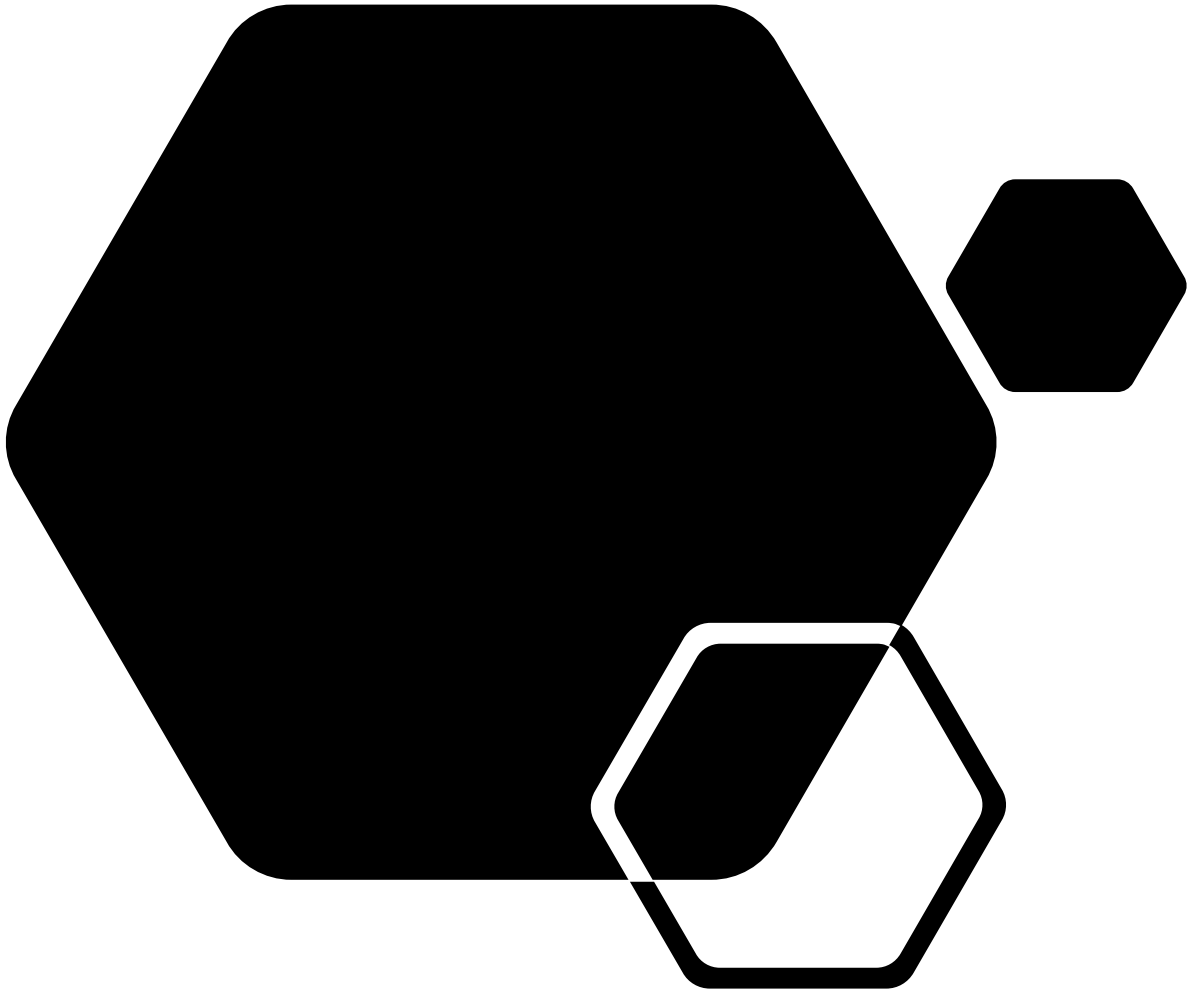
Johanes Kho

Lukman Budi Santoso

Mohammad Fadhil

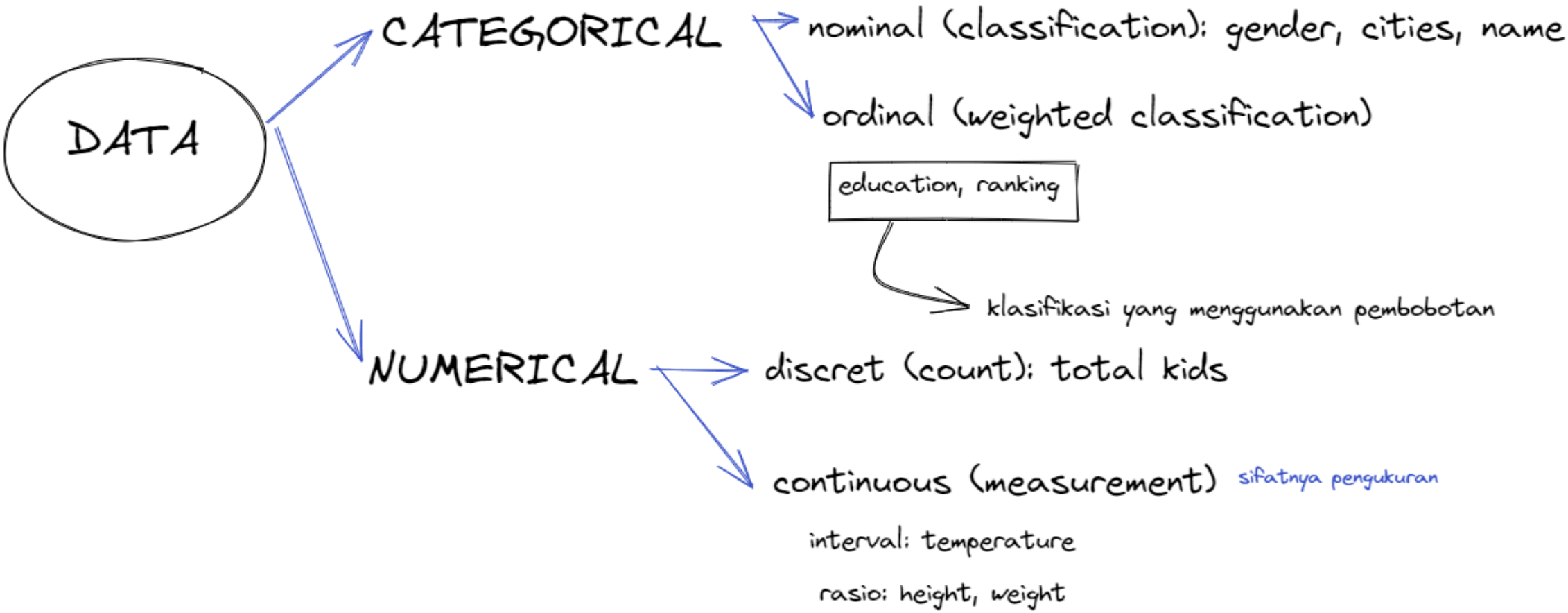
DigitalSkola

Introduction to Data and Database



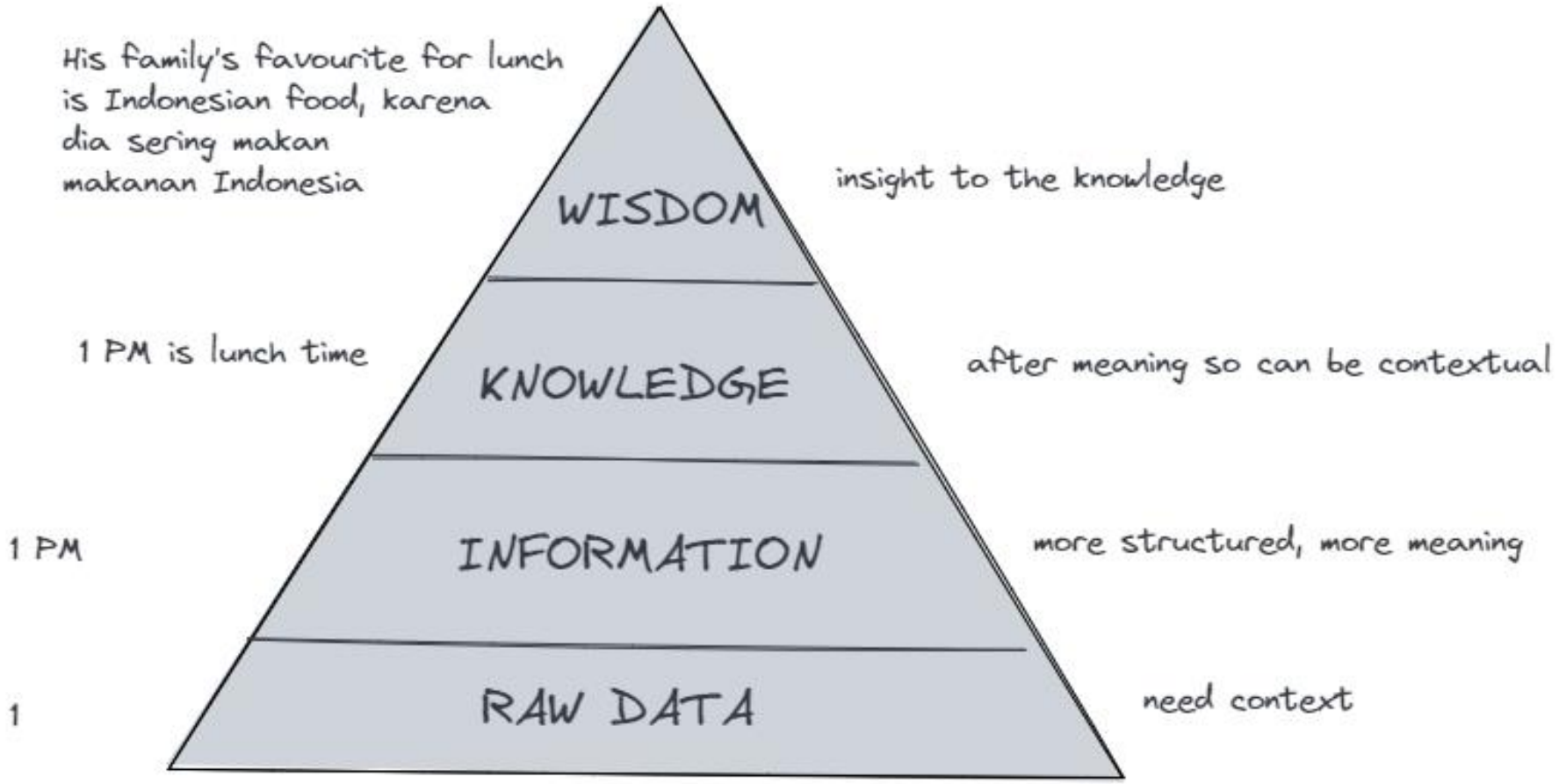
DigitalSkola

Data : collection of information that contains facts and numbers that is examined to support decision-making and available in multiple format



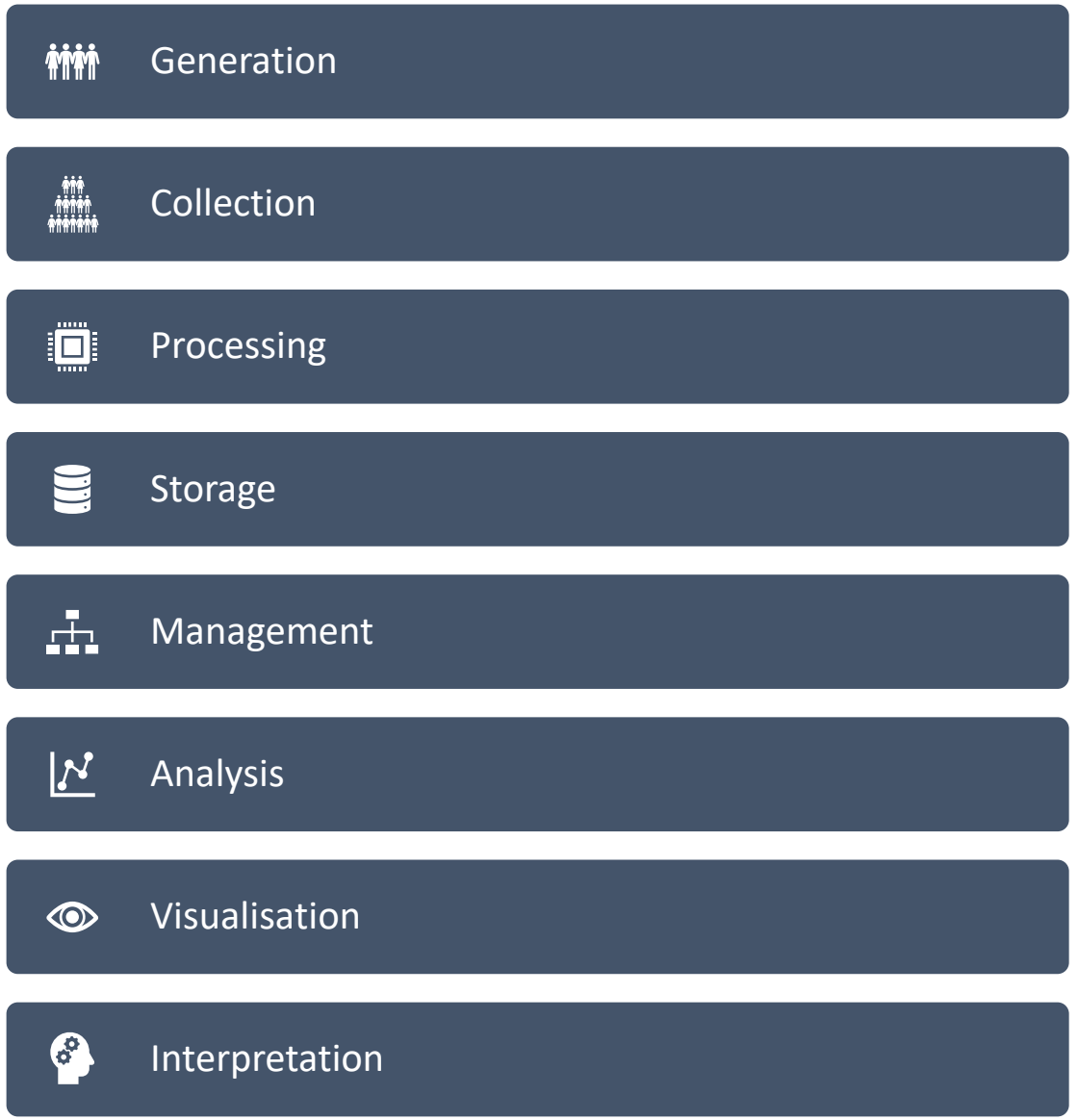
DigitalSkola

The level of data utilisation on how raw is examined to achieve insight



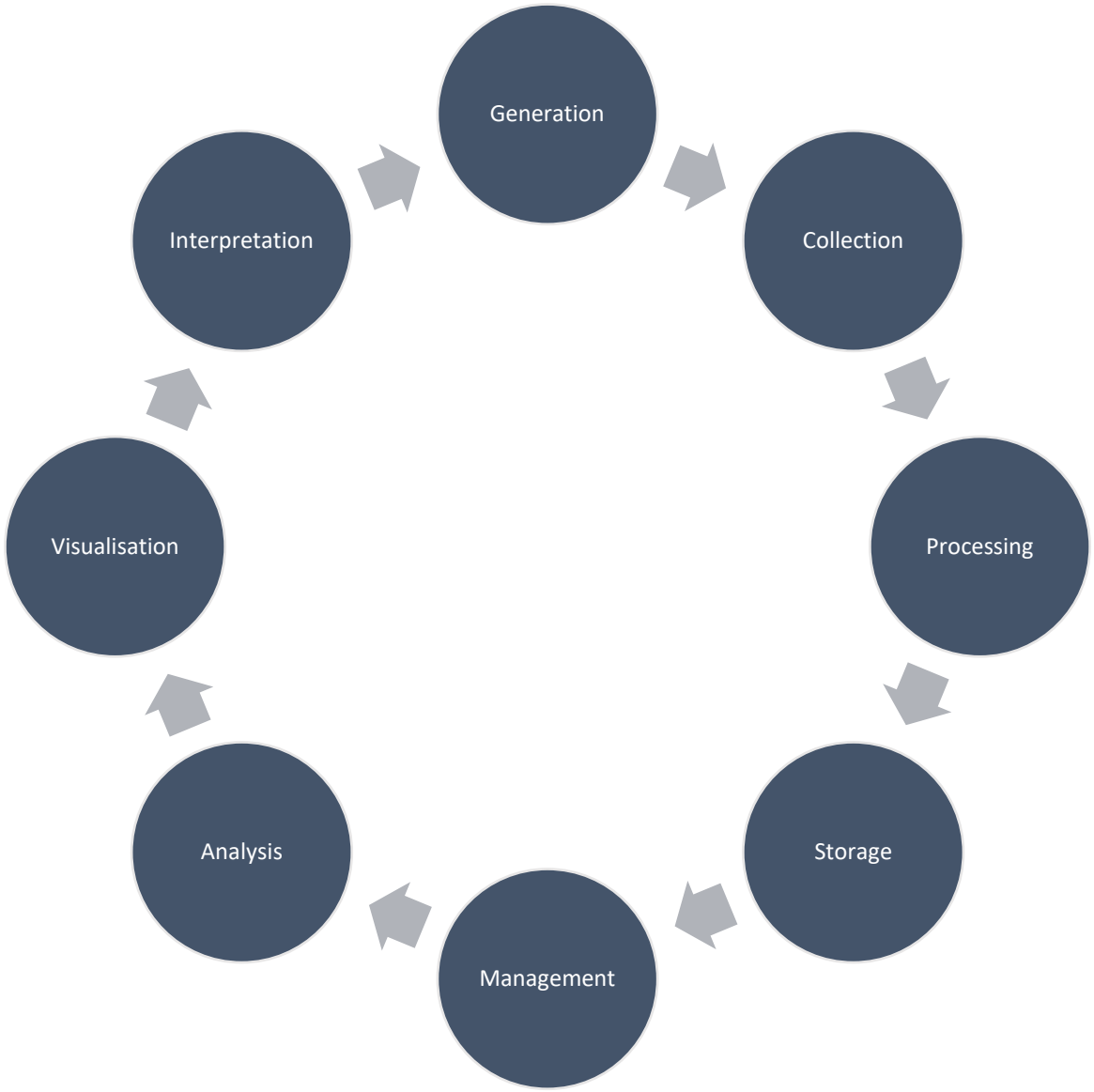
DigitalSkola

Data Lifecycle : the process from recording the data in your system to its utilisation



DigitalSkola

Data Lifecycle : the process from recording the data in your system to its utilisation



DigitalSkola

Database

A **database is** an organized collection of structured information, or data, typically stored electronically in a computer system

Some of the technical definition

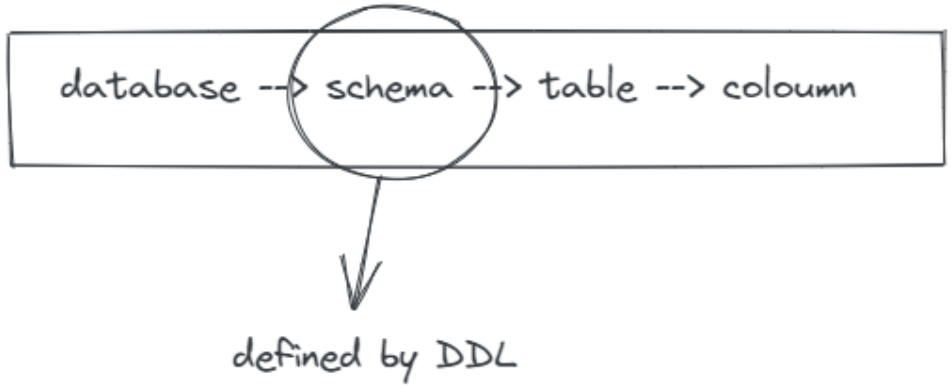
DBMS (Database Management System)

SQL (Structured Query Language)

DDL (Data Definition Language)

DML (Data Manipulation Language)

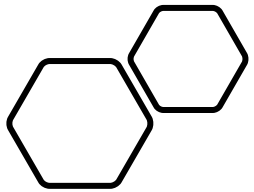
	STRUCTURED	UNSTRUCTURED
(+)	easier implementation for ML easily used by business users/end users compatible with lot of tools paling banyak dipakai orang	Keep original format as it is faster processing time save more storage while saved in data lake jadinya nyimpen sesuatu yang ga ada relasinya
(-)	Limited to relational database	need extra technical skills need specific tools
Tools	MySQL, MS SQL, PostgresSQL, SQLite	MongoDB, DynamoDB, Hadoop



DigitalSkola

Data Types

Numeric	Date/Time	Character/String	Unicode Characted	Binary	Miscellaneous
• Int • Bigint • Smallint • Float • Decimal • Real • bit	• Date • Time • Datetime • Timestamp • Year	• Char • Varchar • Text	• Nchar • Nvarchar • Ntext	• Binary • Varbinary	• Clob • Blob • xml



Basic SQL

SQL is important because of its ubiquity

- SQl is the second most-used programming language in data science after python
- SQL is now common for business analyst, with role ranging from Product to Marketing
- Since everything is stored in database, SQL becomes more important to learn

Schema

database --> schema --> table --> coloumn

located under database and created to cover specific purpose

```
create schema if not exists batch 11
;
```

DigitalSkola

Table

database --> schema --> table --> coloumn

located under schema and created to save data in a tabular format

```
create table if not exists batch_11.anggota_elizadh_5 (
    id int primary key,
    nama varchar(255) not null,
    no_hp varchar(13) unique not null,
    registered_time timestamp not null
);
```

Functions to modify any value in the table

- insert into: to fill the data manually or from another table
- update: to change any value in the table with condition set

```
insert into batch_11.anggota_elizadh_5 values
    (1,'Eliza Dayinta Harumanti','081295468696', '2022-01-12 20:50:30')
    ;

update batch_11.anggota_elizadh_5
set nama = 'Putri Marino'
;

insert into batch_11.anggota_elizadh_5 values
(2,'Sophia Latjuba','081234545643', current_timestamp)
;

update batch_11.anggota_elizadh_5
set nama = 'Marissa Anita'
where id = 2
;

insert into batch_11.anggota_elizadh_5
values (3,'Chicco Jerrikho', '0987654321234', current_timestamp)
;
```

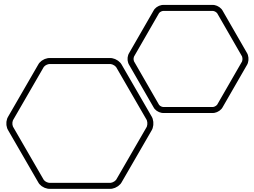
id	nama	no_hp	registered_time
1	Putri Marino	081295468696	2022-01-12 20:50:30.000
2	Marissa Anita	081234545643	2022-01-12 21:10:11.846
3	Chicco Jerrikho	0987654321234	2022-01-12 21:12:09.851

DigitalSkola

Function to modify columns in table

- alter table: to add or delete columns
- delete: to delete rows in the table under specific condition
- truncate: to delete all rows in the table without condition
- drop: to delete the table without condition

```
/*to add column*/
alter table batch_11.anggota_elizadh_5
add hobby varchar (255);

/*to delete column*/
alter table batch_11.anggota_elizadh_5
drop hobby;

/*to delete row with condition*/
delete from batch_11.anggota_elizadh_5
where id = 1;

/*to delete all row*/
truncate batch_11.anggota_elizadh_5 ;

/*to delete table*/
drop batch_11.anggota_elizadh_5 ;
```



Intermediate SQL

```
select *
from datasource.order_details_csv;
```

	order_id	sales	profit	quantity	category	sub_category	cost	total_cost	total_profit
1	AZ-2011-1029887	85	15	2	Furniture	Furnishings	70	70	30
2	AZ-2011-1029887	26	7	2	Office Supplies	Labels	19	19	14
3	AZ-2011-107716	294	109	7	Technology	Accessories	185	185	763
4	AZ-2011-1087704	76	11	3	Furniture	Furnishings	65	65	33
5	AZ-2011-1087704	252	15	5	Office Supplies	Binders	237	237	75
6	AZ-2011-1087704	90	17	3	Office Supplies	Supplies	73	73	51
7	AZ-2011-1114253	1,334	200	8	Technology	Phones	1,134	1,134	1,600
8	AZ-2011-1116129	32	6	3	Office Supplies	Labels	26	26	18
9	AZ-2011-1137571	88	3	2	Furniture	Furnishings	85	85	6
10	AZ-2011-1137571	284	43	5	Office Supplies	Art	241	241	215
11	AZ-2011-1174243	541	156	4	Furniture	Bookcases	385	385	624
12	AZ-2011-122598	576	51	5	Office Supplies	Storage	525	525	255
13	AZ-2011-1229073	307	99	5	Furniture	Chairs	208	208	495
14	AZ-2011-1229073	44	14	3	Office Supplies	Binders	30	30	42
15	AZ-2011-1229073	96	21	4	Office Supplies	Supplies	75	75	84
16	AZ-2011-1240916	152	44	2	Office Supplies	Appliances	108	108	88
17	AZ-2011-1240916	957	316	12	Technology	Phones	641	641	3,792
18	AZ-2011-1253407	87	-78	3	Furniture	Chairs	165	165	-234
19	AZ-2011-1260928	33	10	3	Office Supplies	Binders	23	23	30
20	AZ-2011-1278696	28	10	2	Office Supplies	Fasteners	18	18	20
21	AZ-2011-1279238	25	-11	4	Office Supplies	Art	36	36	-44
22	AZ-2011-1279238	82	-74	3	Office Supplies	Storage	156	156	-222
23	AZ-2011-130330	38	11	2	Furniture	Furnishings	27	27	22
24	AZ-2011-1315772	13	3	2	Office Supplies	Binders	10	10	6
25	AZ-2011-1315772	748	283	4	Office Supplies	Storage	465	465	1,132
26	AZ-2011-1315772	800	168	3	Technology	Machines	632	632	504
27	AZ-2011-1315772	1,908	820	3	Technology	Phones	1,088	1,088	2,460

```
select category
from datasource.order_details_csv;
```

	category
1	Furniture
2	Office Supplies
3	Technology
4	Furniture
5	Office Supplies
6	Office Supplies
7	Technology
8	Office Supplies
9	Furniture
10	Office Supplies
11	Furniture
12	Office Supplies
13	Furniture
14	Office Supplies
15	Office Supplies
16	Office Supplies
17	Technology
18	Furniture
19	Office Supplies
20	Office Supplies
21	Office Supplies
22	Office Supplies
23	Furniture
24	Office Supplies
25	Office Supplies
26	Technology
27	Technology

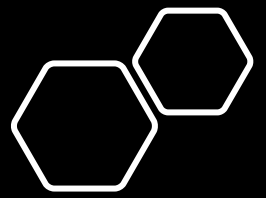
```
select distinct category
from datasource.order_details_csv;
```

	category
1	Furniture
2	Office Supplies
3	Technology

SELECT and DISTINCT

Select: to pick column(s). Use (*) to pick entire column

distinct: to remove duplicate in one column



CASE WHEN: returns value when conditions are met

```
select distinct
category,
        case
        when lower(category) = 'furniture' then 'Meubel’
        when lower(category) = 'office supplies' then
'Stationery’
        else 'Tech’
        end as category_renamed
from datasource.order_details_csv
;
```

category	category_renamed
Office Supplies	Stationery
Technology	Tech
Furniture	Meubel

DigitalSkola

WHERE: to set limitation to the query to filter the data based on needs

```sql
select
    order_id,
    category,
    sub_category,
    quantity,
    sales,
    profit
from datasource.order_details_csv
where
    sub_category in ('Bookcases','Chairs','Paper')
    and
    quantity  <= 5
    ;
```

```sql
select
    order_id,
    category,
    sub_category,
    quantity,
    sales,
    profit
from datasource.order_details_csv
where
    sub_category in ('Bookcases','Chairs','Paper')
    ;
```

GROUP BY: to summarise value and group it by specific criteria. Only can be used with AGGREGATE function

```
select
        category,
        sub_category,
        sum(quantity) as
total_quantity,
        sum(sales) as total_sales,
        sum(profit)as total_profit
from datasource.order_details_csv
where
        sub_category = 'Tables'
group by category,sub_category
;
```

category	sub_category	total_quantity	total_sales	total_profit
Furniture	Tables	287	89,478	-20,731

DigitalSkola

HAVING: to set limitation to the query to filter the data based on needs using aggregated columns

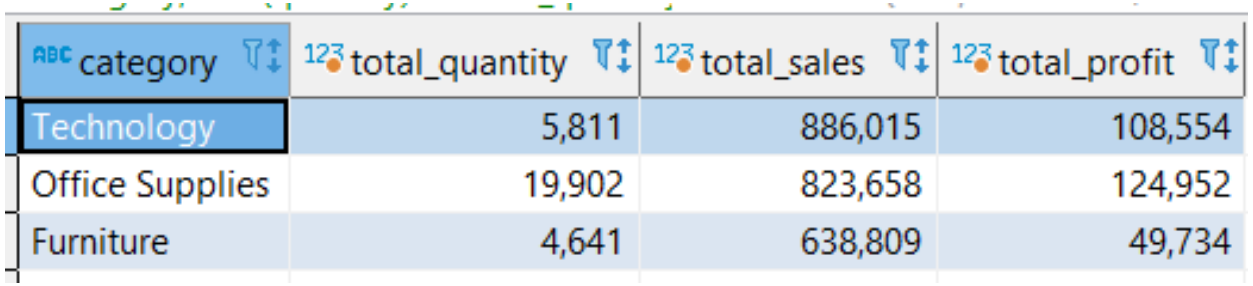
```
select
        category,
        sub_category,
        sum(quantity) as sum_quantity,
        sum(sales) as sum_sales,
        sum(profit)as sum_profit
from datasource.order_details_csv
where
        quantity >= 2
        and quantity <=3
        and sales >= 30
        and sales <= 30
group by category,sub_category
having sum(quantity) >=20
;
```

	category	sub_category	sum_quantity	sum_sales	sum_profit
1	Office Supplies	Art	21	240	62
2	Office Supplies	Binders	25	330	29
3	Office Supplies	Storage	22	240	-18

DigitalSkola

ORDER BY: to order the data when generating query result. Default is ascending

```
select
        category,
        sum(quantity) as total_quantity,
        sum(sales) as total_sales,
        sum(profit)as total_profit
from datasource.order_details_csv
group by category,sub_category
order by category desc
;
```



category	total_quantity	total_sales	total_profit
Technology	5,811	886,015	108,554
Office Supplies	19,902	823,658	124,952
Furniture	4,641	638,809	49,734

DigitalSkola

LIMIT: is used to make the query faster by limiting the total data generated from a query

```
select *
from datasource.order_details_csv
limit 10
;
```

	order_id	sales	profit	quantity	category	sub_category	cost	total_cost	total_profit
1	AZ-2011-1029887	85	15	2	Furniture	Furnishings	70	70	30
2	AZ-2011-1029887	26	7	2	Office Supplies	Labels	19	19	14
3	AZ-2011-107716	294	109	7	Technology	Accessories	185	185	763
4	AZ-2011-1087704	76	11	3	Furniture	Furnishings	65	65	33
5	AZ-2011-1087704	252	15	5	Office Supplies	Binders	237	237	75
6	AZ-2011-1087704	90	17	3	Office Supplies	Supplies	73	73	51
7	AZ-2011-1114253	1,334	200	8	Technology	Phones	1,134	1,134	1,600
8	AZ-2011-1116129	32	6	3	Office Supplies	Labels	26	26	18
9	AZ-2011-1137571	88	3	2	Furniture	Furnishings	85	85	6
10	AZ-2011-1137571	284	43	5	Office Supplies	Art	241	241	215

STRING FUNCTIONS

String Concatenation

```
select
order_id,
'US-'|| order_id as new_order_id
from datasource.order_details_csv
limit 10
;
```

	order_id	new_order_id
1	AZ-2011-1029887	US-AZ-2011-1029887
2	AZ-2011-1029887	US-AZ-2011-1029887
3	AZ-2011-107716	US-AZ-2011-107716
4	AZ-2011-1087704	US-AZ-2011-1087704
5	AZ-2011-1087704	US-AZ-2011-1087704
6	AZ-2011-1087704	US-AZ-2011-1087704
7	AZ-2011-1114253	US-AZ-2011-1114253
8	AZ-2011-1116129	US-AZ-2011-1116129
9	AZ-2011-1137571	US-AZ-2011-1137571
10	AZ-2011-1137571	US-AZ-2011-1137571

Number of characters in string

```
select distinct char_length ('US-
'|| order_id) as tot_karakter;
```

tot_karakter
15
18
17
16

DigitalSkola

STRING FUNCTIONS

Lower character

```
select
order_id,
lower('US-'|| order_id) as
new_order_id
from datasource.order_details_csv
limit 5
;
```

	order_id	new_order_id
1	AZ-2011-1029887	us-az-2011-1029887
2	AZ-2011-1029887	us-az-2011-1029887
3	AZ-2011-107716	us-az-2011-107716
4	AZ-2011-1087704	us-az-2011-1087704
5	AZ-2011-1087704	us-az-2011-1087704

Uppercase character

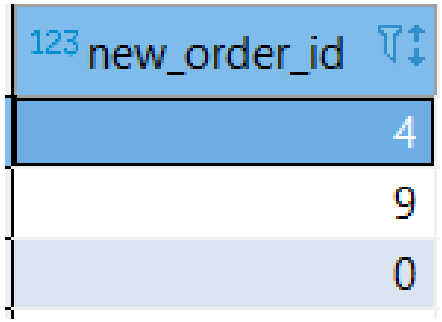
```
select
order_id,
upper ('US-'|| order_id) as
new_order_id
from datasource.order_details_csv
limit 5
;
```

	order_id	new_order_id
1	AZ-2011-1029887	US-AZ-2011-1029887
2	AZ-2011-1029887	US-AZ-2011-1029887
3	AZ-2011-107716	US-AZ-2011-107716
4	AZ-2011-1087704	US-AZ-2011-1087704
5	AZ-2011-1087704	US-AZ-2011-1087704

STRING FUNCTIONS: Index and Substring

To know position of certain character

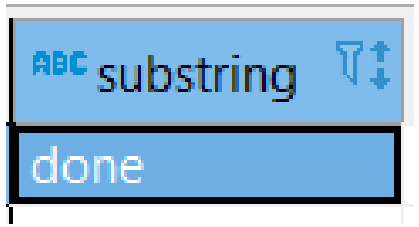
```
select distinct position ('2011'
in (order_id)) as new_order_id

from
datasource.order_details_csv

;
```



123 new_order_id
4
9
0

To return certain character in a string

```
select substring('Indonesia'
from 3 for 4) ;
```



ABC substring
done

AGGREGATE Function: to summarise values

```sql
select avg(total_profit)
from datasource.order_details_csv
;

select sum(total_profit)
from datasource.order_details_csv
;

select
    max(total_profit),
    min(total_profit),
    avg(total_profit)
from datasource.order_details_csv
;
```

```sql
-- how many row in a table
select count(*)
from datasource.order_details_csv
;

-- return same value with count(*) because no null row
select count(order_id)
from datasource.order_details_csv
;

--return total of row
select count(sub_category)
from datasource.order_details_csv
;

--return total distinct value of row
select count(distinct sub_category)
from datasource.order_details_csv
;
```