

Modélisation Base de données et normalisation

Par Richard BONNAMY

Sommaire

- Bases de données relationnelles
- Normalisation
- Types de bases de données

Base de données relationnelles

- ❑ une **base de données relationnelle** est une base de données où l'information est organisée dans des **tableaux** à deux dimensions appelés des **relations** ou **tables**

LIVRES		
TITRE	AUTEUR	EDITEUR
Germinal	Emile Zola	Grasset
Madame Bovary	Gustave Flaubert	Folio
L'assommoir	Zola	Le Livre de Poches

- ❑ Les lignes sont appelées des **enregistrements** ou des **n-uplets**
- ❑ Les colonnes sont appelées des **attributs** ou **champs**

- ❑ **Clé primaire ou PK**
 - ❑ **Attribut ou ensemble d'attributs** qui permet d'identifier un enregistrement de manière unique.
 - ❑ Les attributs constituant la clé primaire sont **soulignés**

LIVRES

<u>TITRE</u>	AUTEUR	<u>EDITEUR</u>
Germinal	Emile Zola	Grasset
Madame Bovary	Gustave Flaubert	Folio
L'assommoir	Zola	Le Livre de Poches

❑ Clé primaire ou PK

- Exemple **KO**

<u>TITRE</u>	AUTEUR	<u>EDITEUR</u>
Germinal	Emile Zola	Grasset
Madame Bovary	Gustave Flaubert	Folio
L'assommoir	Zola	Le Livre de Poches
Germinal	Jean Dupont	Grasset

- Exemple **OK**

<u>TITRE</u>	<u>AUTEUR</u>	<u>EDITEUR</u>
Germinal	Emile Zola	Grasset
Madame Bovary	Gustave Flaubert	Folio
L'assommoir	Zola	Le Livre de Poches
Germinal	Jean Dupont	Grasset

❑ Clé étrangère ou FK : lien d'intégrité entre 2 tables

- Une clé étrangère permet de lier 2 colonnes appartenant à des tables différentes

LIVRES

<u>TITRE</u>	AUTEUR	ID_EDITEUR
Germinal	Emile Zola	1
Madame Bovary	Gustave Flaubert	2
L'assommoir	Zola	3

EDITEURS

<u>ID</u>	<u>NOM</u>	<u>ADRESSE</u>
1	Grasset	61 rue des Saints-Pères, 75006 Paris
2	Folio	5 rue Gaston-Gallimard, 75328 Paris
3	Le Livre de Poches	21 Rue du Montparnasse, 75006 Paris


❑ Impact de la clé étrangère

- Impossible de relier un livre à un éditeur dont l'identifiant n'existe pas
 - Le livre développer en 1h est associé à un éditeur qui n'existe pas : KO
- Impossible de supprimer un éditeur pour lequel des livres existent :
 - on peut supprimer Seuil mais pas Folio

LIVRES

TITRE	AUTEUR	ID_EDITEUR
Germinal	Emile Zola	1
Madame Bovary	Gustave Flaubert	2
L'assommoir	Emile Zola	3
Développer en 1h	Gérard Dupin	5

EDITEURS



ID	NOM	ADRESSE
1	Grasset	61 rue des Saints-Pères, 75006 Paris
2	Folio	5 rue Gaston-Gallimard, 75328 Paris
3	Le Livre de Poches	21 Rue du Montparnasse, 75006 Paris
4	Seuil	57 rue Gaston-Tessier, 75166 Paris

Normalisation

- ❑ Ensemble de principes permettant:
 - d'éviter la redondance de données
 - diminution de la volumétrie
 - Simplification des traitements de mise à jour des données
 - d'éviter de stocker des données difficilement exploitables

❑ Forme normale:

- désigne un type de relation particulier entre les entités.
- est un principe de la normalisation
- 1FN, 2FN, 3FN, ..., 8FN

❑ Le respect d'une forme normale suppose le respect de la précédente

- d'abord 1FN, puis 2FN, etc.

❑ Exemple de table non normalisée:

- Redondance d'informations

LIVRES

ID	TITRE	AUTEUR	EDITEUR	ADRESSE_EDITEUR
1	Vingt mille lieues sous les mers	Jules Verne	Le livre de Poche	21 Rue du Montparnasse, 75006 Paris
2	Germinal	Emile Zola	Le livre de Poche	21 Rue du Montparnasse, 75006 Paris
3	Vipère au poing	Hervé Bazin	Grasset	61 Rue des Saints-Pères, 75006 Paris

❑ Solution:

- 1 table séparée pour stocker les éditeurs

LIVRES

ID	TITRE	AUTEUR	ID_EDI
1	Vingt mille lieues sous les mers	Jules Verne	1
2	Germinal	Emile Zola	1
3	Vipère au poing	Hervé Bazin	2

Clé étrangère

EDITEURS

ID	NOM	ADRESSE
1	Le livre de poche	21 Rue du Montparnasse, 75006 Paris
2	Grasset	61 Rue des Saints-Pères, 75006 Paris

❑ Exemple de table non normalisée:

- Données non atomiques
- SQL peu performant pour analyser des chaînes de caractères
 - Exemple: retrouver un numéro de téléphone

ID	NOM	ADRESSE
1	Le livre de poche	21 Rue du Montparnasse, 75006 Paris, Tél: 0102030405
2	Grasset	61 Rue des Saints-Pères, 75006 Paris, Tél: 0908070605

❑ Solution:

- Plusieurs attributs: RUE, CP, VILLE ET TEL

ID	NOM	RUE	CP	VILLE	TEL
1	Le livre de poche	21 du Montparnasse	75006	Paris	0102030405
2	Grasset	61 des Saints-Pères	75006	Paris	0908070605

ID	NOM	ADRESSE
1	Le livre de poche	21 Rue du Montparnasse, 75006 Paris, Tél: 0102030405
2	Grasset	61 Rue des Saints-Pères, 75006 Paris, Tél: 0908070605

❑ 1FN - Première forme normale:

- Est en première forme normale, une table dont les attributs possèdent tous une valeur **sémantiquement atomique**.
- Attention l'information doit être préservée

❑ **Forme normale 1FN - préservation de l'information**

❑ **Cas du libellé de rue dans l'exemple ci-dessus**

- Si on sépare « du » et « Montparnasse », on perd l'information du nom de la rue, donc la valeur « du Montparnasse » est bien considérée comme atomique.

❑ **Cas particulier d'une date**

- On ne sépare jamais les jours, mois et année
- Une **date** est considérée comme une **information atomique**.

❑ 2FN - Deuxième forme normale:

- Est en première forme normale.
- Un attribut « non clé » doit dépendre uniquement de la clé primaire et non d'une partie de la clé

❑ Dans la table ci-dessous, la clé primaire est constituée d'un **doublet unique** mais **CODE_FOURNISSEUR** ne dépend que du **FOURNISSEUR** et non de la **clé primaire**

ARTICLE	FOURNISSEUR	PRIX	CODE_FOURNISSEUR	SIRET
Boulon de 12	SARL Dupont	2.5	2187	AF102
Boulon de 14	SARL Dupont	2.8	2187	AF102
Perceuse XT2000	Générale de Mécanique	256.99	3128	GM227

❑ Résultat:

- redondance d'information

DEUXIÈME FORME NORMALE

❑ Solution pour être en 2FN:

- Séparer les tables

Clé étrangère

<u>ARTICLE</u>	<u>ID FOUR</u>
Boulon de 12	1
Boulon de 14	1
Perceuse XT2000	2

<u>ID</u>	<u>LIBELLE</u>	CODE	SIRET
1	SARL Dupont	2187	AF102
2	Générale de Mécanique	3128	GM227

☐ **3FN - troisième forme normale:**

- Est en seconde forme normale.
- Un attribut « non clé » ne dépend pas d'un autre attribut « non clé »

☐ Dans la table ci-dessous, le PAYS dépend de la VILLE et on a à nouveau de la redondance d'information

ID	<u>FOURNISSEUR</u>	VILLE	PAYS
1	SARL Dupont	Paris	France
2	Générale de Mécanique	Paris	France
3	Alstom	Mulhouse	France

☐ **Solution pour être en 3FN:**

- Séparer les tables comme dans la 2FN

Bases de données SQL vs NoSQL

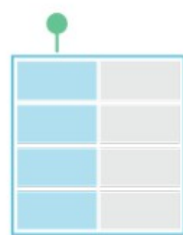
Système de gestion de base de données relationnelles:

- ❑ Les données sont stockées dans des tables, appelées aussi relations

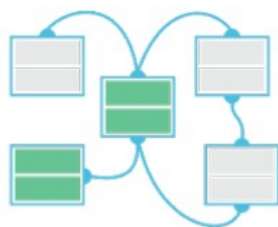
	FirstName	LastName	Email	Phone	Position	Branch	Address
	Andrew	Fuller	afuller@contoso.com	(205) 555 - 9898	CEO	TopManagement	London, 120 Hanover Sq.
	Jeremy	Boather	jboather@contoso.com	(205) 555 - 9888	President QA	QA	London, 120 Hanover Sq.
	Anne	Dodsworth	adodsworth@contoso.com	(205) 555 - 9887	VP QA	QA	London, 120 Hanover Sq.
	Alexander	Tuckings	atuckings@contoso.com	(205) 555 - 9886	Team Lead...	QA	London, 120 Hanover Sq.
	Brenda	Smith	bsmith@contoso.com	(205) 555 - 9885	Senior QA	QA	London, 120 Hanover Sq.
	Mary	Bird	mbird@contoso.com	(205) 555 - 9885	Team Lead...	QA	London, 120 Hanover Sq.
	Steven	Buchanan	sbuchanan@contoso.com	(205) 555 - 9897	President ...	Development	London, 120 Hanover Sq.
	Robert	King	rking@contoso.com	(205) 555 - 9896	VP Dev De...	Development	London, 120 Hanover Sq.
	Laura	Callahan	lcallahan@contoso.com	(205) 555 - 9892	Team Lead...	Development	London, 120 Hanover Sq.
0	Jason	Roland	jroland@contoso.com	(205) 555 - 9872	Senior Dev	Development	London, 120 Hanover Sq.
1	Eric	Danstin	edanstin@contoso.com	(205) 555 - 9882	Team Lead...	Development	London, 120 Hanover Sq.
2	Elizabeth	Lincoln	elincoln@contoso.com	(205) 555 - 9862	Senior Dev	Development	London, 120 Hanover Sq.
3	Margaret	Peacock	mpeacock@contoso.com	(205) 555 - 9852	Senior Dev	Development	London, 120 Hanover Sq.

- ❑ Un langage appelé SQL permet de réaliser des requêtes pour extraire de le l'information

❑ Bases NoSQL, plusieurs types:



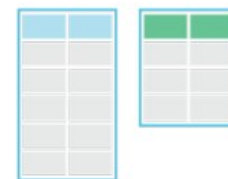
KEY-VALUE



GRAPH



DOCUMENT



COLUMN STORE

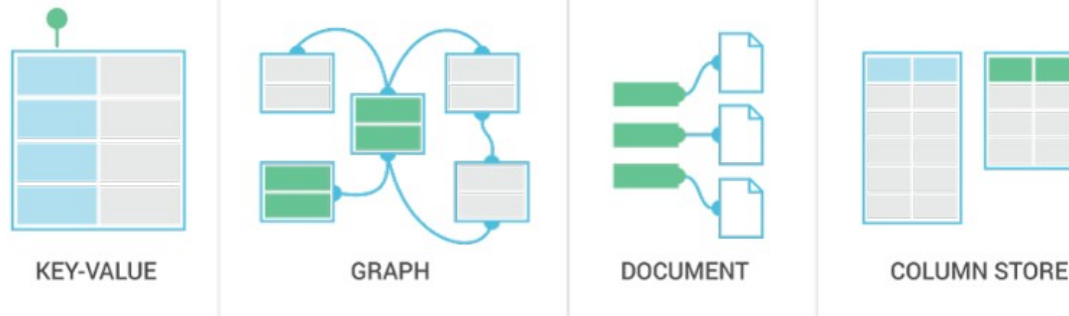
❑ **Key-Value:** les données sont stockées avec une clé unique. Bases les plus simples.

- Bases de ce type: Redis, Amazon DynamoDB

❑ **Document:** comme la Key-Value mais la valeur peut-être structurée. Plusieurs formats reconnus pour la valeur: XML, Json, etc..

- On peut stocker également du Json ou du XML dans une base de données Key-Value mais cette dernière ne proposera aucune fonctionnalité avancée pour l'exploiter.
- Bases de ce type: MongoDB, Elasticsearch

❑ Bases NoSQL, plusieurs types:



- ❑ **Column: données stockées en colonnes plutôt qu'en lignes. Les colonnes peuvent être regroupées par familles.**
 - Exemple de stockage: une colonne "nom du client" contenant tous les noms de famille des clients, une colonne "prénom du client" contenant tous les prénoms des clients, etc..
 - Bases de ce type: Cassandra, Apache Hadoop
- ❑ **Graph: base de données orientée objet dont le fonctionnement repose sur la théorie des graphes**
 - Bases de ce type: DataStax, InfiniteGraph, MarkLogic, etc..

☐ **ACID : Atomicité, Cohérence, Isolation, Durabilité**

☐ **Atomicité:**

- Une transaction de mise à jour d'une base de données se fait en totalité ou pas du tout. Elle ne peut pas être partielle, ceci afin de garantir l'intégrité des données.

☐ **Cohérence:**

- À un instant t , les utilisateurs du monde entier voient les mêmes données.
 - **Essentielle** par exemple pour la finance. A un instant t , le compte bancaire d'un client doit avoir le même solde à Paris et à New-York.

☐ **Isolation:**

- Toutes les transactions sont indépendantes et doivent s'exécuter comme si à un instant t elle était la seule à s'exécuter.

☐ **Durabilité:**

- Une fois une transaction validée, la base de données est dans un état stable, quel que soit l'évènement qui peut se produire, comme une panne d'électricité.

❑ Bases relationnelles (SQL):

- Toutes les bases relationnelles pro sont ACID
- Garantissent la cohérence des données:
 - les personnes du monde entier voient la même chose au même moment.
 - Essentiel pour un système bancaire ou pour les applications de gestion en général.

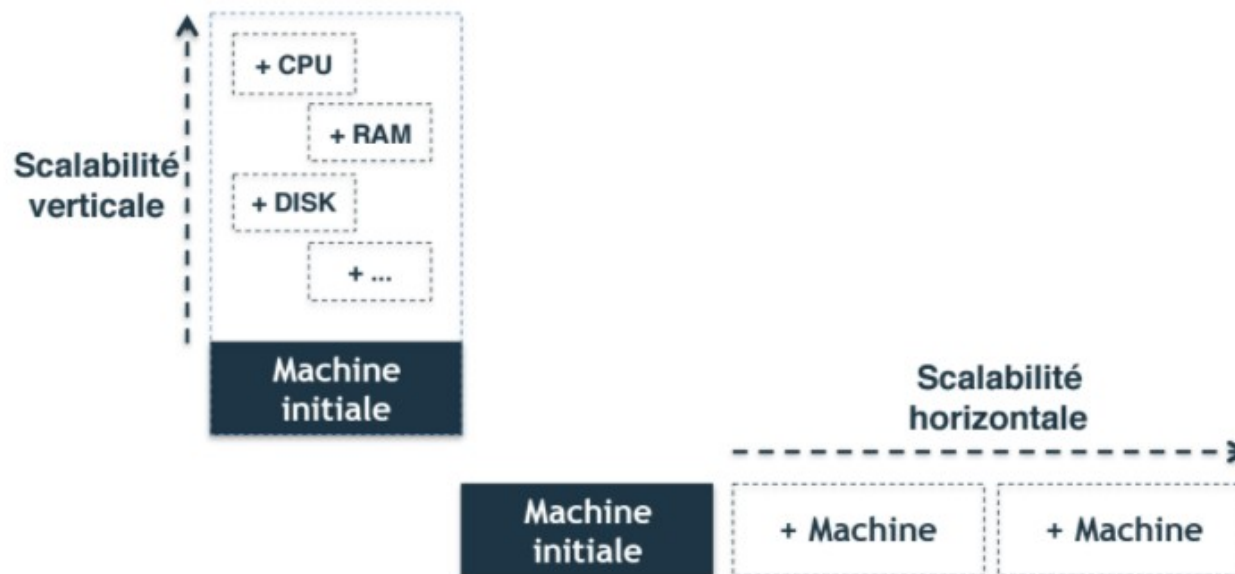
❑ Bases NoSQL:

- Les bases NoSQL sont non ACID
- Ne garantissent pas la cohérence des données:
 - Pour un site comme **www.leboncoin.fr**, peu importe qu'une personne de Lille voit une nouvelle annonce 5 minutes après 1 personne de Nantes...
 - Idem pour un moteur de recherche

❑ Technologie de base de données en fonction du besoin...

❑ Il existe 2 types de scalabilité:

- Scalabilité verticale: en cas de besoin de plus de ressources on augmente les capacités de la machine (+ de CPU, + de RAM, + de Disque).
- Scalabilité horizontale: en cas de besoin de plus de ressources on augmente de nombre de machines.
- Pour une base de données, le type de scalabilité dépend du type de bases



❑ Bases de données relationnelles.

- Exige une scalabilité verticale.
- Pas compatible avec la scalabilité horizontale car on ne peut pas garantir la cohérence des données. En effet si une base de données est répartie sur plusieurs machines (ie. nœud) il y a un temps de réplication des informations d'un nœud à l'autre.

❑ Bases NoSQL.

- sont conçues pour la scalabilité horizontale.

❑ Théorème CAP:

- une base, quelle que soit son type, ne peut pas garantir à la fois la disponibilité (A), la cohérence (C) et la tolérance à la partition, i.e. scalabilité horizontale (P).

❑ Une base est soit :

- ❑ CA: base de données relationnelles
- ❑ Soit CP, ou AP : pour les bases NoSQL selon l'éditeur (cf. schéma ci-dessous)

