

— université
— lumière
— LYON 2

Data face !



**Tout savoir sur
les données**

Définitions (dictionnaire Larousse)

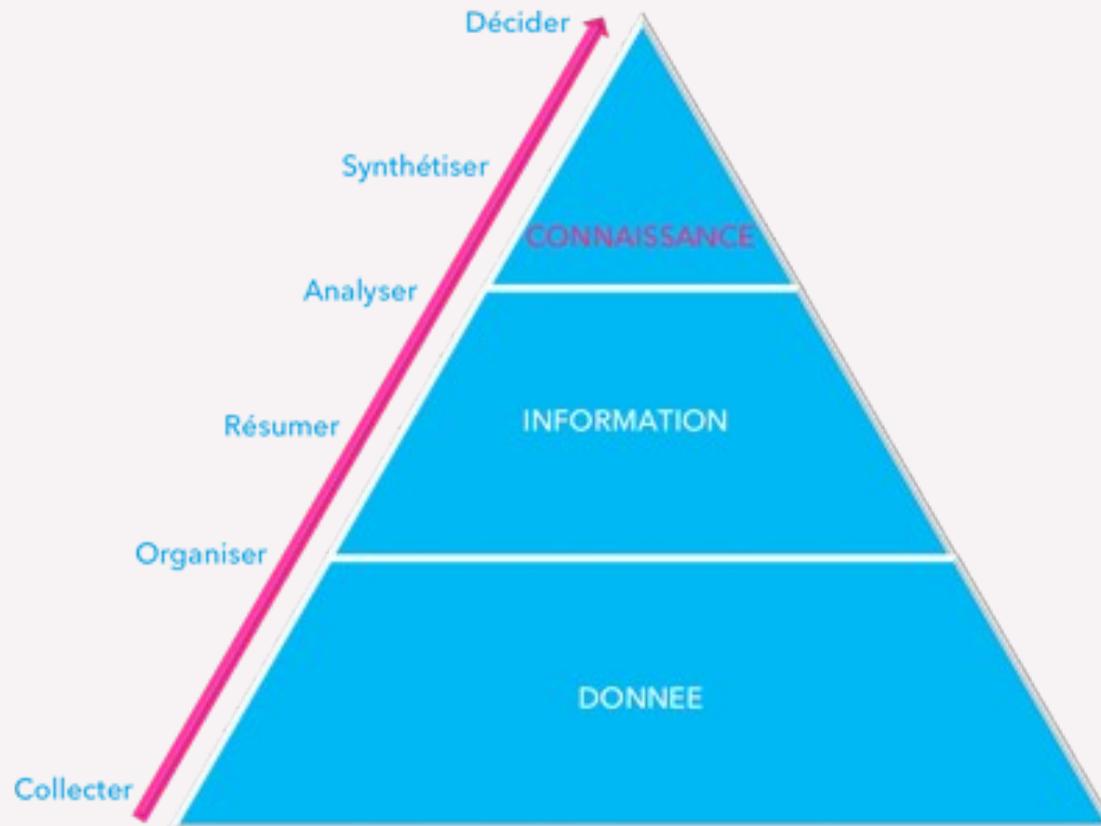
Ce qui est connu ou admis comme tel, sur lequel on peut fonder un raisonnement (...)

Dans un problème de mathématiques, hypothèse figurant dans l'énoncé.

Représentation conventionnelle d'une information en vue de son traitement informatique.

Résultats d'observations ou d'expériences faites délibérément ou à l'occasion d'autres tâches et soumis aux méthodes statistiques.

Donnée, information, connaissance



Information comprise

Donnée interprétée

Élément brut

www.semdee.com

Sondage express

- A) Jusqu'ici, tout va bien
- B) Je suis déjà perdu·e



Répondre sur <https://toreply.univ-lille.fr>

Question n° 427

Sommaire



Différentes facettes des données



Brève histoire de la gestion de données



Traitement des données, IA et compagnie

Principaux types de données



Type	Description
boolean, bool	Booléen (vrai/faux)
smallint, int2	Entier de -32768 à +32767
integer, int, int4	Entier de -2147483648 à +2147483647
bigint, int8	Entier de -9223372036854775808 à +9223372036854775807
real, float4	Réel simple précision (6 décimales)
double precision, float8	Réel double précision (15 décimales)
numeric/decimal, numeric/decimal(P, S)	Nombre jusqu'à 131072 chiffres avant la virgule (P), 16383 après (S)
character[N], char[N]	Chaîne de caractères de longueur N fixe
varchar, varchar[N]	Chaîne de caractères de longueur variable (maximum N si précisé)
text	Texte long de longueur variable (nécessite des opérateurs spécifiques)
date	Date (ex. 'jj/mm/aaaa')
time	Heure (ex. 'hh:mm:ss')
timestamp	Date et heure (ex. 'aaaa-mm-jj hh:mm:ss')



Données structurées

Nom	Prénom	Courriel	Affiliation
Darmont	Jérôme	jerome.darmont@univ-lyon2.fr	Université Lyon 2
Bentayeb	Fadila	fadila.bentayeb@univ-lyon2.fr	Université Lyon 2
Gruenwald	Le	ggruenwald@ou.edu	Oklahoma University



Données non structurées





Données semi-structurées



Langages de requête

Moteurs de recherche

N° employé	Date embauche	Nom	Prénom	Adresse	Ville	Province	Code postal	Date naissance
6080	15-jan-89	Charist	Eric	1075 rue Duath	Montréal	QC	H4T 2L3	05-sep-49
6181	20-mai-89	Roux	Kennedy	888 rue Lavigne	Donat	QC	H9P 3T2	05-oct-53
6150	05-mai-89	Langlois	Claude	368 rue Ste Pierre	Montréal	QC	H2P 1A0	03-mai-58
6181	14-oct-89	Beaudet	Jane	440 rue Fafard	Laval	QC	H7P 3Z8	14-oct-45
6160	07-mai-89	Furts	Jean-Paul	4380 rue Racicot	Montréal-Nord	QC	H2V 2P5	25-juil-59
6030	08-mai-89	Labrecque	Monique	682 rue Victoria	Lévesque	QC	J3P 2T2	02-mai-37
6024	12-avr-89	Therrien	Jacques	320 rue Verchères	Greenfield Park	QC	J4V 2V8	02-avr-71
6087	21-avr-89	Prusa	Manel	8885 St-Je-Basile	Montréal	QC	H3T 2V3	09-mai-56
6080	12-mai-89	Labranche	Andrée	99 rue du Collège	Vale St-Laurent	QC	H4P 2M9	05-juil-54
6034	04-juin-89	Comell	Mark	1219 rue de Gaule	St-Hubert	QC	J4T 3L9	28-juin-69
6080	19-juin-89	Royette	Alain	888 rue Curves	Outremont	QC	H3T 2V2	25-juil-47
6080	15-juin-89	Simard	Martin	2575 rue Victoria	St-Lambert	QC	J4L 8P7	03-juil-74
6028	15-juin-89	Beaudet	Sébastien	2520 rue Rowles	St-Théodore	QC	J7P 3V5	14-juin-83
6039	14-oct-89	Lesun	Josée	684 de Montanille	St-Benoit	QC	J4L 3Y4	12-sep-71
1182	17-oct-89	Foucher	Math	45 rue Havelock	Cardoux	QC	J2T 2W1	31-oct-54
1181	14-oct-89	Le	Juris	521 rue Camille	Laval	QC	H7P 3C7	09-juil-48
1241	07-jan-90	Le	Math	1111 rue St-Jacques	Montréal	QC	J3H 5P9	25-oct-59
1280	01-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	H3V 2V3	05-juil-58
1441	04-mai-90	Le	Math	1111 rue St-Jacques	Montréal	QC	J4P 1B2	13-mai-65
1460	05-mai-90	Le	Math	1111 rue St-Jacques	Montréal	QC	H3T 2P2	28-juil-65
1473	07-mai-90	Vaquin	Jean-Claude	2688 rue Fleury	Montréal	QC	H4H 1C5	12-oct-62
1481	01-août-90	Wasy	Marce	201 de LaSalle	Boucherville	QC	J4C 1T7	07-oct-68
1543	09-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	J4T 2V6	30-mai-66
1560	09-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	H7P 1R8	03-juin-65
1571	14-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	G01 2X0	17-oct-62
2000	14-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	J4P 1C3	28-oct-63
2312	04-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	H3L 5K3	03-juil-70
2325	04-juin-90	Le	Math	1111 rue St-Jacques	Montréal	QC	H4T 2Z2	28-oct-69
2329	04-juin-90	Hamel	Huberte	365 rue Hutchison	Outremont	QC	H2V 2T4	15-mai-61
2321	01-août-90	Mato	Bernard	8987 rue Viger	Montréal	QC	H3V 2M5	10-mai-73
2322	01-août-90	Axel	Gilbert	436 rue Dawson	Montréal	QC	H2L 6L9	10-oct-67
2340	01-août-90	Ouellet	Patricia	7228 rue St-Joseph	Montréal	QC	H2J 3P5	13-août-61
2341	09-juin-90	Achamault	Manon	673 rue Brébeuf	Montréal	QC	H4S 2T1	31-août-61
3089	04-juin-90	Brousseau	Guy	2575 Trisaland	Brossard	QC	J4P 2V3	26-juin-60
3480	04-mai-90	Lapointe	Mirella	682 St-Amand	Montréal	QC	H2T 6S2	26-juil-73
3500	04-mai-90	Piquette	Pierre	3687 Christophe Colomb	Montréal	QC	H3V 5M1	09-juil-72
3501	04-mai-90	Trudeau	Joseph	1300 rue Vintennes	Lévesque	QC	J3P 6T1	11-juil-72
3550	15-mai-90	Trimbou	Alexandra	524 ave. Dumont	Boucherville	QC	J4T 8M2	14-juin-69
3554	15-mai-90	Yves	François	2502 rue Paton	Montréal	QC	H4T 3M9	10-mai-65
3560	21-mai-90	Wester	Hans	52 rue Ste Marie	Verchères	QC	J5G 4X2	20-sep-72
3561	21-mai-90	Coppen	Eric	2220 rue Beaudet	Montréal	QC	H3P 2J5	05-oct-65
3580	09-août-90	Clemens	Marie-Clare	1275 rue Van Home	Outremont	QC	H2V 6P5	05-nov-58
3800	08-oct-90	Rotheval	Céline	7129 Edouard-Montpetit	Montréal	QC	H3V 6L3	29-jan-74

Années 1990



Données semi-structurées

LESTA* la nouvelle procédure d'autorisation de voyage pour aller ou transférer aux États-Unis ¶

De nouvelles formalités d'entrée et de transit sur le territoire américain vont entrer en vigueur dès le 12 janvier 2009. Cela concerne tous les voyageurs qui sont exemptés de visa. ¶

En Europe, ce sont les ressortissants de 22 pays bénéficiaires du Programme d'Exemption de Visa, dont la France, qui sont concernés par cette réforme aux quels s'ajoutent 5 pays asiatiques. ¶

Désormais, il faudra avant d'embarquer pour un voyage à destination des États-Unis ou avec un transit aux États-Unis, obtenir une autorisation de voyage "LESTA"*. ¶

Qui doit soumettre une demande d'autorisation de voyage ESTA ? ¶

A partir du 12 janvier prochain, avant de monter dans un avion ou d'embarquer à bord d'un paquebot à destination des États-Unis, tout voyageur devra remplir via internet le nouveau formulaire d'inscription en ligne, l'ESTA. ¶

Arretenir*: Cette formalité est obligatoire pour: ¶

TOIT-voyageur, qui'il se rende sur le territoire américain ou qu'il y transite. ¶

qui se soit pour un séjour de tourisme ou affaires de moins de 90 jours. ¶

qui il soit majeur ou mineur (donc même pour votre bébé de 8 mois et 1 par enfant), accompagné ou non. ¶

quel que soit le mode de transport, à bord d'une compagnie maritime ou aérienne. ¶

L'autorisation ESTA est un visa. ¶

Non, l'autorisation "ESTA" n'est pas un visa. Elle mesure un contrôle supplémentaire préalable afin de faciliter la tâche des officiers d'immigration pour l'identification des voyageurs à risque. ¶

L'autorisation de voyage ESTA assure-t-elle l'admission sur le territoire américain ? ¶

Ce n'est pas automatique. En effet, tous les candidats à l'ESTA, y compris les membres de la famille, doivent remplir les conditions d'admission à l'entrée aux États-Unis. ¶

Arretenir*: Cette formalité est obligatoire pour: ¶

Dans tous les cas, l'officier d'immigration au poste frontière se prononce sur l'admission comme c'est déjà le cas aujourd'hui. ¶

Quelle est la procédure d'obtention d'une autorisation ESTA ? ¶

Le système "ESTA" utilise Internet uniquement. ¶

Pour soumettre votre demande d'autorisation ESTA depuis le 1er août 2009, vous devez vous rendre sur le site suivant: <https://esta.cbp.dhs.gov> et suivre les instructions pour répondre aux questions posées. ¶

L'accès au site est gratuit et disponible en français (sélectionnez pour cela la langue française dans le menu-derrière en haut à droite). ¶

Arretenir*: ¶

Si vous n'avez pas accès à Internet, vous devez recourir à une tierce personne de votre entourage ou à un agent de voyage. Vous restez légalement responsable des réponses fournies. ¶

La plupart des voyageurs obtiendront la confirmation de l'autorisation de façon quasi-immédiate ce qui permet les voyages de dernière minute. ¶



Données semi-structurées

Nom		Jérôme Darmont
Courriel		jerome.darmont@univ-lyon2.fr
		jerome.darmont@msh-lse.fr
Affiliation		Université Lyon 2
Courriel		fadila.bentayeb@univ-lyon2.fr
Nom	Prénom	Fadila
	Nom de famille	Bentayeb
Affiliation		Université Lyon 2
Nom		Gruenwald
Prénom		Le
Affiliation		Oklahoma University

Qui pèse le plus lourd ?

Données
non structurées



Données
structurées et
semi-structurées





Big (mais pas que) data

Volume

Taille des données

Véracité

Qualité des données
Sécurité



www.lebigdata.fr

Variété

Données (semi-)structurées
et non structurées

Vélocité

Capteurs, réseaux sociaux

Données opérationnelles vs. données décisionnelles



Données opérationnelles en évolution

Image de la base en Mai 2005

Répertoire

Nom	Ville
Dupont	Paris
Durand	Lyon

Image de la base en Juillet 2006

Répertoire

Nom	Ville
Dupont	Marseille
Durand	Lyon

Données décisionnelles historisées

Calendrier

Code	Année	Mois
1	2005	Mai
2	2006	Juillet

Répertoire

Code	Année	Mois
1	Dupont	Paris
1	Durand	Lyon
2	Dupont	Marseille

Schémas de données

Lydie Soler, AgroParisTech/INRA

Quizz

Les *big data* peuvent être ?

- A) Structurées
- B) Semi-structurées
- C) Non structurées
- D) Opérationnelles
- E) Décisionnelles



Répondre sur <https://toreply.univ-lille.fr>

Question n° 278

Sommaire



Différentes facettes des données



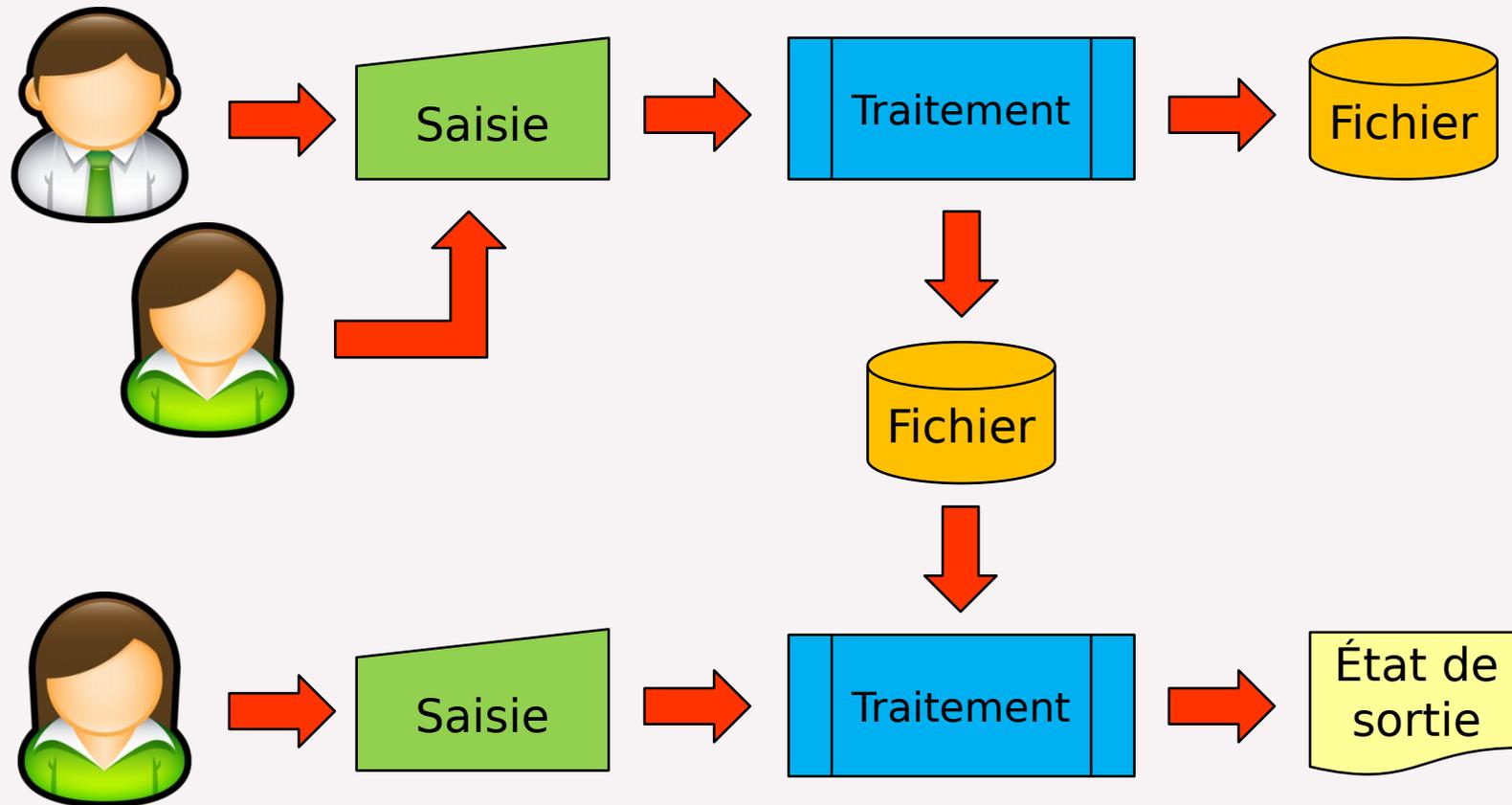
Brève histoire de la gestion de données



Traitement des données, IA et compagnie

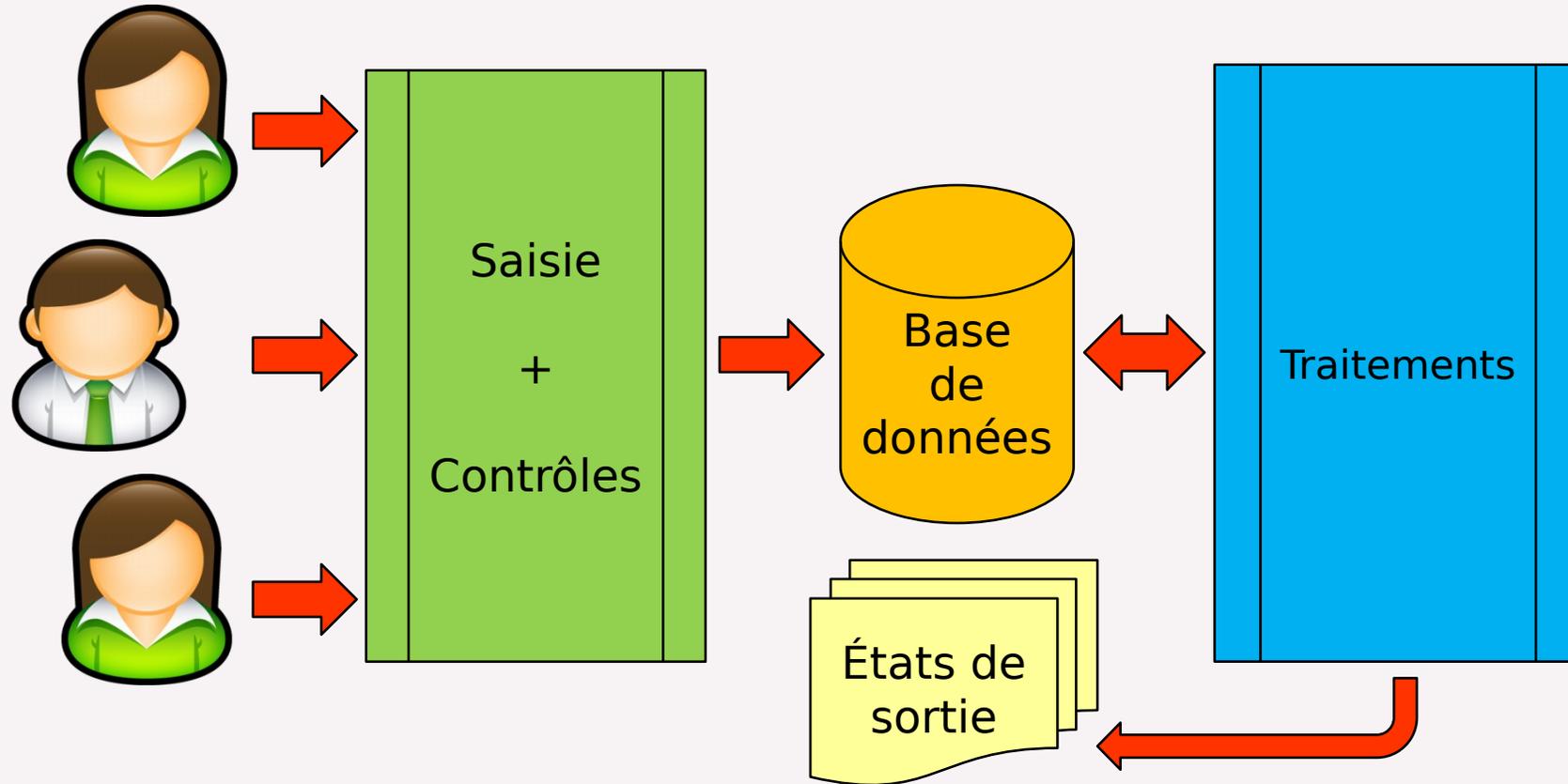


1950-1960 Fichiers





1960-1970 Bases de données



Bases de données relationnelles

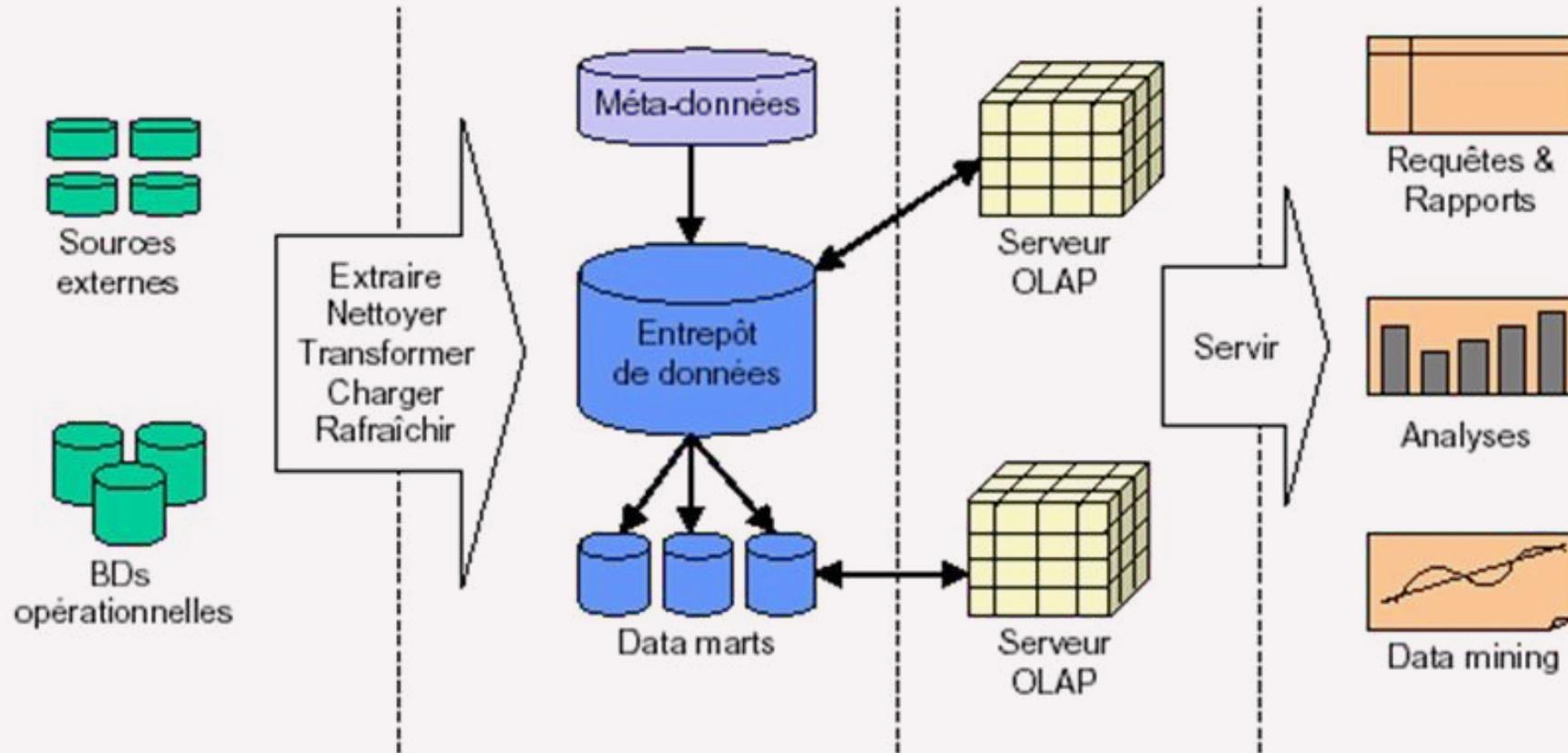


Marie Duflot-Kremer
INRIA

L'HISTOIRE DES BASES DE DONNÉES
ou presque ...



1980-1990 Entrepôts de données

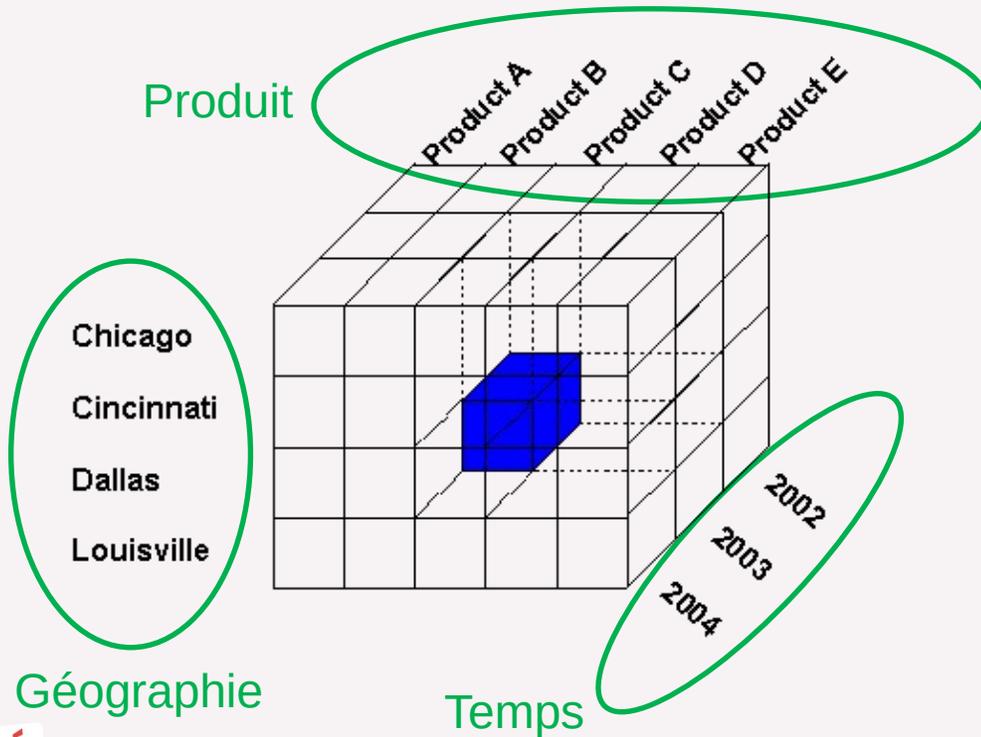


Noura Bouaissa – slideplayer.fr



(Hyper)cube de données

- **Fait** : sujet d'analyse
 - Ensemble de **mesures**
- **Dimensions** : axes d'analyse



	Product A	Product B	Product C	Product D	Product E
Chicago	20M	2M	12M	2M	21M
Cincinnati	30M	4M	10M	8M	26M
Dallas	14M	3M	14M	9M	24M
Louisville	16M	5M	11M	4M	23M

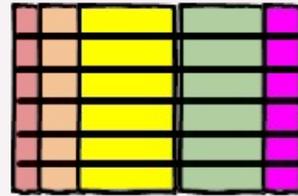
Schémas : docs.oracle.com



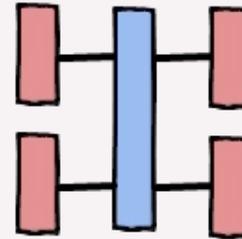
1990-2000 Bases de données NoSQL

Base de données SQL

Relationnelle

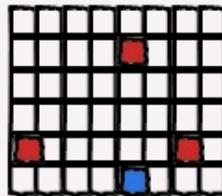


OLAP

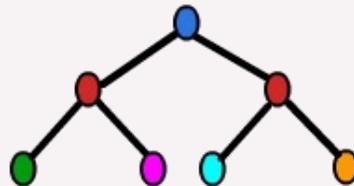


Base de données NoSQL

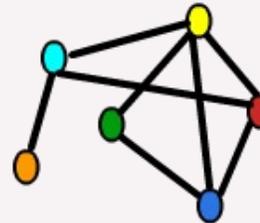
Orienté colonne



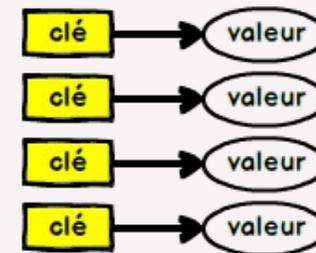
Document



Graphe



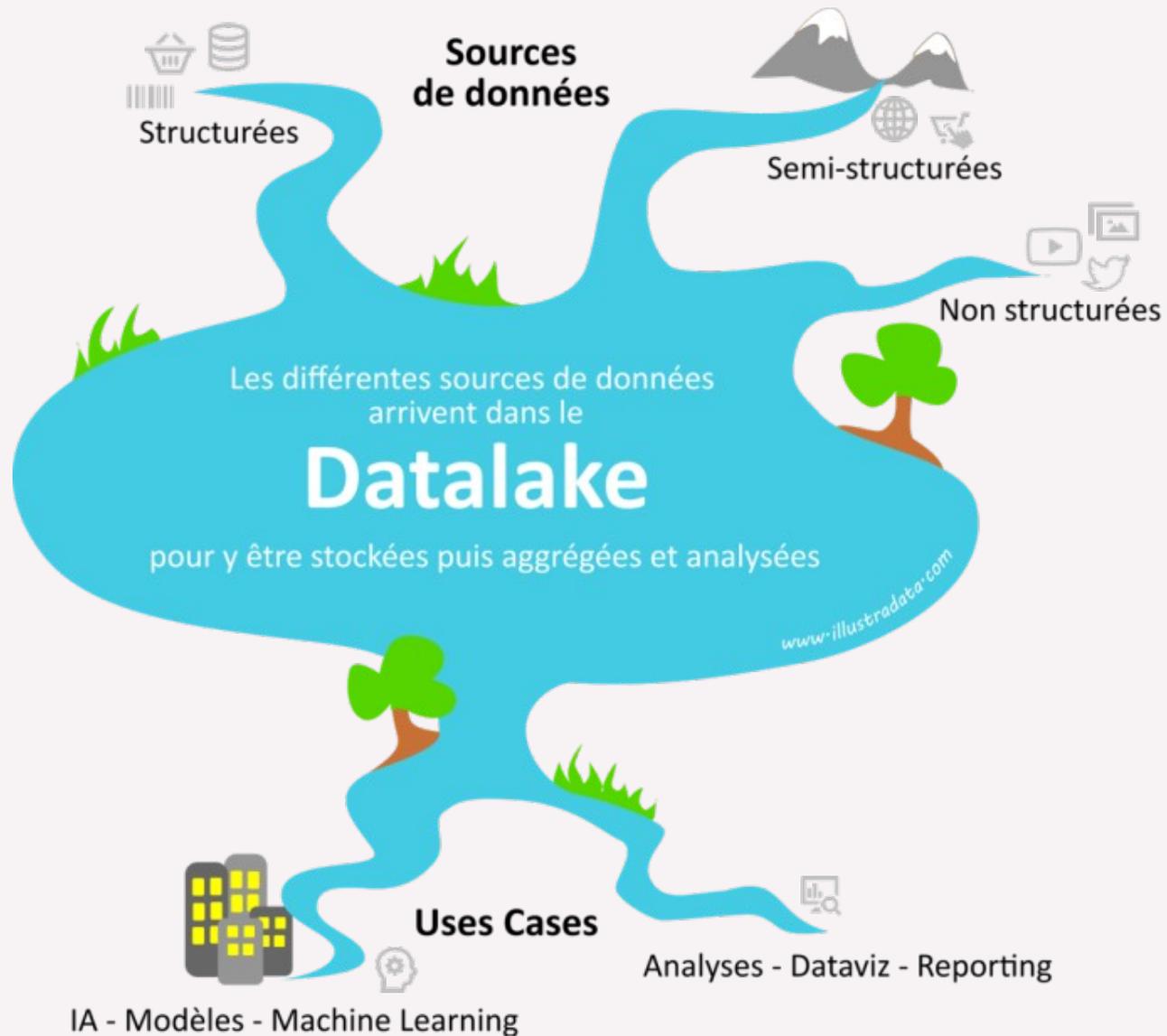
Clé-Valeur



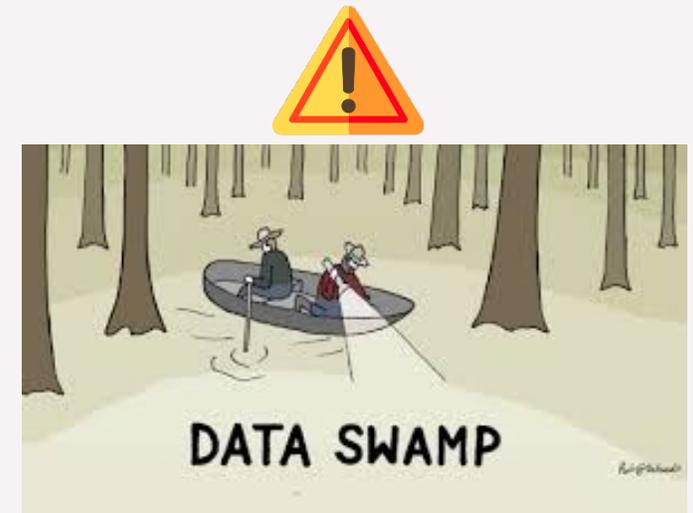
waytolearnx.com

<http://eric.univ-lyon2.fr/jdarmont/>

2010-2020 Lacs de données



illustradata.com



dataedo.com

UNIVERSITÉ
LUMIÈRE
LYON 2

<http://eric.univ-lyon2.fr/jdarmont/>

Quizz

Quels sont les outils de gestion de données les plus courants aujourd'hui ?

- A) Fichiers
- B) Bases de données relationnelles
- C) Entrepôts de données
- D) Bases de données NoSQL
- E) Lacs de données



Répondre sur <https://toreply.univ-lille.fr>

Question n° 925

<http://eric.univ-lyon2.fr/jdarmont/>

Sommaire



Différentes facettes des données



Brève histoire de la gestion de données



Traitement des données, IA et compagnie



Langages de requêtes



Données structurées

```
SELECT nom, prenom  
FROM listeEleves  
WHERE ville = 'Lyon'
```



Données semi-structurées

```
for $e in /listeEleves/eleve  
where ville = 'Lyon'  
return <eleve>  
    $e/nom  
    $e/prenom  
</eleve>
```



Résultat des requêtes



nom	prenom
Lovelace	Ada
Hopper	Grace
Hamilton	Margaret



```
<eleve>
  <nom>Lovelace</nom>
  <prenom>Ada</prenom>
</eleve>

<eleve>
  <nom>Hopper</nom>
  <prenom>Grace</prenom>
</eleve>

<eleve>
  <nom>Hamilton</nom>
  <prenom>Margaret</prenom>
</eleve>
```

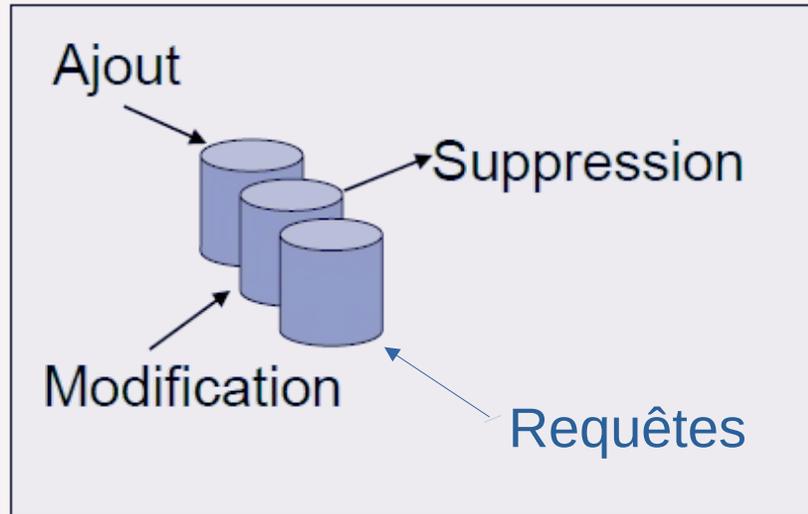


Photos : blog.iakaa.com

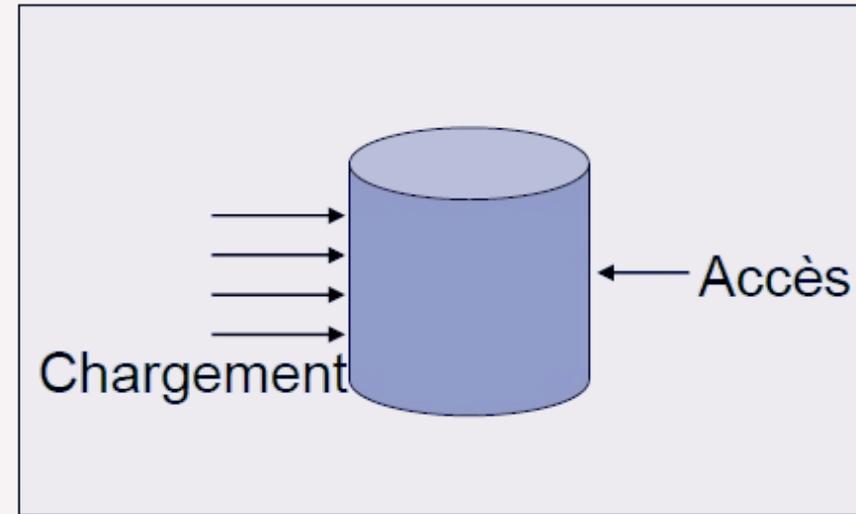


OLTP vs. OLAP

Base de données opérationnelle



Entrepôt de données (décisionnel)



Schémas : Lydie Soler, AgroParisTech/INRA

On **L**ine **T**ransaction **P**rocessing

On **L**ine **A**nalytical **P**rocessing

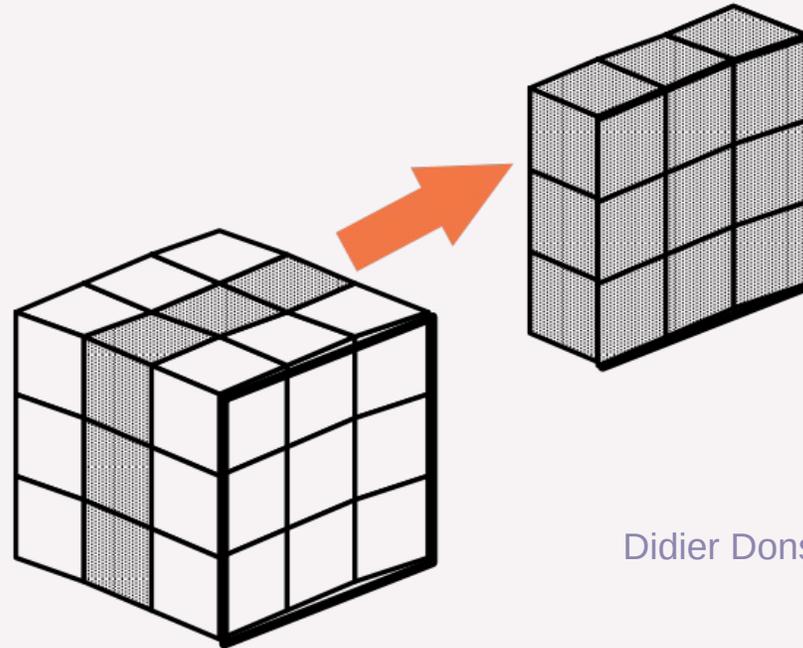
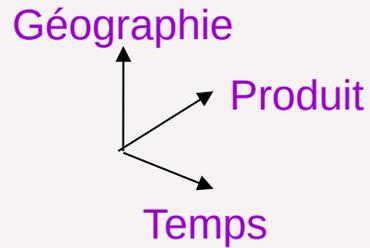


Requêtes OLAP Slice

		1995	1996	1997
Frais	IdF	220	265	284
	Province	225	245	240
Liquide	IdF	163	152	145
	Province	187	174	184



		1996
Frais	IdF	265
	Province	245
Liquide	IdF	152
	Province	174



Didier Donsez, Université Grenoble 1

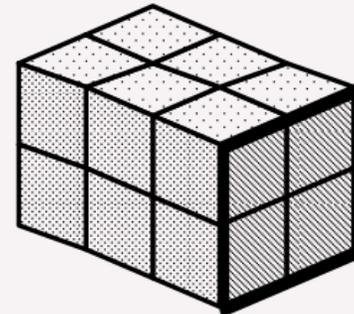
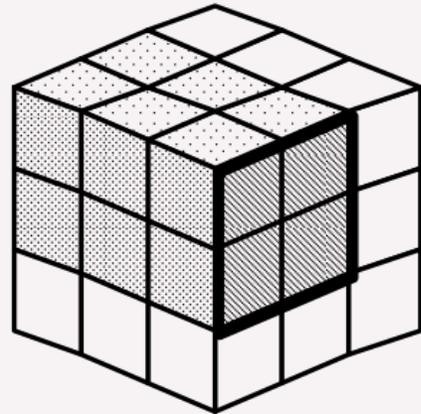


Requêtes OLAP Dice

		1995	1996	1997
Frais	IdF	220	265	284
	Province	225	245	240
Liquide	IdF	163	152	145
	Province	187	174	184



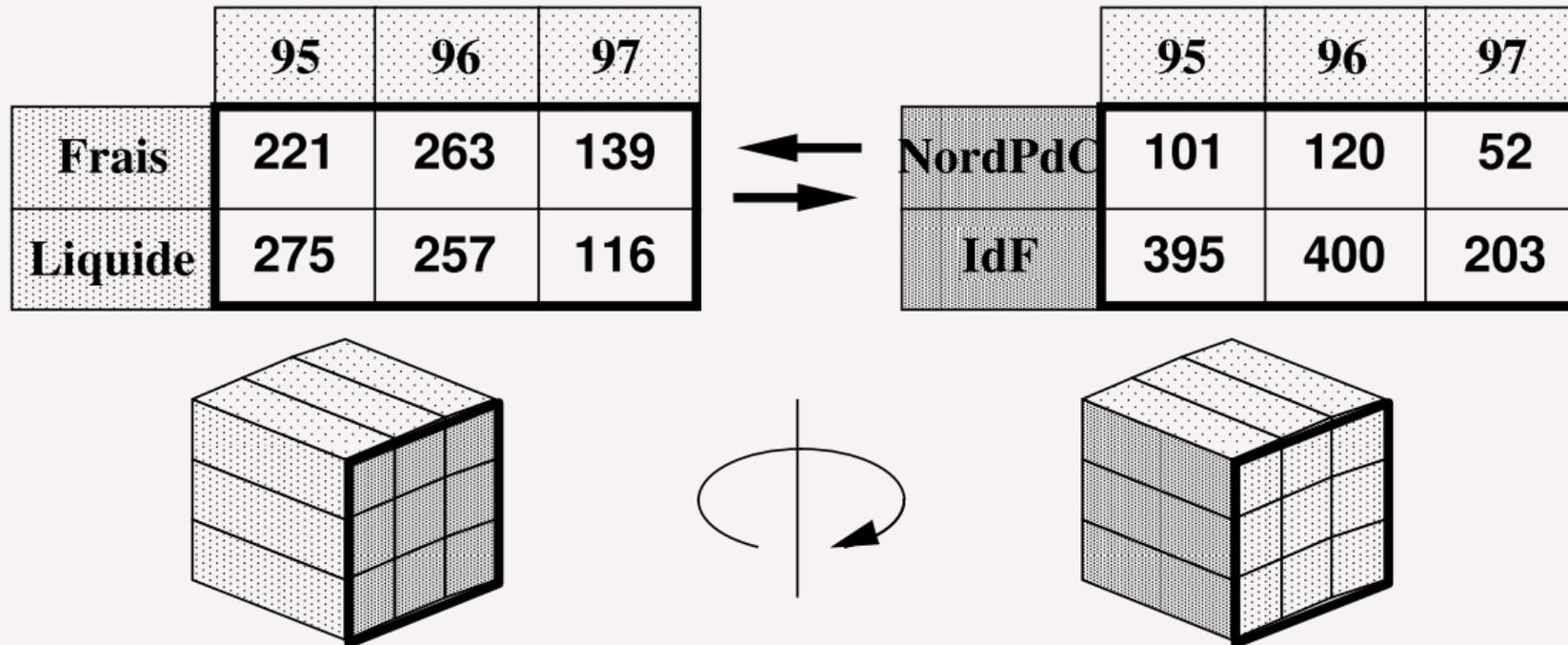
		1995	1996
Frais	IdF	220	265
	Province	225	245



Didier Donsez, Université Grenoble 1



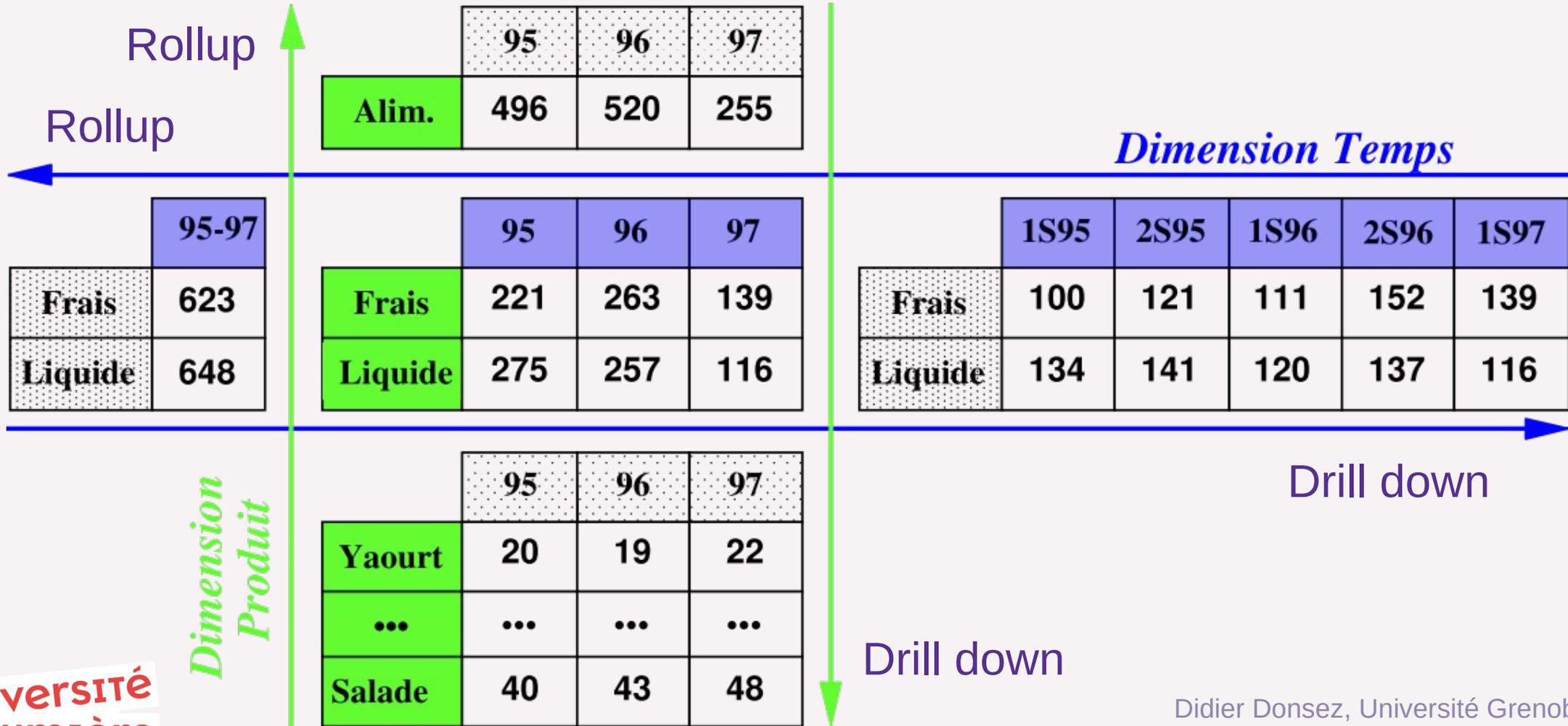
Requêtes OLAP Rotate



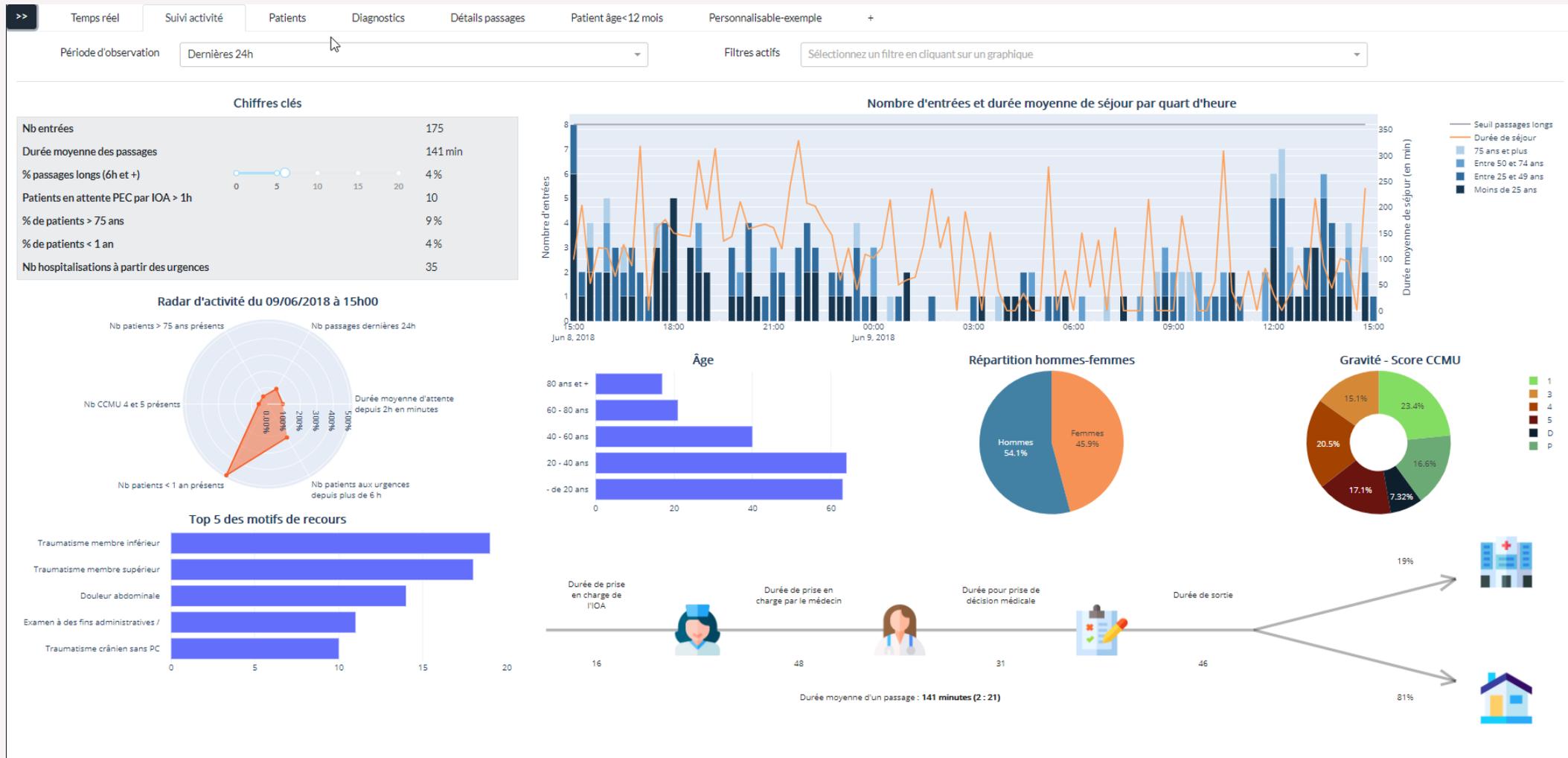
Didier Donsez, Université Grenoble 1



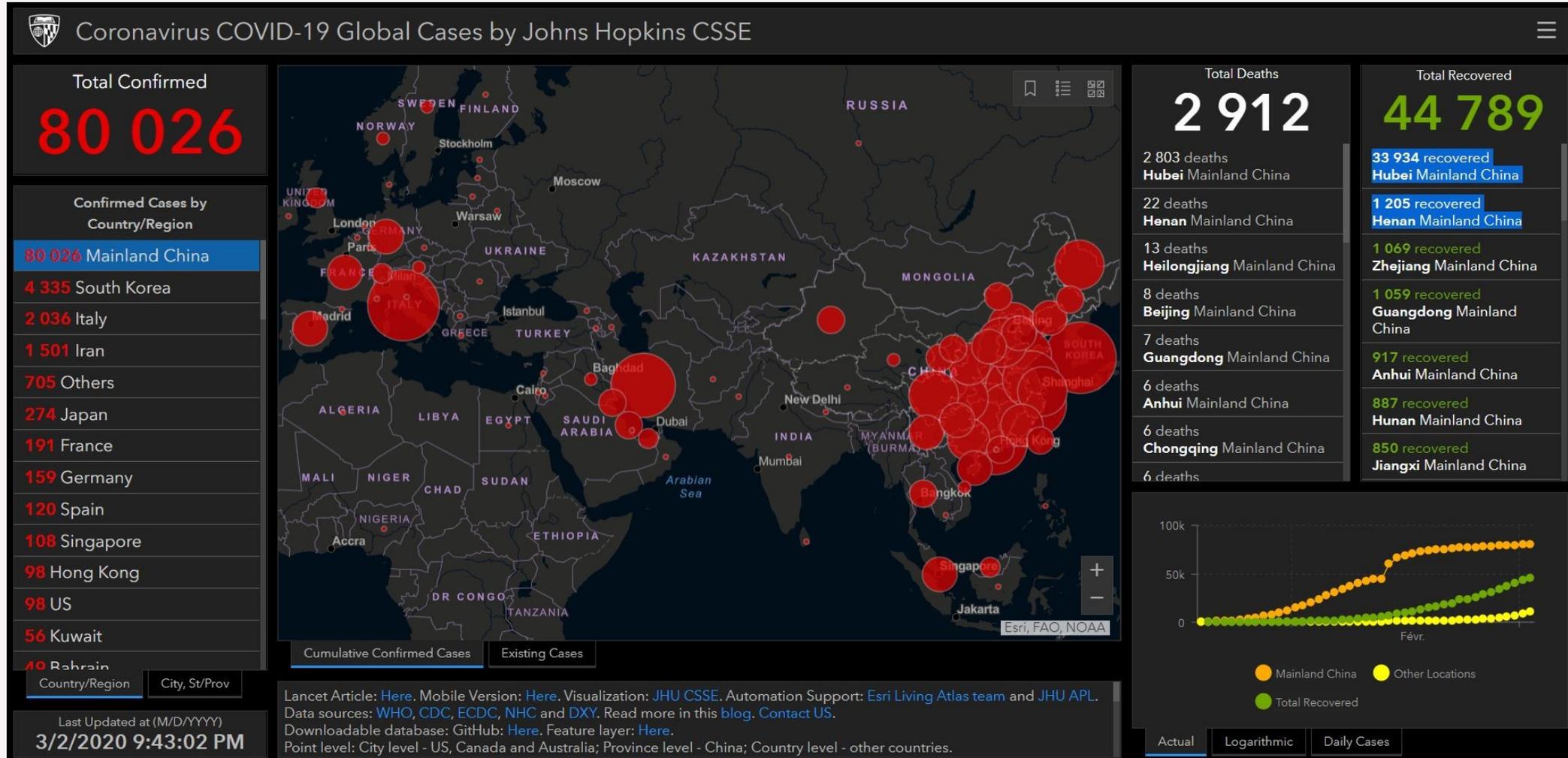
Requêtes OLAP Rollup/Drilldown



Exemple de tableau de bord



Exemple de tableau de bord





Conception d'un tableau de bord

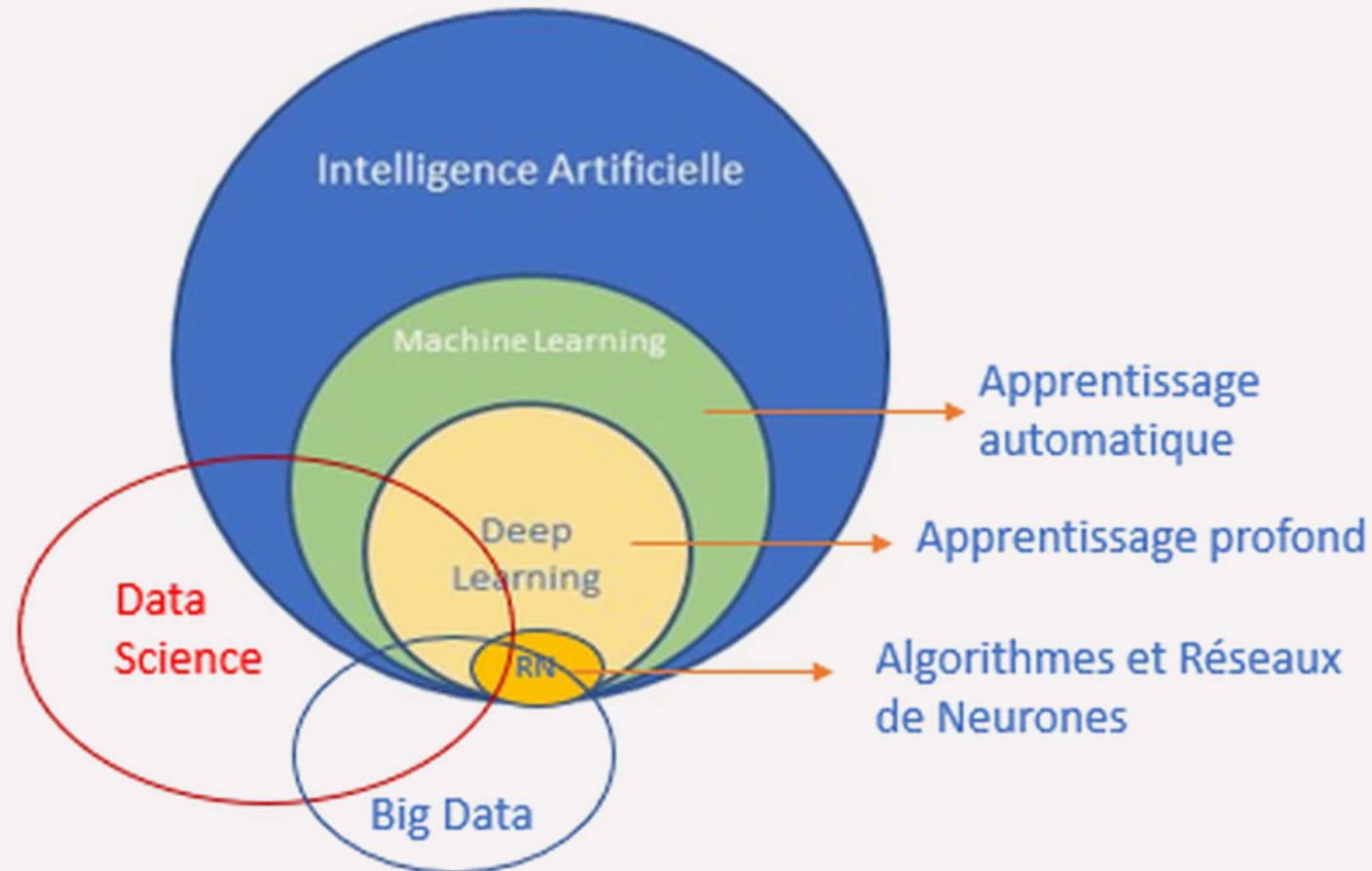
Un reporting réussi



www.piloter.org



Un peu d'intelligence dans ce cours ?



www.journaledunet.com

Apprentissage automatique



neodigital

- université
— LUMIÈRE
— LYON 2

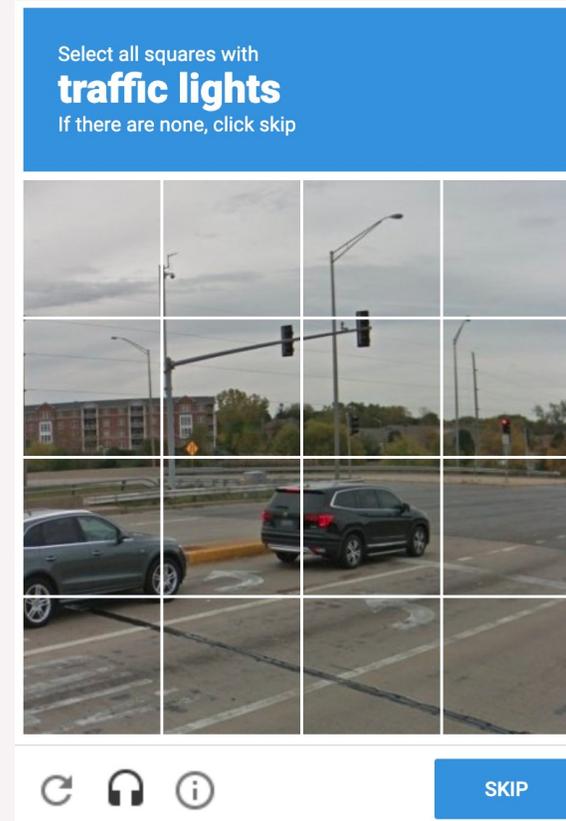
<http://eric.univ-lyon2.fr/jdarmont/>

37/46

Quizz

Quand vous répondez à un captcha, vous participez à de l'apprentissage ?

- A) Supervisé
- B) Non-supervisé



N° de la question : 374



Répondre sur <https://toreply.univ-lille.fr>

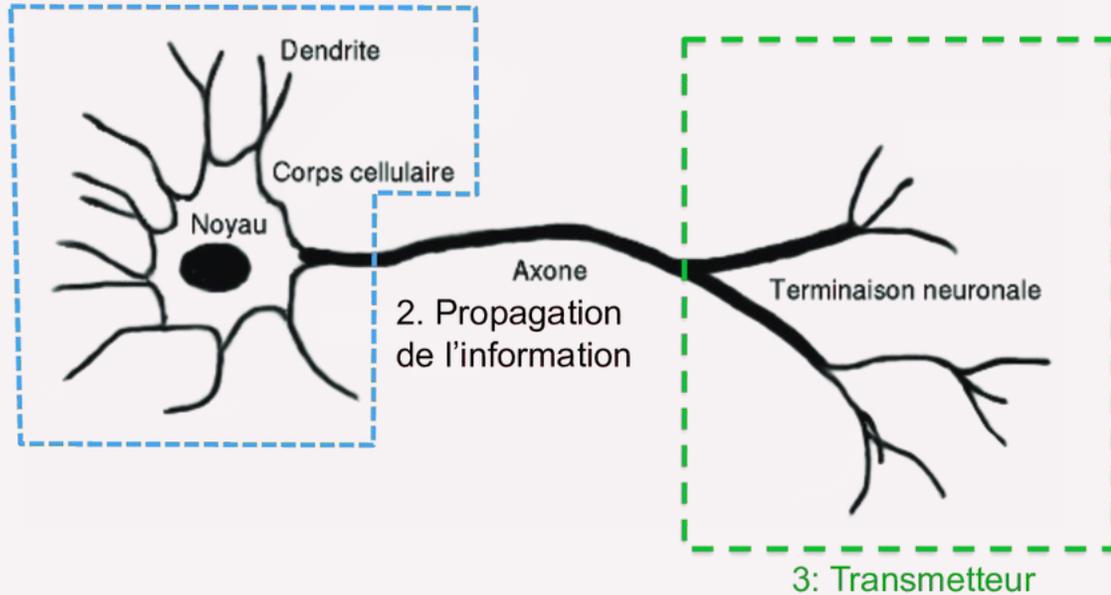
Question n° 374

<http://eric.univ-lyon2.fr/jdarmont/>



Réseaux de neurones... naturels !

1: Récepteur



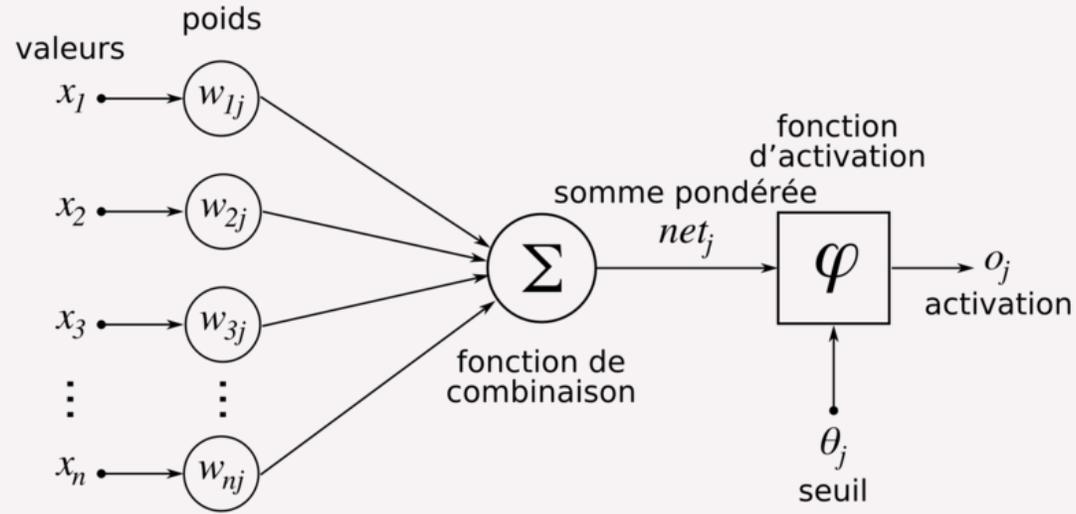
Clément Hébert – www.researchgate.net

inrs.ca

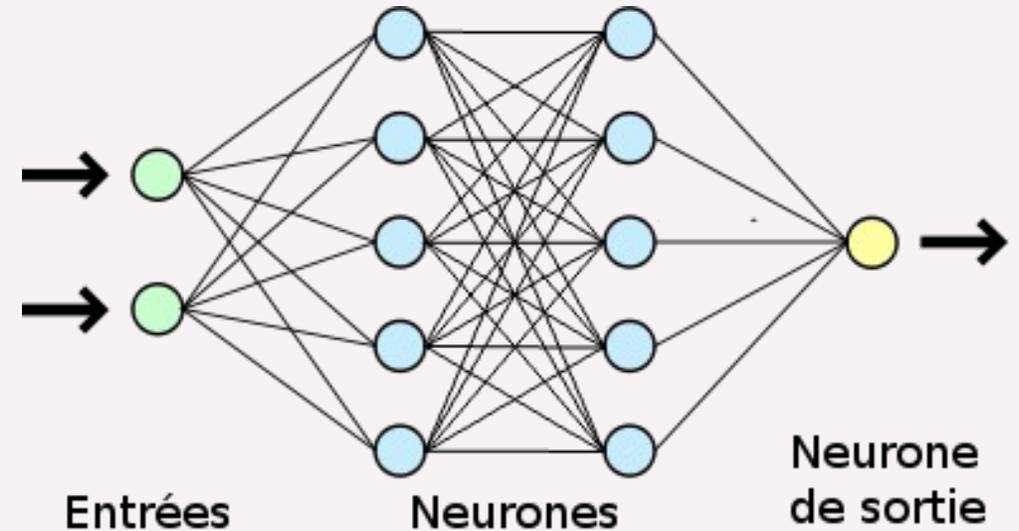




Réseaux de neurones artificiels



www.juripredis.com



Sommaire

	Différents facteurs des données
	Bonheur ou la gestion de données
	Traitement des données, IA et compagnie

BONUS

Exemples de projets de recherche

BI4people



Documents sources



Intégration
sémantique
de données



Modélisation
automatique

Sécurité



Visualisation
automatique



OLAP
simple



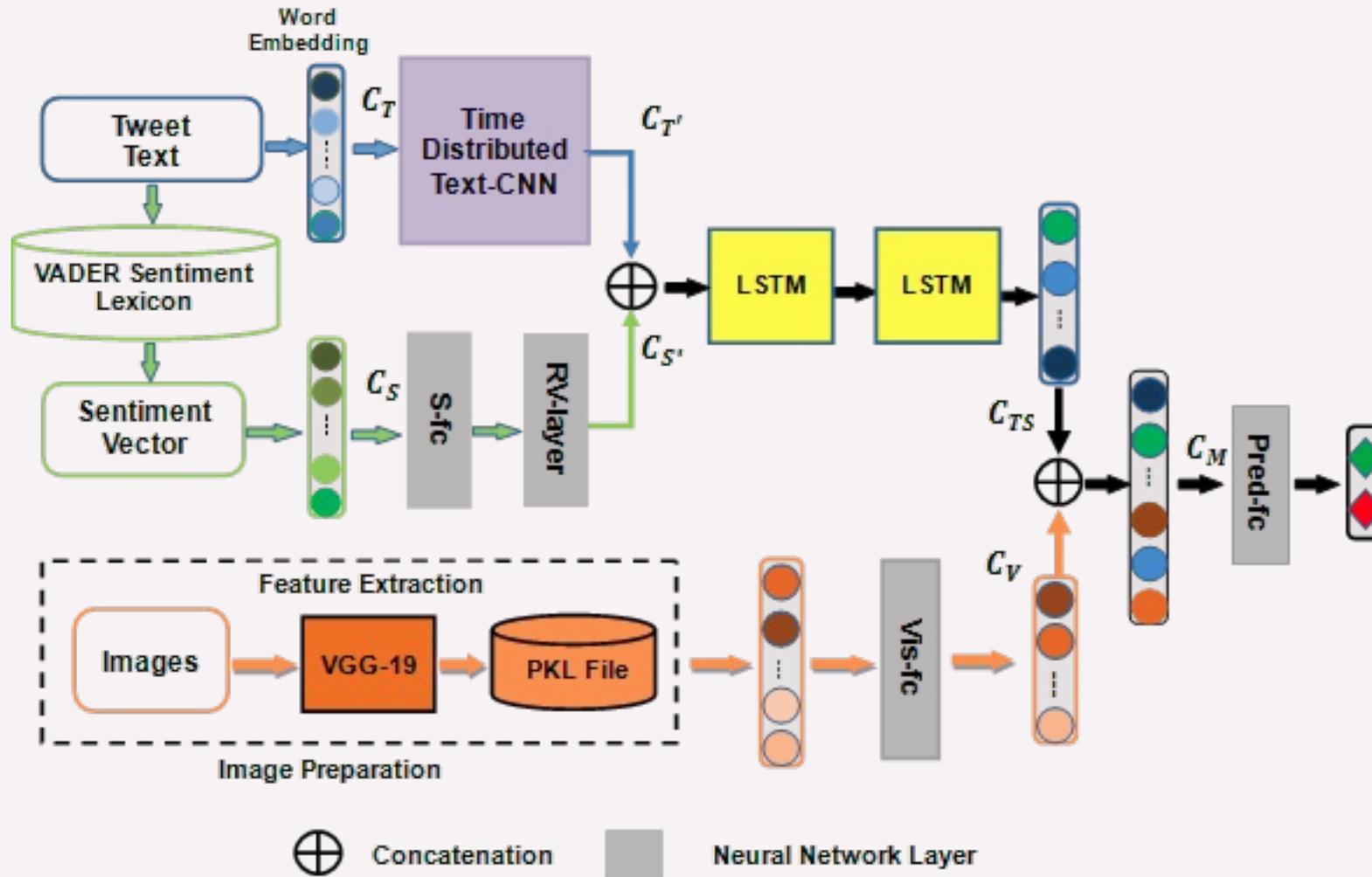
Analyse collaborative



deepMONITOR



Thèse
Abderrazek
Azri



— université
— lumière
— LYON 2



Sondage express

Votre opinion sur ce cours ?

- A) Trop difficile
- B) Bof...
- C) Intéressant
- D) Génial !



Répondre sur <https://toreply.univ-lille.fr>

Question n° 451