



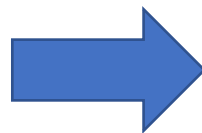
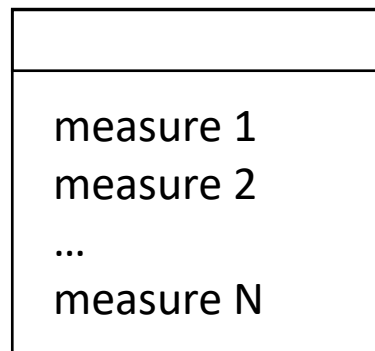
Powered by  
mermaid

# Data Warehouse (I) – Facts table

The following notations should be used to represent the conceptual schema of a Data Warehouse in textual form.

- **Facts table**

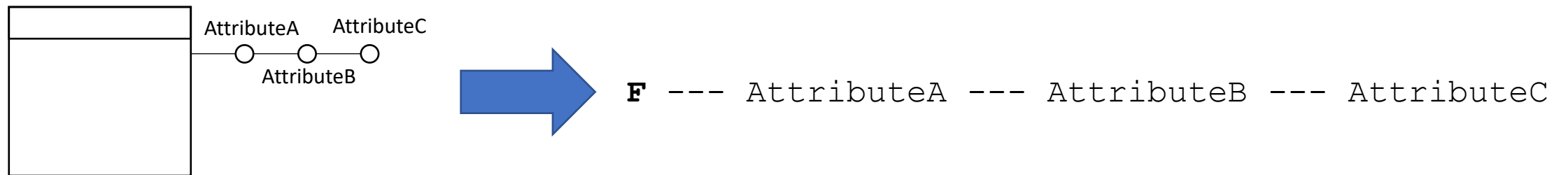
- The facts table will be inserted at the beginning of the reply, specifying the list of related measures and its name (and eventual label)
- Write the label (F), then the name and the measures, separated by “\_\_\_” within doubled square brackets (“[[...]]”)



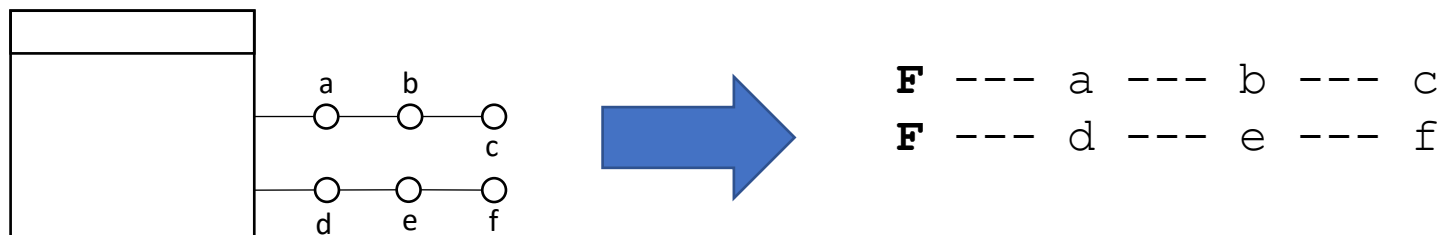
```
F[[Facts  
___  
measure 1  
measure 2  
...  
measure 3]]
```

# Data Warehouse (II) – Dimensions

- Dimensions are represented on the same line
- Arcs that connect attributes to each other and to the fact table are represented with 3 dashes (---) starting from the fact label

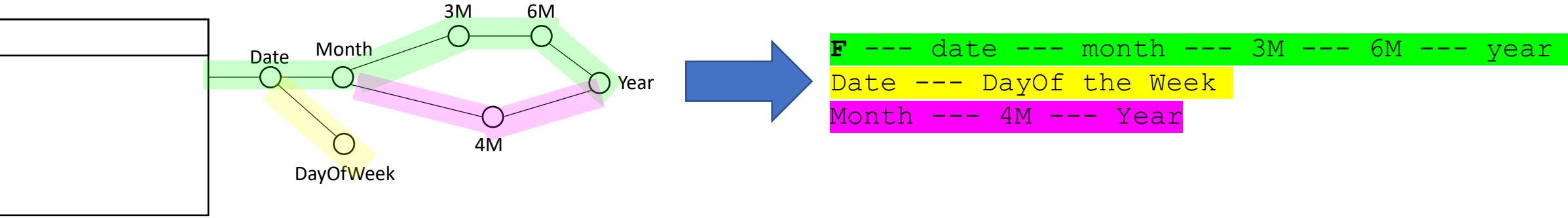
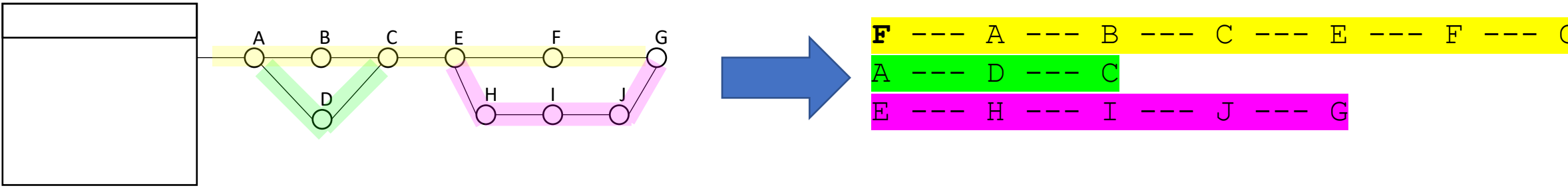


- Use one line for each dimension (path)



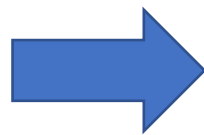
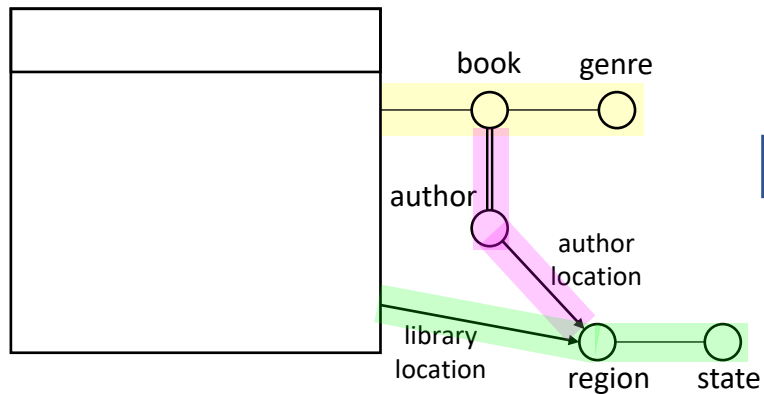
# Data Warehouse (III) – Gerarchies

- To represent divergences and convergences in a hierarchy, structure the dimension on several lines.
- The first row of the dimension represents the longest path within the hierarchy
- The following lines represent "sub-paths" of the hierarchy
- All the lines starts with the first not of the path (or sub-path)



# Data Warehouse (IV) -- Other

- To represent multiple arcs, use === (instead of ---)
- To represent arrows, use -->.
- You can also specify a text within brackets in the arch writing as --(text)->



```
F --- book --- genre  
book === author --(author location)-> region  
F --(library location)-> region --- state
```

# Exercise 1

The management of a multinational wine export company wants to average price per litre of their sales across the world.

Each wine comes from a given **province** in which the grapes were located, belonging to a specific **region**. A **geographical area** groups different regions.

Wines are sold to different **states** across the world, each state belongs to one **continent**.

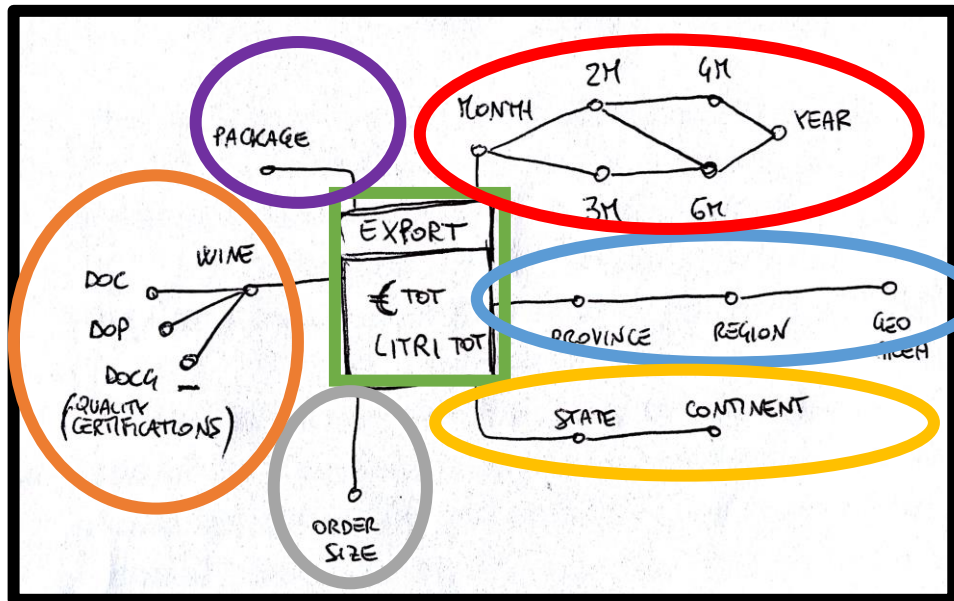
Sales are characterized by an **order size**, which can be small, medium, large, or extra-large, and a specific **package** type.

Wines can have many **quality certifications**, such as DOC, DOP, DOCG, etc. The list of possible quality certifications is known in advance. There are approximately 10 different quality certifications.

Sales must be analyzed for each **month**, 2-month period, 3M, 4M, 6M, year.

Design the conceptual schema of the data warehouse to meet the demands described in the problem specification.

# Data Warehouse (V) – Complete Example



F[[ Export

€ tot

Litri tot]]

F --- Province --- Region --- GeoArea

F --- State --- Continent

F --- OrderSize

F --- Wine --- DOC

Wine --- DOP

Wine --- DOCG

F --- Package

F --- Month --- 2M --- 4M --- Year

Month --- 3M --- 6M --- Year

2M --- 6M



Powered by  
mermaid

# Data Warehouse (VI) – Visualization

You can visualize the corresponding chart pasting the code on [mermaid.live](https://mermaid.live).

Just specify the type of chart by prepending “**flowchart TB**” at the output of the previous slide:

Don't put  
spaces here!

```
flowchart TB
F[ ] Export
€ tot
Litri tot]]

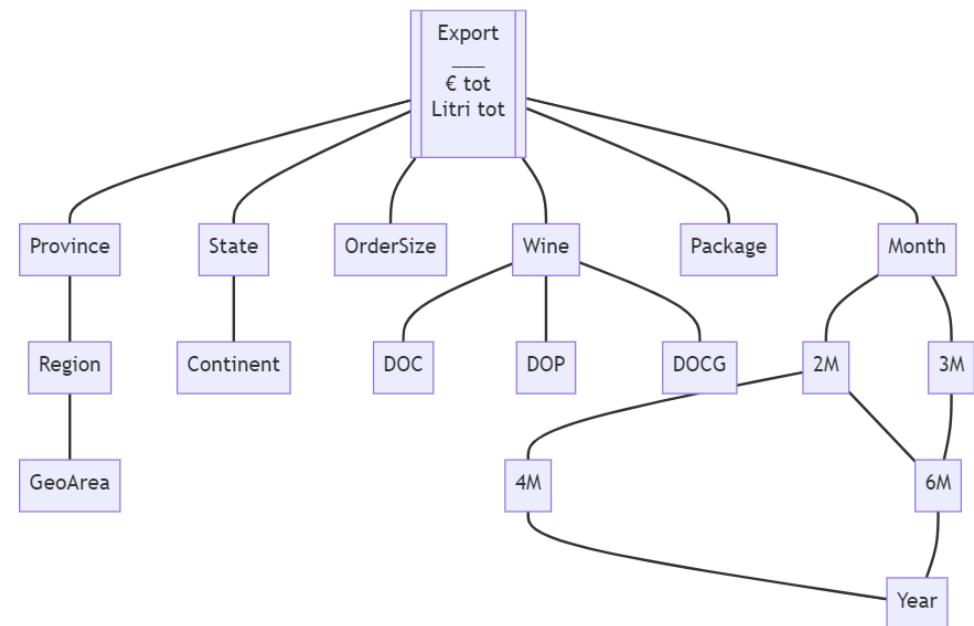
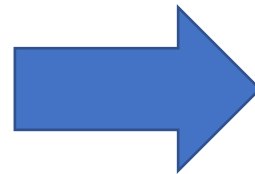
F --- Province --- Region --- GeoArea
F --- State --- Continent

F --- OrderSize

F --- Wine --- DOC
Wine --- DOP
Wine --- DOCG

F --- Package

F --- Month --- 2M --- 4M --- Year
Month --- 3M --- 6M --- Year
2M --- 6M
```



# Exercise 2

The management of a multinational restaurant chain wants to analyze the dishes served in their premises.

For each restaurant is known the **city** in which it is located, the **state** and the **continent**. The **geographical area** where the restaurant is located is also known. Examples of geographical areas are: EMEA (Europe, Middle East, Africa), Americas, Asia-Pacific. A state is assigned to only one geographic area, but states on the same continent can belong to different geographic areas.

Dishes are served at a specific **date** and **time**. Management is interested in analyzing the dishes served also at the level of the **day of the week**, at the level of **month**, **quarter**, **semester** and **year**. You also want to carry out analyses in terms of month of the year and quarter of the year (Q1, Q2, Q3 or Q4).

In addition to the time, of particular interest is also the time **slot of the day** in which a dish is served: in particular, the time from 12 to 15 is called "lunch", while the time from 19 to 22 is called "dinner". Dishes served outside these ranges are considered "out of hours".

For each dish served, the types of **diet** for which it is suitable are known (for example, vegetarian, vegan, pescetarian, gluten-free). One dish can be suitable for different types of diet. It is not known a priori the number of types of diet available. Finally, we want to carry out the analysis based on the type of course (appetizer, first course, second course, side dish, dessert). **Each dish is assigned to only one type of course.**

The management wants to know, as the above factors change:

- the number of dishes served
- the average selling price of each dish

Design the conceptual schema of the data warehouse to meet the demands described in the problem specification.