

# Introduction à la BD

Pascal Lafourcade



Septembre 2022

Google

ALLOCINÉ

facebook

free



waze



IBM

INSEE  
INSTITUT NATIONAL  
DE LA STATISTIQUE  
ET DES ÉTUDES  
ÉCONOMIQUES



DEEZER

OVH

orange™

amazon

SNCF

{ BnF | Bibliothèque nationale de France

LA  
BANQUE  
POSTALE

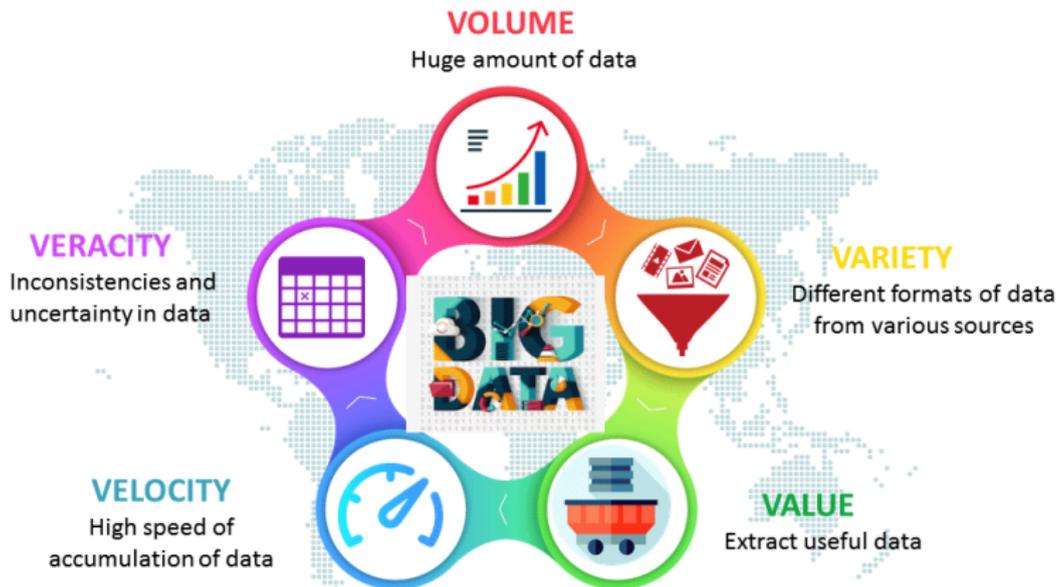
# Base de données



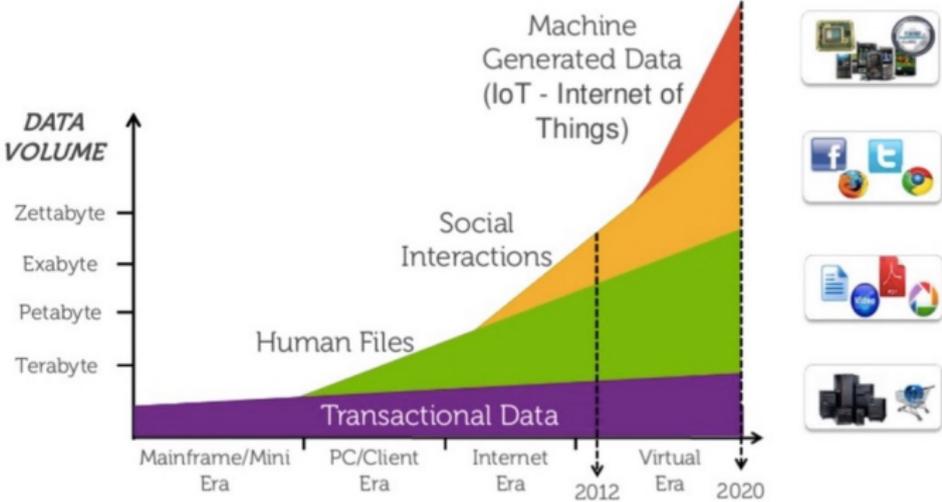
An abstract visualization of data connections. The background is dark, and the scene is filled with a dense network of glowing lines and nodes. The lines are primarily orange and yellow, with some cyan and blue lines interspersed. The nodes are small, glowing circles in orange and cyan. The overall effect is a complex, interconnected web of data points, suggesting a large-scale network or data flow.

# **BIG DATA**

# Les 5 V du Big Data



# Les 5 V : Volume



# 5 V : Vitesse



## 2019 This is What Happens in An Internet Minute





# 5 V : Véracité

## Sources of Data Veracity



Statistical biases



Lack of data lineage



Software bugs



Noise



Abnormalities



Information Security



Untrustworthy data sources



Falsification



Uncertainty and ambiguity of data



Duplication of data



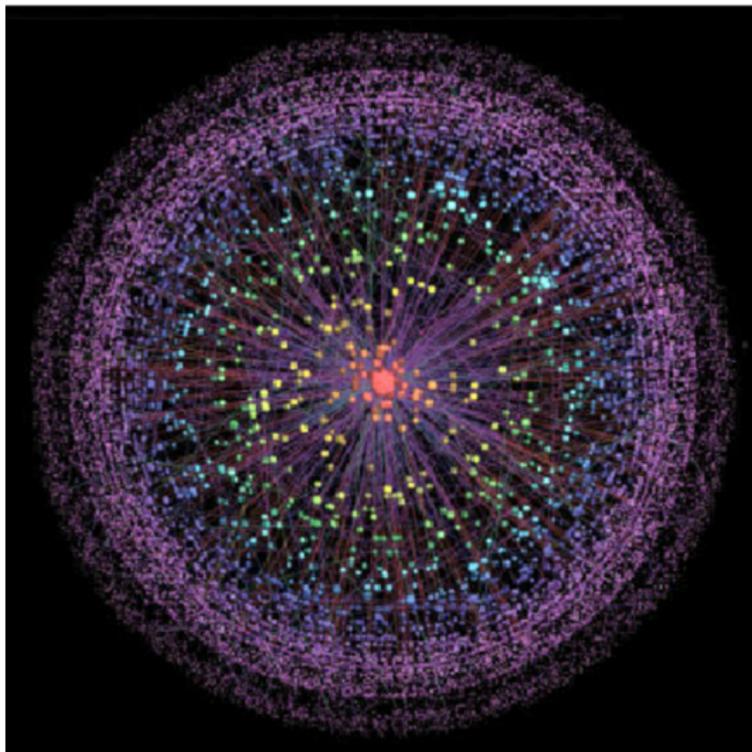
Out of date and obsolete data



Human error



# Visualisation





## Règlement Général sur la Protection des Données

# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

Les données

Introduction à Postgre SQL

NoSQL

Conclusion

# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

Les données

Introduction à Postgre SQL

NoSQL

Conclusion

## Équipe pédagogique (50% industriels)

- ▶ **Jules Azemar**
- ▶ Pascale Brigoulet
- ▶ **Raphaël Delage**
- ▶ Anaïs Durand
- ▶ Sylvie Guillaume
- ▶ **Franck Glaziou**
- ▶ Pascal Lafourcade
- ▶ **Johanna Millet**
- ▶ Léo Robert
- ▶ **Julie Rossignol**

# Organisation des 2 années du BUT

- ▶ S1 : 2h CM, 16h TD, 28h TP
- ▶ S2 : 8h CM, 32h TP
- ▶ S3 : 8h TD, 20h TP
- ▶ S4 : 4h TD, 10h TP

## Semestre 1 : R1.05

- ▶ Approche de la conception des bases de données : modèle conceptuel de données et traduction vers le modèle relationnel + contraintes simples + redondance
- ▶ Algèbre relationnelle
- ▶ Base du SQL (Structured Query Language) : langage de manipulation de données, langage de définition de données
- ▶ Éléments sur les jeux de tests + jeux de données
- ▶ Principes et utilisation d'un SGBD
- ▶ Utilisation d'Atelier de Génie Logiciel
- ▶ Formulaire et état

## Semestre 1

S1 : 2h CM + 1h TD + 2h TP

S2 : 1h TD + 2h TP

S3 : 1h TD + 2h TP

S4 : 1h TD + 2h TP

S5 : 1h TD + 2h TP

S6 : 1h TD + 2h TP

S7\* : 1h TD + 2h TD + 2h TP

E1 : 2h **Examen**

S8\* : 1h TD + 2h TP

S9\* : 1h TD + 2h TP

S10\* : 1h TD + 2h TP

S11\* : 1h TD + 2h TP

S12\* : 1h TD + 2h TP

S13\* : 1h TD + 2h TP

S14 : 1h TD + 2h TD + 2h TP

E2 : 2h **Examen**

## Semestre 2 : R2.06

- ▶ SQL avancé (Structured Query Language)
- ▶ Visualisation de données
- ▶ 1er niveau de l'administration des SGBD : utilisateurs, rôles, droits

## Semestre 2

S1 : 1h CM + 2h TP

S2 : 2h TP

S3 : 1h CM + 2h TP

S4 : 2h TP

S5 : 1h CM + 2h TP

S6 : 2h TP

S7\* : 1h CM + 2h TP + 2h TP\*

E1 : 2h **Exam**

S8\* : 1h CM + 2h TP

S9\* : 2h TP + 2h TP\*

S10\* : 1h CM + 2h TP

S11\* : 2h TP

S12\* : 1h CM + 2h TP

S13\* : 2h TP

S14 : 1h CM + 2h TP

E2 : 2h **Exam**

## 1ère année

- ▶ S1.04 : Conception d'une BD : gestion de stock de votre choix
- ▶ S2.04 : Visualisation des données de votre choix

### **Groupe de 4 étudiants**

## Rendu la semaine 13 de chaque semestre

- ▶ Mini rapport
- ▶ Oral par groupe : 15 minutes + 10 minutes questions

Oral du second semestre en **ANGLAIS !**

## SAÉ S1.04 : Conception

Vous travaillez dans l'entreprise X. Vous devez revoir le système d'information concernant la **gestion des stocks de ses produits** (vente auprès des clients et réapprovisionnement auprès des fournisseurs) en vue d'améliorer le processus métier. Les produits sont vendus en France, dans l'Union Européenne (UE) et hors UE.

### Productions

- ▶ Jeu de données
- ▶ Modèle de données
- ▶ Rapport sur l'importance de ces données en entreprise
- ▶ Scripts création base de données

### Évaluation

- ▶ MCD : 4 pts
- ▶ MLD : 3 pts
- ▶ scripts + jeu de données : 5 pts
- ▶ Données : 4 pts requêtes + 4 pts données économiques

## SAÉ S2.04 : Visualisation

À partir de données ouvertes et accessibles de votre choix, il faut les intégrer dans la mise en place d'une nouvelle entreprise : étude de marché, mise en place d'un produit, étude de la concurrence.

### Productions

- ▶ Étude des données et visualisation des informations
- ▶ Modèle de données
- ▶ Présentation orale des résultats en anglais
- ▶ Scripts de création de base de données

### Évaluation

- ▶ MLD + MCD : 4 pts
- ▶ Script + jeu de données : 4 pts
- ▶ Visualisation : 3 pts
- ▶ Statistiques : 5 pts
- ▶ Exposé en anglais : 2 pts
- ▶ Gestion d'entreprise : 2 pts

# Évaluation

## Notation S1 pour BD1

- ▶ 50 % DS1 + 50 % DS2

## Notation S2 pour BD2

- ▶ 50 % DS1 + 50 % DS2

# Plan

Motivations

La BD à l'IUT

**Base de données**

Modélisation

Les données

Introduction à Postgre SQL

NoSQL

Conclusion

# Le problème

Comment **stocker, utiliser, mettre à jour**, ...

...un grand nombre de données ?

## Exemples

1. les emprunts dans une bibliothèque,
2. les salaires des employés dans une entreprise,
3. les comptes dans une banque,
4. les réservations de places dans les trains, ...



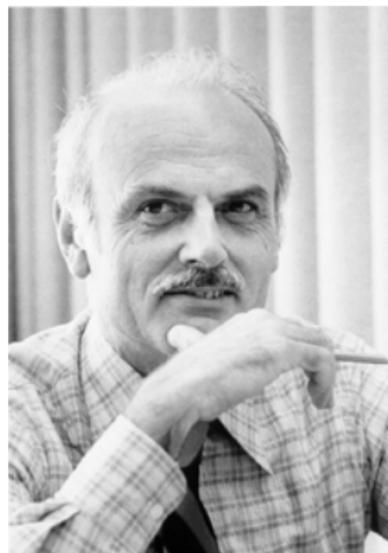
{BnF | Bibliothèque  
nationale de France



# Vocabulaire

- ▶ **BD**, base de données (DB, database).
- ▶ **SGBD**, système de gestion de bases de données (*DBMS, database management system*).
  - ▶ L'ensemble des outils (logiciels) permettant **l'organisation, le contrôle, la consultation et la modification** d'une base de données.
- ▶ **SGBDR**, système de gestion de base de données *relationnelle* (RDBMS, relational database management system)
  - ▶ un SGBD, qui en plus gère les relations, c'est-à-dire qu'on peut définir des contraintes qui garantissent l'intégrité référentielle et fonctionnelle des données.
- ▶ **Requête** (*query*) est une interrogation d'une base de données pour obtenir une certaine information.

## Modèle relationnel de données



Prix Turing en 1981

Modèle le plus utilisé, inventé par **Edgar Frank Codd** en 1970

## SQL, *Structured Query Language*.

SQL a été adopté comme **norme internationale** par l'ISO en 1987.

Révisions, en 1992 (SQL2), 1999 (SQL3), et en 2003, 2008 et 2011.

**Remarque** : SQL n'est pas un langage de programmation !

Pas de condition ni de boucle,  
mais s'intègre dans un langage de programmation.

En 1979, **Relational Software, Inc.** (actuellement **Oracle Corporation**) présente la **première version commercialement disponible de SQL** : **Oracle Database**.

Il y a de légères différences entre les différentes versions de SQL selon les SGBDR (e.g., MySQL, PostgreSQL, SQL Server, ...).

# Algèbre Relationnelle

Définit et étudie les opérations sur les **relations**.

Une **base de données relationnelle** est une base de données structurée suivant les principes de l'algèbre relationnelle.

# Structure d'un SGBD

# Structure d'un SGBD

1. LMD, langage de manipulation des données (DML, Data Manipulation Language) : pour consulter ou modifier le **contenu** de la base.

# Structure d'un SGBD

1. LMD, langage de manipulation des données (DML, Data Manipulation Language) : pour consulter ou modifier le **contenu** de la base.
2. LCT, langage de contrôle des **transactions** (TCL, Transaction Control Language) : pour gérer les transactions.

Une transaction regroupe une série de modifications de la base dans une seule opération logique.

# Structure d'un SGBD

1. LMD, langage de manipulation des données (DML, Data Manipulation Language) : pour consulter ou modifier le **contenu** de la base.
2. LCT, langage de contrôle des **transactions** (TCL, Transaction Control Language) : pour gérer les transactions.

Une transaction regroupe une série de modifications de la base dans une seule opération logique.

3. LDD, langage de définition des données (DDL, Data Definition Language) : pour créer ou modifier la **structure** de la base.

# Structure d'un SGBD

1. LMD, langage de manipulation des données (DML, Data Manipulation Language) : pour consulter ou modifier le **contenu** de la base.
2. LCT, langage de contrôle des **transactions** (TCL, Transaction Control Language) : pour gérer les transactions.

Une transaction regroupe une série de modifications de la base dans une seule opération logique.

3. LDD, langage de définition des données (DDL, Data Definition Language) : pour créer ou modifier la **structure** de la base.
4. SQL procedural (PL/SQL pour Oracle Database) : ensemble d'outils pour développer des procédures, déclencheurs (**triggers**) et fonctions utilisateurs (UDF : User Define Function) et pour que SQL s'**interface** avec des langages hôtes.

# Structure d'un SGBD

1. LMD, langage de manipulation des données (DML, Data Manipulation Language) : pour consulter ou modifier le **contenu** de la base.
2. LCT, langage de contrôle des **transactions** (TCL, Transaction Control Language) : pour gérer les transactions.

Une transaction regroupe une série de modifications de la base dans une seule opération logique.

3. LDD, langage de définition des données (DDL, Data Definition Language) : pour créer ou modifier la **structure** de la base.
4. SQL procedural (PL/SQL pour Oracle Database) : ensemble d'outils pour développer des procédures, déclencheurs (**triggers**) et fonctions utilisateurs (UDF : User Define Function) et pour que SQL s'**interface** avec des langages hôtes.
5. LCD, langage de contrôle des données (DCL, Data Control Language) : pour gérer les **privilèges**, *i.e.*, les utilisateurs et les actions qu'ils peuvent entreprendre.

# Logiciels

ORACLE®



Microsoft®  
SQL Server™



SQLite



mongoDB.



*cassandra*



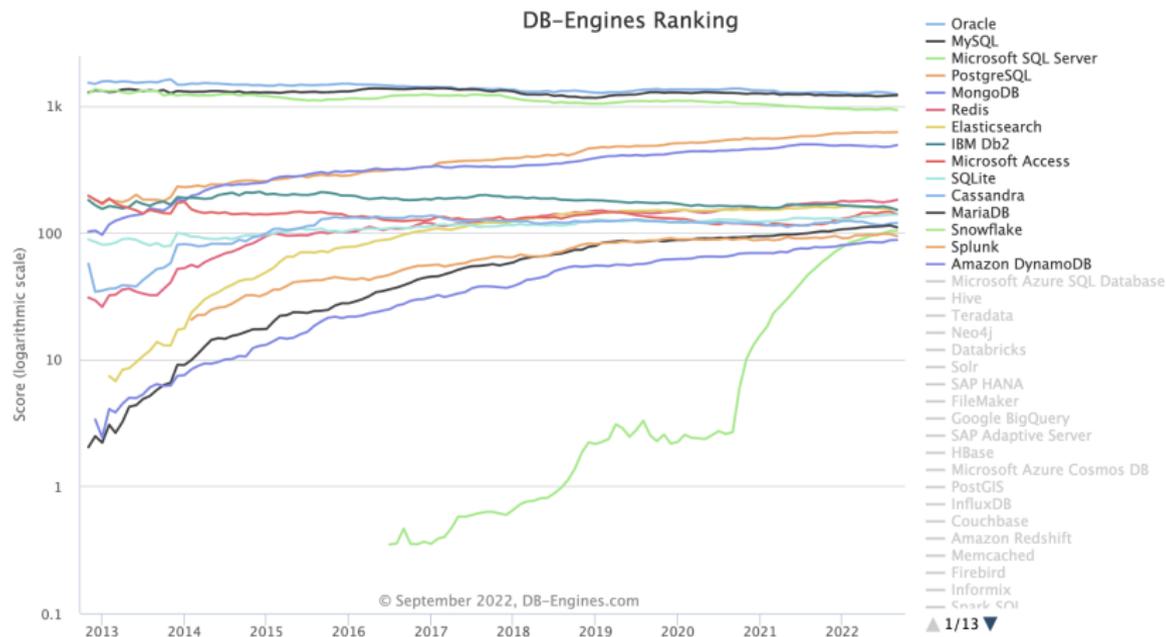
# Logiciels

ORACLE®

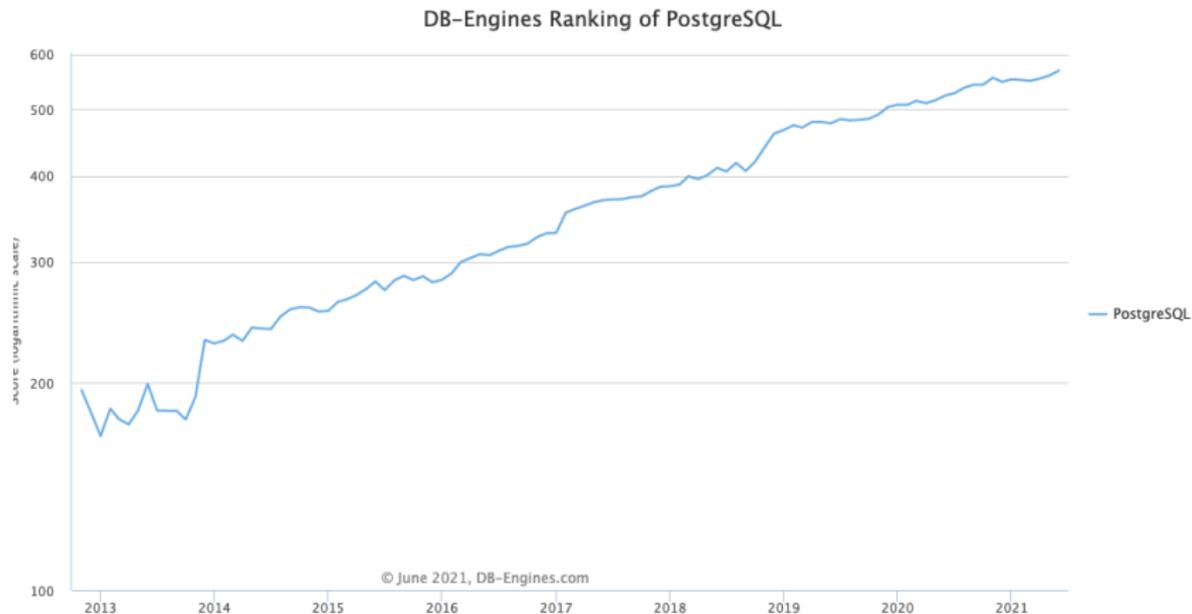
Microsoft®  
SQL Server®



# SGBD évolution



# SGBD évolution



# Plan

Motivations

La BD à l'IUT

Base de données

**Modélisation**

Les données

Introduction à Postgre SQL

NoSQL

Conclusion

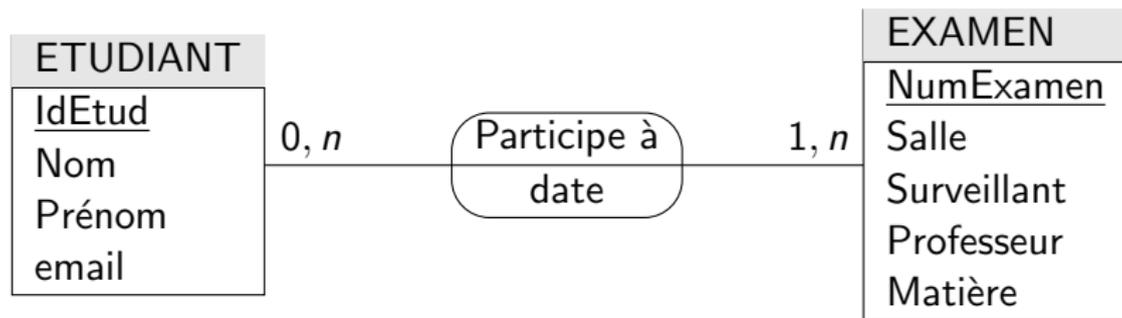
# Un besoin d'une entreprise

## Gestion du stock des produits

- ▶ Comprendre le besoin du client
- ▶ Comment structurer les données ?
- ▶ Comment garder les données cohérentes ?
- ▶ Comment optimiser les requêtes ?

# Modèle Conceptuel de Données (MCD)

- ▶ Entité
- ▶ Attribut
- ▶ Association / Relation
- ▶ Identifiant
- ▶ Cardinalité

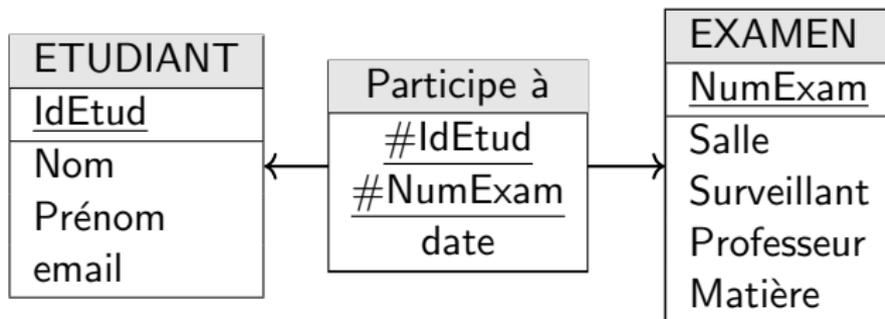


## Déterminer les cardinalités

- ▶ Tous les élèves ont un numéro unique d'étudiant.
- ▶ Pour chaque élève est inscrit à au moins un cours.
- ▶ Une salle de cours peut ne pas être utilisée.

0, 1 ou  $n$ ?

# Modèle Logique de Données (MLD)



- ▶ Clé primaire
- ▶ Clé étrangère
- ▶ Types des données

## Compétence visée : Conception d'une BD

- ▶ Nécessite de bien comprendre les besoins
- ▶ Ça prend du temps
- ▶ Permet la cohérence des données
- ▶ Crucial
- ▶ Difficile à changer en cas de mauvais choix

S'aider des Formes Normales.

# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

**Les données**

Introduction à Postgre SQL

NoSQL

Conclusion

# Les données ouvertes

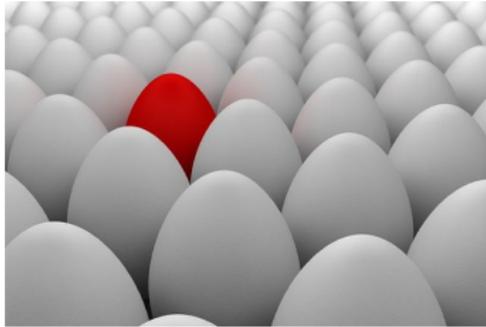


# Qualité des données



Layout (c) Dan Myers, 2018  
Conformed Dimensions of Data Quality, CC, BY-NC Dan Myers, 2018,  
See <http://dimensionsofdataquality.com/>

# Nettoyage des données

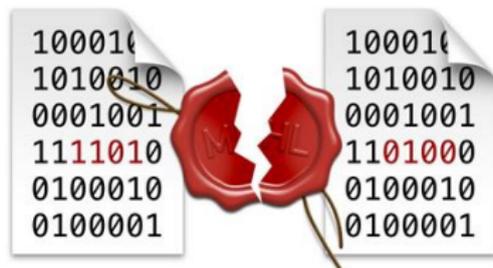


Aberrantes



Manquantes

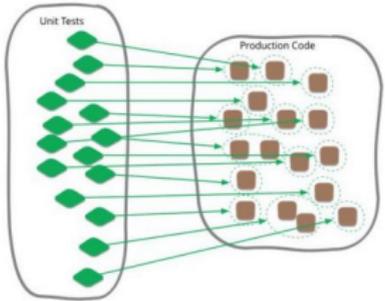
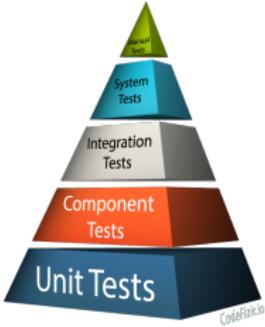
# Cohérences des données



Adresse fournisseur A :

- ▶ 42 allée Alan Turing
- ▶ 51 rue Nicolas Bourbaki

# Jeux de test



# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

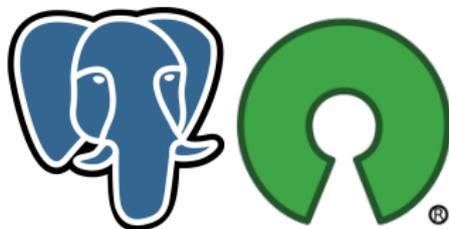
Les données

**Introduction à Postgre SQL**

NoSQL

Conclusion

# Postgre (1989)

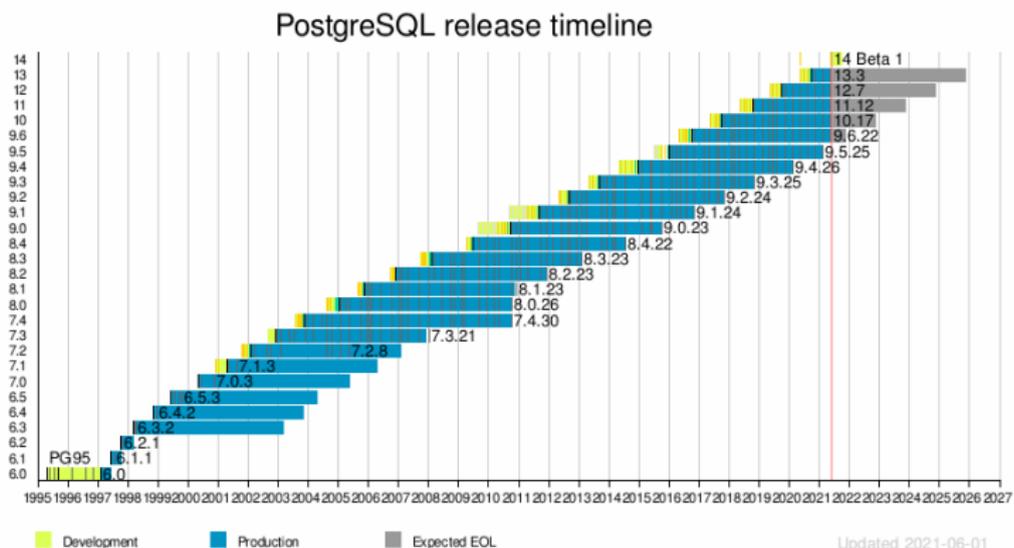


Codé en C

OS supportés

FreeBSD, HP-UX, Linux, NetBSD, OpenBSD, OS X, Solaris, Unix,  
Windows

# Postgre time life



## Commandes de connexion

```
ssh ssh.iut-clermont.uca.fr
```

```
ssh londres
```

```
psql -h londres -d dbpalafour -U palafour -W
```

Mot de passe : achanger

```
ALTER ROLE username
```

```
WITH PASSWORD 'password';
```

## CREATE / DROP TABLE

```
CREATE TABLE NOTES(  
    Nom VARCHAR(30),  
    Prenom VARCHAR(30),  
    Note number,  
    Matiere VARCHAR(20)  
    Prof VARCHAR(20)  
);  
  
DESCRIBE TABLE;
```

## Voir vos tables

Pour voir vos tables :

```
\d;
```

```
\d NOTES;
```

```
SELECT
```

```
    table_name,  
    column_name,  
    data_type
```

```
FROM
```

```
    information_schema.columns
```

```
WHERE
```

```
    table_name = 'NOTES';
```

```
DROP TABLE NOTES;
```

# INSERT

```
INSERT INTO NOTES VALUES(  
Lafourcade,  
Pascal,  
19,  
BD,  
Durand  
);
```

# UPDATE

```
UPDATE NOTES  
SET Prenom='Pascale'  
WHERE Nom='Lafourcade'
```

# DELETE

```
DELETE FROM NOTES  
WHERE Note < 10
```

# SELECT

```
SELECT Nom, Prenom  
FROM NOTES  
WHERE Note<10
```

# SELECT \* FROM NOTES

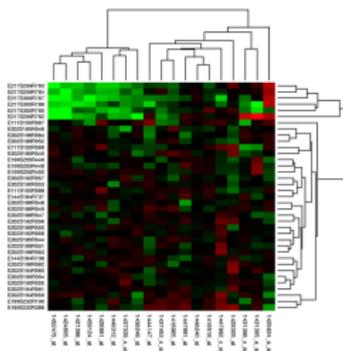
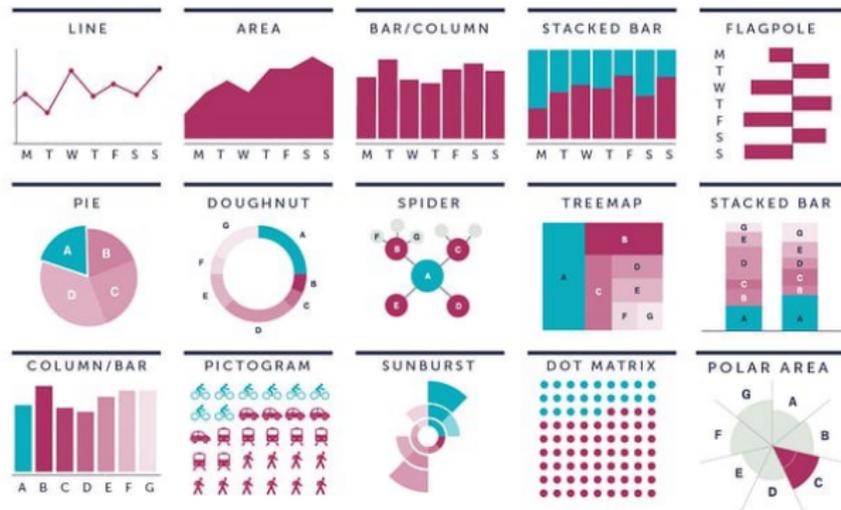
Nom	Prenom	Note	Matiere	Prof
Lux	Guy	12	BD	Durand
Lux	Guy	10	Bio	Pieplu
Frederic	Dard	16	BD	Durand
Lambert	Gerard	11	Math	More

- ▶ Modification du prof de BD
- ▶ Ajout note sans connaître le prof
- ▶ Effacement de Gérard Lambert

# Objectifs

- ▶ Conception de BD
- ▶ Création de scripts de déploiement
- ▶ Requêtes de consultation
- ▶ Exception
- ▶ Déclencheur
- ▶ Compteur
- ▶ Gestion des utilisateurs
- ▶ Optimisation
- ▶ Visualisation des données

# Visualisation



# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

Les données

Introduction à Postgre SQL

**NoSQL**

Conclusion

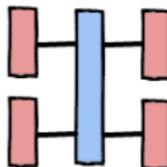
# SQL vs NoSQL

## Base de données SQL

Relationnelle



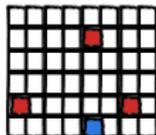
OLAP



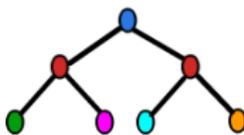
---

## Base de données NoSQL

Orienté colonne



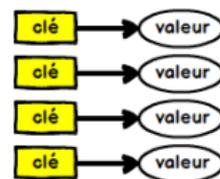
Document

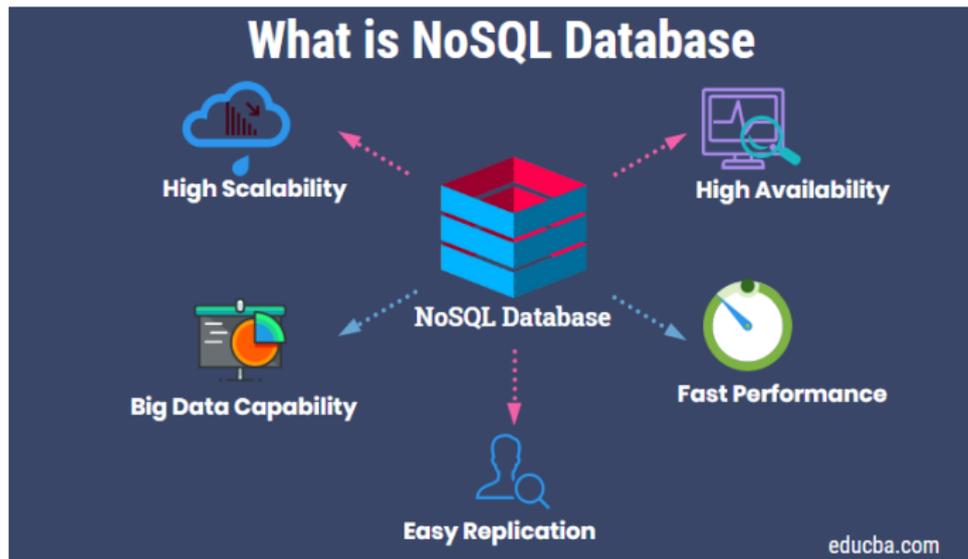


Graphe



Clé-Valeur





# No SQL SGBD

## key-value

Amazon  
DynamoDB (Beta)

ORACLE  
BERKELEY DB 11g



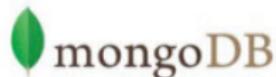
## graph



## column



## document



# Plan

Motivations

La BD à l'IUT

Base de données

Modélisation

Les données

Introduction à Postgre SQL

NoSQL

Conclusion

## A retenir

- ▶ BD est partout
- ▶ SQL une compétence de base d'un informaticien
- ▶ Nouvelle façon de penser
- ▶ Ouvre de nombreuses possibilités d'emploi
- ▶ BD toute l'année

**Merci pour votre attention.**

**Questions ?**