CONNAISSANCE DES SYSTÈMES INFORMATIQUES ET DES RÉSEAUX

E. DUBOIS – 2024

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

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
4. Apparition de l'informatique quantique

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
4. Le chiffrement 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

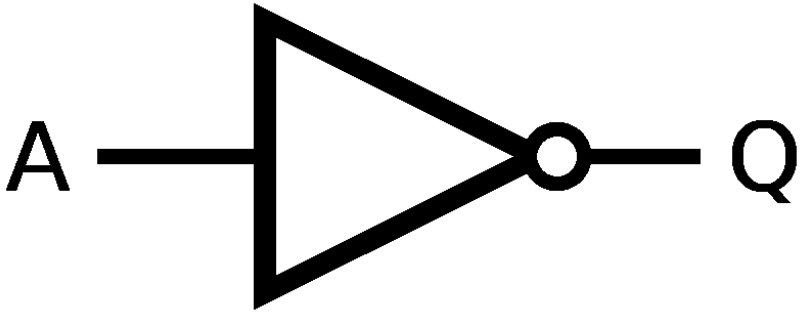
- Peuvent comporter 1 entrée, 2 entrées ou plus mais se caractérisent toujours par une seule sortie.

Simulation sur https://logic.ly/demo

Porte **NOT**

Simulation sur https://logic.ly/demo

- Réalise l'inversion d'une valeur binaire

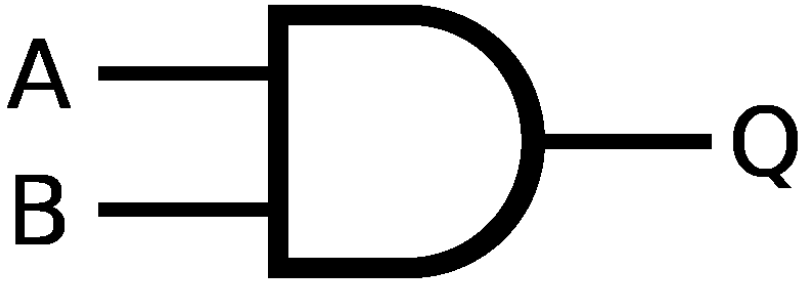


A
0
1

Q
1
0

Porte **AND**

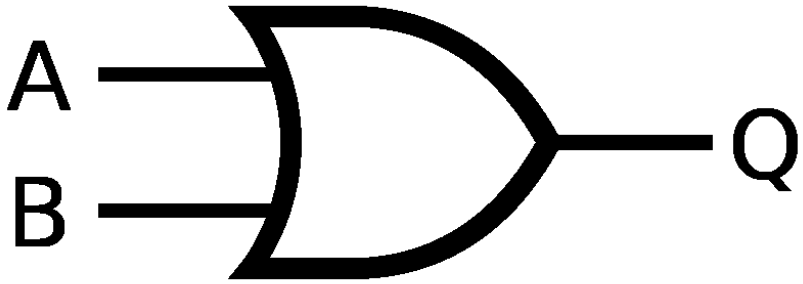
- Réalise une opération ET logique entre deux valeurs binaires



A	B	Q
0	0	0
0	1	0
1	0	0
1	1	1

Porte **OR**

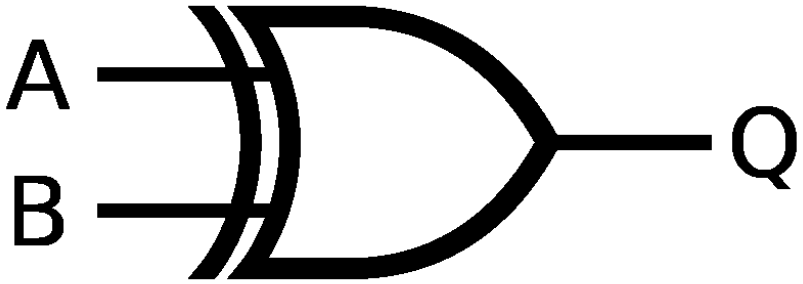
- Réalise une opération OU entre 2 valeurs binaires



A	B	Q
0	0	0
0	1	1
1	0	1
1	1	1

Porte **XOR**

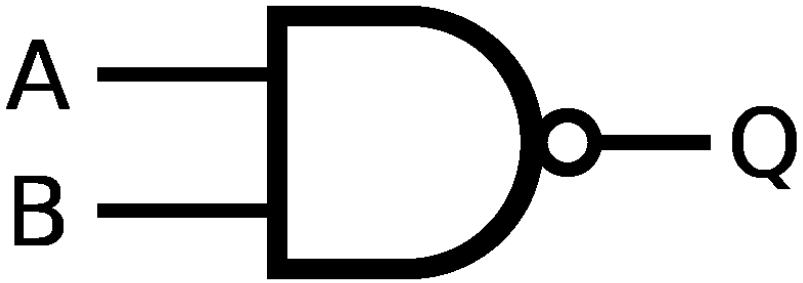
- Réalise une opération logique OU EXCLUSIF entre 2 entrées



A	B	Q
0	0	0
0	1	1
1	0	1
1	1	0

Porte **NAND**

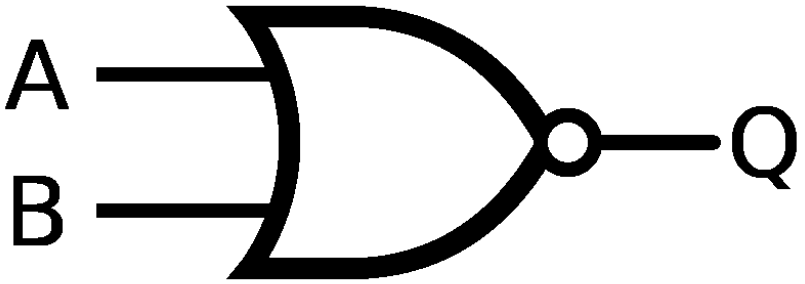
- NOT AND Réalise un ET logique entre 2 valeurs suivi d'une inversion du résultat



A	B	Q
0	0	1
0	1	1
1	0	1
1	1	0

Porte **NOR**

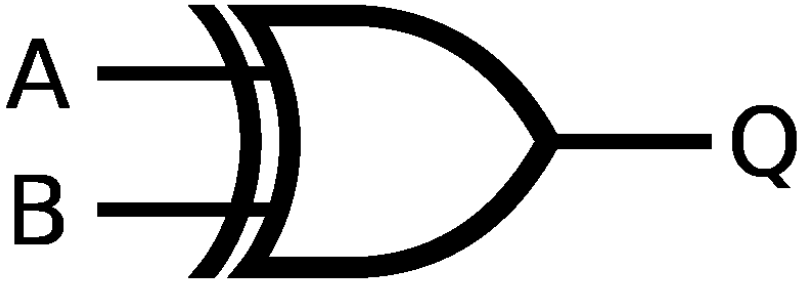
- NOT OR : Réalise un OU logique entre 2 valeurs suivi d'une inversion du résultat



A	B	Q
0	0	1
0	1	0
1	0	0
1	1	0

Porte **XNOR**

- Réalise une opération OU EXCLUSIF puis inverse le résultat

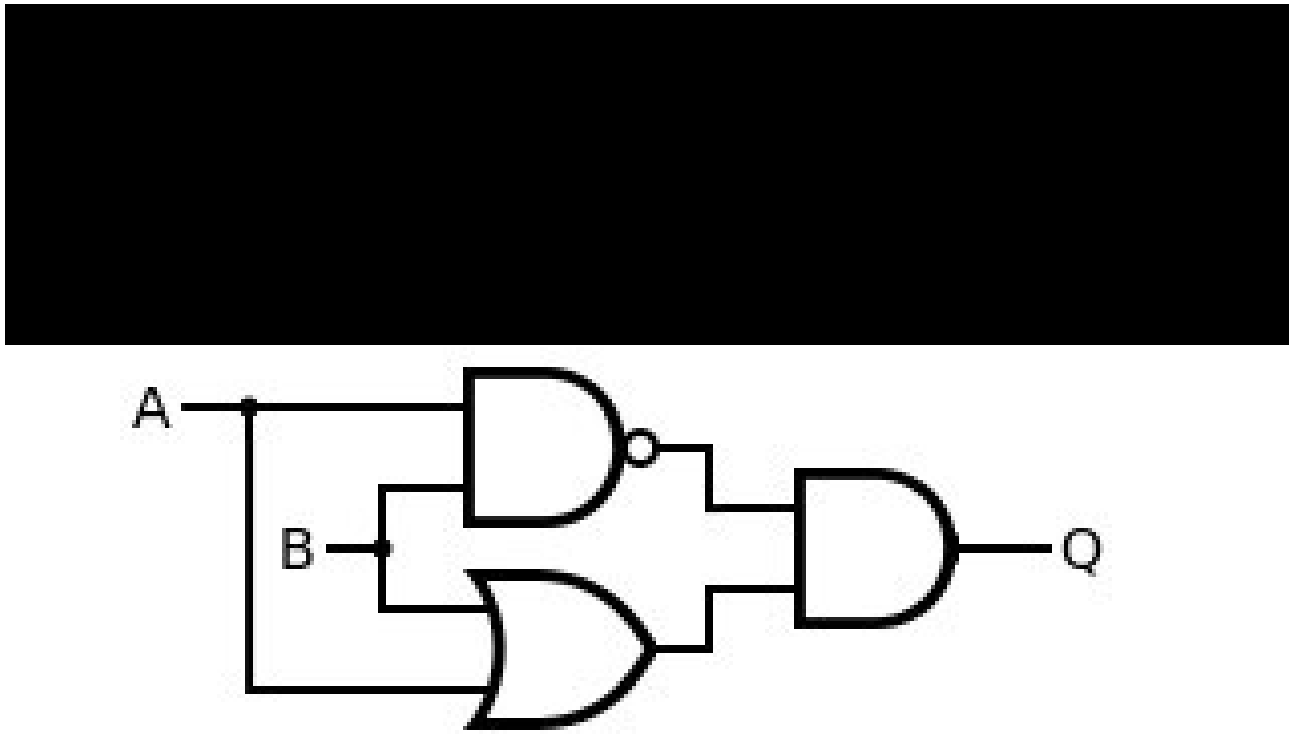


A	B
0	0
0	1
1	0
1	1

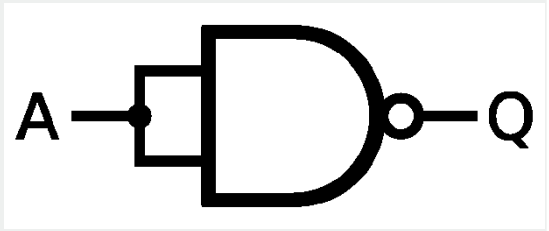
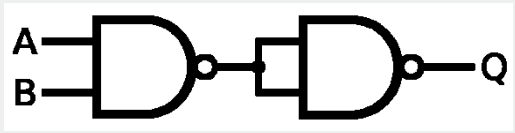
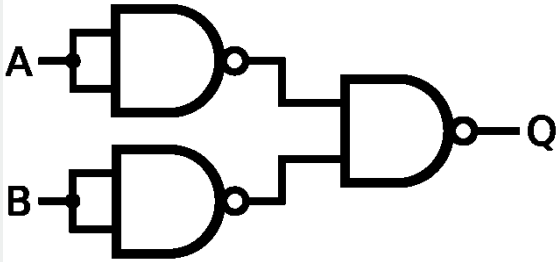
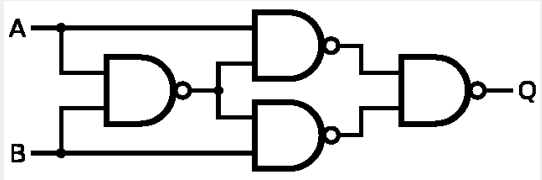
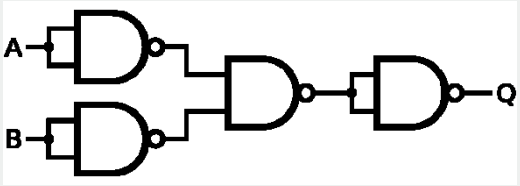
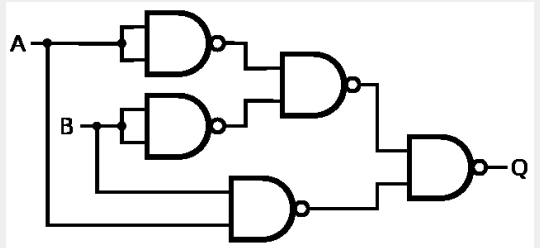
Q
1
0
0
1

Les portes AND NOT OR forment un ensemble logique complet car peuvent composer toutes les autres portes

- NAND
- XOR

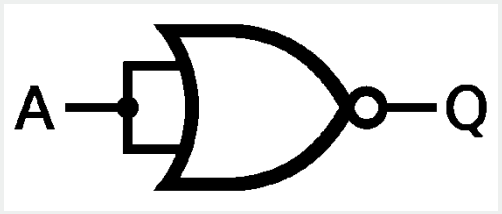
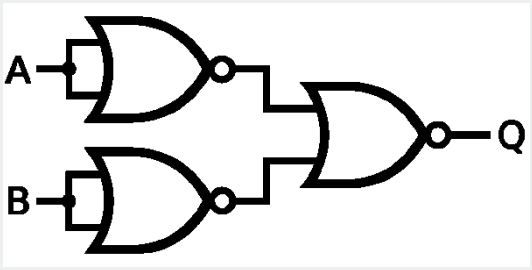
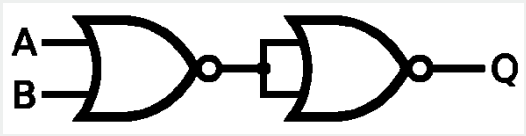
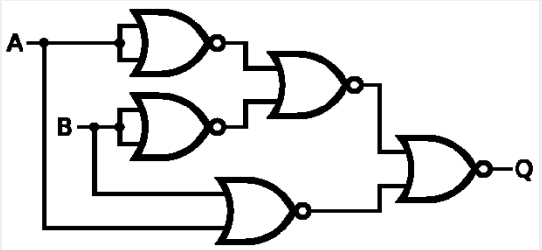
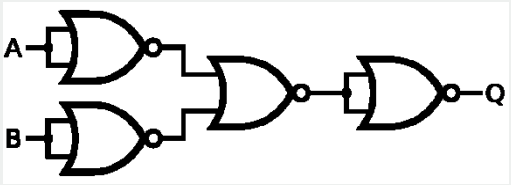
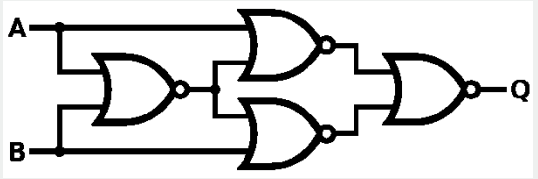


NAND : une porte logique universelle

NOT	AND	OR
		
XOR	**NOR**	**XNOR**
		

Source : https://en.wikibooks.org/wiki/Digital_Circuits/NAND_Logic

NOR : une porte logique universelle

NOT	AND	OR
		
XOR	**NAND**	**XNOR**
		

Source : https://en.wikibooks.org/wiki/Digital_Circuits/NOR_Logic

Comment sont réalisés les circuits ?

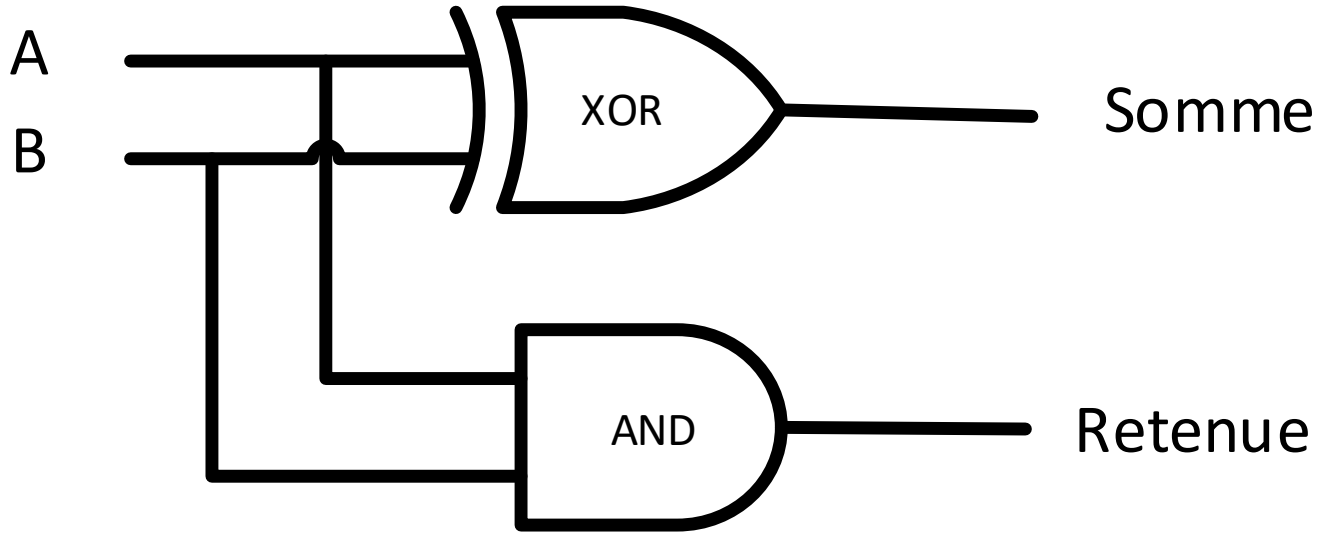
- Circuits non reconfigurables
  - Par association ou assemblage de portes :
    - AND OR NOT
    - NAND
    - NOR
- Circuits reconfigurables: ex FPGA (field programmable gate array)
  - Par utilisation de circuits logiques présentés sous forme de matrice où sont configurables le comportement des portes logiques émulées et leurs connexions

Exemple réaliser l'addition de 2 bits valant 1 ou 0

- Logique :

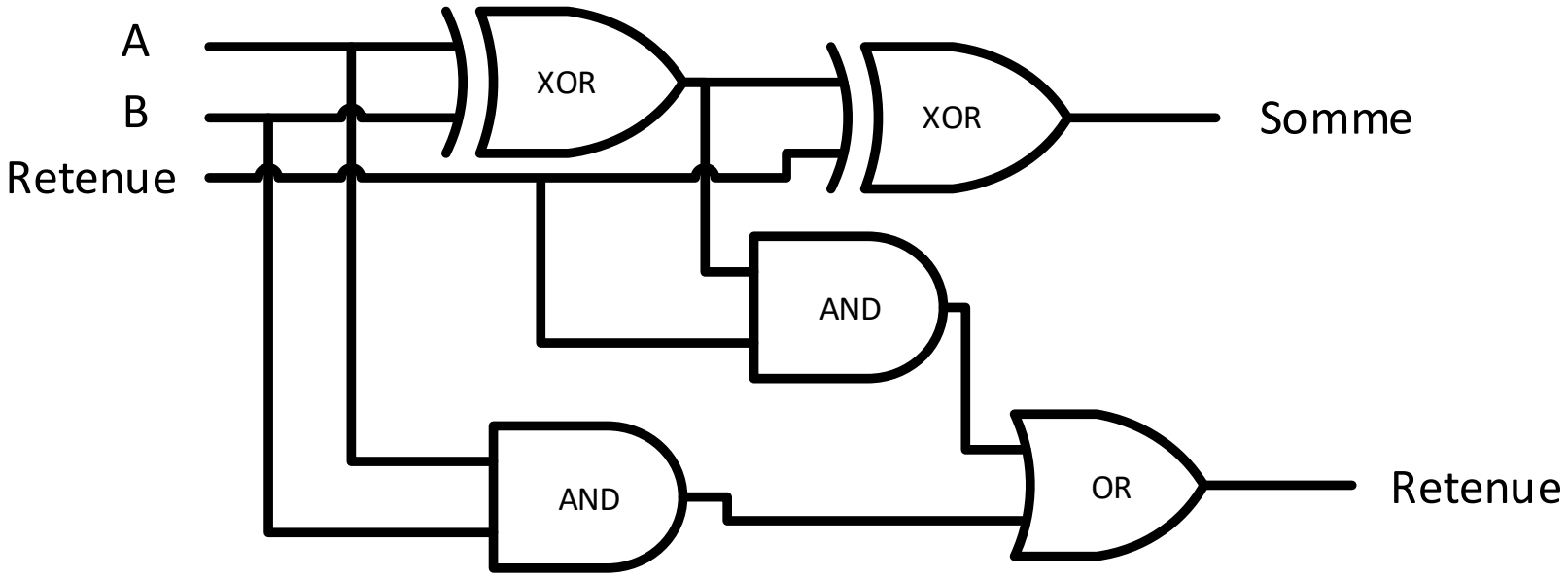
```
0 +  0 =  0
0 +  1 =  1
1 +  0 =  1
1 +  1 =  1  0     soit 0 avec une retenue de 1 pour la position suivante (à gauche)
```

- Semi additionneur (ne prend pas de retenue en entrée)



Exemple réaliser l'addition de 2 bits valant 1 ou 0

- Additionneur complet (intègre également une retenue en entrée)



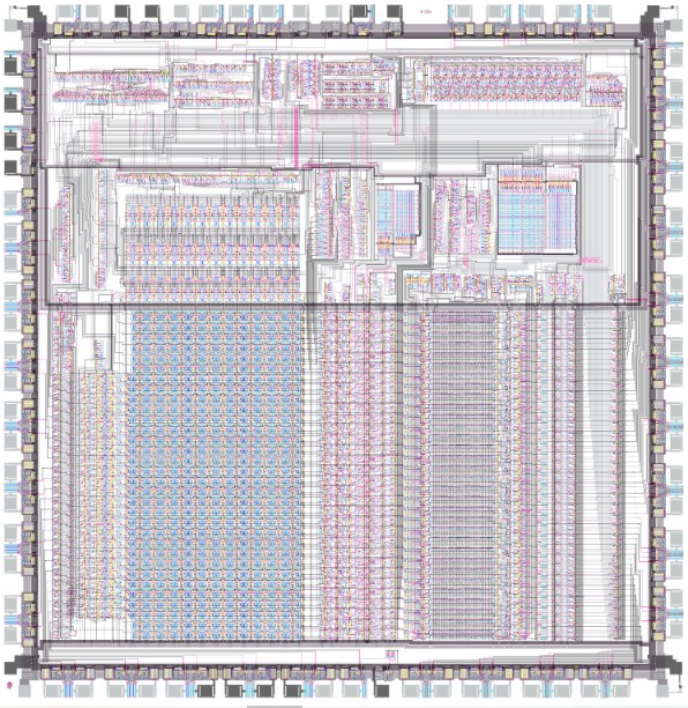
Les différents types de circuits

- DSP (Digital signal processor) : circuits à fonctionnement déterminé à l'avance.
- Le processeur (Central Processing Unit) est un circuit extrêmement intégré qui constitue le composant essentiel d'un ordinateur. A la différence d'un autre circuit intégré qui a un comportement déterminé, le processeur a la capacité de fonder son comportement sur l'exécution d'un programme chargé depuis la mémoire

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

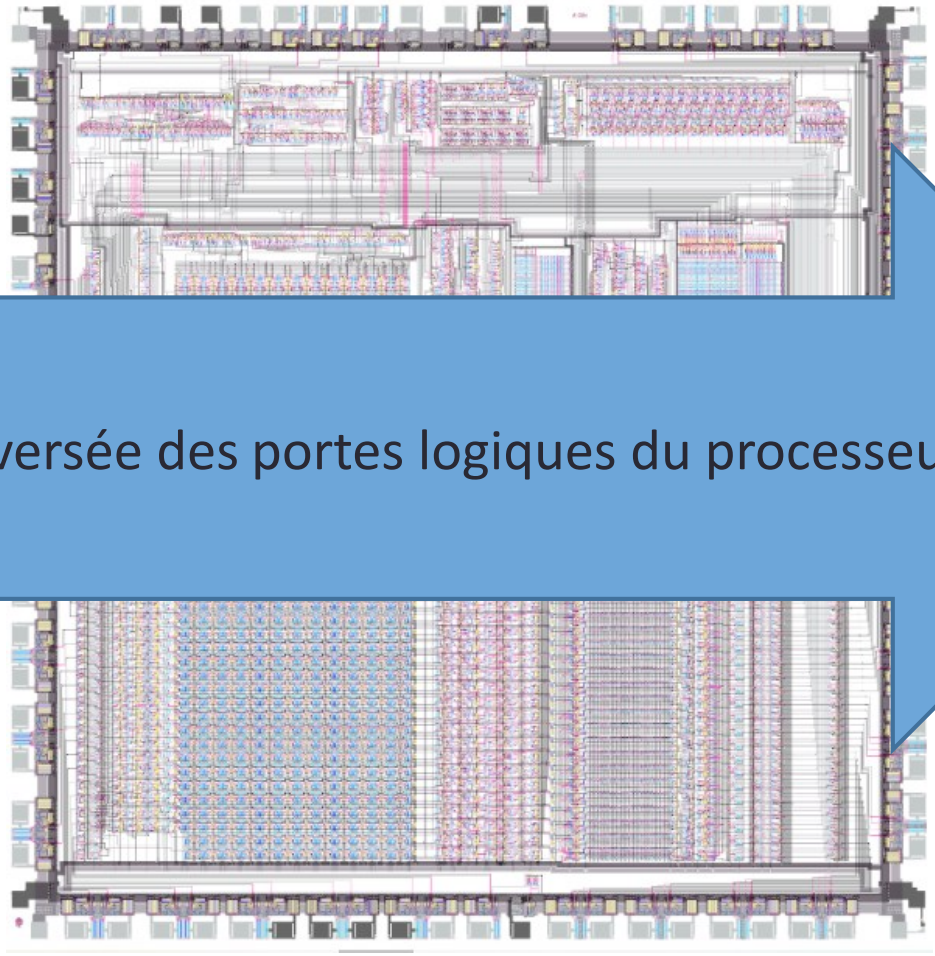


Commande

0
1
0
0
1
0
0
1

Opérande

0
1
1
1
0
1
0
0



Résultat dans un registre

0
1
0
0
1
0
0
1

Etat

0
1
0
0
1
0
0
1

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> Donne à un Qbit une probabilité de 1 de valoir 0
    - X : porte X de Pauly qui inverse le Qbit en opérande
    - + : Porte de NON 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 entrainant la nécessité de recréer les librairies 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

4. Le chiffrement des données

Les grandeurs quantifiant les données

Bit : Acronyme de BInary 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 *kilo* octet binaire (kibioctet)	**Kio**	=1024 octets	2^{10} octets	1 024 octets
1 *méga* octet binaire (mébioctet)	**Moi**	=1024 Kio	2^{20} octets	1 048 576 octets
1 *giga* octet binaire (Gibioctet)	**Gio**	=1024 Moi	2^{30} octets	1 073 741 824 octets
1 *tera* octet binaire (Tébioctet)	**Tio**	=1024 Gio	2^{40} octets	1 099 511 627 776 octets
1 *péta* octet binaire (Pébioctet)	**Pio**	=1024 Tio	2^{50} octets	1 125 899 906 842 624 octets
1 *exa* octet binaire (Exbioctet)	**Eio**	=1024 Pio	2^{60} octets	1 152 921 504 606 846 976 octets
1 *zetta* octet binaire (Zébioctet)	**Zio**	=1024 Eio	2^{70} octets	1 180 591 620 717 411 303 424 octets
1 *yotta* octet binaire (Yobioctet)	**Yio**	=1024 Zio	2^{80} octets	1 208 925 819 614 629 174 706 176 octets

Unités usuelles

1 *kilo* octet	**Ko**	=1000 octets	10^{3} octets	1 000 octets
1 *méga* octet	**Mo**	=1000 Ko	10^{6} octets	1 000 000 octets
1 *giga* octet	**Go**	=1000 Mo	10^{9} octets	1 000 000 000 octets
1 *tera* octet	**To**	=1000 Go	10^{12} octets	1 000 000 000 000 octets
1 *péta* octet	**Po**	=1000 To	10^{15} octets	1 000 000 000 000 000 octets
1 *exa* octet	**Eo**	=1000 Po	10^{18} octets	1 000 000 000 000 000 000 octets
1 *zetta* octet	**Zo**	=1000 Eo	10^{21} octets	1 000 000 000 000 000 000 000 octets
1 *yotta* 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 **BI**nary 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

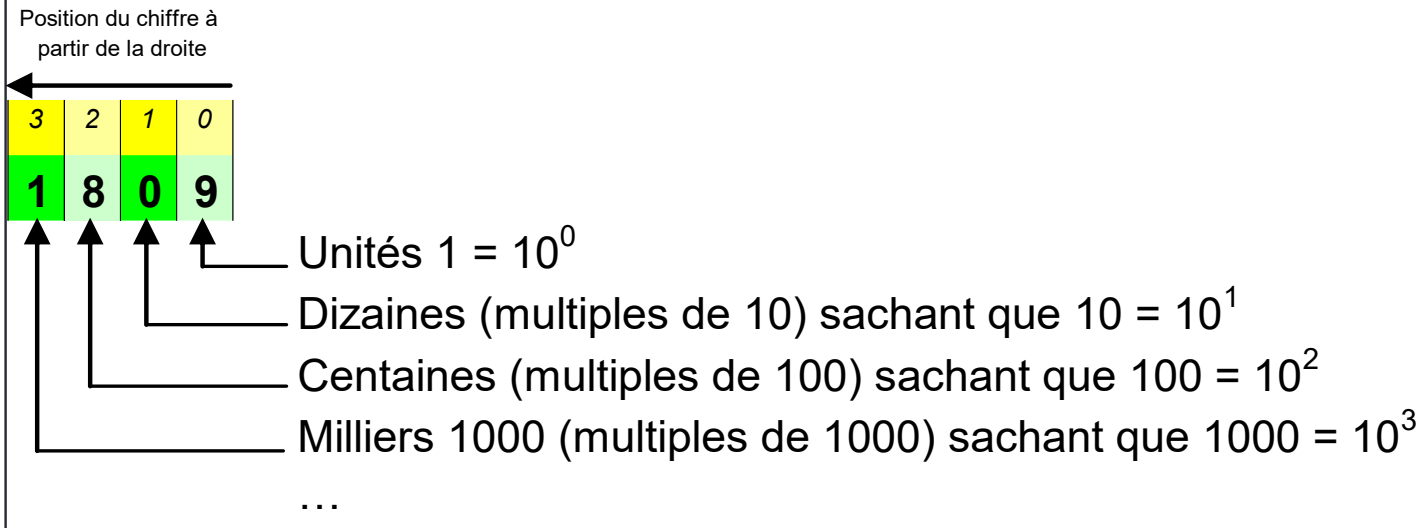
```
    9
+   1
= 1 0
```

- 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)

```
    1
+   1
= 1 0
```

a - Représentation binaire des nombres entiers positifs

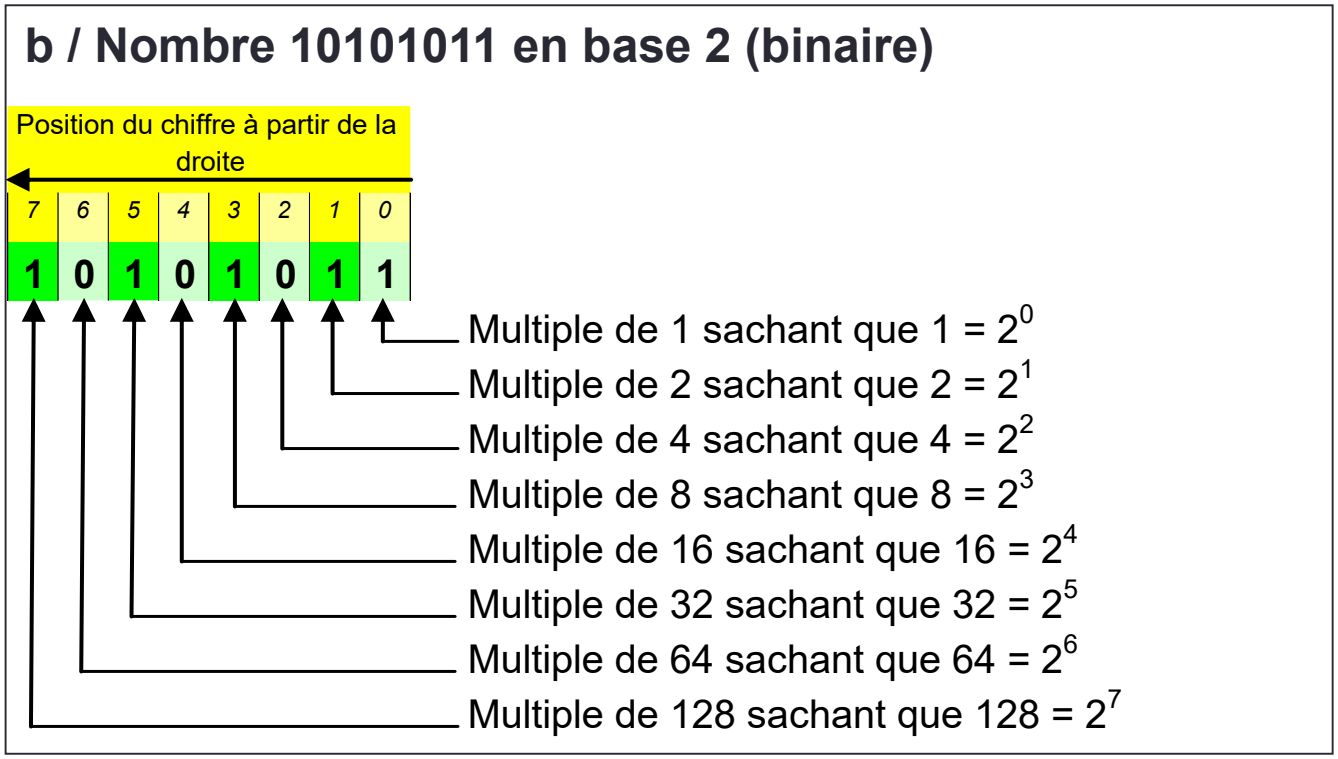
a / Nombre 1809 en base 10



Donc ce nombre est égal à

$1\times1000+8\times100+0\times10+9\times1=1809$

a - Représentation binaire des nombres entiers positifs



Donc ce nombre est égal à 1×128+0×64+1×32+0×16+1×8+0×4+1×2+1×1=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

2^0	1
2^1	2
2^2	4
2^3	8
2^4	16
2^5	32
2^6	64
2^7	128
2^8	256
2^9	512
2^{10}	1 024
2^{11}	2 048
2^{12}	4 096
2^{13}	8 192
2^{14}	16 384
2^{15}	32 768
2^{16}	65 536

1

40 000

Puissance de 2		Valeur		Résultat
32 768	≤	40 000	?	Oui → 1
16 384	≤	7 232	?	Non → 0
8 192	≤	7 232	?	Non → 0
4 096	≤	7 232	?	Oui → 1
2 048	≤	3 136	?	Oui → 1
1 024	≤	1 088	?	Oui → 1
512	≤	64	?	Non → 0
256	≤	64	?	Non → 0
128	≤	64	?	Non → 0
64	≤	64	?	Oui → 1
32	≤	0	?	Non → 0
16	≤	0	?	Non → 0
8	≤	0	?	Non → 0
4	≤	0	?	Non → 0
2	≤	0	?	Non → 0
1	≤	0	?	Non → 0

2

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.

5 / 2 = 2

avec un reste = 5 - 2 × 2

= 1

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 soit 0 avec une retenue de 1 pour la position suivante (à gauche)

			+1	*+1*	*+1*			
		0	1	0	1	0	1	0
+		1	0	0	1	1	1	1
=	0	1	1	1	1	0	0	1

b - Les opérations en système binaire

Soustraction

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

0	-	0	=	0
0	-	1	=	1
1	-	0	=	1
1	-	1	=	0

soit 1 avec une retenue de -1 pour la position suivante (à gauche)

	-1		*-1*	*-1*	*-1*	*-1*	*-1*	
	1	0	1	0	1	0	1	0
-		1	0	0	1	1	1	1
=	0	1	0	1	1	0	1	1

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

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

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

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

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

	1	1	1	1	0	0	0	0
XOR	1	0	1	0	1	0	1	0
=	0	1	0	1	1	0	1	0

	0	1	0	1	1	0	1	0
XOR	1	0	1	0	1	0	1	0
=	1	1	1	1	0	0	0	0

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

b - Les opérations en système binaire

Le complément à 2

Inversion des bits de l'opérande puis ajout de 1 en ignorant les dépassements (une éventuelle retenue au-delà de la position la plus à gauche est ignorée)

		1	0	1	0	1	0	1	0
Complément à 1		0	1	0	1	0	1	0	1
								+1	
	+								1
	=	0	0	1	0	1	0	1	1	0

ignoré

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	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 échantillons / sec × 60 secondes × 16 bits × 2 canaux =

84 672 000 bits ≈ 10 Mo

Une minute de son stéréo en haute qualité=

192 000 échantillons / sec × 60 secondes × 24 bits × 2 canaux =

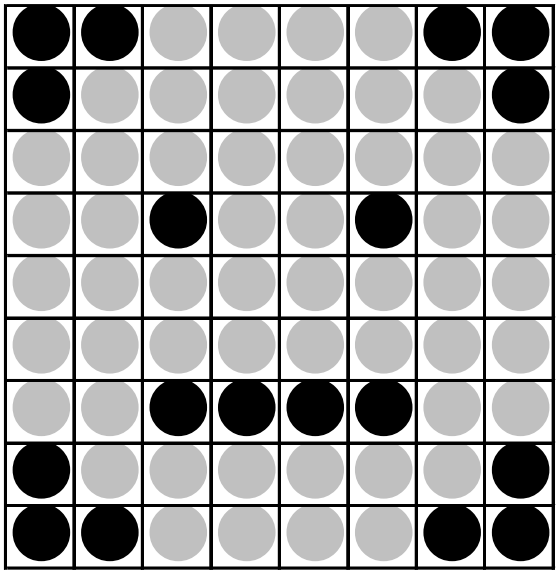
552 960 000 bits ≈ 69 Mo

Les déterminants de l'image numérique et volumétrie des données

Les déterminants de l'image numérique et volumétrie des données

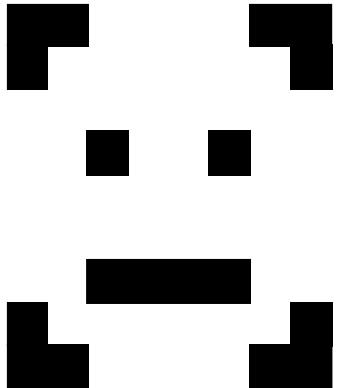
- La **résolution horizontale**, mesurée en nombre de points par pouce
- La **résolution verticale**, mesurée en nombre de points par pouce
- La **profondeur de couleurs**, mesurée en nombre de bits par points
  - 1 bit par point : noir et blanc
  - 8 bits par point : 256 niveaux de gris ou 256 couleurs 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



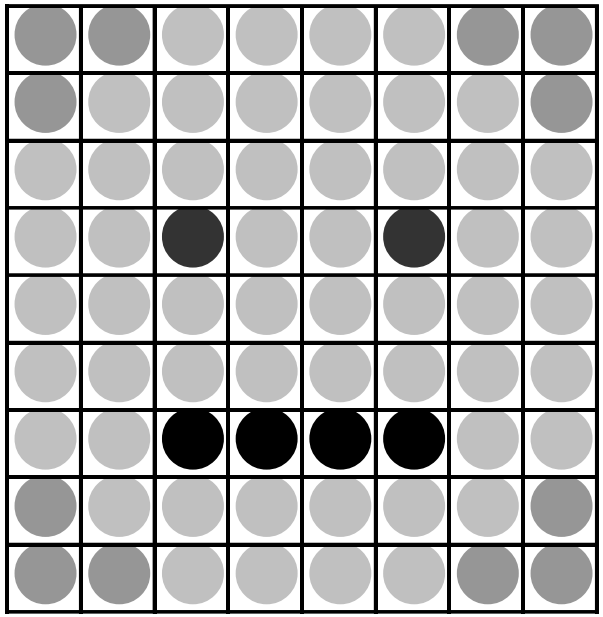
1	1	0	0	0	0	1	1
1	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0
1	0	0	0	0	0	0	1
1	1	0	0	0	0	1	1

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 :

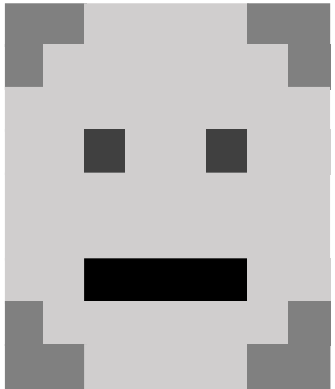
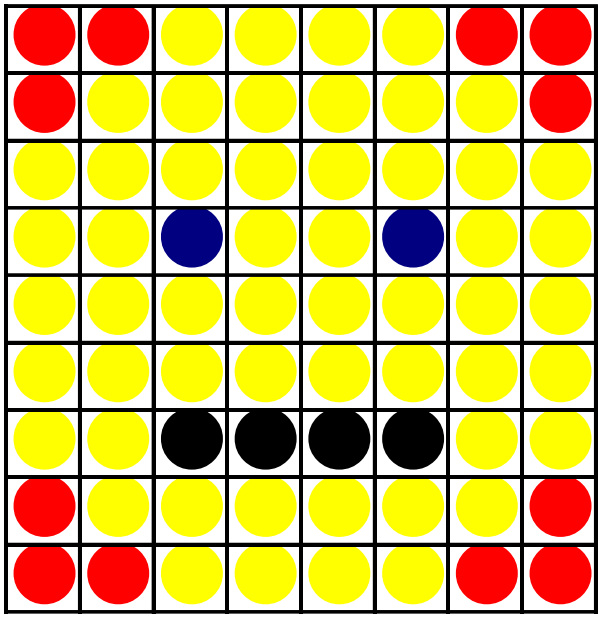
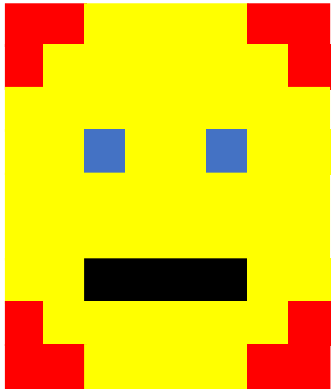


Image en couleurs



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	255	255	255	255	255	255	0
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

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)

	3 840	Points en largeur
×	2 160	Points en hauteur
×	3	Octets par points
×	25	Images par seconde
×	60	secondes
≈	8,3	Go de données

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.

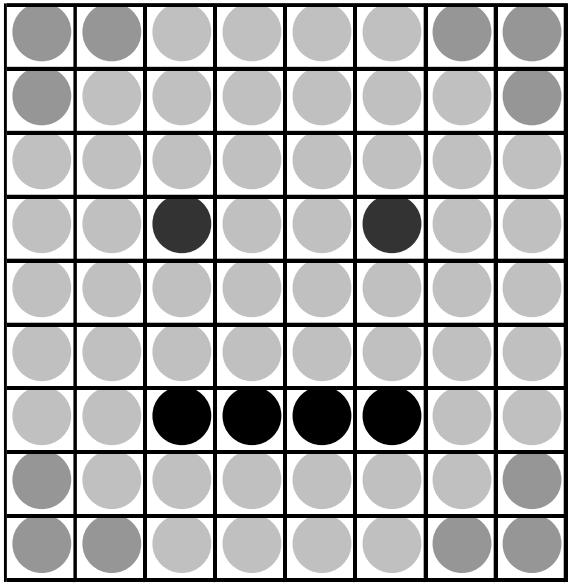
C - La compression sans perte

- 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, Liv 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

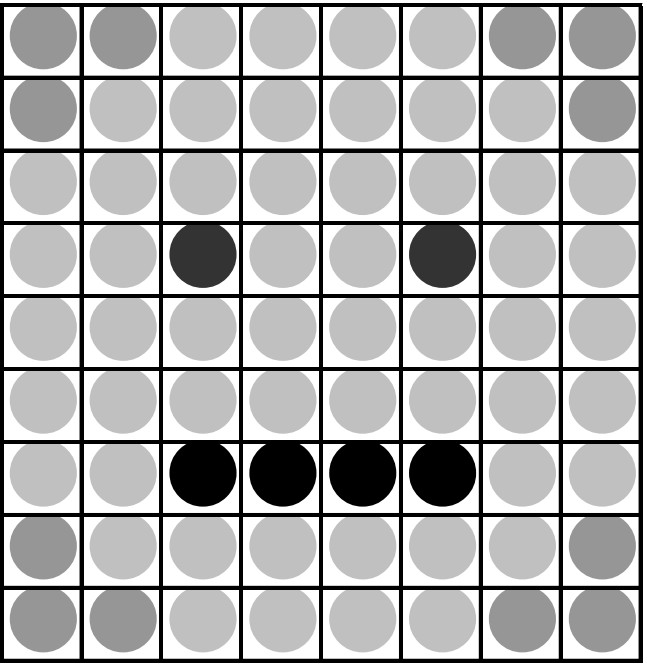
	Marqueur de répétition
	Nombre de répétitions
	Valeur à répéter

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



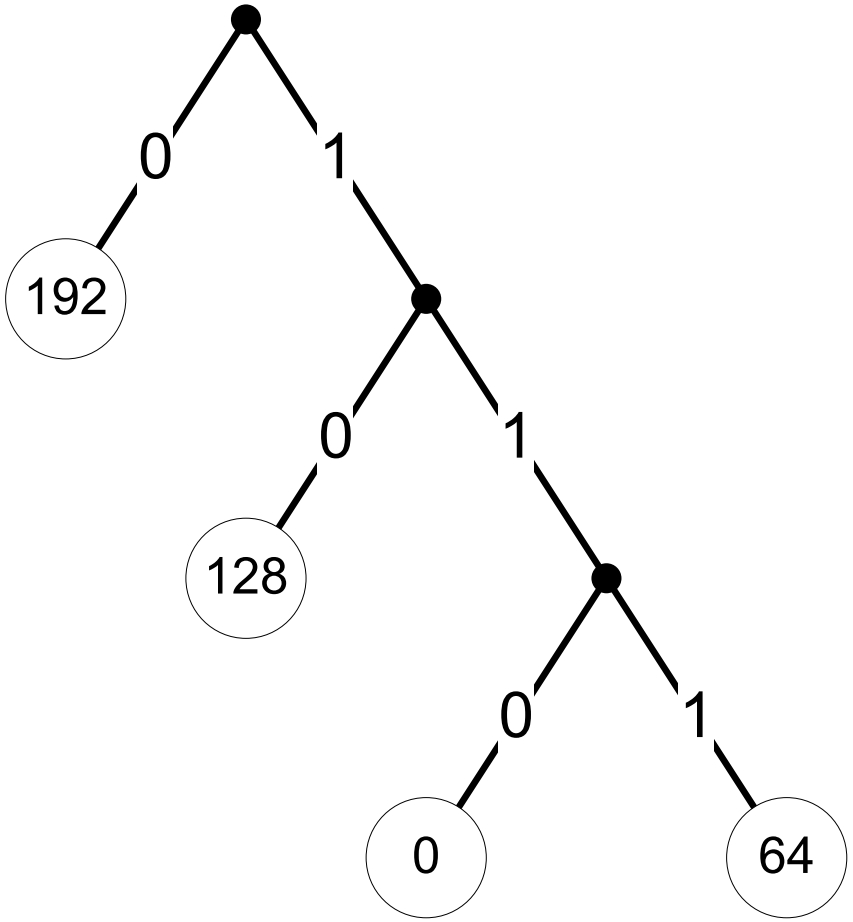
128	128	192	192	192	192	128	128
128	192	192	192	192	192	192	128
192	192	192	192	192	192	192	192
192	192	64	192	192	64	192	192
192	192	192	192	192	192	192	192
192	192	192	192	192	192	192	192
192	192	0	0	0	0	192	192
128	192	192	192	192	192	192	128
128	128	192	192	192	192	128	128

Valeur	Occurrences
192	46
128	12
0	4
64	2

La compression par dictionnaire

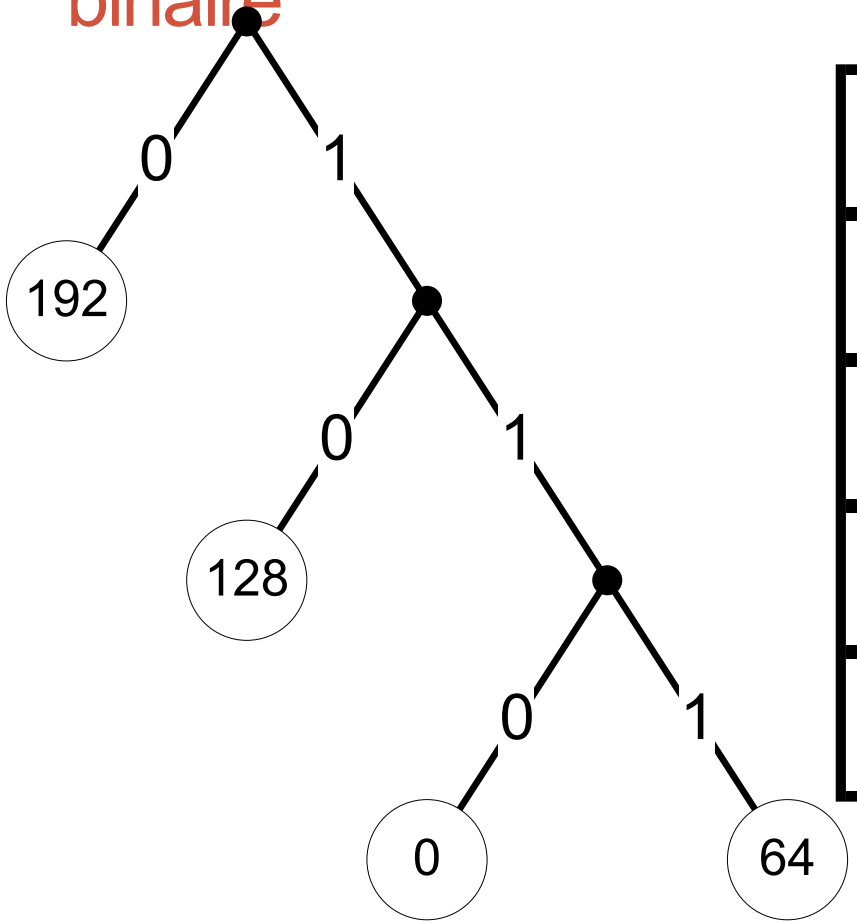
Etape 2 – Construction de l'arbre

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 :

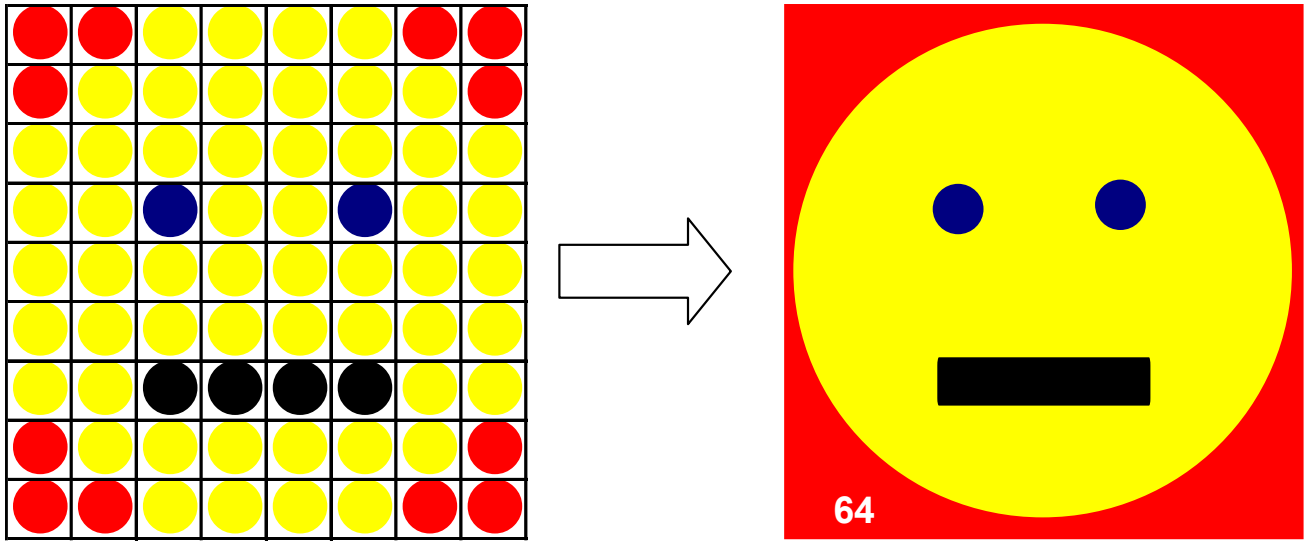
- Élimination des sons inaudibles
- Application d'un modèle psychoaccoustique de masquage des sons et des canaux
- Suppression des informations communes à plusieurs canaux
- Application d'une méthode de compression sans perte sur le résultat

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

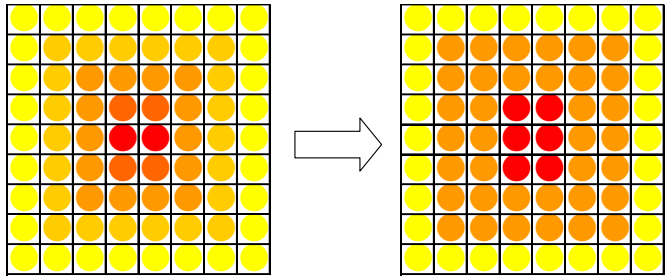
Variantes : Compression fractale, compression par ondelettes.

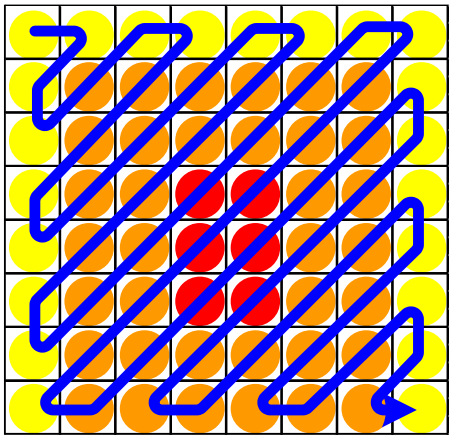


Compression avec perte d'images fixes : JPEG

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

4 - Les procédés de chiffrement

Trois usages des procédés de chiffrement

1 – Contrôle d'intégrité des messages transmis
2 – Mise en œuvre de la confidentialité des transmissions
3 – Authentification de l'émetteur d'un message

- Deux concepts dérivés :

4 – Certificat électronique
5 – Clé de session

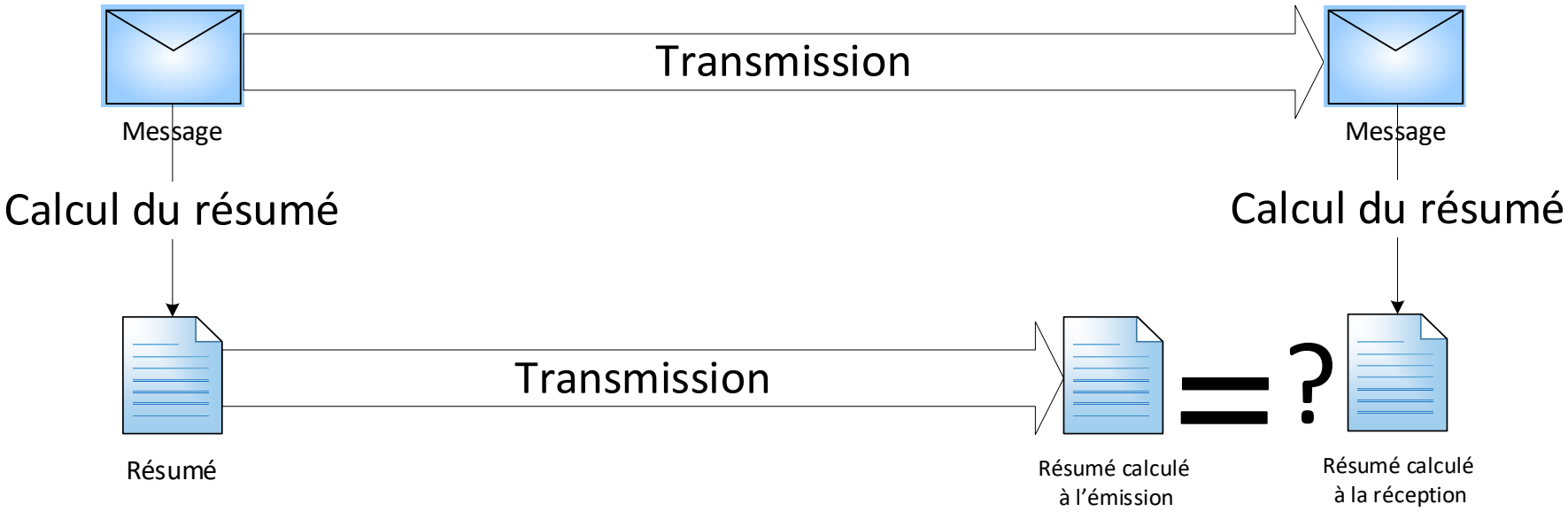
Chiffrements révolus :

- Systèmes fondés sur des méthodes secrètes
  - Code de César
  - Système Enigma
- Systèmes fondés sur des tables de correspondance
  - Sensibles à la cryptanalyse
- Leur seront préférés des méthodes publiques dans lesquelles le secret est porté par une clé. La taille de la clé, exprimée en bits, sera une mesure de la difficulté à tenter de se passer de cette dernière en tentant tous les cas possibles

Contrôle de l'intégrité d'un message

- Création d'un résumé (ou haché ou empreinte digitale)
- Utilisation d'un algorithme public tel que MD5, SHA1, SHA128, SHA512

Texte d'exemple	Résumé MD5	Résumé SHA1	Résumé SHA256	Résumé SHA512
Bonjour	ebc58ab2cb4848d04ec23d83f7ddf985	f30ecbf5b1cb85c631fdec0b39678550973cfcbc	9172e8eec99f144f72eca9a568759580edadb2cfd154857f07e657569493bc44	c447dff0d671f62ad580b255b64f7a8f6a30d1b828569cee08b7c861239f8d4856ef38a1166718b045a9713876336c1f623619f6a78fc891d48d0b98c703def3
bonjour	f02368945726d5fc2a14eb576f7276c0	1f71e0f4ac9b47cd93bf269e4017abaab9d3bd63	2cb4b1431b84ec15d35ed83bb927e27e8967d75f4bcd9cc4b25c8d879ae23e18	3041edbcdd46190c0acc504ed195f8a90129efcab173a7b9ac4646b92d04cc80005acaa3554f4b1df839eacadc2491cb623bf3aa6f9eb44f6ea8ca005821d25d
Bonjour !	ab557953e6057cbeddc3a1124ef2a587	245469f3c41b3222441056b6bc6e789710ef2bfb	bc7444869484dd1ed34bfc4465724b887fecf604fcad0d0eaf40d3746d0bfeb7	24f5990db401d3b3f1161188435aefab4f6450d6f4ffbdc14fdbab65ec7124c8d80faffdcf86615b317911d20b4c5084dadf7997a90f1fb5caba20770a541c2f


Transmission
Message
Message
Calcul du résumé
Calcul du résumé
Transmission
Résumé
Résumé calculé
à l'émission
=
?
Résumé calculé
à la réception

Mise en œuvre de la confidentialité : les systèmes symétrique ou à clé privée

- On utilise un algorithme public utilisant notamment des opérations informatiques de rotation de bits et d'application de OU exclusifs
- Exemples d'algorithmes : Triple D.E.S., A.E.S.
- Ce sont des systèmes dits symétriques dans le sens où la même clé est utilisée pour chiffrer et déchiffrer le message. La taille fixe de la clé (512 bits, 1024 bits, 2048 bits) est une mesure de difficulté de déchiffrement du message en l'absence de la connaissance de la clé.
- Le chiffrement comme le déchiffrement sont très rapides à mettre en œuvre (une vidéo peut facilement être chiffrée et déchiffrée en temps réel sur un système informatique standard).
- Se pose le problème du transfert sécurisé de la clé privée.

Mise en œuvre de la confidentialité et de l'authentification de l'émetteur : les systèmes asymétrique ou à clé publique

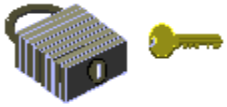
- 2 clés crées simultanément par des algorithmes publics (RSA,…):
  - 1 clé privée, jalousement gardée secrète
  - 1 clé publique, très largement diffusée
- Le principe de base est qu'une donnée chiffrée par une clé, publique ou privée, ne peut être déchiffrée que par la clé associée.
- Les clés sont de même nature et leurs rôles peuvent être inversés : ce que l'une chiffre est déchiffrable par uniquement par l'autre. Il n'est en outre pas possible de déduire la clé privée à partir de la clé publique, même associée à un message.
- Les algorithmes de chiffrement/déchiffrements sont très nettement moins performants que dans le cas des systèmes symétriques, mais il n'y a pas besoin de canal sécurité de transmission de clé.
- Ces systèmes permettent d'assurer également la signature électronique

Cryptographie à clé publique


Alice

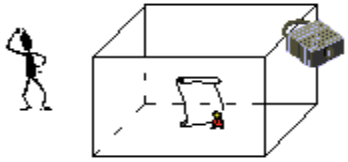

Bob

Etape 1 Fabrication des clés : Bob fabrique une clé publique qui permet de sceller le message dans une boite (ici, le cadenas) et une clé privée qui permet d'ouvrir le cadenas

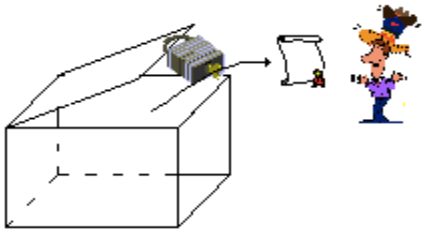


Etape 2 Distribution des clés : Bob fait parvenir à Alice le cadenas, mais garde la clé pour lui

Etape 3 Envoi du message : Alice met son message dans une boite qu'elle ferme à l'aide du cadenas



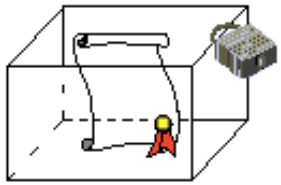
Etape 4 Réception du message : Bob ouvre la boite à l'aide de sa clé et récupère le message. Personne n'a pu l'intercepter car lui seul avait la clé du cadenas.



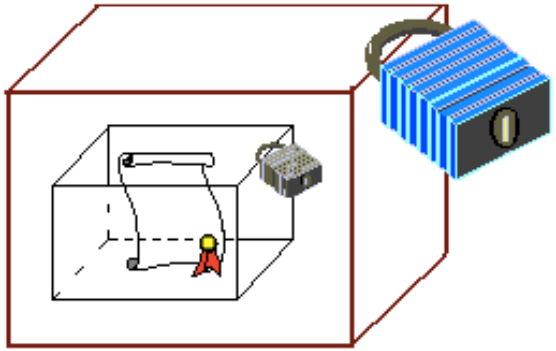
Source : bibmath.net

Source : http://www.bibmath.net

Alice

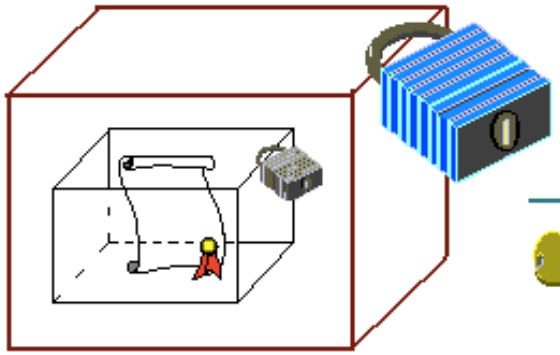


enferme le message dans une boite à l'aide de sa clé privée (authentification de l'expéditeur)

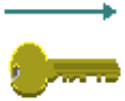


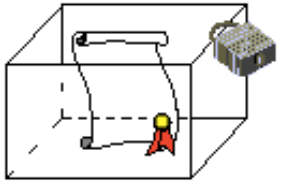
enferme cette boite dans une autre boite avec la clé publique de Bob (chiffrement, seul Bob peut ouvrir)

Bob



reçoit le message chiffré et authentifié





ouvre la première boite avec sa clé privée (déchiffrement)

ouvre la deuxième boite avec la clé publique d'Alice (authentification)

Le certificat électronique

Bob

Bob fournit une fiche d'identification à l'autorité

Autorité de certification

L'autorité :
*vérifie les informations de Bob
*ajoute ses propres données
*signe le certificat.

Alice

clé : 398430355903407

Alice télécharge le certificat auprès de l'autorité, vérifie la signature, et récupère la clé publique de Bob..

Source : http://www.bibmath.net

La clé de session

Méthode symétrique à clé secrète
Chiffrement et déchiffrement rapide mais nécessite un canal sécurisé de transmission des clés

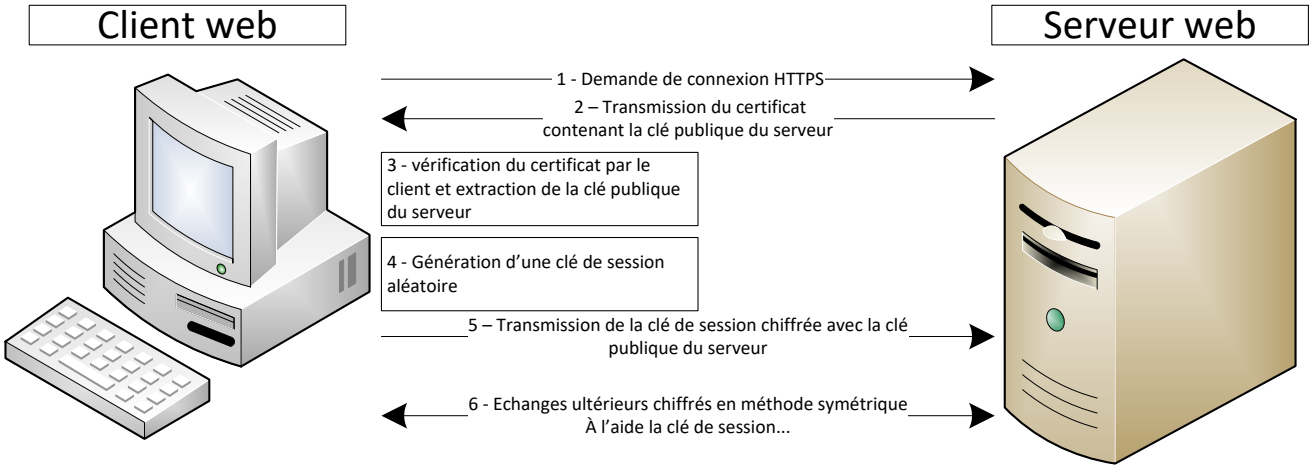
VS

Méthode asymétrique à clé publique
Chiffrement et déchiffrement lent mais ne nécessite pas de canal sécurisé et permet d'authentifier l'émetteur.

⇒

Clé de session : Le canal sécurisé de transmission nécessaire à la clé privée est assuré par un chiffrement à clé publique de cette dernière. La session se poursuit ensuite par un chiffrement à clé privée.

La séquence des opérations est la suivante :



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.

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

Présentation en intention : Relation

ETUDIANT(Num, Nom, Prénom, Courriel)

Présentation en extension : Table

ETUDIANT

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

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

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

Num	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

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

DIPLOME

N°Diplôme	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 ⟝ DIPLÔME

ETUDIANT

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

DIPLOME

N°Diplôme	Libellé	Niveau
1	Licence	Bac+3
2	Master	Bac+5
3	Doctorat	Bac+8

Relations *Plusieurs à plusieurs* : ETUDIANT ⟝⟞ COURS

ETUDIANT

Num	Nom	Prénom	...
20230001	DUPONT	Henri	...
20230002	DUBOIS	Paul	...
20230003	MARTIN	Marie	...

INSCRIPTION PEDAGOGIQUE

Num étudiant	Code Cours
20230002	LIC2
20230002	LIC1
20230001	LIC1

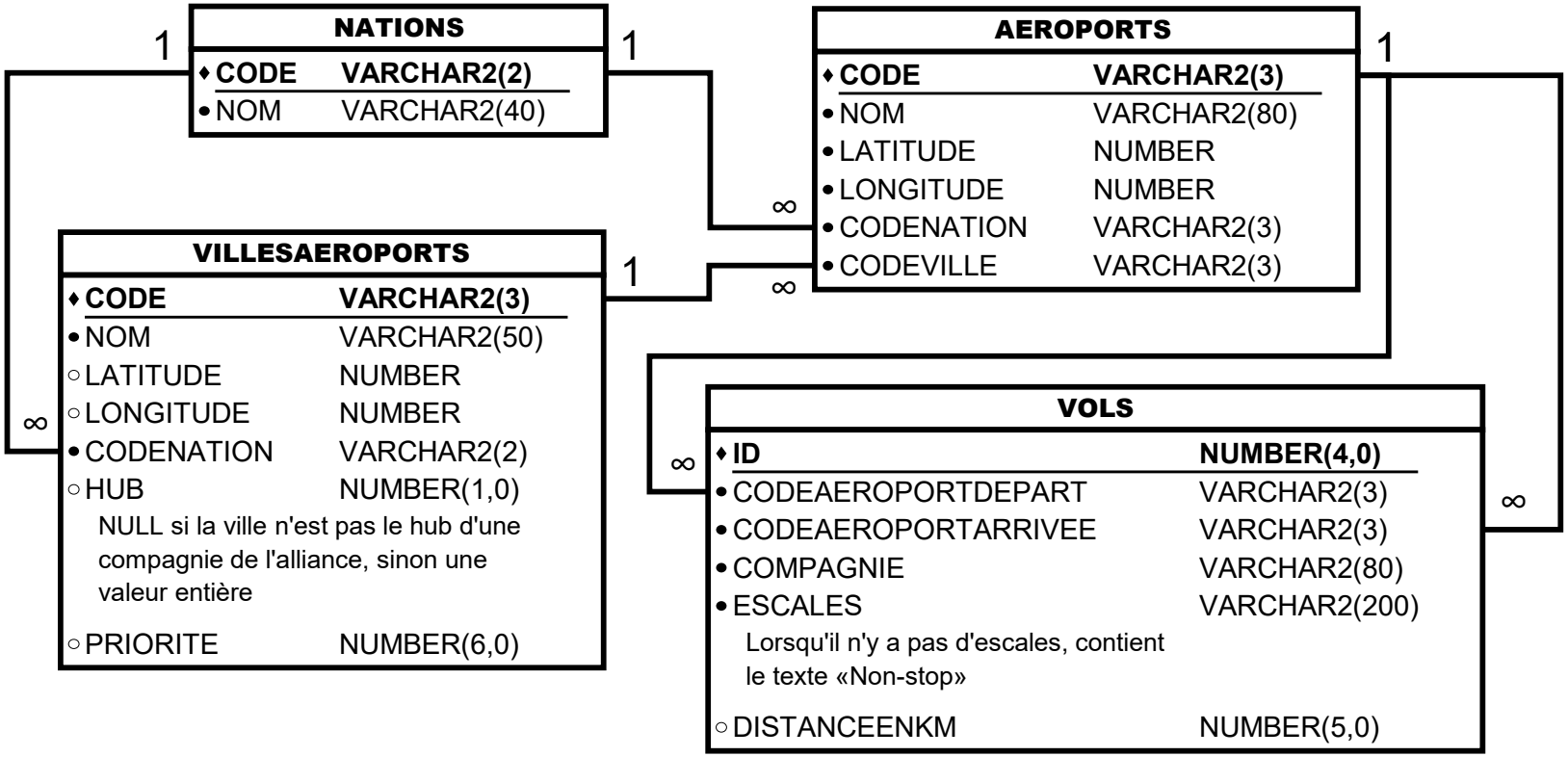
COURS

Code Cours	Libellé	ECTS
LIC1	Droit	3
LIC2	Economie	2
LIC3	Informatique	2

a – Les concepts du modèle relationnel

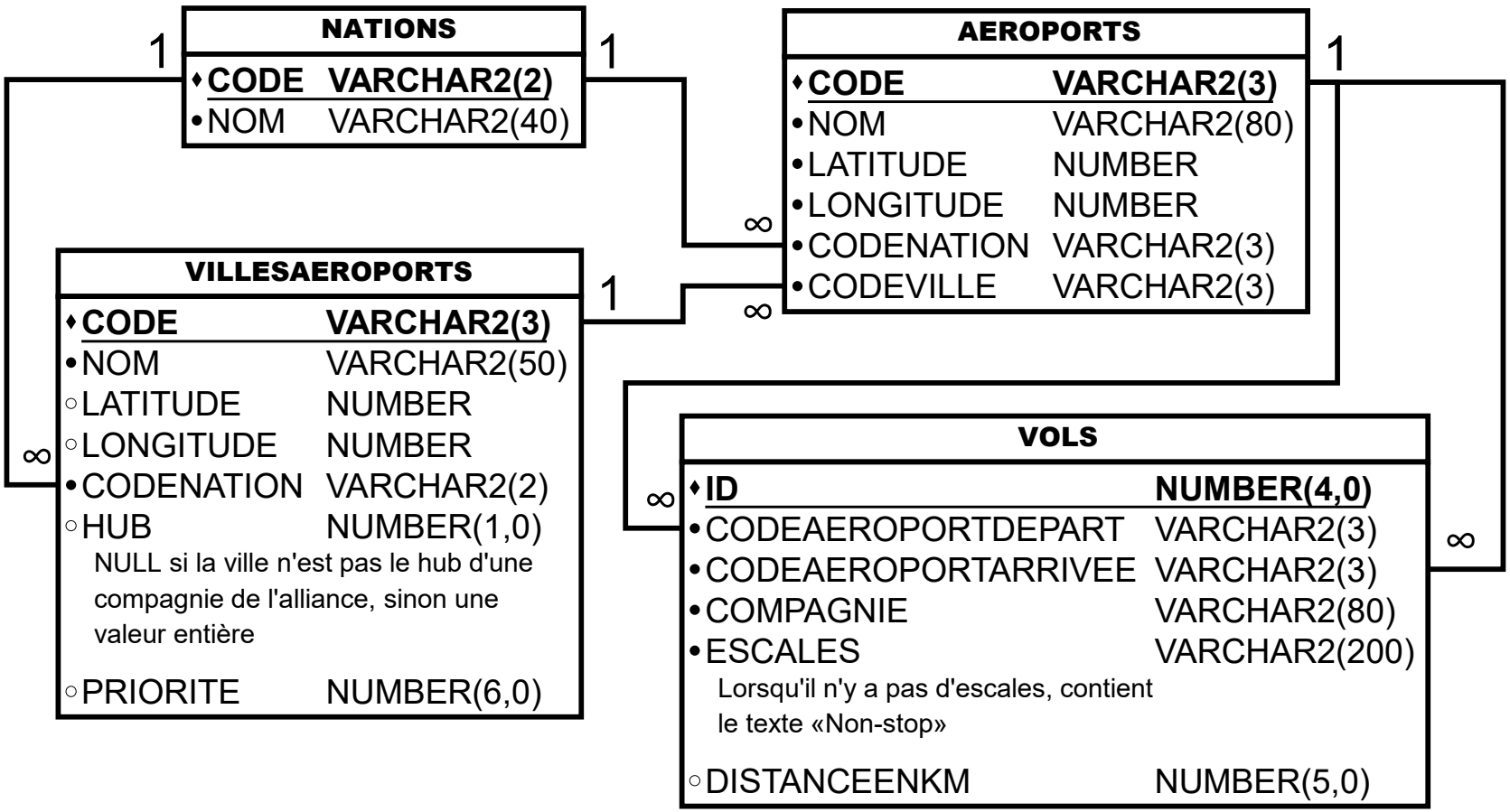
a4 - Schéma relationnel

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 VILLESAEROPORTS

La table **VILLESAEROPORTS** 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 **VILLESAEROPORTS** :

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 – 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

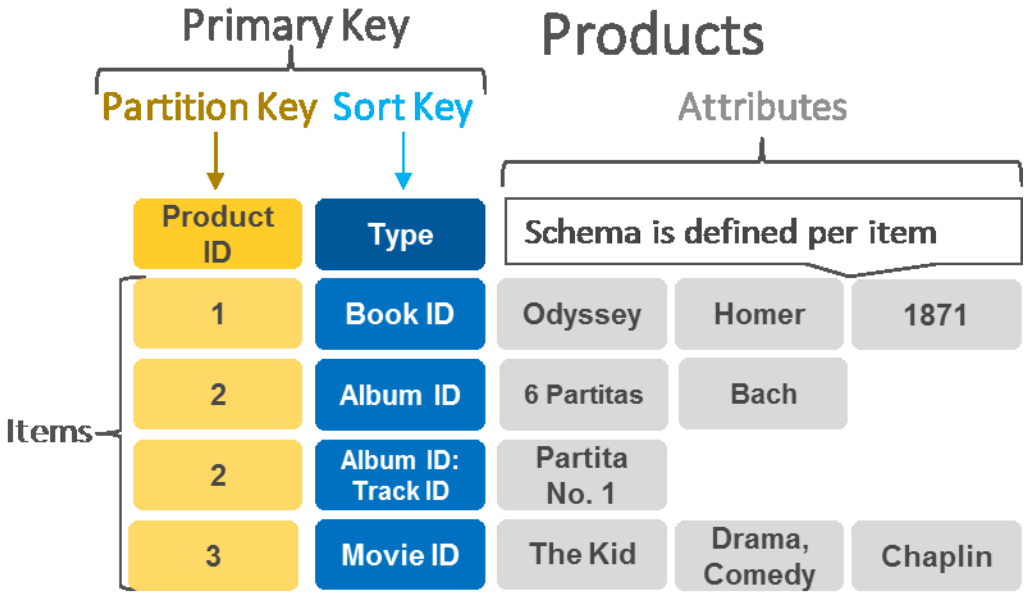
Bases de structure extrêmement simple (une seule table) associant à une clé (présentant une contrainte d'unicité) une valeur pouvant prendre n'importe quel type (simple ou plus complexe).

Clé	Valeur
1	Simple texte
2	3,14159
3	Num : 20230001 / Nom : DUPONT / Prénom : Henri / Courriel : henri.dupont@...
…	

a – Bases clé-valeur

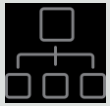
Ex : DynamoDB Amazon

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

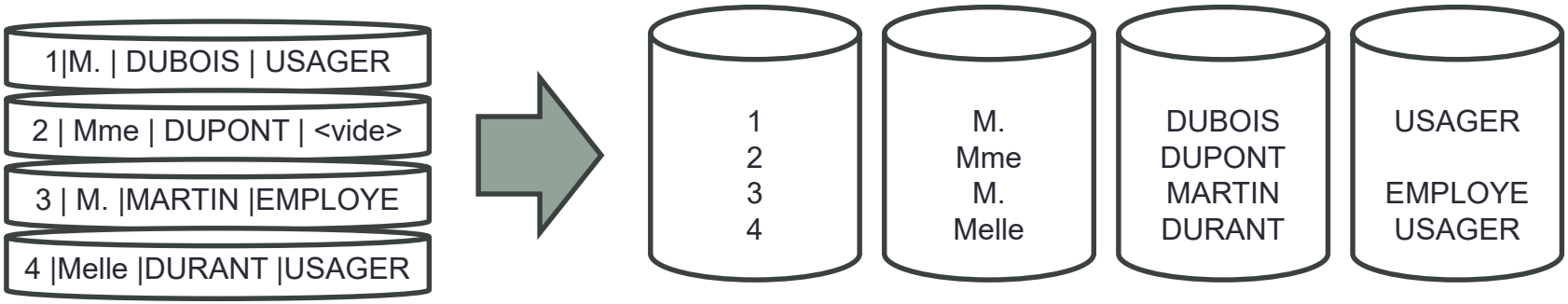
https://www.mongodb.com/fr-fr

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*

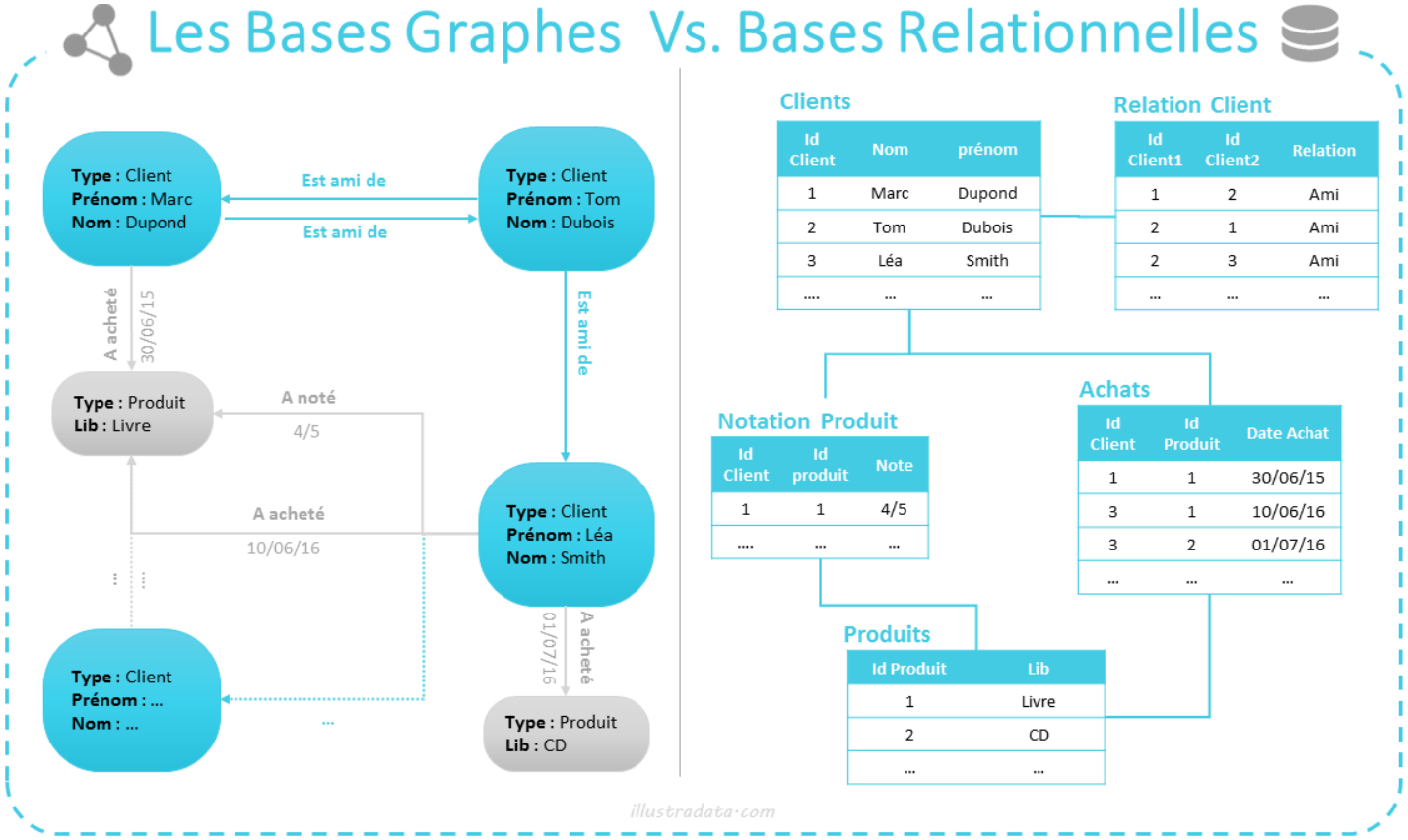
N°	Civilité	NOM	QUALITE
1	M.	DUBOIS	USAGER
2	Mme	DUPONT	
3	M.	MARTIN	EMPLOYE
4	Melle	DURANT	USAGER



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.

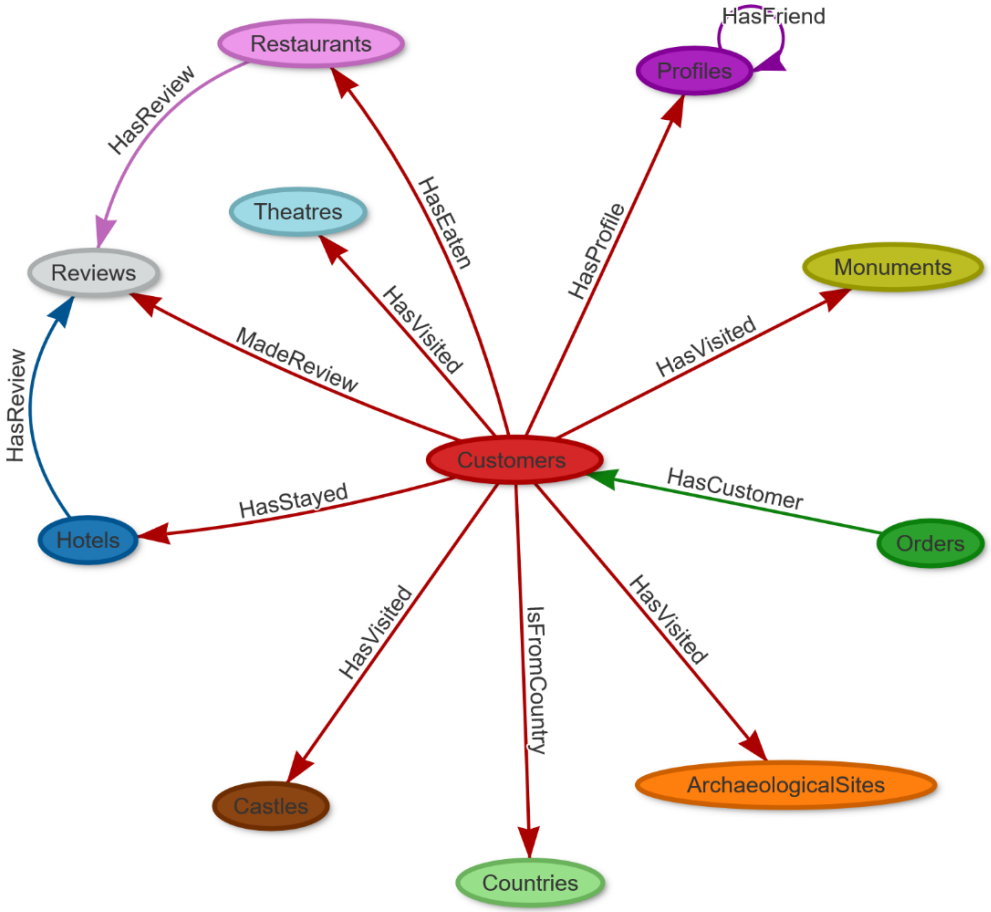
Source : illustradata.com



d – Bases orientées graphes

Ex : SAP OrientDB

Schéma de sa base de démonstation



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
+ **V**ERSATILITE : Bases structurées comme non structurées (NOSQL)
+ **V**ITESSE : Un traitement efficace grâce à la distribution et l'IA
= **V**ALEUR : Un 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

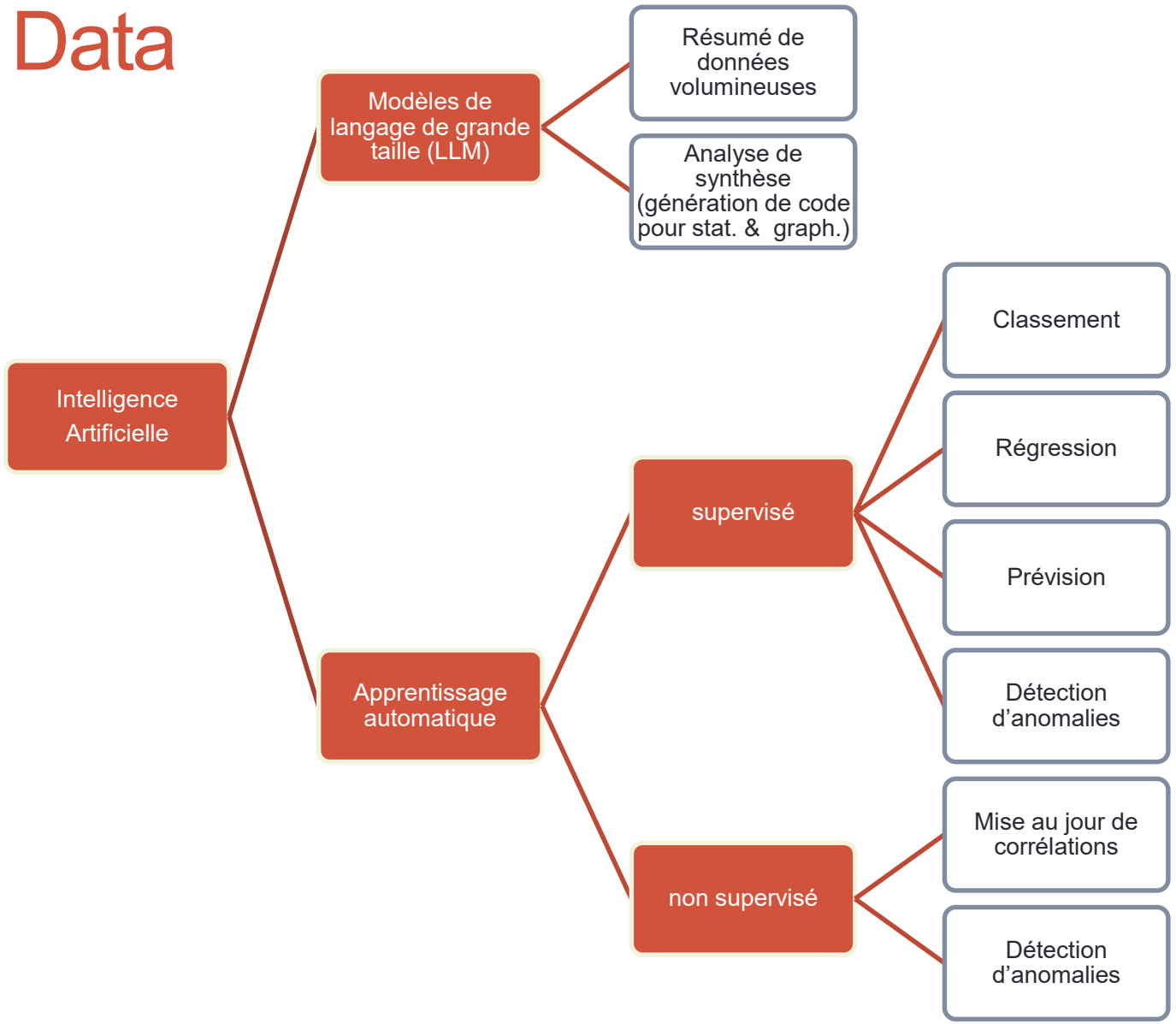
2 – Le Cloud Computing

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 règlementations sur les données personnelles et RGPD

3 – Apports de l'Intelligence Artificielle au Big Data



Les LLM Large Langage 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 Langage 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

L'apprentissage automatique

Domaines d'application :

- Classification
- Prévision
- Détection d'anomalies

Etapes : Entrainement → 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 (reconnaissances 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 maintien 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'immutabilité 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

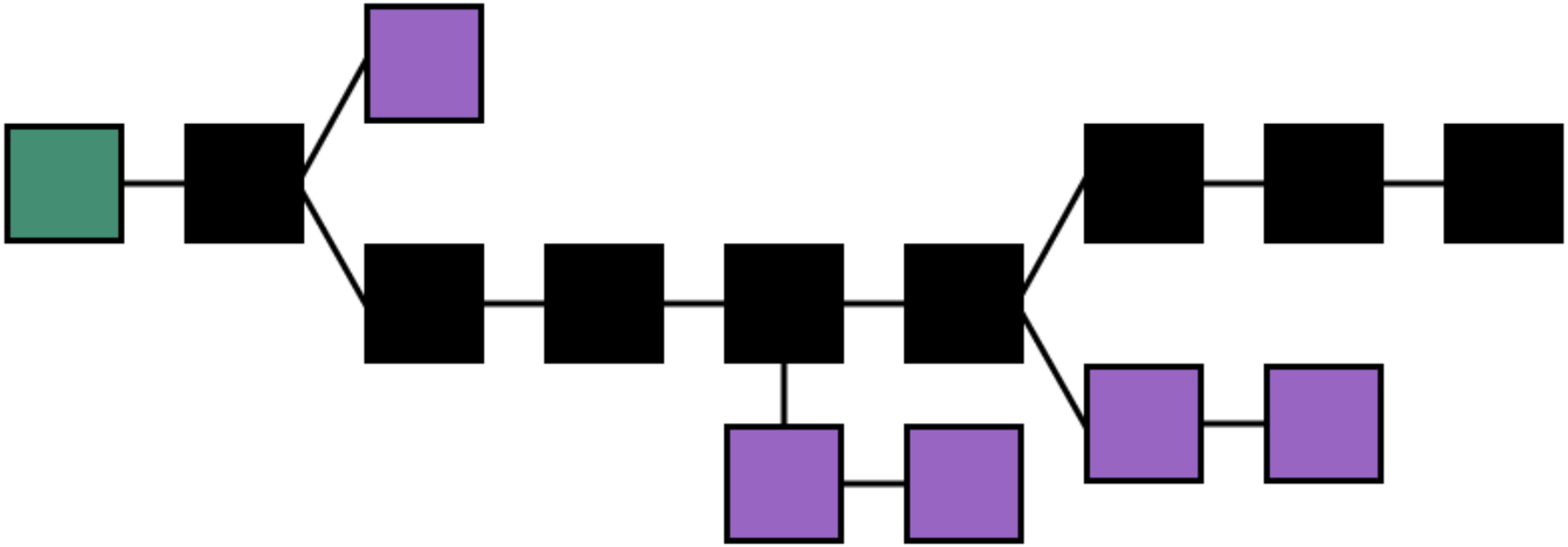
Privée

- 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 – Les réseaux

A – Introduction : terminologie et concepts

B – Protocoles internet

C – Les services apportés par les réseaux

A - Introduction

- **Réseau** : Ensemble d'ordinateurs et de périphériques connectés les uns aux autres en vue d'un partage d'informations ou de périphériques
- **Serveur**: Ordinateur délivrant une information à un autre au travers d'un réseau ou proposant des périphériques en partage
- **Client** : Ordinateur recevant un information ou bénéficiant de l'usage d'un périphérique au travers d'un réseau
- **Peer to Peer (P2P)** : Réseau dans lequel le partage s'effectue sans serveur dédié central

A – Introduction

- 1 – Composition basique d'un réseau local
- 2 – Caractéristiques d'un réseau
- 3 – Du réseau local au réseau à large étendue
- 4 – Différentes technologie de mise en relation
- 5 – Ubiquité au travers des liaisons sans fil
- 6 – Convergence des réseaux

1 – Composition basique d'un réseau local

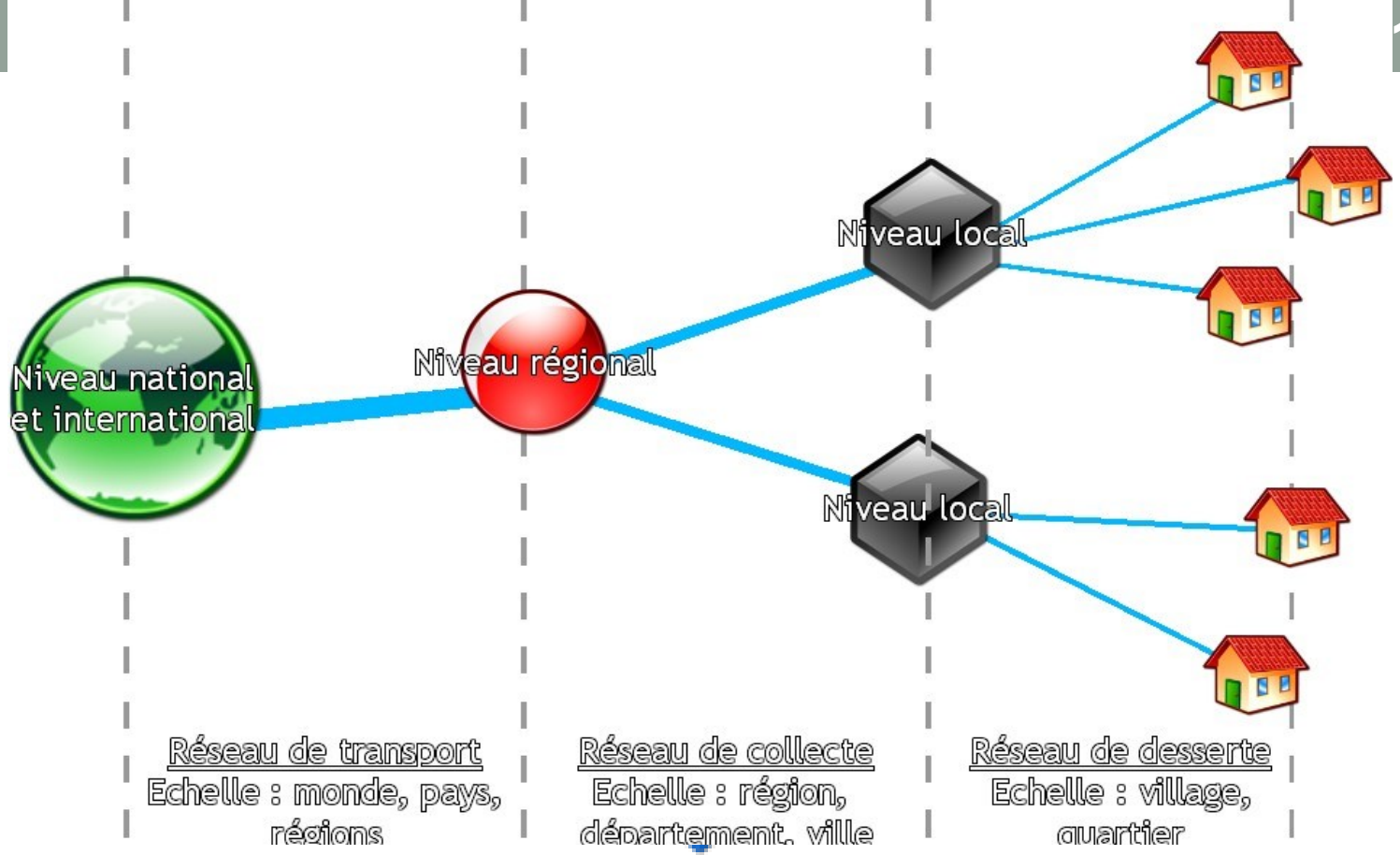
- Au moins 2 ordinateurs
- Une carte réseau par ordinateur
- Un moyen de communication, filaire ou sans fil
- Un logiciel d'exploitation compatible réseau: comprenant les protocoles (dialectes) assurant une communication correcte sur le réseau
- Éventuellement des dispositifs servant à assurer l'intermédiation des ordinateurs : borne Wifi / Concentrateur ou *Hub* / Commutateur ou *Switch*

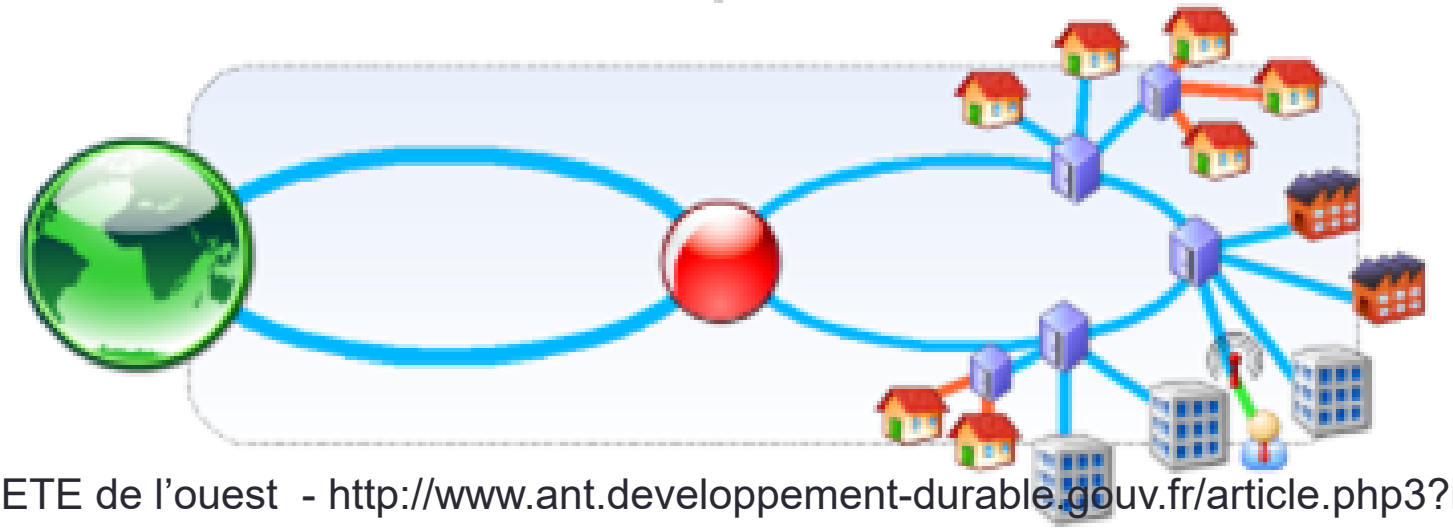
2 – Caractéristiques

- **Débit** : Quantité de données transmise pendant une unité de temps
- **Délai de transmission (latence) :** Temps nécessaire à la traversée du réseau
- **Variation du délai (gigue) :** Écart entre les délais de transmission des différents paquets consécutifs
- **Taux d'erreur :** Pourcentage de paquets de données perdus/altérés sur le total de paquets envoyés

3 – Du réseau local au réseau à large étendue

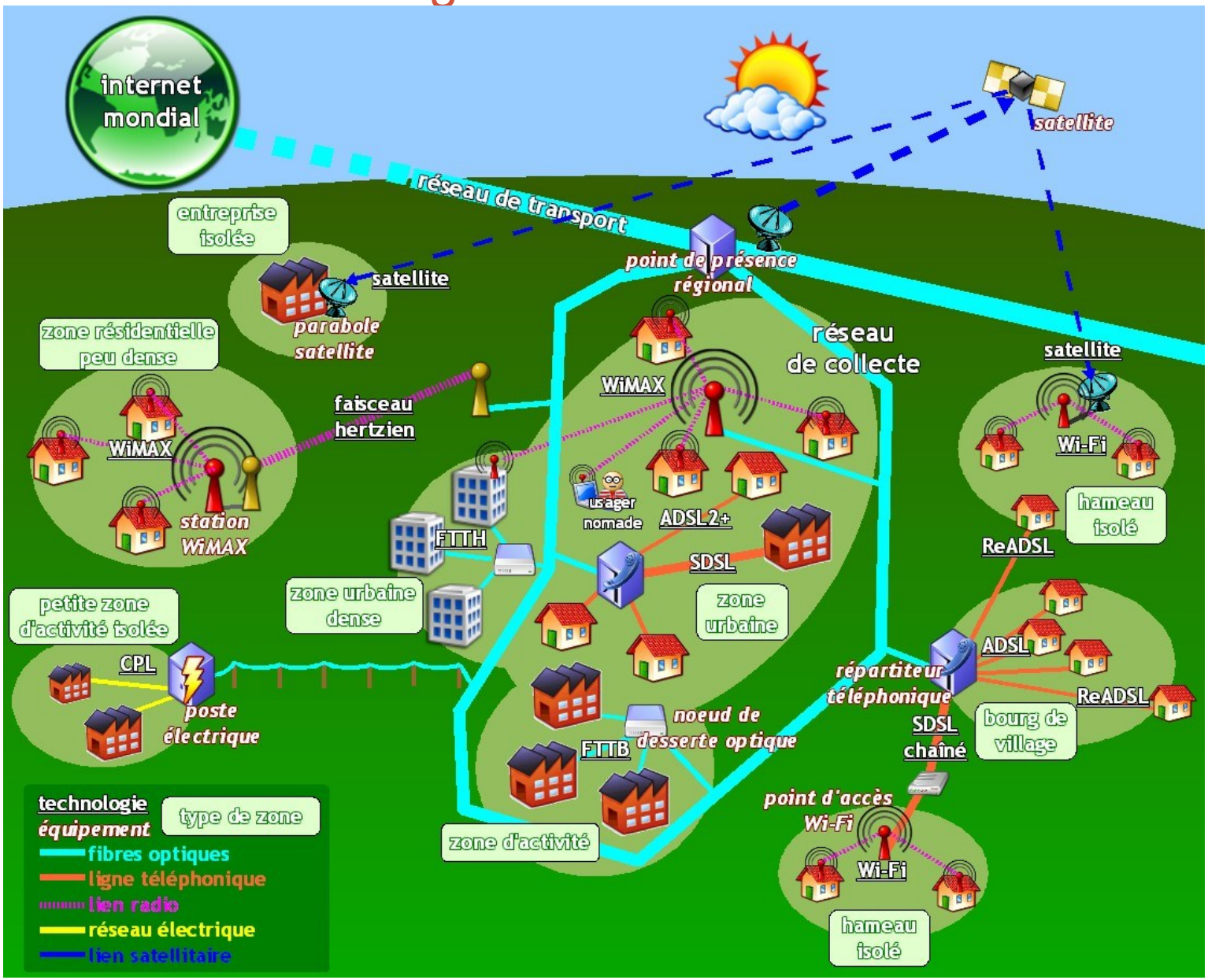
- Utilisation de routeurs
  - Dispositifs capables de réaliser une interconnexion de réseaux de technologies identiques ou différentes.
  - Dispositifs ayant une connaissance de leur environnement lointains et de ce fait aptes à orienter les messages vers la prise la plus efficace.



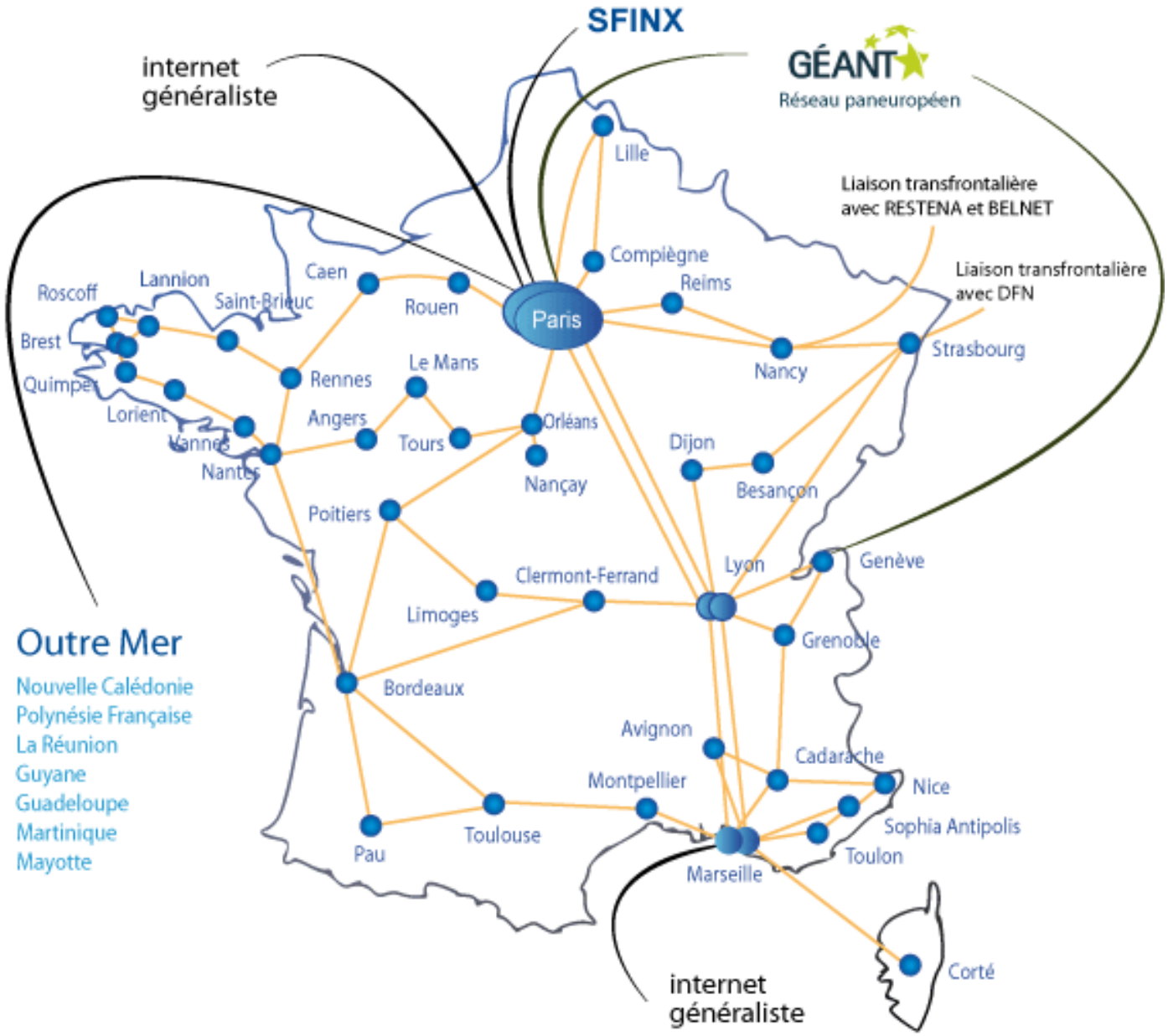


Source CETE de l'ouest - http://www.ant.developpement-durable.gouv.fr/article.php3?id_article=99

4 – Différentes technologies de mise en relation

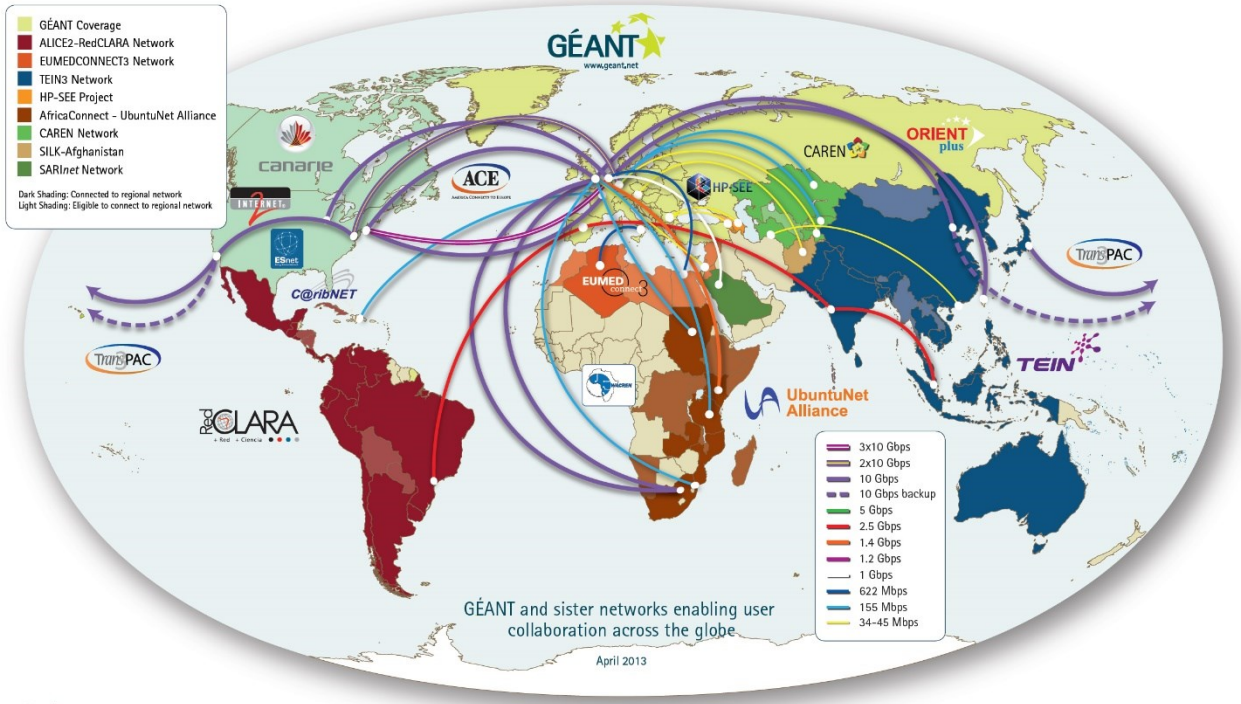


Source CETE de l'ouest - http://www.ant.developpement-durable.gouv.fr/article.php3?id_article=99



Source : www.geant.net

GÉANT At the Heart of Global Research Networking



DANTE www.dante.net GÉANT www.geant.net

connect • communicate • collaborate

GÉANT is co-funded by the European Union within its 7th R&D Framework Programme.
This document has been produced with the financial assistance of the European Union. The contents of this document are the sole responsibility of DANTE and can under no circumstances be regarded as reflecting the position of the European Union.

SEVENTH FRAMEWORK PROGRAMME

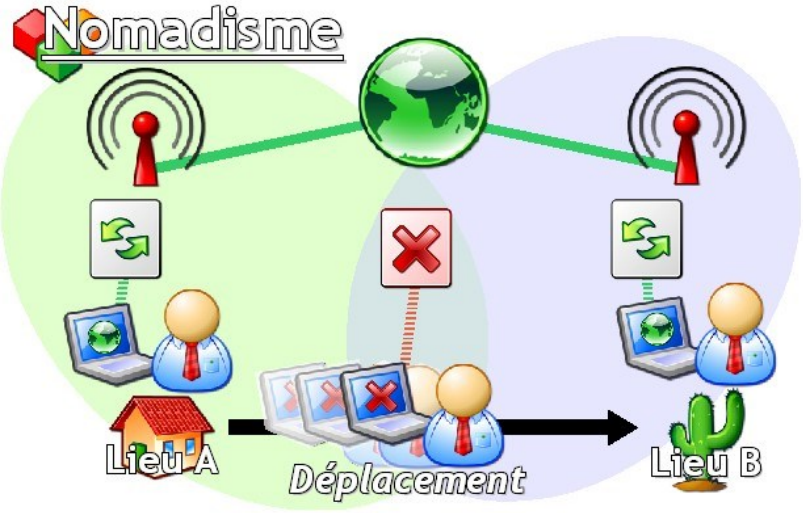
Co-funded by the European Union

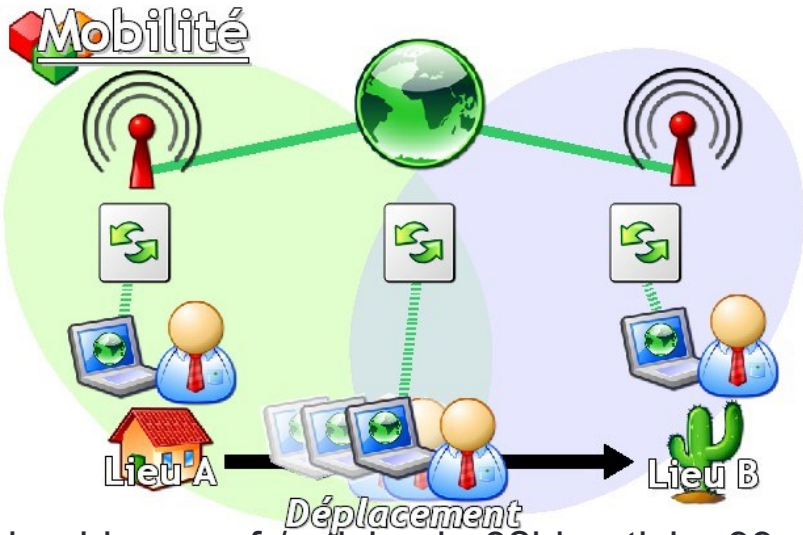
5 – Ubiquité au travers des liaisons sans fil

L'**ubiquité** du réseau désigne la capacité, pour l'usager, de se connecter depuis différents lieux.

le **nomadisme** : permet à l'usager de se connecter depuis différents lieux sans toutefois pouvoir maintenir la connexion pendant un déplacement,

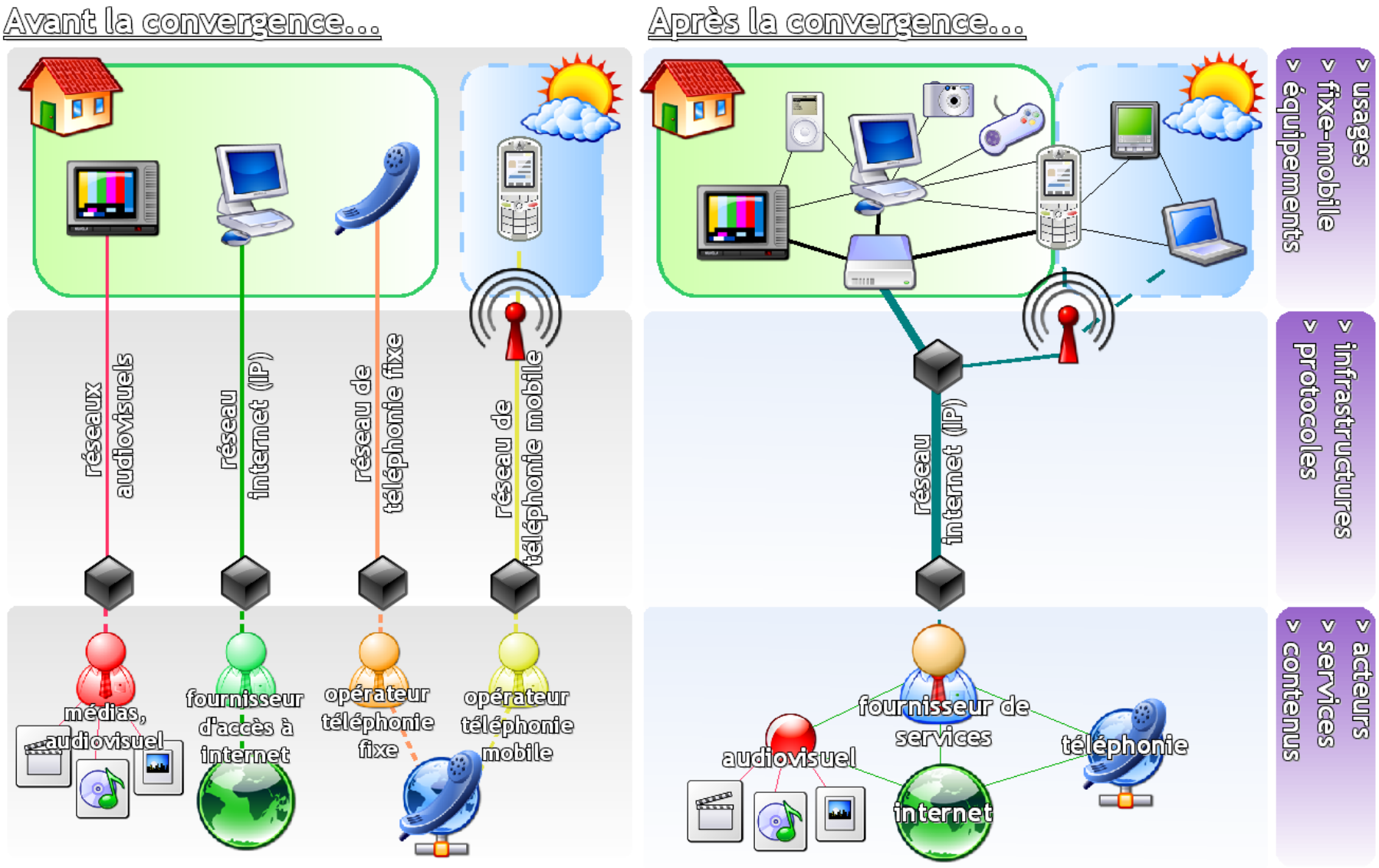
la **mobilité**, qui permet de rester connecté même pendant un déplacement (en train, en voiture...).





Source CETE de l'ouest - http://www.ant.developpement-durable.gouv.fr/article.php3?id_article=99

6 – Convergence des réseaux



Source CETE de l'ouest - http://www.ant.developpement-durable.gouv.fr/article.php3?id_article=99

B – Les protocoles d'internet

- 1 – Protocole
- 2 – Organisation des protocoles
- 3 – L'encapsulation de données
- 4 – Le multiplexage / démultiplexage
- 5 – L'adressage et la résolution de noms

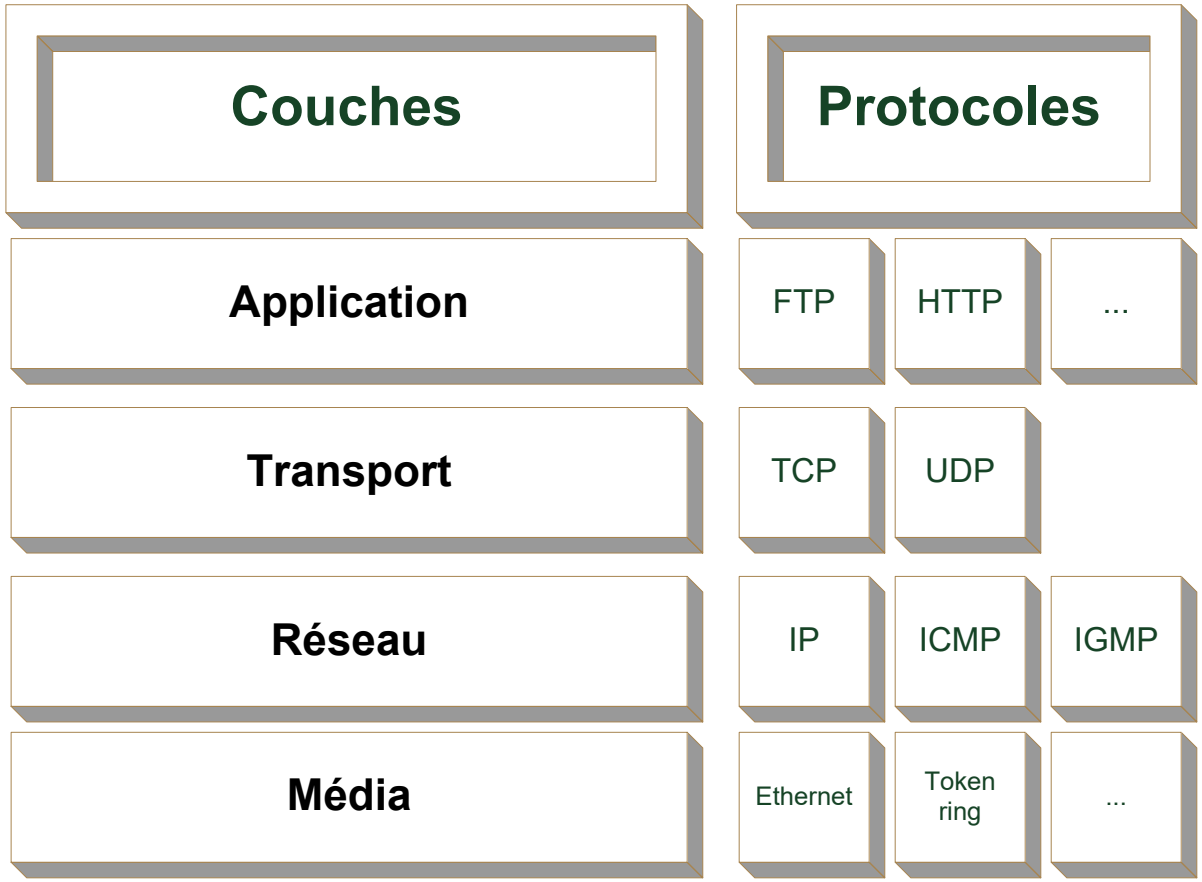
B – Les protocoles d'internet

1 – Protocole

Un **protocole** est une méthode standard qui permet la communication entre des processus (s'exécutant éventuellement sur différentes machines), c'est-à-dire un ensemble de règles et de procédures à respecter pour émettre et recevoir des données sur un réseau. Il en existe plusieurs selon ce que l'on attend de la communication. Sur internet, l'ensemble des protocoles utilisé est appelé TCP-IP (Transport Control Protocol – Internet Protocol)

B – Les protocoles d'internet

- 2. Organisation des protocoles (Modèle ISO) :



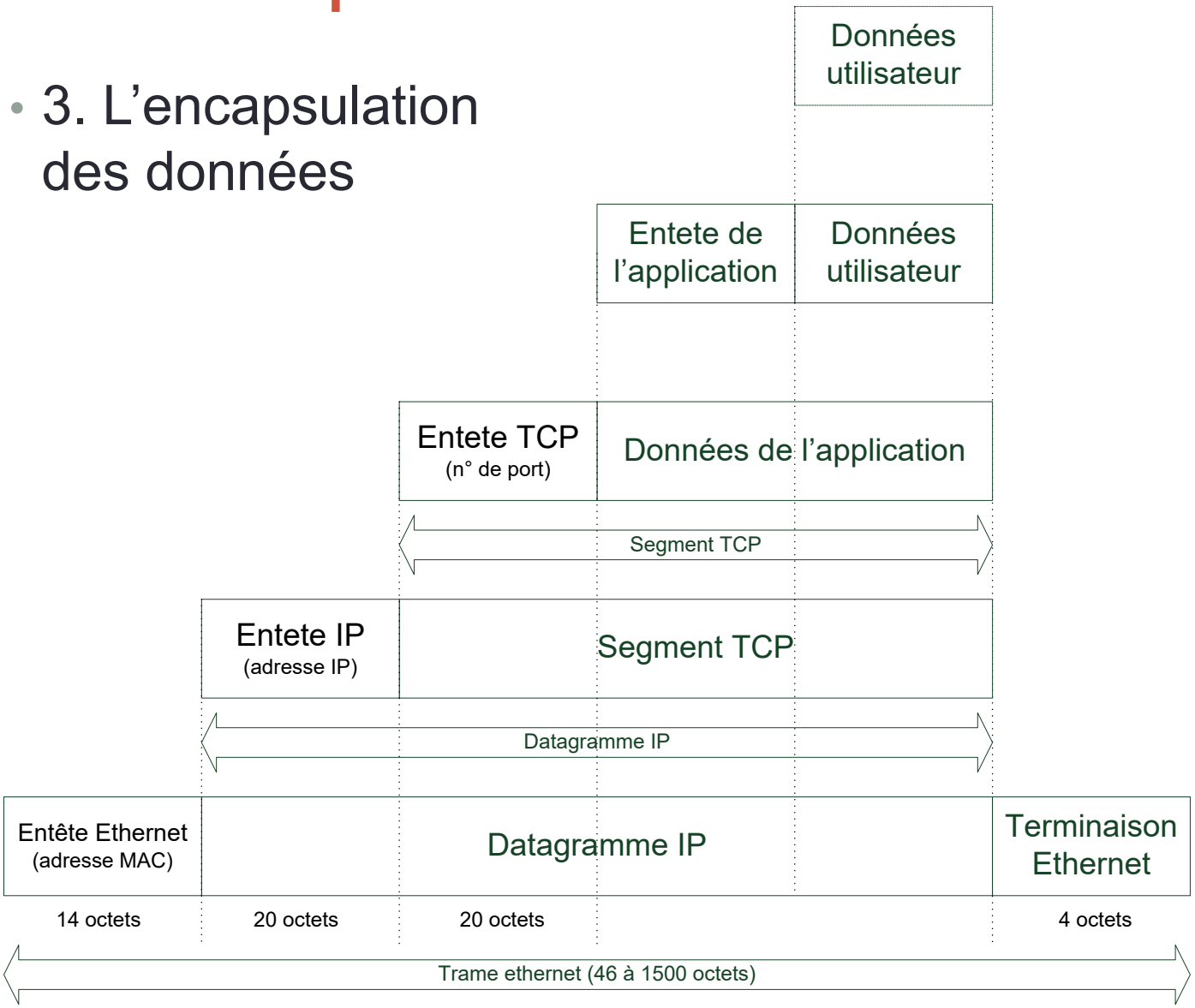
B – Les protocoles d'internet

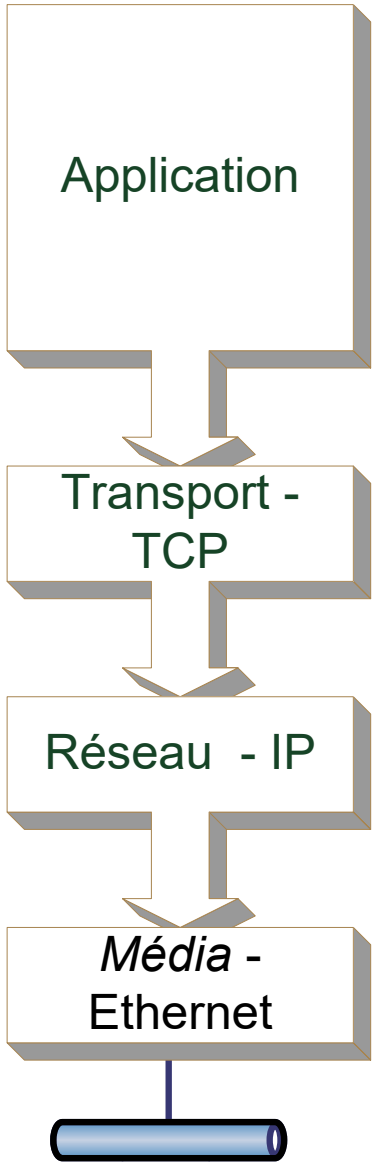
2. Organisation des protocoles (Modèle ISO) :

- **Application** : Gère les données propres à chaque application ou service.
- **Transport** : Gère un flux de données entre deux machines afin de les transmettre à la couche application. Elle valide les messages (sommes de contrôle et délais, accusés de réception et délais d'attente).
- **Réseau** : Gère le déplacement et le routage de paquets de données sur le réseau
- **Média** : Mise en œuvre de différentes technologies de communication pour mettre en relation les systèmes

B – Les protocoles internet

- 3. L'encapsulation des données





B – Les protocoles internet

- 4 – Le multiplexage / démultiplexage :
  Comment recomposer l'information transmise entre 2 applications sur 2 machines décomposée en paquets transmis en vrac sur le réseau.

  - Si la machine destinataire n'est pas sur le même réseau local, l'adresse MAC (Media Access Chanel) de la trame ethernet émise est celle du routeur.
  - L'adresse IP sert à déterminer le routeur du correspondant (réseau local).
  - L'adresse MAC de destination (traduite par le routeur de destination à partir de l'adresse IP) permet de choisir la machine destinataire sur le réseau local.
  - Le type de trame éthernet en réception permet de choisir le protocole de la couche réseau (IP,ICMP,IGMP).
  - Le type de datagramme permet de déterminer le protocole de transport à utiliser (TCP ou UDP).
  - Le port de destination permet de déterminer l'application destinataire.

B – Les protocoles internet

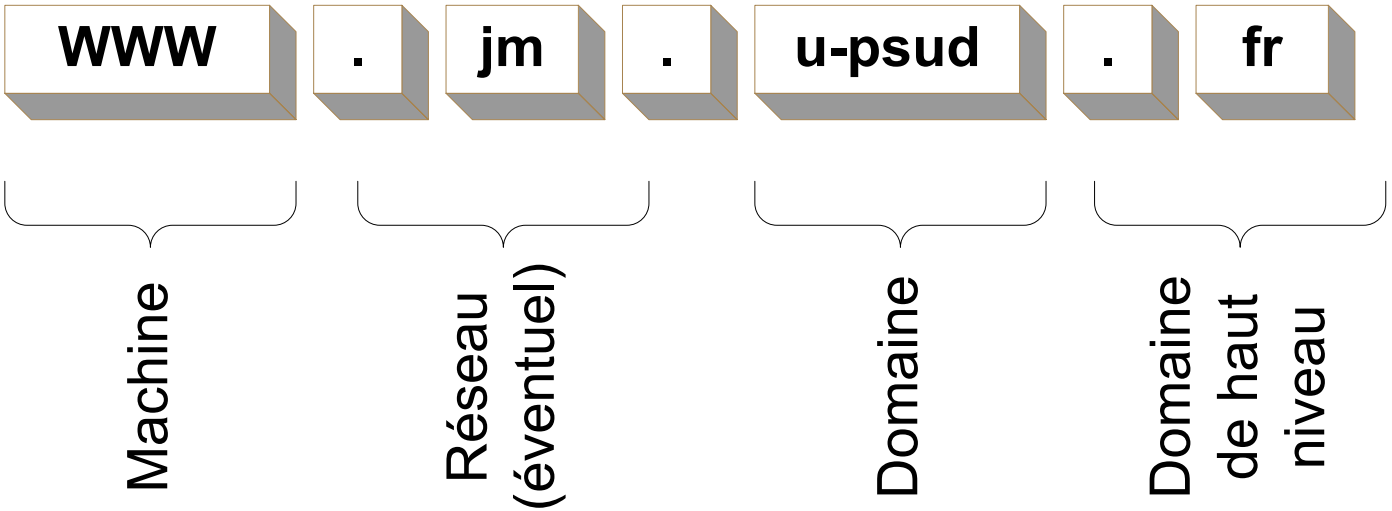
5. L'adressage et la résolution des noms

- Adresse IP (IPv4):

  Nombre de 4 octets ex :129.175.169.5

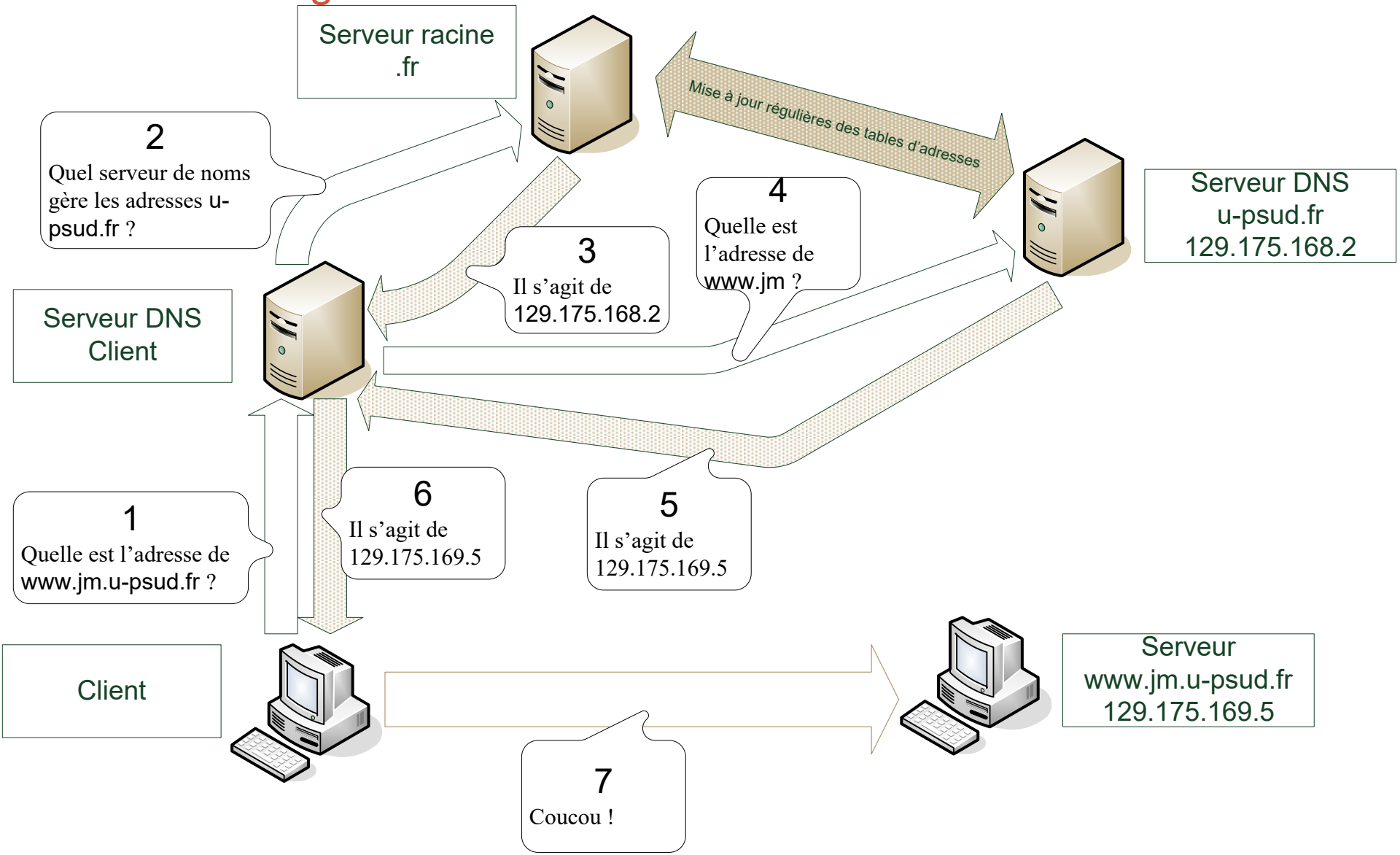
  Masque de réseau : ex : 255.255.255.0

- Nom :



B – Les protocoles internet

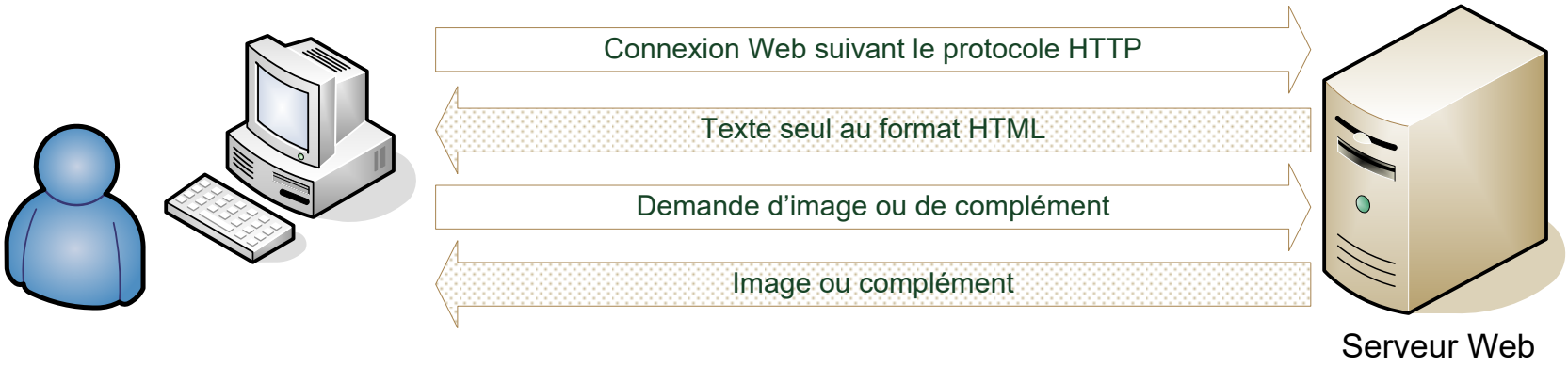
5. L'adressage et la résolution des noms



C – Les services apportés par les réseaux

- 1 – Les services de communication

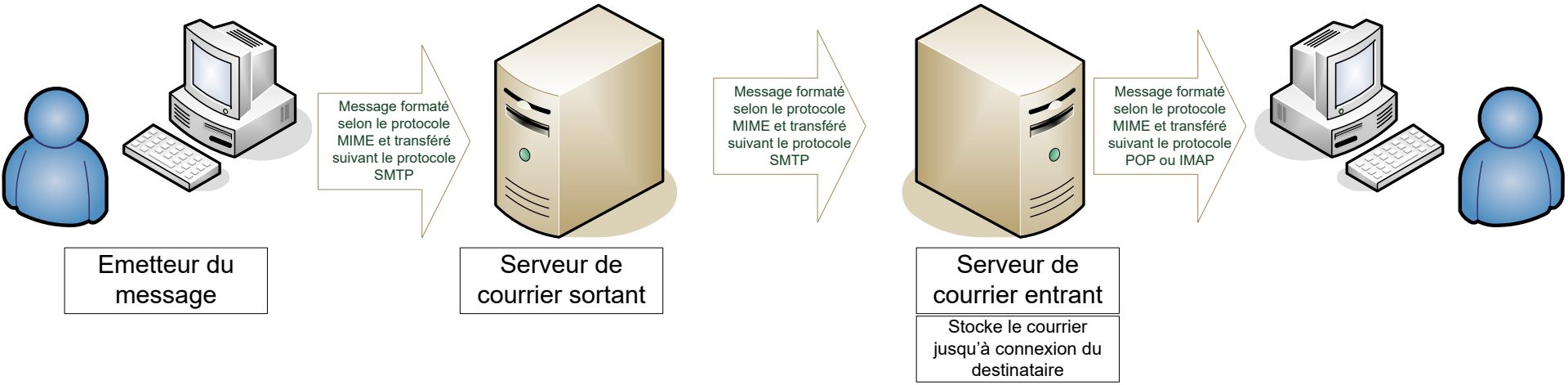
- 1.1 : Le WWW



C – Les services apportés par les réseaux

- 1 – Les services de communication

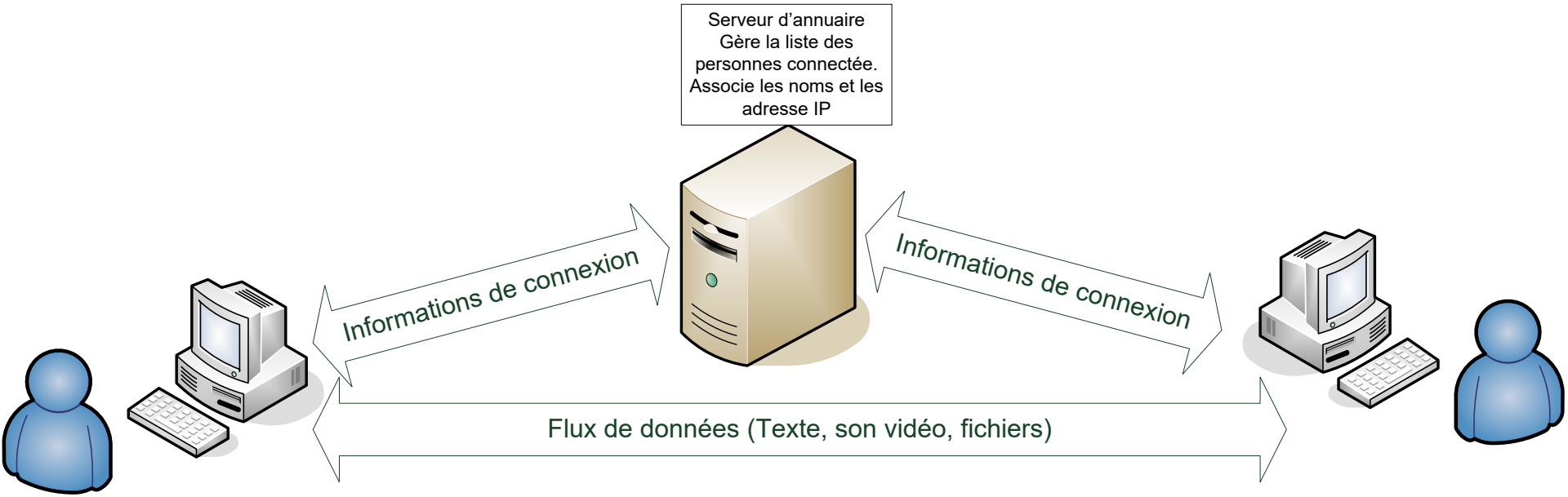
- 1.2 : La messagerie en temps différé



C – Les services apportés par les réseaux

- 1 – Les services de communication

- 1.3 : La messagerie en temps réel et les système d'audio ou de visio-conférence



2 – Les services de diffusion

2.1. Le téléchargement

Transmission d'un fichier à partir d'un serveur vers un ordinateur client, initiateur de l'opération. Sur ce dernier, le fichier obtenu est la copie conforme de celui du serveur.

- Avantages et contraintes
  - Pas de contrainte de débit des réseaux
  - Le fichier peut être rejoué ultérieurement si aucun dispositif ne s'y oppose, et cela sans connexion internet
  - Nécessité pour l'ordinateur client de posséder un espace de stockage suffisant pour recevoir l'intégralité du fichier
  - Le fichier n'est généralement exploitable qu'une fois téléchargé dans son intégralité.
- Les modes de téléchargement

  Parmi tous les modes possibles de téléchargement, les protocoles HTTP (Hyper Text Transfert Protocol) et FTP(File transport Protocol) sont les plus courants, car gérés de manière transparente.

2.2. La diffusion en continu

La diffusion en continu ou *streaming* est une communication entre un ordinateur client et un ordinateur serveur dans laquelle le client se met à l'écoute des émissions dont l'envoi s'effectue en permanence à l'initiative du serveur.

Vrai et faux Streaming

- Le faux streaming :

  Il s'agit de la lecture progressive d'un fichier pendant son téléchargement.

- Le vrai streaming

  Consiste à envoyer au client uniquement les informations dont il a besoin à un instant donné, sans qu'il n'ait jamais à les stocker. Une séquence sera ainsi envoyée sous forme de morceaux, immédiatement effacés après audition à mesure qu'arrivent les suivants

Diffusion *unicast* et *multicast* :

Unicast (un à un): Diffusion du serveur d'un flux de données par client.

Avantages : Serveur et client peuvent négocier le meilleur choix de bande passante, possibilité de diffusion à la demande

Les connexions unicast sur internet utilisent généralement le protocole TCP(Transfert Control Protocol)

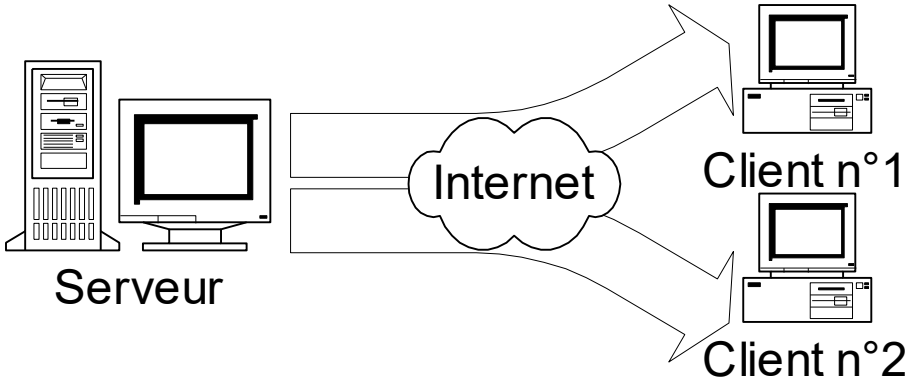
Multicast (un à plusieurs): Diffusion du serveur d'un seul flux de données à l'attention de plusieurs clients, lesquels lisent simultanément les mêmes informations

Avantages : Moins de ressources consommées sur le serveur et le réseau, possibilité d'une plus forte bande passante à moyens égaux.

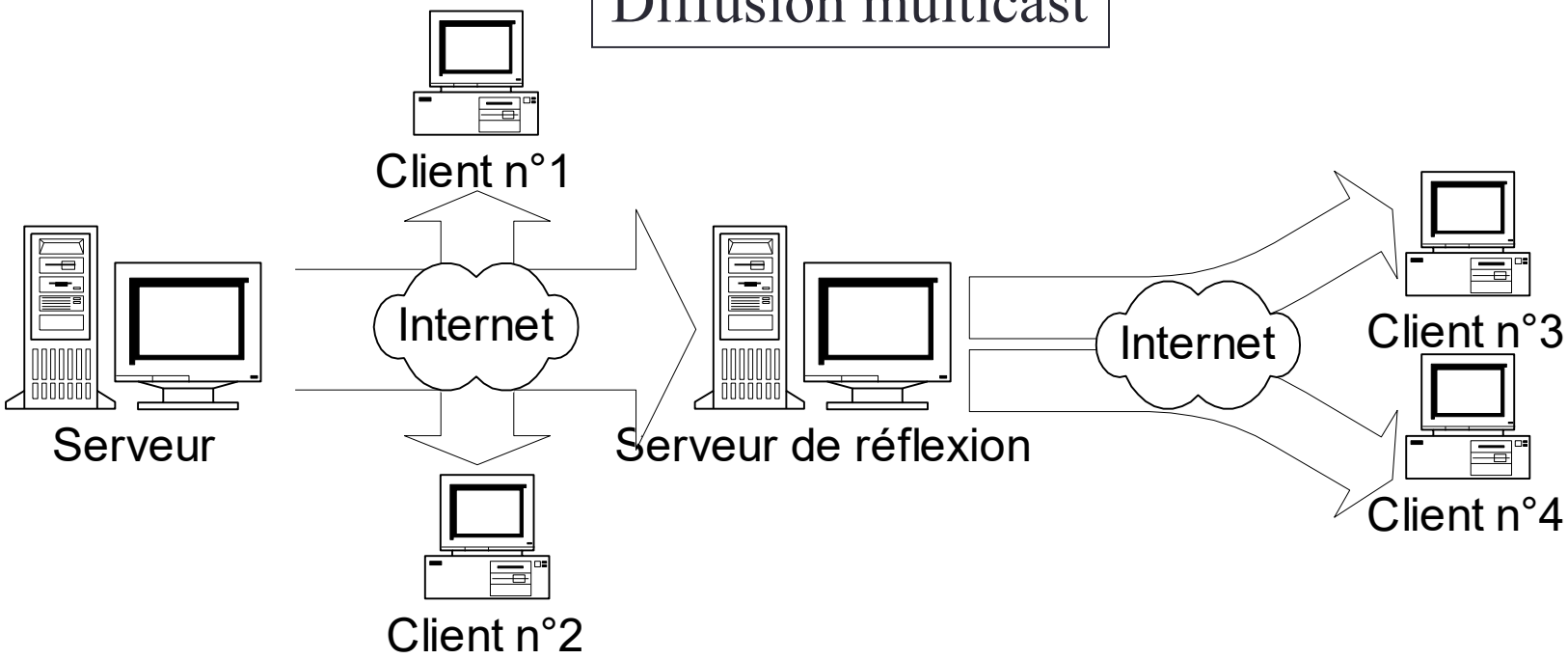
Sur les réseaux qui n'acceptent que l'unicast, les diffusions multicast sont assurées par des serveurs de réflexion.

Sur le réseau internet, les diffusions en multicast utilisent souvent le protocole UDP (User Data Protocol)

Diffusion unicast



Diffusion multicast



Avantages et contraintes

- Pas de necessité de stockage importante
- Nécessité d'une bande passante de réseau suffisante pour le débit attendu
- Impossibilité (souvent souhaitée mais pas toujours obtenue) de rejouer le morceau sans effectuer une nouvelle connexion internet.

3 - Le peer to peer (point à point) ou P2P

- S'oppose au paradigme client-serveur qui définit précisément les rôles de machines dans la fourniture d'un service numérique.
- Dans un système P2P, tous les composants du réseau partagent leurs capacités de traitement et de stockage de manière à fournir le service rendu par le système. Les composants du réseau peuvent s'échanger des informations sans l'intermédiation d'un serveur central et chaque composant du réseau peut être sollicité pour fournir un élément du service rendu par le système.

Les générations de P2P

- 1ère génération (NAPSTER)
  - Centralisation de la recherche
  - Décentralisation du téléchargement / streaming

2e génération :

- Décentralisation (Kazaa, gnutella, Emule) voire abandon (Bittorrent) de la recherche

- Décentralisation du téléchargement / streaming

3e génération :

- Objectif :

  Eviter la surveillance du réseau, éviter l'enregistrement de l'activité du réseau (*honeypots*)

- Moyens
  - Utilisation de proxys (Tor)
  - Réseaux Amis à amis (Darknet,…): adhésion par cooptation
  - Gestion de l'anonymat (freenet,…)
    - Accès indirect
    - Chiffrement du contenu stocké par les composants du réseau