

3.1.3.2. Variables et instructions de base en Python

L'interpréteur reçoit et exécute les instructions de manière interactive.

L'interpréteur agit comme une simple calculatrice. Vous pouvez y taper une expression et il écrira la valeur. La syntaxe des expressions est simple. Les opérateurs +, -, * et / fonctionnent comme dans la plupart des autres langages (par exemple, Pascal ou C). Les parenthèses (()) peuvent être utilisées pour le regroupement.

Le mode interactif de Python implémente la variable spéciale "_" pour contenir le résultat de la dernière expression émise.

Les variables sont des zones de mémoire étiquetées qui sont utilisées pour stocker les données du programme d'exécution. Pour attribuer des valeurs aux variables en Python, utilisez le signe égal à (=). Aucun résultat n'est affiché avant l'invite interactive suivante.

Les tentatives d'utilisation d'une variable non définie (sans valeur assignée) entraîneront une erreur. Les chaînes de caractères, définies comme une séquence de caractères, peuvent également être traitées par le mode interactif. Utilisez le caractère backslash (\) pour échapper aux caractères. Par exemple, une chaîne de caractères utilise des guillemets doubles mais doit également utiliser un guillemet double à l'intérieur de la chaîne. Si la chaîne est saisie comme suit : **"J'ai vraiment "besoin" de ça."** Python sera confus et pensera que le premier guillemet double dans la chaîne termine réellement la chaîne. Si vous placez une barre oblique inverse (\) avant les guillemets doubles à l'intérieur de la chaîne, comme suit : **"J'ai vraiment "besoin" de ça"**, la barre oblique inverse (\) fait que Python échappe ou ignore le caractère qui suit.

Les guillemets simples ou doubles peuvent être utilisés pour envelopper les chaînes de caractères.

L'instruction print imprime le résultat de l'expression qui lui a été donnée. Elle diffère de la simple écriture de l'expression que vous souhaitez écrire (comme nous l'avons fait précédemment dans les exemples de calculatrices) par la manière dont elle gère les expressions et chaînes de caractères multiples. Les chaînes de caractères sont imprimées sans guillemets et un espace est inséré entre les éléments, ce qui vous permet de formater les choses de manière a

```
# Function to add two numbers:
def add_nums():
    a = 5
    b = 11
    return a+b
>>> print (add_nums())
16
>>>
```

Les fonctions sont un élément important de nombreux langages de programmation. Les fonctions permettent de donner un nom à un bloc de code et de le réutiliser si nécessaire. La figure définit une fonction permettant d'additionner deux nombres et d'imprimer le résultat.

3.1.3.3. Fonctions et types de données utiles en python

Python prend en charge de nombreuses fonctions et types de données utiles. Voici quelques-unes des plus importantes :

Gamme()

La fonction `range()` génère une liste de nombres généralement utilisée pour itérer avec les boucles FOR.

- **`range(stop)`** - Il s'agit du nombre d'entiers (nombres entiers) à générer, en partant de zéro.
- **`range([start], stop, [step])`** - Il s'agit du numéro de départ de la séquence, du numéro d'arrivée de la séquence et de la différence entre chaque numéro de la séquence.

Tuples

Un tuple est une séquence d'objets Python immuables. Les tuples sont des séquences, séparées par des parenthèses.

Listes

Les listes sont une séquence d'objets Python modifiables. Les listes peuvent être créées en plaçant différentes valeurs séparées par des virgules entre des crochets.

Sets

Les ensembles sont des collections non ordonnées d'éléments uniques. Les utilisations courantes comprennent les tests d'appartenance, la suppression des doublons d'une séquence et le calcul d'opérations mathématiques standard sur les ensembles, telles que l'intersection, l'union, la différence et la différence symétrique.

Dictionnaire

Un dictionnaire est une liste d'éléments séparés par des virgules. Chaque élément est une combinaison d'une valeur et d'une clé unique. Chaque clé est séparée de sa valeur par deux points. L'ensemble du dictionnaire est écrit entre accolades. Les éléments du dictionnaire peuvent être consultés, mis à jour et supprimés. Il existe également de nombreuses fonctions de dictionnaire intégrées, telles qu'une fonction qui compare les éléments de différents dictionnaires et une autre qui fournit un compte du nombre total d'éléments dans un dictionnaire.

3.1.3.4. Structure de programmation en Python

Comme d'autres langages, Python implémente une structure IF-THEN. Les blocs IF-THEN peuvent être utilisés pour permettre au code de prendre des décisions en fonction du résultat d'une expression.

Le code effectue quelques tests et imprime un message en fonction des résultats du test. Remarquez que Python implémente également deux sous-structures nommées ELSE et ELIF. ELSE permet au programmeur de spécifier les instructions à exécuter si l'expression est fausse. Abréviations de ELSE IF, ELIF est utilisé pour effectuer un deuxième test au cas où la première expression est fausse et qu'un autre test est nécessaire. Il peut y avoir zéro ou plusieurs ELIF, et la partie ELSE est facultative.

```
>>> x = int(input("Please enter an integer: "))
Please enter an integer: 42
>>> if x < 0:
...     x = 0
...     print ('Negative changed to zero')
... elif x == 0:
...     print ('Zero')
... elif x == 1:
...     print ('Single')
... else:
...     print ('More')
...
More
```

La boucle FOR en Python itère les éléments d'une séquence (une liste ou une chaîne), dans l'ordre où ils apparaissent dans la séquence.

```
>>> # Measure some strings:
... words = ['cat', 'window', 'defenestrate']
>>> for w in words:
...     print (w, len(w))
...
cat 3
window 6
defenestrate 12
```

La boucle WHILE exécute un bloc de code si l'expression est vraie. Le programme illustré dans la figure utilise une boucle WHILE pour calculer et imprimer une sous-séquence initiale d'une série de Fibonacci dans laquelle chaque nombre de la série est la somme des deux précédents.

La troisième ligne contient un opérateur d'affectation multiple. Les variables **a** et **b** obtiennent les nouvelles valeurs 0 et 1 en une seule instruction.

La boucle WHILE calcule le terme suivant de la série de Fibonacci tant que la condition $b < 10$ est vraie. Comme en C, Python considère que toute valeur entière non nulle est vraie et que zéro est faux. Le test utilisé dans la figure est une simple comparaison.

```
>>> # Fibonacci series:
... # the sum of two elements defines the next
... a, b = 0, 1
>>> while b < 10:
...     print (b)
...     a, b = b, a+b
...
1
1
2
3
5
8
```

Remarquez que le corps de la boucle est indenté. L'indentation est le moyen utilisé par Python pour regrouper les instructions. À l'invite interactive, vous devez taper une tabulation ou un ou plusieurs espaces pour chaque ligne indentée. Les saisies plus complexes pour Python doivent être effectuées avec un éditeur de texte. Lorsqu'une instruction composée est saisie de manière interactive, elle doit être suivie d'une ligne vide pour indiquer qu'elle est terminée (car l'analyseur syntaxique ne peut pas deviner quelle sera la dernière ligne). Notez que chaque ligne d'un bloc de base doit être indentée de la même manière.

3.2. Prototypage de votre idée

3.2.1. Qu'est-ce que le prototypage ?

3.2.1.1. Définition du prototypage

Le prototypage est le processus qui consiste à créer un modèle fonctionnel rudimentaire d'un produit ou d'un système. Pour le prototypage dans l'IdO, il est utile d'avoir des compétences en conception, en électricité, en physique/mécanique (travailler avec ses mains pour assembler des objets), en programmation et de comprendre le fonctionnement de TCP/IP. Mais il n'est pas nécessaire d'être un expert dans l'un de ces domaines. En fait, le prototypage vous aide à affiner ces compétences.

L'IdO étant encore en développement, il reste des tâches inconnues à découvrir. C'est le moment idéal pour inventer quelque chose qui fasse partie de l'IdO. Parce que l'IdO combine les personnes, les processus, les données et les objets, les inventions que l'IdO peut aider à créer et à intégrer sont illimitées.

Prototyping

- Is fully functional, but not fault-proof.
- Is an actual, working version of the product.
- Is used for performance evaluation and further improvement of product.
- Has a complete interior and exterior.
- May be relatively expensive to produce.
- In the IoT, is often used as a technology demonstrator.

3.2.1.2. Comment faire un prototype ?

Comment faire un prototype ? Il existe plusieurs façons de commencer. Une équipe de Google a utilisé la "méthode de prototypage rapide" pour créer les Google Glass.

<http://ed.ted.com/lessons/rapid-prototyping-google-glass-tom-chi> pour voir un TedTalk sur ce processus.

Bien sûr, Google dispose d'un grand nombre de ressources pour payer les personnes et les matériaux nécessaires au prototypage. La plupart d'entre nous ont besoin d'une aide financière pour faire sortir leurs idées de leur tête et les transformer en prototype. Pour nous, il y a le financement par la foule. Kickstarter, Indiegogo et Crowdfunder ne sont que trois des nombreux programmes de financement par la foule en ligne.

Quelle invention IoT allez-vous créer ?



3.2.2. Ressources pour le prototypage

3.2.2.1. Matériaux physiques

Un bon point de départ est, bien sûr, l'Internet. Les gens échangent des idées depuis des lustres, mais l'Internet permet des échanges d'idées à un tout autre niveau. Des personnes qui ne se sont jamais rencontrées physiquement peuvent désormais collaborer et travailler ensemble. Il existe plusieurs sites web que vous pouvez visiter pour vous connecter avec d'autres fabricants.

Maker Media est une plateforme mondiale permettant de mettre en relation des créateurs pour échanger des projets et des idées. La plateforme offre également un lieu où les fabricants peuvent trouver et acheter des produits pour leurs projets. Pour plus d'informations, rendez-vous sur Makezine à l'adresse <http://makezine.com>.

Il est utile d'avoir des compétences pratiques pour travailler avec certains matériaux ; par exemple, le bois et le métal sont des matériaux de prototypage courants, mais ils peuvent être trop difficiles à utiliser pour un débutant. Vous pourriez être surpris de ce que vous pouvez faire avec du plastique, de l'argile, du papier et des fils. Cherchez sur Google de plus amples informations ou des idées sur la façon de travailler avec les différents matériaux de prototypage.

LEGO Mindstorms a une grande communauté de contributeurs et de fans. Avec LEGO Mindstorms, vous pouvez créer des robots LEGO et les contrôler à l'aide d'une application. Les kits sont livrés avec tout ce dont vous avez besoin pour les faire fonctionner. Rendez-vous sur LEGO Mindstorms à l'adresse <http://mindstorms.lego.com>.

Le Meccano, ou Erector Set, est un système de construction de modèles réduits constitué de bandes métalliques réutilisables, de plaques, de poutres d'angle, de roues, d'essieux et d'engrenages, avec des écrous et des boulons pour relier les pièces. Il vous permet de construire des prototypes et des dispositifs mécaniques fonctionnels. Consultez le site Erector Set à l'[adresse](http://www.erector.us) www.erector.us.

L'impression 3D est le processus de fabrication d'un objet solide à partir d'un fichier informatique de modèle 3D. Une machine, appelée imprimante 3D, est connectée à l'ordinateur. Un certain nombre d'entreprises fabriquent et vendent désormais des imprimantes 3D. Rendez-vous sur le site de Makerbot à l'adresse <https://www.makerbot.com>.



3.2.2.2. Boîtes à outils électroniques

Les programmes informatiques ne peuvent pas fonctionner sans ordinateur. Si vous pouvez créer des programmes pour presque tous les ordinateurs, certaines plates-formes sont conçues pour les débutants. Vous trouverez ci-dessous quelques-unes des plateformes les plus populaires.

Arduino est une plate-forme informatique physique à code source ouvert basée sur une carte microcontrôleur simple et un environnement de développement permettant d'écrire des logiciels pour la carte. Vous pouvez développer des objets interactifs qui prennent en compte les entrées d'une variété d'interrupteurs ou de capteurs pour contrôler des lumières, des moteurs et d'autres objets physiques. Consultez le site Arduino à l'adresse <http://arduino.cc>.

Bien que l'Arduino ne puisse pas être utilisé comme un ordinateur, sa faible consommation d'énergie le rend capable de contrôler efficacement d'autres appareils.

Le Raspberry Pi est un ordinateur bon marché, de la taille d'une carte de crédit, qui se branche sur un écran d'ordinateur ou un téléviseur. Vous le faites fonctionner à l'aide d'un clavier et d'une souris standard. Il est capable de faire tout ce qu'un ordinateur peut faire, de la navigation sur Internet à la lecture de vidéos haute définition, en passant par la création de feuilles de calcul, le traitement de texte et les jeux. Consultez le site Raspberry Pi à l'[adresse](http://www.raspberrypi.org) <http://www.raspberrypi.org>.

Le Beaglebone est très similaire au Raspberry Pi en termes de taille, de besoins en énergie et d'applications. Le Beaglebone a une puissance de traitement supérieure à celle du Raspberry Pi ; il constitue donc un meilleur choix pour les applications nécessitant un traitement plus important. Consultez le site Beaglebone à l'adresse <http://beagleboard.org>.

3.2.2.3. Ressources de programmation

La programmation est essentielle pour l'IdO. La création de code personnalisé est très utile lors du développement d'une solution IoT. Vous avez déjà appris à connaître Blockly et Python. Il existe de nombreuses autres ressources gratuites qui peuvent vous aider à développer vos compétences en programmation.

L'OpenCourseWare (OCW) du MIT est une publication en ligne de la quasi-totalité du contenu des cours du MIT. Ouvert et disponible au monde entier, l'OCW est un endroit idéal pour se familiariser gratuitement avec la programmation informatique. Les cours OCW liés à la programmation peuvent être trouvés sur <http://ocw.mit.edu/courses/intro-programming>.

Khan Academy est un site web éducatif à but non lucratif créé en 2006 pour fournir "une éducation gratuite de classe mondiale pour n'importe qui, n'importe où". Les cours relatifs à la programmation informatique peuvent être trouvés à l'adresse <https://www.khanacademy.org/computing/cs>.

Code Academy est une autre excellente ressource. Elle s'appuie sur l'interactivité pour aider les gens à apprendre à écrire des programmes informatiques. Vous pouvez les trouver sur <http://www.codecademy.com>.

3.3. Résumé

Ce chapitre a débuté par une discussion sur la manière d'appliquer la programmation de base pour prendre en charge les appareils IoT. Les organigrammes sont des diagrammes utilisés pour représenter des processus. Il existe deux types courants de logiciels informatiques : les logiciels système et les logiciels d'application. Les logiciels d'application sont créés pour accomplir une certaine tâche. Le logiciel système fonctionne entre le matériel informatique et le programme d'application. Les variables de programmation peuvent être classées en deux catégories :

- **Variables locales** - Ce sont des variables qui se trouvent dans le champ d'application d'un programme / d'une fonction / d'une procédure.
- **Variables globales** - Ce sont des variables qui sont dans le champ d'application pour le moment de l'exécution du programme. Elles peuvent être récupérées par n'importe quelle partie du programme.
-

Les structures logiques les plus courantes sont les boucles IF - THEN, FOR et WHILE.

Blockly est un outil de programmation visuelle créé pour aider les débutants à comprendre les concepts de la programmation. Blockly met en œuvre la programmation visuelle en attribuant différentes structures de programmation à des blocs colorés. Python est un langage très populaire, conçu pour être facile à lire et à écrire. Python est un langage interprété ; un interpréteur est donc nécessaire pour analyser et exécuter le code Python. Les variables sont des zones de mémoire étiquetées qui sont utilisées pour stocker les données du programme d'exécution. Python prend en charge de nombreuses fonctions et types de données utiles, notamment Range(), Tuples, Listes, Sets, Dictionary. Python implémente également deux sous-structures nommées ELSE et ELIF.

Ensuite, le chapitre détaille le prototypage. Le prototypage est le processus qui consiste à créer un modèle de travail rudimentaire d'un produit ou d'un système. Une équipe de Google a utilisé la "méthode de prototypage rapide" pour créer les Google Glass. L'internet permet d'échanger des idées à un tout autre niveau. Il existe plusieurs sites web que vous pouvez visiter pour vous connecter avec d'autres fabricants.

4. Tout génère des données

4.1. Big Data

4.1.1. Qu'est-ce que le big data ?

4.1.1.1. Introduction

Les données sont des informations qui proviennent de diverses sources, telles que des personnes, des images, des textes, des capteurs et des sites web. Les données proviennent également d'appareils technologiques comme les téléphones portables, les ordinateurs, les kiosques, les tablettes et les caisses enregistreuses. Plus récemment, on a assisté à un pic du volume de données générées par les capteurs. Les capteurs sont maintenant installés dans un nombre toujours croissant d'endroits et d'objets. Il s'agit notamment de caméras de sécurité, de feux de circulation, de voitures intelligentes, de thermomètres et même de vignes !

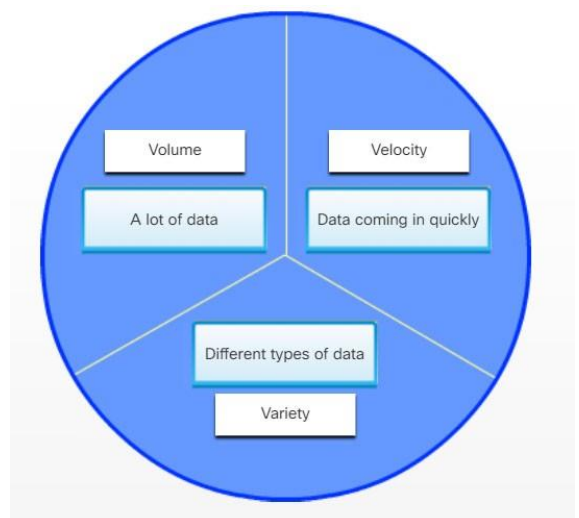
Le Big Data, c'est beaucoup de données, mais qu'est-ce qu'un lot ? Personne ne dispose d'un chiffre exact indiquant à quel moment les données d'une organisation sont considérées comme du "Big Data". Voici trois caractéristiques qui indiquent qu'une organisation peut avoir affaire à du Big Data :

- Ils disposent d'une grande quantité de données qui nécessitent de plus en plus d'espace de stockage (volume).
- Ils disposent d'une quantité de données dont la croissance est exponentielle (vélocité).
- Ils disposent de données qui sont générées dans différents formats (variété).

Combien de données les capteurs collectent-ils ? Voici quelques exemples estimés :

- Les capteurs d'une voiture autonome peuvent générer 4 000 gigabits (Go) de données par jour.
- Un moteur d'Airbus A380 génère 1 pétaoctet (PB) de données lors d'un vol entre Londres et Singapour.
- Les capteurs de sécurité dans les exploitations minières peuvent générer jusqu'à 2,4 téraoctets (TB) de données par minute.
- Les capteurs d'une maison connectée intelligente peuvent produire jusqu'à 1 gigaoctet (Go) d'informations par semaine.

Si le Big Data pose des défis aux organisations en termes de stockage et d'analyse, il peut également fournir des informations précieuses pour affiner les opérations et améliorer la satisfaction des clients.



4.1.1.2. Grands ensembles de données

Les entreprises ne doivent pas nécessairement générer leur propre Big Data. Les petites organisations n'ont peut-être pas les capteurs, le volume de clients ou la capacité de générer la variété d'informations qui pourraient profiter à leur entreprise. Il existe des sources d'ensembles de données gratuites, prêtes à être utilisées et analysées par quiconque est prêt à les chercher.

De nombreuses entreprises de toutes tailles pensent qu'elles doivent collecter leurs propres données pour bénéficier des avantages de l'analyse des big data, mais ce n'est tout simplement pas vrai.



4.1.2. Où sont stockées les données volumineuses ?

4.1.2.1. Quels sont les défis du big data ?

Les estimations d'IBM en matière de Big Data concluent que "chaque jour, nous créons 2,5 quintillions d'octets de données". Pour mettre cela en contexte, chaque minute de chaque jour :

- Nous téléchargeons plus de 300 heures de vidéos sur YouTube.
- Nous envoyons plus de 3,5 millions de messages texte.
- Nous diffusons plus de 86 000 heures de vidéos Netflix.
- Nous aimons plus de 4 millions de publications sur Facebook.
- Nous demandons plus de 14 millions de prévisions à The Weather Channel.

Allez voir : <https://www.internetlivestats.com/>

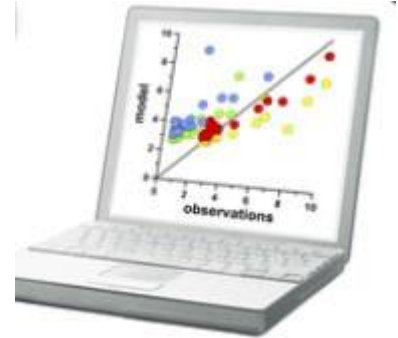
La croissance rapide des données peut être un avantage ou un obstacle lorsqu'il s'agit d'atteindre des objectifs commerciaux. Pour réussir, les entreprises doivent être en mesure d'accéder facilement à leurs données et de les gérer.

Avec cette énorme quantité de données créées en permanence, les technologies traditionnelles et les entrepôts de données ne peuvent pas répondre aux besoins de stockage. Même avec les installations de stockage en nuage proposées par des sociétés comme Amazon, Google, Microsoft et bien d'autres, la sécurité des données stockées devient un gros problème. Les solutions Big Data doivent être sécurisées, avoir une tolérance élevée aux pannes et utiliser la réplication pour s'assurer que les données ne se perdent pas. Le stockage du Big Data ne consiste pas seulement à stocker des données, mais aussi à les gérer et à les sécuriser.

Le Big Data pose cinq problèmes majeurs de stockage de données, comme le montre la figure.



- **Gestion** : Les données peuvent être générées et collectées à partir de multiples sources différentes. Un système de gestion doit donc être utilisé pour organiser et rassembler toutes ces sources. Il existe peu de normes de partage des données et des milliers d'outils de gestion des données.



- **La sécurité** : Les données ayant une valeur pour les entreprises doivent être sécurisées et accessibles uniquement aux utilisateurs autorisés.



- **La redondance** : Des mesures de protection doivent être mises en place pour maintenir l'intégrité des données stockées. Des processus de sauvegarde, de redondance et de reprise après sinistre sont nécessaires.

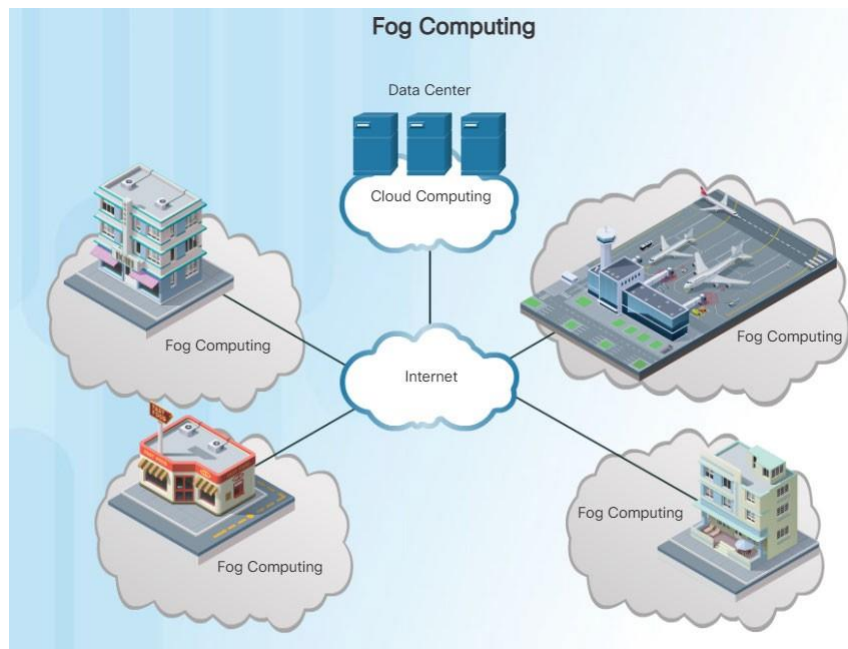


- **Analytique** : Le big data contient à la fois des données structurées et non structurées. Les données non structurées, comme les courriels ou les photos, sont difficiles à analyser et consomment beaucoup d'espace de stockage. Les données structurées, telles que les feuilles de calcul, sont plus faciles à analyser et à stocker. Les systèmes de stockage de données doivent être capables de gérer les deux types de données.



- **L'accès** : Les big data doivent être accessibles de partout et à tout moment. Les solutions de stockage doivent supporter la quantité de demandes d'entrée et de sortie. Les entreprises doivent également être conscientes de la pression exercée sur les liaisons WAN.

4.1.2.2. Où peut-on stocker les big data ?



Les données volumineuses sont généralement stockées sur plusieurs serveurs, généralement hébergés dans des centres de données. Pour des raisons de sécurité, d'accessibilité et de redondance, les données sont généralement distribuées et/ou répliquées sur de nombreux serveurs différents dans de nombreux centres de données différents.

Informatique de brouillard

L'informatique en brouillard est une architecture qui utilise des clients utilisateurs ou des dispositifs "périphériques" pour effectuer une grande partie du prétraitement et du stockage requis par une organisation. L'informatique en brouillard a été conçue pour que les données soient plus proches de la source pour le prétraitement.

Les données des capteurs, en particulier, peuvent être prétraitées plus près de l'endroit où elles ont été recueillies. Les informations obtenues à partir de cette analyse pré-traitée peuvent être réinjectées dans les systèmes de l'entreprise pour modifier les processus si nécessaire. Comme les données des capteurs sont prétraitées par les appareils finaux au sein du système de l'entreprise, les communications vers et depuis les serveurs et les appareils sont plus rapides. Cela nécessite moins de bande passante que de se rendre constamment dans le nuage.

Une fois que les données ont été prétraitées, elles sont souvent expédiées pour un stockage à plus long terme, une sauvegarde ou une analyse plus approfondie dans le nuage.

4.1.2.3. Le nuage et l'informatique dématérialisée

Comme mentionné précédemment, le nuage est un ensemble de centres de données ou de groupes de serveurs connectés. L'accès aux logiciels, au stockage et aux services disponibles sur les serveurs se fait par Internet via une interface de navigateur. Les services de cloud computing sont fournis par de nombreuses grandes entreprises telles que Google, Microsoft et Apple. Les services de stockage en nuage sont fournis par différents fournisseurs tels que : Google Drive, Apple iCloud, Microsoft OneDrive et Dropbox.

Du point de vue de l'individu, l'utilisation des services en nuage vous permet :

- Pour stocker toutes vos données, telles que des photos, de la musique, des films et des courriels, ce qui libère de l'espace sur le disque dur local.
- Pour accéder à de nombreuses applications au lieu de les télécharger sur votre appareil local
- Pour accéder à vos données et applications partout, à tout moment et sur n'importe quel appareil.

L'un des inconvénients de l'utilisation du cloud est que vos données peuvent tomber entre de mauvaises mains. Vos données sont à la merci de la robustesse de la sécurité du fournisseur de services en nuage que vous avez choisi.

Du point de vue d'une entreprise, les services et l'informatique en nuage prennent en charge toute une série de questions relatives à la gestion des données :

- Il permet d'accéder aux données de l'organisation partout et à tout moment.
- Il rationalise les opérations informatiques d'une organisation en s'abonnant uniquement aux services nécessaires.
- Il élimine ou réduit le besoin d'équipement informatique sur site, de maintenance et de gestion.
- Elle réduit le coût des équipements, de l'énergie, des installations physiques et des besoins en formation du personnel.
- Il permet de répondre rapidement aux exigences croissantes en matière de volume de données.

L'informatique en nuage est un autre moyen de gérer, de stocker et d'accéder aux données.

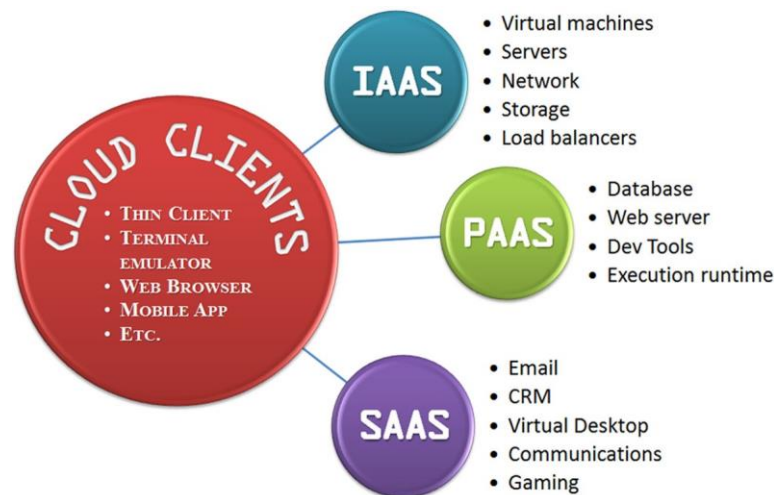
L'informatique en nuage implique un grand nombre d'ordinateurs reliés par un réseau. Les fournisseurs de cloud computing s'appuient fortement sur la virtualisation pour fournir leurs services. Ils peuvent également réduire les coûts opérationnels en utilisant les ressources de manière plus efficace. Ces entreprises fournissent quatre catégories distinctes de services. Cliquez sur les catégories dans la figure pour plus d'informations.

L'informatique en nuage permet aux utilisateurs d'accéder à leurs données partout et à tout moment. Vous utilisez probablement déjà une forme de cloud computing si vous utilisez des services de courrier électronique basés sur le web.

Le cloud computing permet également aux organisations de rationaliser leurs opérations informatiques en s'abonnant uniquement aux services nécessaires. En utilisant l'informatique en nuage, les organisations peuvent également éliminer le besoin d'équipement informatique sur site, de maintenance et de gestion. Le cloud computing réduit les coûts pour les organisations. Il réduit les coûts d'équipement, les coûts énergétiques, les besoins en installations physiques et les besoins en formation du personnel de soutien.

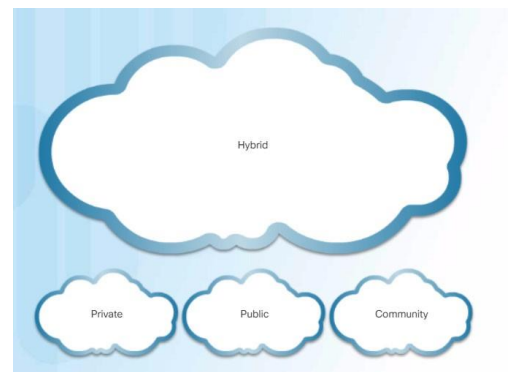
- **SaaS** : Applications fournies aux utilisateurs finaux via le web

- **PaaS** : Outils et services utilisés pour fournir les applications
- **IaaS** : matériel et logiciel pour alimenter les serveurs, le stockage, les réseaux et le système d'exploitation.
- **ITaaS** : Les professionnels de l'informatique prennent en charge les applications, les plateformes et l'infrastructure



Types de nuages :

- Privé
- Public
- Communauté
- Hybride



- **Un cloud privé** est créé exclusivement pour une seule organisation. L'infrastructure peut être physiquement situé sur ou hors site, et peut être détenu par un fournisseur distinct. Le Cloud privé fournit des services uniquement aux membres de l'organisation unique.
- **Un cloud public** est créé pour être utilisé par le grand public. L'infrastructure est physiquement située sur le site du fournisseur, mais peut être détenue par une ou plusieurs organisations, notamment des entreprises, des établissements universitaires ou des gouvernements.
- **Un cloud communautaire** est créé pour une utilisation exclusive par une communauté spécifique. La communauté se compose de plusieurs organisations qui ont des préoccupations communes (par exemple, la mission, les exigences de sécurité, la politique et les considérations de conformité). L'infrastructure peut être physiquement située sur ou hors site et peut être détenue par un fournisseur distinct ou par une ou plusieurs organisations de la communauté. Les différences entre les clouds publics et les clouds communautaires résident dans les besoins fonctionnels qui ont été personnalisés pour la communauté. Par exemple, les organismes de santé doivent rester conformes aux politiques et aux lois (par exemple, HIPAA) qui exigent une authentification et une confidentialité particulières. Les organisations peuvent partager l'effort de mise en œuvre de ces exigences dans un déploiement de cloud commun.
- **Une infrastructure de cloud hybride** est une composition de deux ou plusieurs infrastructures de cloud distinctes (privées, communautaires ou publiques) qui sont des entités uniques. Ces entités sont liées entre elles par une technologie qui permet la portabilité des données et des applications. Cette portabilité permet à une organisation de conserver une perspective unique d'une solution de cloud tout en profitant des points forts disponibles chez les différents fournisseurs de cloud. Par exemple, la géographie (emplacement par rapport aux utilisateurs finaux), la bande passante, les exigences politiques ou légales, la sécurité et le coût sont autant de caractéristiques qui peuvent

différencier les fournisseurs. Un cloud hybride offre la flexibilité d'ajuster et de réagir à ces services de fournisseurs, à la demande.

Centres de données :

Les centres de données sont un élément essentiel de l'informatique en nuage. Un centre de données est une installation qui fournit les services nécessaires pour héberger les plus grands environnements informatiques qui existent aujourd'hui. Sa principale fonction est d'assurer la continuité des activités en maintenant les services informatiques disponibles à tout moment.

Pour fournir le niveau de service nécessaire, plusieurs facteurs doivent être pris en compte dans le déploiement d'un centre de données :

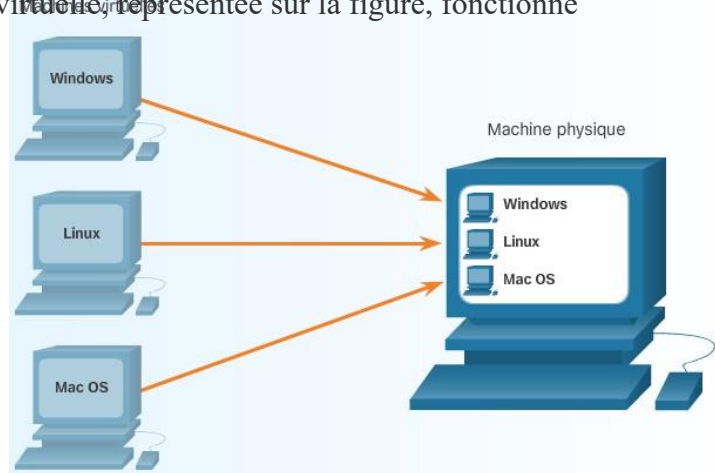
- **Emplacement** : Les centres de données doivent être situés dans des endroits où les risques de catastrophes naturelles sont réduits et suffisamment éloignés des zones à fort trafic de personnes (par exemple, les aéroports, les centres commerciaux, etc.) et des zones d'importance stratégique pour les gouvernements et les services publics (par exemple, les raffineries, les barrages, les réacteurs nucléaires, etc.)
- **Sécurité** : Un centre de données doit exercer des contrôles stricts sur l'accès physique et le personnel sur place.
- **Électricité** : l'accès à l'énergie électrique doit être suffisant. Il doit y avoir une alimentation de secours composée de systèmes d'alimentation sans coupure, de batteries et de générateurs électriques.
- **Environnement** : Un environnement physique étroitement contrôlé qui maintient une température et une humidité appropriées. Il doit également comporter des systèmes sophistiqués d'extinction des incendies.
- **Réseau** : L'infrastructure du réseau doit être évolutive et fiable avec une connectivité redondante.

Actuellement, il existe plus de 3 000 centres de données dans le monde qui offrent des services d'hébergement général (IaaS) aux particuliers et aux organisations. Il y a beaucoup plus de centres de données qui sont détenus et exploités par des industries privées pour leur propre usage.

Virtualisation

Historiquement, chaque ordinateur possède son propre système d'exploitation, ses propres applications et des composants matériels dédiés. Aujourd'hui, grâce à l'émulation logicielle, plusieurs ordinateurs virtuels peuvent fonctionner sur un seul ordinateur physique. Cela signifie que chaque ordinateur virtuel possède son propre système d'exploitation, ses applications et ses composants matériels dédiés. C'est ce qu'on appelle la virtualisation en informatique. Chaque machine virtuelle, représentée sur la figure, fonctionne indépendamment.

Dans le monde de l'entreprise, une seule infrastructure physique peut faire fonctionner plusieurs infrastructures virtuelles. En virtualisant les serveurs et les réseaux, les entreprises peuvent réduire les coûts opérationnels et administratifs. Les économies opérationnelles peuvent provenir de la



réduction des besoins en énergie et en refroidissement et du nombre de machines physiques. Un serveur virtuel peut être ajouté pour prendre en charge des applications supplémentaires.

Vous pouvez également utiliser la virtualisation pour vos besoins informatiques personnels. Vous pouvez essayer un nouveau système d'exploitation sur votre ordinateur sans endommager votre système actuel. Vous pouvez naviguer sur Internet en toute sécurité avec votre machine virtuelle. La machine virtuelle peut être supprimée en cas de problème.

4.1.2.4 Traitement distribué

Du point de vue de la gestion des données, l'analyse était simple lorsque seuls les humains créaient des données. La quantité de données était gérable et relativement facile à passer au crible. Cependant, avec l'explosion des systèmes d'automatisation des entreprises et la croissance exponentielle des applications web et des données générées par les machines, l'analytique devient de plus en plus difficile à gérer. En fait, 90 % des données qui existent aujourd'hui ont été générées au cours des deux dernières années seulement. Cette augmentation du volume en un court laps de temps est une propriété de la croissance exponentielle. Ce volume élevé de données est difficile à traiter et à analyser dans un laps de temps raisonnable.

Plutôt que de traiter les grandes bases de données par de gros et puissants ordinateurs centraux et de les stocker dans des réseaux de disques géants (mise à l'échelle verticale), le **traitement distribué des données** prend le grand volume de données et le décompose en plus petits morceaux. Ces volumes de données plus petits sont répartis en de nombreux endroits pour être traités par de nombreux ordinateurs dotés de processeurs plus petits. Chaque ordinateur de l'architecture distribuée analyse sa partie de l'image du Big Data (mise à l'échelle horizontale).

La plupart des systèmes de fichiers distribués sont conçus pour être invisibles aux programmes clients. Le système de fichiers distribué localise les fichiers et déplace les données, mais les utilisateurs n'ont aucun moyen de savoir que les fichiers sont répartis entre plusieurs serveurs ou nœuds différents. Les utilisateurs accèdent à ces fichiers comme s'ils étaient locaux, sur leur propre ordinateur. Tous les utilisateurs ont la même vue du système de fichiers et peuvent accéder aux données en même temps que les autres utilisateurs.

Hadoop a été créé pour traiter ces volumes de Big Data. Le projet Hadoop a débuté avec deux facettes : Le système de fichiers distribué Hadoop (HDFS), qui est un système de fichiers distribué et tolérant aux pannes, et MapReduce, qui est un moyen distribué de traiter les données. Hadoop a maintenant évolué pour devenir un écosystème très complet de logiciels pour la gestion du Big Data.



Hadoop est un logiciel libre permettant le traitement distribué de grands ensembles de données pouvant atteindre des téraoctets et stockés dans des grappes d'ordinateurs. Hadoop est conçu pour passer d'un seul serveur à des milliers de machines, chacune offrant un calcul et un stockage locaux. Pour plus d'efficacité, Hadoop peut être installé et exécuté sur de nombreuses machines virtuelles. Ces machines virtuelles peuvent toutes travailler ensemble en parallèle pour traiter et stocker les données.

Hadoop possède deux caractéristiques principales qui en ont fait la norme industrielle pour le traitement des Big Data :

- Évolutivité - Des clusters de plus grande taille améliorent les performances et offrent des capacités de traitement des données plus élevées. Avec Hadoop, la taille des clusters peut facilement passer d'un cluster de cinq nœuds à un cluster de mille nœuds sans augmenter

excessivement la charge administrative.

- Tolérance aux pannes - Hadoop réplique automatiquement les données entre les clusters pour s'assurer que les données ne seront pas perdues. Si un disque, un nœud ou un rack entier tombe en panne, les données sont en sécurité.

4.1.3. Soutenir les entreprises grâce au big data

4.1.3.1. Pourquoi les entreprises analysent-elles les données ?

Chaque organisation doit devenir plus efficace et innovante pour rester compétitive et pertinente dans le monde numérisé. L'IdO fait partie intégrante de la réalisation de cette efficacité et de cette innovation. L'objectif de nombreuses entreprises est de collecter et d'analyser les quantités massives de données relatives à l'utilisation des nouveaux produits afin d'obtenir des informations précieuses. L'analyse des données permet aux entreprises de mieux comprendre l'impact de leurs produits et services, d'adapter leurs méthodes et leurs objectifs et de fournir plus rapidement de meilleurs produits à leurs clients. La capacité de tirer de nouvelles informations de leurs données apporte de la valeur à l'entreprise.

Pour les entreprises, les données sont le nouveau pétrole. Comme le pétrole brut, il est précieux, mais s'il n'est pas raffiné, il ne peut pas être utilisé facilement. Le pétrole brut doit être transformé en essence, plastique, produits chimiques et autres substances pour créer un produit de valeur. Il en va de même pour les données. Les données doivent être décomposées et analysées pour avoir de la valeur.

La valeur provient de deux principaux types de données traitées, transactionnelles et analytiques. Les informations transactionnelles sont capturées et traitées au fur et à mesure que les événements se produisent. Les informations transactionnelles sont utilisées pour analyser les rapports de vente quotidiens et les calendriers de production afin de déterminer le volume des stocks à conserver. Les informations analytiques permettent d'effectuer des analyses de gestion, par exemple pour déterminer si l'entreprise doit construire une nouvelle usine de fabrication ou embaucher du personnel de vente supplémentaire.

4.1.3.2. Source d'information

La source des données dans les grands ensembles de données est variée. Outre les données des capteurs, d'autres données proviennent de tout ce qui a été scanné, saisi et diffusé sur Internet à partir de sources telles que :

- Sites de médias sociaux - Facebook, YouTube, eHarmony et Twitter
- HTTP, pages Web et moteurs de recherche sur Internet
- Données historiques provenant d'archives publiques et privées
- Les métadonnées qui sont jointes aux courriels, aux documents transmis et aux images.
- Formulaires médicaux, formulaires d'assurance et formulaires fiscaux
- Recherche génomique utilisant l'ADN

Les données collectées peuvent être classées comme étant structurées ou non structurées.

Les données structurées sont créées par des applications qui utilisent des entrées au format "fixe", comme les feuilles de calcul ou les formulaires médicaux. Même si les données sont considérées comme structurées, différentes applications créent des fichiers dans des formats différents qui ne sont pas nécessairement compatibles entre eux. Les données structurées peuvent avoir besoin d'être manipulées dans un format commun tel que le CSV.

Les fichiers CSV (Comma-separated values) sont un type de fichier en texte clair qui utilise des virgules pour séparer les colonnes d'un tableau de données et le caractère retour chariot pour séparer les lignes. Chaque ligne est un enregistrement. Bien qu'ils soient couramment utilisés pour l'importation et l'exportation dans les bases de données et les feuilles de calcul traditionnelles, il n'existe pas de norme

spécifique. JSON et XML sont également des types de fichiers en texte clair qui utilisent une manière standard de représenter les enregistrements de données. Ces formats de fichier sont compatibles avec un large éventail d'applications. La conversion des données dans un format commun est un moyen précieux de combiner des données provenant de sources différentes.

Les données non structurées sont générées dans un style "libre", comme l'audio, la vidéo, les pages web et les tweets. Les données non structurées nécessitent différents outils pour préparer les données en vue de leur traitement ou de leur analyse. En voici deux exemples :

- Les pages Web sont créées pour fournir des données aux humains, pas aux machines. Les outils de "Web scraping" extraient automatiquement les données des pages HTML. C'est similaire à un Web Crawler ou spider d'un moteur de recherche. Il explore le web pour extraire les données et créer la base de données pour répondre aux requêtes de recherche. Les logiciels de scraping web peuvent utiliser le protocole de transfert hypertexte ou un navigateur web pour accéder au World Wide Web. En général, le web scraping est un processus automatisé qui utilise un robot ou un crawler web pour faire de l'exploration de données. Des données spécifiques sont recueillies et copiées du Web vers une base de données ou une feuille de calcul. Les données peuvent ensuite être facilement analysées.



- De nombreux grands fournisseurs de services web, tels que Facebook, proposent des interfaces normalisées pour collecter automatiquement les données à l'aide de programmes d'application interfaces (API). L'approche la plus courante consiste à utiliser des API RESTful. Les API RESTful utilisent HTTP comme protocole de communication et la structure JSON pour encoder les données. Les sites Internet comme Google et Twitter rassemblent de grandes quantités de données statiques et de séries chronologiques. La connaissance des API de ces sites permet aux analystes de données et aux ingénieurs d'accéder aux grandes quantités de données qui sont constamment générées sur Internet.



4.1.3.3. Visualisation des données

L'exploration de données est le processus qui consiste à transformer des données brutes en informations significatives en découvrant des modèles et des relations dans de grands ensembles de données.

Pour avoir de la valeur, les données extraites doivent être analysées et présentées aux gestionnaires et aux décideurs. Il existe de nombreuses visualisations différentes qui peuvent être utilisées pour présenter la valeur des données. La détermination du meilleur graphique à utiliser varie en fonction des éléments suivants :

- Nombre de variables à afficher
- Nombre de points de données dans chaque variable
- Les données représentent-elles une ligne de temps



- Éléments qui nécessitent des comparaisons
-

Les types de graphiques les plus populaires sont les suivants : ligne, colonne, barre, tarte et dispersion.

4.1.3.4. Analyser les big data pour une utilisation efficace dans les affaires

Les données volumineuses ne sont rien d'autre que des données volumineuses ! Elles sont plus utiles si vous pouvez les analyser pour en tirer de la valeur. L'analyse des données consiste à inspecter, nettoyer, transformer et modéliser les données afin de découvrir des informations utiles. L'analyse des big data nécessite généralement des outils et des applications créés à cette fin. Ces outils d'analyse ont été conçus pour fournir aux entreprises des informations détaillées, des modèles et des renseignements précieux.

Avant de commencer toute analyse, il est essentiel de connaître le problème que l'entreprise tente de résoudre ou les informations qu'elle recherche. S'intéressent-elles au comportement des clients dans des États spécifiques, aux habitudes de consommation d'énergie dans différents quartiers de la ville ou au nombre de "likes" sur Facebook en fonction de l'âge ?

Une stratégie aide l'entreprise à déterminer le type d'analyse nécessaire et le meilleur outil pour effectuer cette analyse. Une stratégie permet également de déterminer la manière la plus efficace de présenter les résultats à la direction.

Les outils et les applications vont de l'utilisation d'une feuille de calcul Excel ou de Google Analytics pour des échantillons de données de petite ou moyenne taille, aux applications dédiées à la manipulation et à l'analyse de très grands ensembles de données.

Il existe de nombreux outils d'analyse du Big Data qu'une entreprise peut choisir, tels que : Knime, OpenRefine, Orange et RapidMiner.

- **Knime** : La plateforme d'analyse KNIME est la principale solution ouverte pour l'innovation basée sur les données, conçue pour découvrir le potentiel caché dans les données, extraire de nouvelles informations ou prédire l'avenir. Les organisations peuvent faire passer leur collaboration, leur productivité et leurs performances au niveau supérieur grâce à une gamme robuste d'extensions commerciales de notre plateforme open source.
- **OpenRefine** : est un outil puissant pour travailler avec des données désordonnées. OpenRefine prend les données et les nettoie, les transforme d'un format à un autre et les étend avec des services Web et des données externes. Veuillez noter que depuis le 2 octobrend, 2012, Google ne soutient plus activement ce projet, qui a été rebaptisé OpenRefine. Le développement, la documentation et la promotion du projet sont désormais entièrement pris en charge par des bénévoles.
- **Rapidminer** : Tout comme KNIM, Rapidminer fonctionne par programmation visuelle et est capable de manipuler, analyser et modéliser des données. La plateforme unifiée de science des données de Rapidminer accélère la construction de flux de travail analytiques complets, de la préparation des données à l'apprentissage automatique en passant par la validation des modèles et le déploiement. Tout cela se passe dans un seul environnement, ce qui améliore considérablement l'efficacité et réduit le temps de valorisation des projets de science des données.
- **Orange** : Orange est un logiciel open source de visualisation et d'analyse de données destiné aux novices et aux experts. Il fournit une large boîte à outils permettant de créer des flux de travail interactifs pour analyser et visualiser des données. Orange est rempli de différentes visualisations, de diagrammes de dispersion, de diagrammes à barres, d'arbres, de dendrogrammes, de réseaux et de cartes thermiques.

4.2.Résumé

Le Big Data présente généralement trois caractéristiques. Il s'agit d'une grande quantité de données qui nécessite de plus en plus d'espace de stockage (volume), qui croît à une vitesse exponentielle (vélocité) et qui est générée dans différents formats (variété).

L'informatique en brouillard est une architecture qui utilise des clients utilisateurs ou des dispositifs "périphériques" pour effectuer une grande partie du prétraitement et du stockage requis par une organisation. L'informatique en brouillard a été conçue pour que les données soient plus proches de la source pour le prétraitement.

Le nuage est un ensemble de centres de données ou de groupes de serveurs connectés permettant d'accéder partout et à tout moment à des logiciels, à du stockage et à des services au moyen d'une interface de navigateur. Les services en nuage permettent d'augmenter le stockage des données selon les besoins et réduisent le besoin d'équipement informatique sur site, la maintenance et la gestion. Ils réduisent également le coût des équipements, de l'énergie, des installations physiques et des besoins en formation du personnel.

Le traitement distribué des données prend de gros volumes de données d'une source et les décompose en plus petits morceaux. Ces volumes de données plus petits sont distribués en de nombreux endroits pour être traités par de nombreux ordinateurs dotés de processeurs plus petits. Chaque ordinateur de l'architecture distribuée analyse sa partie de l'image du Big Data.

Les entreprises gagnent de la valeur en collectant et en analysant des quantités massives de données sur l'utilisation des nouveaux produits pour comprendre l'impact de leurs produits et services, ajuster leurs méthodes et leurs objectifs et fournir plus rapidement de meilleurs produits à leurs clients.

Les données collectées peuvent être classées comme étant structurées ou non structurées. Les données structurées sont créées par des applications qui utilisent des entrées de format "fixe", comme les feuilles de calcul ou les formulaires médicaux. Les données non structurées sont générées dans un style "libre" comme l'audio, la vidéo, les pages web et les tweets. Les deux formes de données doivent être manipulées dans un format commun pour être analysées. CSV, JSON et XML sont des types de fichiers en texte clair qui utilisent une méthode standard pour représenter les enregistrements de données. La conversion des données dans un format commun est un moyen précieux de combiner des données provenant de sources différentes.

L'exploration de données est le processus qui consiste à transformer des données brutes en informations significatives en découvrant des modèles et des relations dans de grands ensembles de données. La visualisation des données est le processus qui consiste à prendre les données analysées et à utiliser des graphiques tels que des lignes, des colonnes, des barres, des camemberts ou des diagrammes de dispersion pour présenter des informations significatives. Une stratégie aide une entreprise à déterminer le type d'analyse requis et le meilleur outil pour effectuer cette analyse. Une stratégie permet également de déterminer la manière la plus efficace de présenter les résultats à la direction.

5. Tout peut être automatisé

5.1. Qu'est-ce qui peut être automatisé ?

5.1.1. Automatisation

5.1.1.1. Qu'est-ce que l'automatisation ?

On entend par automatisation tout processus qui se déroule de manière autonome et qui réduit, puis finit par éliminer, le besoin d'intervention humaine.

L'automatisation était autrefois confinée à l'industrie manufacturière. Des tâches très répétitives, comme l'assemblage de voitures, ont été confiées à des machines et la chaîne de montage moderne est née. Les machines sont excellentes pour répéter la même tâche sans fatigue et sans les erreurs que les humains sont enclins à faire dans ce genre de travail. Il en résulte un rendement supérieur, car les machines peuvent travailler 24 heures sur 24 sans pause. Les machines fournissent également un produit plus uniforme.

L'IoT ouvre un nouveau monde dans lequel des tâches nécessitant auparavant une intervention humaine peuvent être automatisées. Comme nous l'avons vu, l'IoT permet de collecter de grandes quantités de données qui peuvent être rapidement analysées pour fournir des informations susceptibles d'orienter un événement ou un processus.

Alors que nous continuons à adopter les avantages de l'IoT, l'automatisation devient de plus en plus importante. L'accès à d'énormes quantités de données de capteurs traitées rapidement a fait réfléchir les gens sur la façon d'appliquer les concepts d'apprentissage automatique et d'automatisation aux tâches quotidiennes. De nombreuses tâches de routine sont en cours d'automatisation pour améliorer leur précision et leur efficacité.

L'automatisation est souvent liée au domaine de la robotique. Les robots sont utilisés dans des conditions dangereuses telles que l'exploitation minière, la lutte contre les incendies et le nettoyage des accidents industriels, réduisant ainsi le risque pour les humains. Ils sont également utilisés dans des tâches telles que les chaînes de montage automatisées.

Aujourd'hui, l'automatisation est omniprésente, qu'il s'agisse des caisses en libre-service dans les magasins, des contrôles automatiques de l'environnement des bâtiments ou des voitures et avions autonomes. Combien de systèmes automatisés rencontrez-vous en une seule journée ?