



Data Lake

Bootstrapping your Digitalization Journey

Big Data's 4 Vs



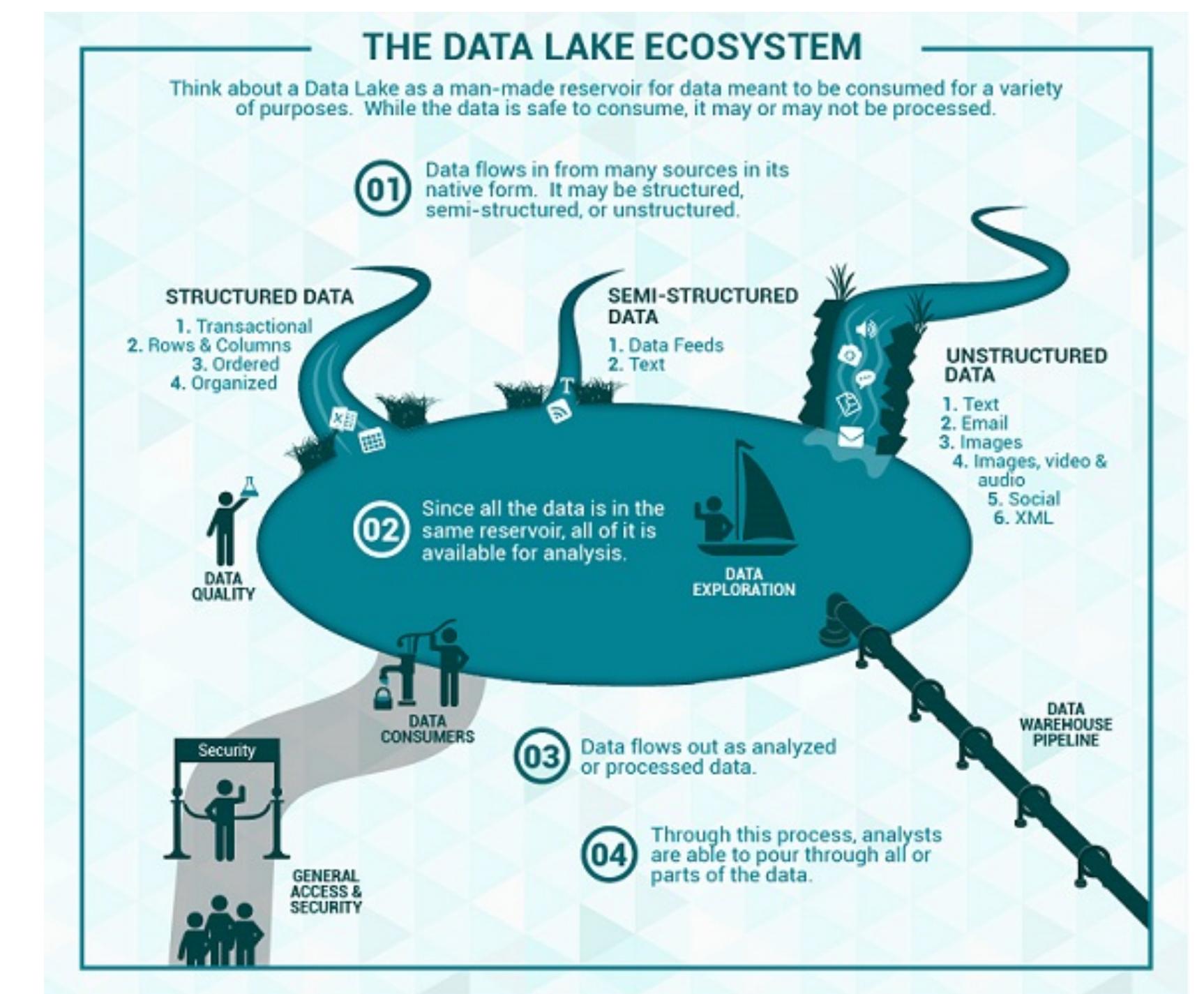
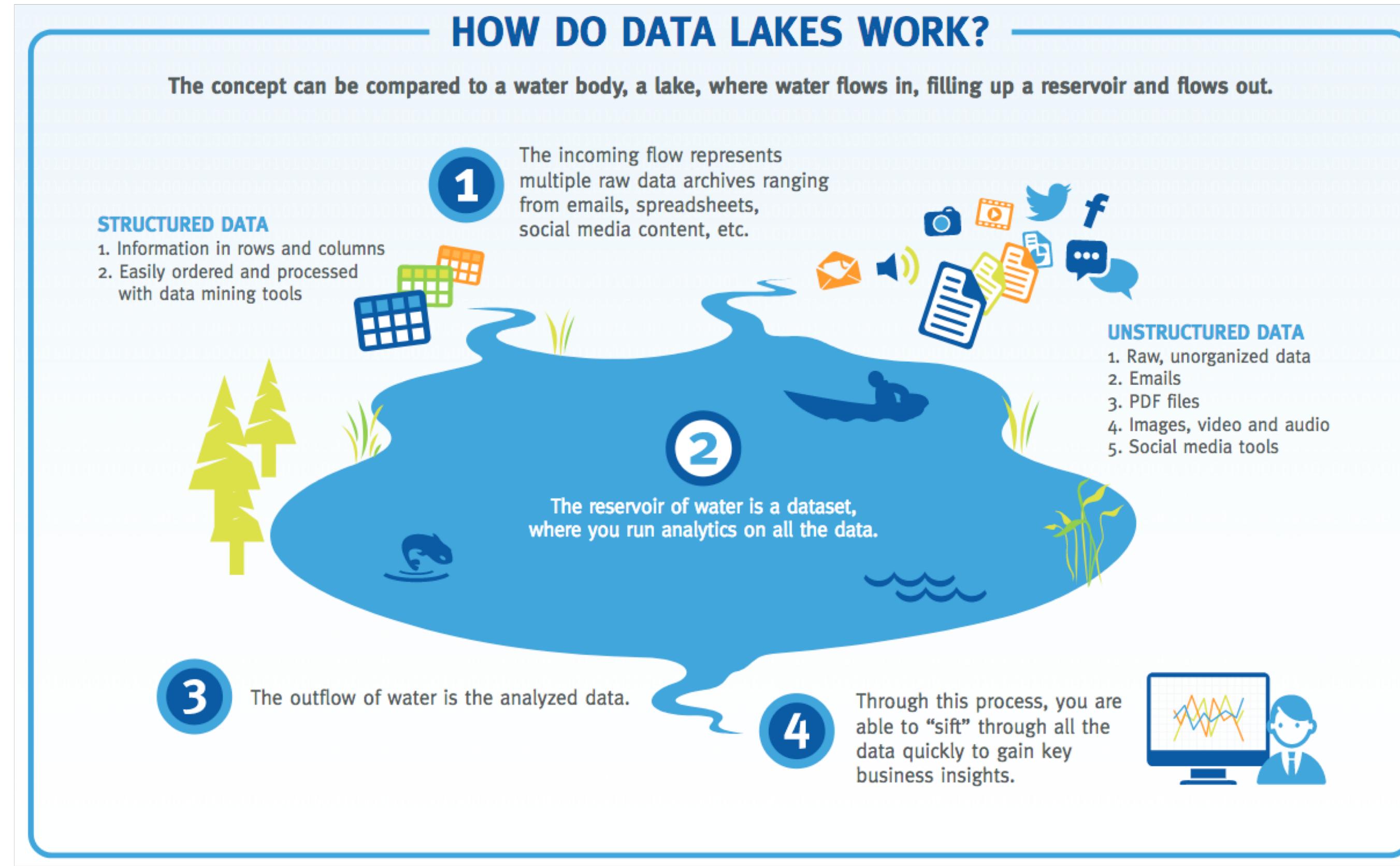
How shall I start the journey?

- ~~Easy! I'll go ahead and start building ML models, right? Wrong!~~
- Instead, start by...
 - Ingesting data
 - Cleaning data
 - Integrating data
- In most companies, this actually represents **75%** of the work
- Only then can you make the last **25% (analytics)** successful
- Build a **data lake** to tame your data first!

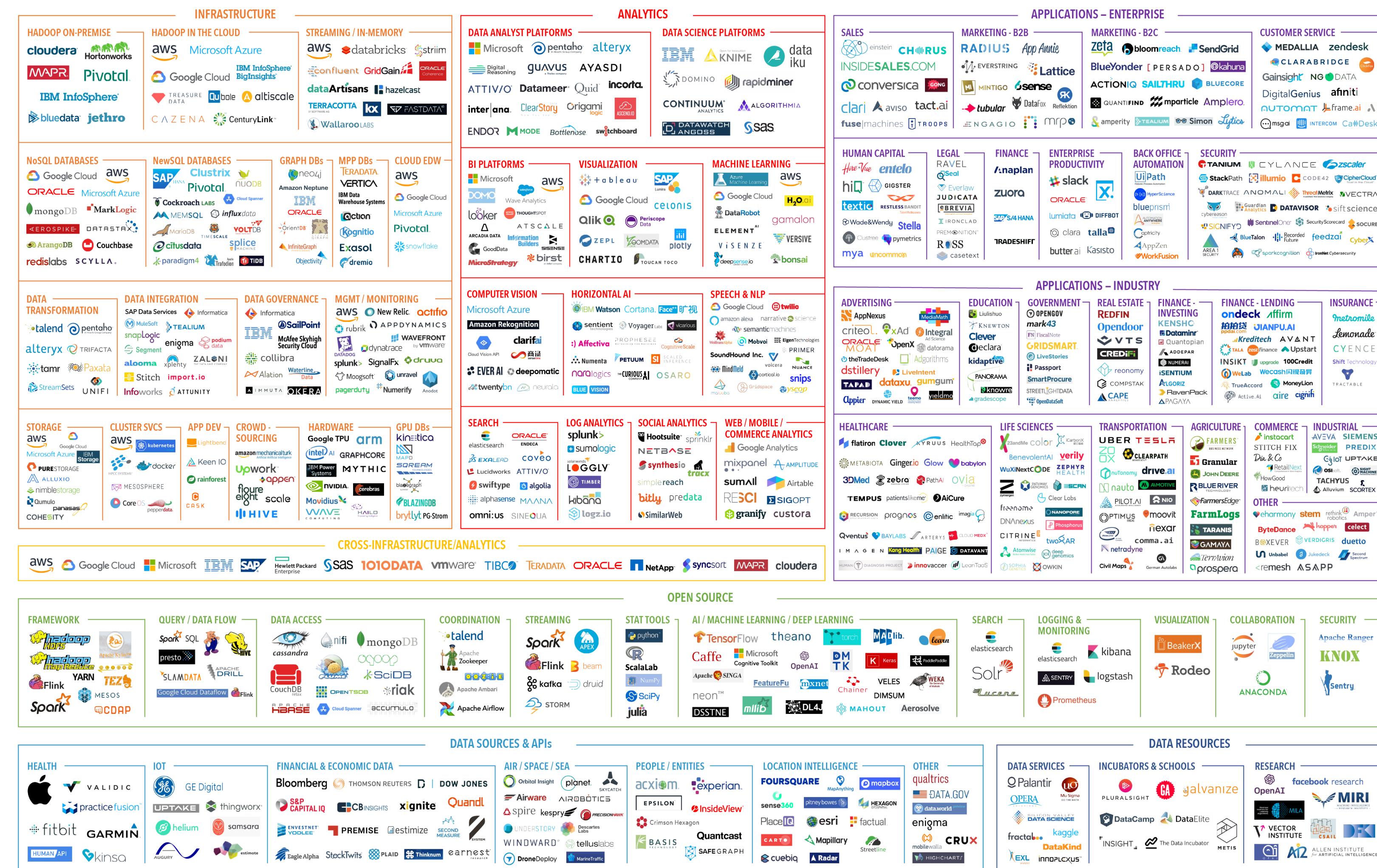
© P. Cudre-Mauroux <https://exascale.info>

What is a Data Lake?

Conceptually: solution to **co-locate** and share data in its native form



The 2018 Big Data landscape



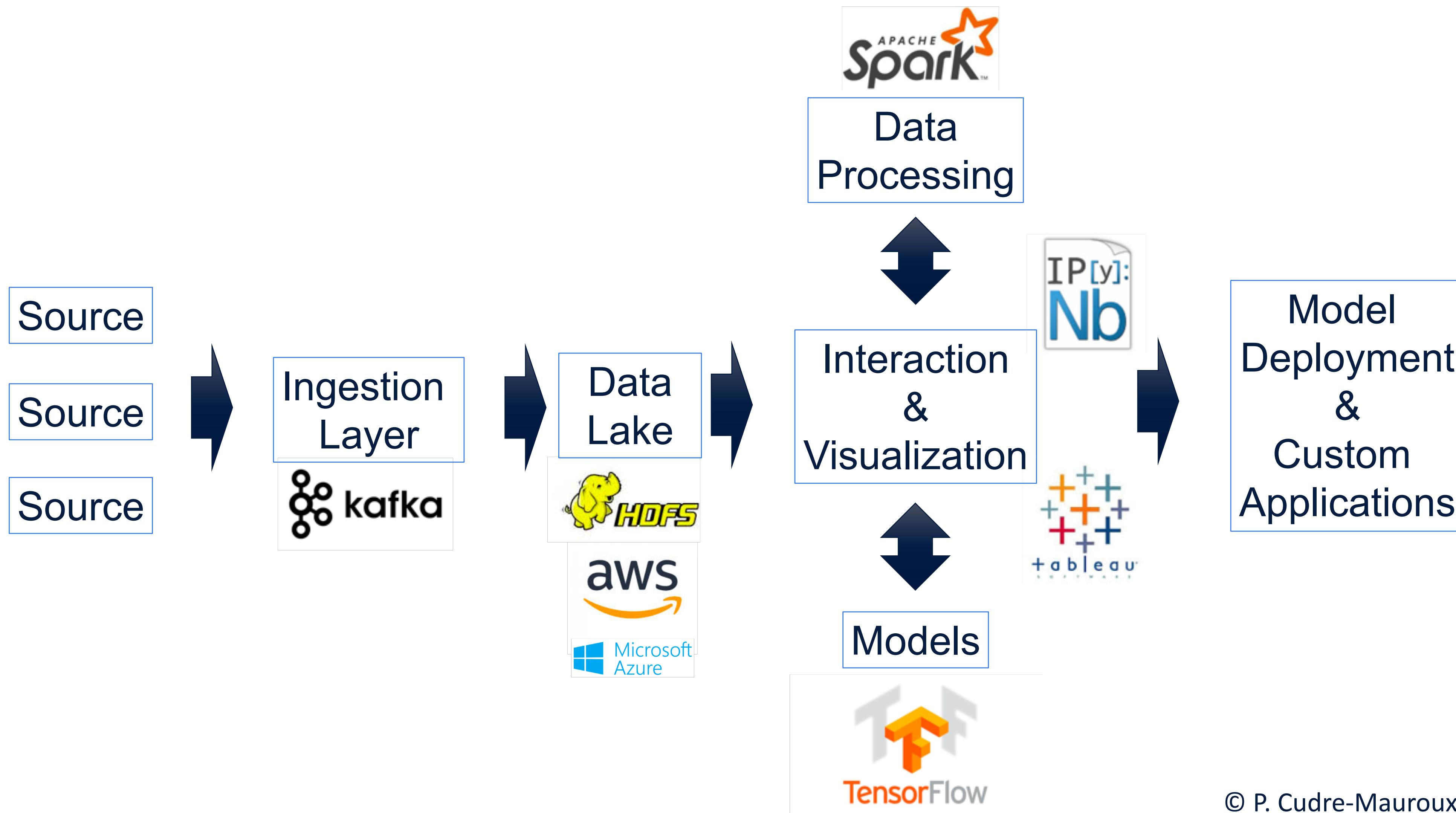
Final 2018 version, updated 07/15/201

© Matt Turck (@mattturck), Demi Obavomi (@demi_ obavomi), & FirstMark (@firstmarkcap)

mattturck.com/bigdata2018

FIRSTMARK 
EARLY STAGE VENTURE CAPITAL

Typical Big Data Architecture (2018)



© P. Cudre-Mauroux <https://exascale.info>

QUESTIONS?

