Data overload: Opportunities and Challenges

Yapi Donatien Achou

Semcon

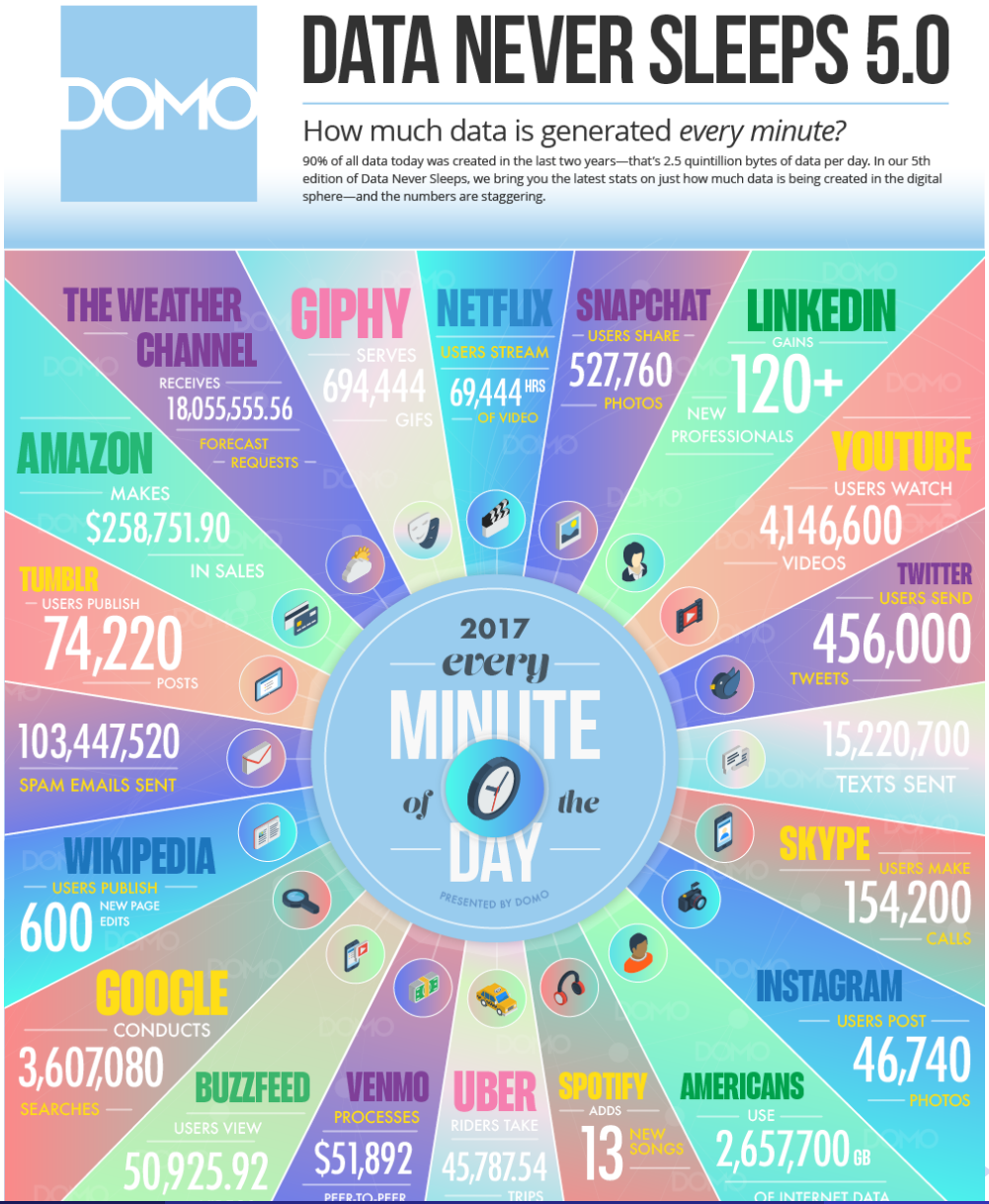
yapi-donatien.achou@semcon.com

August 4, 2021

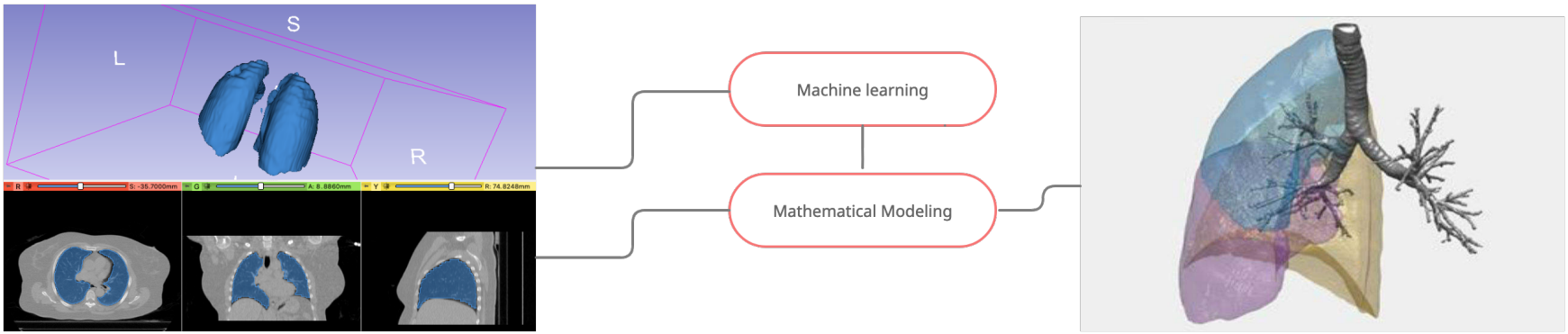
Overview

1. The ubiquity of data
2. Opportunities
3. Challenges
4. Workflow and tool

The ubiquity of data



Opportunities in healthcare



Accurately predict Cancer tumour movement by using

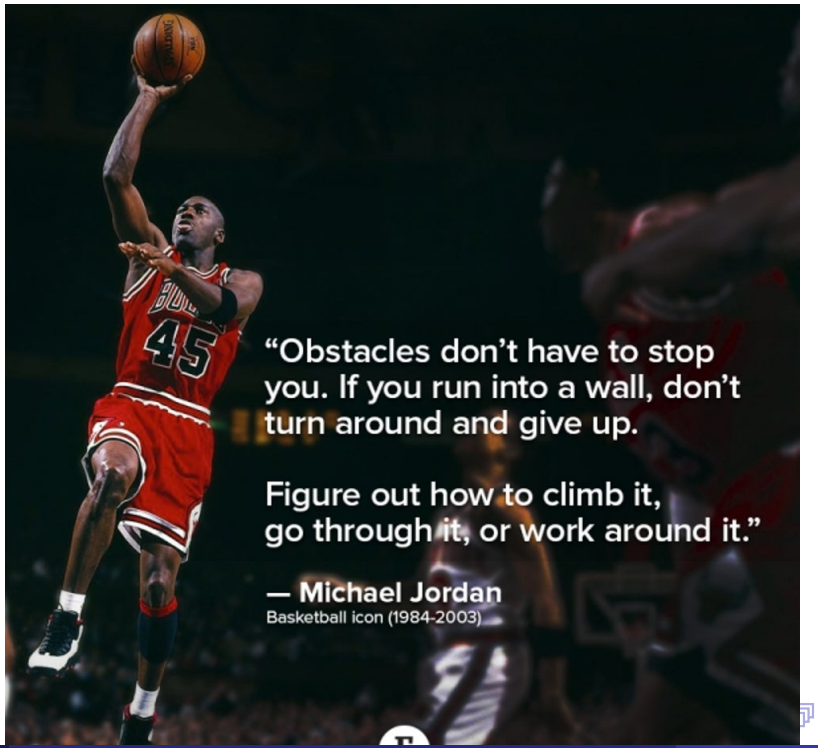
- Medical images (CT, 4DCT, MRI, . . .)
- External sensor data collected from a patient

Opportunities food industry

"We have the ability to not only look at a consumer as an individual and assess their buying patterns, but also look at the multiple consumers residing within a household, understand who is the dominant buyer, who reacts to our coupons, and, foremost, understand how they react to the channel that they're coming to us on," says Djuric.

Challenges

- Thousand of data sources to extract information from
- Data quality issue
- From Gigabits to Terabits of data to process and analyse
- Data privacy concern
- Data quality concern
- A **scalable**, **secure** , **well governed** and **reliable** data management platform



Data pipeline

The data journey from its source to the value that it brings:

- Data ingestion ✔
- Data storing ✔
- Data quality assessment ✔
- data modelling ✔
- Data transformation
- Data analytics

While tracking Data Lineage: Set of transformation that data goes through

Data pipeline

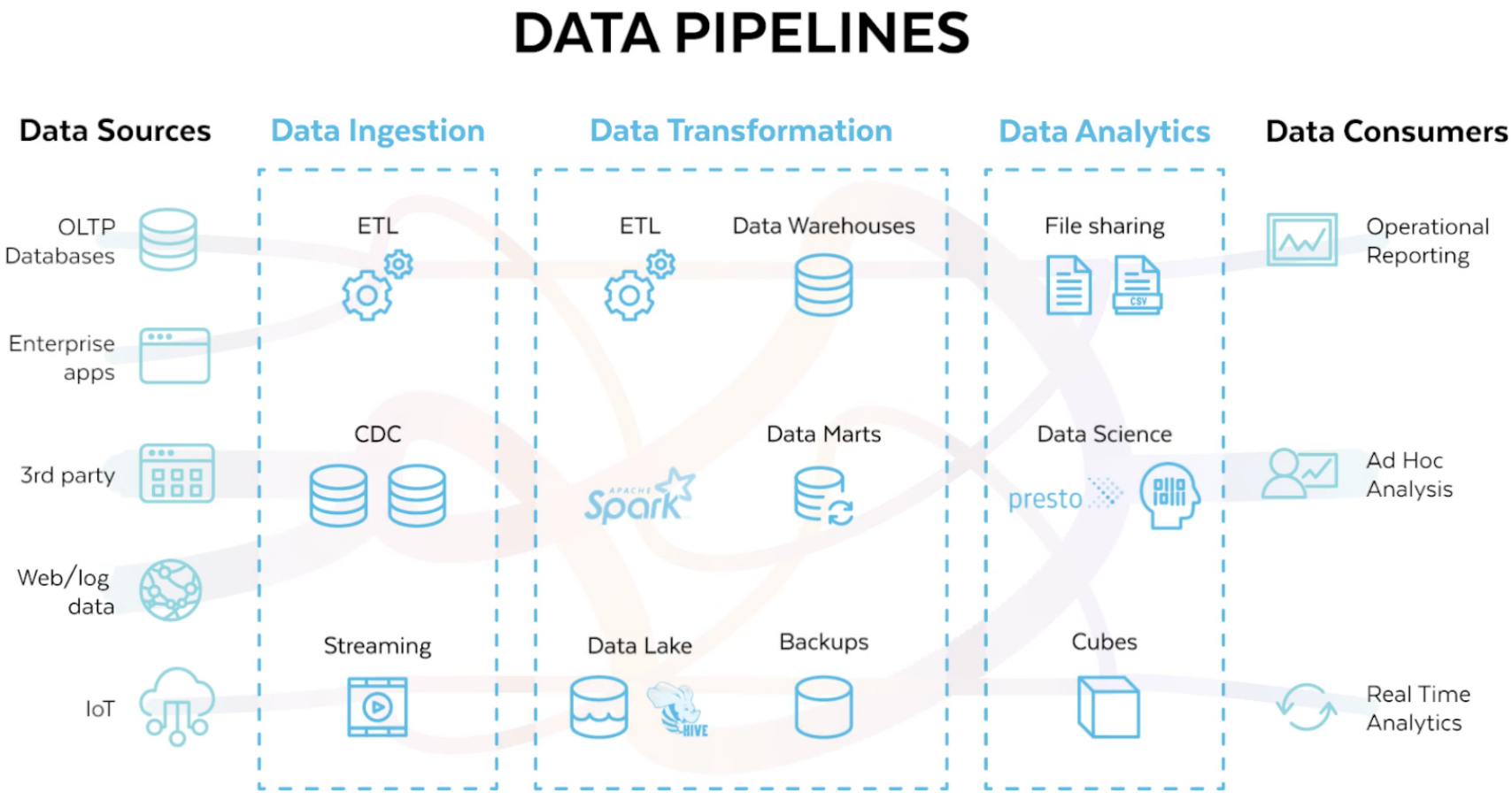
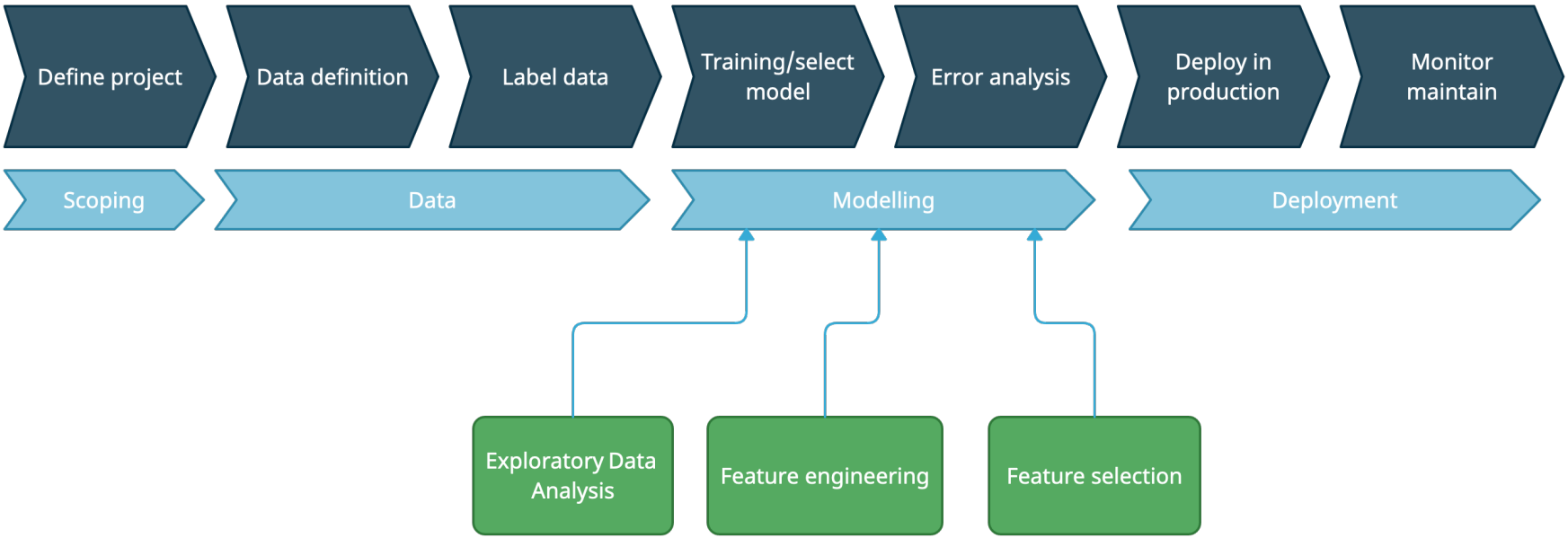


Figure: Cortesie https://www.snowflake.com/guides/etl-pipeline

Tools

Data ingestion

Data storage

Data science

Machine learning pipeline