

# Cours 2 : La donnée

# Introduction : Retour sur vos questions

Y'a t-il des notions que vous avez abordé en entreprise ou en cours qui vous semblent encore vagues ?

# Introduction : Retour sur vos questions

Notion de Build / Run  
Notion d'application / logiciel

“Comment est ce qu'on crée un programme informatique ?”

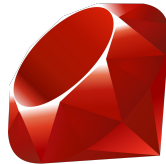
# Créer un programme informatique

IDE  
“Integrated Development  
Environment”

Langage de Programmation

# Créer un programme informatique

Langage de Programmation









Swift

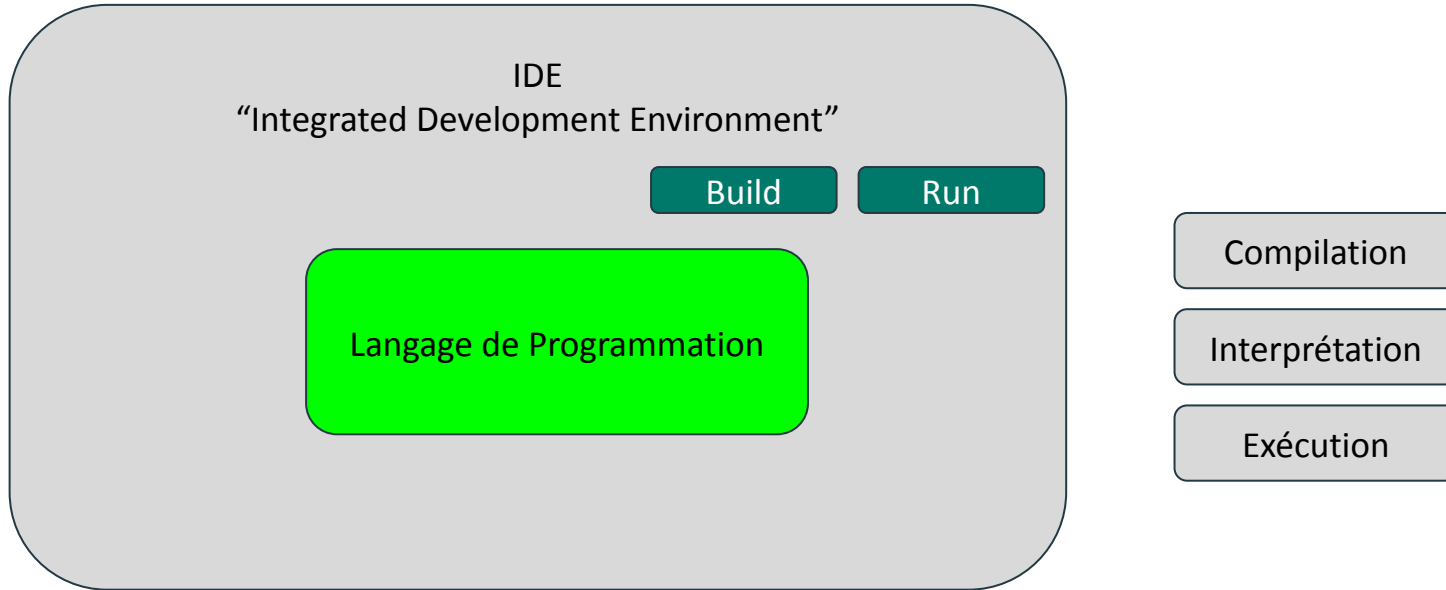


<https://www.scriptol.fr/programmation/liste-langages-programmation.php>

# Les langages de programmation - le top 10

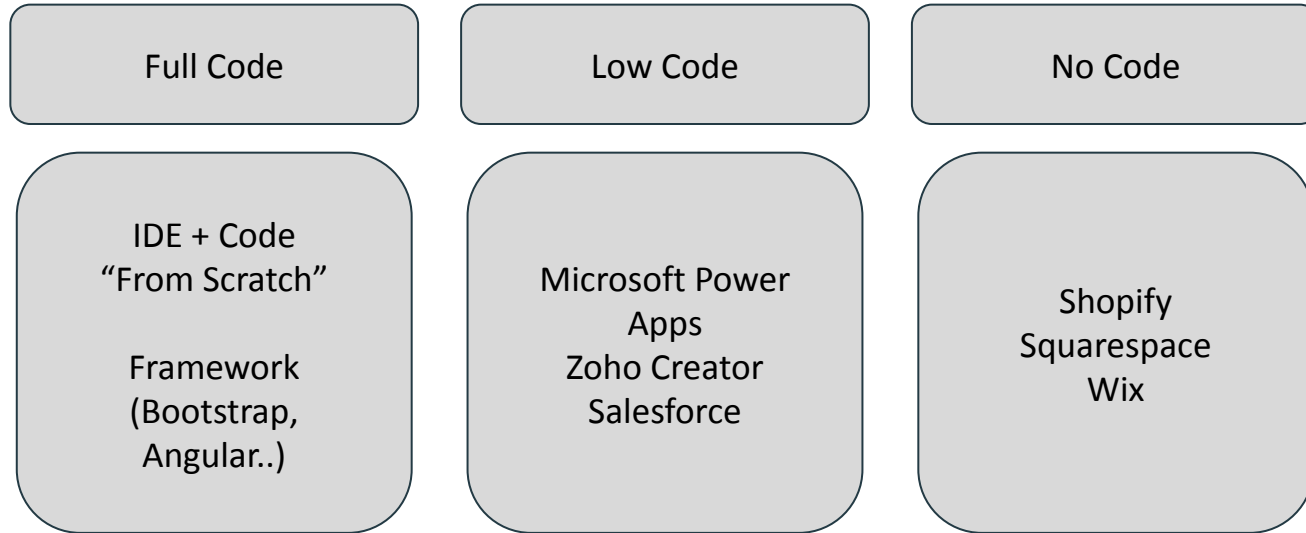
Mar 2023	Mar 2022	Change	Programming Language		Ratings	Change
1	1			Python	14.83%	+0.57%
2	2			C	14.73%	+1.67%
3	3			Java	13.56%	+2.37%
4	4			C++	13.29%	+4.64%
5	5			C#	7.17%	+1.25%
6	6			Visual Basic	4.75%	-1.01%
7	7			JavaScript	2.17%	+0.09%
8	10	▲		SQL	1.95%	+0.11%
9	8	▼		PHP	1.61%	-0.30%
10	13	▲		Go	1.24%	+0.26%

# Créer un programme informatique



# Créer un programme informatique

## Les modes de programmation





# Rappel : Les processus liés aux données

Collecte et gestion des données

"Data"

**Catégorisation**

Données  
Industrielles

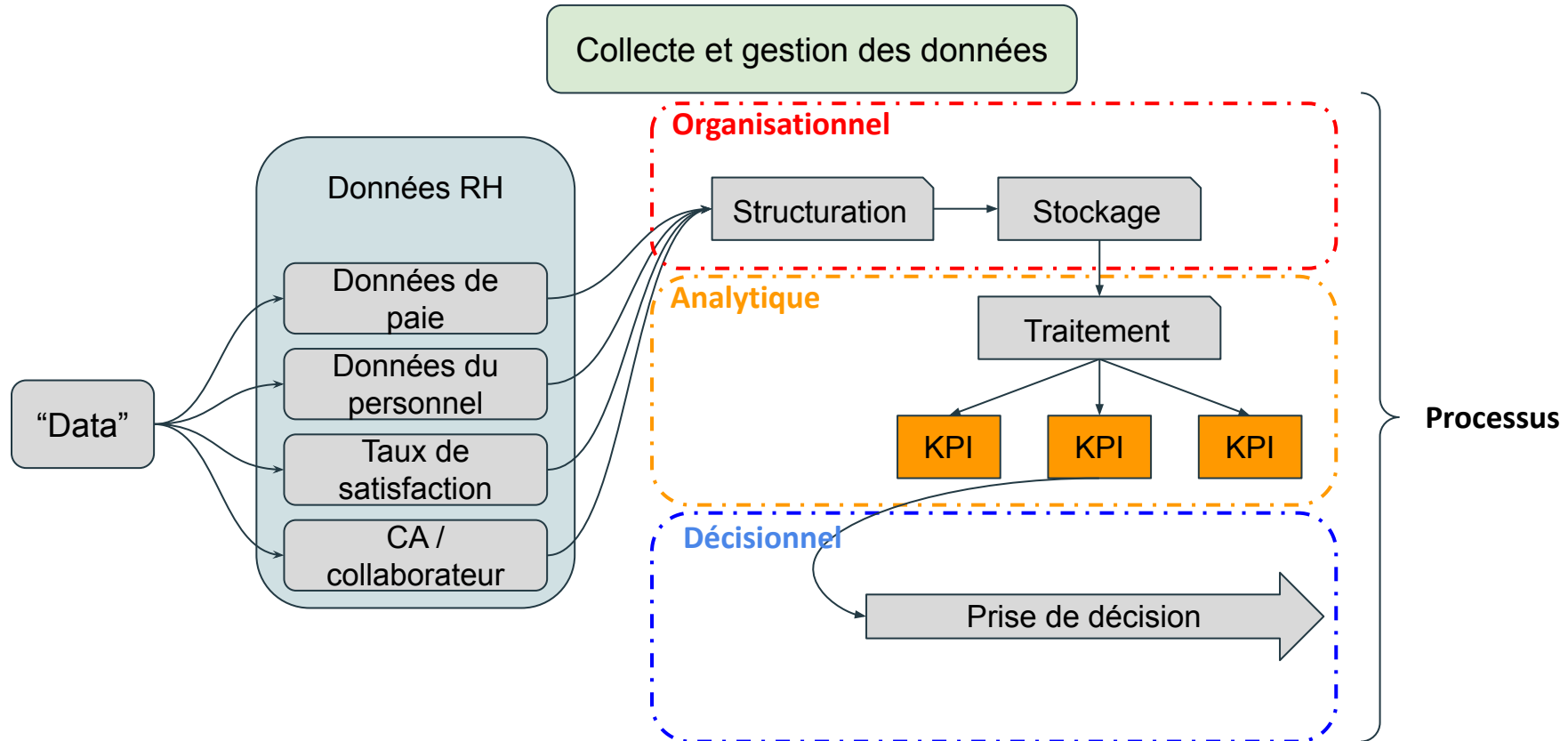
Données RH

Données  
Commerciales

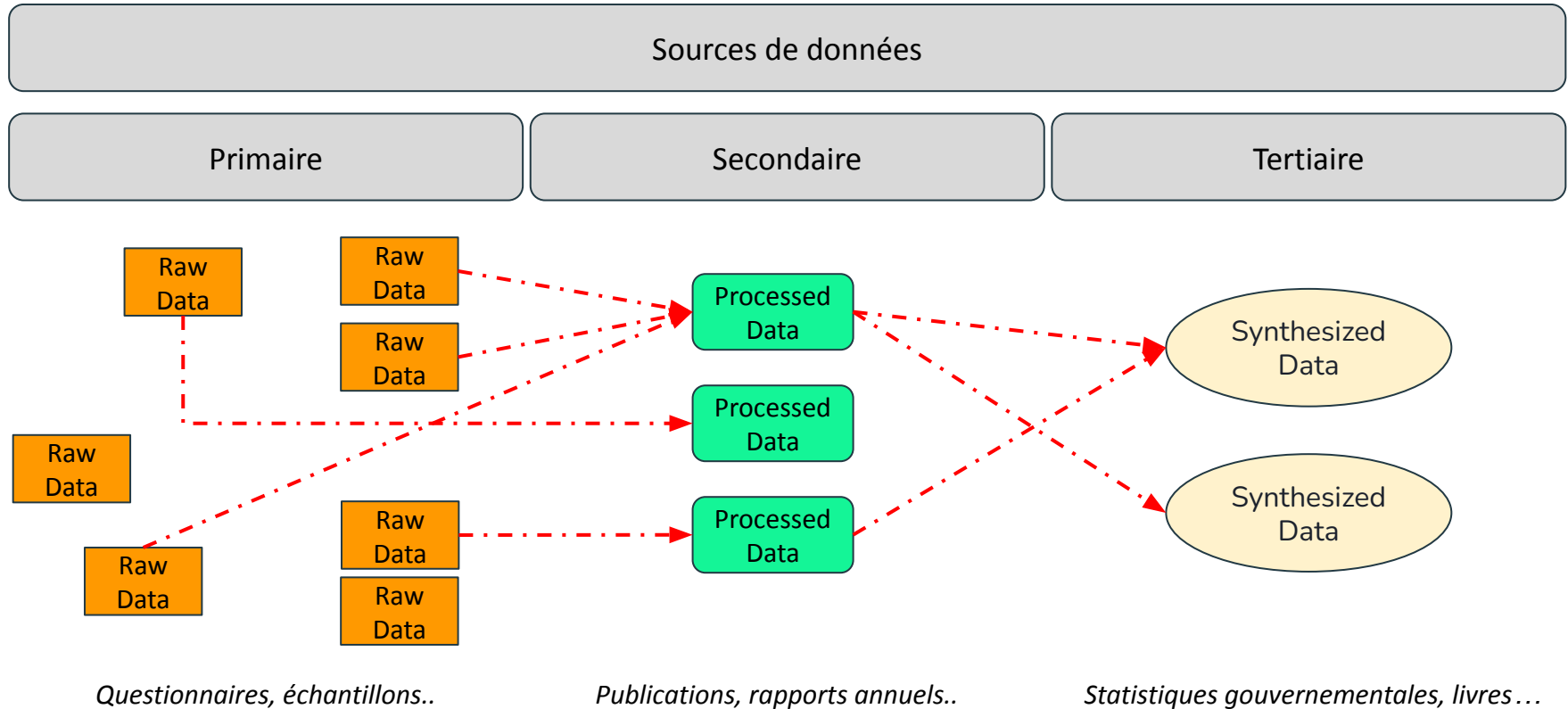
Données  
Financières

Et plus encore...

# Rappel : Les processus liés aux données



# Les sources de données et leur transformation



# La gouvernance des données

La gouvernance des données est un ensemble de **processus**, de **politiques**, de **normes** et de **pratiques** qui régissent la **gestion**, la **qualité**, l'**intégrité**, la **sécurité**, la **confidentialité** et l'**utilisation des données** au sein d'une organisation.

L'objectif principal de la gouvernance des données est de garantir que les données de l'organisation sont gérées de manière efficace, cohérente et conforme aux exigences légales et réglementaires.



# Gouvernance des données

## Définition des responsabilités

La gouvernance des données implique l'**attribution de responsabilités claires** concernant la gestion des données. Cela peut inclure la désignation de responsables de la **gestion des données**, de **responsables de la qualité des données** et d'autres **parties prenantes** chargées de veiller à ce que les données soient gérées de manière appropriée.

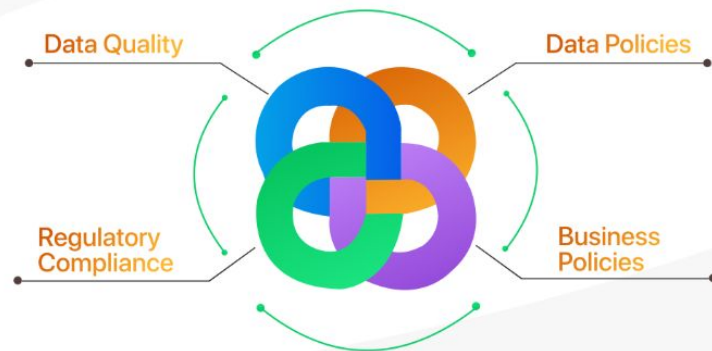
Formaliser les responsabilités de  
chacun face à la donnée

Chaque personne étant en surcharge,  
il était difficile de définir un périmètre  
sur le long terme.

# Gouvernance des données : Politiques et normes

La mise en place de politiques et de **normes spécifiques** pour la **collecte**, le **stockage**, la **classification**, la **diffusion**, la **rétenion** et la **destruction des données**. Ces politiques visent à garantir que les données sont utilisées de manière appropriée et sécurisée.

## Data Governance



# Gouvernance des données : Qualité des données

La gouvernance des données s'attache à maintenir la qualité des données en mettant en place des processus de **contrôle de la qualité**, des **normes de saisie de données**, des **mécanismes de nettoyage des données** et des **procédures de correction des erreurs**.

Instaurer des processus de contrôle et  
de blindage

La conduite du changement



# Sécurité des données

La protection des données sensibles et confidentielles est essentielle. Cela comprend la mise en place de mécanismes de sécurité informatique, de chiffrement des données, de gestion des accès et d'audit pour garantir la confidentialité et l'intégrité des données.

# Conformité légale et réglementaire

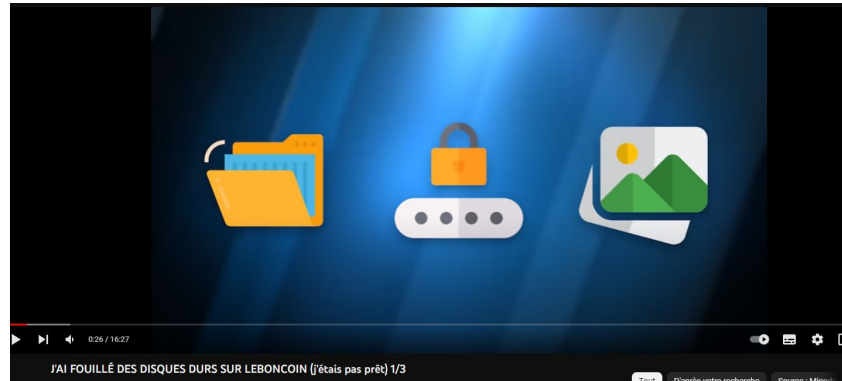
La gouvernance des données s'assure que l'organisation se conforme aux **lois et réglementations** en matière de gestion des données, notamment en ce qui concerne la **protection des données personnelles**, la **confidentialité** et d'autres exigences sectorielles.

DÉFENSE ET SÉCURITÉ NATIONALE	
<b>SECRET</b>	Dénomination de l'entité émettrice
	<b>SPÉCIAL FRANCE</b>
Service émetteur	Lieu, date
	N°
	Déclassifié le :
Objet	: Intitulé



# Gestion du cycle de vie des données

Cela englobe la **gestion de la création**, de **l'utilisation**, de la **modification**, de la **conservation** et de la **suppression des données** tout au long de leur cycle de vie. Il s'agit de s'assurer que les données sont utilisées de manière appropriée et qu'elles sont éliminées de manière sécurisée lorsque cela est nécessaire.



<https://www.youtube.com/watch?v=vt8PyQ2PGxI>

“J’ai fouillé des disques durs sur leboncoin” Micode

# Communication et sensibilisation

La gouvernance des données implique souvent de sensibiliser les employés de l'organisation à l'importance de la gestion des données et de promouvoir les bonnes pratiques en matière de données.

# Technologie et outils

L'utilisation d'outils de gestion des données, tels que les systèmes de **gestion des données**, les **solutions de catalogage des données** et les **outils d'analyse des données**, pour faciliter la gestion efficace des données.

# Les normes européennes et françaises en matière de gestion de données

Que connaissez vous comme normes existantes aujourd'hui ?

# Les normes européennes : Le RGPD

**Règlement général sur la protection des données (RGPD)** : Le RGPD est le principal texte de référence en matière de protection des données personnelles en Europe. Il a été adopté en **2018** et s'applique à toutes les entreprises qui traitent des **données personnelles de résidents européens**. Il définit les principes fondamentaux de **protection des données**, les **droits des individus** et les **obligations des responsables du traitement des données**.

<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

# Les directives NIS

**Directive NIS** : La directive sur la sécurité des réseaux et des systèmes d'information (NIS) vise à renforcer la cybersécurité au sein de l'Union européenne en imposant des obligations de sécurité aux opérateurs de services essentiels et aux fournisseurs de services numériques.

## PRÉSENTATION DE LA DIRECTIVE

Structurée autour de quatre axes, la directive prévoit

- le renforcement des capacités nationales de cybersécurité. Les Etats membres devront notamment se doter d'autorités nationales compétentes en matière de cybersécurité, d'équipes nationales de réponse aux incidents informatiques (CSIRT) et de stratégies nationales de cybersécurité. Respectivement en France, l'ANSSI, le CERT-FR<sup>2</sup> et la stratégie nationale pour la sécurité du numérique ;
- l'établissement d'un cadre de coopération volontaire entre Etats membres de l'UE via la création de
  - un « groupe de coopération » des Etats membres sur les aspects politiques de la cybersécurité ;
  - un « réseau européen des CSIRT » des Etats membres. Ce dernier visera notamment à faciliter le partage d'informations techniques sur les risques, vulnérabilités ;
- le renforcement par chaque Etat de la cybersécurité d'« opérateurs de services essentiels » au fonctionnement de l'économie et de la société via
  - la définition au niveau national de règles de cybersécurité auxquels ces derniers devront se conformer ;
  - l'obligation pour les opérateurs de notifier les incidents ayant un impact sur la continuité de leurs services essentiels.
- l'instauration de règles européennes communes en matière de cybersécurité des prestataires de services numériques dans les domaines de l'informatique en nuage, des moteurs de recherche et places de marché en ligne.



# Le Data Governance Act

Le **Data Governance Act** et le **Data Act** s'inscrivent dans le cadre de la **stratégie européenne pour les données**, présentée par la Commission européenne en février 2020. Cette stratégie vise à développer un marché unique de la donnée en soutenant **l'accès**, le **partage** et la **réutilisation** responsables, dans le respect des valeurs de l'Union européenne et notamment la **protection des données personnelles**.

Elle s'inscrit dans le contexte plus large du plan d'action de la Commission européenne visant à assurer la **souveraineté numérique** de l'Europe à l'horizon **2030**, et est complémentaire de la stratégie européenne en matière d'intelligence artificielle.

# Le Data Governance Act

Première brique de la série de mesures annoncées dans le cadre de la stratégie européenne des données, le Data Governance Act a été adopté en **mai 2022**, et sera applicable en **septembre 2023**. Il vise à favoriser le partage des données personnelles et non personnelles en mettant en place des structures d'intermédiation. Ce règlement comporte 4 grandes mesures.

# Le Data Governance Act

Un encadrement ainsi qu'une assistance technique et juridique facilitant la **réutilisation** de certaines catégories de données protégées du secteur public (informations commerciales confidentielles, propriété intellectuelle, données personnelles)

Une certification **obligatoire** pour les fournisseurs de services d'intermédiation de données.

Une certification **facultative** pour les organismes pratiquant l'altruisme en matière de données.

La création du **"Conseil Européen de l'innovation en matière de données"**

# Le Data Act (Le règlement sur les données)

La proposition législative de la Commission européenne, présentée le **23 février 2022**, a pour objectif d'assurer une meilleure répartition de la valeur issue de l'utilisation des données personnelles et non personnelles entre les acteurs de l'économie de la donnée, notamment liées à l'utilisation des objets connectés et au développement de l'Internet des objets.

# Le Data Act

Faciliter le partage entre entreprises (B2B) et avec le consommateur (B2C) des données, en fixant notamment une obligation de rendre accessibles les données générées par l'utilisation des objets connectés et services connexes, en contrepartie d'une compensation juste et équitable.

Permettre l'utilisation des données détenues par les entreprises et, sous réserve de justifier d'un besoin exceptionnel, par les organismes publics des États membres et les institutions, agences ou organes de l'Union.

Faciliter le changement de fournisseur de services de traitement de données (cloud et edge computing) par l'encadrement des relations contractuelles entre les fournisseurs de services et les consommateurs, et notamment par la suppression progressive des frais liés au changement pour le consommateur.

Prévoir l'élaboration de normes d'interopérabilité pour les données et leurs réutilisations entre les secteurs.

Mettre en place des garanties contre les accès illicites de gouvernements de pays tiers aux données non-personnelles contenues dans le cloud.

# Les normes françaises : La loi informatique et libertés

La France a mis en place sa propre législation nationale sur la protection des données bien avant l'adoption du RGPD. La Loi Informatique et Libertés définit les règles applicables au traitement des données personnelles en France.

<https://www.cnil.fr/fr/la-loi-informatique-et-libertes>

# Les normes ISO

De nombreuses organisations en France se conforment également aux normes ISO en matière de gestion de la sécurité de l'information, notamment ISO 27001, qui définit les exigences pour établir, mettre en œuvre, gérer et améliorer un système de gestion de la sécurité de l'information.

[La norme ISO 27001](#)

# Les normes de l'ANSSI

**Agence nationale de la sécurité des systèmes d'information (ANSSI)** : L'ANSSI est l'autorité française chargée de la sécurité des systèmes d'information. Elle émet des recommandations et des bonnes pratiques en matière de cybersécurité pour les organisations publiques et privées.





# La lutte contre la cybercriminalité

La France a également adopté des lois pour lutter contre la cybercriminalité et garantir la sécurité des données, notamment en renforçant les pouvoirs d'enquête des autorités compétentes.

# Introduction au Big Data

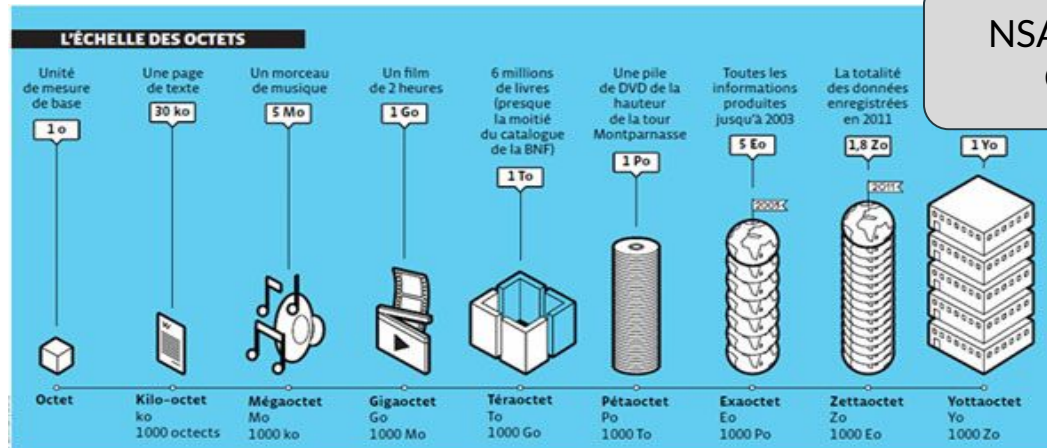
Le Big Data, littéralement "gros volume de données", est un terme qui fait référence à la gestion et à l'analyse de vastes ensembles de données qui dépassent les capacités des outils et des méthodes traditionnels de traitement de données.

# Introduction

La capacité d'un ordinateur "classique" est de quelques TéraOctets (To) au mieux.

Besoin de stockage dans le Big Data d' exaoctets (Eo)

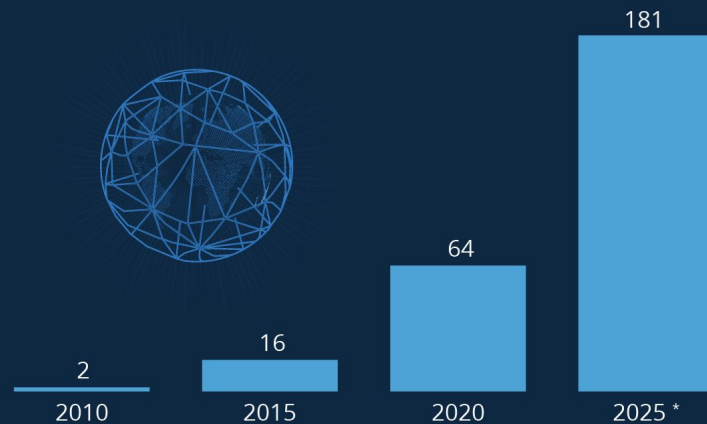
1 Exaoctet = 1024 Pétaoctets (Po)



NSA Utah Data Center ?

## Le Big Bang du Big Data

Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets



Un zettaoctet équivaut à mille milliards de gigaoctets.

\* Prévision en date de mars 2021.

Sources : IDC, Seagate, Statista



# Le Data Center de la NSA dans l'Utah



# Qu'est ce que le Big Data ?

Le Big Data se caractérise par les 3 "V" :

**Volume** : Les données générées aujourd'hui sont massives. Elles proviennent de sources diverses telles que les médias sociaux, les appareils connectés, les capteurs, les transactions en ligne, etc. Les volumes de données sont si importants qu'ils dépassent souvent la capacité de stockage et de traitement des systèmes traditionnels.

# Qu'est ce que le Big Data ?

**Vélocité** : Les données sont générées à une vitesse étonnante. Les médias sociaux, par exemple, produisent constamment de nouvelles informations. Pour être utiles, les données doivent être traitées en temps réel ou presque en temps réel.

# Qu'est ce que le Big Data ?

**Variété :** Les données ne se limitent pas aux simples chiffres. Elles peuvent prendre de nombreuses formes, notamment des textes, des images, des vidéos, des enregistrements audio, des données géospatiales, etc. Cette variété complexifie leur gestion et leur analyse.



# Applications du Big Data

Le Big Data trouve des applications dans de nombreux secteurs, notamment :

**Santé** : Il peut être utilisé pour la recherche médicale, la détection précoce de maladies et la personnalisation des traitements.

**Marketing et publicité** : Les entreprises l'utilisent pour cibler leurs campagnes publicitaires en fonction du comportement des consommateurs.

**Transport** : Il peut améliorer la gestion du trafic, la maintenance préventive des véhicules, et faciliter le développement de véhicules autonomes.

**Finance** : Les banques et les institutions financières l'utilisent pour détecter la fraude, gérer les risques et prendre des décisions d'investissement.