

BIGDATA 1

Informació del curs

Programa

1. Introducció a Big Data. La cadena de valor. Ecosistema d'eines per Big Data.
2. Introducció a estadística per Big Data
3. Introducció a programació per Big Data
4. Ingesta i preparació de dades. Infraestructura i eines emprades
5. Procediments per l'anàlisi de dades
6. Explotació i visualització de dades

Docents

- Natàlia Plana Darnés
- Xavier García Coca
- Marc Comas Cufí

Evolució de la ciència de dades

L'anàlisi de dades

- Incorporació de l'ordinador com a eina d'anàlisi:

For a long time I have thought I was a statistician [...] But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. [...] All in all, I have come to feel that my central interest is in **data analysis**, which I take to include, among other things:

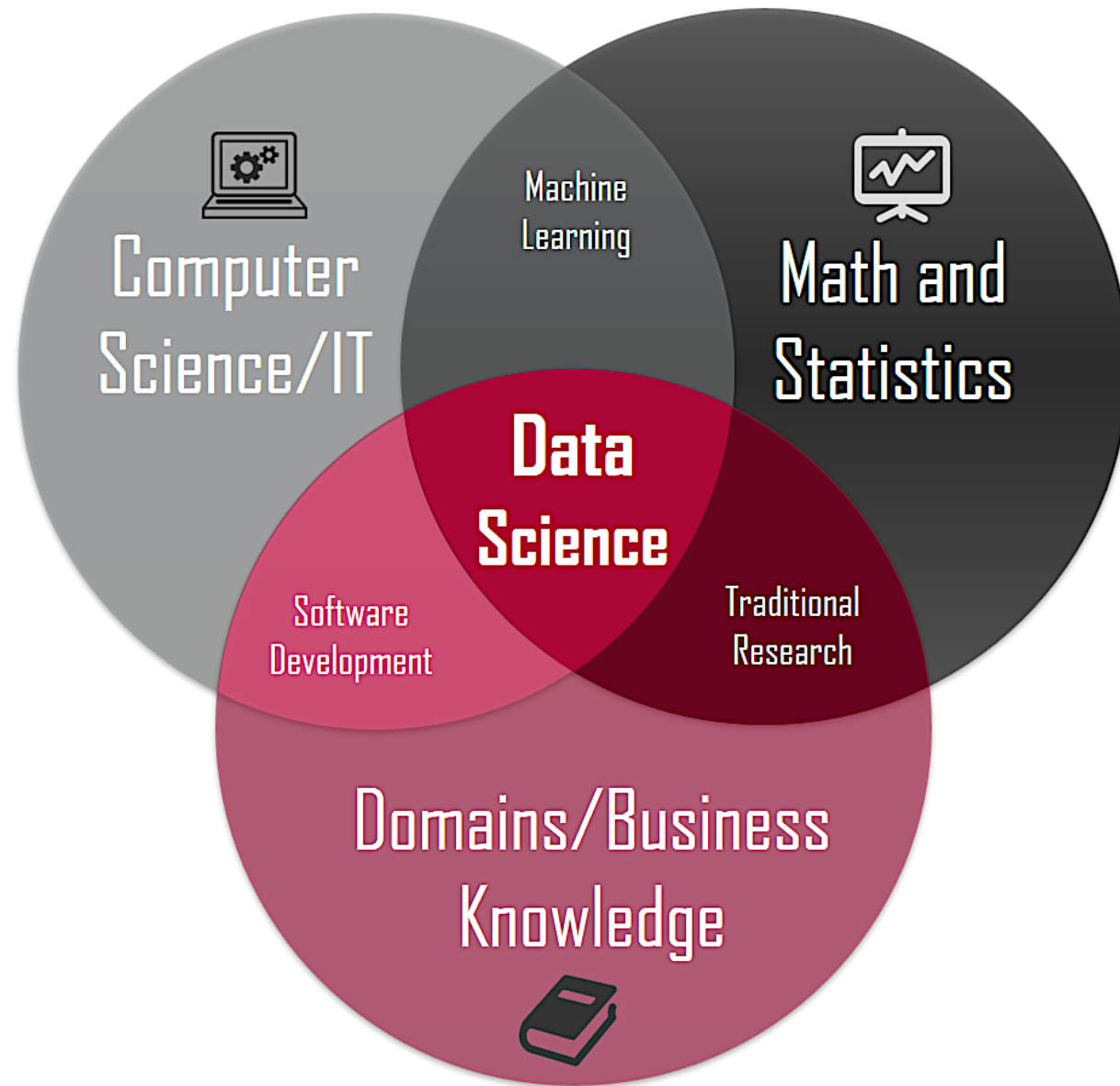
- procedures for analyzing data,
- techniques for interpreting the results of such procedures,
- ways of planning the gathering of data, and
- all the machinery and results of statistics which apply to analyzing data.

(Tukey, 1962)

Evolució fins el dia d'avui

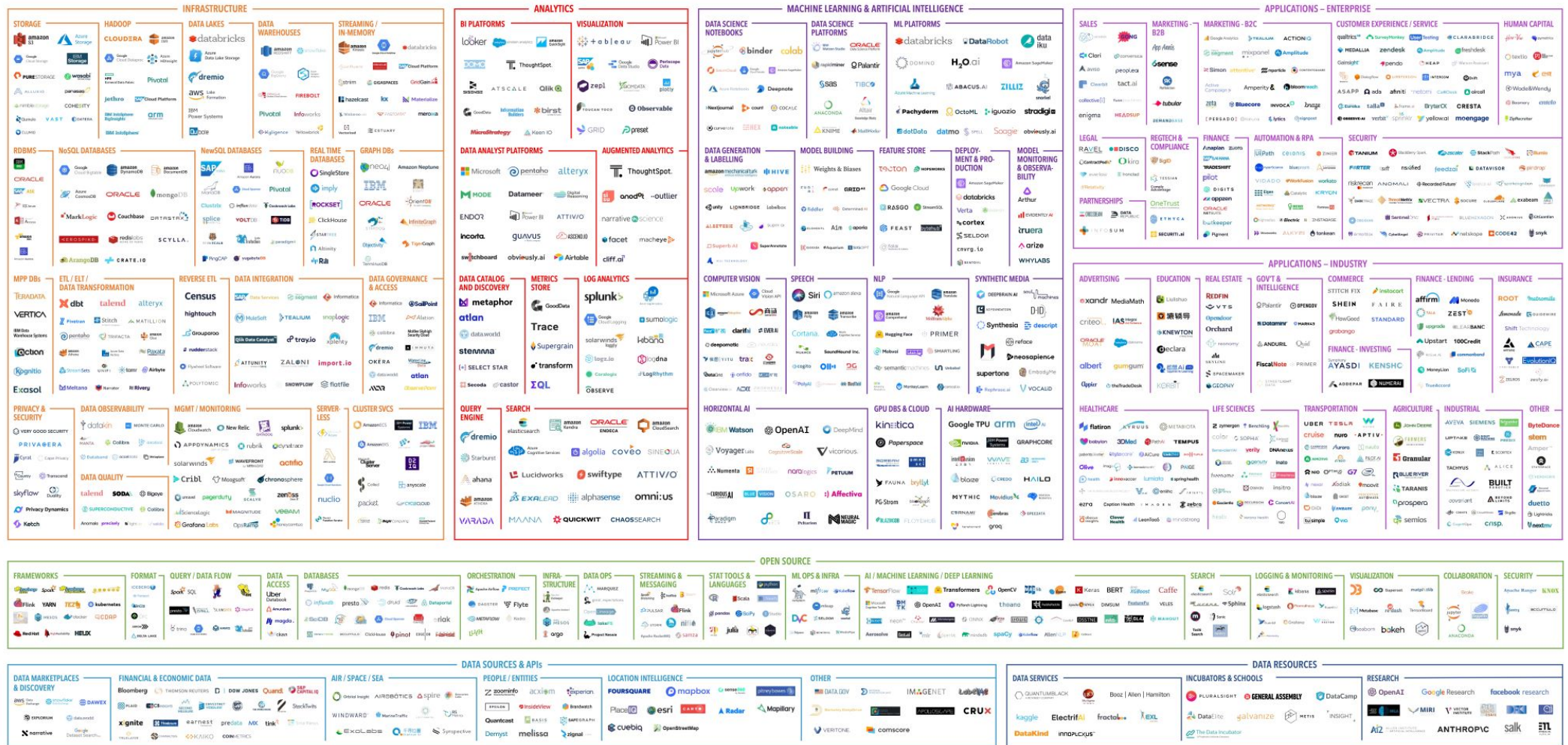
- Els anys 90 es fa una transició cap a les empreses.
 - De la gestió i emmagatzematge de dades, neix la necessitat de prendre decisions a partir de les dades disponibles dins l'empresa (**Business Intelligence**).
 - Eines de l'extracció, la transformació i la càrrega de dades (ETL).
 - Eines d'exploració de dades: consulta de BBDDs, cubs OLAP,...
 - Eines directives: Business Performance Management (BPM), quadres de comandament, ...
- IFCS 1996: **Data Science**, Classification and Related Methods
- Des dels inicis de l'any 2010 fins ara, s'ha anat popularitzant la ciència de dades (Data Science).
 - *Internet of Things*. Segons IBM, el 90% de la informació disponible al 2018 s'ha generat el 2016 i el 2017. Necessitat del processament de dades distribuït. **Big Data**
 - Popularització de tècniques de *Machine Learning* per la presa de decision a grans empreses.

Conceptes de ciència de dades



Panorama de software el 2021

MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND DATA (MAD) LANDSCAPE 2021



Version 3.0 - November 2021

© Matt Turck (@mattturck), John Wu (@john_d_wu) & FirstMark (@firstmarkcap)

mattturck.com/data2021

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

<https://mattturck.com/data2021/>

High resolution

Software que utilizarem



Criteris d'avaluació

- Entregues durant el curs (50%).
- Examen tipus test amb ordinador el darrer dia sobre conceptes bàsics (50%).