

Big Data

Open Data

MMI 2 – TP#1 S4



The background features a dark gray field with two large, bright geometric shapes: a yellow triangle on the left and a pink triangle on the right, both pointing towards the center.

Danielo **JEAN-LOUIS**
Développeur front-end

But du cours

- Sensibilisation aux concepts de Big Data
- Exploitation de données ouvertes
- Initiation au langage Python + Outil Anaconda
- Data visualisation : Python & Javascript

Qu'est-ce que le Big Data ?

*Le Big Data (ou mégadonnées) sont des informations à volume élevé, à grande vitesse et / ou à grande variété qui nécessitent des formes de traitement de l'information rentables et innovantes qui **permettent d'améliorer la compréhension, la prise de décision et l'automatisation des processus.***

Gartner

Sources :

- <https://www.gartner.com/en/information-technology/glossary/big-data> – anglais
- <https://www.oracle.com/fr/big-data/what-is-big-data>

Big Data

- Mégadonnées en français
- Concept pensé dans les années 50
- Exploitation massive de données
 - Au-delà du TB (Terabytes. 1 Tb = 1 000 Gb)
 - Facebook génère ~100 Tb/jour / Twitter ~8Tb

Unités de la donnée

Unités de bits v · d · m						
Système international (SI)			Préfixes binaires (CEI)			Ordre de grandeur
Unité	Notation	Valeur	Unité	Notation	Valeur	
bit	bit	1 bit	bit	bit	1 bit	1
kilobit	kbit ou kb	10^3 bits	kibibit	Kibit (ou Kb, par usage)	2^{10} bits	10^3
mégabit	Mbit ou Mb	10^6 bits	mébibit	Mibit	2^{20} bits	10^6
gigabit	Gbit ou Gb	10^9 bits	gibibit	Gibit	2^{30} bits	10^9
térabit	Tbit ou Tb	10^{12} bits	tébibit	Tibit	2^{40} bits	10^{12}
pétabit	Pbit	10^{15} bits	pébibit	Pibit	2^{50} bits	10^{15}
exabit	Ebit	10^{18} bits	exbibit	Eibit	2^{60} bits	10^{18}
zettabit	Zbit	10^{21} bits	zébibit	Zibit	2^{70} bits	10^{21}
yottabit	Ybit	10^{24} bits	yobibit	Yibit	2^{80} bits	10^{24}

Sources :

- https://fr.wikipedia.org/wiki/Bit#Ensembles_ordonn%C3%A9s_de_bits

3-5-7 V

- Volume – Echelle des données
- Variété – Diversité des données
- Vitesse – Vitesse de données
- Véracité – Exactitude des données
- Variabilité – Changement de la forme des données
- Visualisation – Visualiser les données pour mieux les comprendre
- Valeur – Tirer du profit de ces données

Sources :

- <https://impact.com/marketing-intelligence/7-vs-big-data/> - anglais
- <https://le-datascientist.fr/les-10-v-du-big-data>
- <https://www.oracle.com/fr/big-data/what-is-big-data>
- <https://le-datascientist.fr/6v-bigdata>

Big Data - Source de données

- Navigation web (clics, actions...)
- E-mails
- Images / vidéos / audios
- IoT (Internet of Things, objets connectés)
- Messages
- [...]
- Vous

Big Data - Source de données



Sources :

- <https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/>

**Vous produisez de la donnée,
vous êtes en train d'en produire**

- Où vous regardez
- Votre façon de vous asseoir
- La disposition des fenêtres des logiciels
- Votre parcours pour aller à l'IUT
- ...

Modèle de données

- Structurée
 - SGBDR (ex : MySQL) / Tableur
- Semi-structurée
 - Fichier JSON / XML / log
- Non structurée (pas de schéma)
 - Vidéos / Images...

Sources :

- <https://www.astera.com/fr/type/blog/structured-semi-structured-and-unstructured-data/>
- https://en.wikipedia.org/wiki/Semi-structured_data – anglais
- https://fr.wikipedia.org/wiki/Informations_non_structur%C3%A9es

Le nouvel or noir

- Priorité des entreprises
 - Source de revenus potentielle conséquentes
- Priorité au sein de l'État – E. Macron 2018
- Secteur très porteur dans l'industrie
 - Data-engineer / Data-analyste / Gestionnaire de serveurs...

Sources :

- <https://www.lebigdata.fr/emmanuel-macron-big-data-priorite-europe>

Applications

131 %

C'est le rendement obtenus par les entreprises s'étant lancés de façon intelligente dans le big data.

- Insitut McKinsey 2014

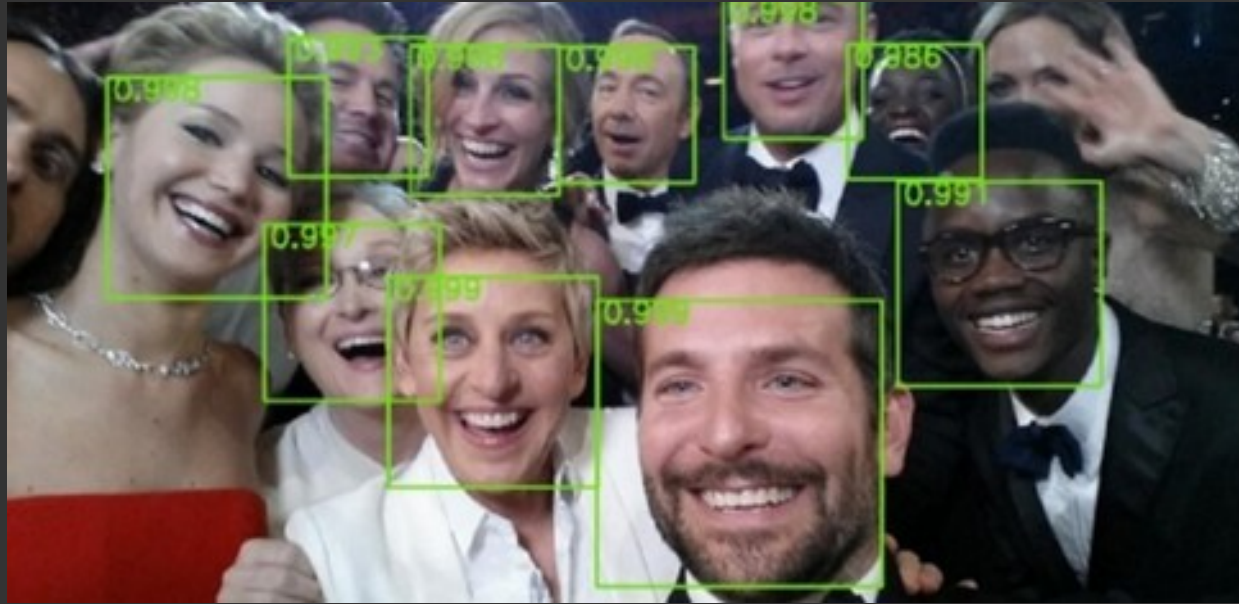
Big Data – Applications

- Gestion des stocks optimisée
- Prix dynamique sur un site d'e-commerce
- Recommandations
- Météorologie (Limité par la puissance de calcul)
- Détection de fraudes
- Médecine – Ex : Détection de cancers
- Sport
- ...

Sources :

- <http://www.01net.com/editorial/623742/le-big-data-le-douzieme-homme-de-lequipe-allemende-de-foot/>
- <https://digital.hbs.edu/platform-rctom/submission/paypals-use-of-machine-learning-to-enhance-fraud-detection-and-more/> - anglais

Big Data – Applications



La reconnaissance faciale : Un exemple parmi tant d'autres mais le plus connu du big data

- Volume
- Variété
- Visualisation

Big Data – Applications



Assistant vocal : une application du big data
(ici Siri d'Apple)

- Volume
- Variété
- Vitesse

Big Data – Applications

Enter both words below, separated by a space.

LYNN *flextime*

Provided by reCAPTCHA™

Submit

Captcha : une application du big data

Big Data – Applications

**Et vous connaissez-vous des
application du Big Data ?**

Big Data – Enjeux / Défis

- Optimiser la logistique
- Stocker les données
- Récolter des données de qualités
 - Big Data : 80% nettoyage / 20 % d'actions
- Éthique de la donnée
- Sécurité de la donnée
- Sensibiliser les employés (surtout les chefs)

Questions ?

Big Data Open Data

MMI 2 – TP#1 S4






Danielo **JEAN-LOUIS**
Développeur front-end

But du cours

- Sensibilisation aux concepts de Big Data
- Exploitation de données ouvertes
- Initiation au langage Python + Outil Anaconda
- Data visualisation : Python & Javascript



Qu'est-ce que le Big Data ?

*Le Big Data (ou mégadonnées) sont des informations à volume élevé, à grande vitesse et / ou à grande variété qui nécessitent des formes de traitement de l'information rentables et innovantes qui **permettent d'améliorer la compréhension, la prise de décision et l'automatisation des processus.***

Gartner

Sources :

- <https://www.gartner.com/en/information-technology/glossary/big-data> – anglais
- <https://www.oracle.com/fr/big-data/what-is-big-data>

Big Data

- Mégadonnées en français
- Concept pensé dans les années 50
- Exploitation massive de données
 - Au-delà du TB (Terabytes. 1 Tb = 1 000 Gb)
 - Facebook génère ~100 Tb/jour / Twitter ~8Tb

Unités de la donnée

Système international (SI)			Préfixes binaires (CIE)			Ordre de grandeur
Unité	Notation	Valeur	Unité	Notation	Valeur	
bit	bit	1 bit	bit	bit	1 bit	1
kilobit	kbit ou kb	10^3 bits	kibibit	Kibit (ou Kb, par usage)	2^{10} bits	10^3
mégabit	Mbit ou Mb	10^6 bits	mebibit	Mibit	2^{20} bits	10^6
gigabit	Gbit ou Gb	10^9 bits	gibibit	Gibit	2^{30} bits	10^9
téraoctet	Tbit ou Tb	10^{12} bits	tebibit	Tbit	2^{40} bits	10^{12}
pétabit	Pbit	10^{15} bits	pebibit	Pbit	2^{50} bits	10^{15}
exabit	Ebit	10^{18} bits	exbibit	Ebit	2^{60} bits	10^{18}
zettabit	Zbit	10^{21} bits	zebibit	Zbit	2^{70} bits	10^{21}
yottabit	Ybit	10^{24} bits	yobibit	Ybit	2^{80} bits	10^{24}

Sources :
 - https://fr.wikipedia.org/wiki/Bit#Ensembles_ordonn%C3%A9s_de_bits

3-5-7 V

- Volume – Echelle des données
- Variété – Diversité des données
- Vitesse – Vitesse de données
- Véracité – Exactitude des données
- Variabilité – Changement de la forme des données
- Visualisation – Visualiser les données pour mieux les comprendre
- Valeur – Tirer du profit de ces données

Sources :

- <https://impact.com/marketing-intelligence/7-vs-big-data/> - anglais
- <https://le-datascientist.fr/les-10-v-du-big-data>
- <https://www.oracle.com/fr/big-data/what-is-big-data>
- <https://le-datascientist.fr/6v-bigdata>

Big Data - Source de données

- Navigation web (clics, actions...)
- E-mails
- Images / vidéos / audios
- IoT (Internet of Things, objets connectés)
- Messages
- [...]
- Vous

Big Data - Source de données



Sources :
- <https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/>

**Vous produisez de la donnée,
vous êtes en train d'en produire**

- Où vous regardez
- Votre façon de vous asseoir
- La disposition des fenêtres des logiciels
- Votre parcours pour aller à l'IUT
- ...

Modèle de données

- Structurée
 - SGBDR (ex : MySQL) / Tableur
- Semi-structurée
 - Fichier JSON / XML / log
- Non structurée (pas de schéma)
 - Vidéos / Images...

Sources :

- <https://www.astera.com/fr/type/blog/structured-semi-structured-and-unstructured-data/>
- https://en.wikipedia.org/wiki/Semi-structured_data – anglais
- https://fr.wikipedia.org/wiki/Informations_non_structur%C3%A9es

Le nouvel or noir

- Priorité des entreprises
 - Source de revenus potentielle conséquentes
- Priorité au sein de l'État – E. Macron 2018
- Secteur très porteur dans l'industrie
 - Data-engineer / Data-analyste / Gestionnaire de serveurs...

Sources :

- <https://www.lebigdata.fr/emmanuel-macron-big-data-priorite-europe>

Les données "non structurées" étaient initialement prévues pour l'être humain, un ordinateur de base ne fait que lire un fichier audio, il ne sait pas par qui il a été chanté tout seul

Applications

131 %

C'est le rendement obtenu par les entreprises s'étant lancés de façon intelligente dans le big data.

- Insitut McKinsey 2014

Big Data – Applications

- Gestion des stocks optimisée
- Prix dynamique sur un site d'e-commerce
- Recommandations
- Météorologie (Limité par la puissance de calcul)
- Détection de fraudes
- Médecine – Ex : Détection de cancers
- Sport
- ...

Sources :

- <http://www.01net.com/editorial/623742/le-big-data-le-douzieme-homme-de-lequipe-allemande-de-foot/>
- <https://digital.hbs.edu/platform-rctom/submission/paypals-use-of-machine-learning-to-enhance-fraud-detection-and-more/> - anglais

Les données "non structurées" étaient initialement prévues pour l'être humain, un ordinateur de base ne fait que lire un fichier audio, il ne sait pas par qui il a été chanté tout seul

Big Data – Applications



La reconnaissance faciale : Un exemple parmi tant d'autres mais le plus connu du big data

- Volume
- Variété
- Visualisation

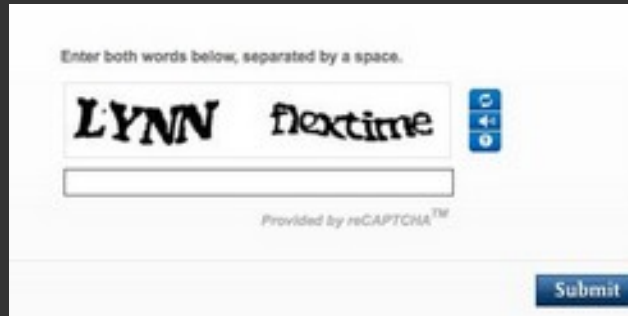
Big Data – Applications



Assistant vocal : une application du big data
(ici Siri d'Apple)

- Volume
- Variété
- Vitesse

Big Data – Applications



Captcha : une application du big data

Big Data – Applications

**Et vous connaissez-vous des
application du Big Data ?**

Big Data – Enjeux / Défis

- Optimiser la logistique
- Stocker les données
- Récolter des données de qualités
 - Big Data : 80% nettoyage / 20 % d'actions
- Éthique de la donnée
- Sécurité de la donnée
- Sensibiliser les employés (surtout les chefs)

Convaincre de l'utilité du bigdata auprès des chefs
Définir un pôle data
Comprendre que les résultats peuvent prendre du temps

Questions ?

