



Frédéric Gisbert
Cloud Solution Architect
Azure Data & Analytics

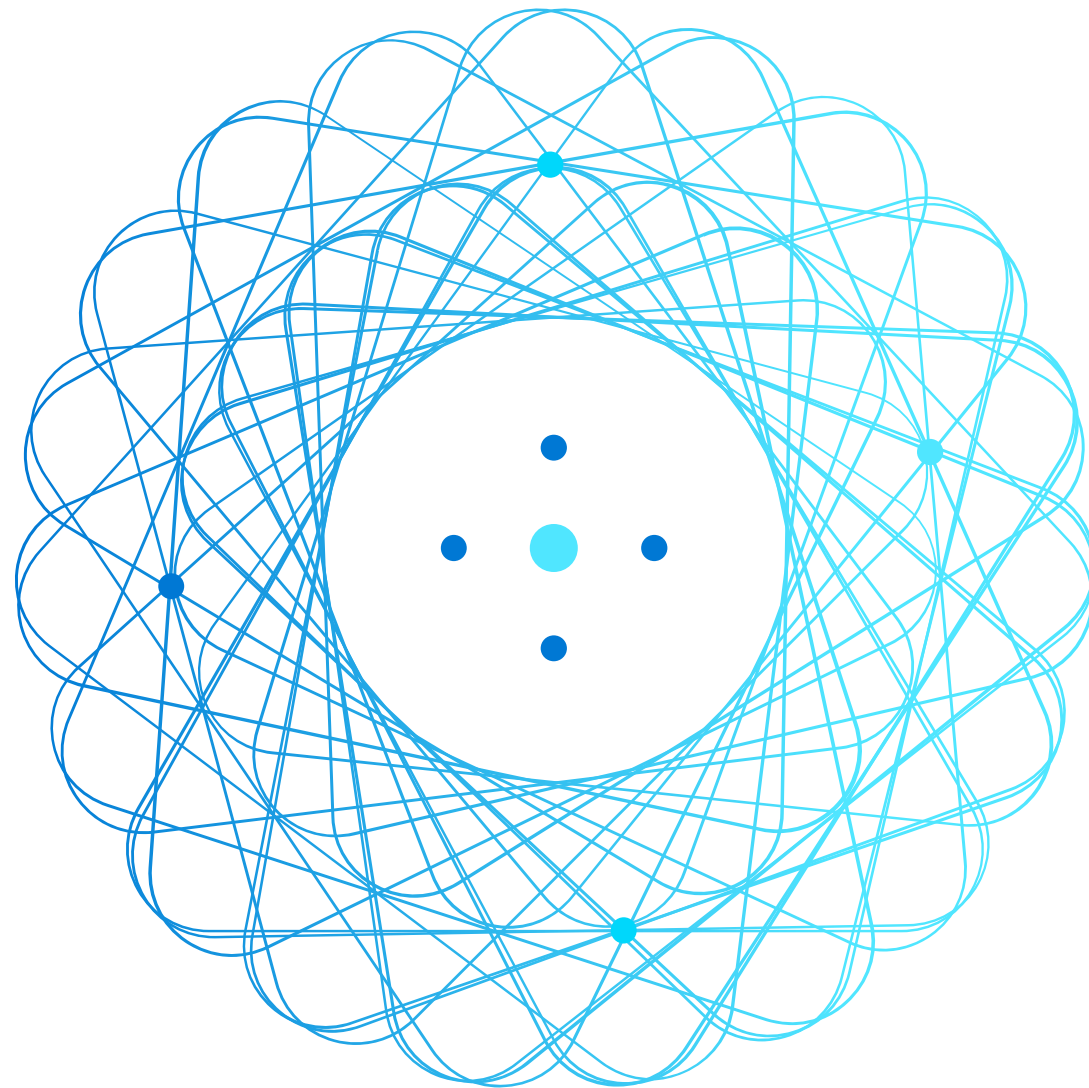


Ali Bouhaddou
Cloud Solution Architect
Azure Data & Analytics

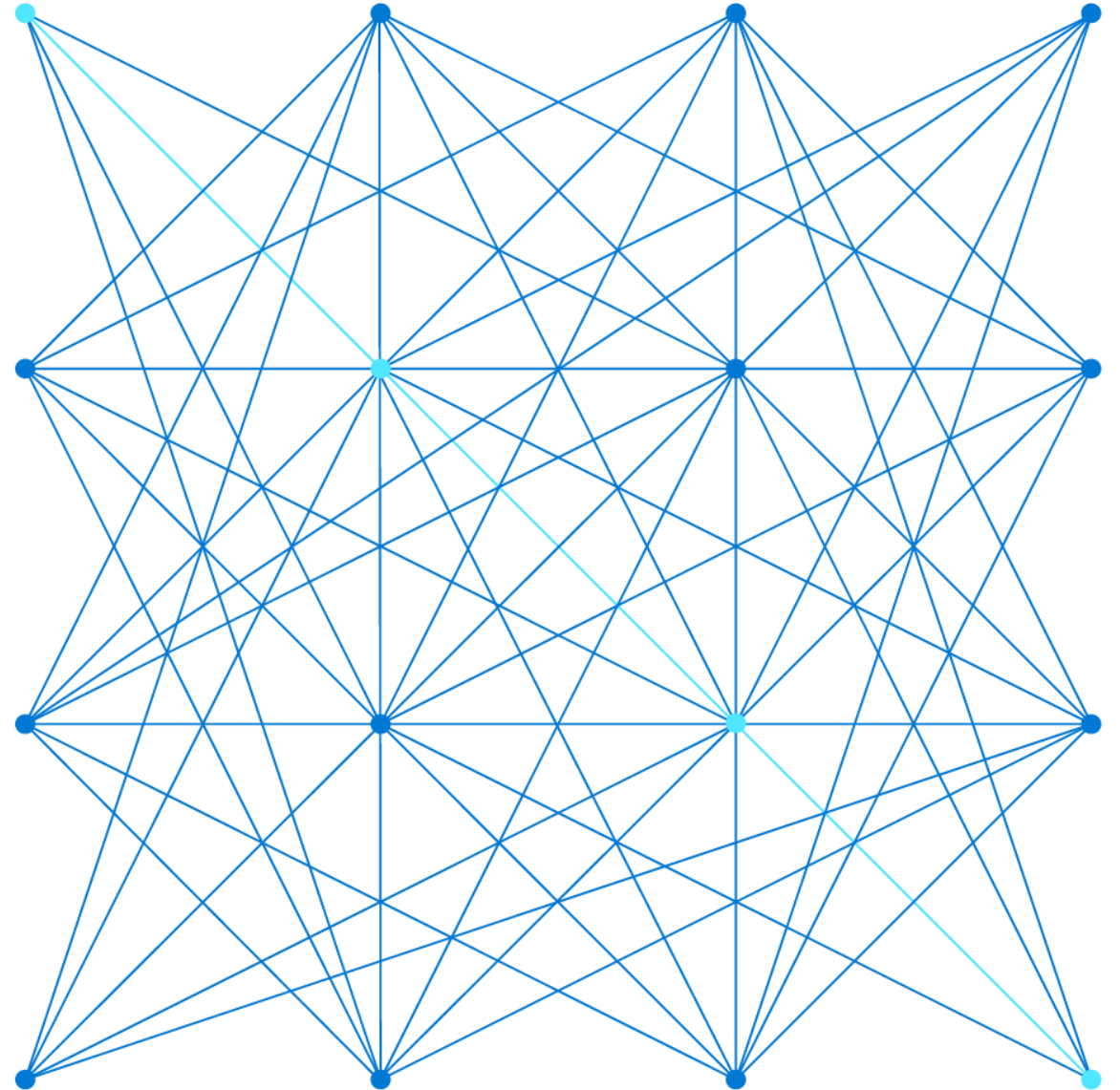
Training session DP900



- 1 Explore core data concepts**
- 2 Explore relational data in Azure**
- 3 Explore non relational Data in Azure**
- 4 Explore modern datawarehouse in Azure**



Module 1 : Explorer les concepts essentiels des données



Agenda



Explorer les concepts essentiels des données



Explorer les rôles et les responsabilités dans le monde des données



Décrire les concepts de données relationnelles

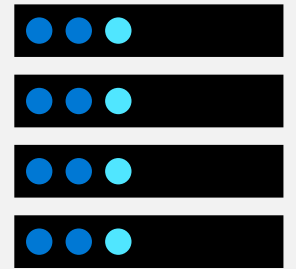


Explorer les concepts de données non relationnelles

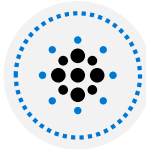


Explorer les concepts de l'analyse des données

Leçon 1 : Explorer les concepts essentiels des données



Leçon 1 - Objectifs



Identifier comment les données sont définies et stockées



Identifier les caractéristiques des données relationnelles et non relationnelles



Décrire et différencier les charges de travail de données



Décrire et différencier les données par lots et en flux continu

Que sont les données ?

Collection de faits, nombres, descriptions, objets, stockés de manière structurée, semi-structurée et non structurée

Données structurées

Tableau

Données semi-structurées

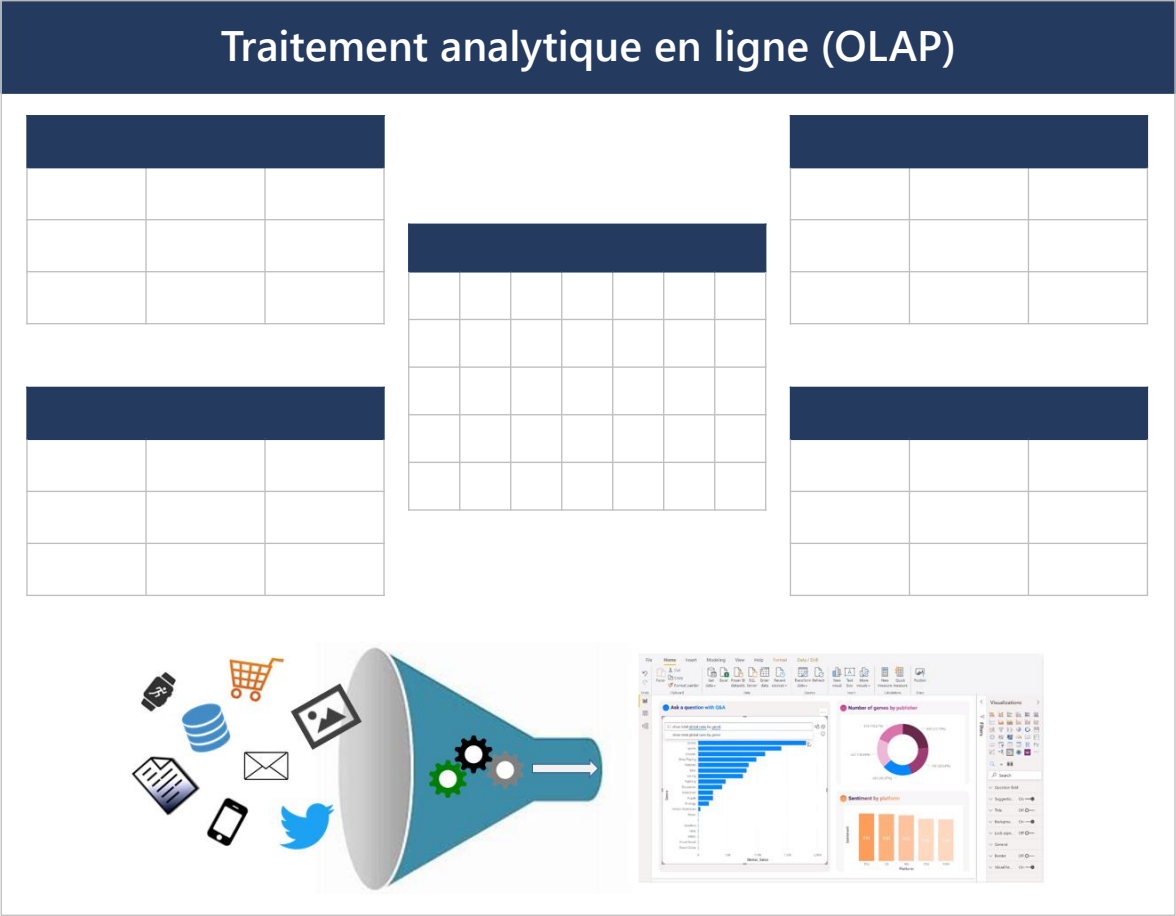
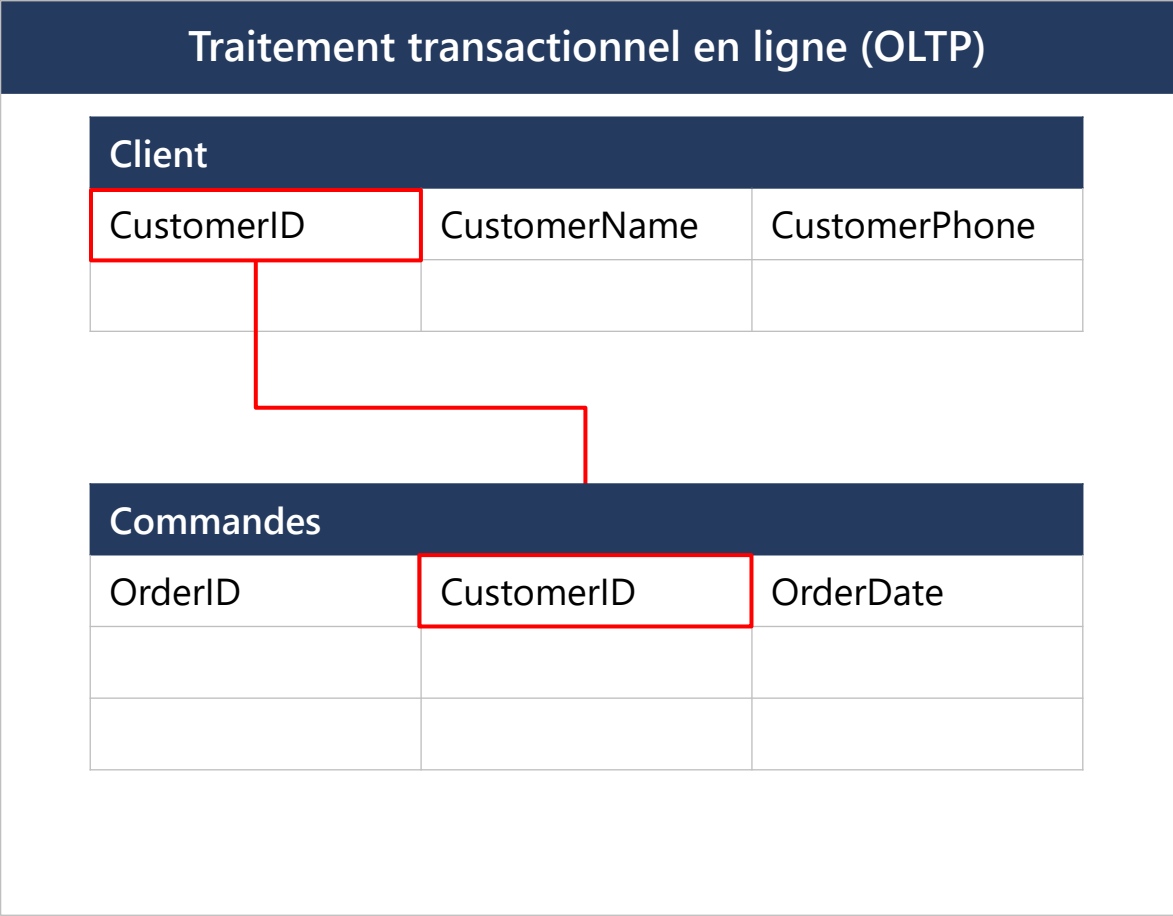
```
## Document 1 ## {
"customerID": "103248",
"name": { "first": "AAA",
"last": "BBB" }, "address": {
"street": "Main Street",
"number": "101", "city":
"Acity", "state": "NY" },
"ccOnFile": "yes",
"firstOrder": "02/28/2003" }
## Document 2 ## {
"customerID": "103249",
"name": { "title": "Mr",
"forename": "AAA",
"lastname": "BBB" },
"address": { "street":
"Another Street", "number":
"202", "city": "Bcity",
"county": "Gloucestershire",
"country-region": "UK" },
"ccOnFile": "yes" }
```

Données non structurées





Banques de données transactionnelles ou analytiques



Charges de travail transactionnelles

Client		
CustomerID	CustomerName	CustomerPhone

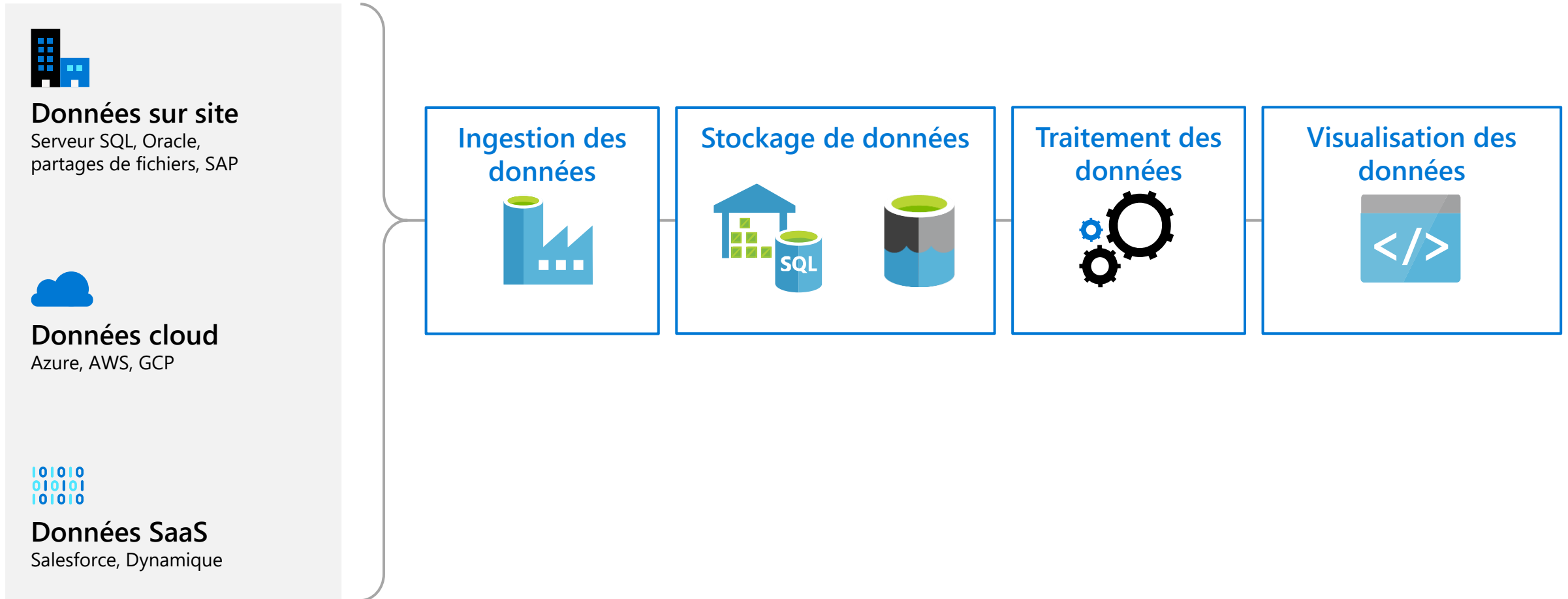
Commandes		
OrderID	CustomerID	OrderDate

Compte	
CustomerID	Solde
5558	1000
6023	1500

Transferts					
TransactionID	FromAccount	ToAccount	TransactionAmount	OrderDate	TransactionDescription
982801	6023	5558	500	JJ/MM/AA	Transférer 500 du compte 6023 vers le compte

```
BEGIN TRANSACTION
UPDATE Account
SET Balance = Balance -500
WHERE CustomerID=6023;
UPDATE Account
SET Balance = Balance +500
WHERE CustomerID=5558;
INSÉRER DANS Transferts (DeCompte, VersCompte, MontantTransaction, DescriptionTransaction)
VALEURS (6023,5558,500,'Transférer 500 du compte 6023 vers le compte 5558)
EXÉCUTER TRANSACTIONTransaction
```

Systeme analytique



Données par lots / Données en diffusion continue

Lot

 101010 101010 101010
010101 010101 010101
101010 101010 101010

 101010 101010 101010
010101 010101 010101
101010 101010 101010

 101010 101010 101010
010101 010101 010101
101010 101010 101010

Diffusion

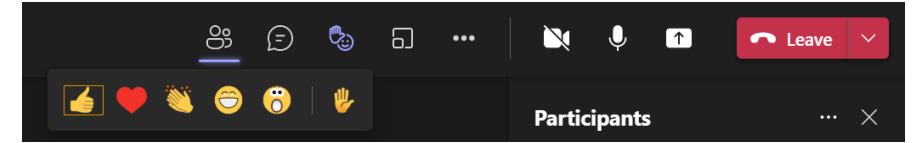
101010 101010 101010
010101 010101 010101
101010 101010 101010

101010 101010 101010
010101 010101 010101
101010 101010 101010

101010 101010 101010
010101 010101 010101
101010 101010 101010



Leçon 1 : Contrôle des connaissances



Comment les données d'une table relationnelle sont-elles organisées ?

- ☒ Lignes et colonnes
- ☐ En-tête et pied de page
- ☐ Pages et paragraphes



Parmi les propositions suivantes, laquelle est un exemple de données non structurées ?

- ☐ Une table Employé avec les colonnes ID d'employé, Nom d'employé et Désignation d'employé
- ☒ Fichiers audio et vidéo
- ☐ Une table dans la base de données SQL Server

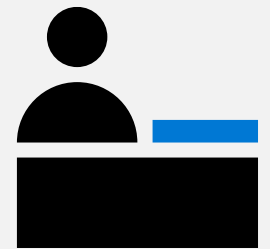


Parmi les propositions suivantes, laquelle est un exemple de jeu de données en flux continu ?

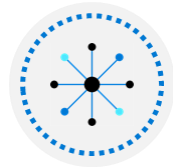
- ☒ Données des flux du capteur
- ☐ Données de vente du mois dernier
- ☐ Liste des employés travaillant pour une entreprise



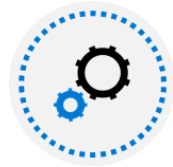
Leçon 2 : Explorer les rôles et les responsabilités dans le monde des données



Leçon 2 - Objectifs



Examiner les rôles de travail des données



Examiner les tâches et les outils courants pour les rôles de travail des données

Rôles dans les données

Administrateur de base de données

Gestion des bases de données

Implémente la sécurité des données

Sauvegardes

Accès utilisateur

Surveille les performances



Ingénieur de données

Pipelines de données et processus

Stockage de l'ingestion des données

Préparer des données pour l'analyse

Préparer les données pour le traitement analytique



Analyste de données

Fournit des informations sur les données

Rapports visuels

Modélisation des données pour l'analyse

Combine les données à des fins de visualisation et d'analyse



Outils courants : Administrateur de base de données

Azure Data Studio

Interface graphique pour la gestion des Data Services sur site et informatiques

Fonctionne sous Windows, macOS, Linux

SQL Server Management Studio

Interface graphique pour la gestion des Data Services sur site et informatiques

Fonctionne sous Windows

Outil d'administration de base de données complet

Portail Azure / CLI

Outils de gestion et d'approvisionnement d'Azure Data Services

Manuel et automatisation des scripts utilisant les scripts Azure Resource Manager ou d'interface de ligne de commande

Outils courants : Engineering données

Azure Synapse Studio

Portail Azure intégré pour gérer Azure Synapse

Ingestion des données (Azure Data Factory)

Gestion des actifs Azure Synapse (SQL Pools/Spark Pool)

SQL Server Management Studio

Interface graphique pour la gestion des Data Services sur site et informatiques

Fonctionne sous Windows

Outil d'administration de base de données complet

Portail Azure / CLI

Outils de gestion et d'approvisionnement des ressources Azure

Manuel et automatisation des scripts utilisant les scripts Azure Resource Manager ou d'interface de ligne de commande

Outils courants : Analyste de données

Power BI Desktop

Outil de visualisation des données
Modéliser et visualiser les données
Gestion des actifs Azure Synapse
(SQL Pools/Spark Pool)

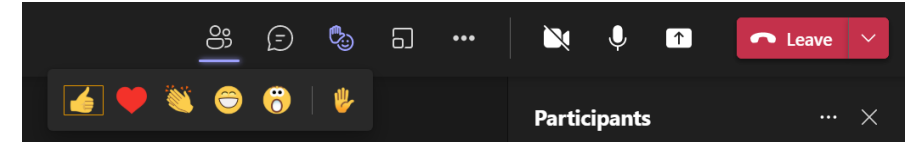
Portail Power BI / Service Power BI

Création et gestion des rapports
Power BI
Création de tableaux de bord
Power BI
Partager des rapports ou des jeux
de données

Power BI Report Builder

Outil de visualisation des données
pour les rapports paginés
Modéliser et visualiser des rapports
paginés

Leçon 2 : Contrôle des connaissances



Laquelle des tâches suivantes est un rôle d'administrateur de base de données ?

- ☒ Sauvegarde et restauration des bases de données
- ☐ Création de tableaux de bord et de rapports
- ☐ Identification des problèmes de qualité des données



Lequel des outils suivants est un outil de visualisation et de rapport ?

- ☐ SQL Server Management Studio
- ☒ Power BI
- ☐ Bases de données

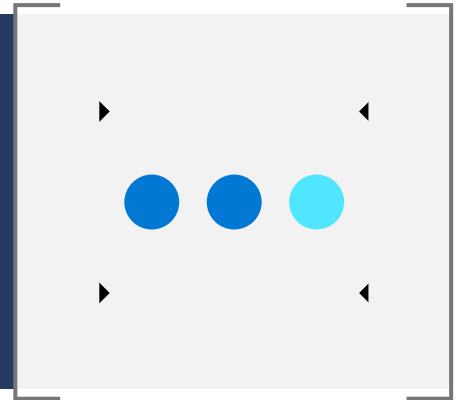


Lequel des rôles suivants n'est pas un rôle de tâche de données ?

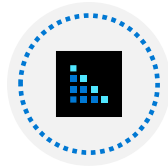
- ☒ Administrateur système
- ☐ Analyste de données
- ☐ Administrateur de base de données



Leçon 3 : Décrire les concepts de données relationnelles



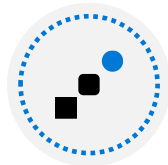
Leçon 3 - Objectifs



Découvrez les caractéristiques des données relationnelles



Définir des tables, des index et des vues



Découvrir les offres de charge de travail de données relationnelles dans Azure

Tables

Clients		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX
107	Francis Ribeiro	XXX-XXX-XXXX

Les données sont stockées dans une table

La table se compose de lignes et de colonnes

Toutes les lignes ont le même nombre de colonnes

Chaque colonne est définie par un type de données

Normalisation

Clients		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Commandes		
OrderID	CustomerName	CustomerPhone
AD100.	Noam Maoz	XXX-XXX-XXXX
AD101.	Noam Maoz	XXX-XXX-XXXX
AD102.	Noam Maoz	XXX-XXX-XXXX
AX103.	Qamar Mounir	XXX-XXX-XXXX
AS104.	Qamar Mounir	XXX-XXX-XXXX
AR105.	Claude Paulet	XXX-XXX-XXXX
MK106.	Muisto Linna	XXX-XXX-XXXX

Les données sont normalisées en :

Réduire le stockage

Éviter la duplication des données

Améliorer la qualité des données

Relations

Clients		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Commandes		
OrderID	CustomerID	IDVendeur
AD100.	101	200
AD101.	101	200
AD102.	101	200
AX103.	103	201
AS104.	104	201
AR105.	105	200
MK106.	105	201



Dans un schéma de base de données normalisé :

Les clés primaires et les clés étrangères sont utilisées pour définir des relations

Aucune duplication de données (autre que les valeurs clé dans le 3^e formulaire normal (3NF))

Les données sont récupérées en joignant des tables dans une requête

Index

Clients		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

IDX-RégionClient	
CustomerID	Région
100	France
101	Brésil
102	Croatie
103	Jordanie
104	Espagne
105	France
106	USA

Un index :

Optimise les requêtes de recherche pour une extraction de données plus rapide

Réduit la quantité de pages de données à lire pour récupérer les données dans une instruction SQL

Les données sont récupérées en joignant des tables dans une requête

Afficher

Clients		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Commandes		
OrderID	CustomerID	IDVendeur
AD100.	101	200
AD101.	101	200
AD102.	101	200
AX103.	103	201
AS104.	103	201
AR105.	105	200
MK106.	105	201
DB205.	100	205

Créer la définition d'une vue :

```
CREATE VIEW
vw_customerorders AS
SELECT Customers.CustomerID,
Customers.CustomerName,
Orders.OrderID FROM
Customers JOIN Orders on
Customers.CustomerID =
Orders.CustomerID
```

Récupérer les commandes passées par le client 102 à l'aide de la vue :

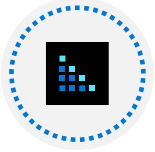
```
SELECT CustomerName, OrderID
from vw_customerorders WHERE
CustomerID=102
```

Une vue est une table virtuelle basée sur l'ensemble des résultats de la requête :

Des vues sont créées pour simplifier la requête

Combiner des données relationnelles dans une seule vue de volet

Leçon 3 : Contrôle des connaissances



Laquelle des affirmations suivantes est une caractéristique d'une base de données relationnelle ?

- ☐ Toutes les données doivent être stockées sous forme de chaînes de caractères
- ☒ Une ligne dans une table représente une seule entité
- ☐ Les différentes lignes d'une même table peuvent contenir plusieurs colonnes



Qu'est-ce qu'un index ?

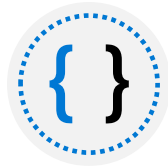
- ☒ Une structure qui vous permet de localiser rapidement des lignes dans une table, à l'aide d'une valeur indexée
- ☐ Une table virtuelle basée sur l'ensemble des résultats d'une requête
- ☐ Une structure comprenant des lignes et des colonnes permettant de stocker des données



Leçon 4 : Explorer les concepts de données non relationnelles



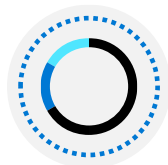
Leçon 4 - Objectifs



Découvrir les caractéristiques des données non relationnelles



Définir des types de données non relationnelles



Décrire NoSQL et les types de bases de données non relationnelles

Explorer les caractéristiques des données non relationnelles

Entités

```
## IDClient 1 : 1
Nom : Mark Hanson
Téléphone : [ Domicile : 1-999-9999999, Professionnel : 1-888-8888888, Portable : 1-777- 7777777 ]
Adresse : [ Domicile : 121 Main Street, Some City, NY, 10110,
            Entreprise 87 Big Building, Some City, NY, 10111 ]
## IDClient 2 : 2
Titre: M.
Nom : Jeff Hay
Téléphone : [ Domicile : 0044-1999-333333, Portable : 0044-17545-444444 ]
Adresse : [ Royaume-Uni : 86 High Street, Some Town, A County, GL8888, UK,
            USA : 777 7th Street, Another City, CA, 90111 ]
```

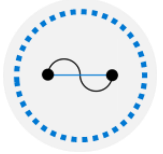
Les collections non relationnelles peuvent avoir :

Plusieurs entités dans la même collection ou le même conteneur, avec différents champs

Un schéma différent et non tabulaire

Sont souvent définies en étiquetant chaque champ avec le nom qu'il représente

Identifier les cas d'utilisation de bases de données non relationnelles



IoT et télématique :

Nécessite souvent d'ingérer de grandes quantités de données dans de fréquentes rafales d'activités, les données sont semi-structurées ou structurées et nécessitent souvent un traitement en temps réel



Vente au détail et marketing :

Scénarios courants pour les données distribuées à l'échelle mondiale, stockage de documents



Gaming :

Statistiques pendant le jeu, intégration des réseaux sociaux, classements, applications à faible latence



Web et mobile :

Fréquemment utilisé avec l'analyse des clics web et les applications modernes, y compris les bots

Types de données non relationnelles

Qu'est-ce que les données semi-structurées ?

La structure des données est définie dans les données réelles, par des champs. Les types de format/fichier sont les suivants :



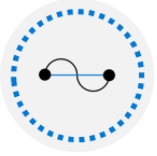
JSON

AVRO

ORC

Parquet

Qu'entend-on par « données non structurées » ?

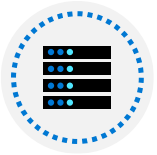


Par défaut, ne contient pas de champs :

Exemples : vidéo, audio, flux média, documents



Souvent utilisé pour extraire un formulaire de données et classer ou identifier des « structures »



Fréquemment utilisé en combinaison avec les fonctionnalités de Machine Learning ou Cognitive Services pour « extraire des données » en utilisant :

Analyse de texte

L'Analyse des sentiments, avec les API cognitives

L'API Vision

Qu'est-ce que NoSQL ?

Un terme générique, pour décrire des valeurs non relationnelles concernant les éléments suivants :



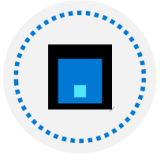
Bases de
données
clé-valeur

Document

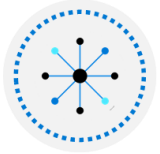
Bases de
données de
familles de
colonnes

Bases de
données de
graphiques

Quelles sont les fonctions d'une base de données de graphiques ?



Elle stocke des entités centrées sur les relations



Elle permet aux applications d'effectuer des requêtes traversant un réseau de nœuds et d'arêtes

Quelles sont les applications qui nécessitent une base de données de graphiques ?

Besoins de l'entreprise :

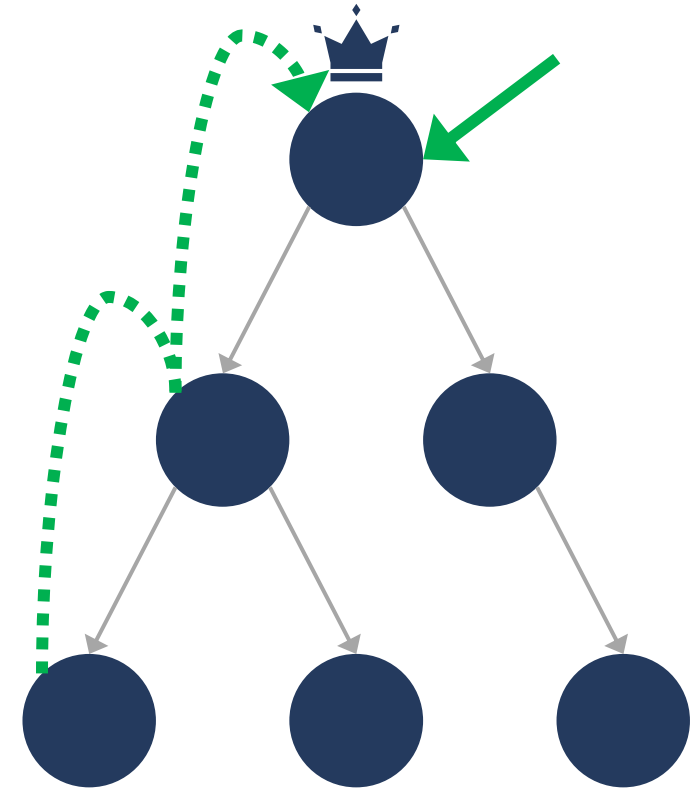
Applications OLTP avec des **données hautement corrélées**

Mises à jour **aisées** d'objets uniques ou de plusieurs objets

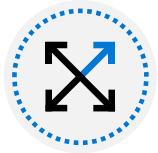
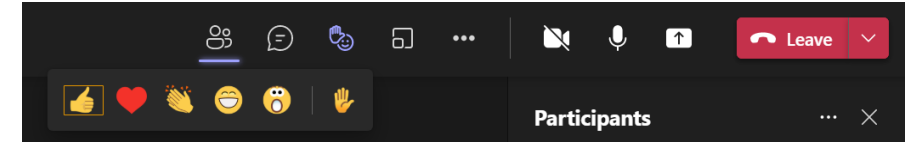
Modélisation **de données flexibles**

Exigences des données qui **évoluent**

Structures hiérarchiques des données



Leçon 4 : Contrôle des connaissances



Lequel des services suivants doit-on utiliser pour implémenter une base de données non relationnelle ?

- ☒ Azure Cosmos DB
- ☐ Base de données Azure SQL
- ☐ L'API Gremlin



Laquelle des caractéristiques suivantes s'applique aux bases de données non relationnelles ?

- ☐ Les bases de données non relationnelles contiennent des tables avec des enregistrements plats à colonne fixe
- ☐ Les bases de données non relationnelles requièrent l'utilisation de techniques de normalisation des données pour réduire la duplication des données
- ☒ Les bases de données non relationnelles sont sans schéma ou ont des schémas assouplis



Vous construisez un système qui contrôle la température dans un ensemble d'immeubles de bureaux et règle la climatisation dans chaque pièce de chaque bloc pour maintenir une température ambiante agréable. Votre système doit gérer la climatisation dans plusieurs milliers de bâtiments répartis à travers le pays ou la région, et chaque bâtiment contient généralement au moins 100 pièces climatisées. Quel est le type de stockage de données NoSQL le plus approprié pour capturer les données de température afin de permettre leur traitement rapide ?

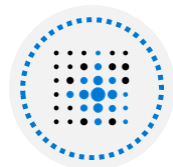
- ☒ Une base de données clé-valeur
- ☐ Une base de données de familles de colonnes
- ☐ Écrire les températures dans un objet blob dans Stockage Blob Azure



Leçon 5 : Explorer les concepts de l'analyse des données



Leçon 5 - Objectifs



En savoir plus sur l'ingestion et le traitement des données

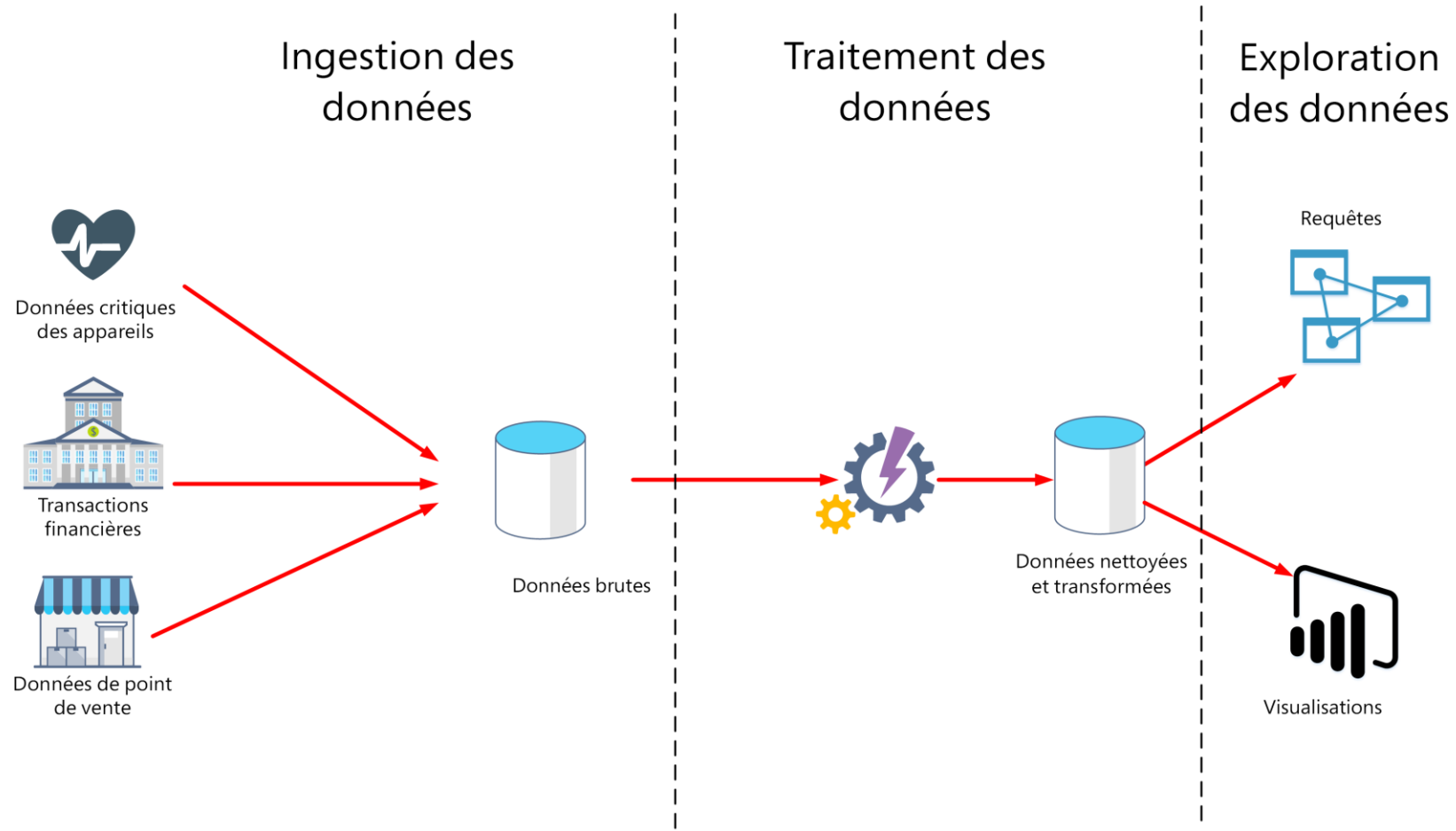


Découvrir la visualisation des données

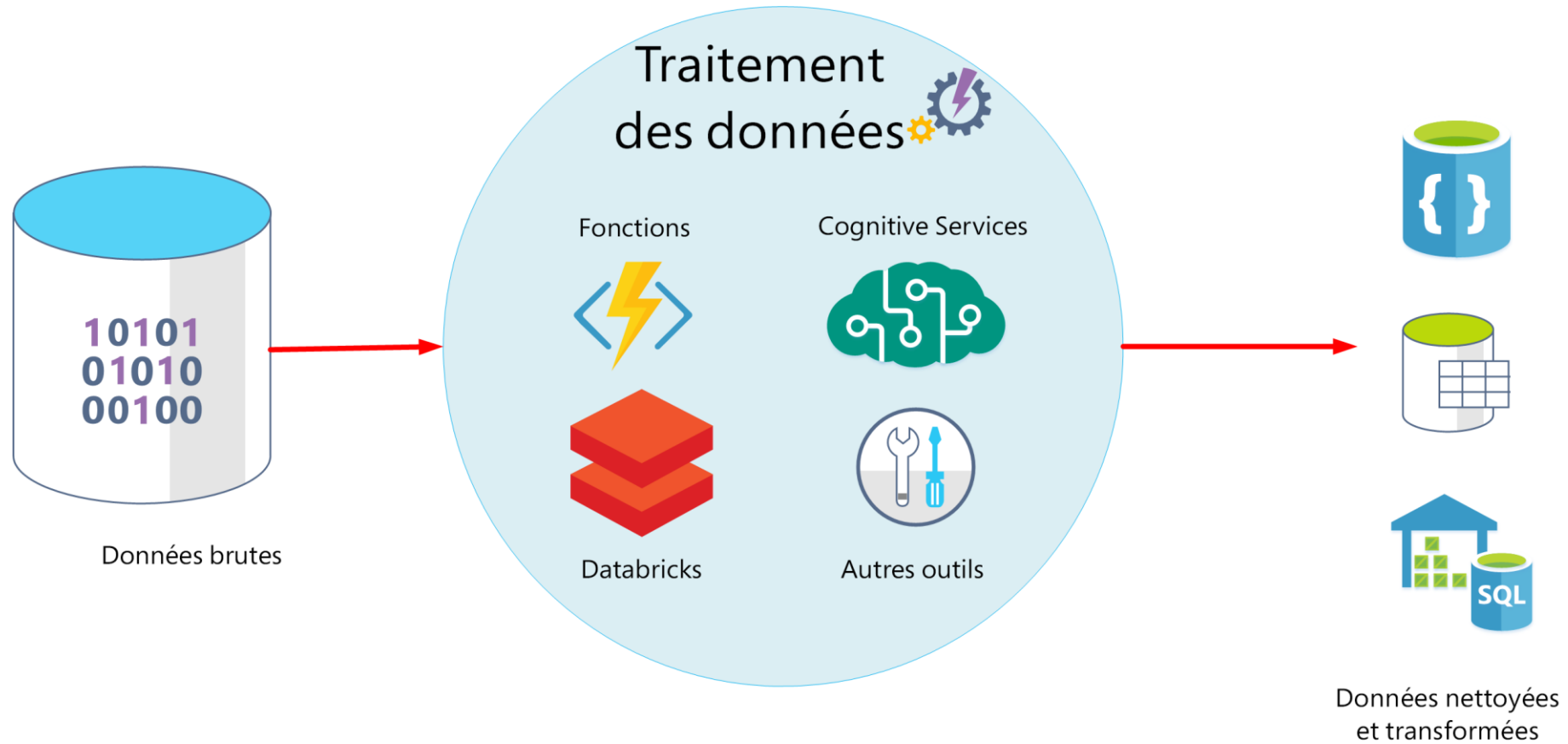


Découvrir l'analyse des données

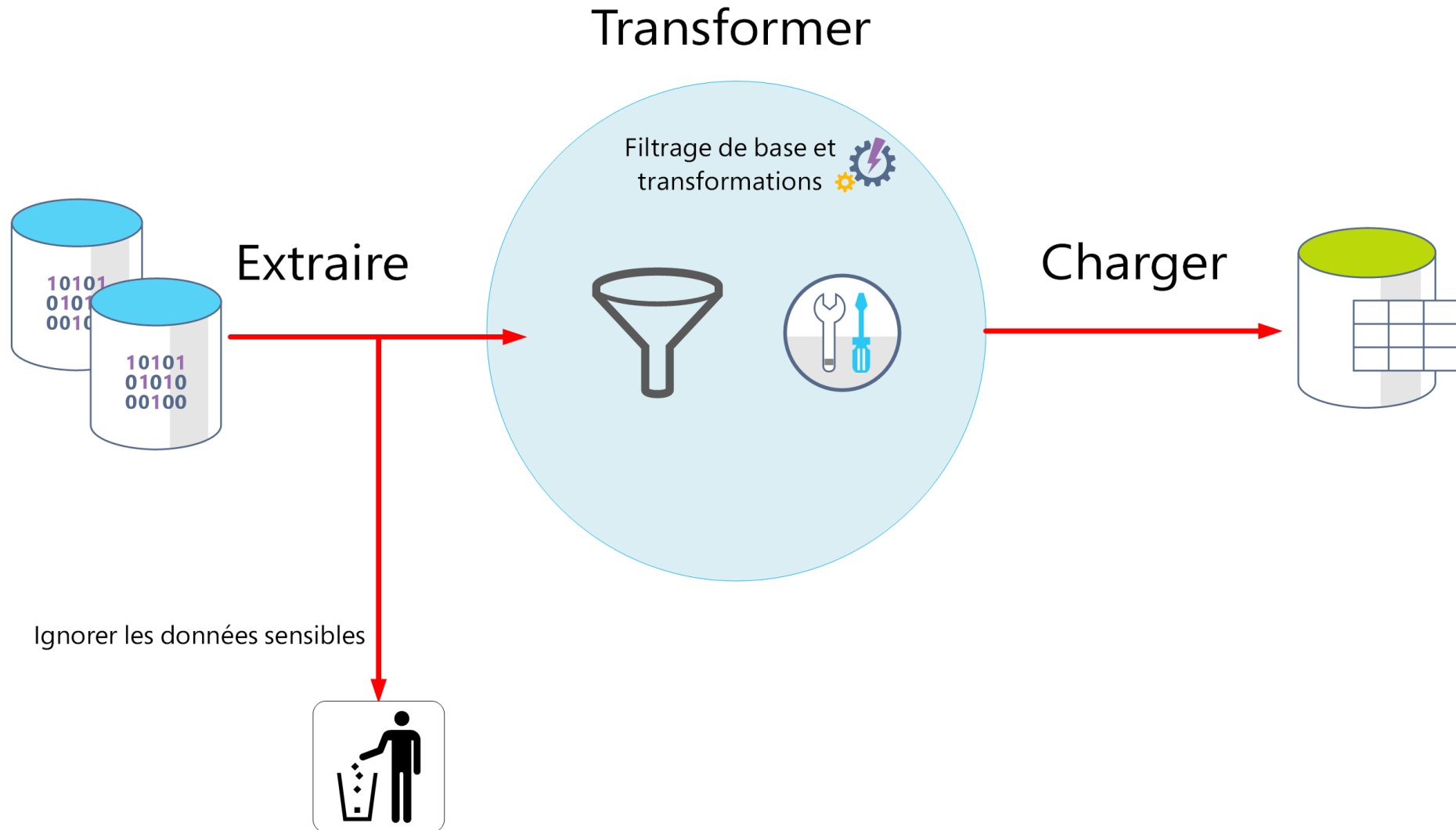
Qu'est-ce que l'ingestion de données ?



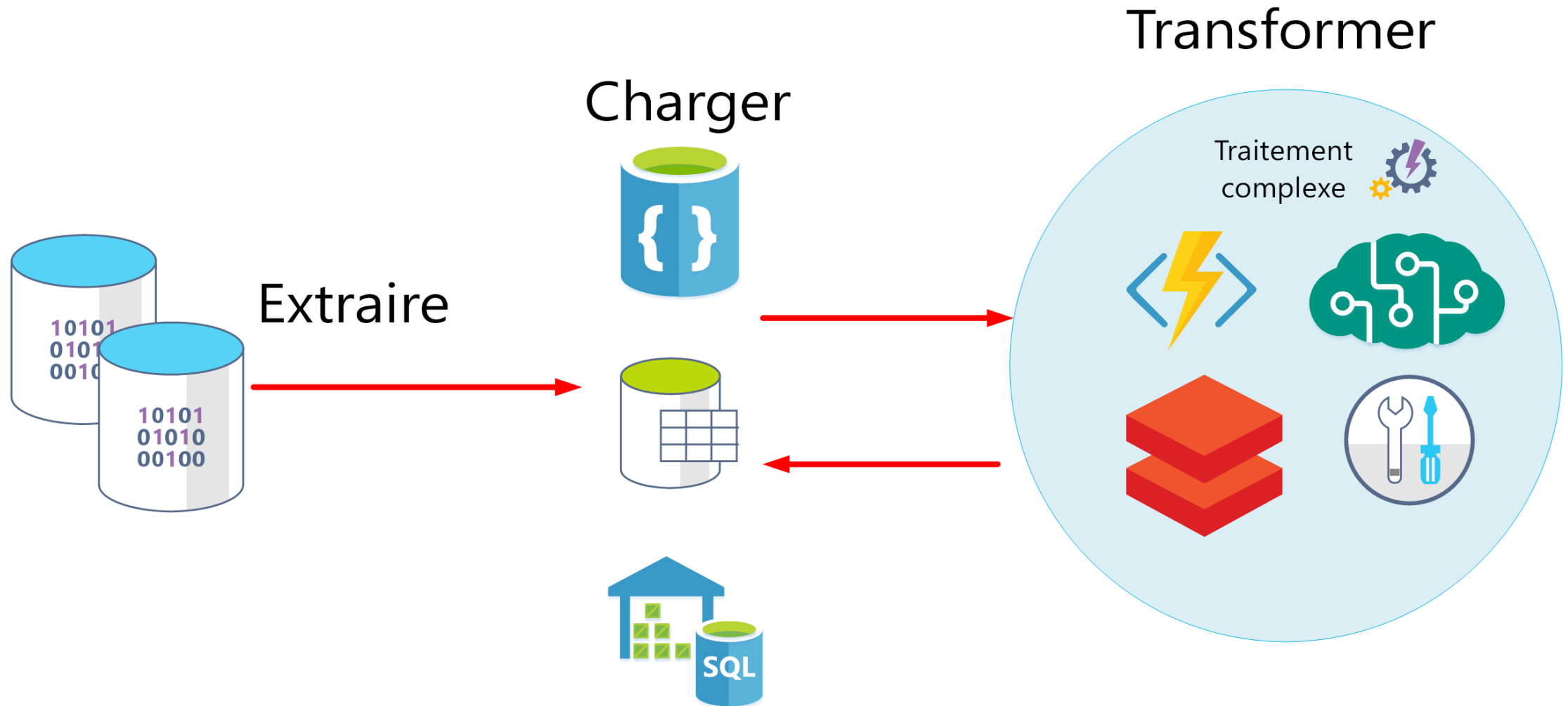
Qu'est-ce que le traitement des données ?



Qu'est-ce que le ETL ?

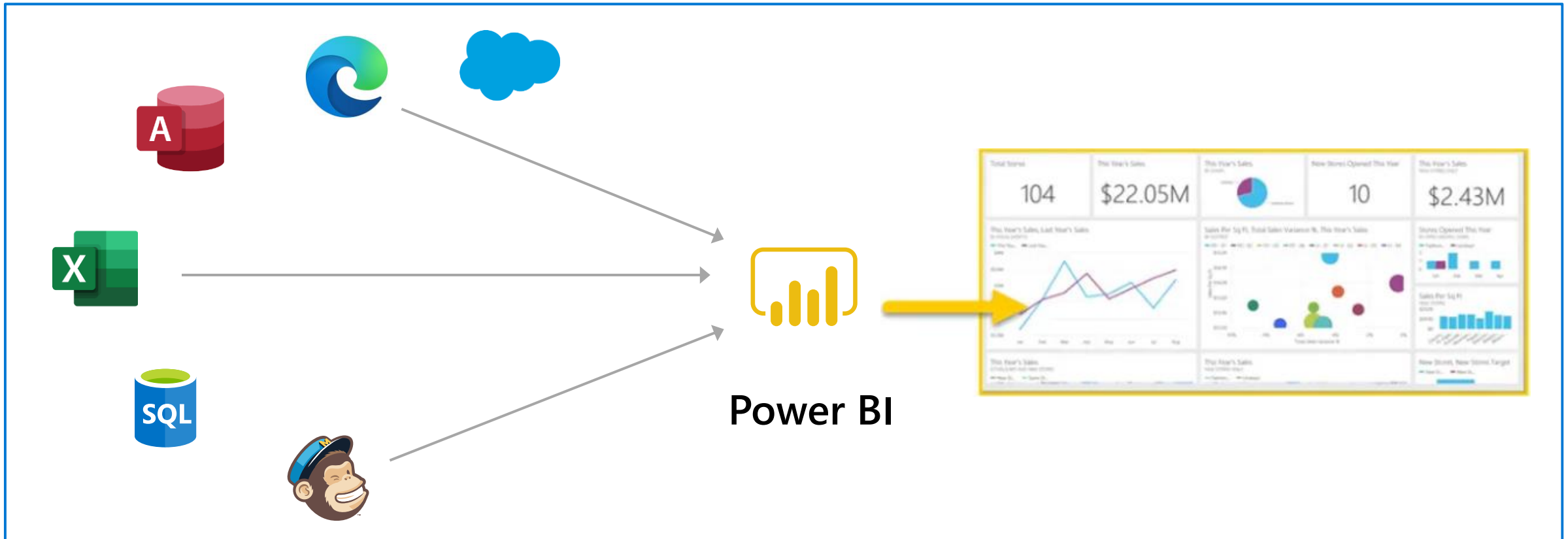


Qu'est-ce que le ELT ?



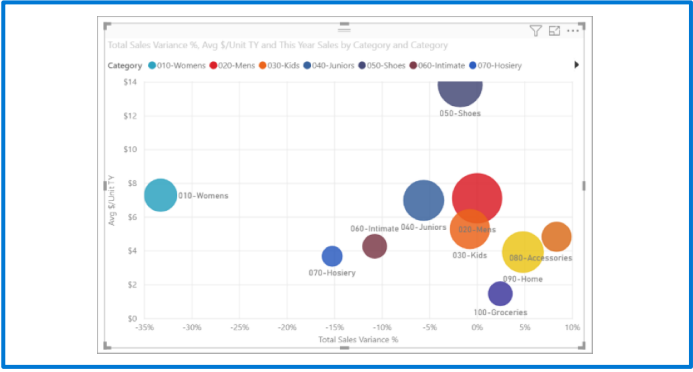
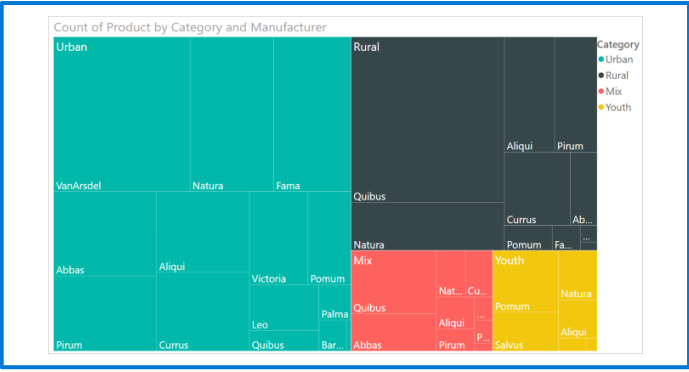
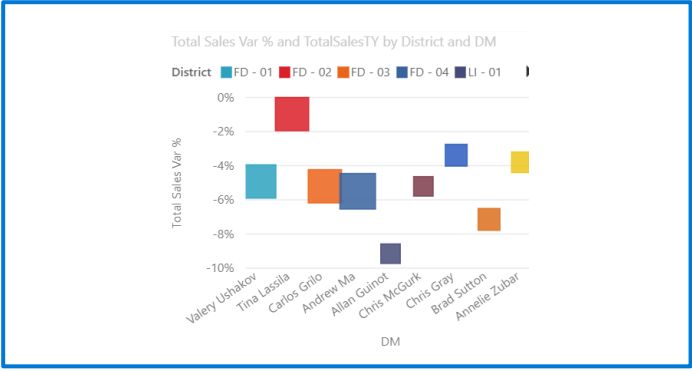
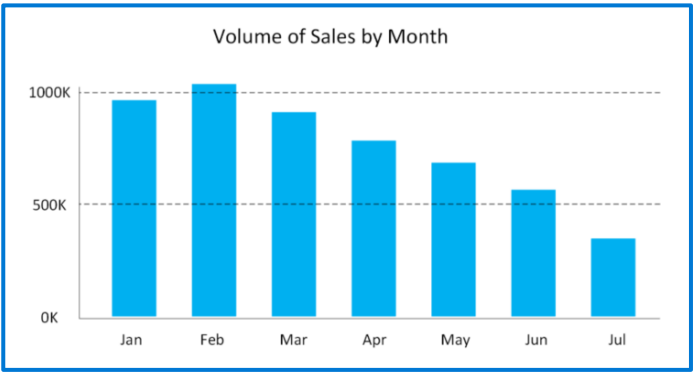
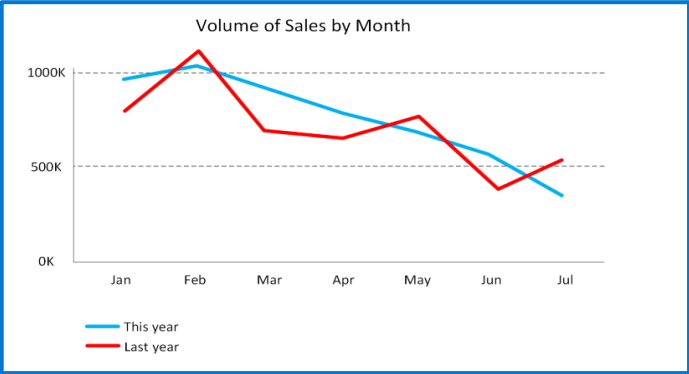
Découvrir la visualisation des données

Power BI : Collection de logiciels, services, applications et connecteurs



Power BI

Visualisations

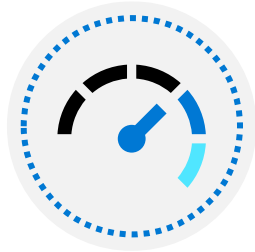


Quarter	Q1	YTD	Q2	YTD
Year	Revenue	Revenue	Revenue	Revenue
2015	\$45,186	\$45,186	\$70,609	\$115,795
2016	\$52,154	\$52,154	\$73,542	\$125,696
2017	\$51,388	\$51,388	\$68,149	\$118,537
2018	\$48,281	\$48,281	\$66,853	\$115,134
2019	\$53,145	\$53,145	\$49,135	\$102,280

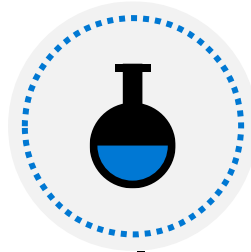
Découvrir l'analyse des données



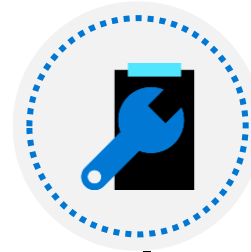
**Analyse
descriptive**



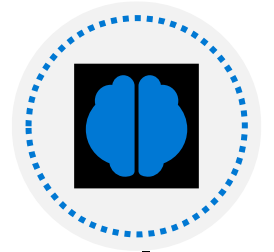
Diagnostic



**Analyse
prédictive**

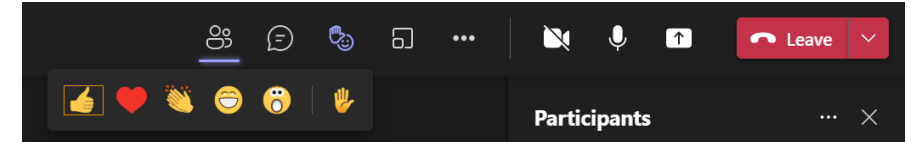


**Analyse
prescriptive**



**Analyse
cognitive**

Leçon 5 : Contrôle des connaissances



Qu'est-ce que l'ingestion de données ?

- ☐ Le processus de transformation des données brutes en modèles contenant des informations sensées
- ☐ Analyse des données pour les anomalies
- ☒ Capturer la diffusion en continu des données brutes à partir de diverses sources et les stocker



Lequel des visuels suivants affiche les principaux contributeurs à un résultat ou à une valeur sélectionnée ?

- ☒ Principaux influenceurs
- ☐ Colonne et graphique à barres
- ☐ Graphique matriciel



Quel type d'analyse permet de répondre aux questions sur ce qui a eu lieu dans le passé ?

- ☒ Analyse descriptive
- ☐ Analyse prescriptive
- ☐ Analyse prédictive



