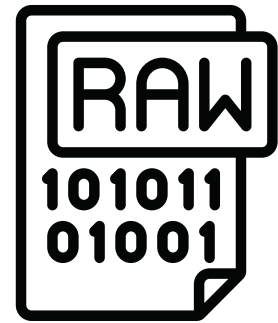


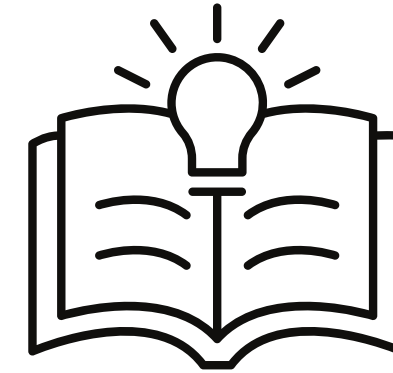
01 – Introduction

Infrastructure de données 1

Données



Faits, observations,
mesures brutes



Information et
connaissance

Types

- **Structurées** : tableaux, bases de données.
- **Semi-structurées** : JSON, XML
- **Non structurées** : textes, images, vidéos.

Bases de données

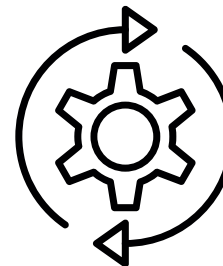
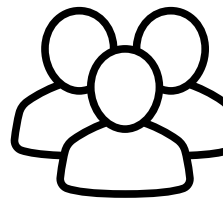
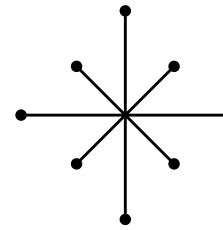
- **Base de données** : stockage structuré, performant, sécurisé.
- **Excel** : adapté aux petites données, mais limité en scalabilité, collaboration, et fiabilité.



Avantages bases de données

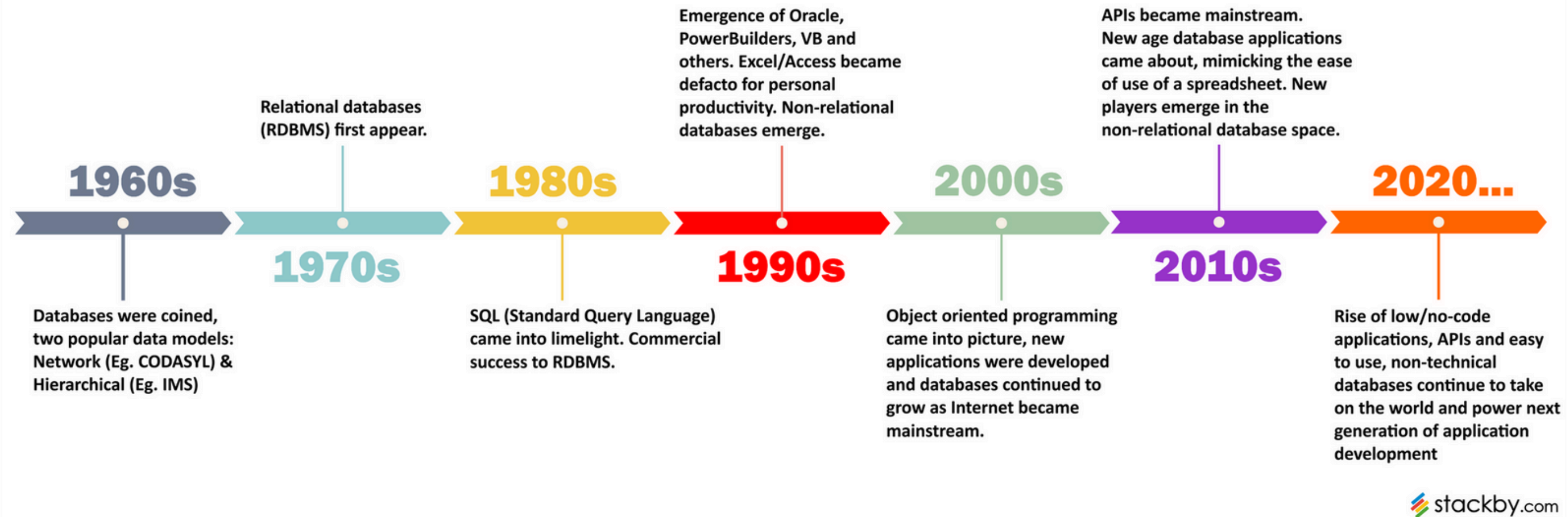
Avantages bases de données

- **Fiabilité** : contraintes d'intégrité, transactions
- **Scalabilité** : gestion efficace des grandes données
- **Multi-utilisateur** : accès concurrent, gestion des droits.
- **Sécurité** : permissions, cryptage, audit.
- **Automatisation** : requêtes SQL, procédures stockées.



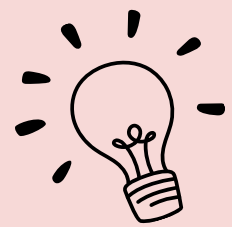
Histoire

History of Databases (1960-2020)

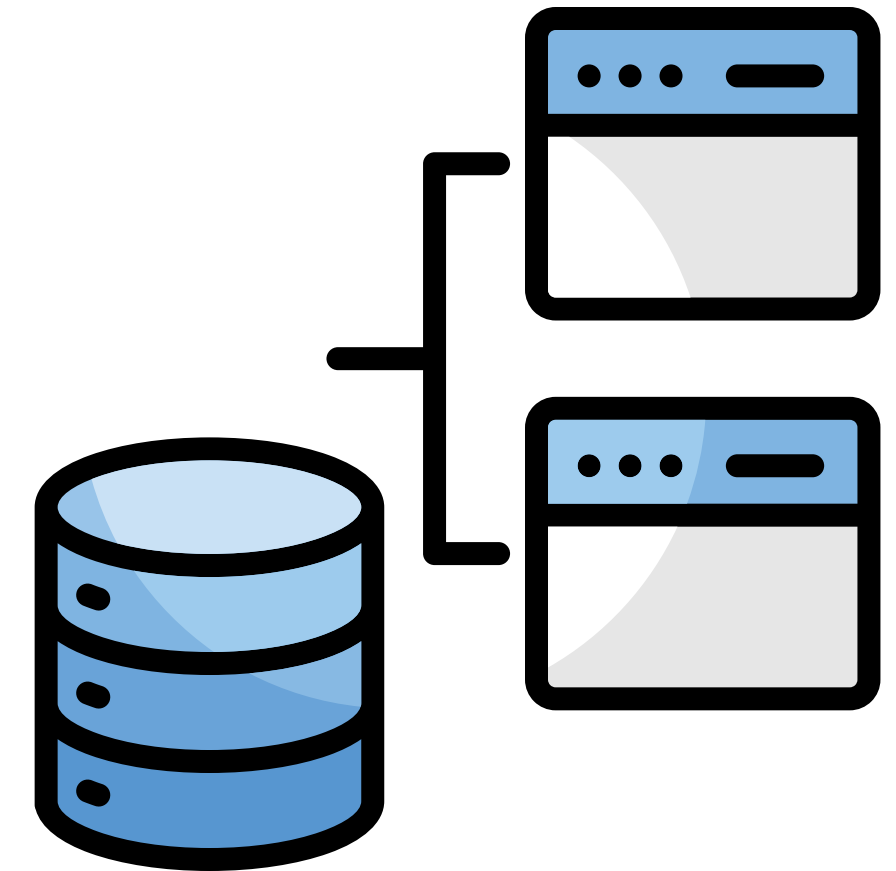


Base de données relationnelles

- **Tables (relations)** : lignes = enregistrements, colonnes = attributs.
- **Clés primaires et étrangères** : intégrité et relations entre tables.
- **Normalisation** : éviter la redondance, optimiser la structure.
- **SQL (Structured Query Language)** : interrogation et manipulation des données.



Base de données relationnelles toujours dominantes, malgré la montée des NoSQL et bases cloud.




OLTP vs OLAP

OLTP (Online Transaction Processing)

Bases optimisées pour les transactions rapides et fréquentes


- Petites transactions fréquentes.
- Réponse rapide et faible latence.
- Forte cohérence et intégrité des données.
- Accès concurrentiel multi-utilisateur.

 **Utilisations majeures:** systèmes bancaires, e-commerce, gestion des stocks, applications mobiles et web

OLAP (Online Analytical Processing)

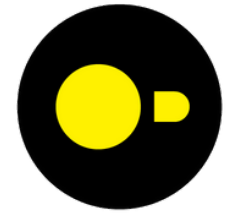
Bases optimisées pour l'analyse de grands volumes de données.

- Requêtes complexes et agrégations.
- Traitement de gros volumes de données historiques.
- Moins d'écritures, principalement des lectures.
- Performances optimisées pour les analyses avancées.

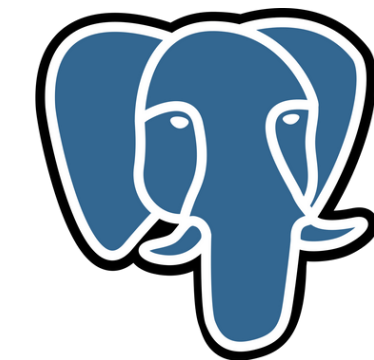
 **Utilisations majeures:** reporting d'entreprise, business intelligence, prévisions et analyses, tableaux de bords décisionnels

Systemes de Gestion de Base de Données (SGBD)

- **MySQL** : Open-source, rapide, très utilisé pour le web.
- **PostgreSQL** : Open-source, robuste, conforme aux standards SQL.
- **DuckDB** : Base analytique en mémoire, optimisée pour OLAP.
- **Autres** : Oracle, SQL Server, SQLite, MongoDB (NoSQL), BigQuery, Snowflake.



DuckDB



...



Choix selon performance,
scalabilité, et cas d'usage.

Objectifs du cours

- Connaître la problématique générale des **infrastructures** de données et du **stockage** de l'information en fonction des différents **types de données**.
- Maîtriser la transformation de la **modélisation conceptuelle** de données dans un **modèle physique**.
- Connaître les problématiques liées à l'**optimisation** d'une base de données relationnelles
- Pratiquer la mise en place d'une **base de données relationnelle**.

Evaluation

EXAMEN INTERMÉDIAIRE
30%

3 avril

Fondamentaux SQL

Format papier

PROJET
40%

Plus de détails dans les
slides d'après...

EXAMEN FINAL
30 %

13 juin

Ensemble du cours

Format papier

Projet

Objectif

Concevoir une **base de données relationnelle** pour un hôpital à partir d'un fichier Excel brut.

Contexte

L'hôpital dispose de plusieurs types de données :

- *Patients* : informations personnelles
- *Médecins* : spécialités, hôpitaux
- *Rendez-vous* : dates, motifs, patients et médecins
- *Médicaments* : nom, dosage, type
- *Prescriptions* : attribution de médicaments aux patients

Projet

Equipes

Groupes de 4/5 personnes

Rôles

Répartitions des rôles comme suit:

- **Chef de projet** : Coordination, documentation et interaction avec client
- **Database Architect**: Analyse et modélisation UML
- **Développeur SQL** : Création et gestion de la base de données
- **Data Engineer** : Rédaction et optimisation des requêtes SQL

Évaluation des pairs

Mesurer l'implication et la contribution de chaque étudiant dans le projet

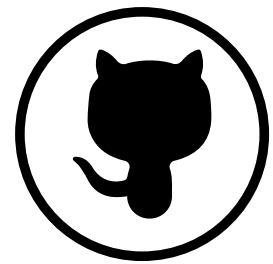
Projet

Livrables

- **Schéma UML** optimisé 7 mars
- **Script SQL** pour la création de la base de données et l'import des données 28 mars
- **Requêtes SQL** répondant aux besoins du client 2 mai
- **Présentation** orale du projet 5/6 mai
- **Rapport final** détaillant l'ensemble du projet et les différentes étapes du projet (schéma UML, scripts SQL, optimisation etc). 6 mai

Projet

Accès aux données



github.com/MediaComem/comem-infradon-1/

Aujourd'hui

- Création des groupes
- Première lecture des données Excel

Questions ?