

# Outils pour le Big Data

---

MASTER SEP 2022-2023

[Bart.Lamiroy@univ-reims.fr](mailto:Bart.Lamiroy@univ-reims.fr)

# Big Data ?

---

Evolution technologique

Changement de culture

Opportunité d'évolution



Source: giphy

# Big Data pour quoi faire ?

---

Grande masse de données

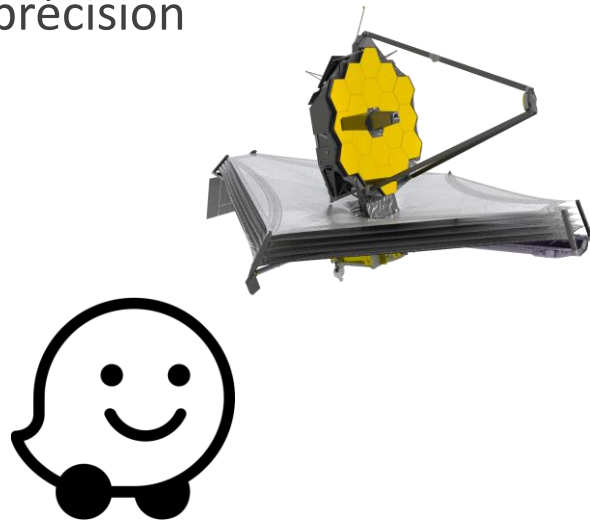
Meilleurs modèles

Meilleure précision

Grande masse de données

Personnalisation des données marketing

Meilleure accroche auprès des clients

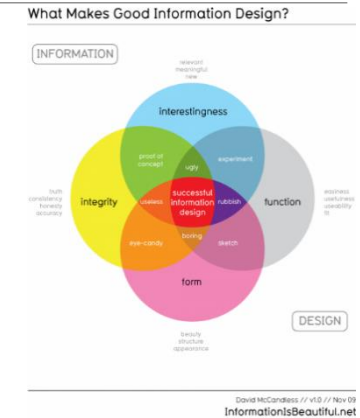
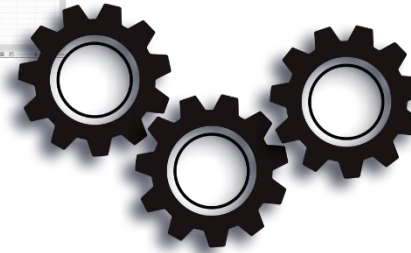
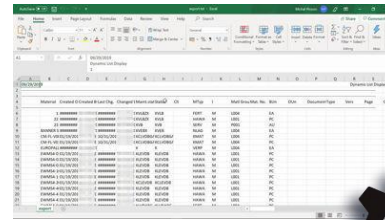


**amazon**

**NETFLIX**

**facebook**

# Analyse – Science des Données



# Logiciels (exemples)

---



Logos appartenant à leurs propriétaires respectifs

# Changement de culture

---

- Rendre accessibles les données
- Partager les données
- Open Data
  
- Risques de sécurité
- Risques éthiques
- Risques business
  
- RGPD

# Pilotage par la donnée

---

« le chiffre fait peur »

Différence fondamentale entre

- KPI
- Optimisation des indicateurs
- Culture du « chiffre »
- Pilotage par la donnée



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE

# Analyse de données *vs.* *Machine Learning*

---

## ***Analyse de données***

Mathématiques/informatique

Détection de structures, régularités, corrélations

Applications

- Analyse prédictive
- Classification
- Régression
- Réduction de dimensions
- Statistique inférentielle

## **Apprentissage Numérique**

Mathématiques/informatique

Construction de fonctions, algorithmes à partir d'exemples

Domaines

- Supervisé
- Non-supervisé
- *Deep Learning*

Applications

- Classification
- Intelligence artificielle



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE

# Corrélation n'est pas causalité !

---

*Les personnes avec un grand pouce maîtrisent mieux leur langue maternelle et parlent plus de langues étrangères.*

*L'application de crème solaire augmente le risque de mélanome*



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE

# Différence avec l'informatique décisionnelle

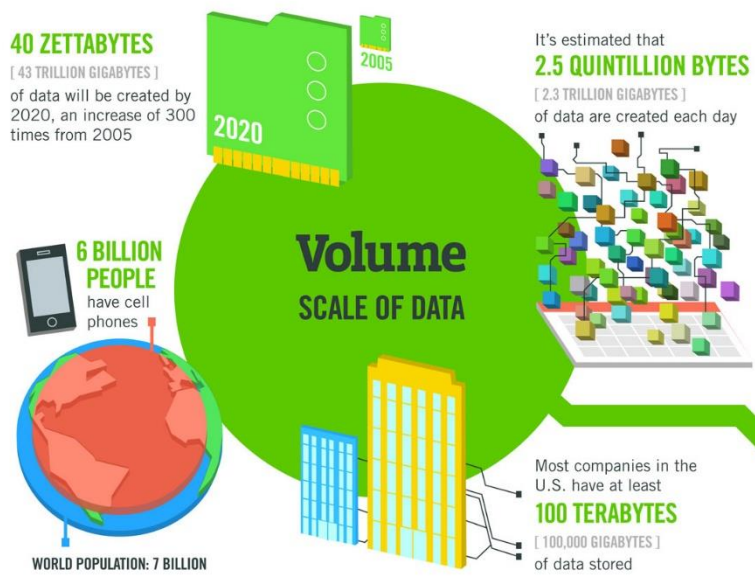
---

- Données structurées
- Données consolidées
- Données issus de processus connus/identifiés/maitrisés
  
- Analyse orientée question/réponse
- Processus analytique « humain » pré-établi

# Apport « Big Data »

---

- Découverte de structures non anticipées
- Remise en question de « certitudes »
- Capacité de gestion de données faiblement structurées
- Réactivité



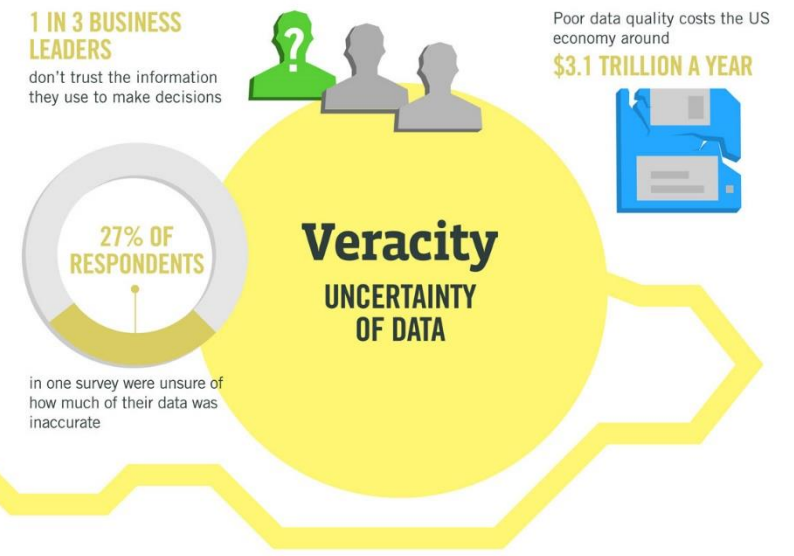
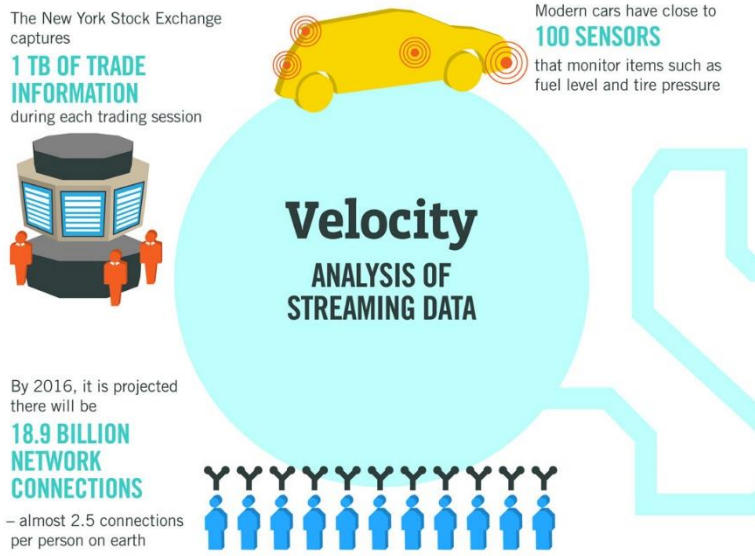
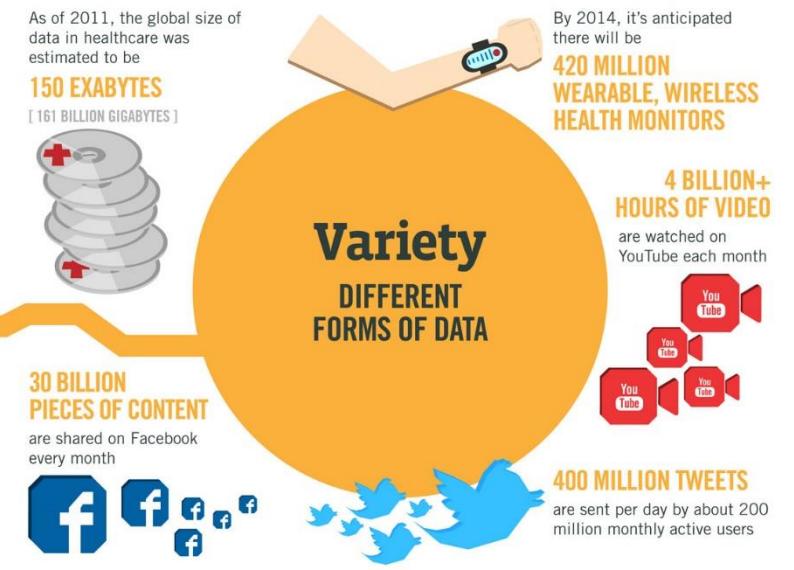
# The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS



# En 2021



Source: <http://www.domo.com>

# Exemple non-commercial

---

(source: LHC – CERN) Analyse de données issues d'une campagne de bombardement de particules à très grande énergie

- ~1PB/s (1PB =  $10^{15}$  octets)
- Préfiltrage 1/10 000 évènements = 10TB/s
- 1% du restant est considéré comme scientifiquement "intéressant" = 1GB/s
- Résultat = 25PB/an (les campagnes ne se font pas 24/7)

# Bases mathématiques de l'analyse de données

---

Régression linéaire

Analyse en composantes principales

Analyse factorielle des correspondances

Classification automatique

Analyse discriminante

Régression logistique



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE

# Bases algorithmiques du Machine Learning

---

*Clustering* hiérarchique

K-means

Espérance-Maximisation

Classification de Bayes naïf

Classification au p-proches voisins

Machines à vecteurs support (SVM)

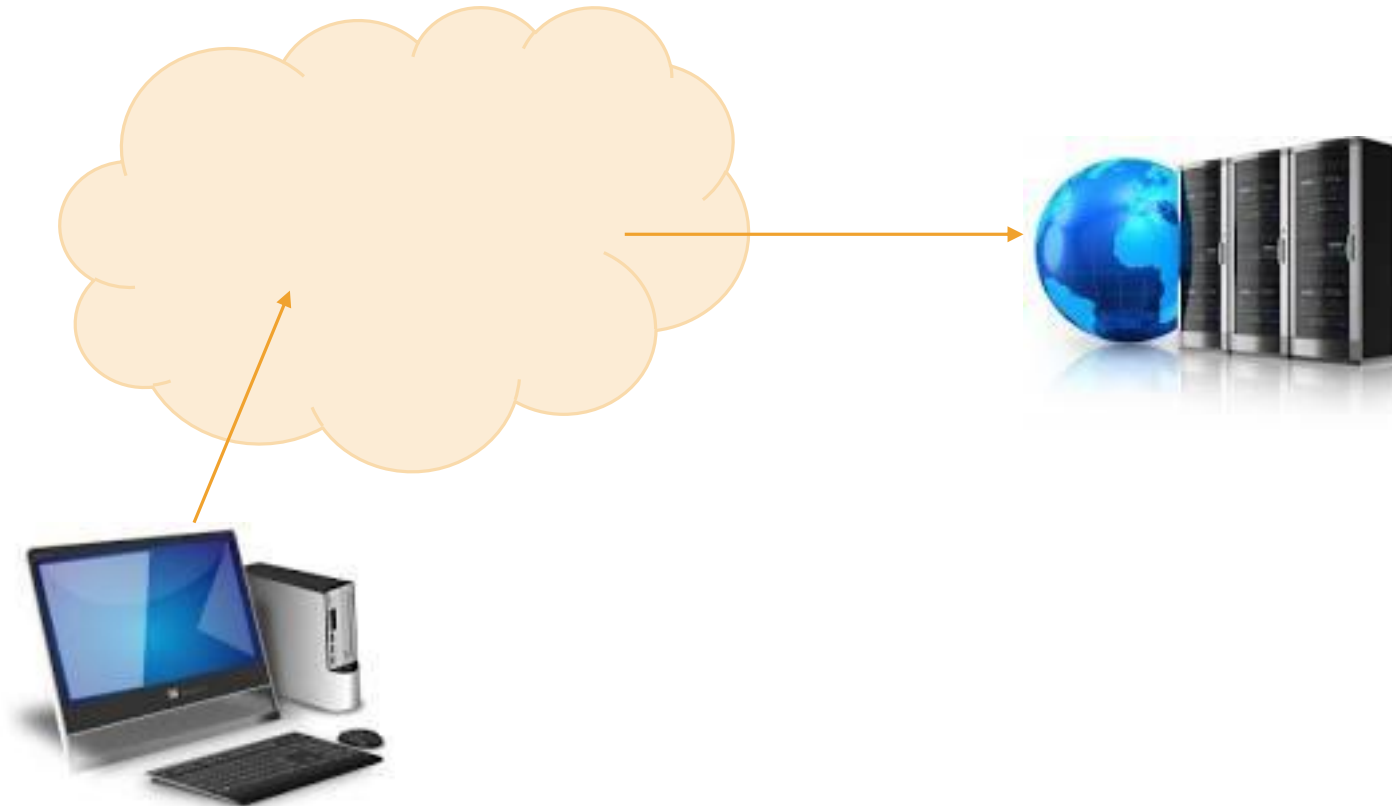
Réseaux de neurones

# Rapide introduction aux fondamentaux des architectures numériques et réseaux

---

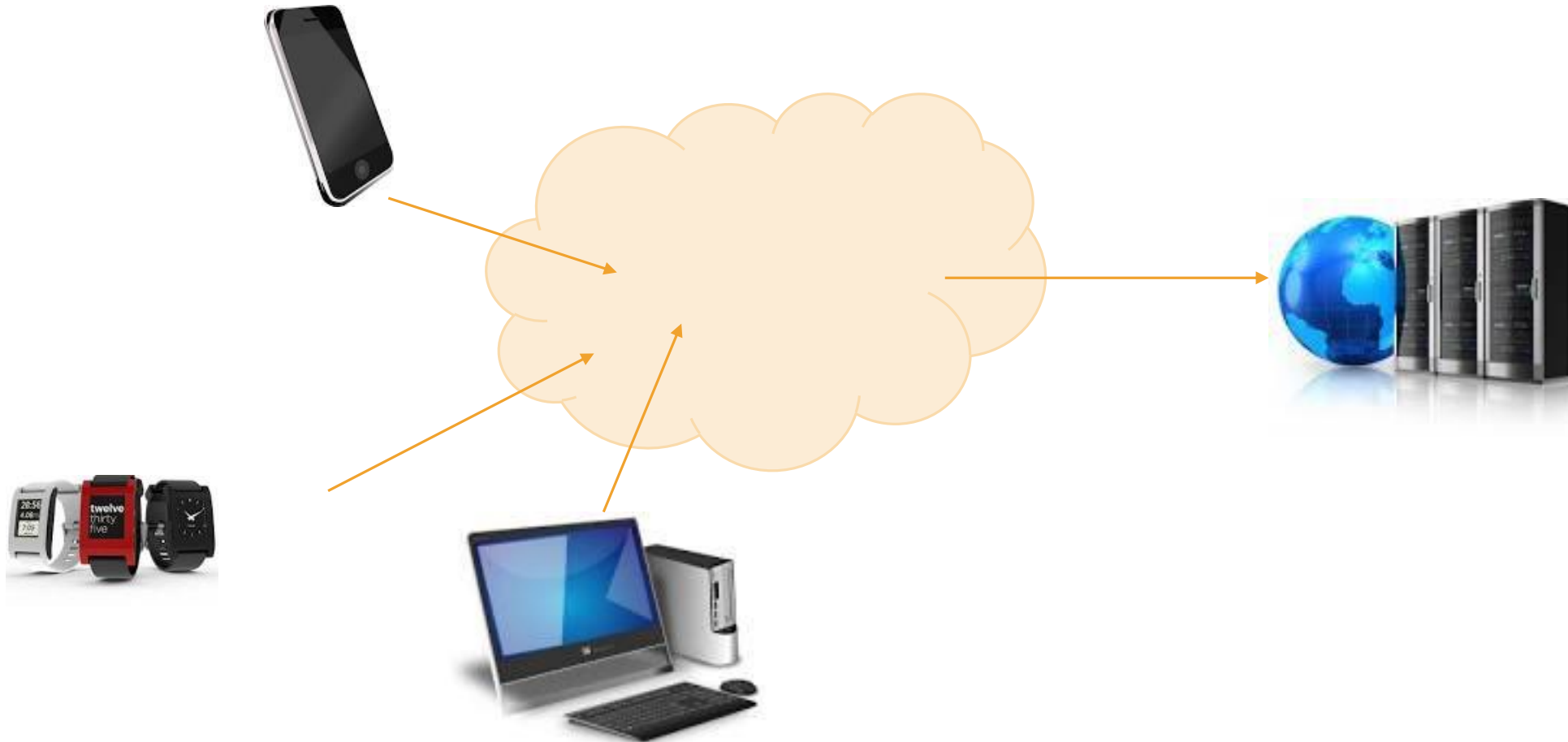
# Architecture générale

---



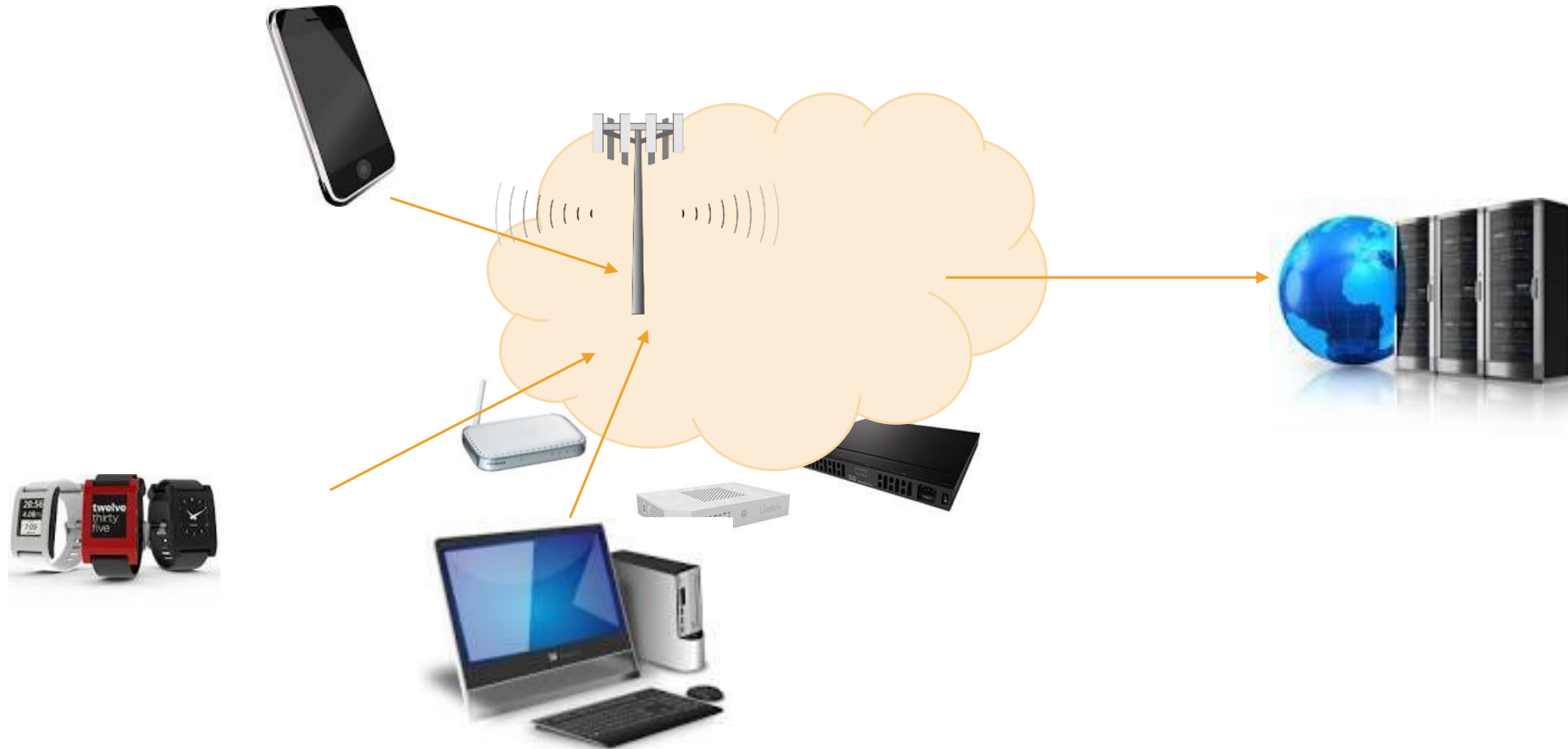
# Architecture générale

---



# Architecture générale

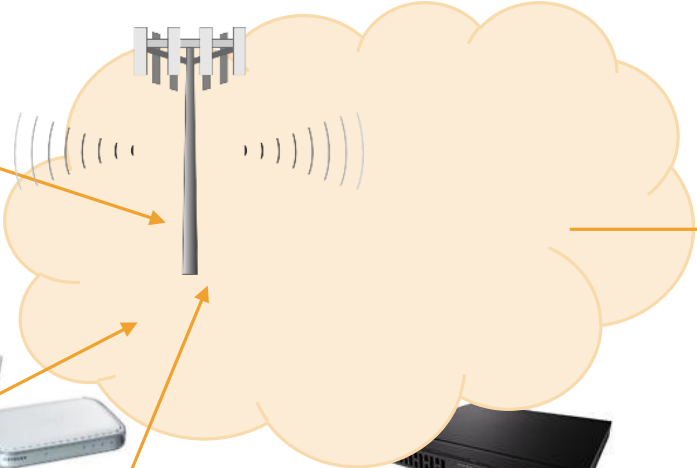
---



# Architecture générale



ORACLE



UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE



# Architecture générale



ORACLE



# Architecture générale

Applicatif



ORACLE

API



Back End

App

Microservice

Server Side

Client – Serveur

n-Tiers



Front End



Interface

Framework

Navigateur

REST

Intranet



UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE



Mobilité

Client Side



Client Léger



C'est trop  
complexe



# Informatique = gestion d'information

---

Grille de lecture à trois niveaux d'interaction

1. Stockage de l'information
2. Acheminement de l'information
3. Traitement de l'information



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE



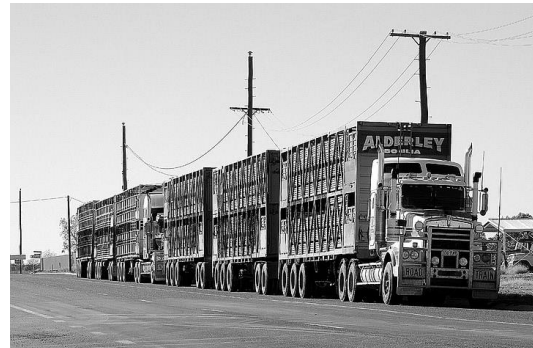
# Parallèle avec le monde physique

---

Stockage



Transport



Traitement



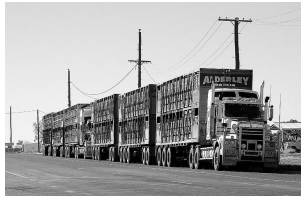
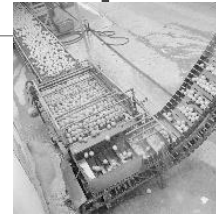
# Parallèle avec le monde physique

---



# Parallèle avec le monde physique

---



# Architecture client

---

Stockage

Transport

Traitement

# Architecture client

↳ Informatique > Ordinateur de bureau > PC de bureau



## PC de bureau HP HP PAVILION DESKTOPS 590-P0012NF I3-8100 4G 1TB W10



**429,99€** ~~549,99€~~ **VENTE FLASH - 21%**

Dont 1,44€ d'éco-part. DEEE

**3XCB** **4XCB** Payez en 3x ou 4x par carte bancaire

8 Neufs dès 429,99 €

Livré dès demain  
Livraison OFFERTE  
[Voir toutes les options de livraison](#)

Le Dimanche 22/12  
Retrait Darty Strasbourg Les Halles

AJOUTER AU PANIER

RETIRER EN MAGASIN

Le prix DARTY, service compris

- ✓ Garantie 2 ans
- ✓ Livraison offerte (voir conditions)
- ✓ 15 jours pour changer d'avis

**Ajouter les services Darty sans limite**  
(30 jours offerts puis 49,00 €/an). [Voir détails](#)

[Ajouter au comparateur](#) | [Ajouter à ma liste d'envies](#)



### Les points forts :

- Processeur Intel® Core™ i3-8100
- Carte graphique Intel UHD 630
- RAM 4 Go - 1 To HDD
- Windows 10 - HDMI - Wifi 802.11 ac - Bluetooth 4.0

[Voir toutes les caractéristiques](#)

DESCRIPTION

LA MARQUE VOUS PARLE

CARACTÉRISTIQUES

CONSEILS ET SERVICES

## CARACTÉRISTIQUES

### PROCESSEUR

Marque du processeur	Intel
Modèle du processeur	i3-8100
Nombre de coeur(s)	4
Fréquence	3,6 GHz

### MEMOIRE ET STOCKAGE

Mémoire vive (RAM) installée	4 Go
Type de mémoire vive (RAM)	DDR4
Type de disque dur	HDD
Capacité du disque dur HDD	1000 Go
Vitesse de rotation du disque dur	7200 tr/min

### PROCESSEUR GRAPHIQUE

Marque de la carte graphique	Intel
Modèle du processeur graphique	UHD Graphics 630

### CONNECTIQUE

Lecteur de cartes	3-en-1
Casque	1
Microphone	1
USB 2.0 (arrière)	4
Port(s) USB 3.1	4
VGA (D-sub)	1
HDMI	1
Réseau	1 x RJ45
Autres	1 USB 3.1 type C, 1 prise jack combinée casque/micro



# Architecture client

Informatique > Ordinateur de bureau > PC de bureau



## PC de bureau HP HP PAVILION DESKTOPS 590-P0012NF I3-8100 4G 1TB W10



**429,99€** ~~549,99€~~ **VENTE FLASH - 21%**

Dont 1,44€ d'éco-part. DEEE

**3XCB** **4XCB** Payez en 3x ou 4x par carte bancaire

8 Neufs dès 429,99 €

Livré dès demain  
Livraison OFFERTE

[Voir toutes les options de livraison](#)

Le Dimanche 22/12  
Retrait Darty Strasbourg Les Halles

**AJOUTER AU PANIER**

**RETIRER EN MAGASIN**

Le prix DARTY, service compris

- ✓ Garantie 2 ans
- ✓ Livraison offerte (voir conditions)
- ✓ 15 jours pour changer d'avis

**Ajouter les services Darty sans limite**  
(30 jours offerts puis 49,00 €/an). [Voir détails](#)

[Ajouter au comparateur](#) | [Ajouter à ma liste d'envies](#)



### Les points forts :

- Processeur Intel® Core™ i3-8100
- Carte graphique Intel UHD 630
- RAM 4 Go - 1 To HDD
- Windows 10 - HDMI - Wifi 802.11 ac - Bluetooth 4.0

[Voir toutes les caractéristiques](#)

DESCRIPTION LA MARQUE VOUS PARLE **CARACTÉRISTIQUES** CONSEILS ET SERVICES

## CARACTÉRISTIQUES

### PROCESSEUR

Marque du processeur	Intel
Modèle du processeur	i3-8100
Nombre de coeur(s)	4
Fréquence	3,6 GHz

Traitement

### MEMOIRE ET STOCKAGE

Mémoire vive (RAM) installée	4 Go
Type de mémoire vive (RAM)	DDR4
Type de disque dur	HDD
Capacité du disque dur HDD	1000 Go
Vitesse de rotation du disque dur	7200 tr/min

Stockage

### PROCESSEUR GRAPHIQUE

Marque de la carte graphique	Intel
Modèle du processeur graphique	UHD Graphics 630

### CONNECTIQUE

Lecteur de cartes	3-en-1
Casque	1
Microphone	1
USB 2.0 (arrière)	4
Port(s) USB 3.1	4
VGA (D-sub)	1
HDMI	1
Réseau	1 x RJ45
Autres	1 USB 3.1 type C, 1 prise jack combinée casque/micro

Transport



# Questions guides

---

Stockage : où sont mes données ?

Transport : comment sont elles acheminées

Traitement : où mes données sont-elles traitées ?



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE



# Principe général

**Interface  
Utilisateur**



**Client**



**Requête**



**Serveur**



**Front End**

**Réponse**



**Back End**

# Principe général



**HTML/CSS**  
**Javascript**



**HTTP**

**PHP**  
**Python**



**JSX**  
**ASP**  
**JSP**

**SQL**  
**NoSQL**



# Consultation des données

---

- Consommation/Consultation des données
- Production/Mise à disposition des données
- Contrôle d'accès aux données
- Mise en forme des données



# Version 1 : web statique



**Interprétation  
HTML**

**Requête HTTP**

**GET URL**

**HTML/CSS**



**Fichiers HTML**

# Version 2 : web dynamique (server side)



**Interprétation  
HTML**

**Requête HTTP**

**GET URL**

**HTML/CSS**

**PHP  
Python**



**JavaScript**

**SQL**



UNIVERSITÉ  
DE REIMS  
CHAMPAGNE-ARDENNE



# Version 3 : web dynamique (client side light)



# Version 4 : web dynamique *(full client side)*



**Exécution  
Javascript**

**Requête HTTP**

**GET URL**

**HTML/CSS  
Javascript**

**Requête HTTP**

**REST - WebSocket**

**PHP  
Python**



**JavaScript**

**SQL**

