

SYSTÈMES D'INFORMATION

E. DUBOIS – 2024

Introduction

- Définition
- Qu'est ce qu'une information ?
- Qu'est ce qu'un système ?

Définition

« Un système d'information est un ensemble organisé de ressources permettant d'acquérir, de traiter, de stocker des informations dans et entre des organisations. »

Robert REIX

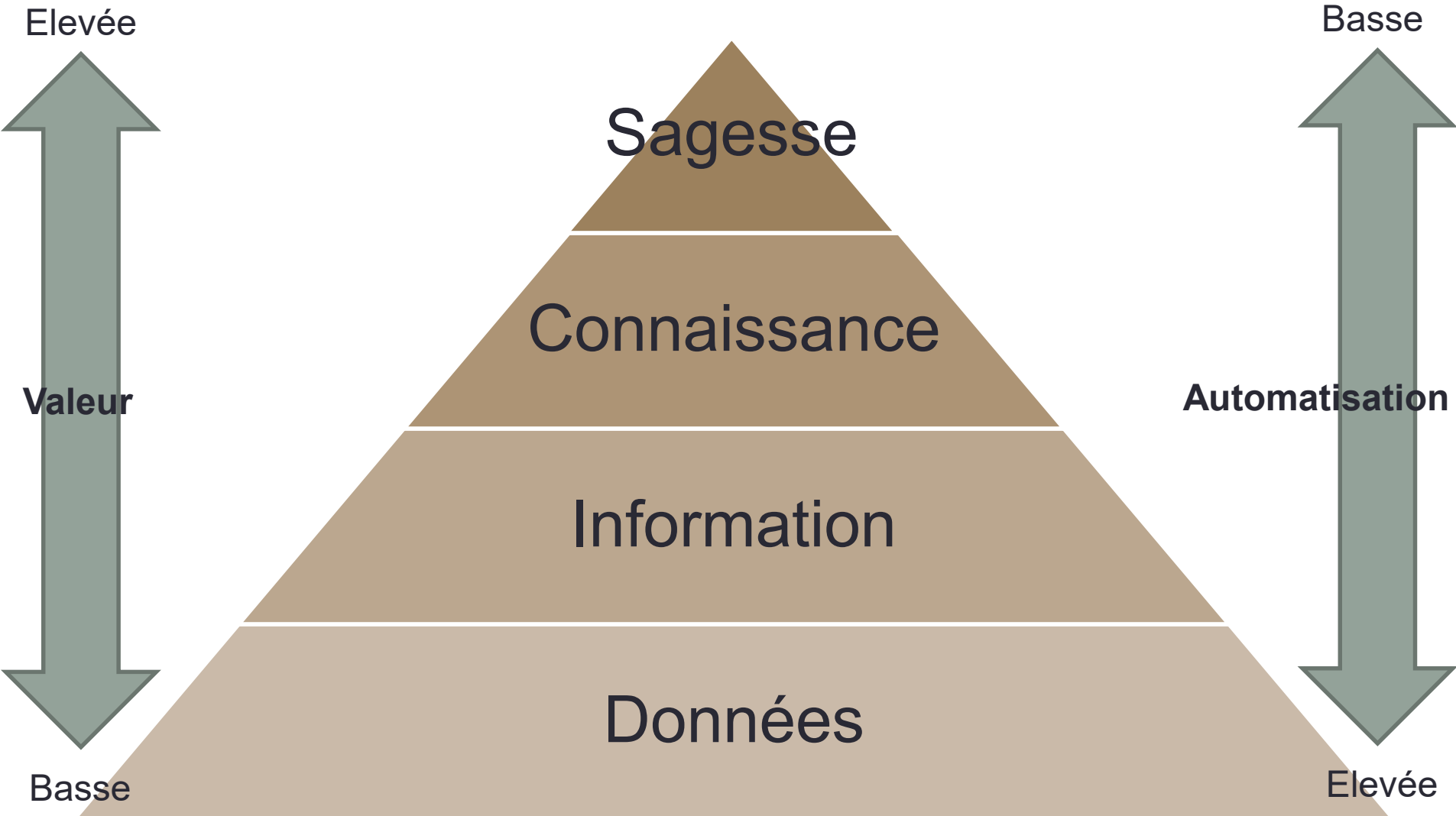
Qu'est ce qu'une information ?

Claude Shannon :

« L'information est ce qui réduit l'incertitude »

- Qualités d'une information :
 - Précision et vérifiabilité ► Fiabilité
 - Complétude
 - Coût raisonnable
 - Flexibilité (utilisation à plusieurs fins)
 - Pertinence
 - Disponibilité
 - Intégrité
 - Confidentialité

Pyramide DICS (Russel Ackoff)



Information vs Donnée

- Les données sont des suites non aléatoires de nombres, valeurs ou textes, obtenues à partir d'observations ou créées par un traitement, et enregistrées.
- **Constat** : Le volume des données augmente régulièrement dans les entreprises et le coût de la gestion d'une donnée y diminue régulièrement
- Une information est :
 - Une donnée rendue signifiante après traitement
 - Une donnée interprétée et comprise par son destinataire

Qu'est ce qu'un système ?

Un ensemble de composants interdépendants travaillant conjointement

Composants :

- **Systèmes informatiques :**
 - Matériels
 - Logiciels
 - Réseaux
- **Ressources humaines**
- **Procédures organisationnelles**

Système informatique : Matériel (Hardware)

- Processeur
 - CISC
 - RISC
- Mémoire centrale
 - ROM
 - RAM
- Bus d'échanges de données
 - Interne
 - Externe (Usb, ...)

Unité centrale



- Périphériques d'entrée
 - Clavier
 - Webcam, ...
- Périphériques de sortie
 - Ecran
 - Imprimante, ...
- Périphériques de stockage
 - Disques SSD
 - Disques magnétiques
 - Disques optiques
 - Bandes magnétiques, ...

Périphériques



- Filaire : ethernet, fibre optique, CPL
- Sans-fil : Wifi, 5G ...

Carte réseau



+ Logiciels (Software)

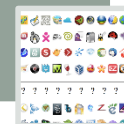
- Noyau
- Pilotes de périphériques
- Interface utilisateur
 - Graphique
 - Ligne de commandes

Système d'exploitation



- Applications bureautiques
- Applications métier
- Applications de divertissement
- ...

Applications



...+ Services réseaux

Le système d'information d'entreprise

Source : Igor ANSOFF

Niveau Stratégique :
Besoin d'information
synthétique et consolidée

Niveau Pilotage :
Besoin de tableaux de bord

**Niveau
administratif/opérationnel**
Besoin d'outils en soutien des
activités

I - Acquisition de données

- L'acquisition des données pose 2 problèmes :
 1. Un problème de représentation : Quel type de donnée conviendra la mieux, avec quels arbitrages correspondants ?
 2. Un problème de codage dans le principal système utilisé par les ordinateurs : le système binaire.

Plan :

A - Un environnement essentiellement gouverné par la logique binaire

1. Pourquoi le binaire ?
2. Des équipements de taille très hétérogènes
3. Evolution : Loi de Moore, circuitisation, downsizing

B - Des arbitrages autour du système binaire

1. Le système binaire et ses opérations
2. La représentation binaire des informations
3. La compression des données

A - Un environnement essentiellement gouverné par la logique binaire

1. Pourquoi le binaire ?
2. Des équipements de taille très hétérogènes
3. Evolution : Loi de Moore, circuitisation, downsizing

1 – Pourquoi le binaire ?

Les portes logiques à la base des circuits

- Basées sur la présence de tension électrique sur un conducteur (état 1) ou son absence (
- Elles sont réalisées par des composants électroniques appelés transistors :
 - 2 transistors pour une porte NOT
 - 4 transistors pour une porte NAND ou NOR
 - 6 transistors pour portes AND, OR, XOR
- Peuvent comporter 1 entrée, 2 entrées ou plus mais se caractérisent toujours par une seule sortie.

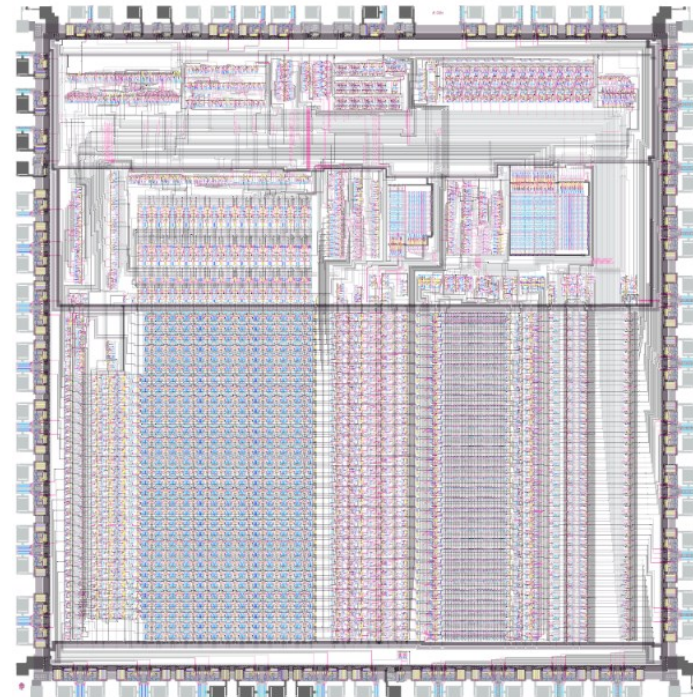


Simulation sur <https://logic.ly/demo>

Un processeur, composant principal d'un ordinateur, est un concentré de portes logiques ayant la faculté de suivre un programme qui détermine son comportement dans la réalisation d'un traitement.

Il reçoit un signal électrique sur certaines pattes composant commandes et données puis produit en retour des résultats sur d'autres pattes connectées aux autres composants de l'ordinateur.

<http://visual6502.org/sim/varm/armgl.html>





Opérande								Commande							
0	0	1	1	1	0	1	0	1	0	0	1	0	0	1	0

	1	0	0	1	0	0	1	0	0	1	0	1	0
	Etat								Résultat dans un registre				

2. Des équipements de tailles très hétérogènes

Les systèmes binaires sont de tailles très variées :

- Grande informatique : *Supercalculateurs* et *mainframes*
<https://www.top500.org/>
- Informatique départementale : *Mini-ordinateurs*
- Informatique du poste de travail : *Micro-ordinateurs*
- Smartphones et tablettes
- Internet of Things

3. Evolution : Loi de Moore, circuitisation, downsizing

- Loi de Moore

Tous les deux ans, on double le nombre de transistors dans un processeur, doublant la puissance d'un ordinateur pour un coût à peu près constant

- Circuitisation

Disparition des éléments électromécaniques au sein des ordinateurs et intégration de toujours plus de fonctions au sein des processeurs

- «system on a chip SoC» : système complet embarqué sur un seul circuit intégré

- Downsizing

- Lors du renouvellement de systèmes informatique, préférence économique pour un ensemble de systèmes de taille inférieure en réseau.

4. Apparition l'informatique Quantique

- En informatique quantique, un **Qbit** ou bit quantique peut valoir à la fois 0 et 1. Des algorithmes spécifiques traitent cette propriété en exploitant des opérations quantiques.
- L'informatique quantique utilise un autre système d'opérations que celui des portes logiques, dont certaines peuvent émuler les portes logiques.
- Pour se ramener à une logique binaire, on peut utiliser des opérations qui maximisent la probabilité qu'un Qbit ait la valeur 0 ou 1.
 - On peut citer par exemple :
 - $|0\rangle$ Donne à un Qbit une probabilité de 1 de valoir 0
 - X : porte X de Pauli qui inverse le Qbit en opérande
 - + : Porte de NOT contrôlé qui nécessite un Qbit opérande et un Qbit de contrôle qui inverse le Qbit opérande si le Qbit de contrôle est dans l'état 1
- Toutefois, les principaux concepts (enchevêtrement, superposition) de l'informatique quantique sont très différents de ceux véhiculés par les portes logiques entraînant la nécessité de recréer les bibliothèques de fonctions à la base des applications

B - Des arbitrages autour du système binaire

- Introduction : Les grandeurs quantifiant les données binaires
 1. Le système binaire et ses opérations
 - a – Représentation binaire des nombres entiers positifs
 - b – Les opérations en système binaires
 2. La représentation binaire des informations
 3. La compression des données

Les grandeurs quantifiant les données

Bit : Acronyme de Blnary digiT, plus petite grandeur informatique pouvant avoir, de
Octet : Association de 8 bits accolés les uns derrière les autres. Son sigle est la lettre o
Mot : Association de 8, 16, 32 ou 64 bits correspondant à la taille de mémorisation des données à l'intérieur d'un processeur (cerveau d'un ordinateur)

Unités binaires

1 <i>kilo</i> octet binaire (kibiocet)	Kio	=1024 octets	2^{10} octets	1 024 octets
1 <i>méga</i> octet binaire (mébiocet)	Moi	=1024 Kio	2^{20} octets	1 048 576 octets
1 <i>giga</i> octet binaire (Gibiocet)	Gio	=1024 Moi	2^{30} octets	1 073 741 824 octets
1 <i>tera</i> octet binaire (Tébiocet)	Tio	=1024 Gio	2^{40} octets	1 099 511 627 776 octets
1 <i>péta</i> octet binaire (Pébiocet)	Pio	=1024 Tio	2^{50} octets	1 125 899 906 842 624 octets
1 <i>exa</i> octet binaire (Exbiocet)	Eio	=1024 Pio	2^{60} octets	1 152 921 504 606 846 976 octets
1 <i>zetta</i> octet binaire (Zébiocet)	Zio	=1024 Eio	2^{70} octets	1 180 591 620 717 411 303 424 octets
1 <i>yotta</i> octet binaire (Yobiocet)	Yio	=1024 Zio	2^{80} octets	1 208 925 819 614 629 174 706 176 octets

Unités usuelles

1 <i>kilo</i> octet	Ko	=1000 octets	10^3 octets	1 000 octets
1 <i>méga</i> octet	Mo	=1000 Ko	10^6 octets	1 000 000 octets
1 <i>giga</i> octet	Go	=1000 Mo	10^9 octets	1 000 000 000 octets
1 <i>tera</i> octet	To	=1000 Go	10^{12} octets	1 000 000 000 000 octets
1 <i>péta</i> octet	Po	=1000 To	10^{15} octets	1 000 000 000 000 000 octets
1 <i>exa</i> octet	Eo	=1000 Po	10^{18} octets	1 000 000 000 000 000 000 octets
1 <i>zetta</i> octet	Zo	=1000 Eo	10^{21} octets	1 000 000 000 000 000 000 000 octets
1 <i>yotta</i> octet	Yo	=1000 Zo	10^{24} octets	1 000 000 000 000 000 000 000 000 octets

1 – Le système binaire (en base 2) et ses opérations

- Le système binaire (système base 2) connaît seulement 2 chiffres 0 et 1.
... alors que le système décimal (système en base 10) que nous utilisons en connaît 10 (0,1,2,3,4,5,6,7,8,9).
- Un chiffre binaire, valant donc 0 ou 1 est appelé **bit** (contraction de **B**inary digi**T**)
- Un nombre binaire est une association de bits, dont le poids est fonction de leur position (le poids d'un **bit** est d'autant plus important qu'il est positionné à **gauche**)

a - Représentation binaire des nombres entiers positifs

Question : Quand en base 10 on dépasse d'une unité le dernier chiffre de la base (9), que fait-on ?

- **Réponse :** On met 0 que l'on précède du 1 (à sa gauche)

$$\begin{array}{r} 9 \\ + 1 \\ \hline = 10 \end{array}$$

- En binaire, c'est pareil, dès que l'on dépasse d'une unité le dernier chiffre de la base, on met 0 précédé du 1 (à sa gauche)

$$\begin{array}{r} 1 \\ + 1 \\ \hline = 10 \end{array}$$

a - Représentation binaire des nombres entiers positifs

a / Nombre 1809 en base 10

Position du chiffre à
partir de la droite

3	2	1	0
1	8	0	9

Unités 1 = 10^0

Dizaines (multiples de 10) sachant que $10 = 10^1$

Centaines (multiples de 100) sachant que $100 = 10^2$

Milliers 1000 (multiples de 1000) sachant que $1000 = 10^3$

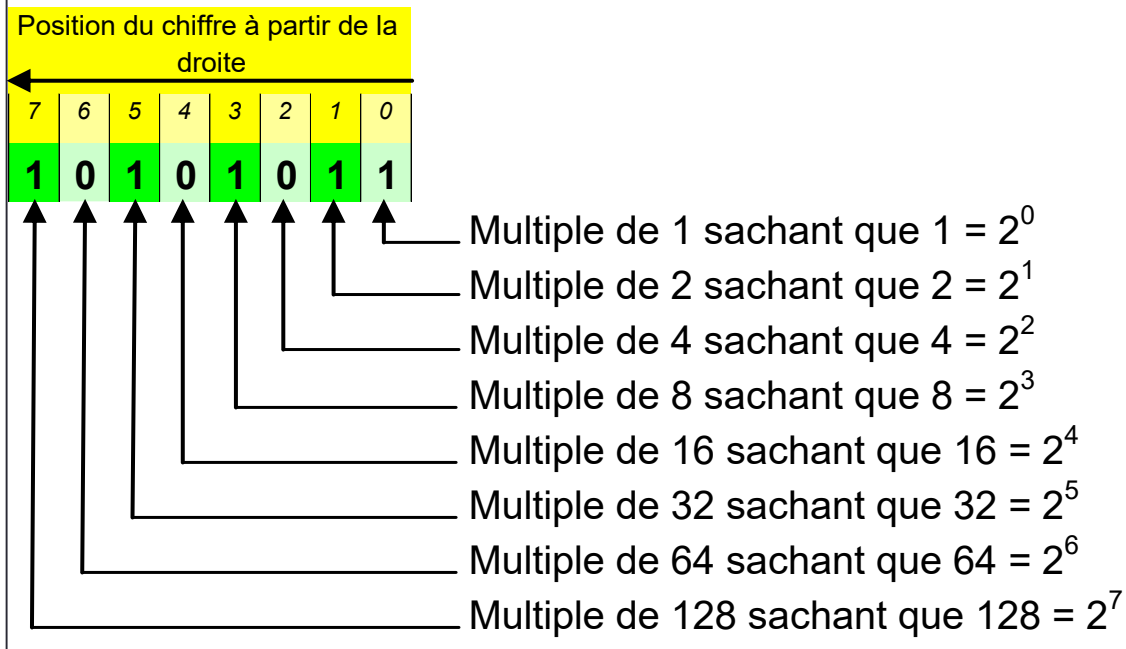
...

Donc ce nombre est égal à

$$1 \times 1000 + 8 \times 100 + 0 \times 10 + 9 \times 1 = 1809$$

a - Représentation binaire des nombres entiers positifs

b / Nombre 10101011 en base 2 (binaire)



Donc ce nombre est égal à $1 \times 128 + 0 \times 64 + 1 \times 32 + 0 \times 16 + 1 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1 = 171$ en système décimal

a - Représentation binaire des nombres entiers positifs

Conversion Décimal / Binaire

Ex : Conversion de 40 000 en représentation binaire.

- **Méthode 1 :**
 - **Etape 1 :** Quelle est la plus grande puissance de 2 inférieure ou égale à 40 000
 - **Etape 2 :** Pour chaque puissance de 2 prise par ordre décroissant à partir de celle obtenue, si cette puissance est inférieure à la valeur, on positionne 1 et on retranche cette puissance de 2 à la valeur, sinon on positionne un 0. Nous avons ainsi une succession de bits que nous positionnons de la gauche vers la droite.
- **Méthode 2 :**

On obtient la valeur binaire par une succession de divisions entières du nombre par la valeur 2. Le reste de chacune des divisions nous donne les bits à positionner de la droite vers la gauche

a - Représentation binaire des nombres entiers positifs

			Puissance de 2		Valeur		Résultat
2^0	1	1					
2^1	2		32 768	≤	40 000	?	Oui → 1
2^2	4		16 384	≤	7 232	?	Non → 0
2^3	8		8 192	≤	7 232	?	Non → 0
2^4	16		4 096	≤	7 232	?	Oui → 1
2^5	32		2 048	≤	3 136	?	Oui → 1
2^6	64		1 024	≤	1 088	?	Oui → 1
2^7	128		512	≤	64	?	Non → 0
2^8	256		256	≤	64	?	Non → 0
2^9	512	2	128	≤	64	?	Non → 0
2^{10}	1 024		64	≤	64	?	Oui → 1
2^{11}	2 048		32	≤	0	?	Non → 0
2^{12}	4 096		16	≤	0	?	Non → 0
2^{13}	8 192		8	≤	0	?	Non → 0
2^{14}	16 384		4	≤	0	?	Non → 0
2^{15}	32 768		2	≤	0	?	Non → 0
2^{16}	65 536		1	≤	0	?	Non → 0

3

La valeur binaire obtenue est ainsi :

1 0 0 1 1 1 0 0 0 1 0 0 0 0 0 0

a - Représentation binaire des nombres entiers positifs

Méthode n°2 :

La division entière utilise deux opérandes (un dividende et un diviseur) entières et produit un résultat (quotient) obligatoirement entier, éventuellement accompagné d'un reste sous la forme d'un entier compris entre 0 et la valeur du diviseur minorée de 1.

$$\begin{array}{l} 5 / 2 = 2 \\ \text{avec un reste} = 5 - 2 \times 2 \\ = 1 \end{array}$$

a - Représentation binaire des nombres entiers positifs

40 000	/ 2 =	20000	avec un reste de : 0
20 000	/ 2 =	10000	avec un reste de : 0
10 000	/ 2 =	5000	avec un reste de : 0
5 000	/ 2 =	2500	avec un reste de : 0
2 500	/ 2 =	1250	avec un reste de : 0
1 250	/ 2 =	625	avec un reste de : 0
625	/ 2 =	312	avec un reste de : 1
312	/ 2 =	156	avec un reste de : 0
156	/ 2 =	78	avec un reste de : 0
78	/ 2 =	39	avec un reste de : 0
39	/ 2 =	19	avec un reste de : 1
19	/ 2 =	9	avec un reste de : 1
9	/ 2 =	4	avec un reste de : 1
4	/ 2 =	2	avec un reste de : 0
2	/ 2 =	1	avec un reste de : 0
1	/ 2 =	0	avec un reste de : 1
0		Stop	

La valeur binaire obtenue est ainsi : 1 0 0 1 1 1 0 0 0 1 0 0 0 0 0 0

b - Les opérations en système binaire

Addition

Comme en système décimal, sachant que dès que le résultat sur une position des opérandes dépasse 1, il convient de poser une retenue sur la position suivante.

L'opération se réalise de droite à gauche en appliquant à chaque position les règles suivantes :

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 1 \ 0 \quad \text{soit 0 avec une retenue de 1 pour la position suivante (à gauche)}$$

$$\begin{array}{rcccccccc} & & & & & +1 & +1 & +1 \\ & & & & & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ + & & & & & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ \hline = & & & & & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}$$

$$\begin{array}{r}
 \begin{array}{ccccccc}
 & -1 & & -1 & -1 & -1 & -1 \\
 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
 - & & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\
 \hline
 = & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1
 \end{array}
 \end{array}$$

b - Les opérations en système binaire

Et (And)

L'opération logique ET est une opération qui s'effectue bit à bit sans retenue et qui consiste à considérer que le résultat ne vaut 1 que si les deux bits correspondants des opérandes sont simultanément à 1, le résultat est zéro dans les 3 autres cas.

0	ET	0	=	0
0	ET	1	=	0
1	ET	0	=	0
1	ET	1	=	1

$$\begin{array}{r} \text{ET} \\ = \end{array} \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

b - Les opérations en système binaire

OU (Or)

L'opération logique OU est une opération qui s'effectue bit à bit sans retenue et qui consiste à considérer que le résultat vaut 1 si au moins l'un des deux bits correspondants des opérandes est à 1, le résultat est zéro dans le dernier cas.

0	OU	0	=	0
0	OU	1	=	1
1	OU	0	=	1
1	OU	1	=	1

$$\text{OU} \begin{array}{r} 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0 \\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0 \\ \hline 1\ 1\ 1\ 1\ 1\ 0\ 1\ 0 \end{array}$$

b - Les opérations en système binaire

OU EXCLUSIF (eXclusive OR – XOR)

L'opération logique OU EXCLUSIF est une opération qui s'effectue bit à bit sans retenue et qui consiste à considérer que le résultat ne vaut 1 que si les deux bits correspondants des opérandes différents (l'un à 1 et l'autre à zéro), et zéro dans les deux autres cas.

Une nouvelle application de cette opération sur le résultat avec l'une des deux opérandes permet de retrouver l'autre opérande.

0	OU EXCLUSIF	0	=	0
0	OU EXCLUSIF	1	=	1
1	OU EXCLUSIF	0	=	1
1	OU EXCLUSIF	1	=	0

$$\begin{array}{r} \text{XOR} \quad \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{array} \\ = \quad \begin{array}{cccccccc} 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{array} \end{array}$$

$$\begin{array}{r} \text{XOR} \quad \begin{array}{cccccccc} 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{array} \\ = \quad \begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{array} \end{array}$$

b - Les opérations en système binaire

Le Complément à 1

Inversion des bits de l'opérande

0	→	1
1	→	0

	1	0	1	0	1	0	1	0
Complément à 1	0	1	0	1	0	1	0	1

$$\begin{array}{r}
 \text{Complément à 1} \quad \begin{array}{cccccccccc} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{array} \\
 \hline
 \begin{array}{cccccccccc} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array} \\
 \begin{array}{r} +1 \\ \hline \end{array} \\
 \begin{array}{r} + \\ \hline \end{array} \begin{array}{cccccccccc} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{array} \\
 = \begin{array}{cccccccccc} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{array} \quad \text{ignoré}
 \end{array}$$

b - Les opérations en système binaire

Le complément à 2

Lorsque l'on a un nombre binaire x de n bits, complémenter à 2 la valeur x est équivalent à faire $(2^n - x)$ en ignorant les dépassements

	-1	-1	-1	-1	-1	-1	-1		
	1	0	0	0	0	0	0	0	$=2^8$
-		1	0	1	0	1	0	1	0
	0	0	1	0	1	0	1	1	0

ignoré

b - Les opérations en système binaire

Le décalage à droite

Chaque bits est décalé d'une position vers la droite Le bit initialement le plus à droite est perdu (sauf, en pratique, s'il est considéré en tant que retenue pour une opération ultérieure) et le bit le plus à gauche est positionné à zéro.

Le décalage à droite est équivalent à une division entière par 2.

1 1 0 1 0 1 0 1	valeur en base 10 :	213
→ → → → → → → →		
0 1 1 0 1 0 1 0	valeur en base 10 :	106

b - Les opérations en système binaire

Le décalage à gauche

Chaque bits est décalé d'une position vers la gauche. Le bit initialement le plus à gauche est perdu (sauf, en pratique, s'il est considéré en tant que retenue pour une opération ultérieure) et le bit le plus à droite est positionné à 0.

Le décalage à gauche est équivalent à une multiplication par 2 (sauf en cas de perte du bit positionné le plus à gauche ayant la valeur 1)

0 1 0 1 0 1 0 1

valeur en base 10 : 85

← ← ← ← ← ← ← ←

1 0 1 0 1 0 1 0

valeur en base 10 : 170

b - Les opérations en système binaire

La rotation à droite

Chaque bits est décalé d'une position vers la droite. La valeur du bit initialement le plus à droite est positionnée dans le bit le plus à gauche.

La rotation à gauche

Chaque bits est décalé d'une position vers la gauche. La valeur du bit initialement le plus à gauche est positionnée dans le bit le plus à droite.

Ces deux opérations n'ont aucune correspondance en arithmétique et sont utilisés essentiellement dans le cadre d'opérations de chiffrement.

2 – La représentation binaire des informations

- a. Relation entre taille de codage et plage de valeurs
- b. Nombre entiers dont négatifs
- c. Nombre réels
- d. Textes
- e. Données temporelles
- f. Données analogiques
- g. Données graphiques

3 – La compression

- A – Les déterminants des données multimédia et leur volumétrie
- B – Nature de la compression
- C – La compression sans perte
- B – La compression avec perte

A - Les déterminants des données multimédia la volumétrie des données

Les déterminants du son numérique et volumétrie des données

Les données audio sont caractérisées par les paramètres suivants, conditionnés par les réglages du convertisseur analogique/digital qui les a enregistrées :

- *Taux d'échantillonnage*, en échantillons par seconde ou en (kilo-)Hertz
- *Nombre de bits par échantillon*, par exemple 8 ou 16
- *Nombre de canaux* (1 pour mono, 2 pour stéréo, etc...)

Exemple 1 :

Une minute de son stéréo en qualité CD =

$$44100 \text{ échantillons / sec} \times 60 \text{ secondes} \times 16 \text{ bits} \times 2 \text{ canaux} = \\ 84\,672\,000 \text{ bits} \approx 10 \text{ Mo}$$

Une minute de son stéréo en haute qualité=

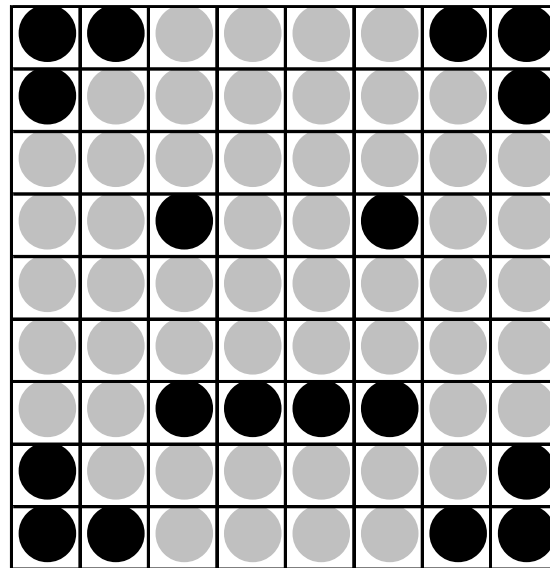
$$192\,000 \text{ échantillons / sec} \times 60 \text{ secondes} \times 24 \text{ bits} \times 2 \text{ canaux} = \\ 552\,960\,000 \text{ bits} \approx 69 \text{ Mo}$$

Les déterminants de l'image numérique et volumétrie des données

Les déterminants de l'image numérique et volumétrie des données

- La **résolution horizontale**, mesurée en nombre de points par pouce
- La **résolution verticale**, mesurée en nombre de points par pouce
- La **profondeur de couleurs**, mesurée en nombre de bits par points
 - 1 bit par point : noir et blanc
 - 8 bits par point : 256 niveaux de gris ou 256 couleurs différentes choisies dans une palette
 - 24 bits : 16 millions de couleurs caractérisées par 3 intensités de 0 à 255 pour chaque composante *Rouge Vert Bleu* de la couleur.

Image en noir et blanc



1	1	0	0	0	0	1	1
1	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0
1	0	0	0	0	0	0	1
1	1	0	0	0	0	1	1

195
129
0
36
0
0
60
129
195

1 bit par points soit 9 octets pour l'image en basse résolution :

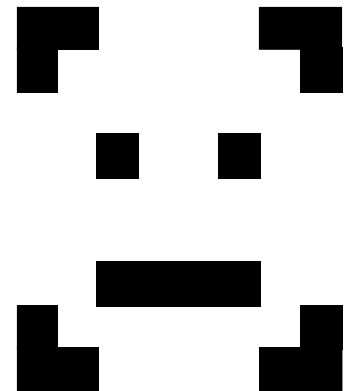
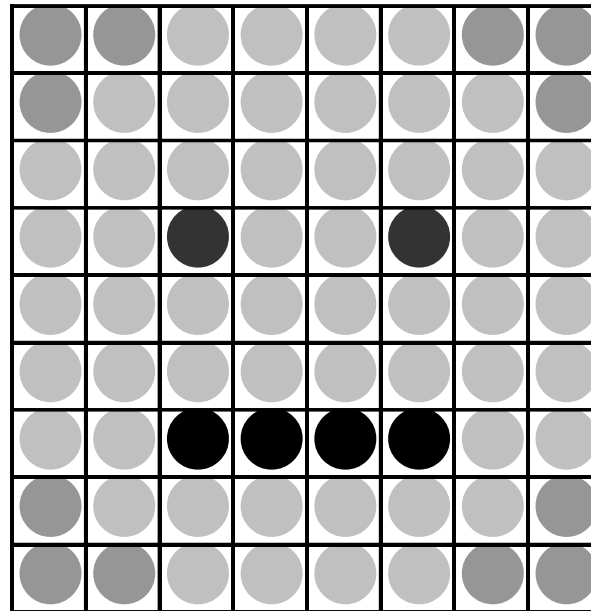


Image en niveaux de gris



128	128	192	192	192	192	128	128
128	192	192	192	192	192	192	128
192	192	192	192	192	192	192	192
192	192	64	192	192	64	192	192
192	192	192	192	192	192	192	192
192	192	192	192	192	192	192	192
192	192	0	0	0	0	192	192
128	192	192	192	192	192	192	128
128	128	192	192	192	192	128	128

8 bits (ou 1 octet) par points soit 72 octets pour l'image :

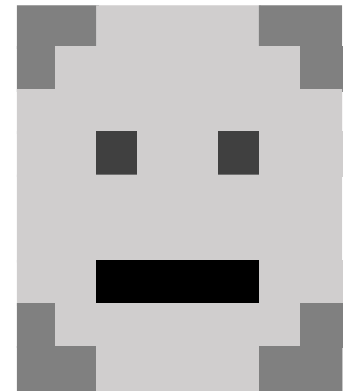
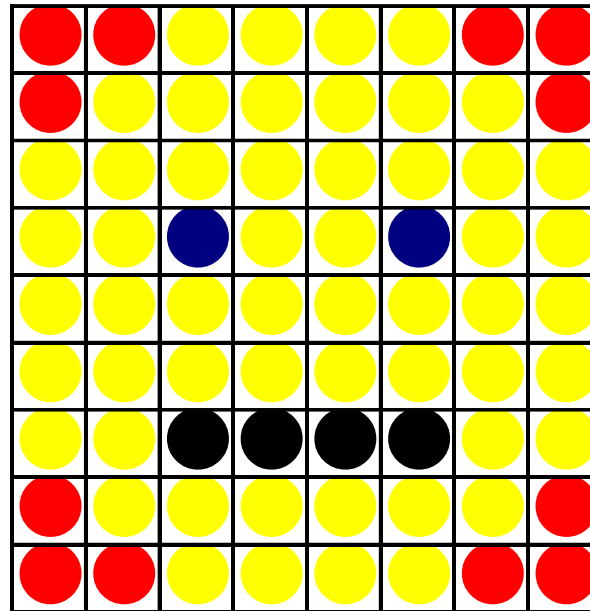
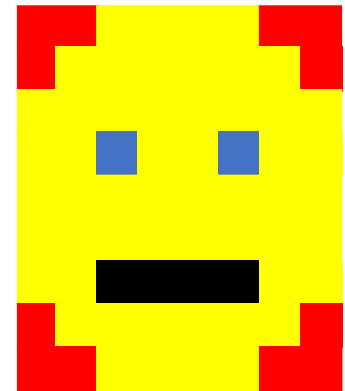


Image en couleurs



255	255	255	255	255	255	255	255
0	0	255	255	255	255	0	0
0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255
0	255	255	255	255	255	255	0
0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
0	0	0	0	0	0	0	0
255	255	0	255	255	0	255	255
255	255	0	255	255	0	255	255
0	0	255	0	0	255	0	0
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
0	0	0	0	0	0	0	0
255	255	0	0	0	0	255	255
255	255	0	0	0	0	255	255
0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255
0	0	255	255	255	255	0	0
0	0	0	0	0	0	0	0
255	255	255	255	255	255	255	255
0	0	0	0	0	0	0	0

24 bits (ou 3 octets) par points soit 216 octets pour l'image :



Nécessité de compression

Exemple 2 : 1 minute de vidéo en résolution 4K
(sans sons)

$$\begin{array}{rcl} & 3\,840 & \text{Points en largeur} \\ \times & 2\,160 & \text{Points en hauteur} \\ \times & 3 & \text{Octets par points} \\ \times & 25 & \text{Images par seconde} \\ \times & 60 & \text{secondes} \\ \hline \approx & 8,3 & \text{Go de données} \end{array}$$

B – Nature de la compression

Compression = Techniques de codage permettant de diminuer la taille des informations à transmettre ou à stocker.

- **La compression sans perte :**

A l'issue du processus de décompression, le flux de données obtenu est strictement identique à celui précédant le processus de compression.

- **La compression avec perte**

A l'issue du processus de décompression, le flux de données obtenu ressemble mais n'est plus identique à celui précédant le processus de compression.

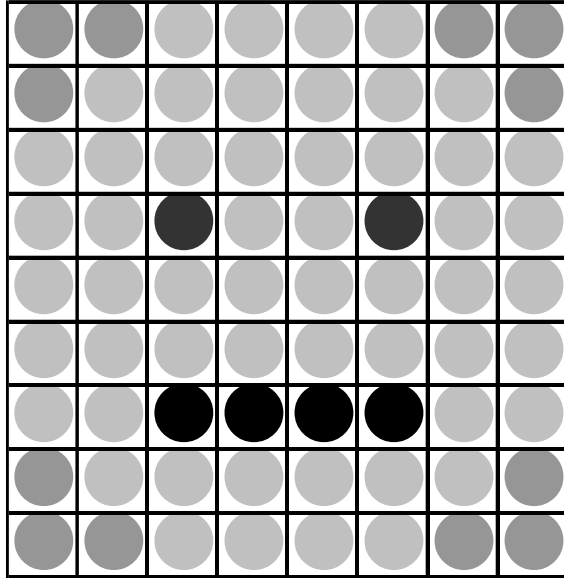
C - La compression sans perte

- Méthode :
 - Des algorithmes spéciaux reconnaissent les ensembles similaires au sein des données à traiter (des redondances) et les codent de manière à ce que l'ensemble soit réduit.
 - Exemple : *S'il y a une série de 100 valeurs '0' successives, cette série est codée par 100 x '0' ce qui occupe moins de place.*
 - Exemples de Méthodes : Run Length Encoding, Huffman, Lempel Zempel Welch, ADPCM
- Maintien de la qualité
- Faible taux de compression : division de la taille en par 2 à 5.

C - La compression sans perte

- Méthode RLE
 - Les motifs redondants sont remplacés par un marquage spécial, détectable lors de la décompression, permettant d'exprimer de manière concise leur multiplicité.
- Méthode de compression à partir d'un dictionnaire
 - Un dictionnaire (optimisé au regard de la fréquence d'apparition) des motifs contenus est constitué et le contenu du document est remplacé par des chemins d'accès binaires vers les éléments de ce dictionnaire optimisé.

La compression RLE (Run length Encoding)



128	128	192	192	192	192	128	128
128	192	192	192	192	192	192	128
192	192	192	192	192	192	192	192
192	192	64	192	192	64	192	192
192	192	192	192	192	192	192	192
192	192	192	192	192	192	192	192
192	192	0	0	0	0	192	192
128	192	192	192	192	192	192	128
128	128	192	192	192	192	128	128

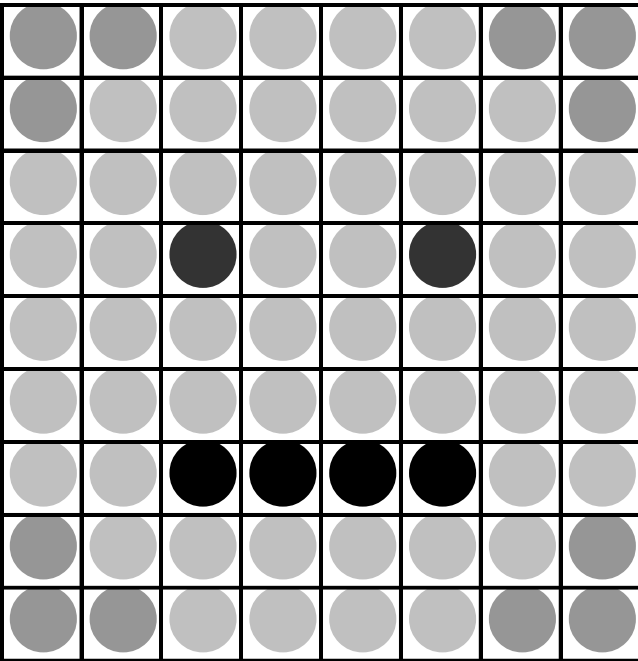
	Marqueur de répétition
	Nombre de répétitions
	Valeur à répéter

128	128	255	4	192	128	128	
128	255	6	192	128			
255	10	192					
64	192	192	64	255	20	192	
255	4	0	192	192			
128	255	6	192	128			
128	128	255	4	192	128	128	

Rapport de compression : 54%

La compression par dictionnaire

Etape 1 – Décompte des motifs



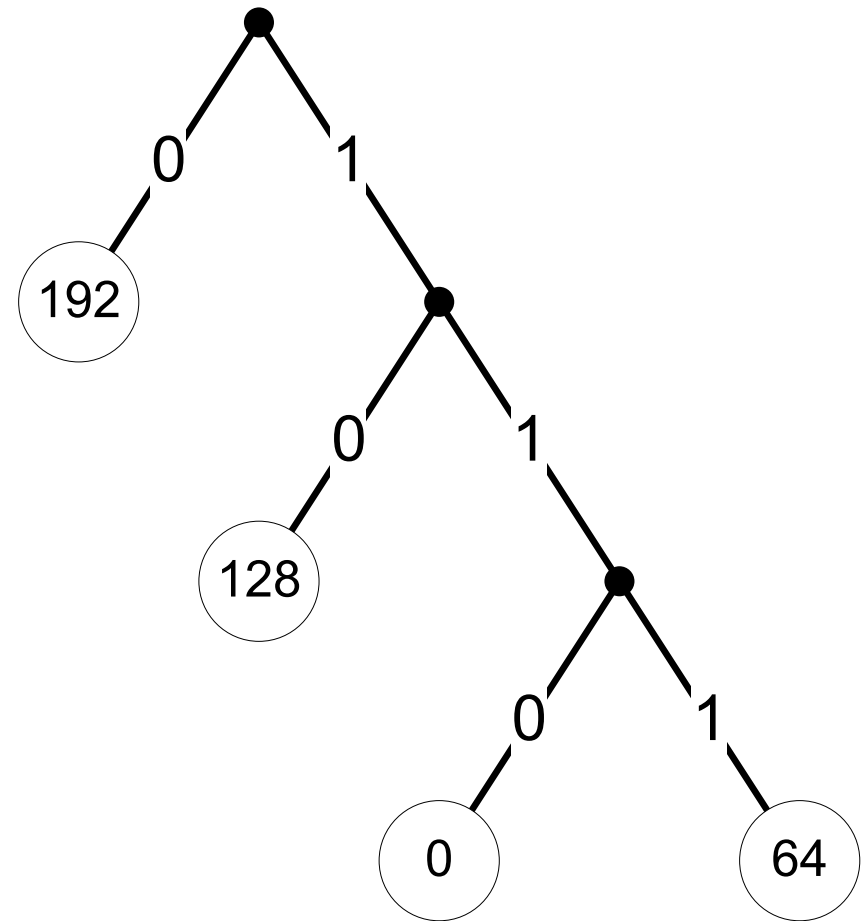
128	128	192	192	192	192	128	128
128	192	192	192	192	192	192	128
192	192	192	192	192	192	192	192
192	192	64	192	192	64	192	192
192	192	192	192	192	192	192	192
192	192	192	192	192	192	192	192
192	192	0	0	0	0	192	192
128	192	192	192	192	192	192	128
128	128	192	192	192	192	128	128

Valeur	Occurrences
192	46
128	12
0	4
64	2

La compression par dictionnaire

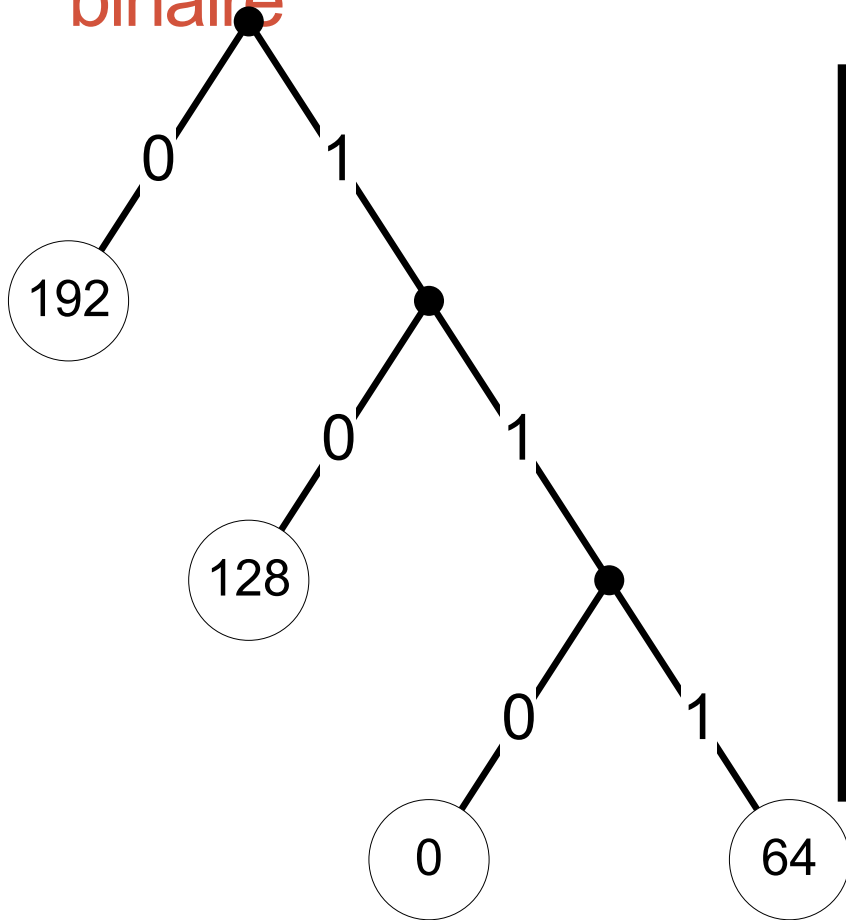
Etape 2 – Construction de l'arbre

Valeur	Occurrences
192	46
128	12
0	4
64	2



La compression par dictionnaire

Etape 3 – Transformation de l'arbre en chemin binaire



Valeur	Chemin
192	0
128	10
0	110
64	111

La compression par dictionnaire

Etape 3 – Codage

128	128	192	192	192	192	128	128
128	192	192	192	192	192	192	128
192	192	192	192	192	192	192	192
192	192	64	192	192	64	192	192
192	192	192	192	192	192	192	192
192	192	192	192	192	192	192	192
192	192	0	0	0	0	192	192
128	192	192	192	192	192	192	128
128	128	192	192	192	192	128	128

1	0	1	0	0	0	0	0	1	0	1	0				
1	0	0	0	0	0	0	0	1	0						
0	0	0	0	0	0	0	0								
0	0	1	1	1	0	0	1	1	1	0	0				
0	0	0	0	0	0	0	0								
0	0	0	0	0	0	0	0								
0	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0
1	0	0	0	0	0	0	0	1	0						
1	0	1	0	0	0	0	0	1	0	1	0				

$$\text{Rapport de compression} = \frac{96 \text{ bits}}{64 \times 9 \text{ bits}} = 17\%$$

D - La compression avec perte

- **Méthode** : Dégradation préalable des données dans un sens qui maximise le gain de compression lors de l'application d'une méthode de compression **sans** perte
- Perte de qualité évoluant de pair avec la compression
- Fort taux de compression : division de la taille entre 10 et 50
- Exemples de méthodes : MP3, JPEG, MPEG niveau 1,2 et 3

Compression avec perte de données sonores : méthode MP3

Méthode :

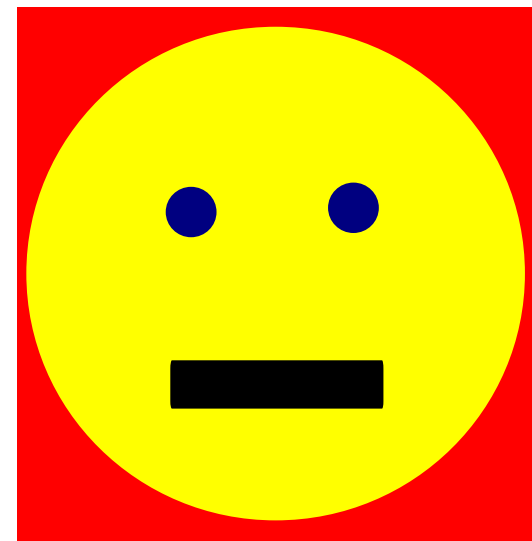
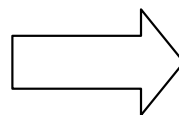
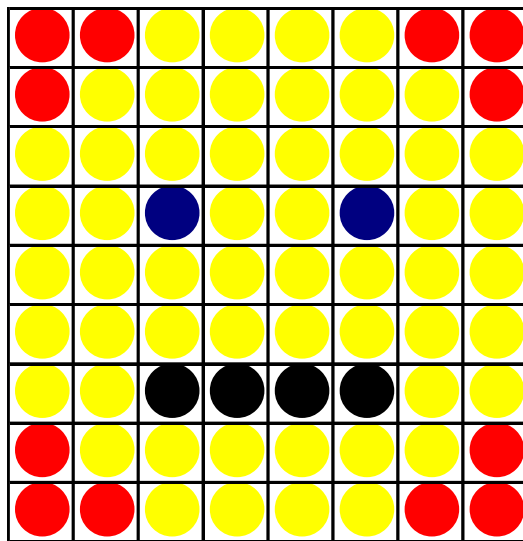
- Élimination des sons inaudibles
- Application d'un modèle psychoacoustique de masquage des sons et des canaux
- Suppression des informations communes à plusieurs canaux
- Application d'une méthode de compression sans perte sur le résultat

Compression avec perte d'images fixes

Vectorisation :

La vectorisation a pour but de décomposer l'image sous forme d'une série de formes géométriques et n'enregistrer que les coefficients des équations mathématiques qui décrivent ces formes

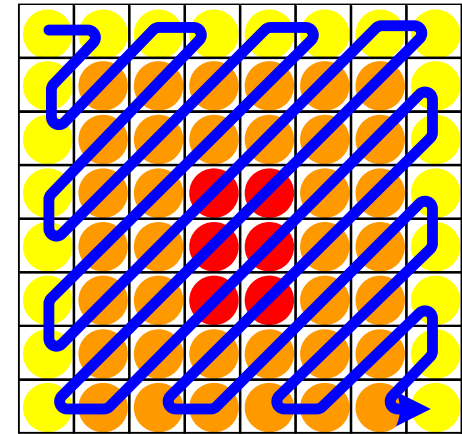
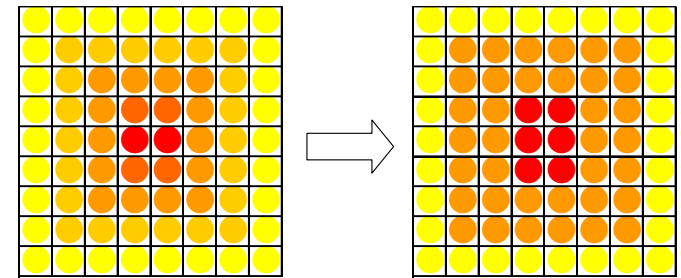
Variantes : Compression fractale, compression par ondelettes.



Compression avec perte d'images fixes : JPEG

Les étapes de la compression JPEG sont les suivantes :

- **Rééchantillonnage de la chrominance** (information sur la couleur, à l'opposé de la luminance, information sur le noir et blanc), car l'oeil ne peut discerner de différences de chrominance au sein d'un carré de 2x2 points
- **Découpage de l'image en blocs de 8x8 points**, puis l'application de la fonction DCT (*Discrete Cosinus Transform*, transformation discrète en cosinus) qui décompose l'image en somme de fréquences
- **Quantification** de chaque bloc, c'est-à-dire qu'il applique un coefficient de perte (qui permet de déterminer le ratio taille/qualité) "annulera" ou diminuera des valeurs de hautes fréquences, afin d'atténuer les détails en parcourant le bloc intelligemment avec un codage RLE (en zig-zag pour limiter les ruptures).
- **Encodage de l'image** puis compression avec la méthode d'Huffman (compression avec dictionnaire)



Compression avec pertes d'images animées : MPEG

Méthode :

- A partir d'une image clé compressée en format JPEG, mémorisation des seules zones en mouvement
- Pour la suite de la séquence : mémorisation des seules différences entre les images (entre une image et sa précédente ou sa suivante) après une estimation et une compensation du mouvement entre les images.
- Compression du son en MP3

II - Structuration des données

A – Les matériels et outils de sauvegarde

B - L'organisation logique des données

1. Le modèle relationnel
2. Les bases NO-SQL

C - Le Big Data et les technologies sous-jacentes

1. Virtualisation
2. Cloud computing
3. Intelligence artificielle et Apprentissage automatique

D - Les blockchains

A – Les matériels et outils de sauvegarde

1- Les matériels

- Les disques SSD (Solid State Disks)
- Les disques magnétiques
- Les systèmes RAID (Redundant Array of Independant Disks)
- Les disques optiques
- Les bandes LTO (*Linear Tape-Open*)

A – Les matériels et outils de sauvegarde

2 – Les méthodes de sauvegarde

- Sauvegarde complète
- Sauvegarde différentielle
 - Sauvegarde des données et fichiers modifiés depuis la dernière sauvegarde complète
- Sauvegarde incrémentielle
 - Sauvegarde des données et fichiers modifiés depuis la dernière sauvegarde (complète ou incrémentielle)
- Règle 3-2-1
 - 3 exemplaires des données avec des sauvegardes sur 2 supports différents dont 1 hors site

B - L'organisation logique des données

1 - Le modèle relationnel

Proposé par CODD en 1970.

Fondé sur la théorie mathématique des ensembles et proposant une algèbre pour manipuler les données.

A mis 10 ans pour s'imposer puis règne depuis les années 80 dans les bases de données utilisées par les entreprises pour la gestion de leurs activités.

Un élément de son succès fut également l'utilisation de langages unifiés d'interrogation de données : SQL et QBE.

Plan :

a – Concepts

b – Propriétés ACID

a – Les concepts du modèle relationnel

a1 / La relation / table

Soit un objet ou un individu à mémoriser en base de données, celui-ci est caractérisé par des propriétés que l'on appelle **attributs** ou **champs**. Une **relation** (ou **table**) est un ensemble d'attributs associés pour décrire un objet ou individu. Pour chaque individu ou objet, il y aura un **tuple**, **n-uplet** ou **enregistrement** qui présentera au plus une valeur au niveau de chaque attribut.

En général, on ne s'intéresse qu'aux attributs que l'on doit nécessairement mémoriser et les attributs calculés (ou déduits d'autres attributs) n'apparaissent pas.

Présentation en intention : Relation

ETUDIANT(Num, Nom, Prénom, Courriel)

Ligne = **tuple**, **n-uplet** ou **enregistrement**

Présentation en extension : Table

ETUDIANT

<u>Num</u>	Nom	Prénom	Courriel
20230001	DUPONT	Henri	henri.dupont@
20230002	DUBOIS	Paul	paul.dubois@
20230003	MARTIN	Marie	marie.martin@

Colonne = **attribut** ou **champ**

a – Les concepts du modèle relationnel

a2 / La clé primaire

La clé primaire est l'attribut ou les attributs d'une relation qui permettent d'identifier sans ambiguïté les différents enregistrements de la relation.

- Deux enregistrements d'une même relation/table ne peuvent avoir la même valeur de clé primaire.
- Un enregistrement ne peut pas avoir une clé primaire sans valeur.

Ainsi pour une valeur de la clé primaire, on peut déterminer les valeurs de tous les autres attributs de la relation

ETUDIANT			
<u>Num</u> 	Nom	Prénom	Courriel
20230001	DUPONT	Henri	henri.dupont@
20230002	DUBOIS	Paul	paul.dubois@
20230003	MARTIN	Marie	marie.martin@

a – Les concepts du modèle relationnel

a3 / La clé étrangère

Une clé étrangère est un attribut d'une relation dont les valeurs font référence à celles de la clé primaire d'une autre relation.

ETUDIANT

<u>Num</u> 	Nom	Prénom	Courriel	<u>N°Diplôme</u>
20230001	DUPONT	Henri	henri.dupont@	1
20230002	DUBOIS	Paul	paul.dubois@	2
20230003	MARTIN	Marie	marie.martin@	1

DIPLOME

<u>N°Diplôme</u> 	Libellé	Niveau
1	Licence	Bac+3
2	Master	Bac+5
3	Doctorat	Bac+8



a – Les concepts du modèle relationnel

Relations *1 à plusieurs* : ETUDIANT $\rightarrow$ DIPLÔME

ETUDIANT

<u>Num</u> 	Nom	Prénom	Courriel	N°Diplôme
20230001	DUPONT	Henri	henri.dupont@	1
20230002	DUBOIS	Paul	paul.dubois@	2
20230003	MARTIN	Marie	marie.martin@	1

DIPLÔME

<u>N°Diplôme</u> 	Libellé	Niveau
1	Licence	Bac+3
2	Master	Bac+5
3	Doctorat	Bac+8



Relations *Plusieurs à plusieurs* : ETUDIANT $\rightleftarrows$ COURS

ETUDIANT

<u>Num</u> 	Nom	Prénom	...
20230001	DUPONT	Henri	...
20230002	DUBOIS	Paul	...
20230003	MARTIN	Marie	...

INSCRIPTION PEDAGOGIQUE

<u>Num étudiant</u> 	<u>Code Cours</u> 
20230002	LIC2
20230002	LIC1
20230001	LIC1

COURS

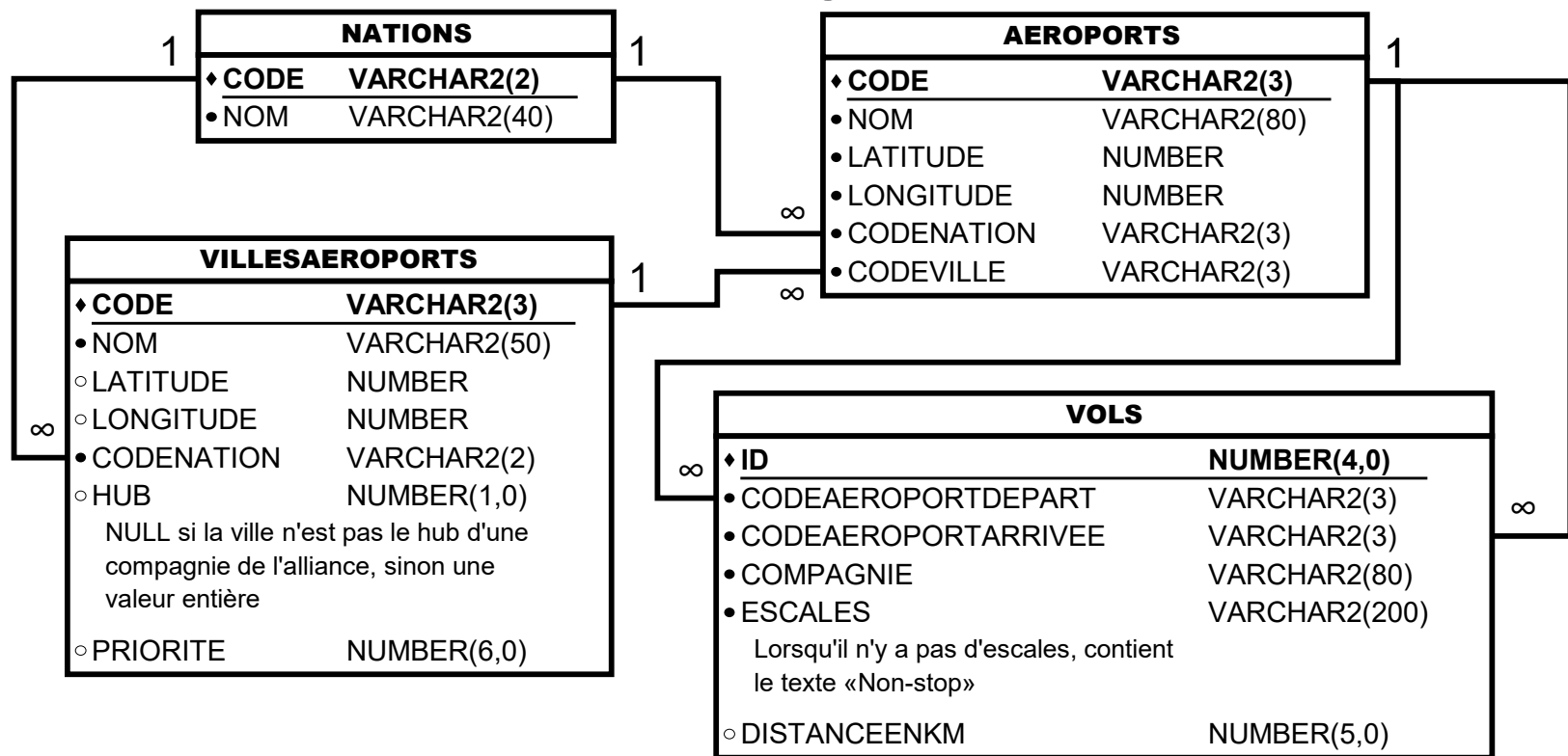
<u>Code Cours</u> 	Libellé	ECTS
LIC1	Droit	3
LIC2	Economie	2
LIC3	Informatique	2



a – Les concepts du modèle relationnel

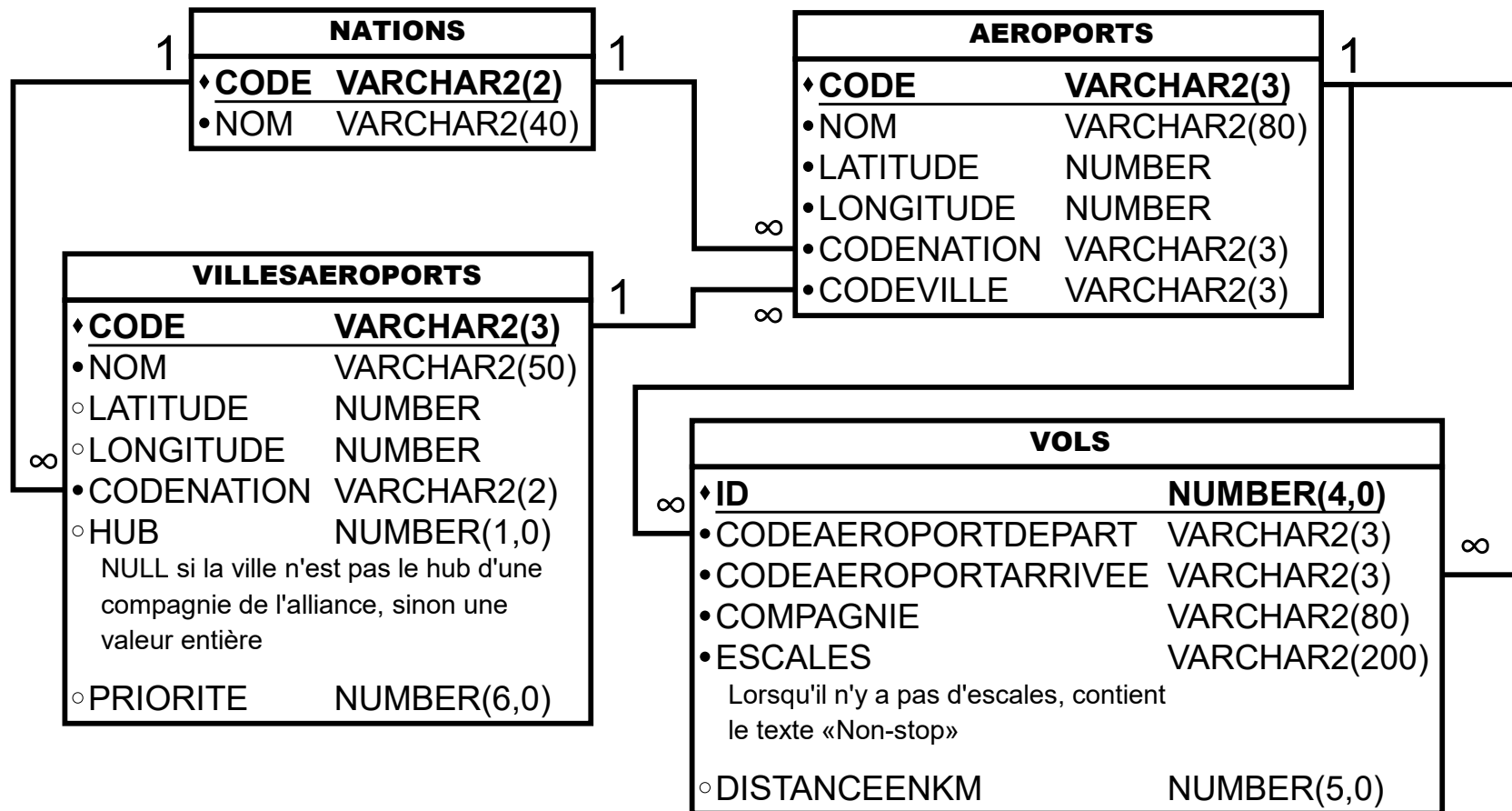
a4 - Schéma relationnel

Dans une base de données, les données sont ainsi réparties en plusieurs tables organisées en schéma



Remarque : les traits 1 ∞ peuvent également être représentés par la notation

Base de données d'application



Remarque : les traits 1 ∞ peuvent également être représentés par la notation 

Base de données d'application

1. Une table NATIONS

La table **NATIONS** présente les pays pouvant être au départ ou à la destination d'un vol.

Extrait de la table NATIONS

CODE	NOM
FR	France
NL	Pays-Bas
DK	Danemark
NO	Norvège
GB	Royaume-Uni
US	États-Unis d'Amérique
...	...

Base de données d'application

2. Une table VILLES AEROPORTS

La table **VILLES AEROPORTS** présente les villes pouvant être au départ ou à la destination d'un vol. Lorsqu'une ville est un hub elle contient une valeur numérique dans le champ **HUB**, sinon ce dernier est vide (NULL). Le champ **PRIORITE** contient une valeur numérique d'autant plus forte que le groupe y est bien implanté, afin de pouvoir mettre ces villes en avant.

Extrait de la table **VILLES AEROPORTS** :

CODE	NOM	LATITUDE	LONGITUDE	CODENATION	HUB	PRIORITE
IXC	Chandigarh	30,73	76,79	IN		0
JAI	Jaïpur	26,91	75,81	IN		0
KIN	Kingston	17,95	-76,79	JM		0
LAI	Lannion	48,75	-3,48	FR		0
LEI	Almería	36,84	-2,46	ES		0
...	...	...	...	...	...	...
PAR	Paris	48,86	2,35	FR	1	10071
AMS	Amsterdam	52,37	4,9	NL	1	10000
...	...	...	...	...	...	...

Base de données d'application

3. Une table AEROPORTS

La table **AEROPORTS** présente les aéroports pouvant être au départ ou à la destination des villes. Chaque aéroport est associé à une ou plusieurs desservies et à leur nation d'appartenance.

Extrait de la table **AEROPORTS** :

Code	Nom	Latitude	Longitude	Codeville	CodeNation
PUJ	Aéroport de Punta Cana	18,5667	-68,3833	PUJ	DO
PUS	Aéroport de Gimbae	35,1731	128,9464	PUS	KR
TAO	Aéroport de Qingdao	36,2661	120,3744	TAO	CN
BQC	Sainte-Foy Bus Station	46,7728	-71,3011	YQB	CA
YQB	Aéroport de Quebec	46,7911	-71,3933	YQB	CA
QRO	Aéroport de Queretaro	20,6203	-100,1881	QRO	MX
UIP	Aéroport de Pluguffan	47,975	-4,1678	UIP	FR
UIO	Aéroport de Mariscal Sucre	-0,1133	-78,3586	UIO	EC
RBA	Aéroport de Sale	34,0514	-6,7514	RBA	MA
...	...	...	...	...	...

Base de données d'application

4. Une table VOLS

La table **VOLS**, présente chacun des vols, reliant un aéroport de départ à un aéroport d'arrivé, assuré par une compagnie de l'alliance, avec ou sans escale pour une distance donnée. Lorsqu'un vol ne comporte pas d'escale, son champ **ESCALES** présente la valeur « Non-stop » sinon contient un texte listant une ou plusieurs escales.

Extrait de la table **VOLS** :

ID	CODEAEROPORTDEPART	CODEAEROPORTARRIVEE	COMPAGNIE	ESCALES	DISTANCEENKM
432	ATL	BMI	KLM (liaison assurée par ExpressJet)	Non-stop	859
433	CDG	BOG	Air France	Non-stop	8653
434	AMS	BOG	KLM	Non-stop	8850
435	SEA	BOI	Air France (liaison assurée par SkyWest Airlines)	Non-stop	642
436	MSP	BOI	Air France (liaison assurée par Delta Air Lines)	Non-stop	1835

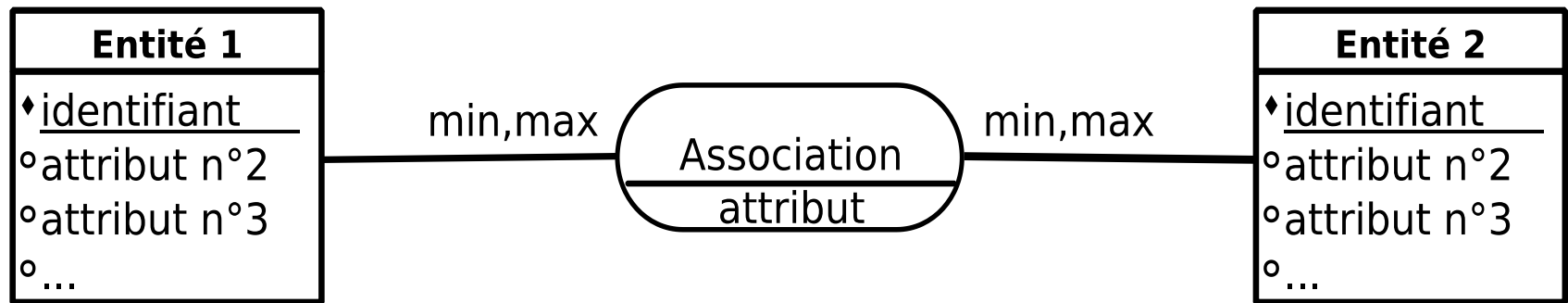
a – Les concepts du modèle relationnel

a5 – La logique du modèle relationnel

1. Lors de la création de la base de données, on répartit l'ensemble d'information sous forme de données dans plusieurs table pour éviter toute redondance au sein de ces tables.
2. Lors de l'exploitation des données dans les tables, on reconstruit grâce des opérations que l'on appelle *jointure* qui mettent en correspondance les données de deux ou plusieurs tables, l'ensemble global d'information.
3. On peut ensuite créer des rapports (états) ou des formulaires de saisie en utilisant les outils du système de base de données ou des applications web.

b – Création d'un schéma relationnel à partir d'un Modèle Conceptuel des Données

Un modèle conceptuel des données associe des entités porteuses d'attributs, représentant et détaillant les objets ou individus représentés dans la base de données, par des associations (également susceptibles de porter des attributs) pour établir les liens entre les données mémorisées. L'identifiant de l'entité a le même rôle que la clé primaire : identifier



Les cardinalités entre une entité et une association dénotent les modalités de sa participation :

Cardinalité minimale :

0 → Participation optionnelle (un membre de l'entité peut ne jamais participer à l'association)

1 → Participation obligatoire (un membre de l'entité participe systématiquement à l'association)

Cardinalité maximale :

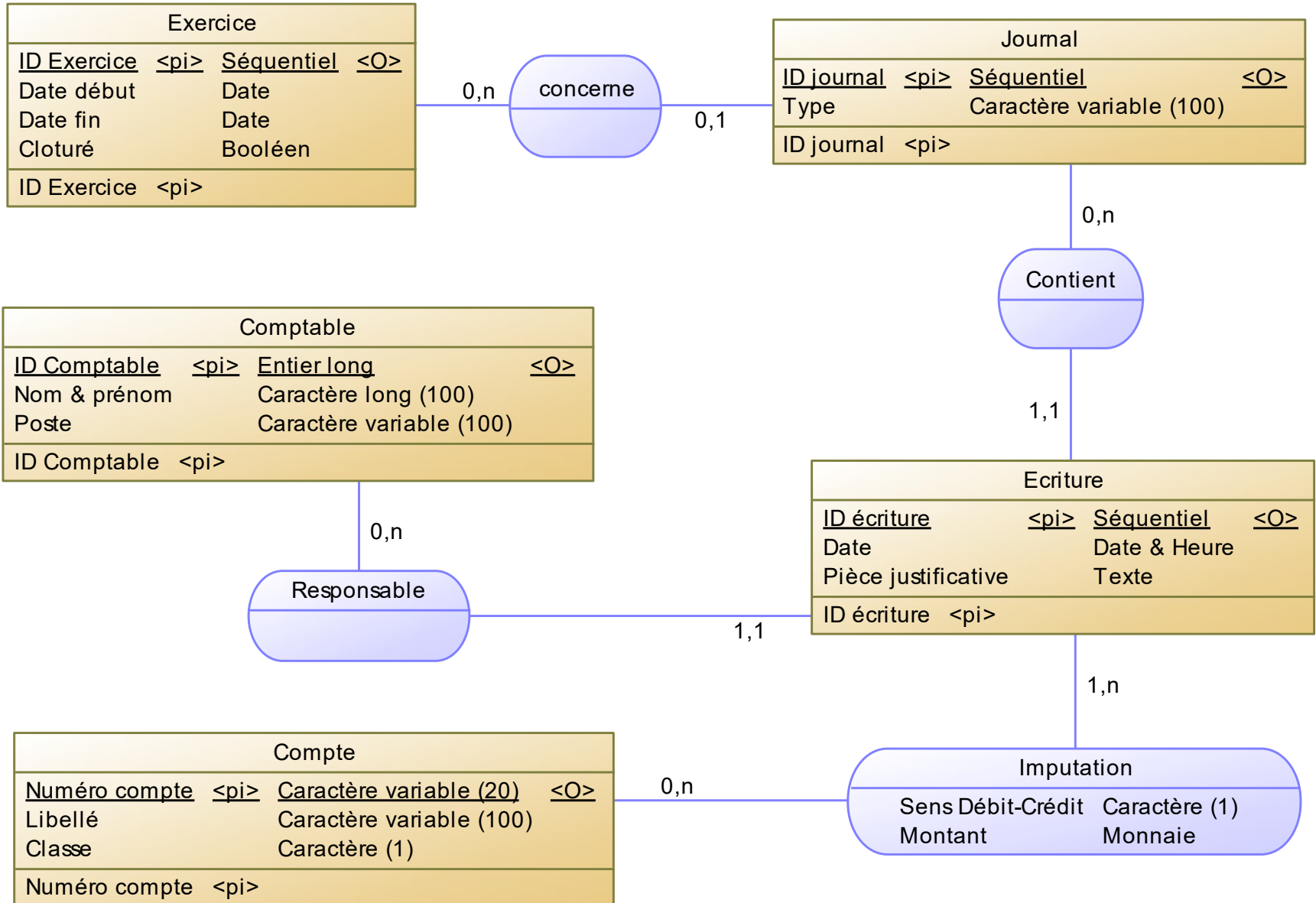
1 → Participation unique (un membre de l'entité participe au plus une seule fois à l'association)

n → Participation multiple (un membre peut participer plusieurs fois à l'association)

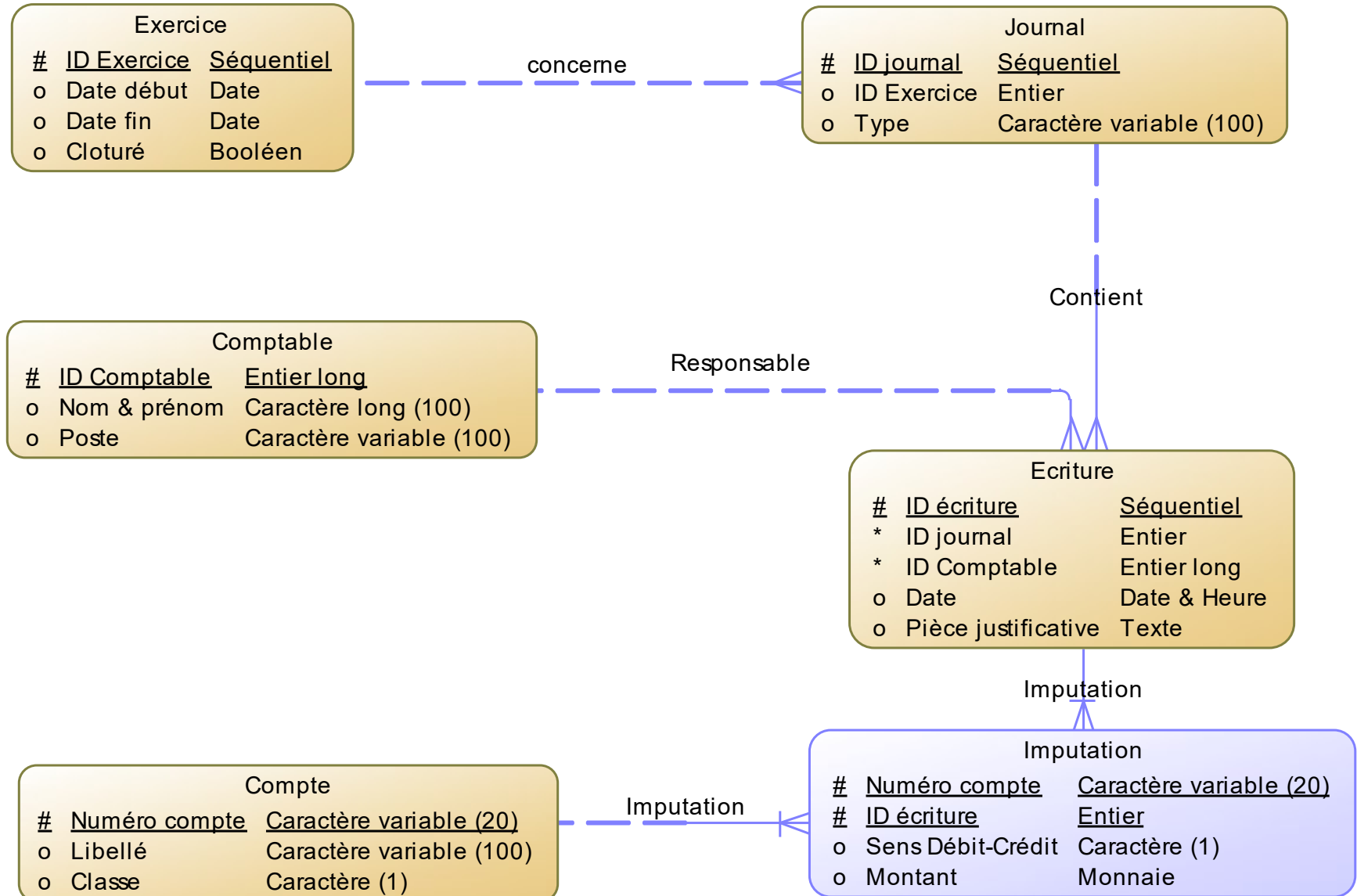
b – Création d'un schéma relationnel à partir d'un Modèle Conceptuel des Données

MCD		Schéma relationnel
<u>Règle n°1</u> : Une entité avec ses identifiant et attributs	➡	Relation avec clé primaire issue de l'identifiant de l'entité et attributs
<u>Règle n°2</u> : Association avec une des cardinalités maximales à 1 et l'autre à n	➡	Clé étrangère dans la relation correspondant au côté du 1 et référençant la relation correspondant au côté du n
<u>Règle n°3</u> : Association avec ses cardinalités maximales à n et ses attributs	➡	Relation avec clé primaire composée des deux clés étrangères référençant les relations correspondant aux entités participant à l'association, complétée des attributs de l'association

Modèle conceptuel de données



Modèle relationnel dérivé



c – Propriétés ACID

Ces propriétés découlent du concept complémentaire de TRANSACTION

- **A** : ATOMICITE
- **C** : COHERENCE
- **I** : ISOLATION
- **D** : DURABILITE

2 – Les bases NOSQL

Crées par les géants du Web (Google, Amazon, Facebook, Microsoft,...) puis mises à disposition en open-source, elles se veulent complémentaires aux bases de données relationnelles. Elles visent à renoncer à des propriétés ACID, notamment en privilégiant la disponibilité à la cohérence, ou proposent une simplification du schéma en vue d'accroître les performances ou de permettre le stockage de données de grands sites Web et réseaux sociaux.

- 1 – Bases clé-Valeur
- 2 – Bases clé-document
- 3 – Bases orientées colonne
- 3 – Bases orientées graphe

1 – Bases clé-valeur

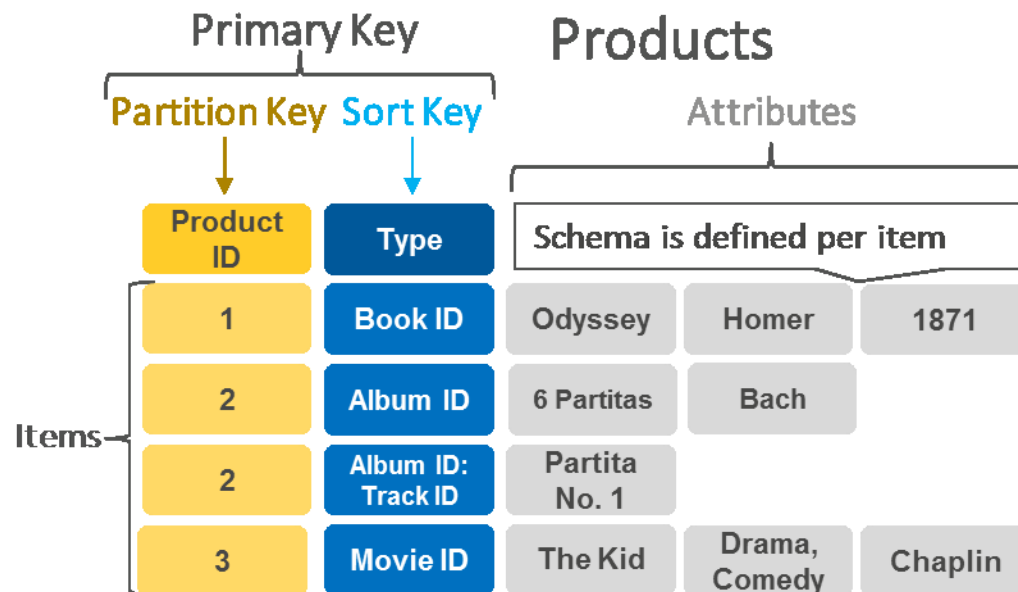
Bases de structure extrêmement simple (une seule table) associant à une clé (présentant une contrainte d'unicité) une valeur pouvant prendre n'importe quel type (simple ou plus complexe).

Clé	Valeur											
1	Simple texte											
2	3,14159											
3	<table><tr><td>Num :</td><td>Nom :</td><td>Prénom :</td><td>Courriel :</td></tr><tr><td>20230001</td><td>DUPONT</td><td>Henri</td><td>henri.dupont@...</td></tr></table>				Num :	Nom :	Prénom :	Courriel :	20230001	DUPONT	Henri	henri.dupont@...
	Num :	Nom :	Prénom :	Courriel :								
20230001	DUPONT	Henri	henri.dupont@...									
...												

a – Bases clé-valeur

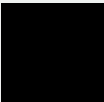
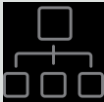
Ex : DynamoDB Amazon

<https://aws.amazon.com/fr/nosql/key-value/>



b – Bases orientées document

Bases associant un document de structure libre à une clé. La base reste simple tout en permettant de stocker des structures de données complexes car c'est l'application cliente qui gère la complexité du document, la base de donnée ne s'intéressant qu'à la mémorisation de son flux d'octets.

Clé	Document
1	
2	
3	
4	

b – Bases orientées document

MongoDB

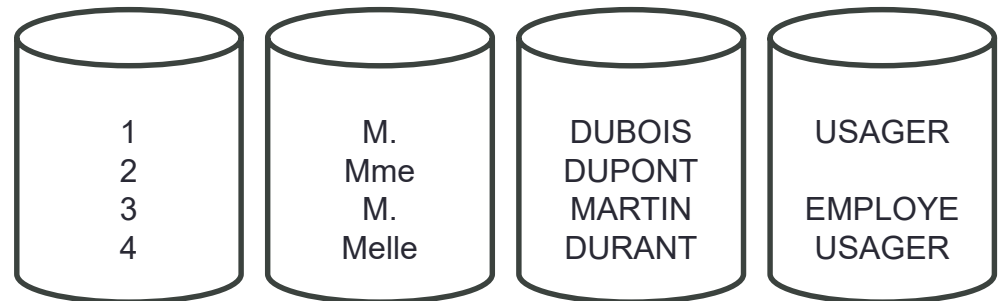
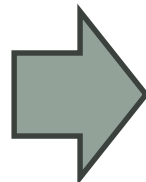
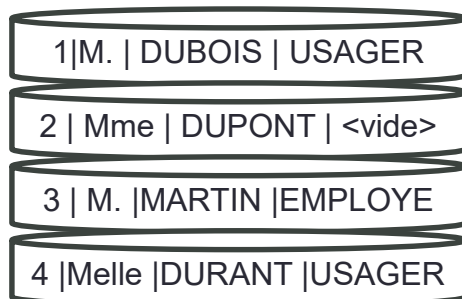
<https://www.mongodb.com/fr-fr>

c – Bases orientées colonne

Ces bases stockent les tables non ligne après ligne mais colonne après colonne et ne stockent pas les valeurs nulles (ou vides)

Ex: *Microsoft PowerBI*, *Microsoft PowerQuery sous Excel*, *Google BigTable*

N°	Civilité	NOM	QUALITE
1	M.	DUBOIS	USAGER
2	Mme	DUPONT	
3	M.	MARTIN	EMPLOYE
4	Melle	DURANT	USAGER

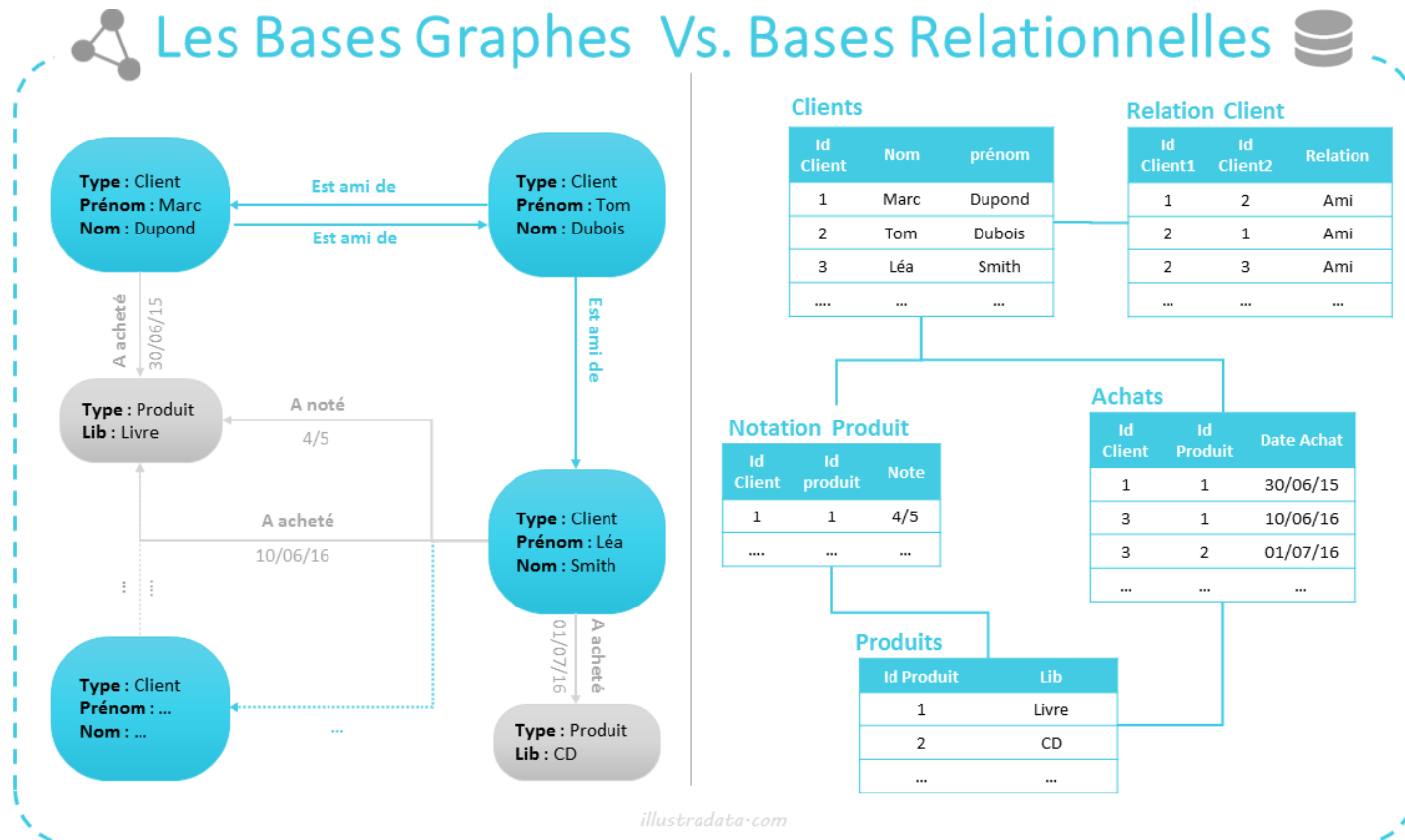


c – Bases orientées graphes

Plutôt que des relations/tables, ces bases mémorisent des graphes formés de sommets porteurs d'attributs reliés par des arcs, orientés ou non, eux-mêmes porteurs d'attributs.

Les Bases Graphes Vs. Bases Relationnelles

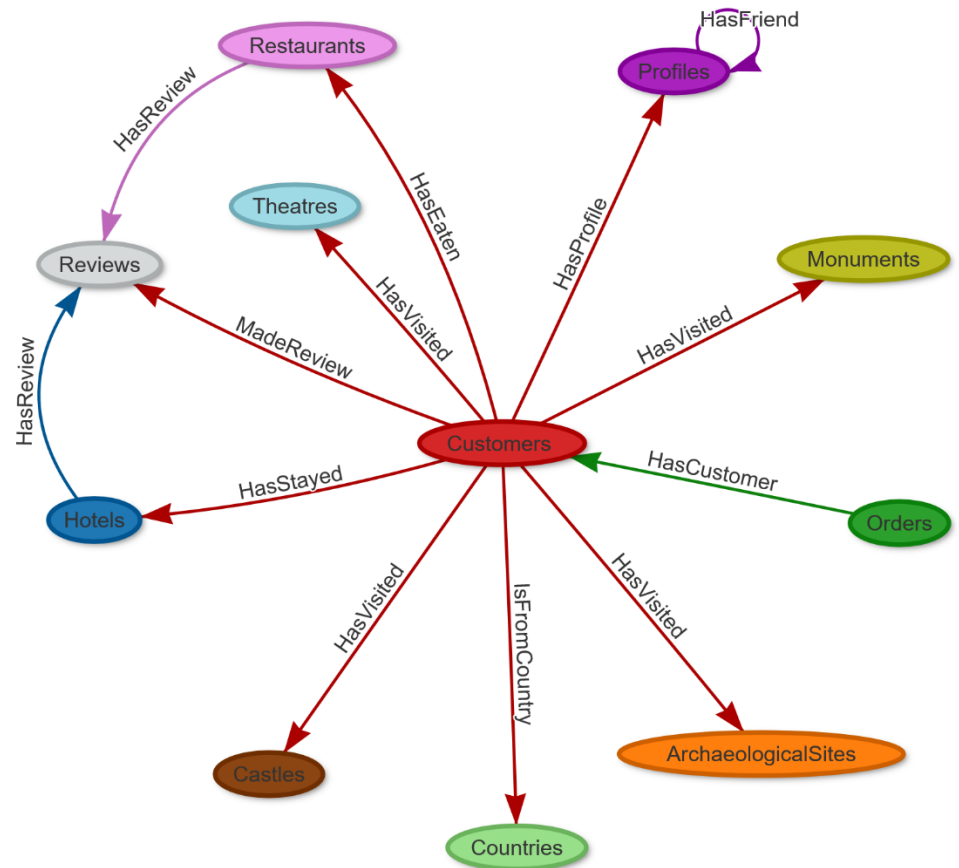
Source :
illustradata.com



d – Bases orientées graphes

Ex : SAP OrientDB

Schéma de sa base de démonstration



C - Le Big Data et les technologies sous-jacentes

Le BigData repose sur le constat que les organisations collectent et conservent un maximum de données dans le cadre du suivi de leur activité, dans le but de tirer un avantage de leur exploitation à grande échelle.

Les 4 **V** du Big Data :

- VOLUME** : Un stockage volumineux abordable
- + **VERSATILITE** : Bases structurées comme non structurées (NOSQL)
- + **VITESSE** : Un traitement efficace grâce à la distribution et l'IA

- = **VALEUR** : Une information utile pour l'activité

Pour cela, il s'appuie sur les bases NOSQL et d'autres évolutions :

1. Virtualisation
2. Cloud computing
3. Intelligence artificielle et Apprentissage automatique

1 – La virtualisation

C'est l'aptitude d'un système à se faire passer pour un autre système dans l'exécution des applications de ce dernier.

Avantages :

- Portabilité
- Sécurité
- Souplesse

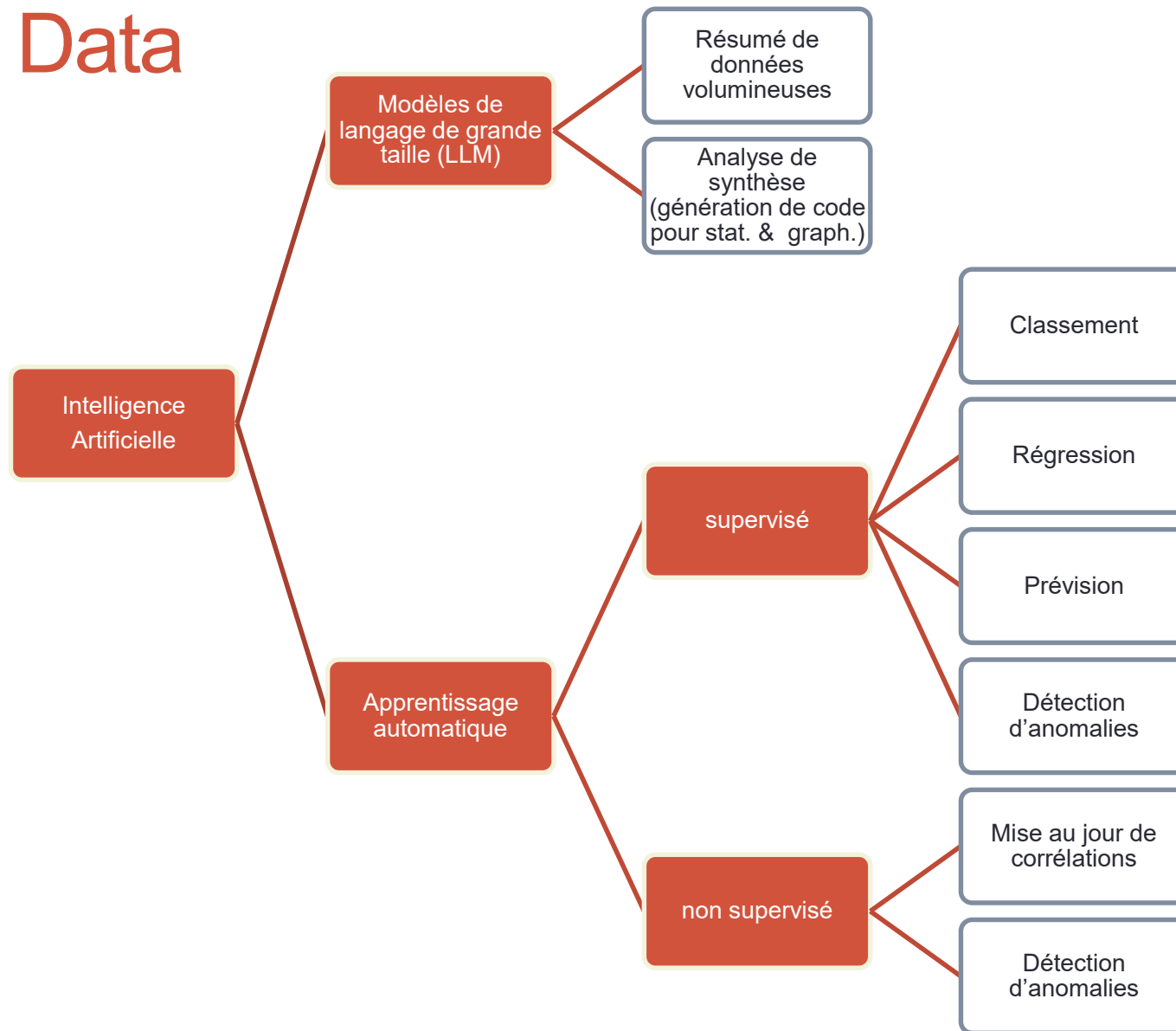
2 – Le Cloud Computing

Le cloud Computing consiste en la mise à disposition de services (stockage, traitement, ...) au travers d'internet entre un fournisseur qui bénéficie de la mutualisation de ses systèmes et un client qui limite ses investissements matériels.

Repose sur la virtualisation

Avantages	Inconvénients
• Informatique à la demande • Transformation de charges d'investissement en charges de fonctionnement plus flexibles. • Pré-connaissance des coûts • Recentrage sur le métier	• Dépendance du réseau • Données chez le prestataire de service - Sécurité - Réversibilité du contrat • Respect des réglementations sur les données personnelles et RGPD

3 – Apports de l'Intelligence Artificielle au Big Data



Les LLM Large Language Models – *Modèles de langage de grande taille*

Ils appliquent les résultats d'une branche en grand développement de l'Intelligence Artificielle

-  MICROSOFT Copilot / OPENAI ChatGPT <https://chatgpt.com>
-  GOOGLE Gemini  <https://gemini.google.com>
-  META Llama <https://llama.meta.com>
<https://huggingface.co/meta-llama>
<https://labs.perplexity.ai>
-  ANTHROPIC Claude Sonnet <https://claude.ai>
-  MISTRAL AI Mistral <https://mistral.ai/fr/>
- ...

Les LLM Large Language Models – *Modèles de langage de grande taille*

- Avantages :
 - Interrogation en langage courant
 - Réponses en langage courant
 - Capacité reconnue à synthétiser en un temps réduit de grands volumes d'information
 - Couplage possible avec des outils de rédaction/diffusion pour aider à la publication des réponses obtenues
 - Couplage possible avec des données privées
- Inconvénients :
 - Possibilité d'amplification de biais
 - Possibilité d'hallucination
 - Relative opacité des sources de la synthèses

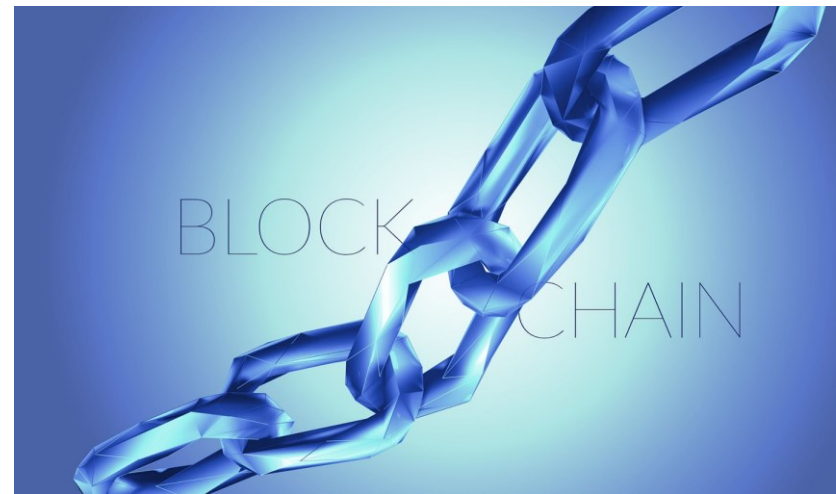
3 – Apports de l'Intelligence Artificielle au Big Data

APPRENTISSAGE AUTOMATIQUE :

- Domaines d'application :
 - Classification
 - Prévion
 - Détection d'anomalies
- Etapes : Entraînement → Validation → Production
- Deux types d'apprentissage automatique :
 - Apprentissage supervisé : A partir d'un jeu d'apprentissage contenant les bonnes conclusions
 - Apprentissage non supervisé : A partir d'un jeu d'apprentissage sans conclusions

D – Les Blockchains

- Définition d'un blockchain :
 - *Base de données distribuée et sécurisée créée en vue de recenser toutes les faits ou transactions établis depuis son origine.*
- 1 – Usage
- 2 – Modèle centralisé vs modèle distribué
- 3 – Deux types de blockchains
- 4 – Problématique inhérente à la distribution des blocs



1 - Usage

Grand Livre de transactions

- Règlements effectuées
- Engagements à payer (reconnaisances de dettes)
- Transferts d'actifs

Certification de droits

- Dépôt d'œuvres en vue d'attester une antériorité
- Contrats et éléments contractuels (engagements à faire ou à recevoir)

2 - Modèle Distribué

Modèle centralisé

- Une autorité de certification maintient une base de donnée centralisée
- Il y a autant d'autorités de certification qu'il y a de types de transaction
- Chaque autorité doit avoir la confiance des usagers concernant l'immuabilité de la base de données

Modèle distribué

- Usage des technologies de bases de données distribuées (NOSQL) rencontrées dans le Big Data
- La base de données est scindée en Blocs, répartis et partagés (dupliqués) sur le réseau
- Il n'y a plus d'autorité de gestion de la base

3 - Deux types de Blockchains

Publique

- Anonymat (relatif)
- Toutes les transactions depuis l'origine sont accessibles à tous.
- ***Preuve de travail*** (intensif) ou ***preuve d'enjeu*** nécessaire à l'endossement (validation) d'une transaction confiée à l'ensemble du réseau
- S'appuie sur une cryptomonnaie
- A l'échelle mondiale

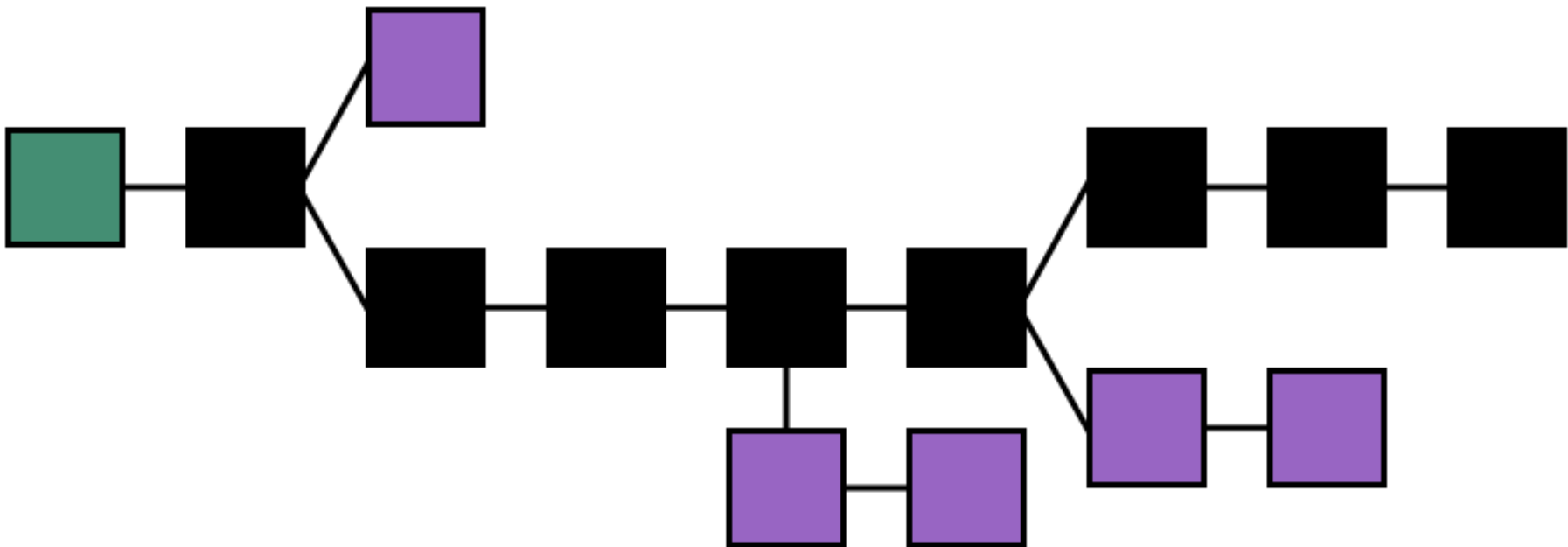
Privée

- Mécanisme de vérification d'identité et confidentialité des transactions.
- Les transactions sont signées et chiffrées.
- Endossement sélectif : un certain nombre d'utilisateurs certifiés peuvent valider une transaction.
- Ne s'appuie pas sur une cryptomonnaie : les transactions modifient les propriétés d'un actif (propriétaire, qualités, choix d'apparence...)
- Limitée au groupe d'affaires concerné

4 – Problématique inhérente à la distribution des blocs

Succession des blocs

Gestion des branches concurrentes et immutabilité de la blockchain



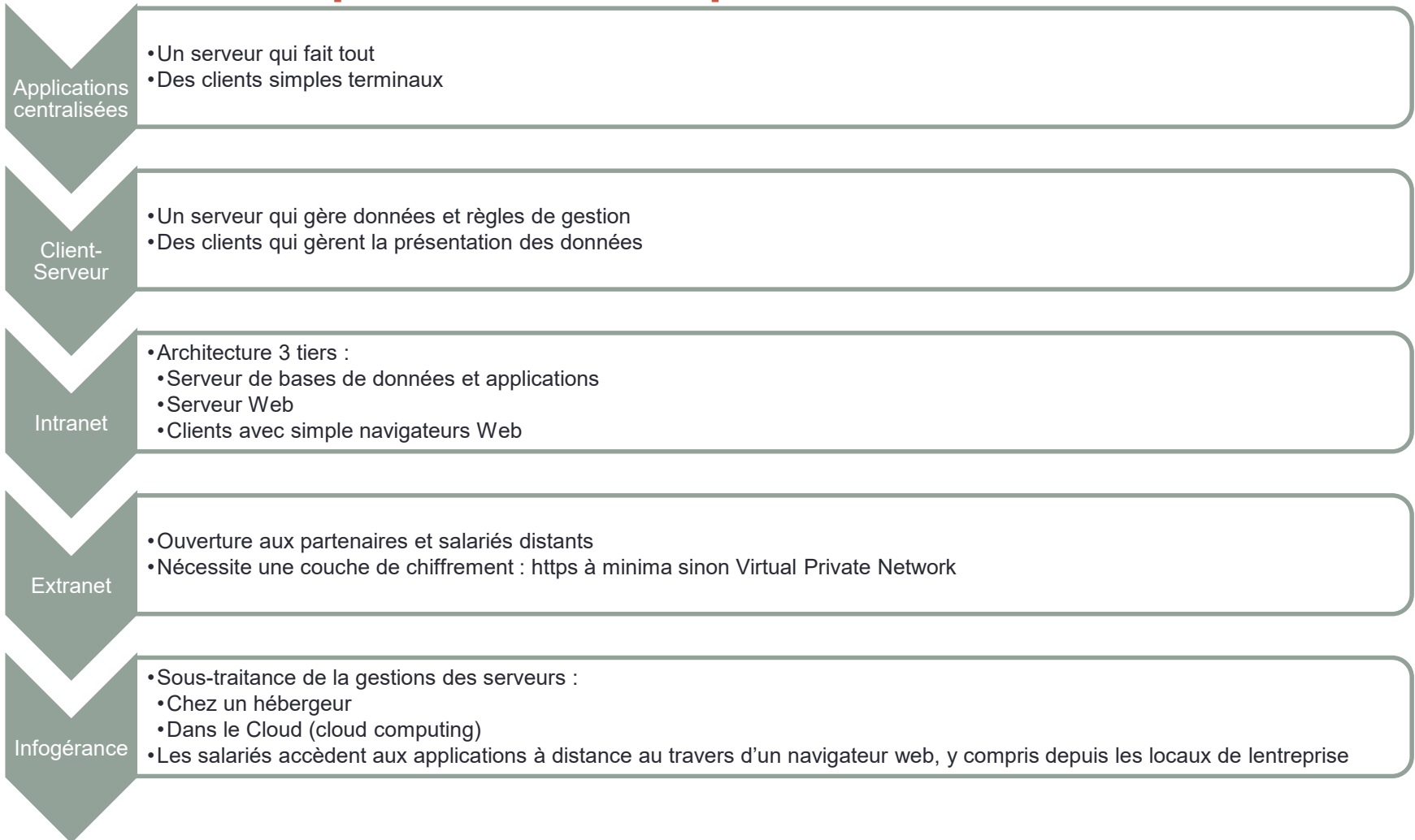
III – Traitement et diffusion de données

A – Structure des applications informatiques d'entreprise

B – L'intégration des données

C – Les ERP/PGI

A – Structure des applications informatiques d'entreprise



B – L'intégration des données

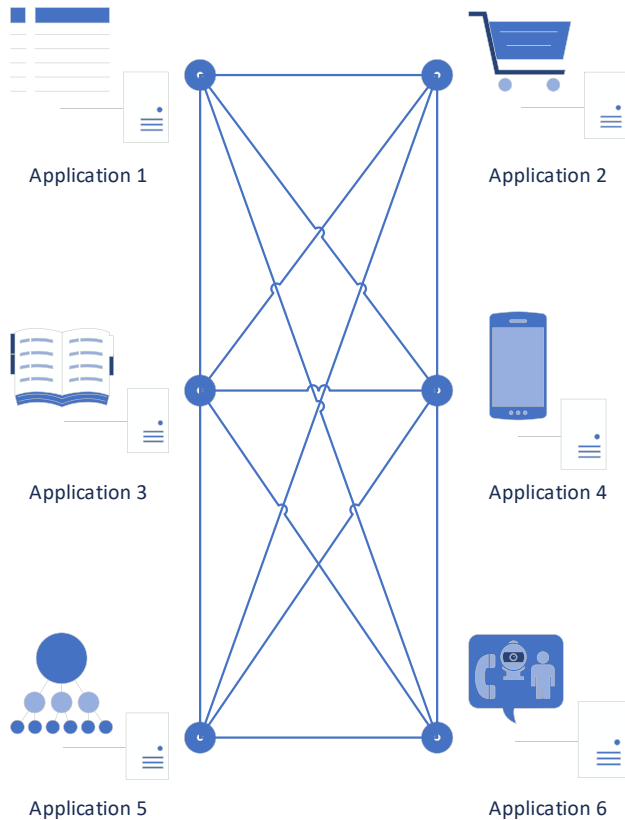
L'intégration des données vise à permettre la transmission d'information de données entre applications sans ressaisie, délai, ni erreur.

2 pistes :

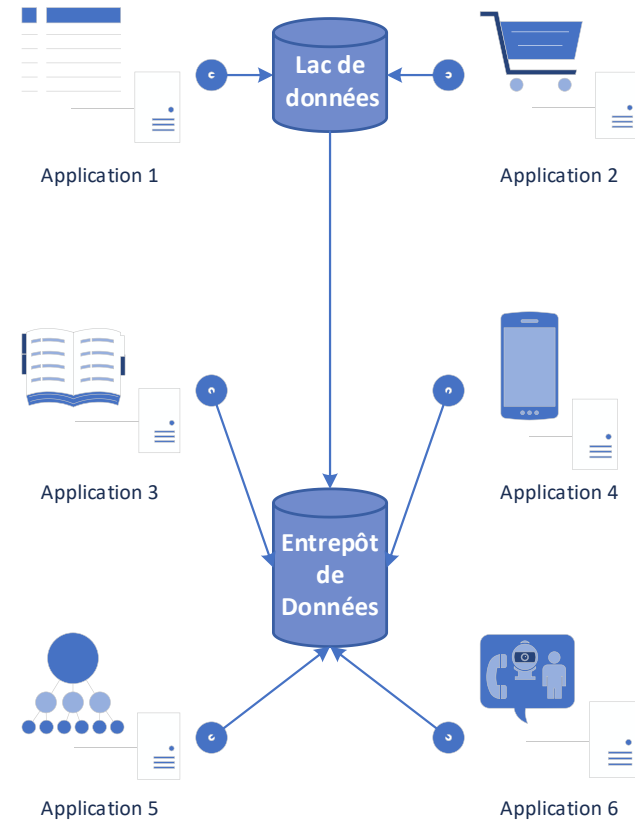
- **Enterprise Application Integration** :
 - Les applications retenues sont initialement indépendantes, choisies pour leur soutien aux spécificités des métiers
 - Des connecteurs ou middleware sont mis en place pour permettre l'échange de données entre applications ou entre les applications et des lacs ou entrepôts de données.
- **Enterprise Resource Planning**
 - Des modules applicatifs organisés autour d'une seule base de données centrale avec une communication en temps réel.

B – L'intégration des données

Connexion point à point des données



Concentration des données



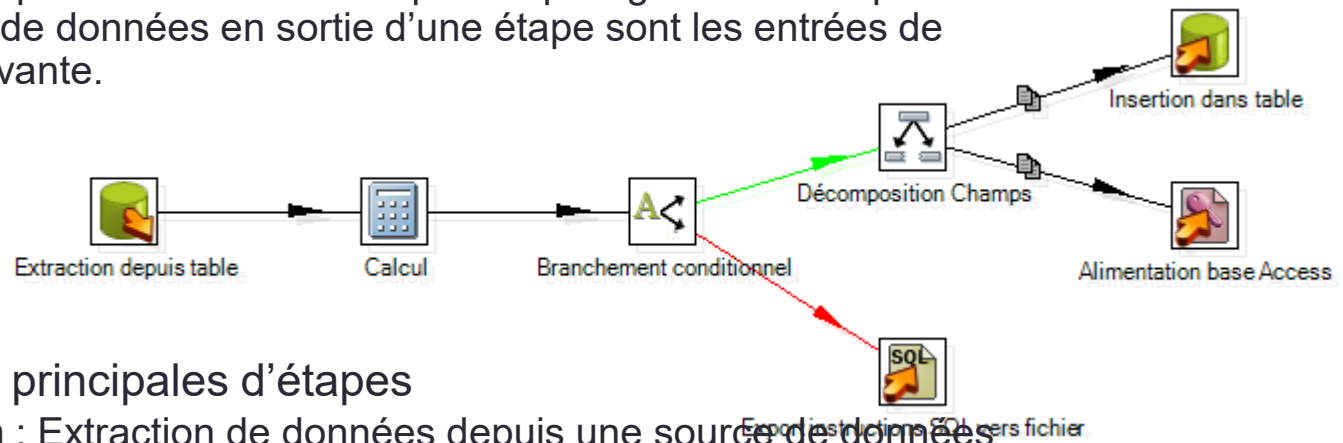
B – L'intégration des données

- Les connecteurs ou *middleware* sont :
 - Des scripts écrits dans un langage informatique
 - Des outils Extract-Transform-Load
- L'analyse des données peut être conduite par :
 - Des requêteurs
 - des outils de génération de rapports et tableaux de bord

B – L'intégration des données

- Outils ETL : Métaphore du Pipeline

- Mise en place d'une séquence d'étapes
 - Chaque étape est choisie dans la palette puis glissée sur le plan de travail
 - Les lignes de données en sortie d'une étape sont les entrées de l'étape suivante.



- 3 catégories principales d'étapes

- **Extraction** : Extraction de données depuis une source de données
- **Transformation** : Ajout /modification de champs
- **Alimentation** : Envoi des lignes transformées vers une base de données ou un fichier

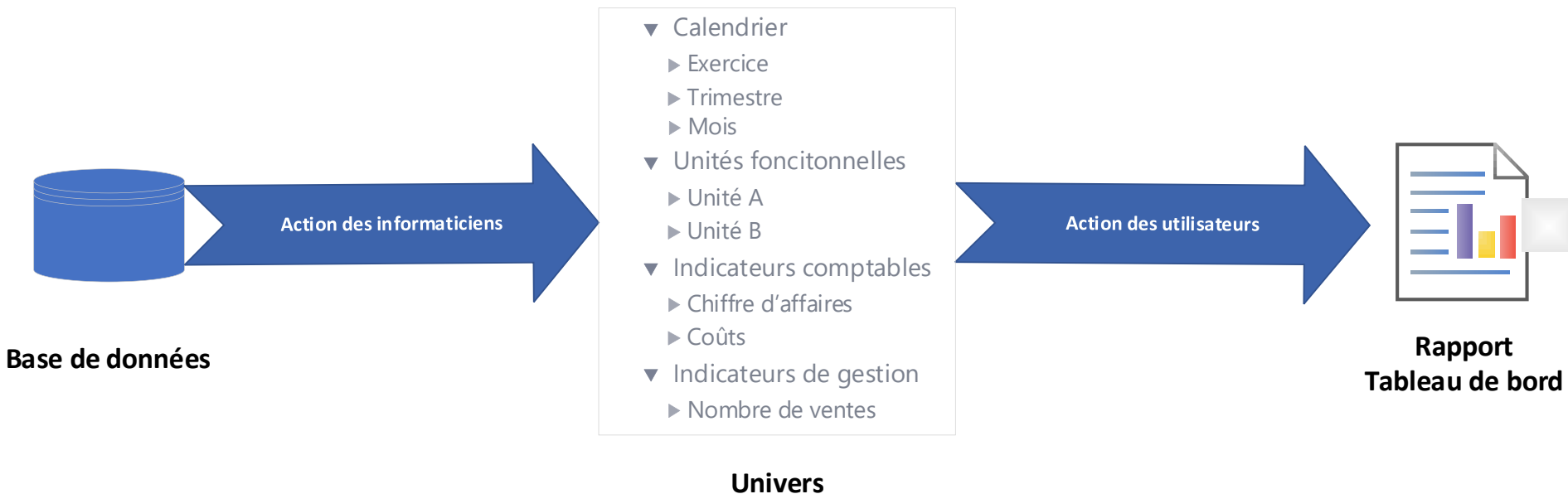
- Catégories complémentaires

- **Contrôle de flux** : Branchements conditionnels ou filtrages de lignes
- **Exécution de scripts** : Transformations à l'aide scripts ou de formules
- **Jointure lignes** : Association de données
- **Statistiques** : Statistiques avec regroupement

B – L'intégration des données

Les requêteurs

- 1^{ere} étape : Les informaticiens créent des univers (vues métier, incorporant indicateurs et axes d'analyse, issues de la base de données dans le respect de la séparation des fonctions)
- 2^e étape : Les utilisateurs créent des requêtes sur les univers et créent des rapports et tableaux de bord personnalisés.



B – L'intégration des données

Génération de rapports (Mise en forme de données pour affichage / impression)

- Etapes de création d'un rapport
 - Etablir la connexion avec la base de données
 - Créer un ou plusieurs jeux de données (requêtes) à mettre en forme
 - Elaborer graphiquement le rapport
 - Publier le rapport (web, PDF, traitement de texte, tableur)

B – L'intégration des données

Une généricité des rapports sur la base de sections

- **Entête de rapport** : section s'affichant une fois en début de rapport
- **Pied de rapport** : section s'affichant une fois en fin de rapport, utilisable pour des statistiques générales
- **Entête de page** : section s'intercalant au sommet de chaque page
- **Pied de page** : section s'intercalant au bas de chaque page
- **Entête de colonne** : section s'intercalant au sommet de chaque page
- **Pied de colonne** : section s'intercalant au bas de chaque page
- **Entête de groupe** : section répétée pour chaque nouvelle valeur d'un groupe
- **Pied de groupe** : section répétée en fin de listage d'un groupe, utilisable pour des statistiques locales au groupe.
- **Détail** : section répétée pour chaque ligne du jeu de données.

Contenu des sections

- **Éléments statiques** : Libellés, images, formes géométriques
- **Champs** : données issues de la base de données ou formules à partir des données ou éléments automatiques (numéro de page,...)
- **Graphiques** : représentation d'un jeu de données sous forme graphique
- **Tableaux croisés** : Tableau de synthèse à plusieurs dimensions analytiques
- **Sous-rapport** : Autre rapport inclus dans le flux de sortie

C – Les ERP/PGI

- 1 – Définition et prolongements
- 2 – Avantages / inconvénients
- 3 – Structure

1 – Définition et prolongements

L'ERP/PGI constitue le socle du système d'information des entreprises et organisations qui y ont recours. Il a pour finalité une gestion interactive de tous ses métiers au travers de la centralisation, du partage, de l'analyse et de la consolidation de l'information stratégique.

1 – Définition et prolongements

ERP =

- **Entreprise Resource Planning**
- **Progiciel de Gestion Intégrée (PGI)**

Enterprise

Les ERP sont destinés à gérer tous les processus d'une entreprise :

- Multi métiers
- Multi plateformes
- Multi sites, multi entreprises dans le cadre d'un groupe
- Multi exercices (mémoire sur plusieurs années)
- Multi devises, multi langues, multi réglementations dans le cadre d'un groupe international

⇒ Outil complexe pour appréhender la complexité de l'entreprise

Resources Planning

- Le terme "*ERP*" provient du nom de la méthode *MRP* (*Material Requirement Planning* => *Manufacturing Resource Planning*) utilisée depuis les années 70 pour la gestion et la planification de la production industrielle
- L'objectif était de rationaliser la production par l'intermédiaire de 3 éléments :
 - Un système informatique
 - Une méthode de gestion et de contrôle de la production en temps réel
 - Un système de minimisation des stocks et d'inventaire en temps réel

Resources Planning

Désormais, les ERP s'affranchissent du cadre restrictif de la production et retiennent le caractère **Temps Réel** :

- l'impact d'une opération dans l'entreprise doit être répercuté sans ressaisie et en temps réel sur toutes les activités qui peuvent en dépendre :
- Les ERP s'appuie sur un moteur de *Workflow* autorisant le transit d'informations en temps réel au travers de procédures automatisées

Progiciel

- Un **progiciel** est un logiciel commercial vendu par un éditeur sous forme d'un produit complet, plus ou moins *clés en main*. Le terme résulte de la contraction des mots *produit* et *logiciel*
- Le recours aux progiciels s'oppose aux développements internes
- Toutefois, dans le cadre de l'ERP, il y a des travaux nécessaires :
 - Amélioration de la couverture fonctionnelle
 - Adaptation à l'entreprise

=> les ERP sont donc des progiciels de série, pour lesquels un certain nombre de développements sur mesure sont nécessaires.

Pour favoriser l'adaptation d'un progiciel, une entreprise peut faire appel à un partenaire plus ou moins indépendant du fournisseur, le consultant en intégration de systèmes.

Gestion intégrée

Unification des fonctions

La gestion de l'entreprise présente plusieurs aspects :

- Production,
- Logistique et gestion des chaînes d'approvisionnement,
- Gestion des ressources humaines,
- Relation client, automatisation de la force de vente,
- Finance, facturation, trésorerie
- ...

Initialement ces métiers sont pris en charge par des logiciels horizontaux (génériques : traitement de texte, tableurs, SGBD, Messagerie) ou verticaux (application Paye, application Compta, application de gestion des stocks...) peu communicants que les ERP visent à remplacer.

Gestion intégrée

Le PGI va donc remplacer les applications métiers, peu communicantes entre elles, par une solution fondée sur une **base de données unique**

Un outil de gestion intégrée vise à présenter des modules (*applications métiers interagissant exclusivement avec la base de données centrale et contenant les automatismes liés à la répercussion des événements sur les autres éléments du système*) propres à chacun des aspect (initialement les premiers ERP étaient monolithiques mais ils suivent désormais une approche modulaire).

Toutefois la notion d'ERP apparaît chez les fournisseurs de solutions dès qu'au moins 2 aspects de la gestion sont couverts, bien que les leaders du marché couvrent un champ très large des activités de l'entreprise.

Gestion intégrée

Unification des outils :

L'ERP unifie les outils de :

- Back-office : outils invisibles des clients et partenaires de l'entreprise
- Front-office : outils visibles des clients (portail d'entreprise)

Le PGI va également unifier les différents éléments du système d'information de l'entreprise :

- Système opérationnel
- Système de pilotage
- Système décisionnel et stratégique

Unification des accès au système :

- Mode client / Serveur multi-plateforme : un seul outil d'accès présentant une interface spécifique au profil de l'utilisateur
- Mode web / intranet / extranet

Rappel des points clés :

- Outils complexes qui gèrent **toute** l'entreprise, dans la multiplicité de ses aspects
- Se présentent sous forme **modulaire** autour d'une **base de données unique**
- Répercutent en **temps réel** les informations saisies pour toutes les activités de l'entreprises qu'elles concernent, permettent le traitement **opérationnel**, mais également les **agrégations** d'informations pour aider au **pilotage** ou à la **stratégie** d'entreprise.

=>S'ils contribuent à résoudre une masse d'enjeux pour l'entreprise, leur installation est en soi un enjeu majeur.

2 – Avantages / Inconvénient

- Avantages des ERP :
Les ERP : Pourquoi ?
- Inconvénients de ERP

Pourquoi de l'ERP :

- Réduction des coûts, délais, et pertes d'intégrité ou de disponibilité des données dues au manque d'intégration et appui de l'entreprise sur un système unique et cohérent autorisant la traçabilité des opérations.
- Prise en charge des opérations à l'échelle Mondiale pour les grands groupes, économies d'échelle dans le cadre d'une centralisation (ex : fonction achat) pour toutes les entreprises
- Couverture totale de la chaîne de valeur (séquence de tous les processus d'une organisation qui ajoutent de la valeurs à un produit ou à un service offert).
- Mise à disposition de fonctions de pilotage et d'analyse décisionnelle
- Organisation globale les flux d'information internes et externes et disponibilité des données utiles à des milliers d'utilisateurs différents (internes ou externes)
- La réflexion autour de l'ERP permet de se caler sur les Meilleures Pratiques: refonte des processus de production et des processus administratifs visant à atteindre les plus hauts niveaux de qualité.
- Amélioration de la réactivité grâce à des données centralisées (face aux clients, entreprises, réglementations)
- Réduction des coûts en personnel (développeurs par rapport à une solution interne), opérationnels dans le cadre de la réduction des tâches manuelles et de duplication (ressaisie) de données

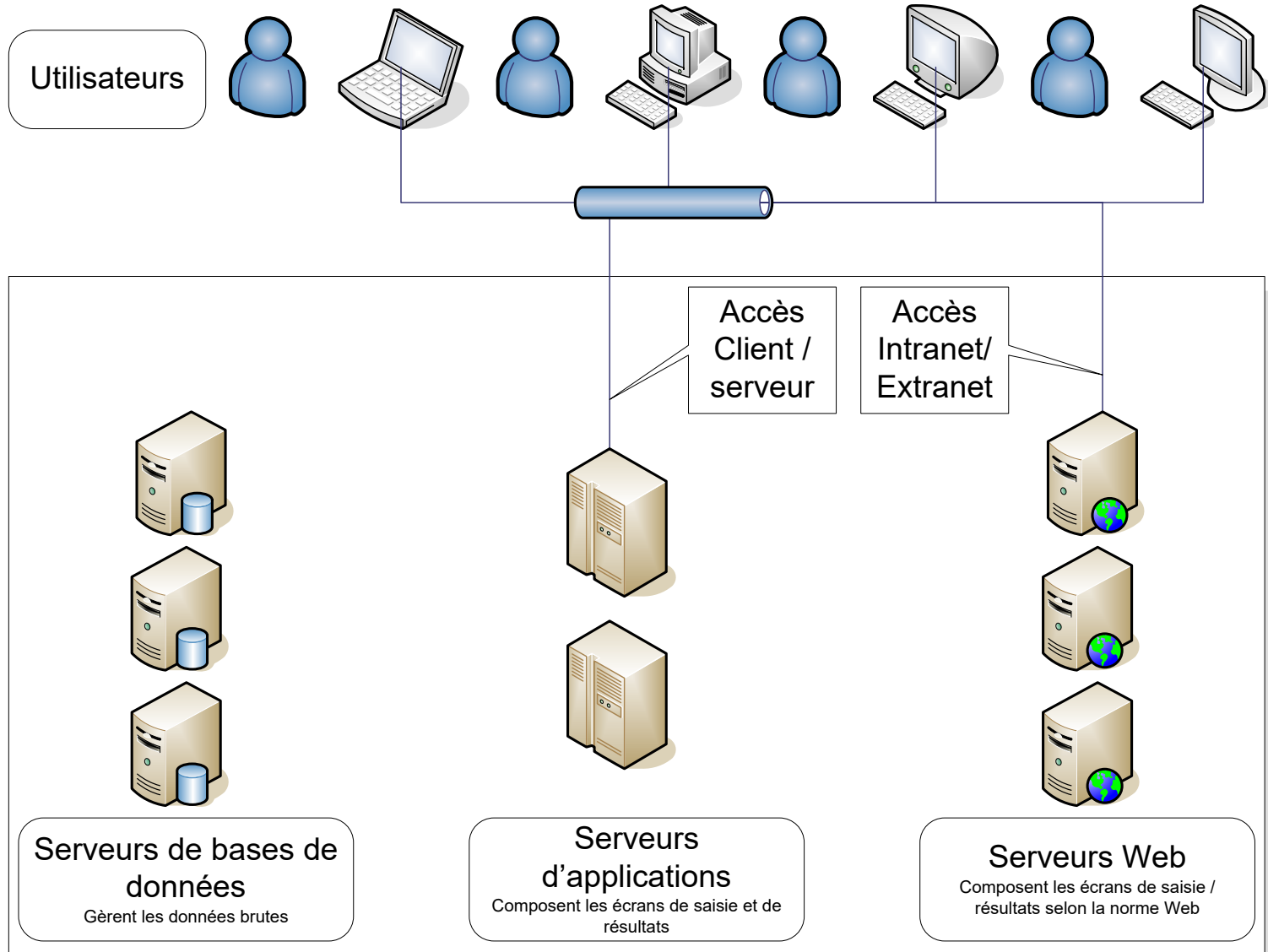
Inconvénients de L'ERP

- Lourdeur d'installation et d'utilisation
- Adaptation contraignante et parfois limitée
- Standardisation des processus pouvant faire perdre en compétitivité
- Goulets d'étranglements au niveau des fonctions insuffisamment couvertes
- Coûts très élevés
- Nécessite un accompagnement organisationnel du changement très sérieux

3 – Structure des ERP

- Structure matérielle
- Structure logicielle

Structure matérielle



Structure logicielle

