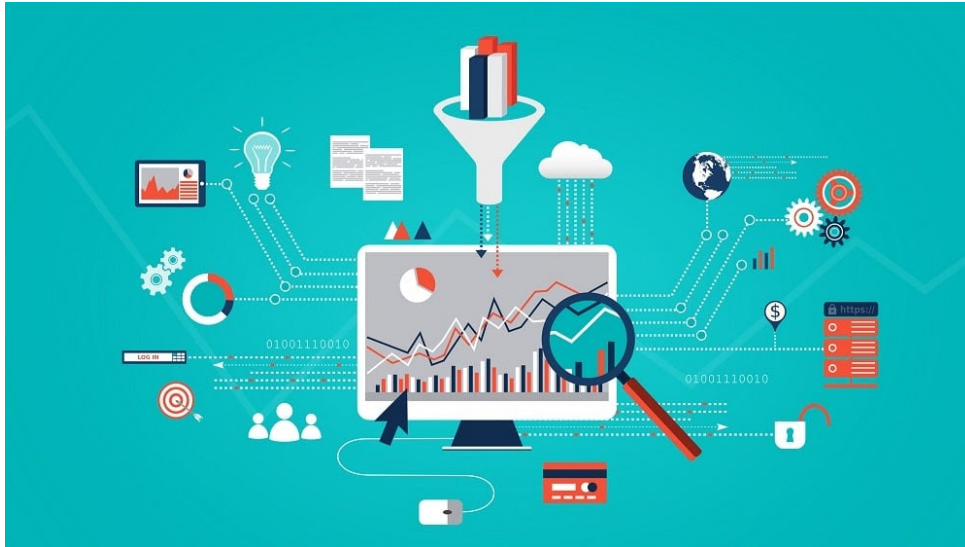


i162 - Analyser et modéliser des données

Loré Godel

06/09/23

Analyser et modéliser des données



Définition :

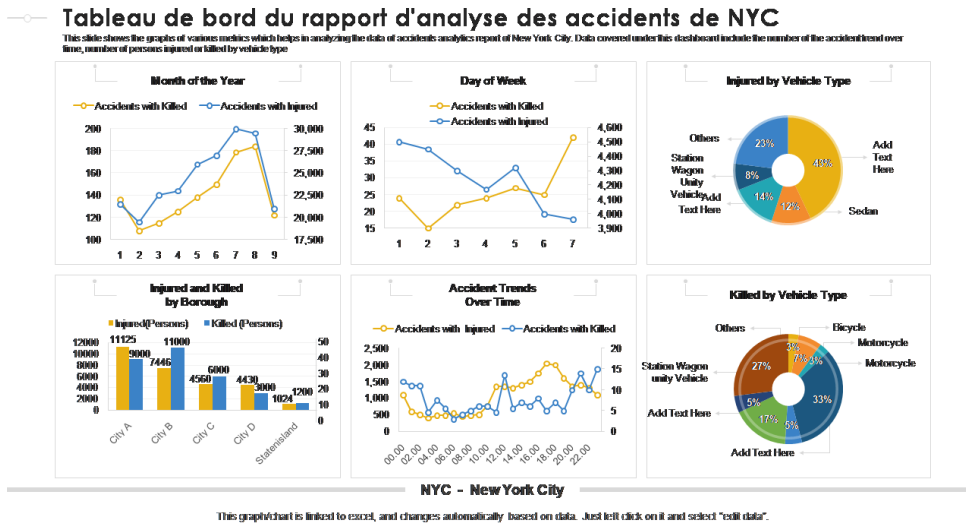
- **Analyser des données** : Examiner/Interpréter les données accompagné de graphiques pour élaborer des réponses à des questions.
- **Modéliser des données** : Etablir la structure, les associations, les relations et les contraintes des données disponibles.

Exemples :

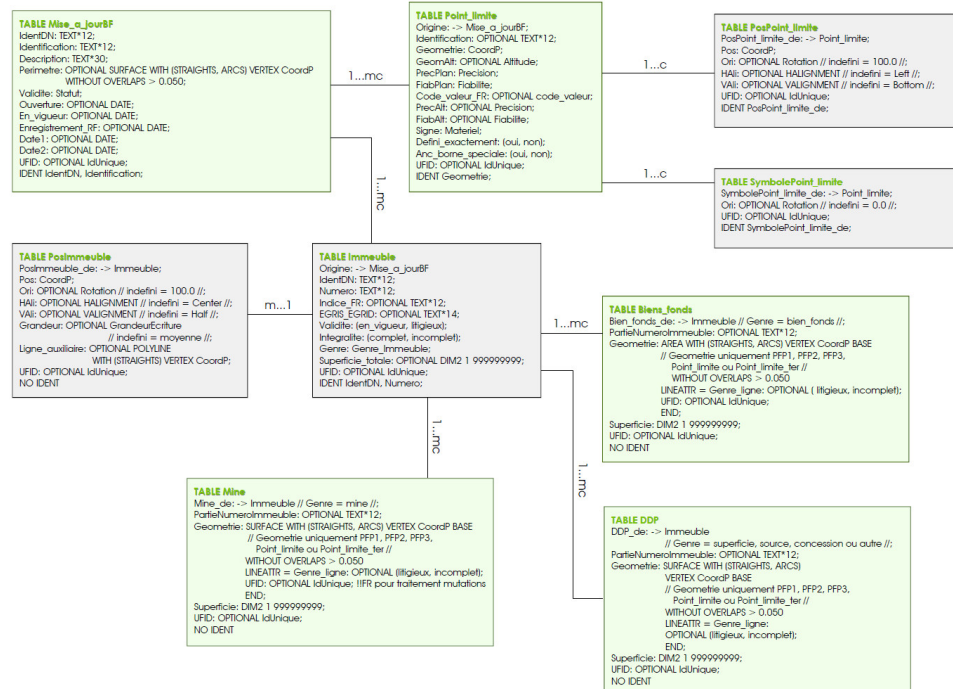
- **Analyse** de base de données ex.:

protected by reCAPTCHA

Privacy - Terms



• Modélisation base de données ex.:



sources images :

(<https://www.imope.fr/data.html> (<https://www.imope.fr/data.html>))

(<https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang=French> (<https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang=French>))

(<https://www.fr.ch/territoire-amenagement-et-constructions/cartes-plans-cadastre-et-geomatique/modeles-de-donnees> (<https://www.fr.ch/territoire-amenagement-et-constructions/cartes-plans-cadastre-et-geomatique/modeles-de-donnees>))

sources text :

(<https://www.talend.com/fr/resources/guide-modelisation-donnees/#:~:text=La modélisation des>

données (<https://www.talend.com/fr/resources/guide-modelisation-donnees/#:~:text=La%20mod%C3%A9lisation%20des%20donn%C3%A9es%20>)

Différence entre données et informations

Les données ne sont PAS des informations !



Définition :

- **Données** : contiennent des chiffres, des énoncés et des caractères sous forme brute.
- **Information** : une nouvelle, un renseignement, une documentation sur quelque chose ou sur quelqu'un, portés à la connaissance de quelqu'un.

Exemples :

- **Données** :
 - Prénom
 - Nom
 - Age
 - Nationalité
 - Médicales
 - Scientifiques
 - Réseaux sociaux (likes, commentaires, partages)
- **Information** : Données traitées, organisées, structurées ou présentées dans un contexte donné afin de les rendre utiles -> "Il fait chaud aujourd'hui!"

source image :

(<https://waytolearnx.com/2018/08/difference-entre-donnee-et-information.html>)

sources text :

Données([https://fr.wikipedia.org/wiki/Donnée_\(informatique\)](https://fr.wikipedia.org/wiki/Donnée_(informatique))) ([https://fr.wikipedia.org/wiki/Donn%C3%A9e_\(informatique\)\)](https://fr.wikipedia.org/wiki/Donn%C3%A9e_(informatique))))

Stocks de données

Evitons la redondance !



CD,HDD,SSD,USB... !



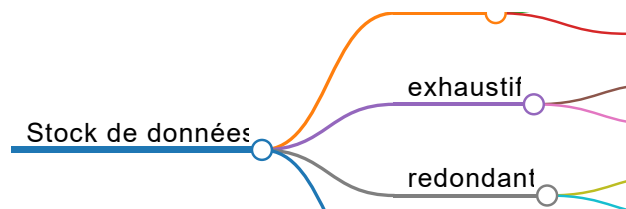
Ex. : Dropbox, Google Drive, et Amazon S3



Data center

Exemples :

- Stockage de médias (musique, films et photos numériques)
- Stockage de fichiers personnels (enregistrer un document word sur son pc pour qu'il reste le même jusqu'à la prochaine modification)
- Données de télémétrie (Les entreprises technologiques collectent des données de télémétrie à partir de leurs produits et services, telles que l'utilisation des applications, les performances des appareils, etc., pour améliorer leurs produits.)



Exemple visu. :

- exhaustif :

ID Produit	Nom du Produit	Catégorie	Marque	Poids (g)	Prix (EUR)	Date d'Expirati
1	Pommes	Fruits	Ferme X	150	1.99	2023- 09-30
2	Lait	Produits laitiers	Marque A	1000	0.99	2023- 10-15

- redondant :

ID	Nom	Adresse	Email	Téléphor
1	John Doe	123 Rue de la Ville	<u>john@example.com</u> (mailto:john@example.com)	555-123- 4567
2	Jane Smith	456 Avenue du Quartier	<u>jane@example.com</u> (mailto:jane@example.com)	555-987- 6543

Les données du client (Nom, Adresse, Email, Téléphone) sont stockées à la fois dans cette table et dans la table "Commandes", créant une duplication inutile.

CommandeID	ClientID	Date de Commande	Total
101	1	2023-09-15	50.00
102	2	2023-09-16	30.00

- incohérent :

ID	Nom	Date de Naissance
1	John Doe	05-15-1990
2	Jane Smith	1992-07-20
3	Alice Lee	12/04/1985
4	Bob White	1988-10-30

sources images :

(<https://quizizz.com/admin/quiz/5e3b6f663bd6ad001bd2679a/storage-media-questions-from-students-cie-ict-igcse-unit-3> (<https://quizizz.com/admin/quiz/5e3b6f663bd6ad001bd2679a/storage-media-questions-from-students-cie-ict-igcse-unit-3>))

<https://medium.com/meet-lima/whats-the-cloud-what-s-a-private-cloud-7fcdd8b535d1> (<https://medium.com/meet-lima/whats-the-cloud-what-s-a-private-cloud-7fcdd8b535d1>)

<https://home.cern/fr/news/news/computing/cern-data-centre-passes-100-petabytes>

(<https://home.cern/fr/news/news/computing/cern-data-centre-passes-100-petabytes>)

sources texte:

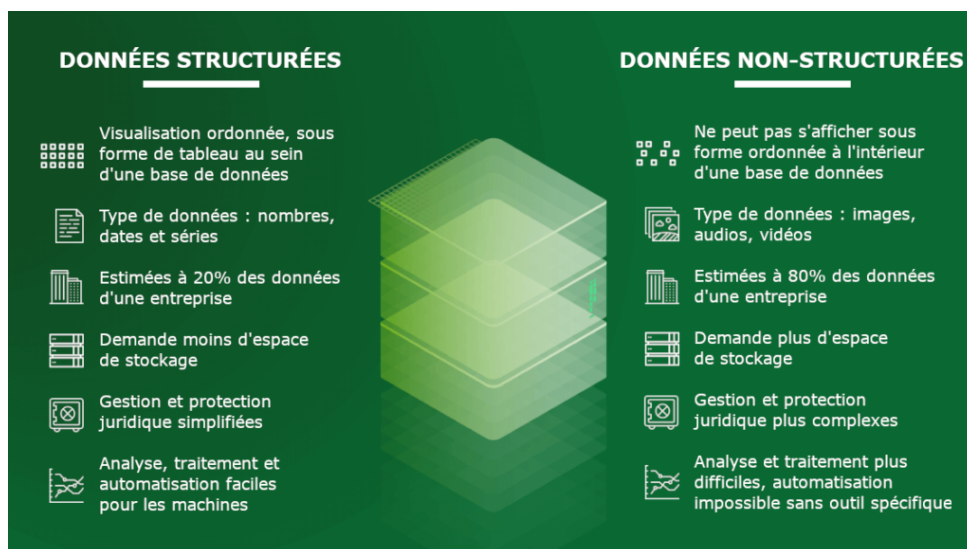
<https://fr.differbetween.com/article/what-is-the-difference-between-data-redundancy-and-data-inconsistency>

(<https://fr.differbetween.com/article/what-is-the-difference-between-data-redundancy-and-data-inconsistency>)

prompt ChatGPT:

"stock de données exhaustif exemple" "peux-tu me donner un exemple de stock de données redondant ?" "peux-tu me donner un exemple de stock de données incohérent ?" "quel est le meilleur exemple de stock de données exhaustif?"

Stocks de données structurés ou non structurés



Définition :

- **Données structurées:** Organisées de manière très formelle et tabulaire, elles sont généralement stockées dans des bases de données relationnelles ou des feuilles de calcul.
- **Données non structurées:** Sans format fixe ou de structure préalablement définie, elles peuvent prendre la forme de texte brut, de médias tels que des images

ou des vidéos.

Exemple :

- **structuré:**
 - carte d'identité
 - factures
 - tableaux de bord financiers
 - bases de données de ressources humaines
- **non structuré:**
 - signature sur la carte d'identité
 - Articles de blog
 - vidéos YouTube
 - images Instagram
 - enregistrements audio de conversations
 - courriels

source image :

(<https://www.group-edt.fr/les-differents-types-de-donnees/>)

Qu'est-ce qu'un format pour une donnée en informatique ?

L'extension d'un fichier n'est PAS le format d'un fichier

Différence : l'extension de fichier indique le type de fichier, tandis que le format de fichier spécifie la manière dont les données sont organisées et encodées à l'intérieur du fichier.



Définition : Le format de données = format de fichier, est une spécification qui définit la structure, l'encodage et l'organisation des données à l'intérieur du fichier. Il est

essentiel pour l'interopérabilité entre les logiciels et les systèmes, ainsi que pour garantir que les données sont lues et interprétées correctement.

Exemple :

- JPEG pour les images
- MP3 pour l'audio
- JSON pour l'échange de données
- CSV pour les données tabulaires

Il existe beaucoup de formats...

source image :

(<http://editorialisation.org/ediwiki/index.php/Accueil>

(<http://editorialisation.org/ediwiki/index.php/Accueil>)

Code ou codage d'une donnée.

Aspect	Code	Codage
Nature	Instructions	Transformation des informations
Language	Langage de programmation	Format ou représentation
But	Indiquer à l'ordinateur quoi faire	Rendre les données compréhensibles par l'ordinateur

Définition:

- **code:** mots-clés/ symboles/chiffres et structures logiques qui permettent à toute sorte de dispositifs électroniques de comprendre les tâches à accomplir et de les exécuter.
 - Dispositifs électroniques capables de comprendre/d'exécuter du code:
 - Ordinateurs personnels (PC et Mac)
 - Serveurs
 - Superordinateurs
 - Automates et dispositifs embarqués
 - Calculatrices

■ IoT (Internet des objets)

- **codage:** Permet de représenter l'information de manière à ce qu'elle puisse être traitée, stockée, transmise et interprétée par des systèmes informatiques.

Les données sont généralement stockées sous forme de séquences de bits (0 et 1) dans les systèmes informatiques.

Exemple:

- **Code:**

- Python :

```
# Ceci est un commentaire en Python
nom = "Alice"
print("Bonjour, " + nom + "!")
```

affiche "Bonjour, Alice!" à l'écran.

- Fortran (Formula Translation) :

```
program bonjour
  print *, "Bonjour, Monde!"
end program bonjour
```

- COBOL (Common Business-Oriented Language):

```
IDENTIFICATION DIVISION.
PROGRAM-ID. BonjourMonde.

PROCEDURE DIVISION.
    DISPLAY 'Bonjour, Monde!'.
```

- HTML :

```
<!DOCTYPE html>
<html>
<head>
  <title>Exemple de page web</title>
</head>
<body>
  <h1>Bienvenue sur ma page web</h1>
  <p>Ceci est un paragraphe de texte.</p>
</body>
</html>
```

- **Codage:**

- L'encodage des caractères textuels en binaire (comme UTF-8 pour les caractères Unicode)
- la compression d'images JPEG

- la conversion d'audio en MP3

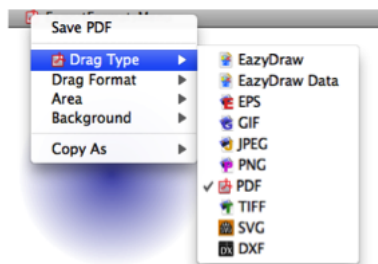
1	1	1	1	1	1	1	1	1	1
1	1	1	0	0	0	0	1	1	1
1	1	0	1	1	1	1	0	1	1
1	0	1	1	1	1	1	1	0	1
1	0	1	0	1	1	0	1	0	1
1	0	1	1	1	1	1	1	0	1
1	0	1	0	1	1	0	1	0	1
1	0	1	1	0	0	1	1	0	1
1	1	0	1	1	1	1	0	1	1
1	1	1	0	0	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1

Codage d'une image.

Source image :

5 (https://www.loutrel.fr/wikisn/doku.php?id=les_exposes:image_format_d_une_image_bmp)

Différents formats de données



Chaque format de données est adapté à des besoins particuliers en termes de stockage, de transmission et de traitement. Voici quelques-uns des formats de données les plus courants :

Catégorie	Format	Description
Documents	PDF	Portable Document Format, utilisé pour les documents avec mise en page fixe.
	DOCX	Format de fichier de document Microsoft Word.
	ODT	Open Document Text, format ouvert de traitement de texte.
Base de données	SQL	Structured Query Language, utilisé pour interagir avec les bases de données relationnelles.
	NoSQL	Divers formats de base de données non relationnelles, tels que MongoDB, Cassandra, etc.
	Shapefile	Format de données géographiques vectorielles largement utilisé.

Source image :

(<https://www.eazdraw.com/fileFormats.htm>

(<https://www.eazdraw.com/fileFormats.htm>)

Prompt chat gpt :

"C'est quoi exactement et techniquement le format des données ?""

Format de numéro téléphone

Pays/zone géographique	Indicatif de pays	Préfixe international	Préfixe national	Numéro national (significatif)	UTC/DST*	Note
Sao Tomé-et-Principe	239	00	...	7 chiffres	UTC	
Sénégal	221	00	...	9 chiffres	UTC	
Serbie	381	00	0	4 à 12 chiffres	+1/+2	
Service de coût partagé international (ISCS)	808	8 chiffres		
Service kiosque international (IPRS)	979	9 chiffres		
Service mobile mondial par satellite (GMSS), indicatif partagé	881					
Seychelles	248	00	...	7 chiffres	+4	
Sierra Leone	232	00	0	8 chiffres	UTC	
Singapour	65	001,008	...	8 à 12 chiffres	+8	9
Sint Maarten (partie néerlandaise)	1	011	1	(721) + 7 chiffres	-4	
Slovaquie	421	00	0	4 à 9 chiffres	+1/+2	
Slovénie	386	00	0	8 chiffres	+1/+2	
Somalie	252	00	...	5 à 8 chiffres	+3	
Soudan	249	00	0	9 chiffres	+3	
Soudan du Sud	211	00	0		+3	
Sri Lanka	94	00	0	9 chiffres	+5.30	
Sudafricaine (Rép.)	27	00	0	9 chiffres	+2	2
Suède	46	00	0	7 à 13 chiffres	+1/+2	
Suisse	41	00	0	4 à 12 chiffres	+1/+2	2

Format File signature :

Hexed.it (<http://Hexed.it>) (éditeur hexadécimal) :

"JPEG" :

Chat_big_head.jpg x	
00000000	FF D8 FF DB 00 43 00 04 03 03 04 03 03 04 04 03
00000010	04 05 04 04 05 06 0A 07 06 06 06 06 0D 09 0A 08
00000020	0A 0F 0D 10 10 0F 0D 0F 0E 11 13 18 14 11 12 17
00000030	12 0E 0F 15 1C 15 17 19 19 1B 1B 1B 10 14 1D 1F
00000040	1D 1A 1F 18 1A 1B 1A FF DB 00 43 01 04 05 05 06
00000050	05 06 0C 07 07 0C 1A 11 0F 11 1A 1A 1A 1A 1A 1A
00000060	1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A
00000070	1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A
00000080	1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A 1A FF C0 00 11

GCK'S FILE SIGNATURES TABLE:

FF D8 yÿ
 JPE, **JPEG**, JPG Generic **JPEG** Image file
 Trailer: FF D9 (yÿ)

Hexed.it (<http://Hexed.it>) (éditeur hexadécimal) :

"JPEG" :

La signature est une séquence spécifique d'octets située au début du fichier qui permet d'identifier son format et sa structure.

Une sorte d'étiquette d'identification essentielle qui permet aux logiciels de reconnaître et de traiter les fichiers de manière appropriée.

Source image :

(https://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-PDF-F.pdf) (https://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-PDF-F.pdf)

PDF-F.pdf)

(https://www.garykessler.net/library/file_sigs.html

(https://www.garykessler.net/library/file_sigs.html)

Prompt Chat GPT:

“peux tu me faire un résumé technique sur la signature des fichier”

Table ASCII (American Standard Code for Information Interchange)

ASCII TABLE											
Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[END OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Définition : La table ASCII est utilisée pour représenter du texte en utilisant des nombres binaires (bits) et est la base de nombreuses autres normes de codage de caractères, notamment UTF-8 et UTF-16, qui prennent en charge un ensemble beaucoup plus large de caractères, y compris les caractères non latins et les emojis.

Source image:

(<https://simple.m.wikipedia.org/wiki/File:ASCII-Table-wide.svg>)

(https://www.garykessler.net/library/file_sigs.html

(https://www.garykessler.net/library/file_sigs.html)

Prompt chat GPT: “C’est quoi une Table ASCII?”

Conversion

Binaire (1101 0000 1100)₂
 ↓ ↓ ↓ ↓
 Hexadécimal (D 0 C)₁₆

Hexadécimal (1 A F 3)₁₆
 ↓ ↓ ↓ ↓
 Binaire (0001 1010 1111 0011)₂

Base 10 (Décimal)

- **Définition** : Dans le système décimal, il y a dix chiffres différents disponibles, de 0 à 9. Système de numération le plus couramment utilisé dans le monde.

- **Conversions:**

1. Pour les nombres binaires : Additionnez les puissances de 2 en utilisant les chiffres binaires (0 et 1).
2. Pour les nombres hexadécimaux : Utilisez la valeur décimale équivalente de chaque chiffre hexadécimal (0-9, A-F).

Base 2 (Binaire)

- **Définition:** Dans le système binaire, il n'y a que deux chiffres différents disponibles, 0 et 1.

- **Conversions:**

1. Pour les nombres décimaux : Divisez successivement par 2, en notant les restes à chaque étape (de droite à gauche) pour obtenir la représentation binaire.
2. Pour les nombres hexadécimaux : Convertissez d'abord en base décimale, puis en base binaire.

Le système hexadécimal

- **Définition:** utilise seize chiffres différents, de 0 à 9 et de A à F (représentant les valeurs de 10 à 15 en décimal)

- **Conversions:**

1. Pour les nombres décimaux : Divisez successivement par 16, en notant les restes à chaque étape, puis lisez les restes de droite à gauche pour obtenir la représentation hexadécimale.

2. Pour les nombres binaires : Divisez les chiffres binaires en groupes de 4 (en ajoutant des zéros à gauche si nécessaire) et convertissez chaque groupe en sa valeur hexadécimale équivalente.

il existe plusieurs systèmes de numération différents, mais les systèmes les plus utilisés sont le décimal (base 10), le binaire (base 2) et l'hexadécimal (base 16).

Source image : <https://www.positron-libre.com/cours/electronique/systeme-numeration/conversion-decimal-binaire-hexadecimal.php> (<https://www.positron-libre.com/cours/electronique/systeme-numeration/conversion-decimal-binaire-hexadecimal.php>)

Source text : <https://www.positron-libre.com/cours/electronique/systeme-numeration/conversion-decimal-binaire-hexadecimal.php> (<https://www.positron-libre.com/cours/electronique/systeme-numeration/conversion-decimal-binaire-hexadecimal.php>)

Prompt Chat GPT : "Que peux-tu me dire à propos de la conversion en base décimale, binaire et hexadécimale? C'est pour un glossaire/aide mémoire technique."

Format des fichiers images, sons, vidéos, textes, etc

Le choix de la base numérique définit le format des données, y compris la structure, le codage et les conventions de stockage

Définition : le choix de la base numérique définit le format des données, y compris la structure, le codage et les conventions de stockage

Source : Nom du site ou de la source₀

 Image 3

Distinction entre format et nom d'extension de fichier

Définition : Une brève explication du mot 3.

Source : Nom du site ou de la source₀

 Image 3

Ouverture de différents fichiers avec un éditeur hexadécimal (<https://hexed.it/>), montrer leurs signatures des formats de fichier (https://www.garykessler.net/library/file_sigs.html)

Définition : Une brève explication du mot 1.

Source : Nom du site ou de la source₀

 Image 1

Changer le titre d'un album, le nom de l'artiste dans un fichier .mp3 à l'aide d'un éditeur hexadécimal. (<https://en.wikipedia.org/wiki/ID3>)

Définition : Une brève explication du mot 2.

Source : Nom du site ou de la source₀

 Image 2

