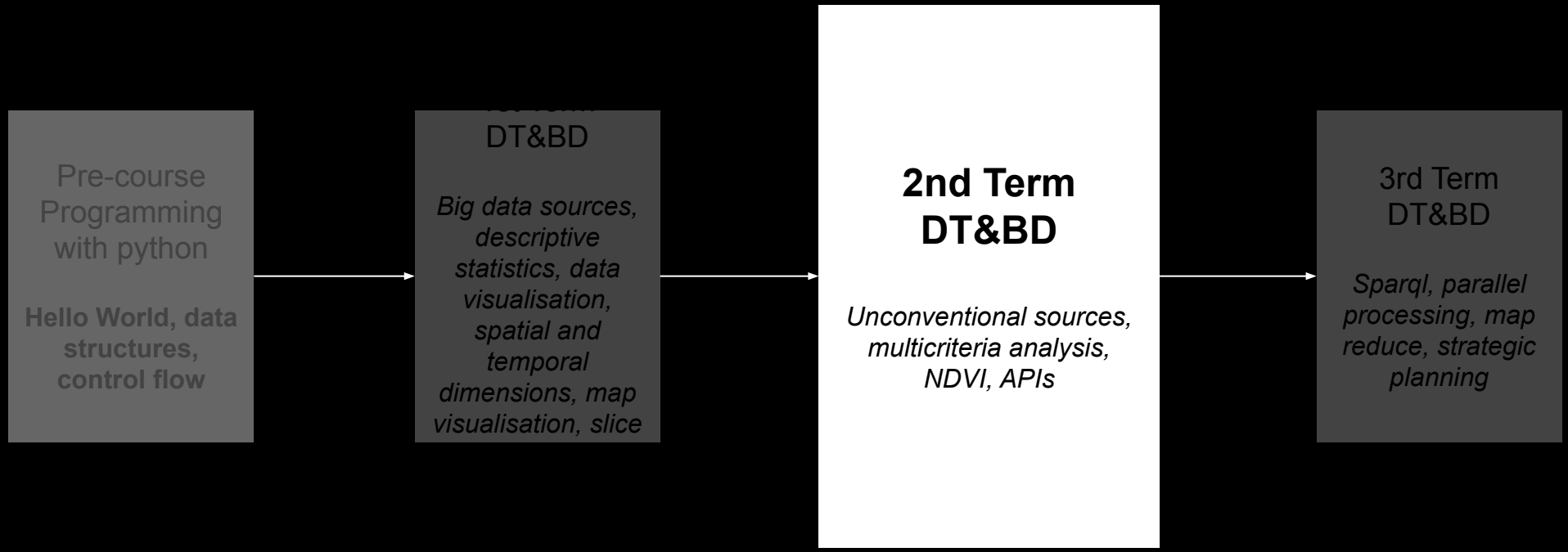
Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA - 2nd Term
2019/2020

FACULTY DIEGO PAJARITO

Data sources, analysis and management


Pre-course
Programming
with python
Hello World, data
structures,
control flow
DT&BD
Big data sources,
descriptive
statistics, data
visualisation,
spatial and
temporal
dimensions, map
visualisation, slice
2nd Term
DT&BD
Unconventional sources,
multicriteria analysis,
NDVI, APIs
3rd Term
DT&BD
Sparql, parallel
processing, map
reduce, strategic
planning

Structured / unstructured

The degree in which a data set follows a set of rules.

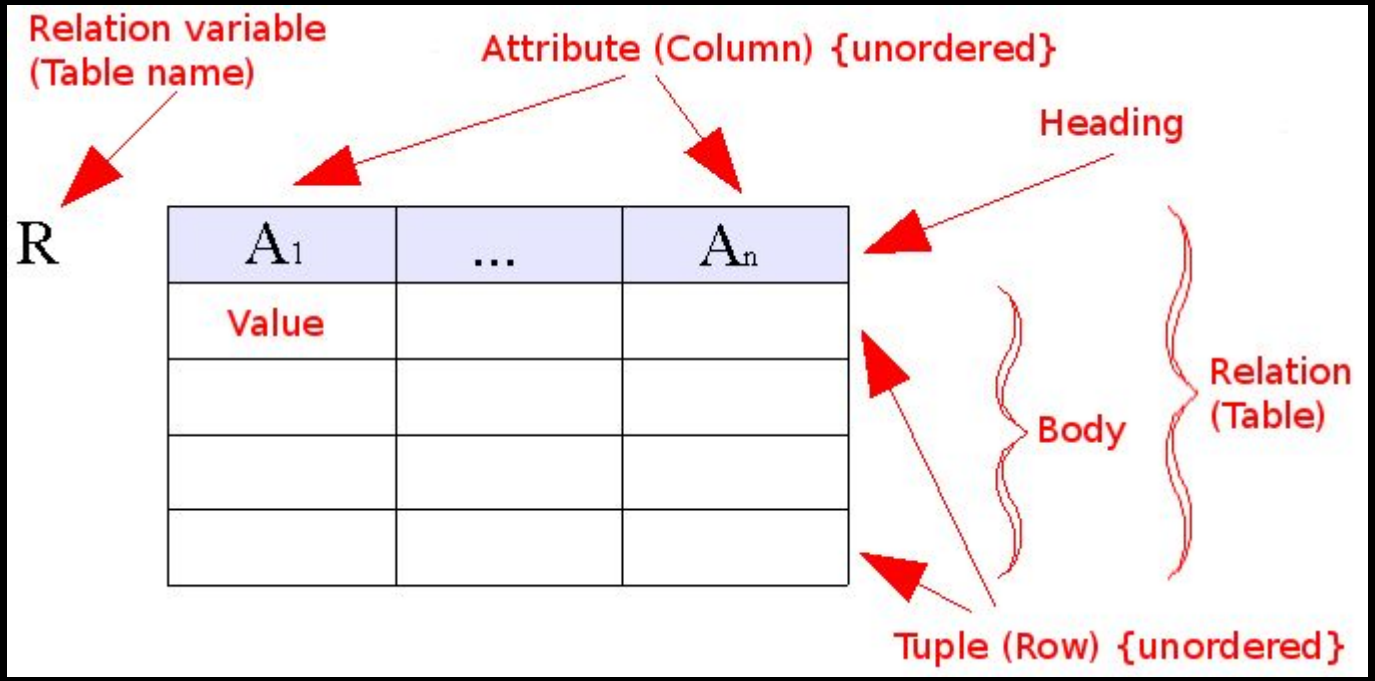
Data with a consistent structure that allows computers to organize (index) and access (query) it. The logical structure aims at fitting into a series of operations that maximize computer processing times and, not necessarily, simplify human comprehension.

Human thinking and language have, by nature, a flexible structure. Therefore, to directly fit a structure is not within human nature and demands a translation process.
The purpose of the relational model is to provide a declarative method for specifying data and queries:

* Relational Database Model

Definition from: Codd, E.F (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. Classics. 13 (6): 377–87. doi:10.1145/362384.362685. Archived from the original on 2007-06-12.

Relational model



A realization of relational algebra.

An example from Wikipedia at: The relational model

Structured / unstructured

(Tradition sticks to the rules)

~~Duplicate rows~~
~~Anonymous columns~~
~~Duplicate column names~~
Column order significance (operations)
Check option
Columnless tables unrecognized

Definition from: Codd, E.F (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. Classics. 13 (6): 377–87. doi:10.1145/362384.362685. Archived from the original on 2007-06-12.

To explore a wide range of techniques that allow *feeding analysis* and design of advanced architecture with data coming from either *traditional or non-traditional* sources.

The course provides a practical perspective of the different types of data sources relevant to urban analytics. Starting from traditional and structured data sets, the students will experience the different approaches to *access and manage massive datasets* to describe city features.

Manage structured and unstructured data sets

Define queries for filtering and cleaning data

Design analysis models involving spatial constraints

Multidimensional plots and dashboards

Web maps and story maps

Data journalism and data-driven multimedia

Source Code

Examples of analysis tasks performed during the course

Final Presentation / web app

Windows/Linux/Mac Computer (Dual-Core + processor, RAM 8Gb)

Python 3.6.x or later (Get the installer at https://www.python.org/downloads/)

Anaconda (Get the installer at https://www.anaconda.com/distribution/)

Pycharm Community (Get the installers at https://www.jetbrains.com/pycharm/download)

QGis Desktop (Get the installers at https://qgis.org/en/site/forusers/download.html)

pandas https://pandas.pydata.org/
geopandas http://geopandas.org/
matplotlib https://matplotlib.org/
seaborn https://seaborn.pydata.org/
scrapy https://scrapy.org/
Cartogram plugin for QGIS https://plugins.qgis.org/plugins/cartogram/
Story map JS https://storymap.knightlab.com/
pgAdmin (get the installer at https://www.pgadmin.org/download)

Let’s set this up

pgAdmin
Pycharm and pandas working
Scrapy
Qgis
* Visual Studio Code

Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA - 2nd Term
2019/2020

FACULTY DIEGO PAJARITO