UE Large Scale Processing @ ESIGELEC
2019/2020

01 – What is “Large Scale Processing”?

Abdel Dadouche
DJZ Consulting

adadouche@hotmail.com
@adadouche

Let’s Google “Large Scale Processing”

- Most hits refers to GDPR : General Data Protection Regulation
- “RGPD” in French

“The aim of the GDPR is to protect all EU citizens from privacy and data breaches in today’s data-driven world”.

This is done by enforcing new rules and rights:

- *Increased Territorial Scope*
- *Penalties*
- *Consent*
- *Breach Notification*
- *Right to Access & to be Forgotten*
- *Data Portability*
- *Privacy by Design*
- *Data Protection Officers*

- More information on https://eugdpr.org

“Large Scale Processing”, just a new word for Big Data

Let’s check "Big Data" from Wikipedia (1/3)

- *"Big data" is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.*

- *Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source.*

- *Big data was originally associated with three key concepts:* ***volume, variety, and velocity****.*

- *Current usage of the term big data tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.*

Source: https://en.wikipedia.org/wiki/Big_data

Let’s check "Big Data" from Wikipedia (2/3)

- *Data sets grow rapidly, in part because they are increasingly gathered by cheap and numerous information-sensing Internet of things devices such as mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks.*

- *The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5 exabytes (2.5×1018) of data are generated.*

- *Based on an IDC report prediction, the global data volume will grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts there will be 163 zettabytes of data.*

Source: https://en.wikipedia.org/wiki/Big_data

Let’s check "Big Data" from Wikipedia (3/3)

- *Relational database management systems, desktop statistics and software packages used to visualize data often have difficulty handling big data.*
- *The work may require "massively parallel software running on tens, hundreds, or even thousands of servers".*
- *What qualifies as being "big data" varies depending on the capabilities of the users and their tools, and expanding capabilities make big data a moving target.*
- *For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration.*

Source: https://en.wikipedia.org/wiki/Big_data

Summary

- A lot of buzz last year due to GDPR
- Large Scale Processing ~ Big Data
- Different needs addressed by Big Data (capture, store, analyze…)