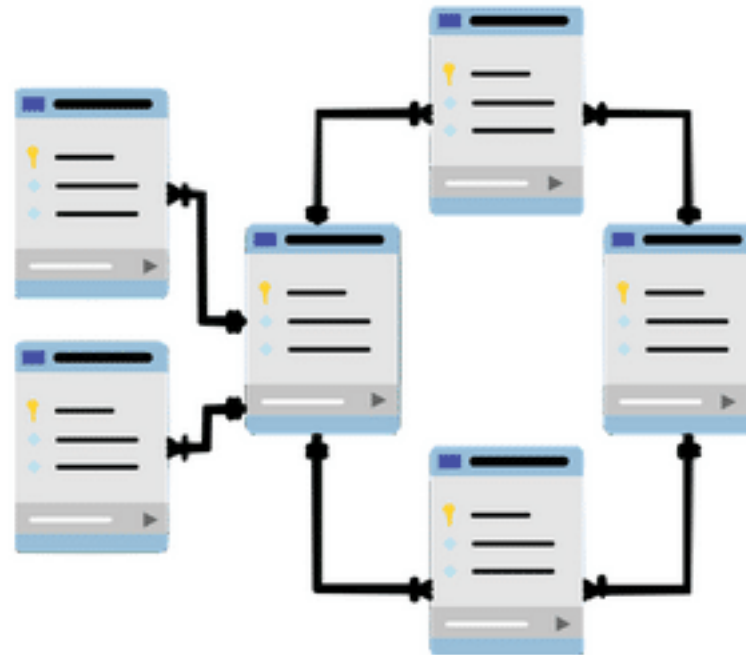


Les bases de données : grands principes

Quelles bases
de données
pour quelles
données ?



Une rapide définition :



https://fr.wikipedia.org/wiki/Base_de_donn%C3%A9es

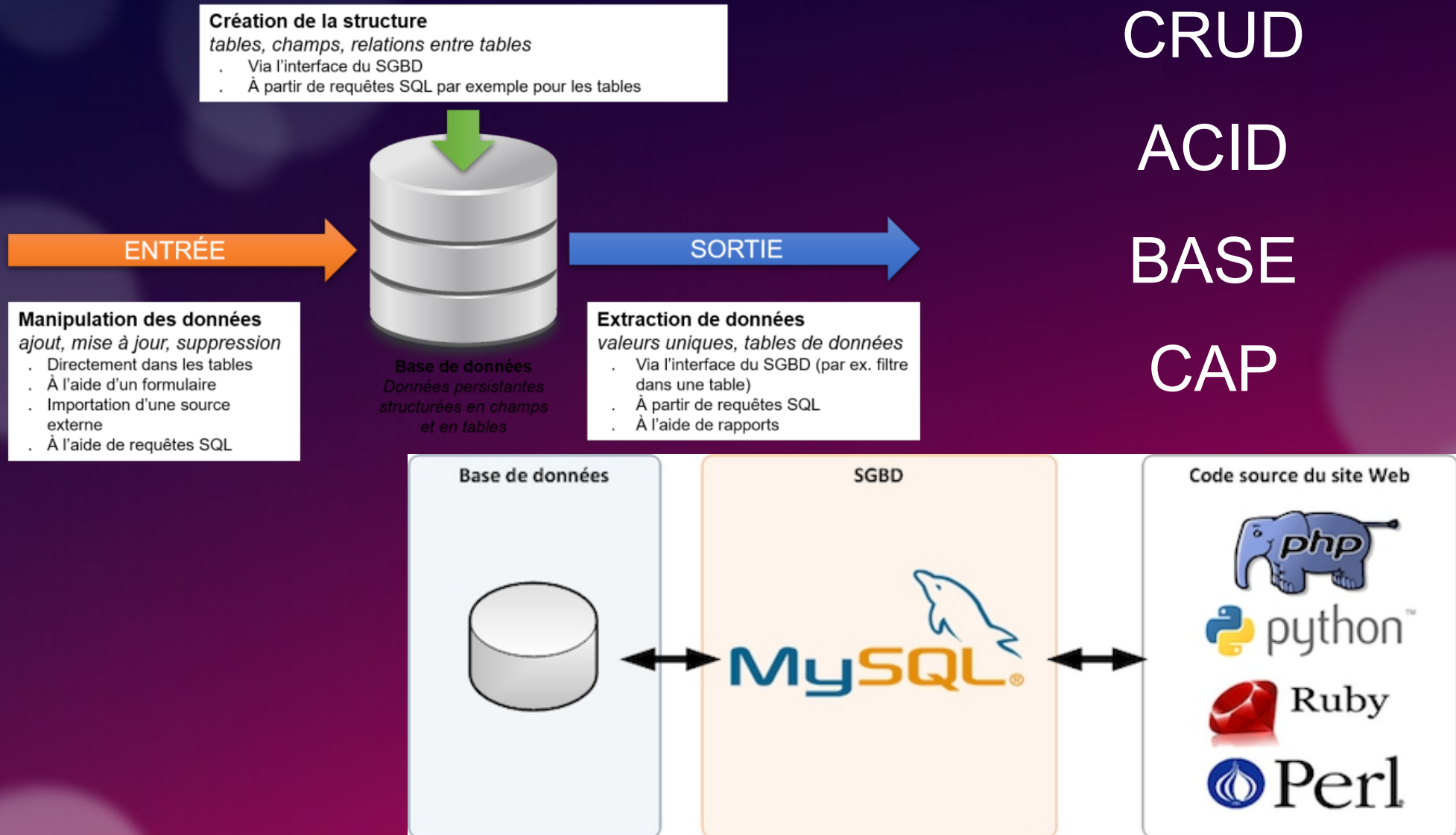
4 notions importante à retenir :

- Stocker et retrouver ?
- Données brutes ou informations ?
- De natures différentes ?
- Reliées plus ou moins entre elles ?

... la magie du SGBD



Le SGBD ... ou Système de Gestion de Base de Données



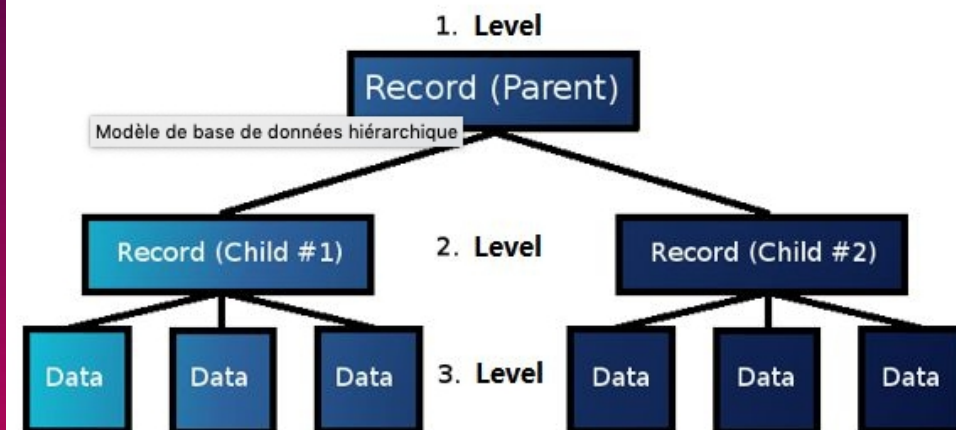
Les 3 grands types de bases de données #1-XML

- Les bases de données hiérarchiques.

Le principe :
l'élément "enfant"
dépend de l'élément
"parent".

Peu permissives,
très rigides, mais
très organisées.

```
<-record>
  <leader> cx a22 3 45 </leader>
  <controlfield tag="001">029205565</controlfield>
  <controlfield tag="003">http://www.idref.fr/029205565</controlfield>
  <controlfield tag="004">19881017</controlfield>
  <controlfield tag="005">20200412004930.000</controlfield>
  <controlfield tag="006">0004</controlfield>
  <controlfield tag="007">341720001</controlfield>
  <controlfield tag="008">Tp5</controlfield>
  <-datafield tag="010" ind1=" " ind2=" ">
    <subfield code="a">0000000453021059</subfield>
    <subfield code="2">ISNI</subfield>
    <subfield code="C">BNF</subfield>
    <subfield code="d">20190711</subfield>
  </datafield>
  <-datafield tag="033" ind1=" " ind2=" ">
    <subfield code="a">http://catalogue.bnf.fr/ark:/12148/cb11886526v</subfield>
    <subfield code="2">BNF</subfield>
    <subfield code="d">20150918</subfield>
  </datafield>
  <-datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">026658542</subfield>
    <subfield code="9">sudoc</subfield>
  </datafield>
  <-datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">frBN000362615</subfield>
  </datafield>
  <-datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">frBN004161296</subfield>
  </datafield>
  <-datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">FRBNF118865265</subfield>
    <subfield code="z">FRBNF11886526</subfield>
  </datafield>
```

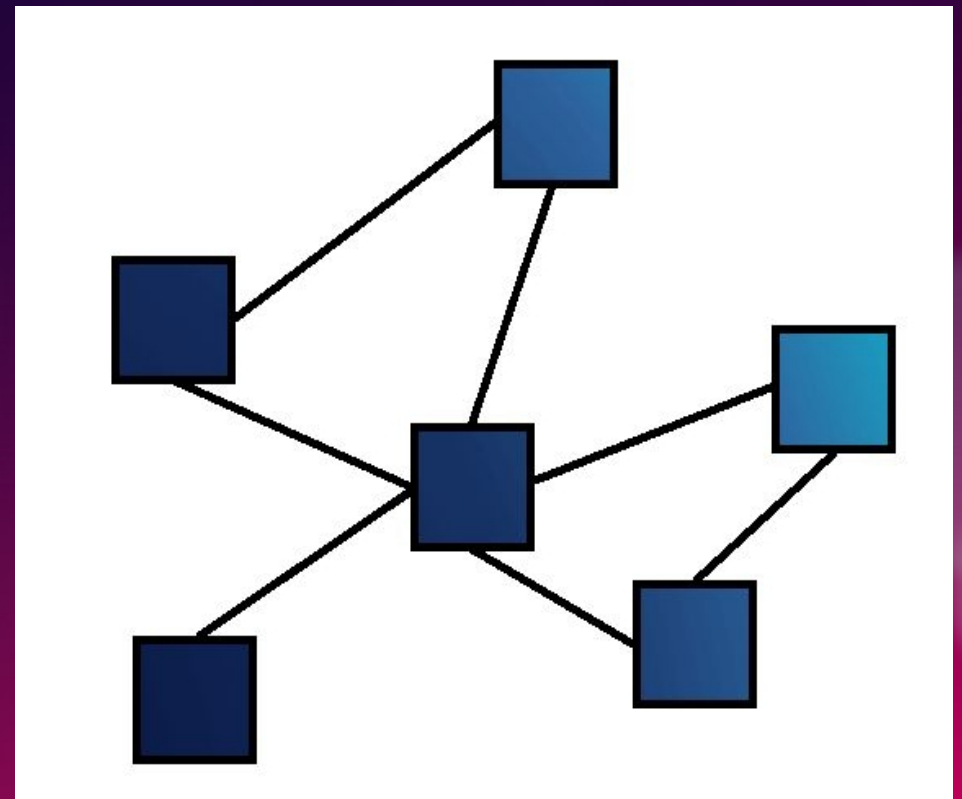


Les 3 grands types de bases de données #2-SQL

- Les bases de données réseau : les documents n'ont pas de relations "parent-enfant" *stricto sensu*

Donc pas UN chemin unique vers un enregistrement, mais plusieurs

Principalement utilisées par les *mainframes*



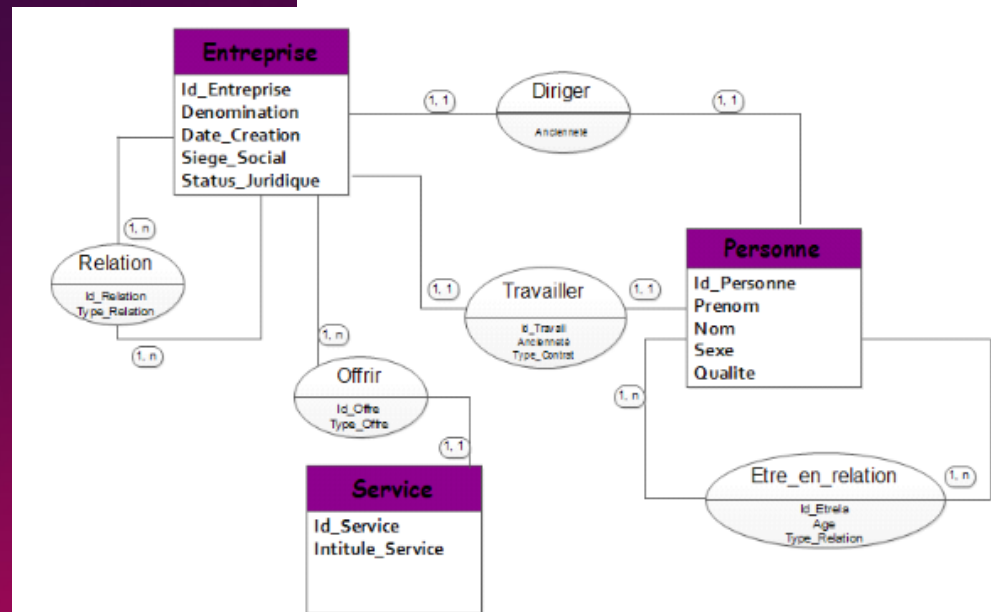
Les 3 types de bases de données #2.1-SQL

Les bases de données relationnelles : basées sur des tables et sur le concept de la relation entre ces tables

C'est sur le principe de l'algèbre relationnel que l'on peut extraire des informations de ces relations.

C'est la base du langage SQL.

| Numéro | Pseudonyme | E-mail | Âge |
|--------|---------------|-------------------------------|-----|
| 1 | Kryptonic | kryptonic@free.fr | 24 |
| 2 | Serial_Killer | serialkiller@unitedgamers.com | 16 |
| 3 | M@teo21 | top_secret@siteduzero.com | 18 |
| 4 | Bibou | bibou557@laposte.net | 29 |
| ... | ... | ... | ... |



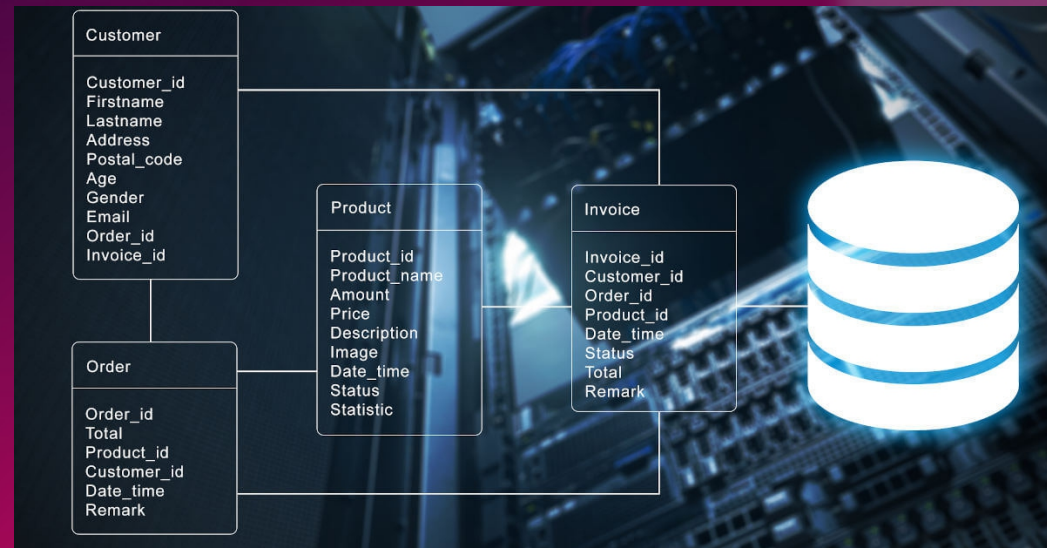
Les 3 grands types de bases de données #3-NoSQL

- Les bases de données orientées objet :

Un objet est un conteneur symbolique et autonome qui contient des informations et des mécanismes concernant un sujet, manipulés dans un programme. C'est également une représentation du monde réel.

4 grands sous-ensembles : les bases de données orientées **documents**, les bases de données « **clé/valeur** », les bases de données **colonnes** et les bases de données orientées **graphes**.

Le NoSQL est à la fois une autre manière d'interroger les données, mais aussi de les stocker.

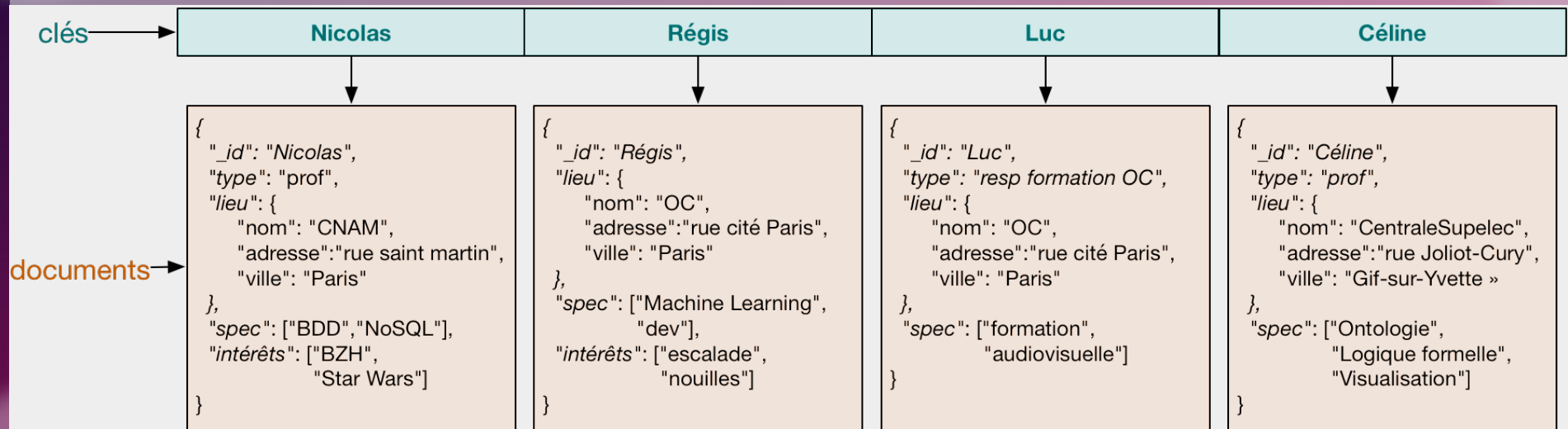


Les 4 sous-types de bases de données NoSQL #1

- Les bases de données orientées documents :

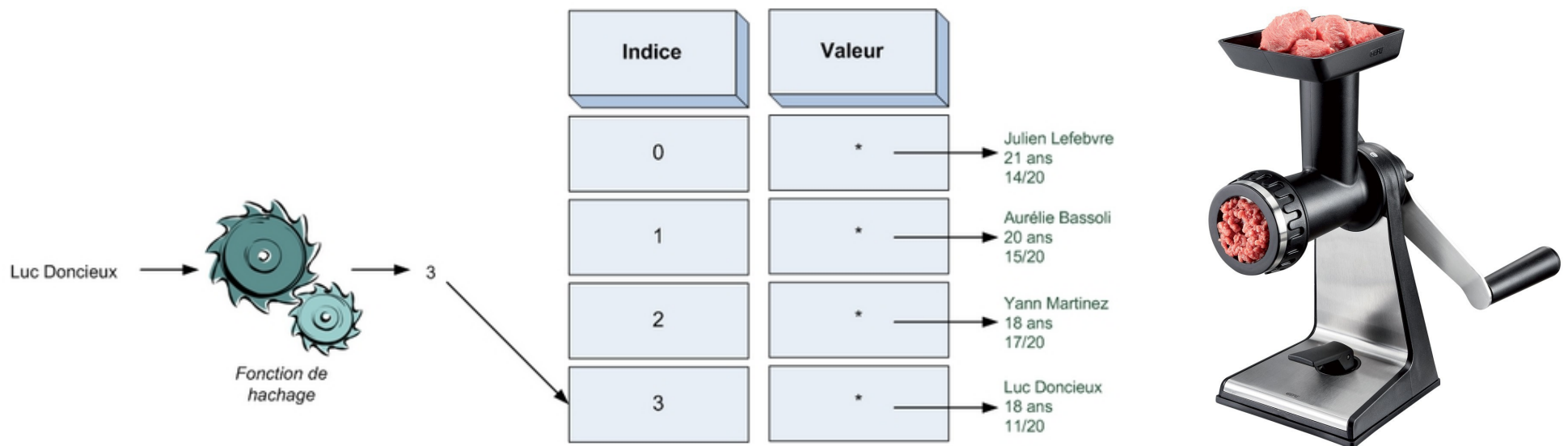
Fonctionnent sur un système clé/valeur étendu aux champs qui composent le document.

Approche structurée de chaque valeur = ensembles, elles forment un document qui a pour identifiant une clé.



Les 4 sous-types de bases de données NoSQL #2

- Le système clé/valeur :
Efficacité et simplicité ; ***Table de hachage***
- Clé : identifiant unique de la donnée, lui indique le comportement qu'elle doit avoir
- Valeur : contient n'importe quel type de données



Les 4 sous-types de bases de données NoSQL #3

- Les bases de données « colonnes » :

Focus fait sur chaque attribut et redistribution de ce dernier. La requête se fait donc en entrant par les colonnes que l'on veut interroger et seulement elles

Gain de temps et souplesse des requêtes
Davantage utile pour les gros calculs analytiques

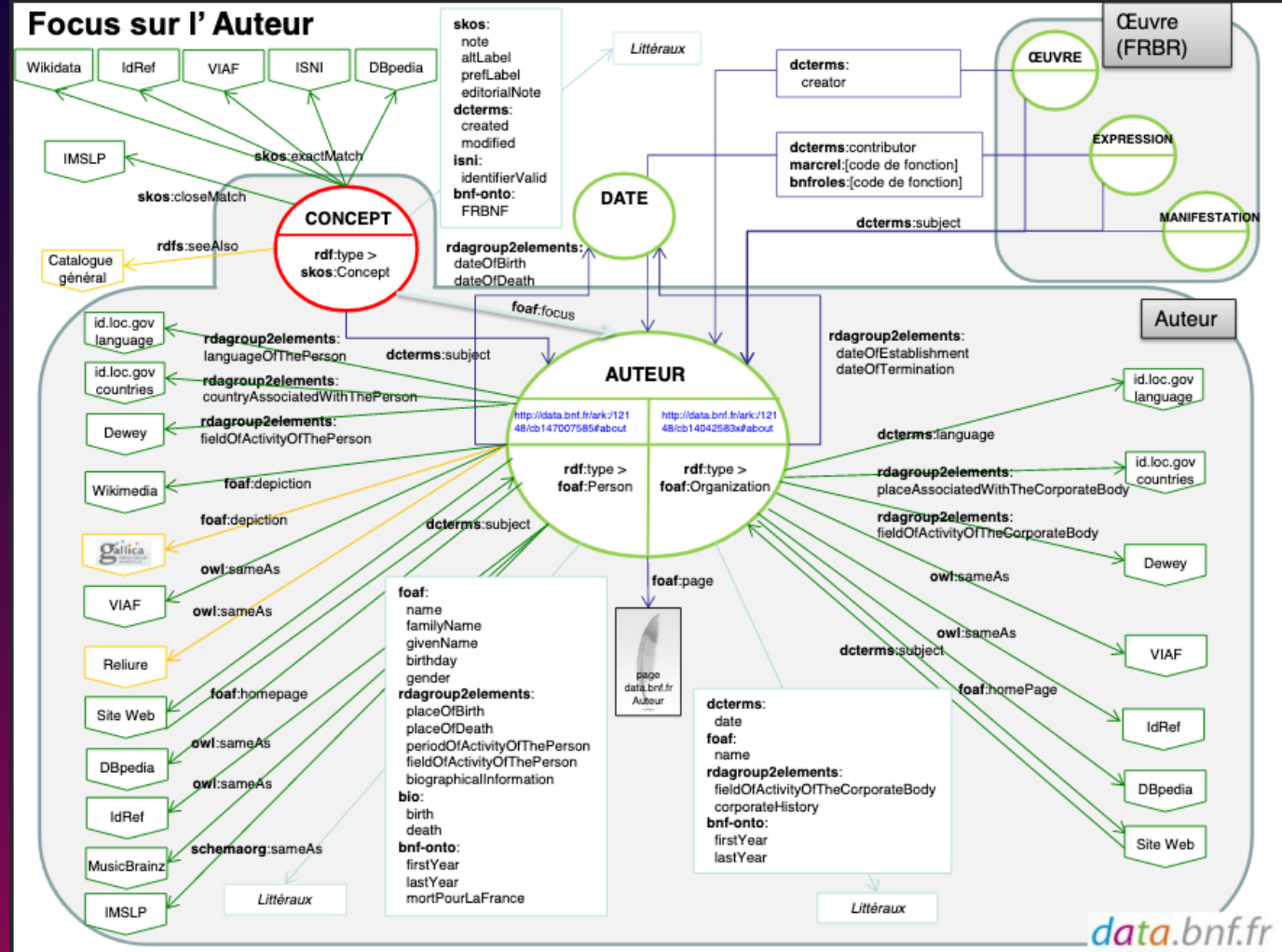
Elasticsearch (exemple)

| Stockage orienté lignes | | | | | Stockage orienté colonnes | | | | | | | |
|-------------------------|-------------------|-----------------|--|------------------------------|---------------------------|-------------------|---------|------------------|---------|------------------|---------|--------------------|
| id | type | lieu | spec | intérêts | id | type | id | lieu | id | spec | id | intérêts |
| Nicolas | prof | CNAM | BDD, NoSQL | BZH, Star Wars | Nicolas | prof | Céline | Centrale Supelec | Nicolas | BDD | Nicolas | BZH |
| Régis | | OC | Machine Learning, Dev | escalade, nouilles chinoises | Céline | prof | Nicolas | CNAM | Nicolas | NoSQL | Nicolas | Star Wars |
| Luc | resp formation OC | OC | formation, audiovisuel | | Luc | resp formation OC | Régis | OC | Régis | Machine Learning | Régis | escalade |
| Céline | prof | CentraleSupelec | Ontologie, logique formelle, visualisation | | Luc | OC | Luc | OC | Régis | Dev | Régis | nouilles chinoises |
| | | | | | | | | | Luc | formation | | |
| | | | | | | | | | Luc | audiovisuel | | |
| | | | | | | | | | Céline | Ontologie | | |
| | | | | | | | | | Céline | logique formelle | | |
| | | | | | | | | | Céline | visualisation | | |

Les 4 sous-types de bases de données NoSQL #4

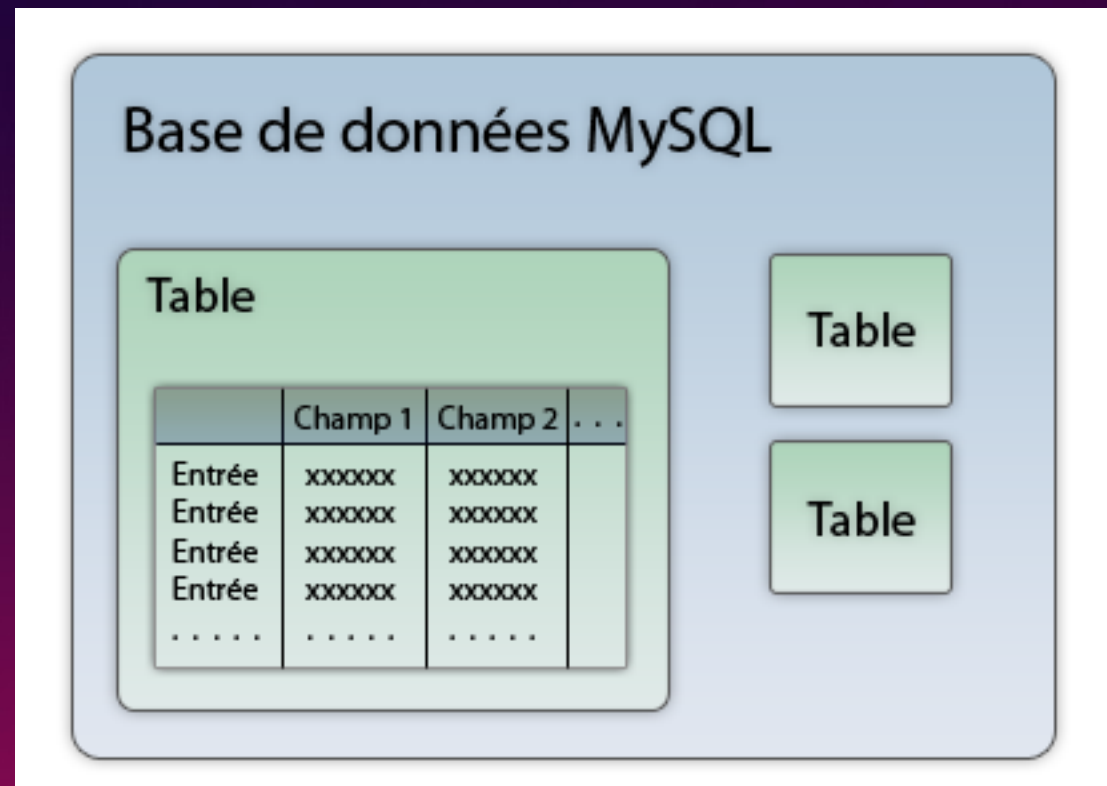
- Les bases de données graphe :
Tente de résoudre le problème de corrélations entre les éléments

Les données stockées sont : les nœuds, les liens et des propriétés sur ces nœuds et ces liens.



Les bases de données SQL

- Les principaux concepts : entité, attribut, association, cardinalité, clé primaire, clé secondaire, table, champ, enregistrement, opérations CRUD



Un premier problème ...

| titre | année | nom_realisateur | prenom_realisateur | annee_naissance |
|----------|-------|-----------------|--------------------|-----------------|
| Alien | 1979 | Scott | Ridley | 1943 |
| Vertigo | 1958 | Hitchcock | Alfred | 1899 |
| Psychose | 1960 | Hitchcock | Alfred | 1899 |
| etc. | | | | |

... et la solution :

films

| titre | annee |
|----------|-------|
| Alien | 1979 |
| Vertigo | 1958 |
| Psychose | 1960 |

realisateurs

| nom_realisateur | prenom_realisateur | annee_naissance |
|-----------------|--------------------|-----------------|
| Scott | Ridley | 1943 |
| Hitchcock | Alfred | 1899 |

A retenir : éviter les redondance !

Mais ... il reste un problème et pas des moindres, le voyez-vous ?

ID, Oh mon ID !

Notions

- Entité
- Attribut
- Clé primaire (identifiant) : attribut qui identifie un enregistrement

films

| id | titre | annee |
|----|----------|-------|
| 1 | Alien | 1979 |
| 2 | Vertigo | 1958 |
| 3 | Psychose | 1960 |

realisateurs

| id | nom_realisateur | prenom_realisateur | annee_naissance |
|----|-----------------|--------------------|-----------------|
| 1 | Scott | Ridley | 1943 |
| 2 | Hitchcock | Alfred | 1899 |

Et le relationnel dans tout ça ?

- C'est la clé étrangère la clé du relationnel !!

films

| id | titre | annee | id_realisateur |
|----|----------|-------|----------------|
| 1 | Alien | 1979 | 1 |
| 2 | Vertigo | 1958 | 2 |
| 3 | Psychose | 1960 | 2 |

realisateurs

| id | nom_realisateur | prenom_realisateur | annee_naissance |
|----|-----------------|--------------------|-----------------|
| 1 | Scott | Ridley | 1943 |
| 2 | Hitchcock | Alfred | 1899 |