

Course Introduction

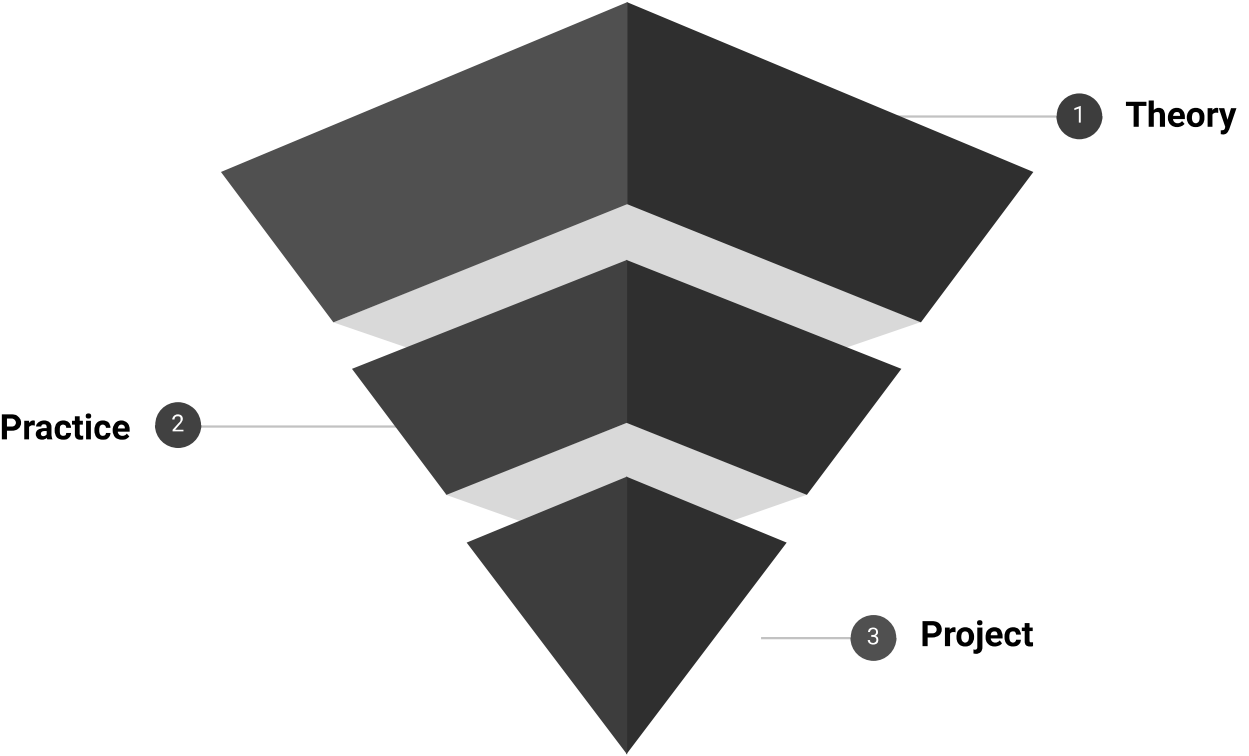


OBJECTIVES

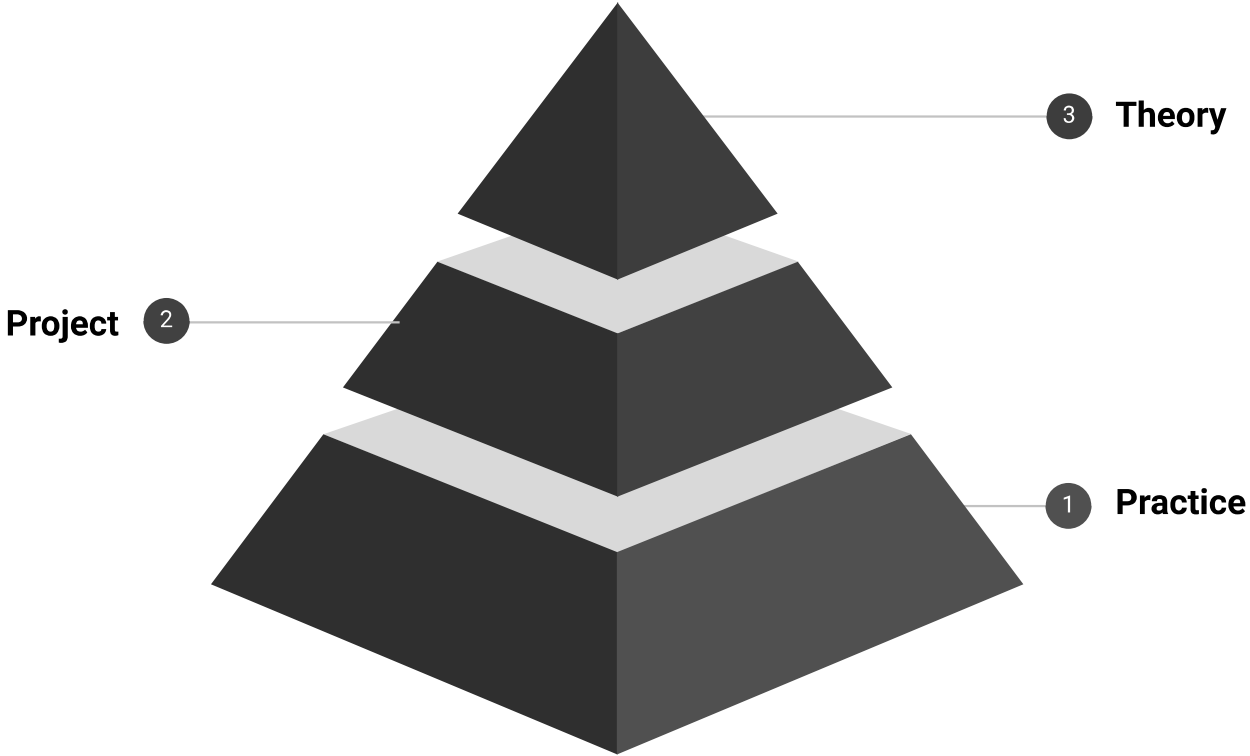
- Data-maturity model
- dbt and data architectures
- Data warehouses, data lakes, and lakehouses
- ETL and ELT procedures
- dbt fundamentals
- Analytics Engineering



TOP-DOWN



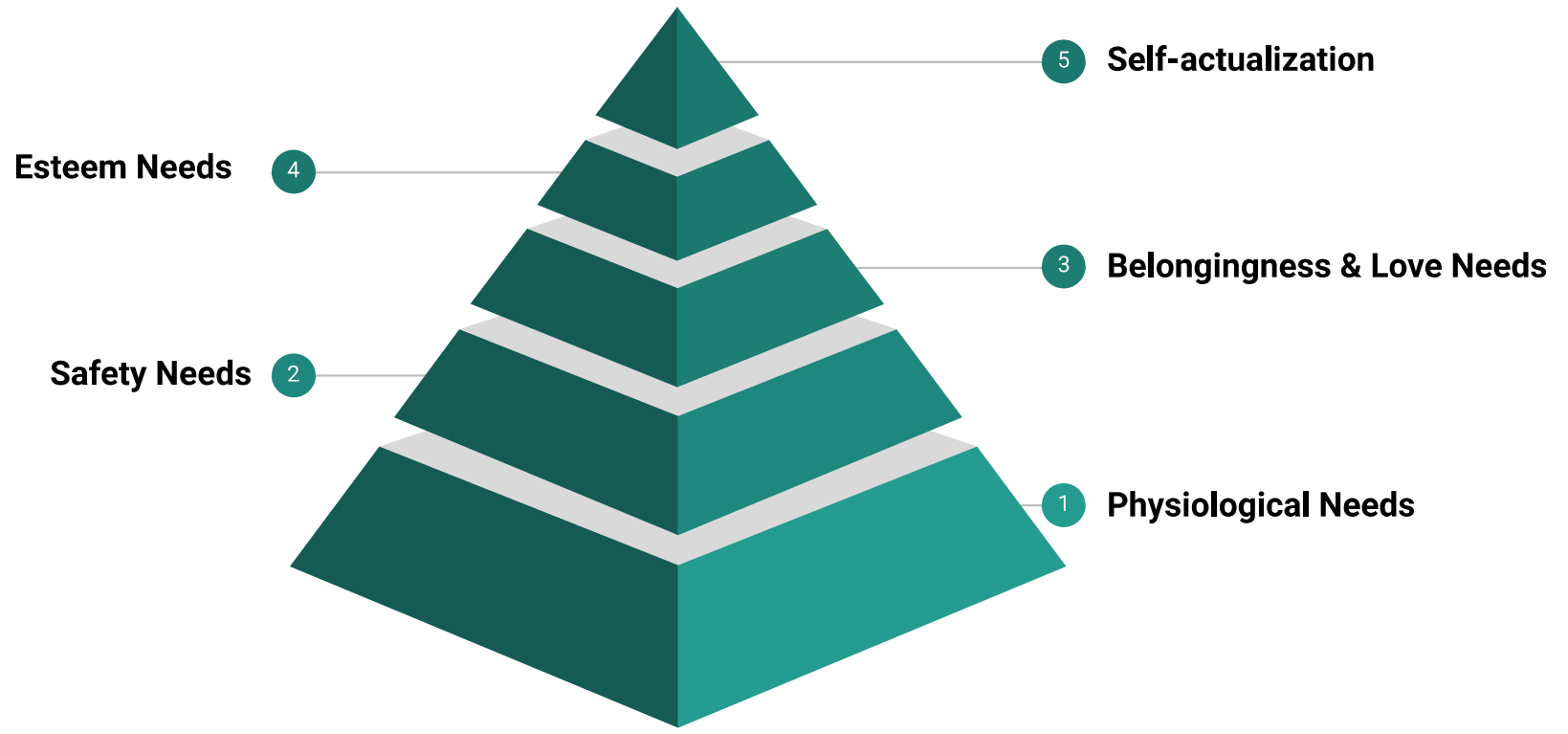
BOTTOM-UP



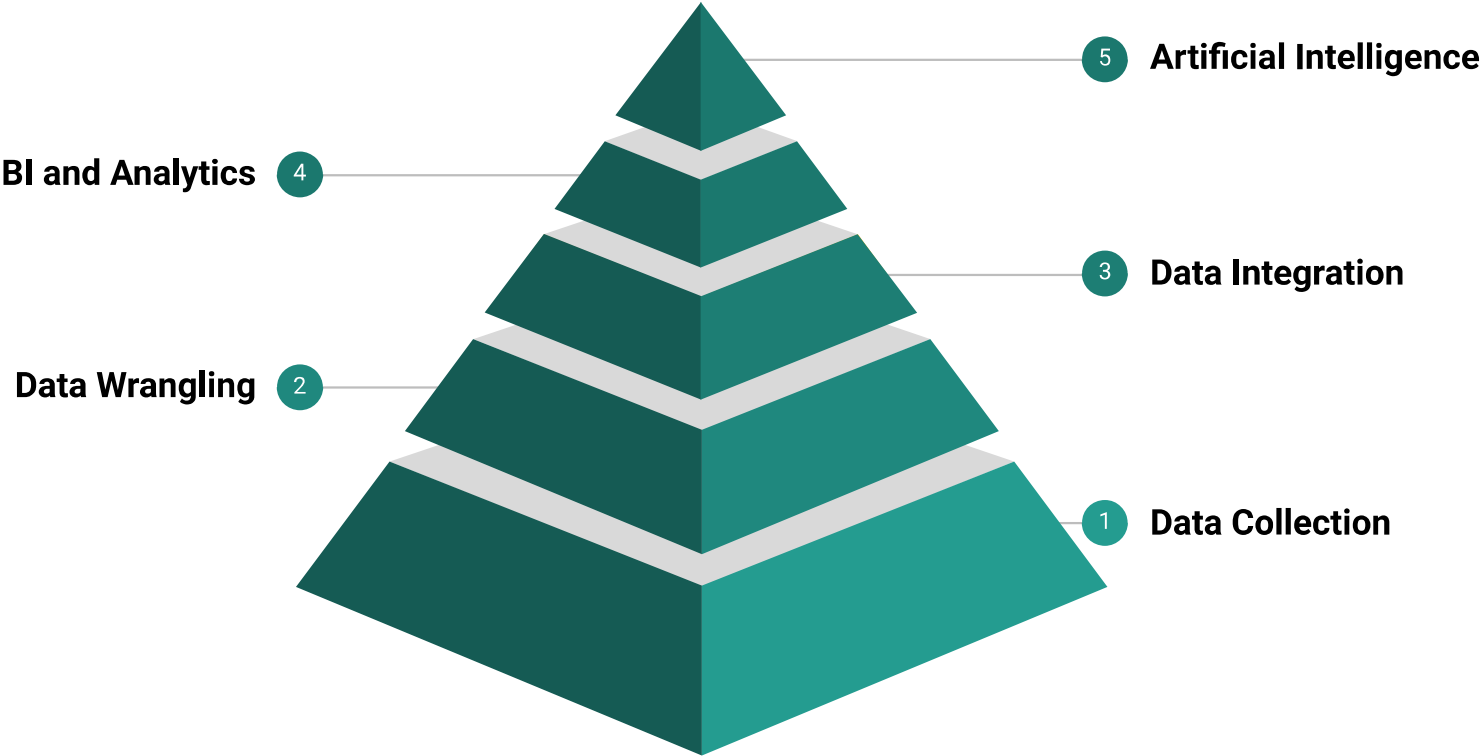
Data Maturity Model



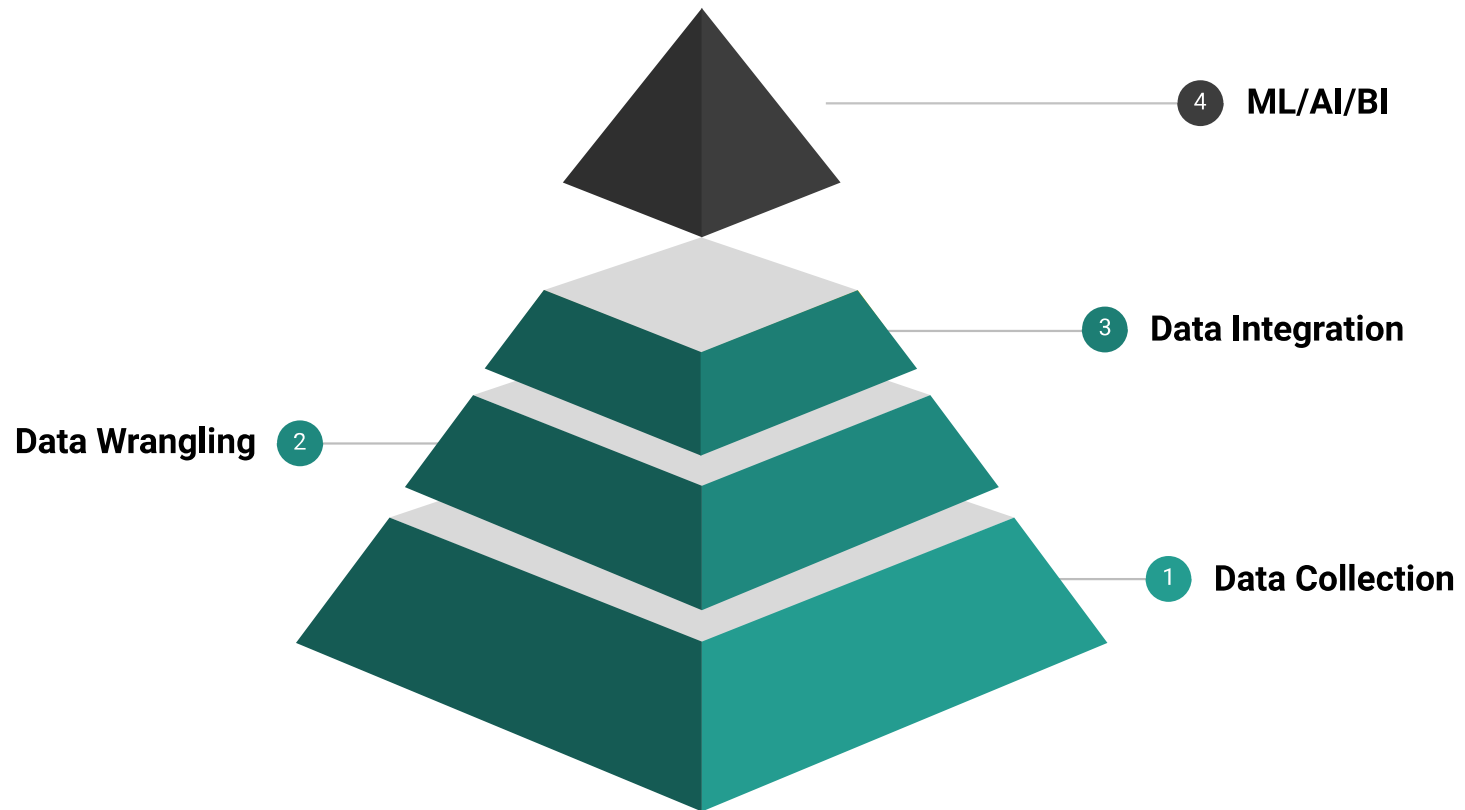
Maslow's Hierarchy of Needs



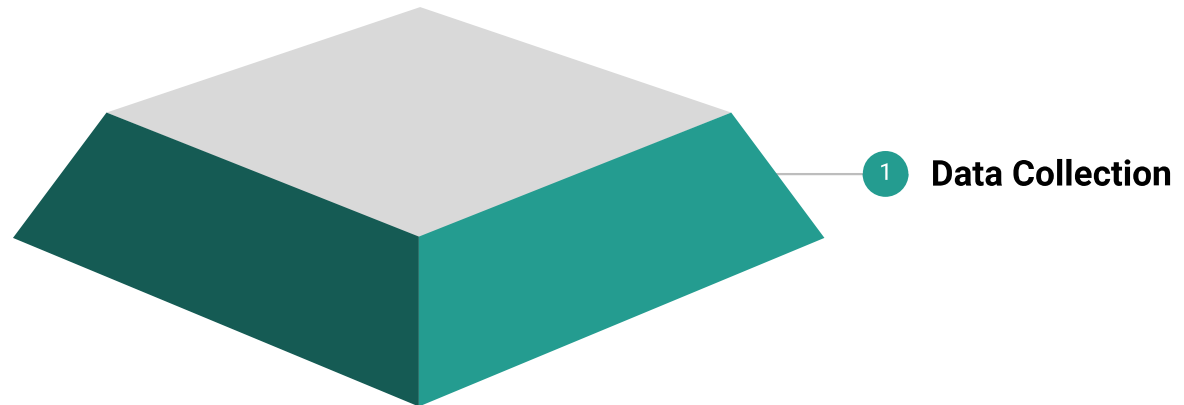
Data-Maturity Model



Typical Data Architecture



Data Collection



Data Wrangling



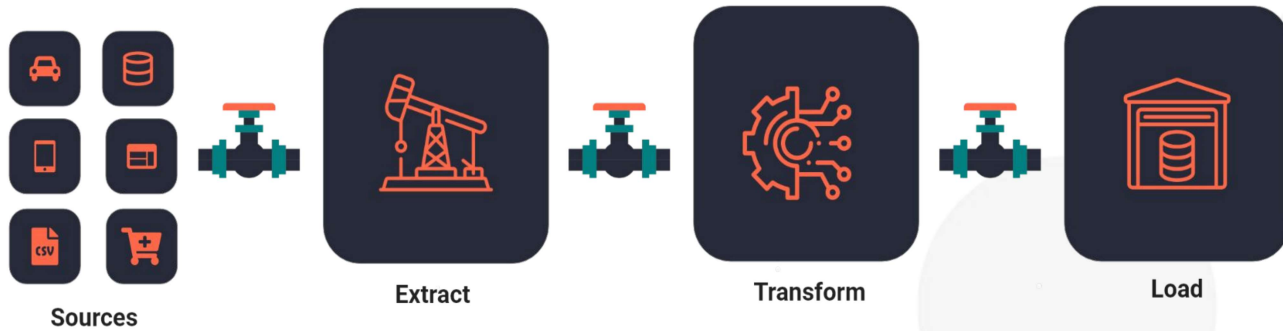
Data Integration



ETL - ELT #

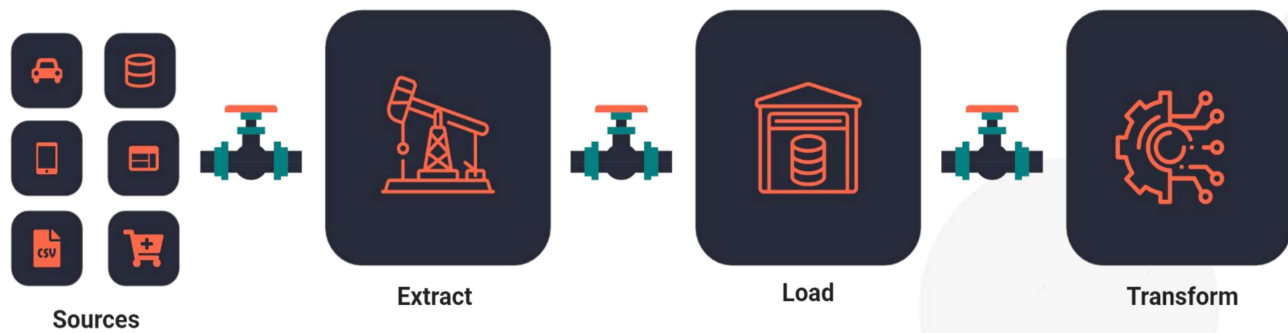
ETL

ETL
Extract, Transform, Load

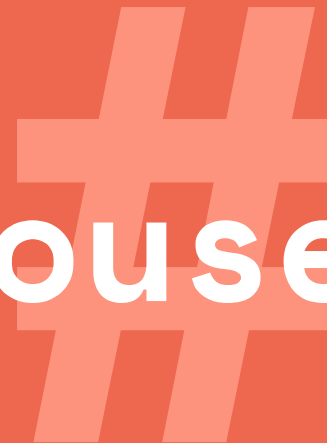


ELT

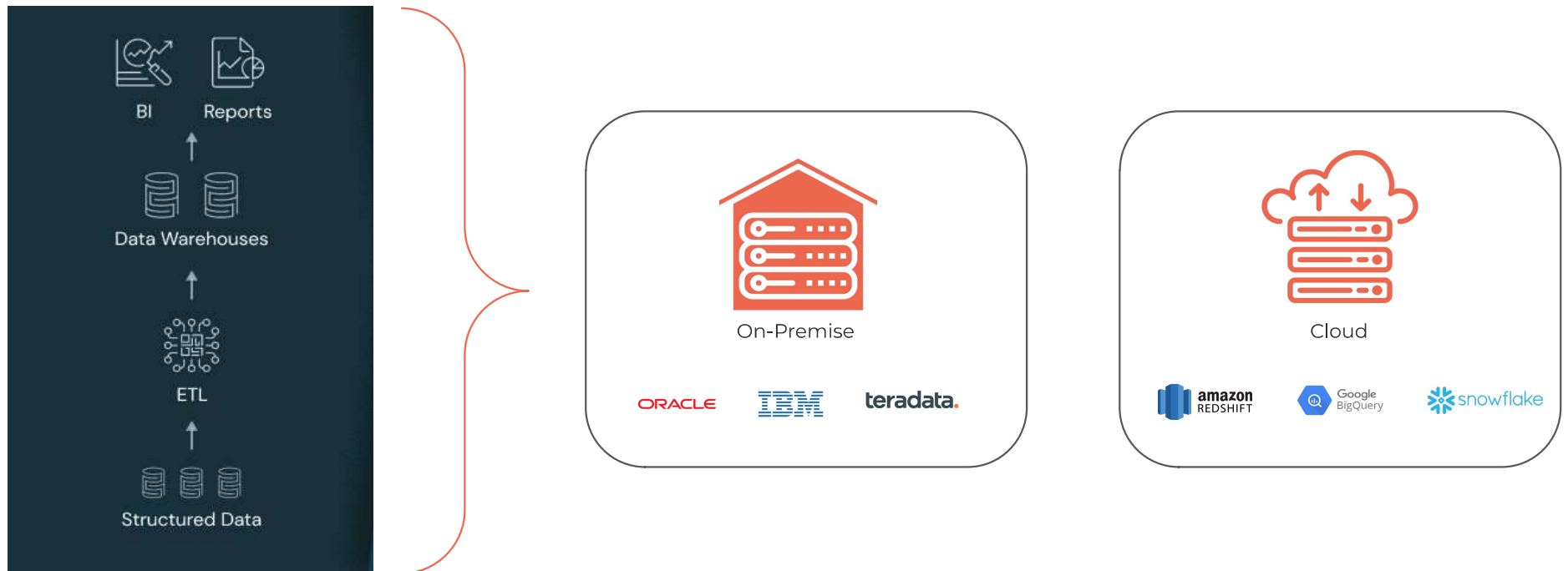
ELT
Extract, Load, Transform



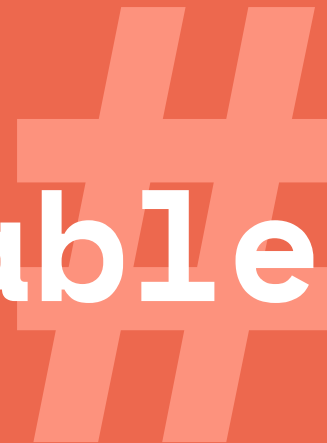
Data Warehouse



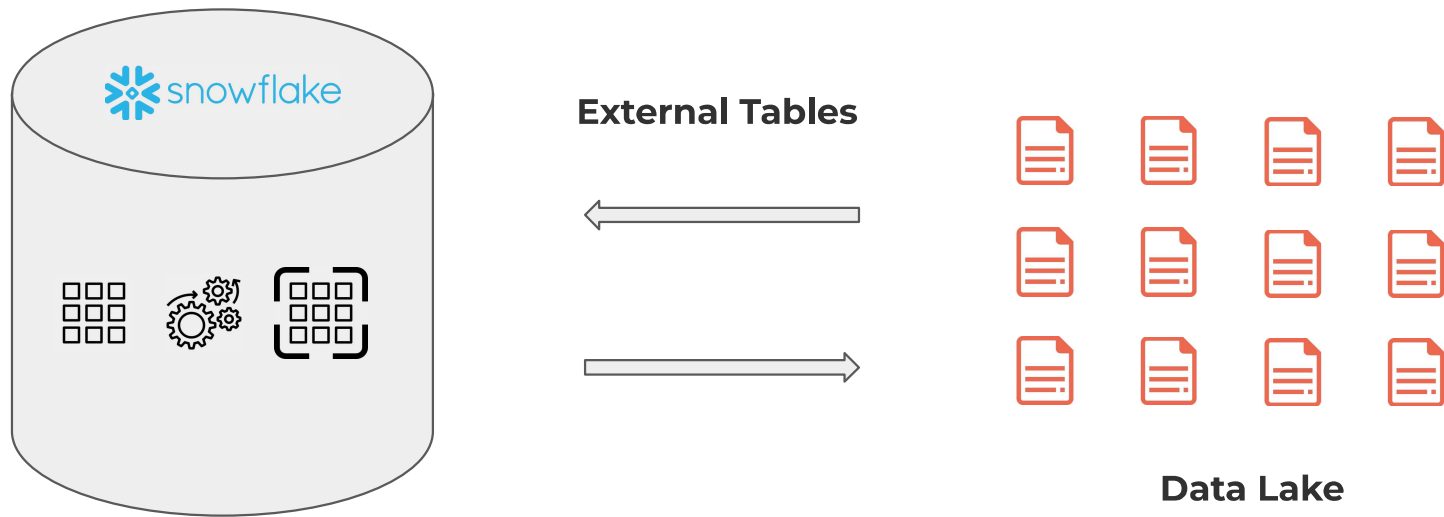
Data Warehouse



External Tables



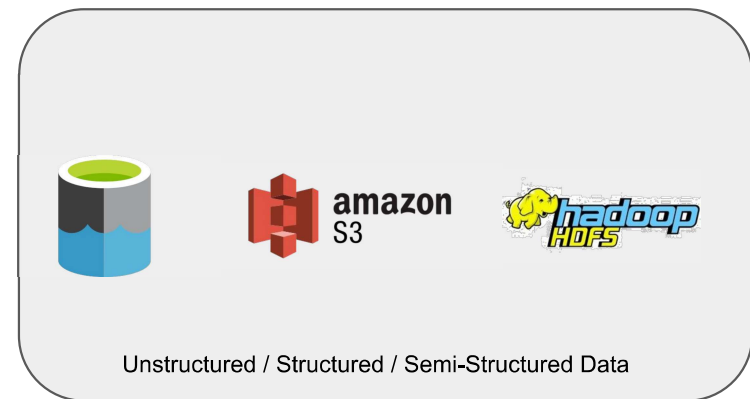
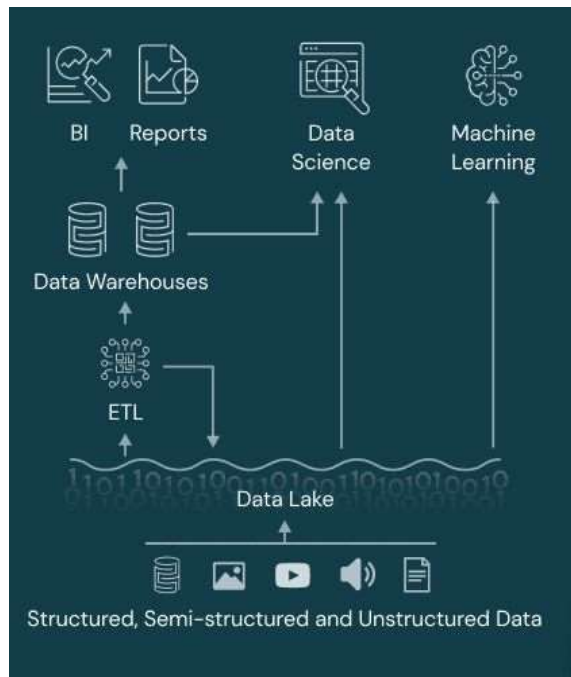
External Tables



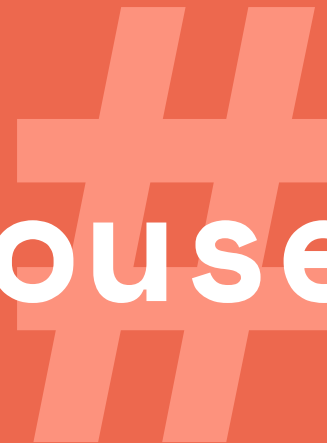
Data Lake



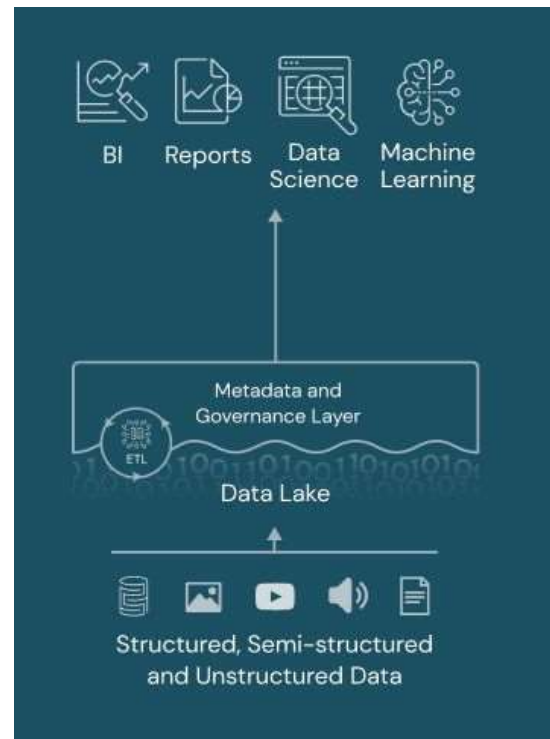
Data Lake



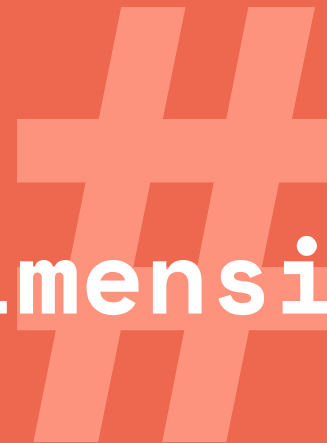
Data Lakehouse



Data Lakehouse



Slowly Changing Dimensions



SCD Type 0

Not updating the DWH table when a Dimension changes

Source			DWH		
Host ID	Name	Fax Number	Host ID	Name	Fax Number
#1	John Doe	1-408-999 8888	#1	John Doe	555-123-4567

SCD Type 1

Updating the DWH table when a Dimension changes, overwriting the original data

	Source	DWH								
No Air-conditioning	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>No</td></tr></table>	Host ID	Air-conditioning	#1	No	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>No</td></tr></table>	Host ID	Air-conditioning	#1	No
Host ID	Air-conditioning									
#1	No									
Host ID	Air-conditioning									
#1	No									
Installed Air-conditioning	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>Yes</td></tr></table>	Host ID	Air-conditioning	#1	Yes	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>No</td></tr></table>	Host ID	Air-conditioning	#1	No
Host ID	Air-conditioning									
#1	Yes									
Host ID	Air-conditioning									
#1	No									
DWH updated	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>Yes</td></tr></table>	Host ID	Air-conditioning	#1	Yes	<table><tr><th>Host ID</th><th>Air-conditioning</th></tr><tr><td>#1</td><td>Yes</td></tr></table>	Host ID	Air-conditioning	#1	Yes
Host ID	Air-conditioning									
#1	Yes									
Host ID	Air-conditioning									
#1	Yes									

SCD Type 2

Keeping full history - Adding additional (historic data) rows for each dimension change

Current rental price (\$300)

Apartment ID	Price
#1	\$300

Change in the rental price (\$450)

Apartment ID	Price
#1	\$450

DWH updated

Source	DWH																			
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$300</td></tr></table>	Apartment ID	Price	#1	\$300	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>NULL</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	1000	#1	\$300	2020-01-01T00:00:00	NULL					
Apartment ID	Price																			
#1	\$300																			
Host Key	Apartment ID	Price	Start_Date	End_Date																
1000	#1	\$300	2020-01-01T00:00:00	NULL																
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$450</td></tr></table>	Apartment ID	Price	#1	\$450	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>2021-01-01T00:00:00</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00					
Apartment ID	Price																			
#1	\$450																			
Host Key	Apartment ID	Price	Start_Date	End_Date																
1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00																
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$450</td></tr></table>	Apartment ID	Price	#1	\$450	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>2021-01-01T00:00:00</td></tr><tr><td>1001</td><td>#1</td><td>\$450</td><td>2021-01-01T00:00:00</td><td>NULL</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00	1001	#1	\$450	2021-01-01T00:00:00	NULL
Apartment ID	Price																			
#1	\$450																			
Host Key	Apartment ID	Price	Start_Date	End_Date																
1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00																
1001	#1	\$450	2021-01-01T00:00:00	NULL																

SCD Type 2

Keeping full history - Adding additional (historic data) rows for each dimension change

Current rental price (\$300)

Apartment ID	Price
#1	\$300

Change in the rental price (\$450)

Apartment ID	Price
#1	\$450

DWH updated

Source	DWH																						
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$300</td></tr></table>	Apartment ID	Price	#1	\$300	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th><th>Is_Current</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>9999-12-12T00:00:00</td><td>Y</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current	1000	#1	\$300	2020-01-01T00:00:00	9999-12-12T00:00:00	Y						
Apartment ID	Price																						
#1	\$300																						
Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current																		
1000	#1	\$300	2020-01-01T00:00:00	9999-12-12T00:00:00	Y																		
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$450</td></tr></table>	Apartment ID	Price	#1	\$450	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th><th>Is_Current</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>9999-12-12T00:00:00</td><td>N</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current	1000	#1	\$300	2020-01-01T00:00:00	9999-12-12T00:00:00	N						
Apartment ID	Price																						
#1	\$450																						
Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current																		
1000	#1	\$300	2020-01-01T00:00:00	9999-12-12T00:00:00	N																		
<table><tr><th>Apartment ID</th><th>Price</th></tr><tr><td>#1</td><td>\$450</td></tr></table>	Apartment ID	Price	#1	\$450	<table><tr><th>Host Key</th><th>Apartment ID</th><th>Price</th><th>Start_Date</th><th>End_Date</th><th>Is_Current</th></tr><tr><td>1000</td><td>#1</td><td>\$300</td><td>2020-01-01T00:00:00</td><td>2021-01-01T00:00:00</td><td>N</td></tr><tr><td>1001</td><td>#1</td><td>\$450</td><td>2021-01-01T00:00:00</td><td>9999-12-12T00:00:00</td><td>Y</td></tr></table>	Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current	1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00	N	1001	#1	\$450	2021-01-01T00:00:00	9999-12-12T00:00:00	Y
Apartment ID	Price																						
#1	\$450																						
Host Key	Apartment ID	Price	Start_Date	End_Date	Is_Current																		
1000	#1	\$300	2020-01-01T00:00:00	2021-01-01T00:00:00	N																		
1001	#1	\$450	2021-01-01T00:00:00	9999-12-12T00:00:00	Y																		

SCD Type 3

Keeping limited data history - adding separate columns for original and current value

Listed as Private

Host changed Private to Entire

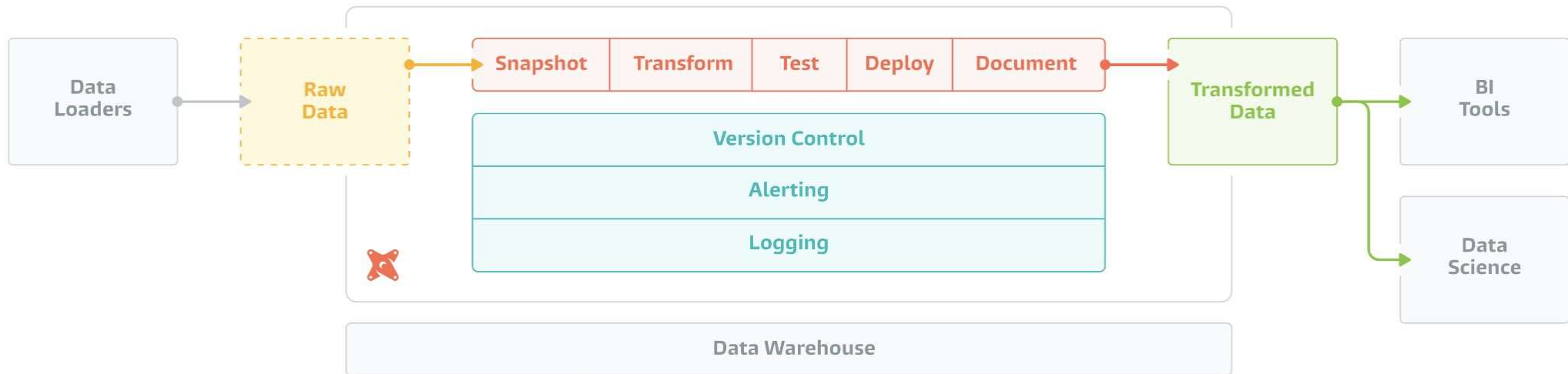
Host changed Entire to Shared

Source	DWH										
<table><tr><th>Apartment ID</th><th>Type</th></tr><tr><td>#1</td><td>Private</td></tr></table>	Apartment ID	Type	#1	Private	<table><tr><th>Apartment ID</th><th>Previous Type</th><th>Current Type</th></tr><tr><td>#1</td><td>Private</td><td>Private</td></tr></table>	Apartment ID	Previous Type	Current Type	#1	Private	Private
Apartment ID	Type										
#1	Private										
Apartment ID	Previous Type	Current Type									
#1	Private	Private									
<table><tr><th>Apartment ID</th><th>Type</th></tr><tr><td>#1</td><td>Entire</td></tr></table>	Apartment ID	Type	#1	Entire	<table><tr><th>Apartment ID</th><th>Previous Type</th><th>Current Type</th></tr><tr><td>#1</td><td>Private</td><td>Entire</td></tr></table>	Apartment ID	Previous Type	Current Type	#1	Private	Entire
Apartment ID	Type										
#1	Entire										
Apartment ID	Previous Type	Current Type									
#1	Private	Entire									
<table><tr><th>Apartment ID</th><th>Type</th></tr><tr><td>#1</td><td>Shared</td></tr></table>	Apartment ID	Type	#1	Shared	<table><tr><th>Apartment ID</th><th>Previous Type</th><th>Current Type</th></tr><tr><td>#1</td><td>Entire</td><td>Shared</td></tr></table>	Apartment ID	Previous Type	Current Type	#1	Entire	Shared
Apartment ID	Type										
#1	Shared										
Apartment ID	Previous Type	Current Type									
#1	Entire	Shared									

dbt Overview



dbt Overview



Analytics Engineering




Common Table Expression (CTE)



CTE

Syntax

```
SQL   
  
WITH <name_of_the_result_set> ([column_names])  
AS  
(  
    <cte_query>  
)  
<reference_the_CTE>
```