

CREATE VIEW GS.CASTROLE(CASTMEMBER CODE, ROlename) AS

SELECT

FROM DS1.CASTROLE AS CR, DS1.CAST AS C, DS1.SHOW AS S

WHERE

UNION

SELECT KEYGENCASTMEMBER (NAME, 'DAZ'), 'STABLE-MEMBER'

FROM DS2.ANCHOR

UNION

SELECT KEYGENCASTMEMBER (NAME, 'DAZ'), 'GUEST'

FROM DS2.GUEST

DATA WAREHOUSE

DEFINITION: A DATA WAREHOUSE IS A DATA COLLECTION BUILT TO SUPPORT DECISION-MAKING PROCESS. IT IS USUALLY BUILT STARTING FROM SEVERAL DATA SOURCES AND WE UPDATE IT ONLY PERIODICALLY

| IN A COMPANY THERE ARE DIFFERENT DATA BASES. WE DISTINGUISH THEM IN:

- OPERATIONAL DB

8, 23

- DATA WAREHOUSES

DATA INTEGRATION



DEFINITION: A FACT IS A CONCEPT OF INTEREST FOR THE DECISION MAKING PROCESS. EXAMPLES: "SALES, SHIPPING, PURCHASES"

DEFINITION: A MEASURE IS A NUMERIC QUANTITY OF A FACT WHICH DESCRIBES A QUANTITATIVE ASPECT OF WHICH IS OF INTEREST FOR THE ANALYSIS

EXAMPLES: NUMBER OF UNIT SOLD, UNIT PRICE, TOTAL REVENUE, ...

DEFINITION: A DIMENSION IS A DISCRETE PROPERTY OF A FACT WHICH DESCRIBES A POSSIBLE PERSPECTIVE OF ANALYSIS

NOTE: TIME IS ALWAYS A DIMENSION

STEPS TO FOLLOW:

- 1) FROM LOGICAL TO CONCEPTUAL SCHEMA
- 2) CONCEPTUAL DESIGN OF THE DATA WAREHOUSE DRIVEN BY THE REQUESTED QUERY
 - IDENTIFY THE FACT
 - FOR EACH FACT, DEFINE:
 - ATTRIBUTE TREE
 - FACT SCHEMA
 - GLOSSARY
- 3) LOGICAL DESIGN OF THE DATA WAREHOUSE
- 4) QUERY

DEFINITION: A PRIMARY EVENT IS A PARTICULAR OCCUPANCY OF A FACT

IDENTIFIED BY SPECIFIC VALUE FOR EACH DIMENSION; MOREOVER, FOR EACH PRIMARY EVENT IS ASSOCIATED

SUMMARIZE:

- FACT → SPECIFY SOMETHING
- DIMENSION → DATA, NAME OF PRODUCT, NAME OF THE SHOP
- MEASURES → NUMBER OF UNIT SOLD AND TOTAL REVENUE

EXAMPLE: ON THE 10/10/2023 WE SOLD 10 UNITS OF THE ALBUM "THE GLOBE" IN THE MONDADORI STORE OF DUOMO FOR A REVENUE OF €200

DEFINITION: A DIMENSIONAL ATTRIBUTE CAN BE A DIMENSION OR ANOTHER TYPE OF ATTRIBUTE

HIERARCHY IS A SPECIFIED TREE WHOSE NODES ARE DIMENSIONAL

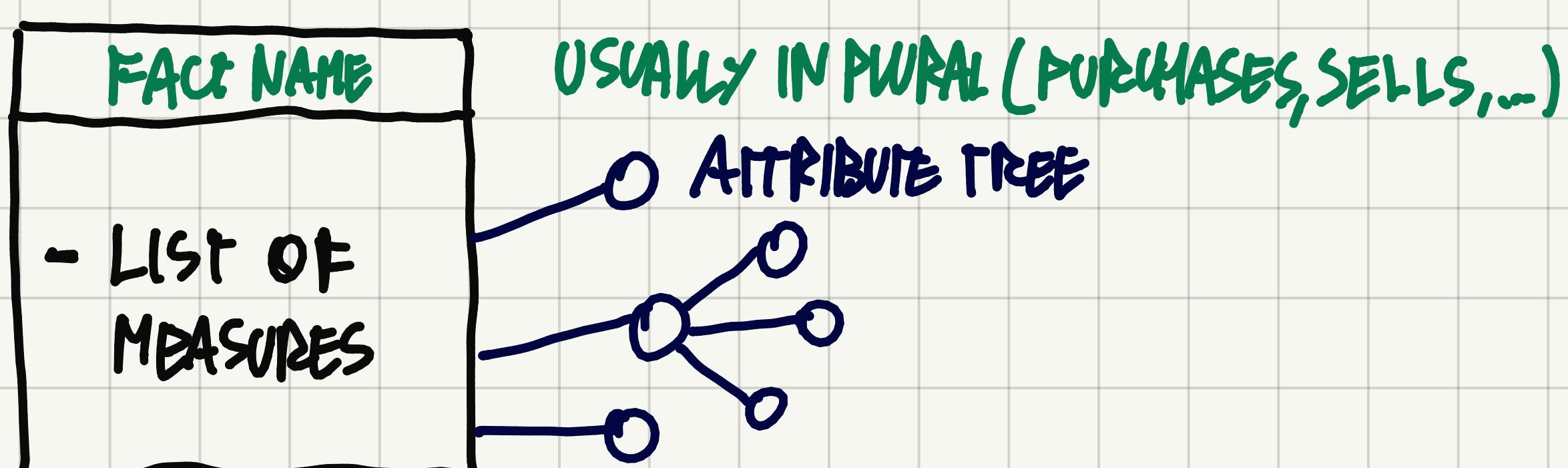
ATTRIBUTES AND EDGES REPRESENT 1:n RELATIONSHIP BETWEEN

COUPLES OF DIMENSIONAL ATTRIBUTES. THE ROOT OF THE TREE

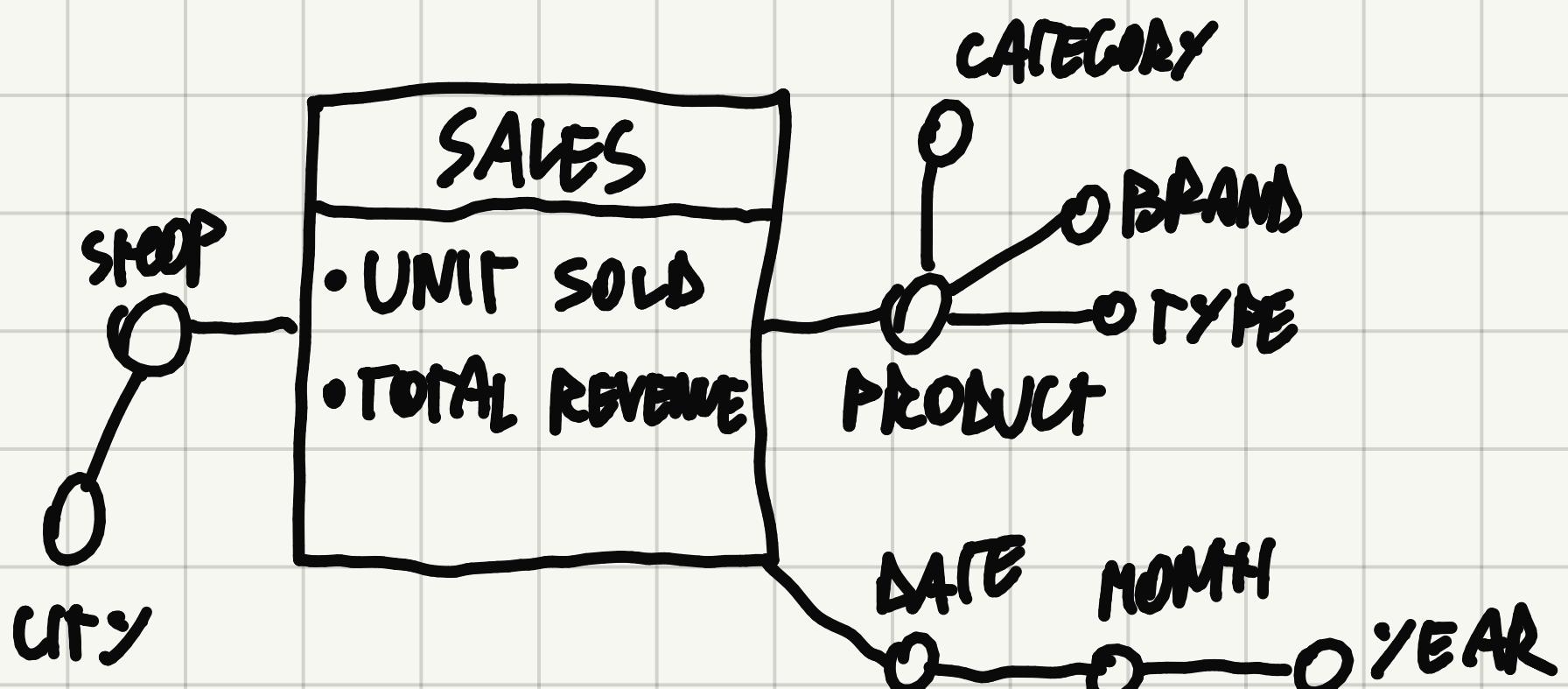
IS THE DIMENSION.

EXAMPLES: { DATE: MONTH, YEAR, ...
PRODUCT: TYPE, CATEGORY, ...
SHOP: CITY, STATE, ...}

FACT SCHEMA



EXAMPLE:



DEFINITION: GIVEN A SET OF DIMENSIONAL ATTRIBUTES (GENERALLY BELONGING TO DIFFERENT HIERARCHIES), SUCH SET IDENTIFIES A **SECONDARY EVENT** WHICH IS THE AGGREGATION OF ALL THE NODES BINDING PRIMARY ELEMENTS.

EXAMPLE:

	SHOP	DATE	PRODUCT	M ₁	M ₂
1		11/12/23	THE GLOBE		
2		12/12/23	THE GLOBE		
	⋮	⋮	⋮	⋮	⋮

A SINGLE ROW IS A PRIMARY EVENT

Pd F-

T MACUP

EXERCISE (Insurance company)

An insurance company requires the data warehouse design for accident analysis of its customers. In particular, the company requires to:

- Evaluate the history of the accidents with respect to the date, the motivation and the policies stipulated by the insurance company, considering the number of accidents and the total cost.
- Evaluate the history of the policies with respect to the class, the start and end dates, the customers of the insurance company and the policy type, considering the number of policies, the total amount and the average policy duration.

Relational schema of the operational database:

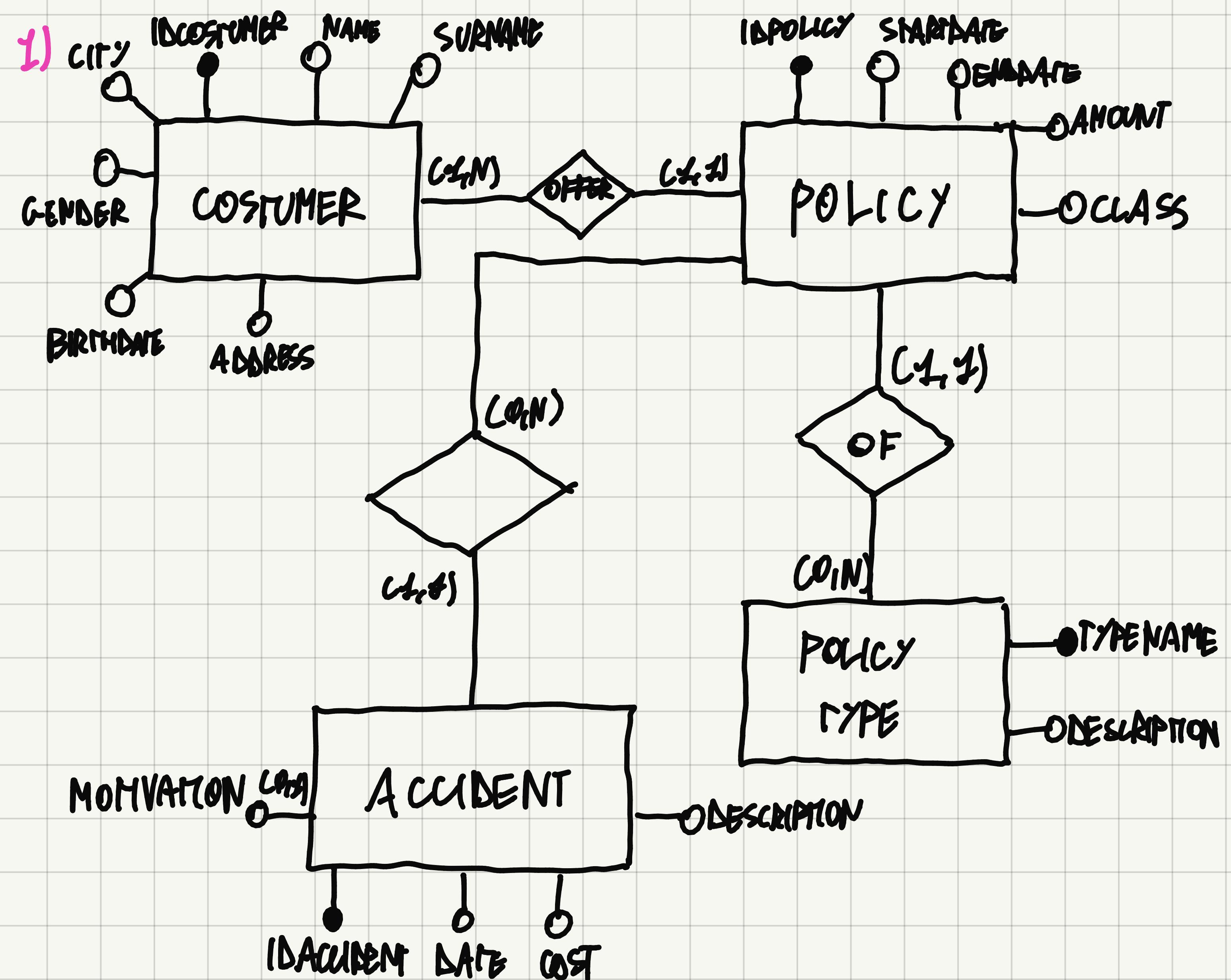
CUSTOMER (IdCustomer, Name, Surname, Address, City, Gender, BirthDate)

POLICY (IdPolicy, StartDate, EndDate, Amount, Class, Owner, TypeName) // *The class of the policy is an integer that depends on the past history of accidents associated with this policy. The type of the policy indicates the kind of accidents that are covered; e.g., the 'Comprehensive' policy type covers all the accidents.*

POLICYTYPE (TypeName, Description)

ACCIDENT (IdAccident, Date, Cost, Description, Motivation*, IdPolicy) // *Sample motivations: rain, drunk driving, speeding, ...*

1. Perform the reverse engineering of the given logical schema into a conceptual schema (Entity-Relationship model).
2. With respect to the produced ER diagram, discover the fact(s) that are useful specifically for answering the queries reported below. For each of these facts:
 - a. Produce the attribute tree (with pruning and grafting).
 - b. Produce the conceptual schema (fact schema).
 - c. Produce the glossary.
3. Produce a logical schema consistent with the conceptual schema.
4. Write in SQL the following queries against the designed logical schema:
 - a. Considering only the accidents happened on Sunday in 2015 and covered by policies of class '14' and type 'Comprehensive', compute the total number of accidents by policy, week and motivation.
 - b. Compute the total cost of the accidents for each customer (specify also name and surname) and date, considering only the accidents happened in March 2017 with customer born in 1975 and residing in Milan.
 - c. Considering only the policies starting in July 2015, compute the total amount of the policies grouped by class, type, start date, and home city and birth year of the customer.
 - d. Considering only the policies ending in February 2017, compute the average duration of the policies by customer (specify also name and surname) and end date.



2) FACT DEFINITION. A FACT CAN BE AN ENTITY, A WEAK ENTITY, AN n:n RELATIONSHIP

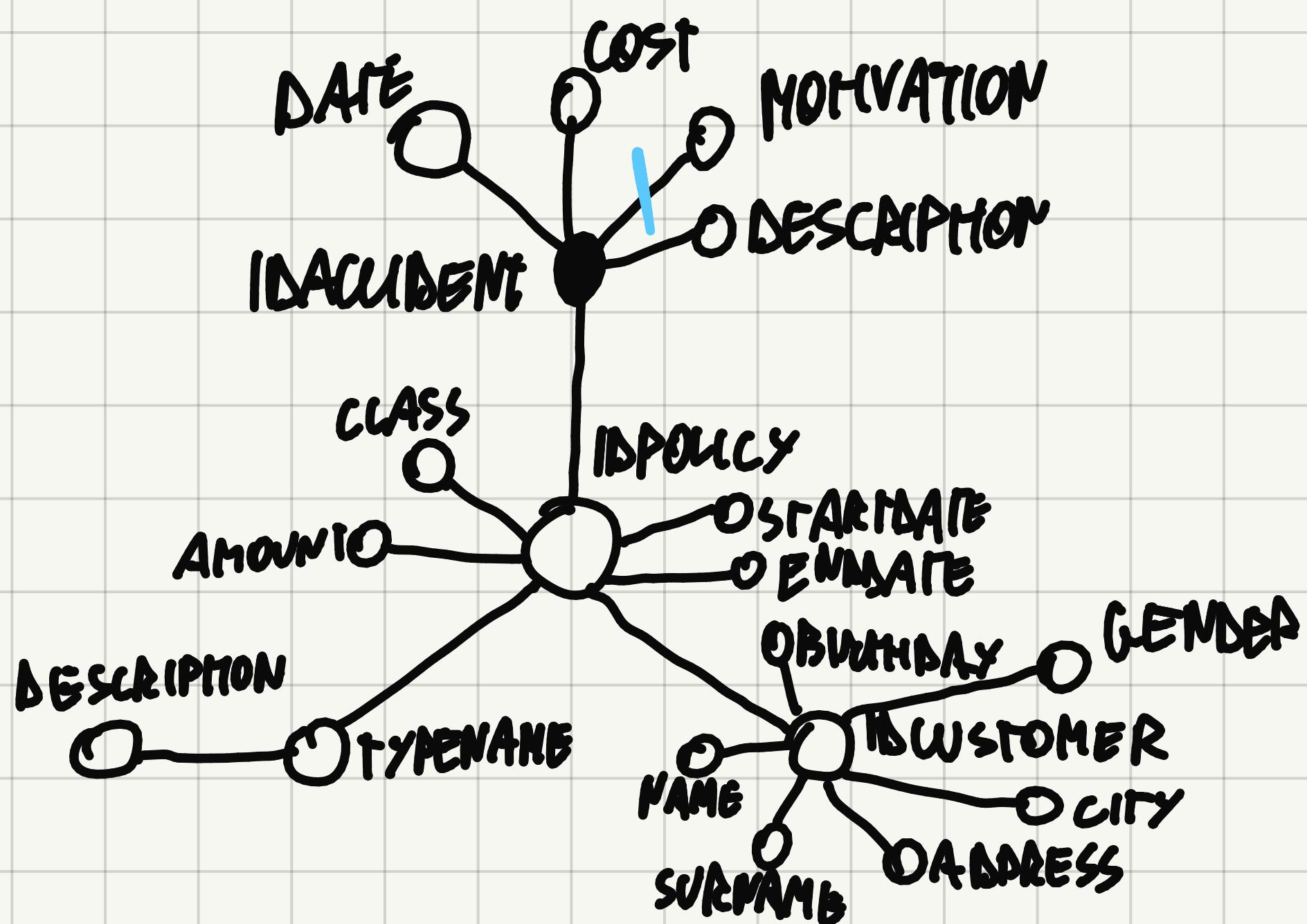
- ACCIDENTS / ACCIDENT ВМНГ
● POLICIES (РОЙСЫ ВМНГ)

ACCIDENTS

@ATTRIBUTE NAME. STEPS:

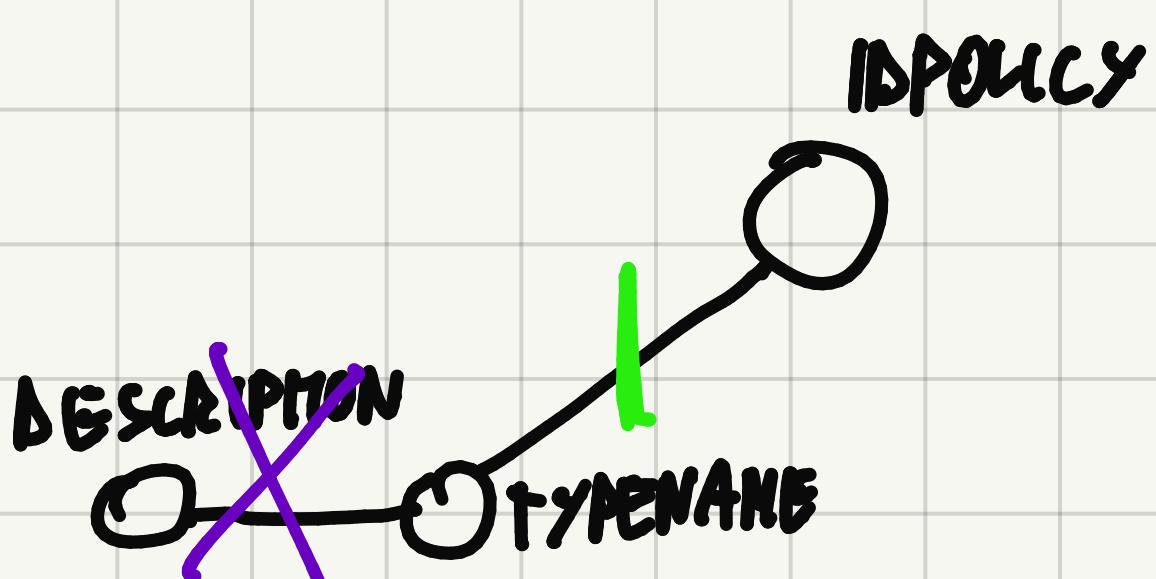
- DRAW A FULL DOT FOR THE PRIMARY KEY(S) OF THE QUERY'S FACT
 - DRAW AN EMPTY DOT FOR EVERY ATTRIBUTE OF THE FACT TABLE AND CONNECT IT TO THE FULL DOT
 - "NAVIGATE" THROUGH THE ENTRIES CONNECTED TO WHERE ARE PROVIDED ON THE "ONE SIDE" AS MAXIMUM CARDINALITY OF THE RELATIONSHIP

• OPTIONAL

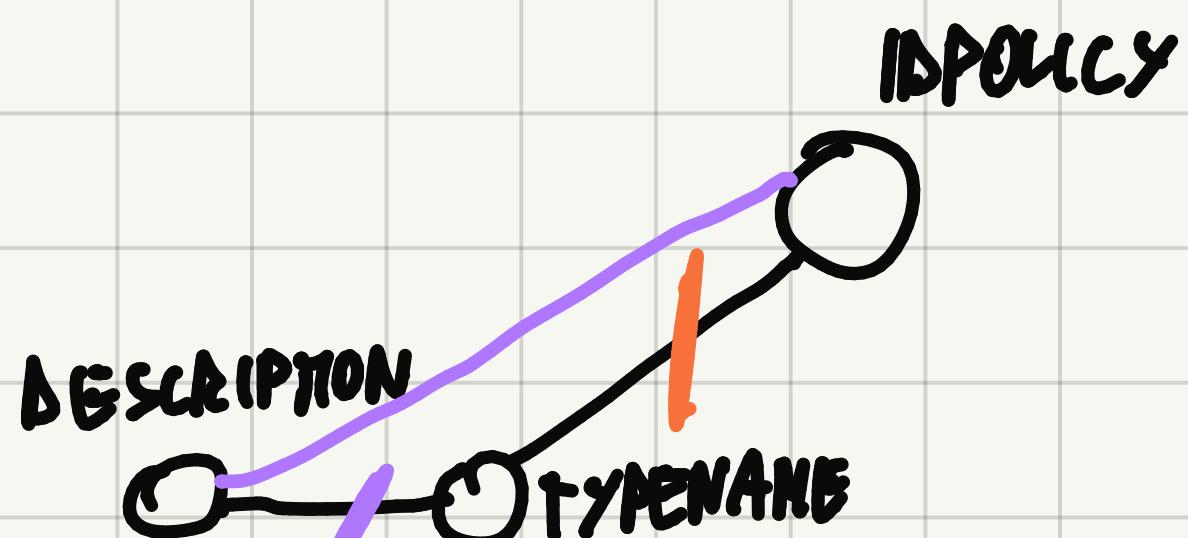


PERFORM OPERATIONS OF PRUNING OR GRAFTING:

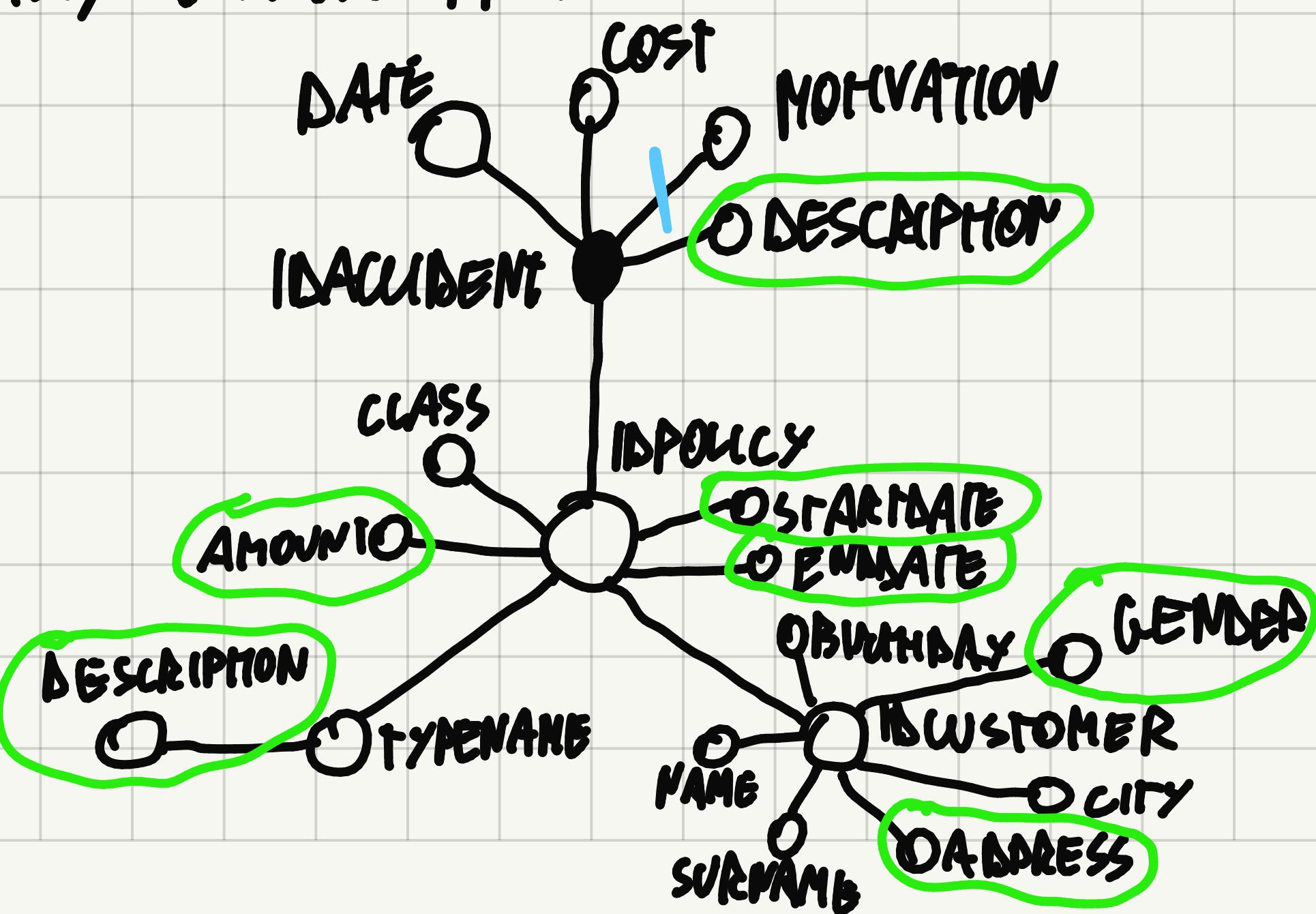
- DELETING A NODE, ALL ITS SUBNODES ARE DELETED

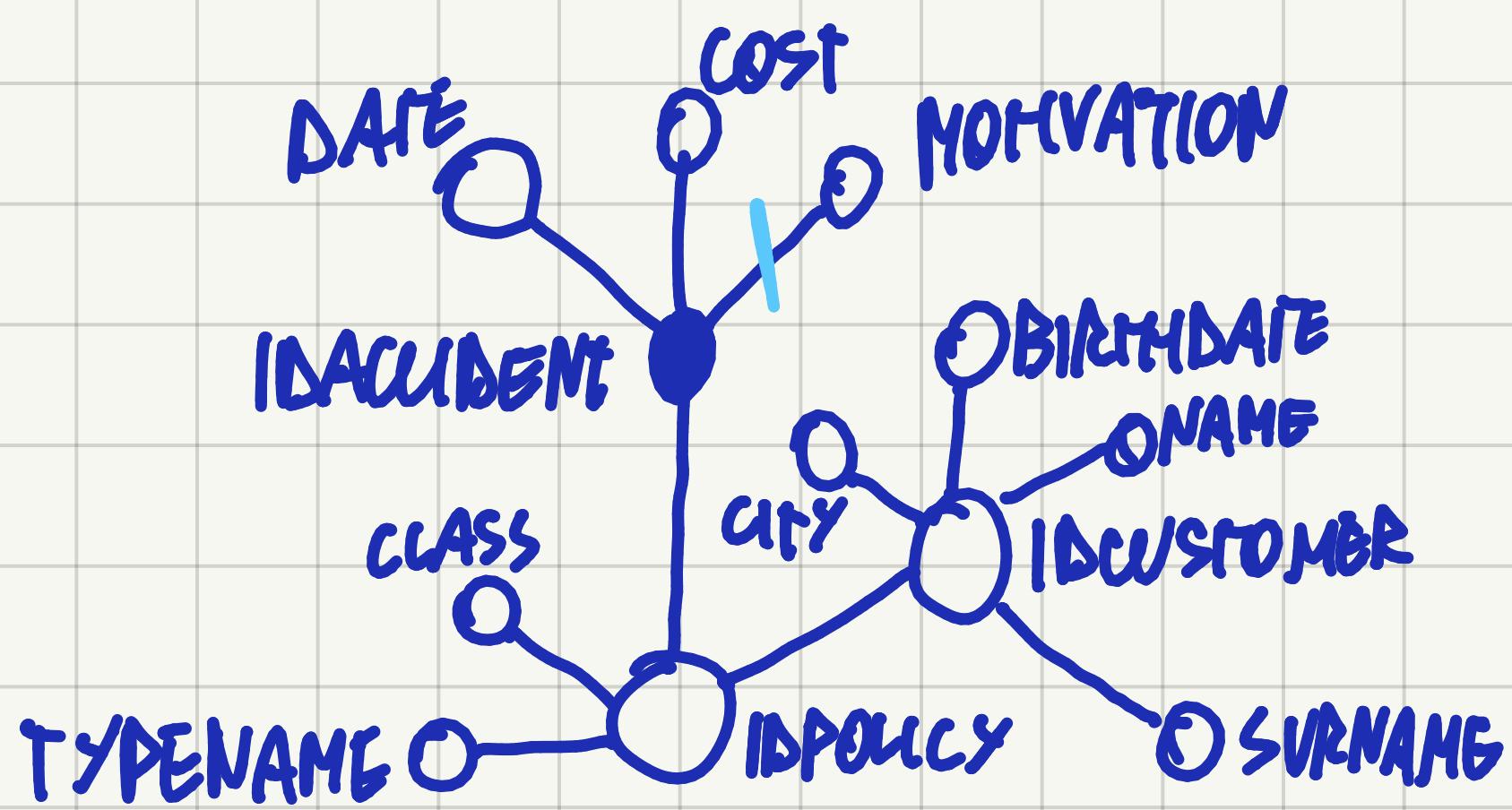


- DELETING A NODE, ITS SUBNODES ARE CONNECTED TO THE MAIN ONE



| IN THIS CASE, WE'LL USE PRUNING

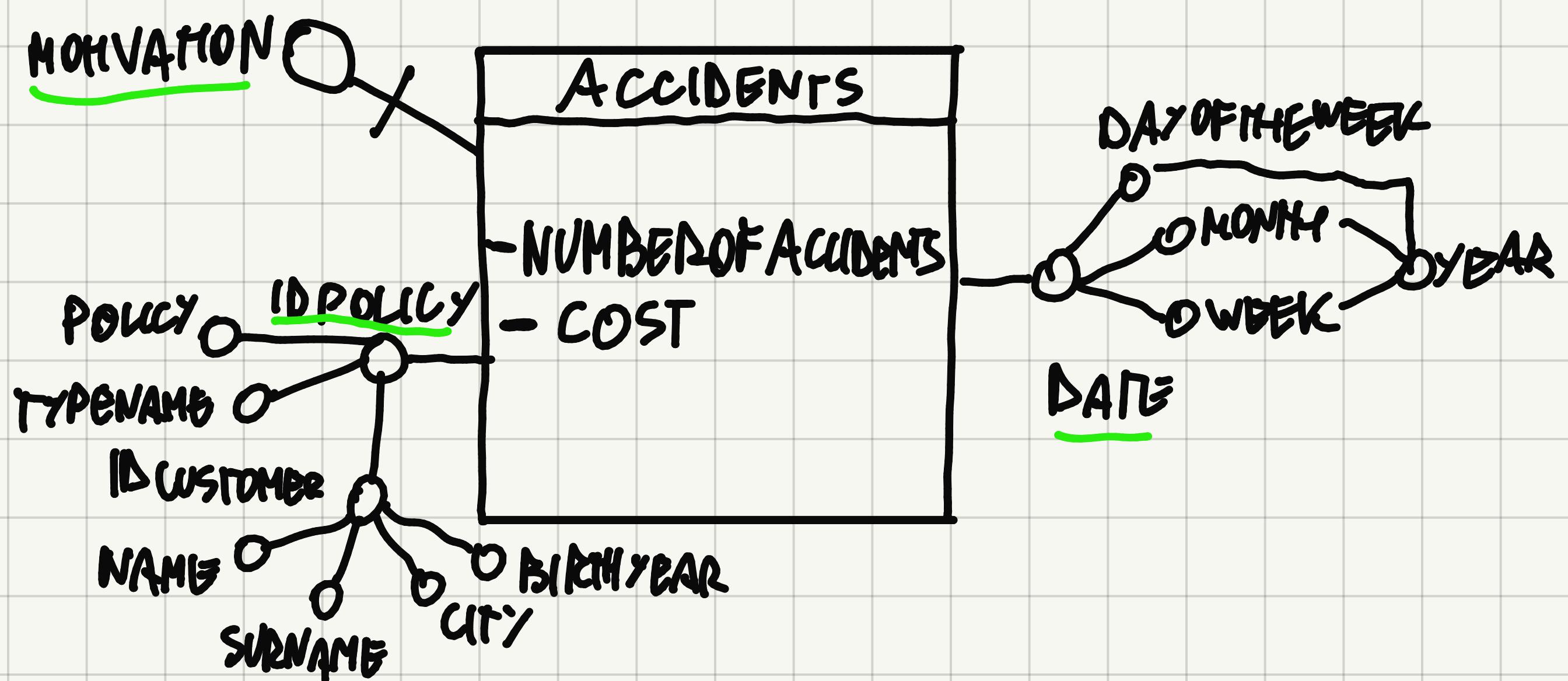




⑥ FACT SCHEMA

NOTE: DIMENSIONS ARE EMPTY DOTS DIRECTLY CONNECTED TO THE FULL DOT

BUT NOT ALL THE CONNECTED EMPTY DOTS ARE DIMENSIONS



DESCRIPTIVE ATTRIBUTES: ATTRIBUTES WITH WHICH IT DOESN'T MAKE MUCH SENSE

TO GROUP BY/AGGREGATE; MOREOVER, IT IS NOT MANDATORY FOR THEM TO RESPECT THE

ONE-TO-MANY RULE OF THE HIERARCHIES

GLOSSARY WILL BE BUILT AS:

MEASURE NAME

SELECT -----, FUNCTION

FROM TABLE(S) USED TO DEFINE THE FACT

WHERE JOINS (IF NEEDED)

GROUP BY

IDCUSTOMER