

Data overload: Opportunities and Challenges

Yapi Donatien Achou

Semcon

yapi-donatien.achou@semcon.com

July 30, 2021

Overview

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

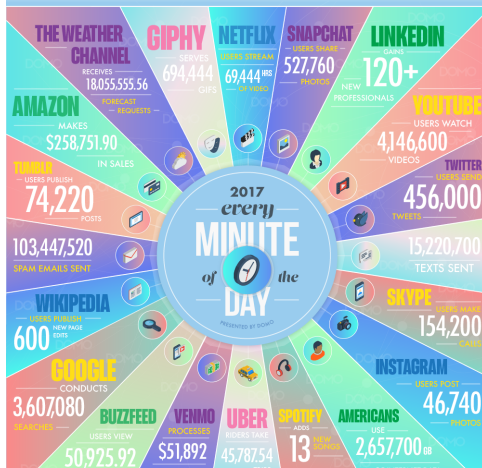
The ubiquity of data



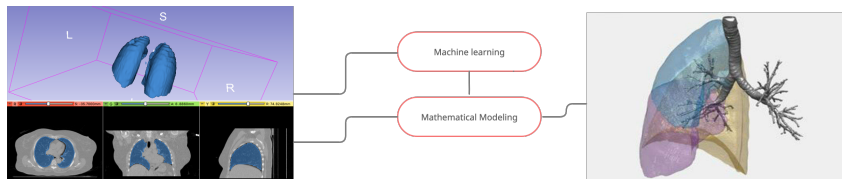
DATA NEVER SLEEPS 5.0

How much data is generated *every minute*?

90% of all data today was created in the last two years—that's 2.5 quadrillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.



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
- 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 workflow

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

- Data ingestion
- Data storing
- Data transformation
- Data analytics

While tracking [Data Lineage](#): Set of transformation that data goes through

Data pipeline workflow

DATA PIPELINES

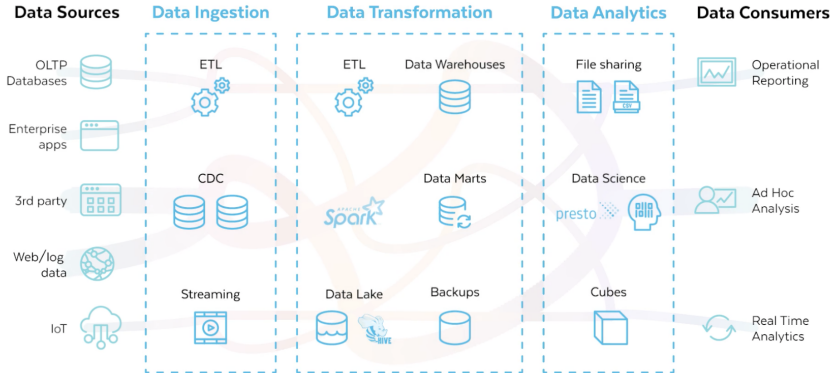
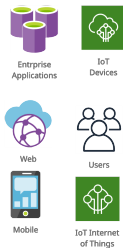


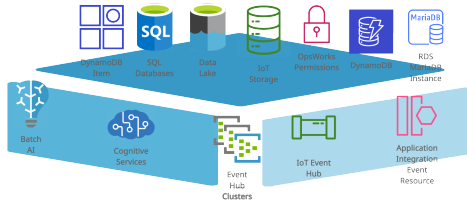
Figure: Cortesie <https://www.snowflake.com/guides/etl-pipeline>

Data Management Platform

Data Sources



Data Management Platform



Data Consumers

