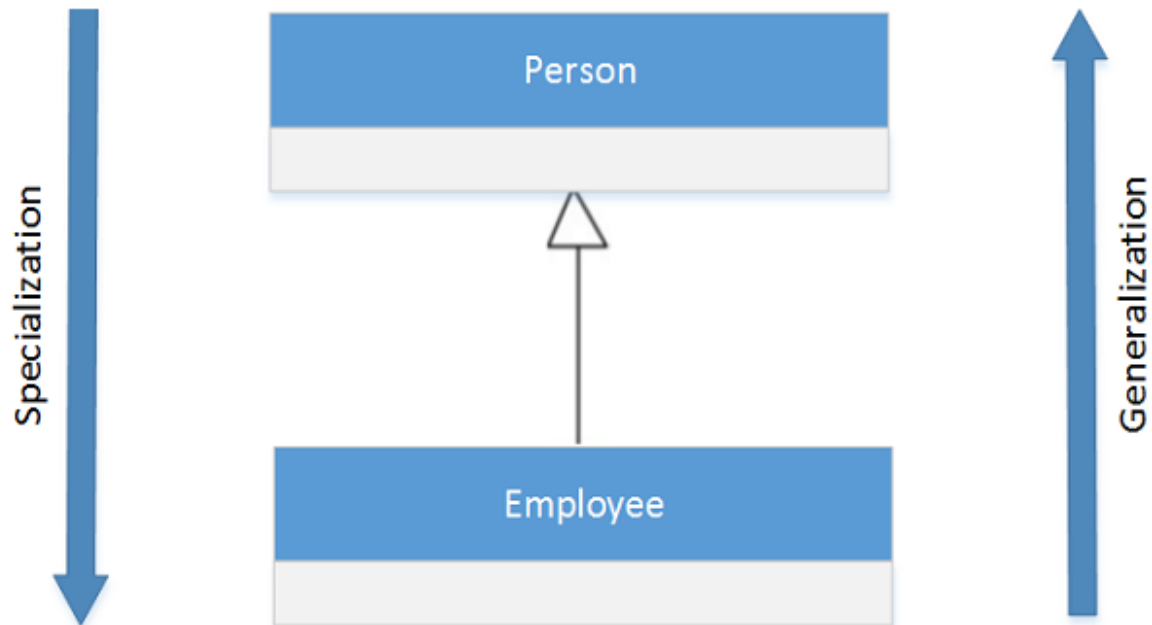


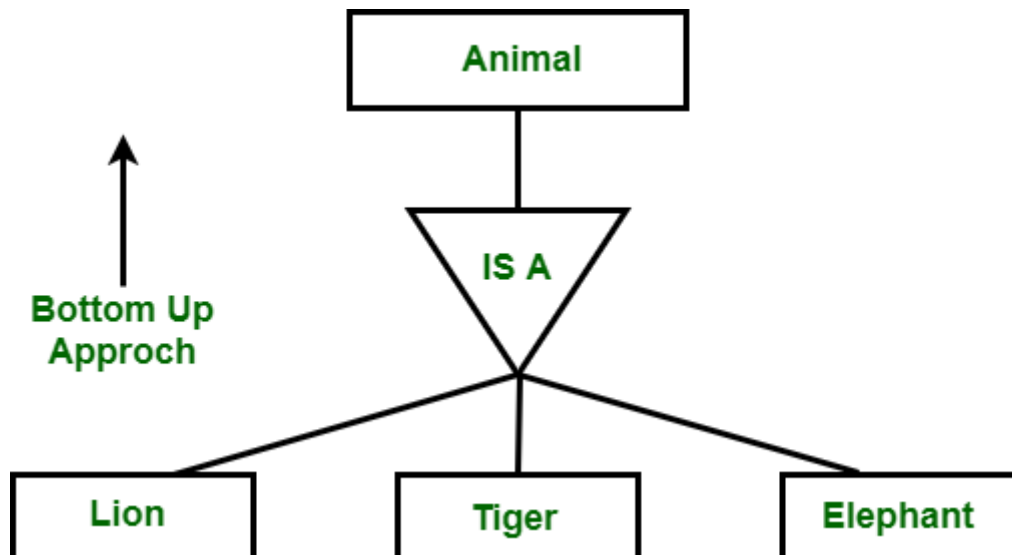
Generalization:

Data Generalization is the process of summarizing data by replacing relatively low level values with higher level concepts. It is a form of descriptive data mining.



The Employee class extends the Person class (Inheritance)

Generalization in data analytics refers to the ability of a predictive model to perform effectively on new, unseen data that it was not specifically trained on. In other words, a model that generalizes well can make accurate predictions or provide meaningful insights beyond the data it was initially exposed to. Generalization is a key goal when developing models because it ensures that they can be applied to real-world situations and not just reproduce patterns present in the training data. Achieving good generalization typically involves techniques like cross-validation, regularization, and careful model selection to prevent overfitting and underfitting.



There are two basic approaches of data generalization :

1. Data cube approach :

- It is also known as OLAP approach.
- It is an efficient approach as it is helpful to make the past selling graph.
- In this approach, computation and results are stored in the Data cube.
- It uses Roll-up and Drill-down operations on a data cube.
- These operations typically involve aggregate functions, such as count(), sum(), average(), and max().
- These materialized views can then be used for decision support, knowledge discovery, and many other applications.

2. Attribute oriented induction :

- It is an online data analysis, query oriented and generalization based approach.
- In this approach, we perform generalization on basis of different values of each attributes within the relevant data set. after that same tuple are merged and their respective counts are accumulated in order to perform aggregation.
- It performs off-line aggregation before an OLAP or data mining query is submitted for processing.
- On the other hand, the attribute oriented induction approach, at least in its initial proposal, a relational database query – oriented, generalized based (on-line data analysis technique).
- It is not limited to particular measures nor categorical data.
- Attribute oriented induction approach uses two method :

(i). Attribute removal.

(ii). Attribute generalization.

EXAMPLE OF GENERALISATION

BOTTOM UP APPROACH

