Supports pour le Traitement de Données

Francieli ZANON BOITO

francieli.zanon-boito@u-bordeaux.fr

ENSEIRB-MATMECA + Université de Bordeaux

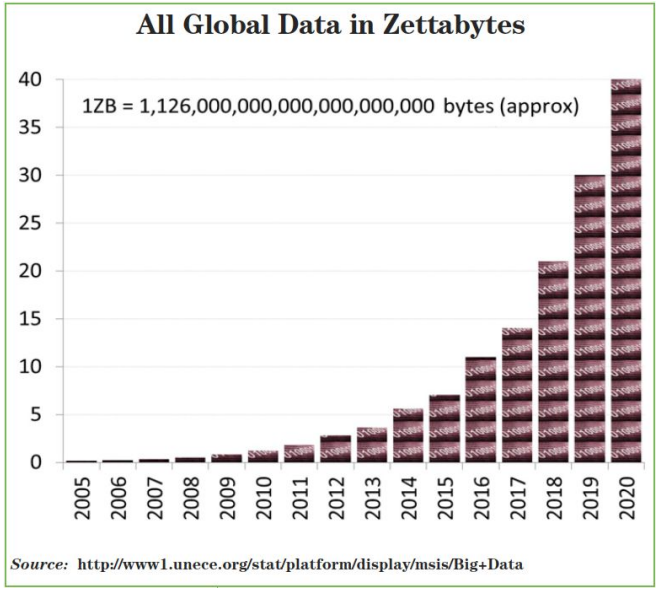
October 2021

Context : data access

	1999	2019	Speed-up 1999 ⇨ 2019
Storage (bytes/s)	~50 MB/s (HDD)	6,4 GB/s (SSD NVMe)	x 128
Processing (flops)	332 MFlops	44 TFlops	x 132000

Context : data access

	1999	2019	Speed-up 1999 ⇨ 2019
Storage (bytes/s)	~50 MB/s (HDD)	6,4 GB/s (SSD NVMe)	x 128
Processing (flops)	332 MFlops	44 TFlops	x 132000



Level	Latency
Register	~0.1 **ns**
L[1,2,3] cache	1~50 **ns**
DRAM memory	~100 **ns**
SSD	10~100 **µs**
Hard disk	~10 **ms**

This module

- Some tools we use to manipulate large amounts of data
- A little bit of history
- We'll talk about
  - MapReduce
  - Spark
  - NoSQL databases
- The goal is **NOT** to become masters of the presented technologies
  - I want you to understand the main ideas behind them

Practical stuff

- We'll use the Plafrim *formation* cluster
- I test it before each class
  - It sometimes (often) stops working between classes
    - Just send me an e-mail
- Don't leave important files in your HDFS space
- You can install your own "single-node cluster" (in your own computer)
  - It will be good for everything except performance measurements

Evaluation

1. Reports (2 or 3) - 90%
   - Individual (you can work in groups but each person has to write their own report)
   - .pdf file
   - Maximum length: 3 pages
   - French or English
   - In your own words, very direct
   - A summary of the things you learned in the class, and about the practical session
   - No code, just a description of the solutions (our focus: the ideas, not the technology)
2. Activities on moodle (QCM) - 10%
   - Unlimited tries
   - To be done outside of class

Beware

Always pay attention to the deadlines

Don’t copy anything from the paper, from the Internet, or from your colleagues