Specialised Database Concepts

Oscar Romero

Facultat d'Informàtica de Barcelona

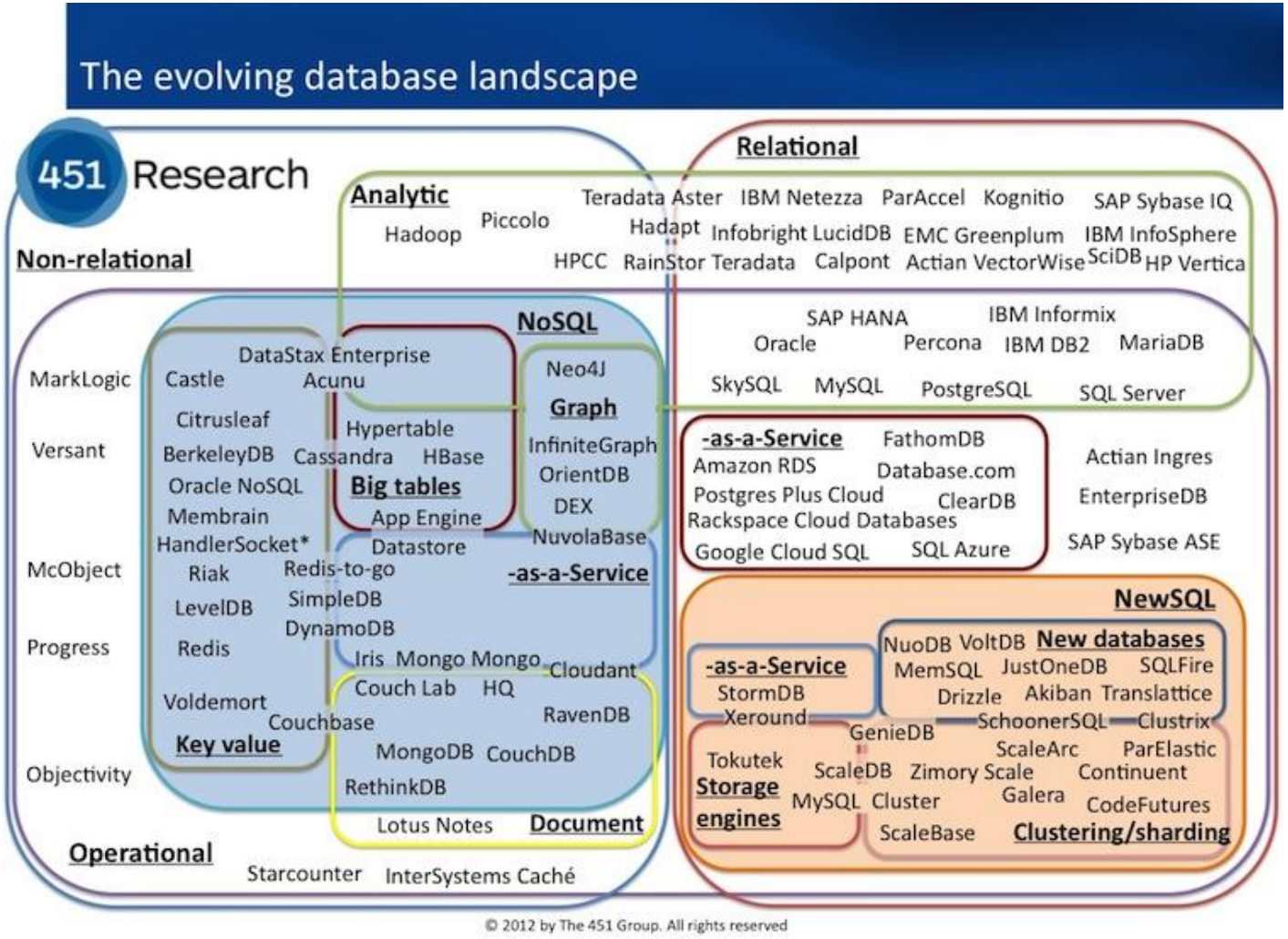
Universitat Politècnica de Catalunya

INTRODUCTION

ONE SIZE DOES NOT FIT ALL

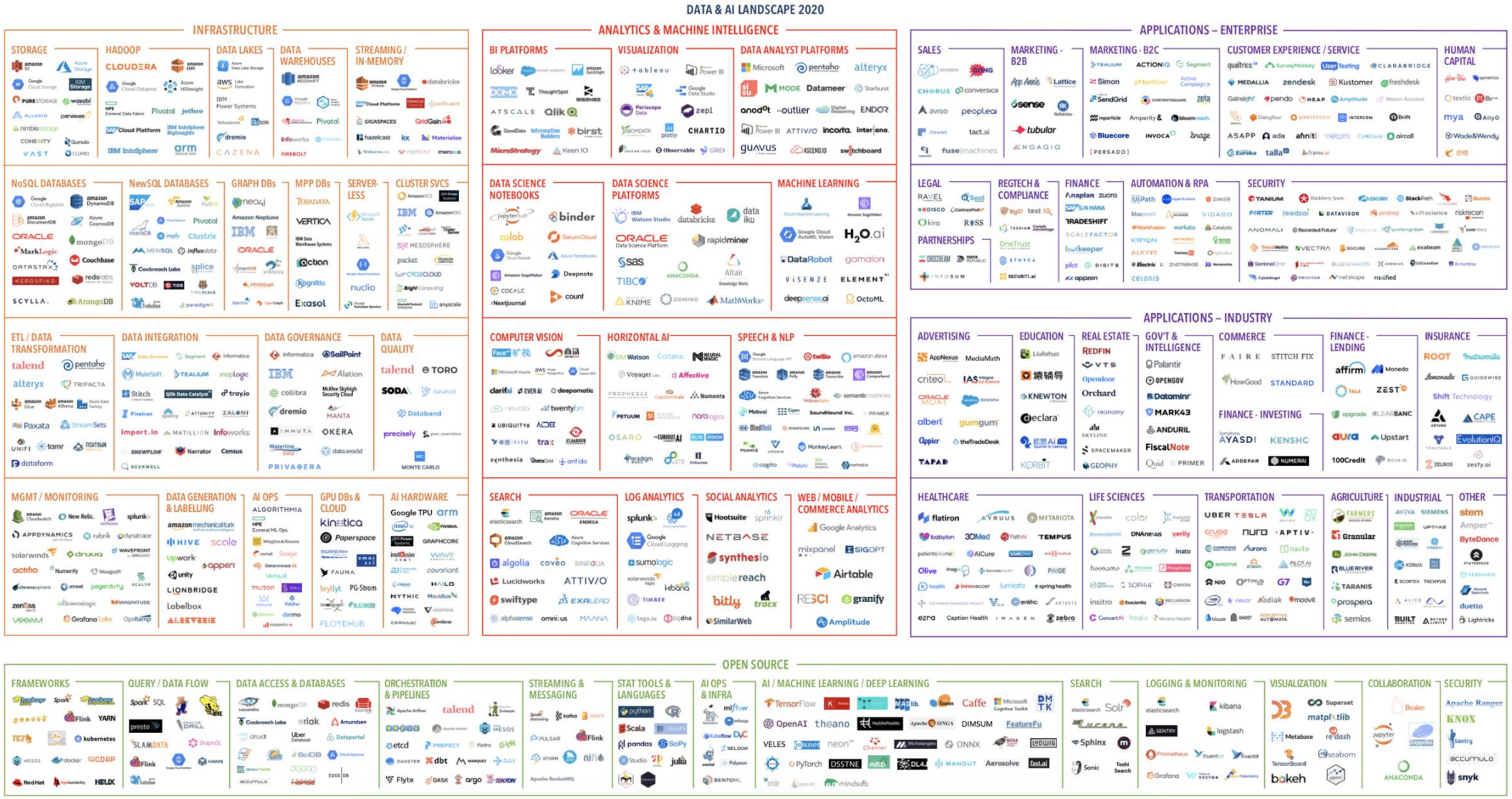
Hey, What Is Going On?!

The Data Landscape (2012)



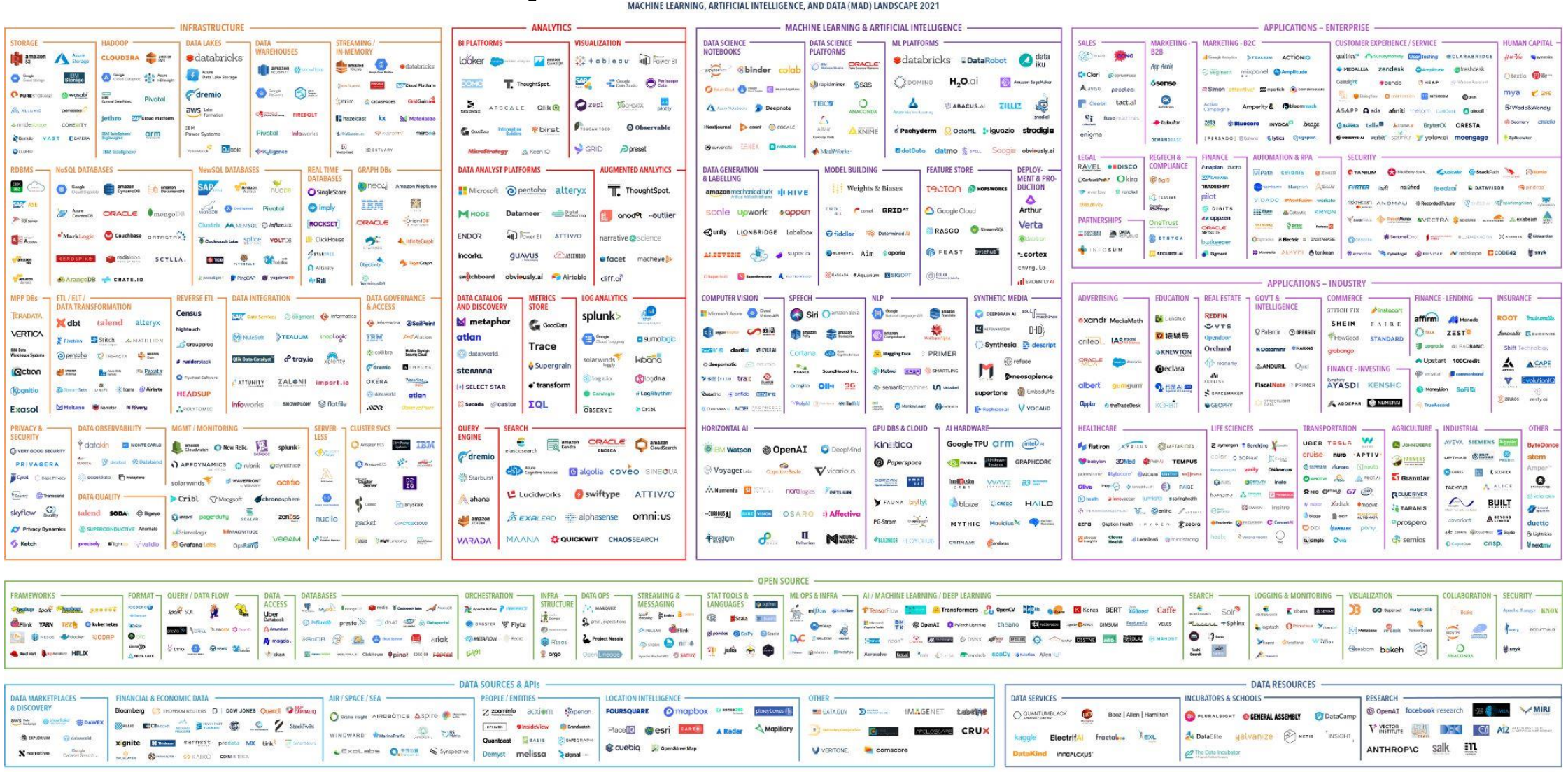
By 451 Research

The Data Landscape (2020)



By Matt Turck

The Data Landscape (2021)



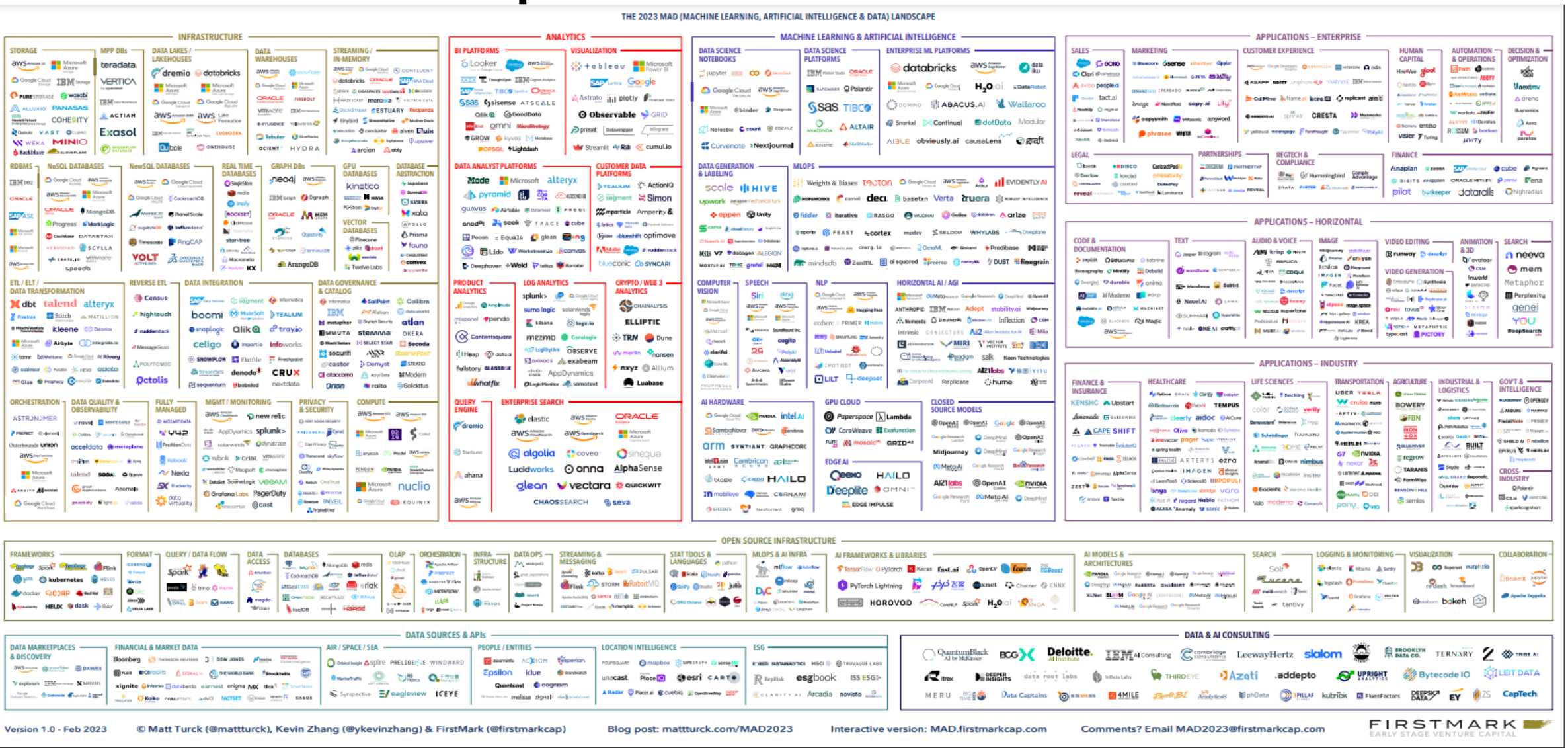
mattturck.com/data2021

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

FIRSTMARK EARLY STAGE VENTURE CAPITAL

DTIM www.essi.upc.edu/dtim

The Data Landscape (2023)



A New Business Model

- Traditionally databases have been seen as a passive asset
  - OLTP systems: Data gathered is structured to facilitate (automate) daily operations
  - The relational model as *de facto* standard
- Soon, many realized data is a valuable asset for any organization. *Use* it!
  - Decisional systems: Stored data is analysed to better understand our activity and make decisions out of them (I want to know and then objectively do)
  - As of today, decisional systems can be divided in two big groups:
    - To perform descriptive statistical analysis of data (aka as *business intelligence*)
    - To perform predictive (e.g., predict or classify) analysis of data (aka as *data science* and typically powered by *Big Data*)

Disclaimer: many people abuse terminology and refer to *ANY* kind of decisional system as *Artificial Intelligence*

"WITHOUT DATA,
YOU'RE JUST
ANOTHER PERSON
WITH AN OPINION"

W. Edwards Deming, American Statistician

Instagram's Fable (xkcd.com)

A New Business Model

- Hello good afternoon. Renato Pizza?
- *No sir, this is Google Pizza.*
- Excuse me, I'll have the wrong number...
- *No sir, Google has bought and renamed it.*
- Oh perfect! Well I would like to order.
- *Very good, Mr. López. The usual order?*
- The usual? Mr. López? Do you know me?
- *According to our caller ID, the last 12 times, you have ordered an individual Quattro Formaggio.*
- Exactly, that's what I want.
- *Can I suggest you try this time our vegetable pizza with ricotta, eggplant, zucchini and dried tomato?*
- No thanks. I hate vegetables.
- *Yeah, but it would be better for your cholesterol whose level is not very good.*
- Excuse me? How do you know that?
- *Through your subscription to the Online Medical Guide, we see your blood tests of the last 5 years.*
- But I do not like that pizza, I hate vegetables. Also, I'm being treated and taking the right medication.
- *Mr. López, you know that you do not take medication regularly, 5 months ago you bought a box of 30 pills at Otero García Pharmacy, and you didn't buy more...*
- That's not true, I bought more at another pharmacy.
- *Well, it does not appear on your credit card statement...*
- Because I paid in cash.
- *Well, according to your balance, you have hardly any cash in your pocket...*
- I have cash at home.
- *Seriously? Well, you have not declared it in your last income declaration... recognizing that you declare less than you earn? That is a crime, Mr. López.*
- But, WHAT DO YOU HAVE ...?! Enough! I'm sick of Google, Facebook, Twitter, WhatsApp, Instagram... I'm going to a deserted island without Internet, where there are no phones, and nobody can spy me!

- *I understand, gentleman. But remember that you must renew your passport, it expired three months ago*

Brainstorm! How Big Data / Data Science Affect your Life?

- Let us discuss!
  - What other examples beyond Google / Facebook can you think of?
  - Can you think of other scenarios like the ones discussed in our day by day?
  - If you are having a hard time trying to find examples, think of specific business domains. For example:
    - Insurances
    - Videogames
    - Entertainment platforms (e.g., Netflix)
    - Retail (physical shops and e-sites)
    - Collaborative economy (e.g., Glovo, Wallapop, etc.)
    - Fake news
    - …

  Can you think of innovative ways to collect and analyse data to make better decisions in these domains?

Data As The New Cornerstone

- We have witnessed the bloom of a new business model based on data analytics: *Data is not a passive but an active asset*
  - «Data is the new oil!» - Clive Humby, 2006
  - «No! Data is the new soil» - David McCandless, 2010
- Organization must adapt their infrastructures to benefit from the data deluge
  - Digital data doubling every 18 months (IDC)
- Innovation and entrepreneurship are mandatory!
- Some numbers:
  - Data Science is the sector with higher expected growth since 2014 in most developed economies (EU and USA)
  - Most optimistic forecasts measure a sustained growth world-wide of around 20% per year (the most pessimistic measure it around 12%)
- The European Commission has developed a data-driven economy strategy: Building a European Data Economy
  - In 2019, the data-driven economy whole impact on the economy was valued as 400 Billion Euro (EU+UK), with a yearly growth of 7,6% (much higher than the total IT sector of 4,9%)
  - The EU identified a gap of 459.000 unfilled positions in the EU27 + UK

The Technological Shake-Down

- The new business model fully relies on data
  - Data must be of quality, to guarantee good decisions
  - When learning through statistical inference, data must be representative, in order to guarantee a good learning
    - Thus, we must guarantee any relevant aspect related to the analysis at hands must be captured and represented in the available data
- As consequence, this new business model means collecting as much relevant data as possible, in whatever format
  - Large volumes
  - Disparate formats (data variety)
  - High ingestion rates / data freshness (data velocity)

Unfortunately, the traditional architecture of relational databases have been shown suboptimal when dealing with the three Vs features: volume, variety and velocity

Summary

- The irruption of Business Intelligence and Data Science (aka data-driven Artificial Intelligence) solutions as first-class citizens have rearranged the whole database landscape
- Although the use of buzzwords is nowadays trendy, few concepts and techniques are introduced. It's more about…
  - Choosing the proper solution for each scenario
  - *Simplifying* and rigorous thinking than about *new* solutions
- We will learn during the course how to understand this new technological scenario at the database level and learn how to know if a database might be appropriate for a project

Bibliography

- https://www.thesocialdilemma.com/
- The 451 Group: *The database landscape*. http://blogs.the451group.com/information_management/2012/11/02/updated-database-landscape-graphic
- Matt Turck: *Resilience and Vibrancy: The 2020 Data & AI Landscape*. https://mattturck.com/data2020/
- Michael Stonebraker: *One size fits all: an idea whose time has come and gone*. Communications of the ACM: 51(12) 2008
- Michael Stonebraker, Samuel Madden, Daniel J. Abadi, Stavros Harizopoulos, Nabil Hachem, Pat Helland: *The End of an Architectural Era (It's Time for a Complete Rewrite)*. Proceedings of VLDB 2007, pp.1150-1160
- Michael Stonebraker: *New opportunities for New SQL*. Communications of the ACM: 55(11) 2012
- Maria Belén Bianchi: *Els beneficis de l'ús de tecnologies NOSQL*. Tesi de màster (UPC, Juny 2012)