



SAPIENZA
UNIVERSITÀ DI ROMA

Data Management and Analysis

Unit 2

Introduction

Dott. Franco Liberati
liberati@di.uniroma1.it



INTRODUCTION

Digital Data: definition

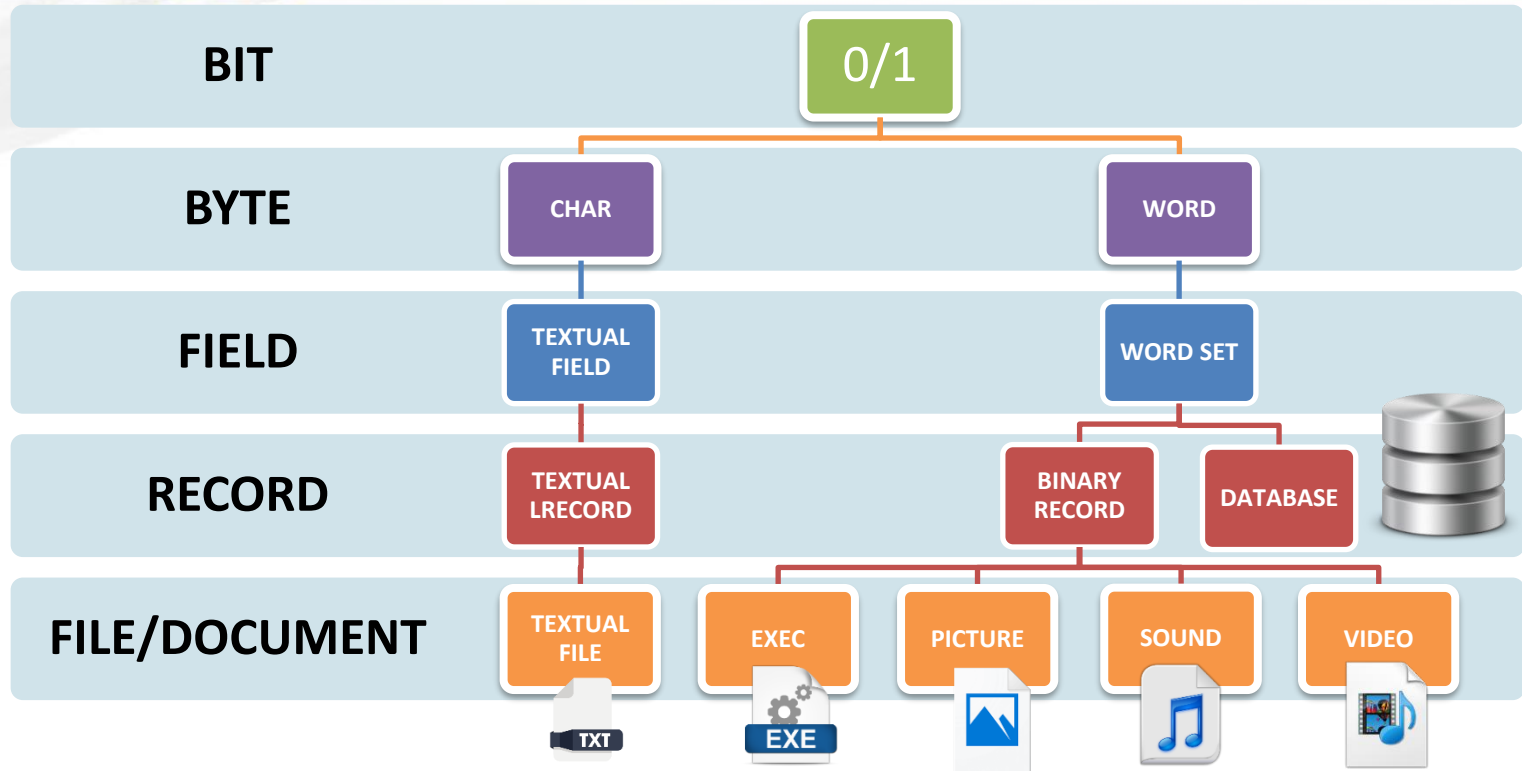
Digital data is information represented as a string of discrete symbols, each of which can take on one of only a finite number of values from some alphabet, such as letters or digits

Example

ASCII CODE:	'A'	01000001	$(65)_{10}$
RGB MODEL CODE:	■	11111111 11111111 00000000	$(16776960)_{10}$

INTRODUCTION

Digital Data: hierarchy



INTRODUCTION

Digital Data: overview

- ❑ In recent years the amount of digital information has grown exponentially
- ❑ According to recent research there are 2.5 quintillion bytes of data created each day



Social Media

- Snapchat users share 527,760 photos (every minute)
- More than 120 professionals join LinkedIn (every minute)
- Users watch 4,146,600 YouTube videos (every minute)
- 456,000 tweets are sent on Twitter (every minute)
- Instagram users post 46,740 photos (every minute)



Communication

- 16 million text messages (every minute)
- 990,000 Tinder swipes (every minute)
- 156 million emails are sent every minute
- 15,000 GIFs are sent via Facebook messenger
- Every minute there are 103,447,520 spam emails sent
- There are 154,200 calls on Skype (every minute)



Digital Photos

- People will take around 3 trillion photos by the end of 2022
- There will be 7.4 trillion photos stored



The Internet of Things

- Around 200 billion IoT device connected by 2020



DATABASE AND DATA MODEL



DATABASE

Definition

A database is a collection of data managed by a *database management system*

A database management system (DBMS) is a software system able to manage collections of data that are large, shared and persistent, and to ensure their reliability and privacy. Like any software product, a DBMS must be efficient and effective.



DATABASE

Data Model

Data models describe the structure, manipulation, and integrity aspects of the data stored in DBMS



DATABASE

Data Model

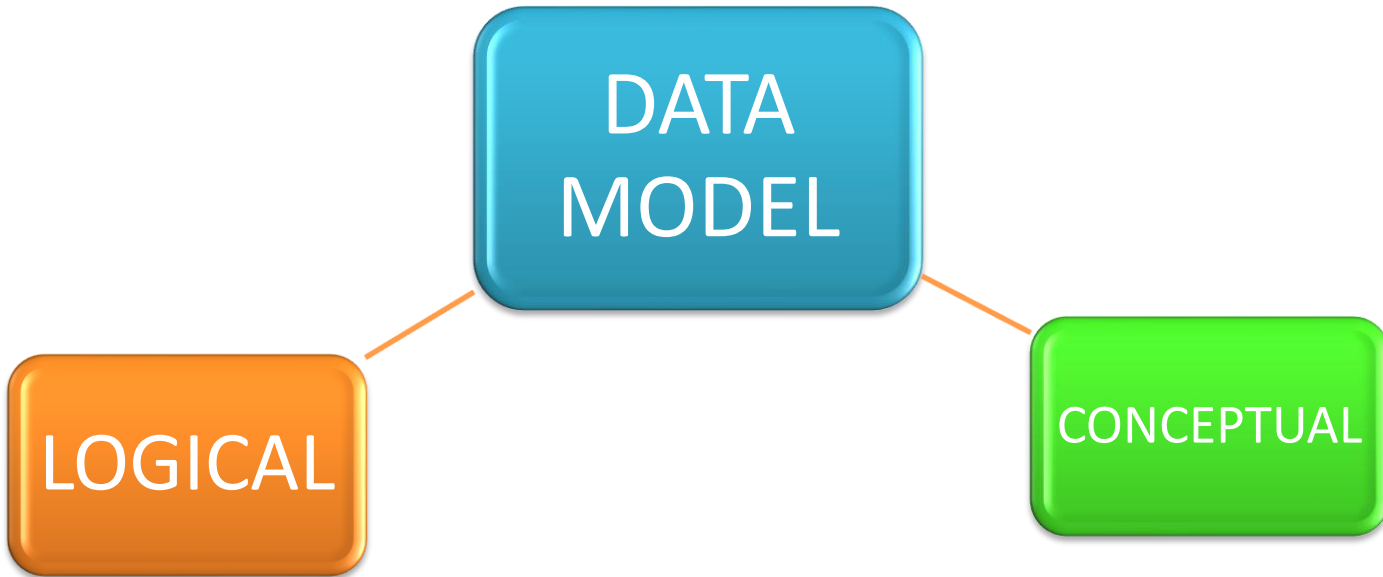
A data model is a combination of **constructs** used to organize data

Each data model provides structuring mechanisms, similar to the type constructors of programming languages, which allow the definition of new data types based on constructors applied to predefined, elementary types

DATABASE

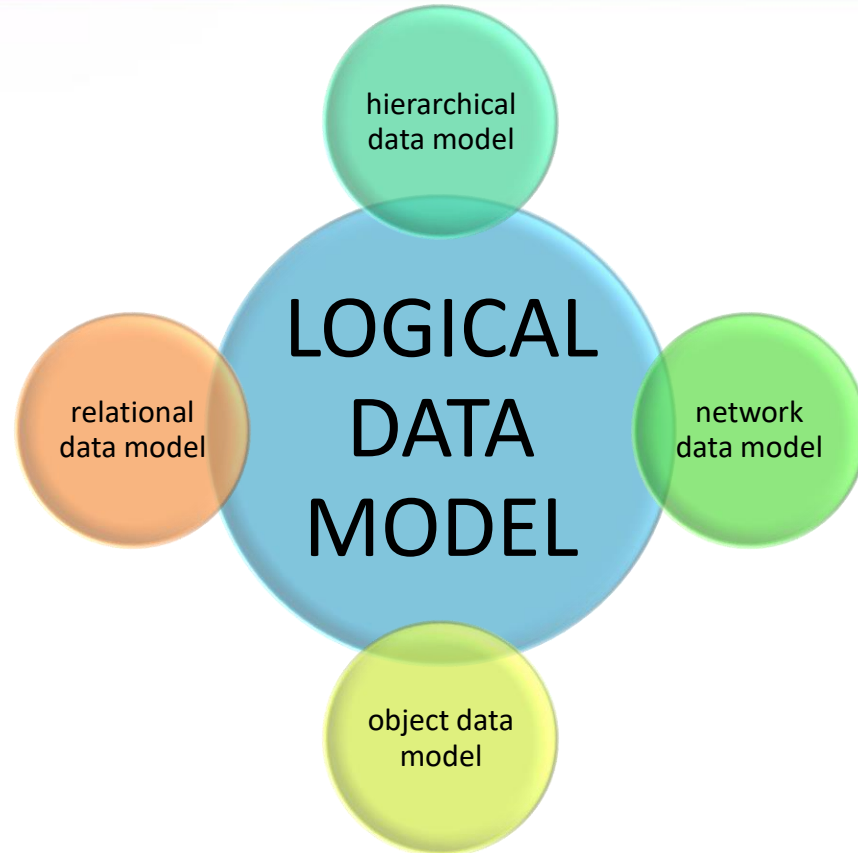
Data Model

There are different data model



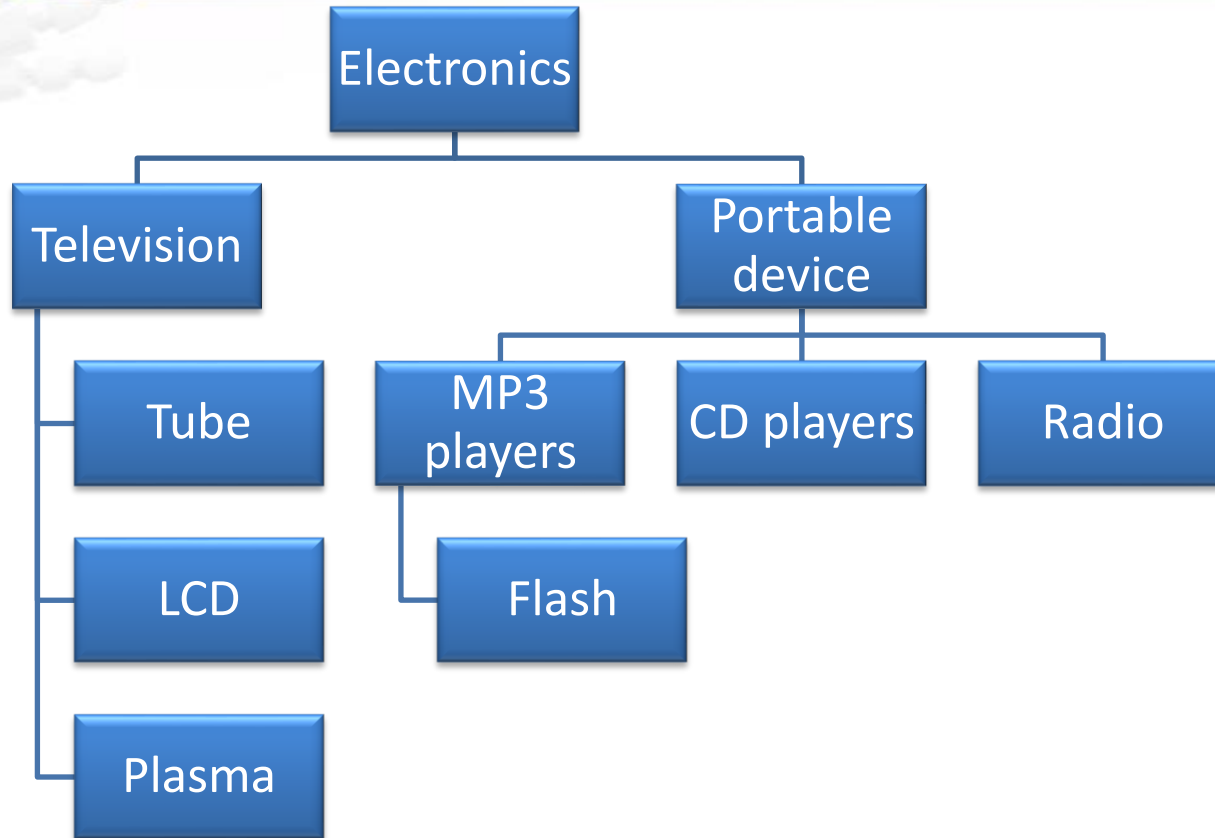
DATABASE

Data Model: LOGICAL



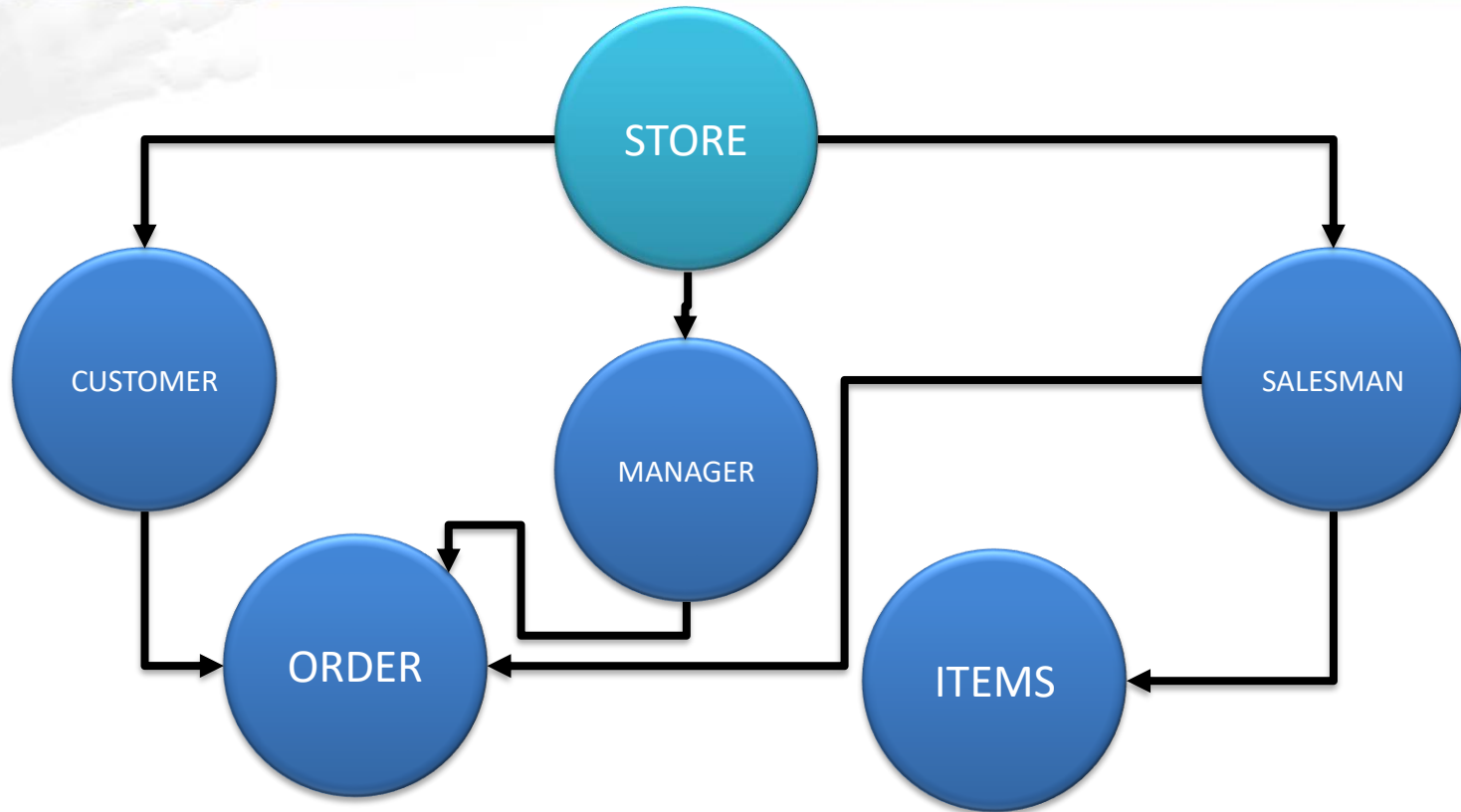
DATABASE

Data Model: LOGICAL-Hierarchical



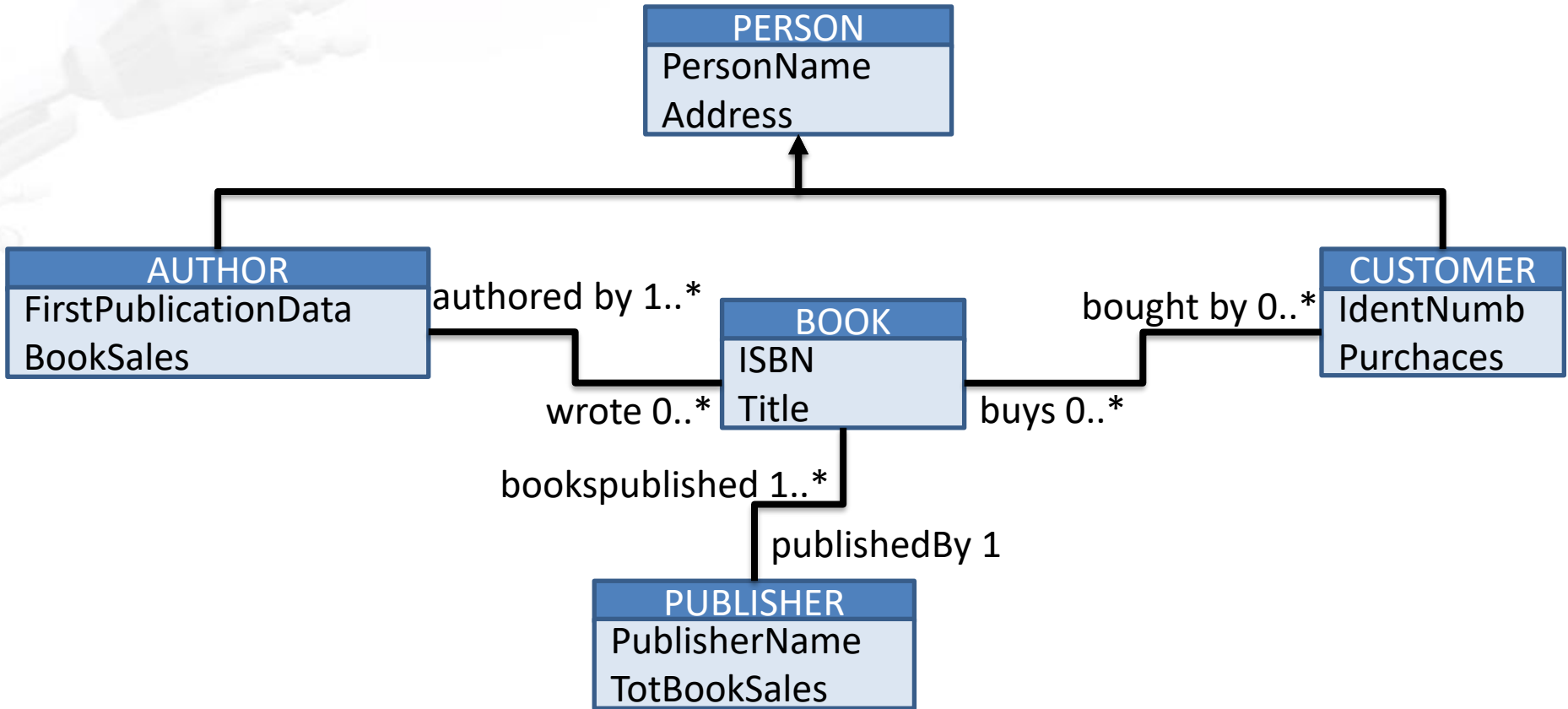
DATABASE

Data Model: LOGICAL-Network



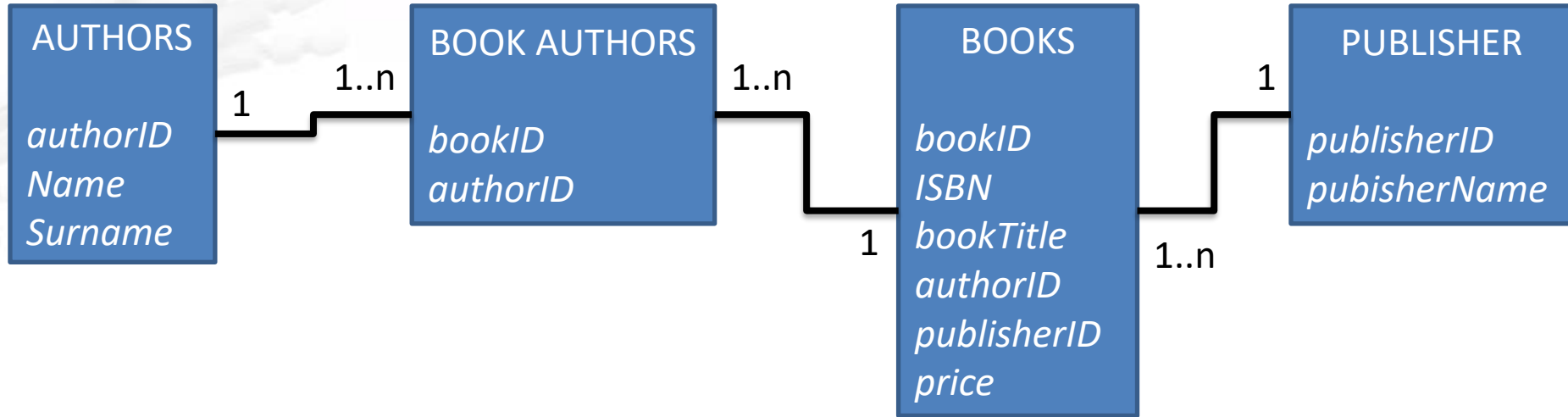
DATABASE

Data Model: LOGICAL-Object



DATABASE

Data Model: LOGICAL-Relational

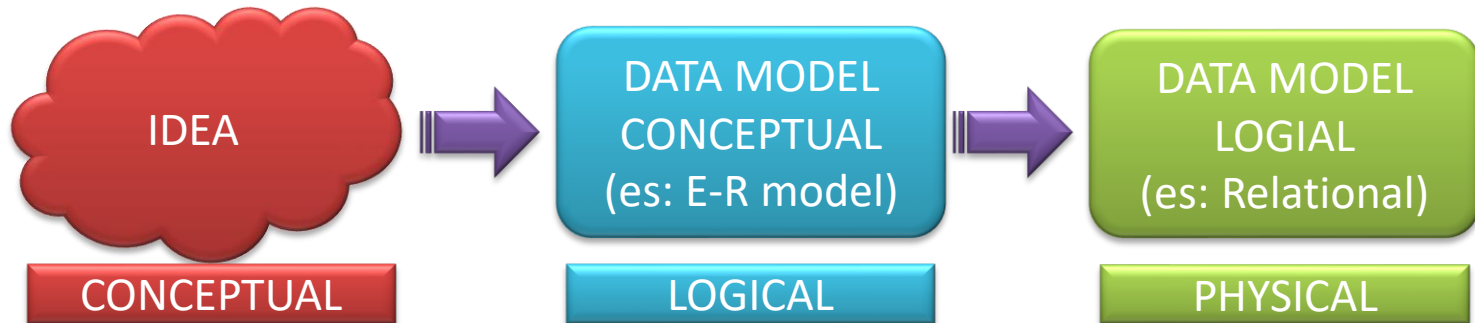


DATABASE

Data Model CONCEPTUAL

Conceptual Data Models have been introduced to describe data in a manner independent of the logical mode

Their name comes from the fact that they tend to describe concepts of the real world, rather than the data needed for their representation





DATABASE

Architecture DBMS

LOGICAL

- Description of the whole database

INTERNAL

- Implementation of the logical schema

EXTERNAL

- Description of a portion of the database by means of the logical model

DATABASE

Database Schema

In a database there is a part that is invariant in time, called ***schema*** made up of the characteristics of the data, and a part that changes with time, called **instance**, made up of the actual values

TEACHING	
COURSE	TEACHER
DMA	Pirrò
PSN	Melatti
Network	Bongiovanni

SCHEMA: TEACHING (Course, Teacher)

INSTANCE: (DMA, Pirrò)

(PSN, Melatti)

(Network, Bongiovanni)

DATABASE

Architecture DBMS external schema example

The external level is not explicitly present, but it is possible to define derived relations (called ***views***)

TEACHING	
COURSE	TEACHER
DMA	Pirrò
PSN	Melatti
Network	Bongiovanni

DEGREE PROGRAMME BIG DATA		
Degree	Course	Year
BigData	DMA	4
Bigdata	Network	4



DATABASE

Architecture DBMS: data independence

Physical independence allows interaction with the DBMS independently of the physical aspects of the data

Logical independence guarantees that the interaction with the external level of the database is independent of the logical level



DATABASE

Roles

The **database administrator** (DBA) is the person responsible for the design, control and administration of the database

The **application designers and programmers** define and create programs that access the database

The **users** employ the database for their own activities



DATABASE

Language

The **data definition language** (DDL) is used to define the logical, external and physical schemas and access authorizations

The **data manipulation language** (DML) is used for querying and updating database instances



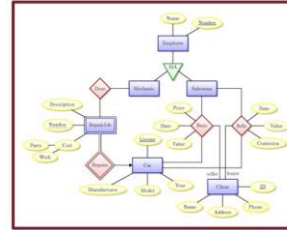
DATABASE

Design of a database

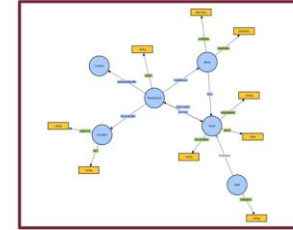
Conceptual
design

Logical
design

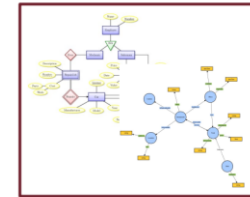
Physical
design



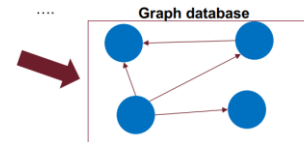
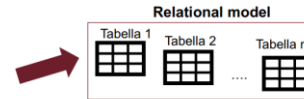
Entity-Relationship Model (ER)



Ontology



Conceptual schema



PostgreSQL





End