Rapport d’étude et de Recherche

Story 2 : Introduction Bases de Données

# Contexte :

Comprendre ce qu’est une base de données. Connaitre les types de bases de données, les types de données qu’elles utilisent, et les notions.

# Mots Clefs :

* Base de Données
  + relationnelle
  + non relationnelle
* Données
  + Structurées
  + Semi-Structurées
* Information
* Connaissance
* Sagesse
* Champ
* Enregistrement
* Table
* SQL
* NoSQL
* Big Data
* Cloud Computing

# Problématique :

1. Quel type de base de données pourraient répondre aux besoins de l'entreprise?
2. Comprendre comment stocker, organiser, et traiter nos données structurées ou non structurer?

# Hypothèses :

1. La base de données permet de stocker la donnée client de l’entreprise.
2. La base de données relationnelle est majoritairement utilisée pour passer les données entre tables.
3. Les activités de l’entreprise peuvent impacter sur le type de base de données.
4. Le type de données ne permet d’utilisée qu’un seul type de base de données.
5. Il existe plusieurs types de base de données
6. Une base de données permet de stocker un nombre important de données
7. Une base de données permet de récolter des données non-ordonnées d’une entreprise pour les replacer de manière ordonnée
8. Les différents types de bases de données présentent des contraintes différentes en terme d'opération qui peuvent être conduites sur les données
9. La base de données ne peut traiter la donnée si cette dernière n’a pu être stockée et organisée au préalable
10. Une base de données peut contenir des données unique à un tableau
11. Toutes les entreprises n'ont pas besoin de base de données relationnelles
12. MySQL est le type de base de données le plus simple à exploiter
13. Les bases de données relationnels fonctionnent avec des champs-clé

# Plan d’action

1. Investigation des ressources en lien avec la problématique
2. Avantage/inconvénients des bases de données relationnelles et non relationnelles
3. Rechercher caractéristiques des données structurées et non structurées et leur différences
4. Définition des mots clés
5. Répondre aux hypothèses
6. Écrire le RER (Livrable)

Rapport d’Étude et de recherche

# Rechercher caractéristiques des données structurées et non structurées et leur différences

Les donnée structurée sont des données prédéfinie et formatée selon une structure précise avant d'être enregistrer dans une base de données

Présente les avantages d'être traiter plus facilement par un algorithmes de machine learning ou par un utilisateur professionnel, et d'avoir davantage d'outils de traitement disponible.

Par contre son but prédéfini en limite l'utilisation et les options de stockage sont limitées étant donné que toute modification des exigences entraîne la mise à jour de toutes ces données structurée pour les adapter au nouveaux besoins ce qui entraîne des dépense massive de ressources et de temps.

Les données non-structurées sont des données stockées dans leur format d'origine et non traitées avant leur utilisations.

Leurs avantages sont la liberté du format natif et taux d'accumulation plus rapide des données ainsi que des solutions de stockage appelé "data lake" qui ont un coût réduis et plus simple à faire évoluer.

Malheureusement, leur exploitation nécessite d'avoir soit une expertise en data science soit accès à un expert en data science car une personne capable d'utiliser des bases de données relationnelle ne peut pas utiliser des données non structurées telles quelles. De plus, l'exploitation de ces données nécessite l'utilisation d'outils spéciaux au nombre limité, dont certains n'en sont encore qu'à leurs débuts.

# Définition des mots clés

* Base de Données : ensemble d'informations structurées accessibles au moyen d'un logiciel.
  + relationnelle : type de base de données où les données sont liées à d'autres informations au sein de la ou des bases de données.
  + non relationnelle : type de base de données qui n’utilise pas le schéma tabulaire sous forme de lignes et de colonnes présent dans la plupart des systèmes de base de données plus traditionnels
* Données : les chiffres, caractères, mots sous forme brut
  + Structurées : données qui a été prédéfinie et formatée selon une structure précise avant d'être placée dans une base de données
  + Semi-Structurées : forme intermédiaire entre les données structurées et non structurées. Elles ne sont pas organisées selon une méthode complexe rendant possible un accès et une analyse sophistiqués ; cependant, certaines informations peuvent leur être associées, telles que des balises de métadonnées, qui permettent l'adressage des éléments qu'elles renferment.
  + Non structurée : données qui ne sont par organisées dans un format permet d'y accéder et de les traiter plus facilement
* Information : ensemble de données collectées, structurées de manière compréhensible dans un contexte spécifique
* Connaissance : lorsqu'on traite des données, pour en tirer des patterns, dans un contexte particullier
* Sagesse : ensemble de connaissance
* Champ : information élémentaire d’une base de données
* Enregistrement : élément d'une base de données
* Table : ensemble de données organisées sous forme d'un tableau où les colonnes correspondent à des catégories d'information
* SQL (Structured Query Language) : langage informatique normalisé servant à exploiter des bases de données relationnelles. Il permet de définir, protéger et manipuler les données de manière simple et schématique.
* NoSQL : famille de systèmes de gestion de base de données (SGBD) qui s'écarte du paradigme classique des bases relationnelles. L'explicitation la plus populaire de l'acronyme est Not only SQL (« pas seulement SQL » en anglais).
* Big Data : ensemble de données volumineux et variées qui sont reçu et traitées si rapidement qu'un logiciel de traitement de données traditionnel ne peut tout simplement pas les gérer.
* Cloud Computing : utilisation de la mémoire et des capacités de calcul des ordinateurs et des serveurs répartis dans le monde entier et liés par un réseau. Les applications et les données ne se trouvent plus sur un ordinateur déterminé mais dans un nuage (cloud) composé de nombreux serveurs distants interconnectés.
* Enregistrements ou tuple : les lignes d’une table de base de données relationnelle, constitué d’un ensemble de valeurs organisées selon des attributs spécifique et possédant un type spécifique
* Table : tableau formé d’un ensemble d’attribut (colonnes) et enregistrements (lignes)

# Réponse aux hypothèses :

1. La base de données permet de stocker la donnée client de l’entreprise.

VRAI

1. La base de données relationnelle est majoritairement utilisée pour passer les données entre tables.

VRAI, via les champs commun

1. Les activités de l’entreprise peuvent impacter sur le type de base de données.

Cela dépend du types de données collectées

1. Le type de données ne permet d’utilisée qu’un seul type de base de données.

VRAI (données structurées avec base de données relationnelle vs données non structurer avec base de données non relationnelle)

1. Il existe plusieurs types de base de données.

VRAI (Base de données relationnelle et non relationnelle)

1. Une base de données permet de stocker un nombre important de données.

Tout est relatif. Il faut prédéfinir la taille de la Base de Données dans un SGBD

1. Une base de données permet de récolter des données non-ordonnées d’une entreprise pour les replacer de manière ordonnée.

VRAI

1. Les différents types de bases de données présentent des contraintes différentes en terme d'opération qui peuvent être conduites sur les données

VRAI

1. La base de données ne peut traiter la donnée que si cette dernière n’a pu être stockée et organisée au préalable

VRAI

1. Une base de données peut contenir des données unique à un tableau

FAUX, Une table doit contenir au moins un champs à valeur unique/champ unique (clé primaire)

1. Toutes les entreprises n'ont pas besoin de base de données relationnelles

FAUX, pour les besoins salariales, organisationnelles, hiérarchiques...

1. MySQL est le type de base de données le plus simple à exploiter

FAUX, il s’agit d’un logiciel de gestion de bases de données relationnelle

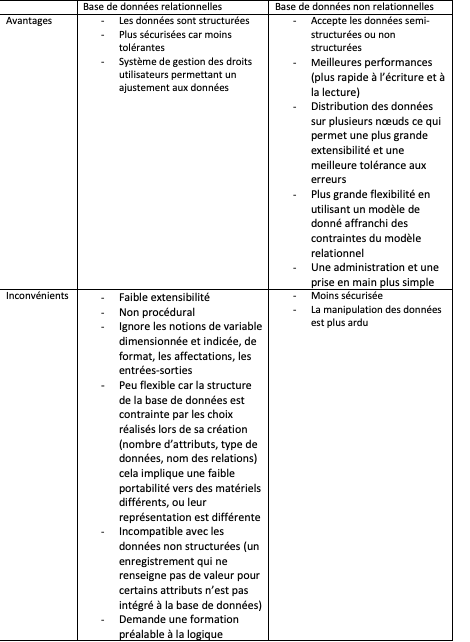
1. Les bases de données relationnels fonctionnent avec des champs-clé

VRAI, on les appelle clés primaire et clés étrangère

# Avantage/inconvénients des bases de données relationnelles et non relationnelles

|  |  |  |
| --- | --- | --- |
| Type de base de données | Avantages | Inconvénients |
| Relationnelle | * Données structurées, cohérente, et sécurisée * Possibilité de gérer des droits d’accès différents à la base de données et aux données * Pas de limite de taille des bases de données | * Pas adaptée aux données non structurées * Difficile à faire évoluer à cause de la structure * La connaissance du SQL est obligatoire pour pouvoir utiliser ce moteur de base de données |
| Non-relationnelle | * Non structurée et semi-structurée * Flexibilité de votre base de données et de vos données * Rapidité de prise en main * Rapidité plus forte dans l’écriture des données que le relationnel | * Sécurité plus faible * Cohérence des données plus faible * Regroupement et croisement de données plus complexe |

|  |  |  |
| --- | --- | --- |
| Données Structurées | Données non structurées | Données semi structurées |
| Accès en libre-service | Expertise en Data Science nécessaire | Expertise en Data Science nécessaire |
| Type de données sélectionnées | Nombreux types différents de conglomérats | Basée sur le XML et RDF |
| Schéma à l’écriture | Schéma à la lecture | Plus flexible que les données structurées mais moins que les données non structurées |
| Généralement stockées dans des Data Warehouses | Généralement stockées dans des Data Lakes |  |
| Format prédéfini | Format natif |  |
| Mise à l’échelle laborieuse | Mise à l’échelle simple | Mise à l’échelle plus simple que pour les données structurées |



# Sources :

* <https://apprendre-la-programmation.net/type-de-base-de-donnees-choisir/> (BDD relationnelles et non relationnelles)
* <https://www.mysql.com/fr/> (MySQL)
* <https://fr.wikipedia.org/wiki/NoSQL> (NoSQL)
* <https://www.talend.com/fr/resources/structured-vs-unstructured-data/> (Données structurées & non structurées)
* <https://www.lemagit.fr/definition/Donnees-semi-structurees> (Données structurées & non structurées)
* <https://docs.oracle.com/en/database/index.html> (Bases de données)
* <https://learn.microsoft.com/fr-fr/azure/architecture/data-guide/big-data/non-relational-data> (Données non relationnelles & NoSQL)
* <https://www.cnil.fr/fr/definition/cloud-computing> (Cloud Computing)
* <https://www.oracle.com/fr/big-data/what-is-big-data/> (Big Data)