Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Modélisation Base de données et normalisation

Par Richard BONNAMY

Sommaire

- Modélisation
- Diagrammes UML
- Bases de données relationnelles
- Normalisation
- Types de bases de données
- TP modélisation

Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Modélisation

MODELISATION

- **un petit dessin vaut mieux qu'un long discours**.

- Un **modèle** est une abstraction permettant de mieux comprendre un objet complexe (bâtiment, économie, atmosphère, cellule, logiciel, …).

- Un **modèle** sert :
  - de document d'échange,
  - d'outil de conception,
  - de référence pour la réalisation,
  - de référence pour la maintenance et l'évolution

- La construction d'un système d'information, d'un réseau, d'un logiciel complexe, de taille importante nécessite une **modélisation**.

MODELISATION

- **Langage de modélisation** : une syntaxe commune, graphique, pour modéliser (UML, MERISE,…).

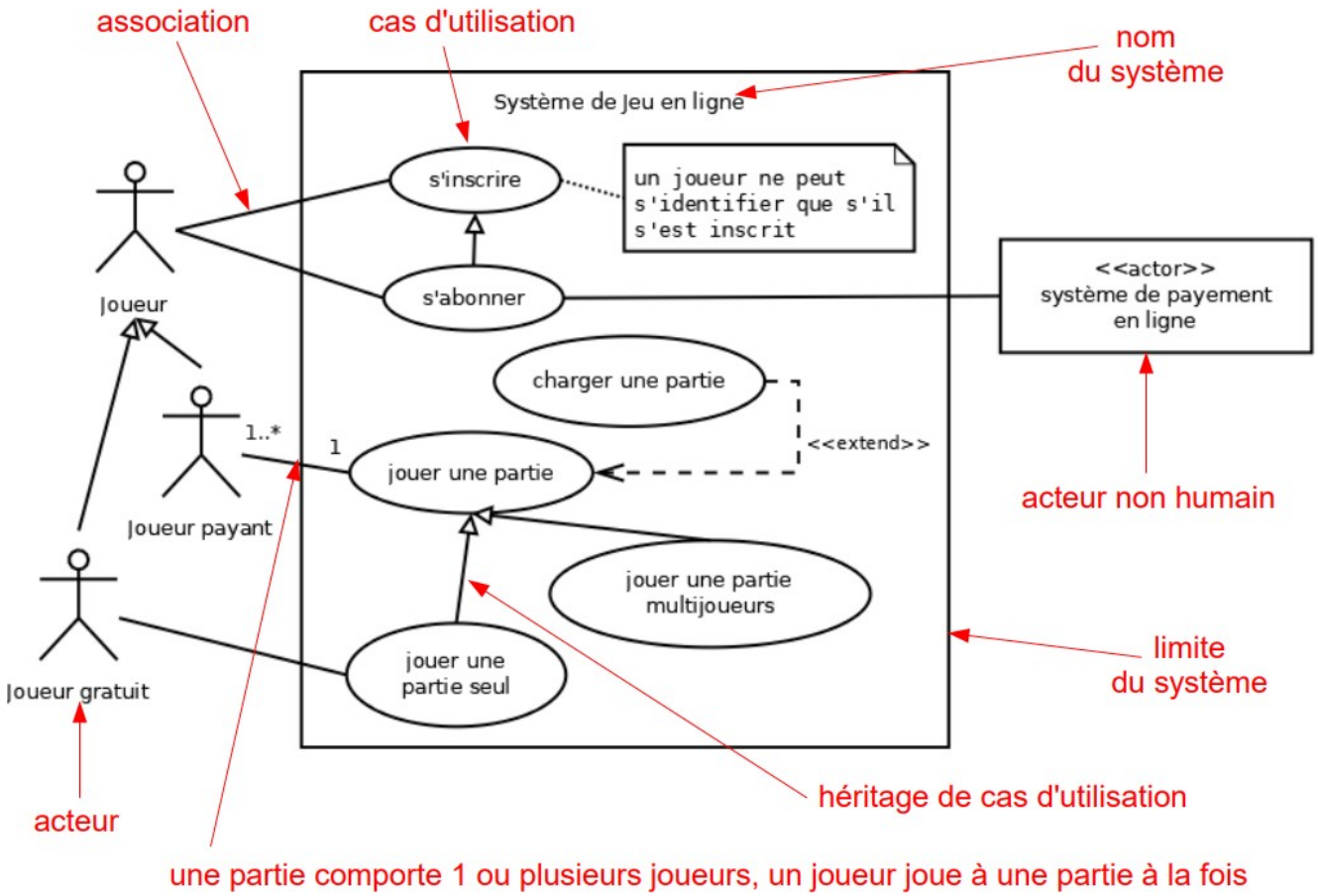
- UML: le langage standard pour la modélisation objet.

- Les spécifications UML font environ 800 pages et couvrent tous les aspects du développement logiciel, offrant des outils de modélisation de l'architecture générale aux plus petits détails.

MODELISATION

- **UML** propose de nombreux diagrammes.
- En java, il existe des outils permettant de générer du code à partir d'un diagramme **UML** correctement renseigné:
  - squelette des classes
  - héritage
  - signatures des méthodes.
- C'est l'utilisation préconisée par le **MDA** (Model Driven Architecture).
- L'intérêt principal d'un tel modèle reste la documentation, la conception et la communication

- Exemple: diagramme de cas d'utilisation



Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Diagrammes UML

DIAGRAMMES

- La modélisation est issue de l'étude du cahier des **spécifications**.

- **Savoir lire une spécification**
  - Une spécification est constituée de « cas d'utilisation » (Use cases en anglais)
    - Exemple: L'administrateur saisit une nouvelle référence de livre dans le système
  - **Cas d'utilisation:**
    - **Sujet** => acteur
    - **Verbe** => méthode
    - **Complément** => entité métier

DIAGRAMMES

- Identifier les **entités métier** dans les spécifications afin de réaliser le **diagramme de classes** (modèle objet avec cardinalités)
- Les **acteurs** vont permettre d'identifier des **profils** et donc la **sécurité** d'accès à votre application
- Les **verbes** employés vont définir les **méthodes**.

- **Un diagramme de cas d'utilisation décrit:**
  - les acteurs (profils)
  - Les actions réalisables par les acteurs

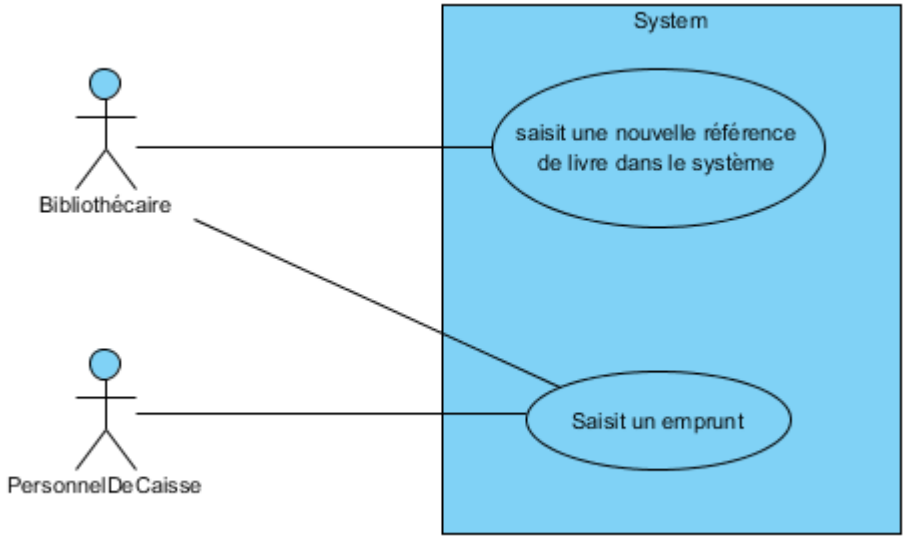


Diagramme de classes:

- Le diagramme le plus utilisé
- Décrit des relations (principalement associations et généralisations) entre **classes.** Les **associations** ont des **cardinalités**

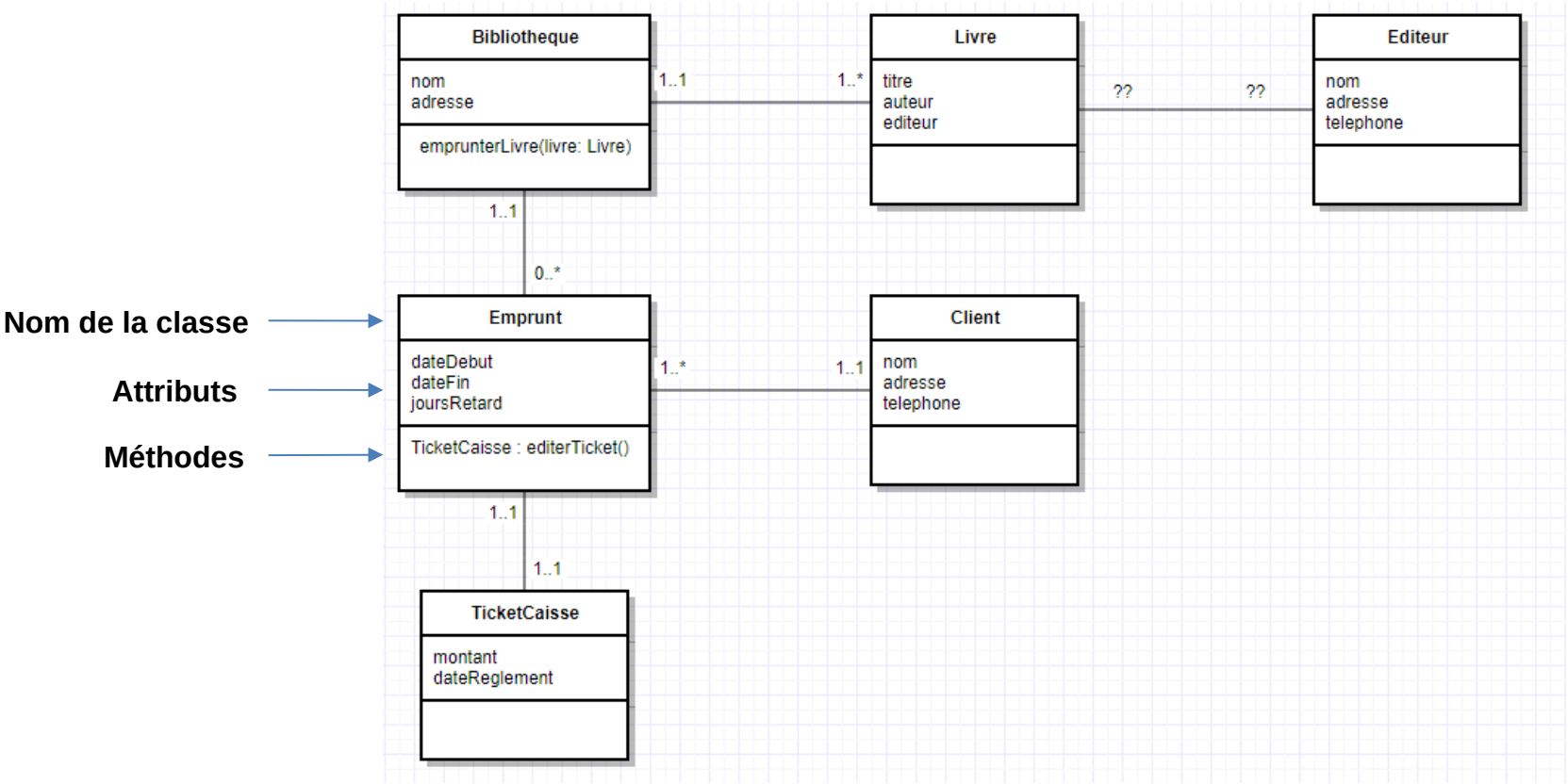
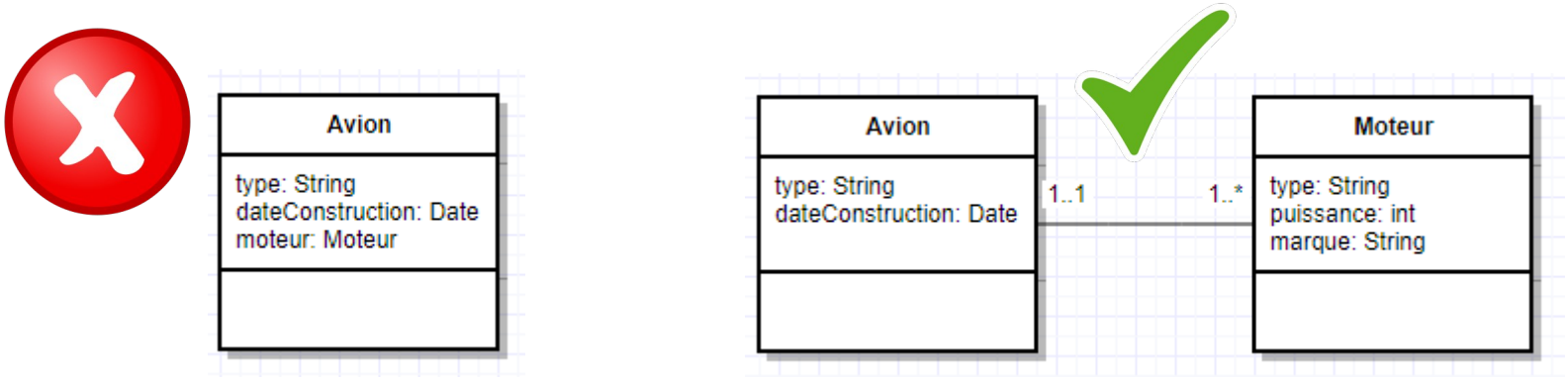


Diagramme de classes

- Seuls les attributs correspondant à un type primitif ou un type commun (date, chaine de caractères) apparaissent dans la classe:
  - Nom
  - Prénom
  - Date de naissance
- Pour les **associations**, on utilise une **liaison entre classes**.



- Seules les méthodes qui ont une importance fonctionnelle ou technique sont indiquées
  - On ne met pas les getters et setters

- **Cardinalités**
  - Décrit la multiplicité **bidirectionnelle** dans une relation.
    - Exemple: 1 voiture possède 4 roues. Une roue n'appartient qu'à une seule voiture.
  - Exemple ci-dessous:
    - Détaillons les cardinalités
    - Quelle est la cardinalité entre Livre et Editeur ?

Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Base de données relationnelles

BASE DE DONNEES

- 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

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

BASE DE DONNEES

Clé primaire ou PK

- **Exemple KO**

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

- **Exemple OK**

TITRE	AUTEUR	EDITEUR
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

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

EDITEURS

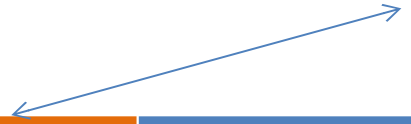
ID	NOM	ADRESSE
1	Grasset	61 rue des Saints-Pères, 75006 Paris
2	Folio	5 rue Gaston-Gallimard, 75328 Paris
3	Le Livre de Poches	21 Rue du Montparnasse, 75006 Paris

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

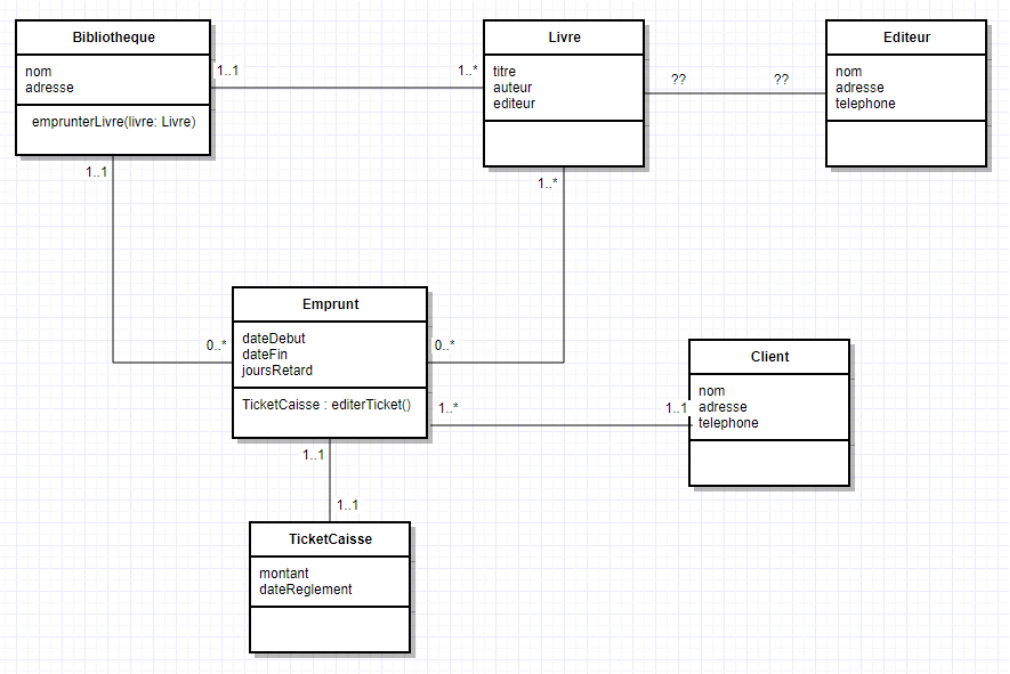
ID	NOM	ADRESSE
1	Grasset	61 rue des Saints-Pères, 75006 Paris
2	Folio	5 rue Gaston-Gallimard, 75328 Paris
3	Le Livre de Poches	21 Rue du Montparnasse, 75006 Paris
4	Seuil	57 rue Gaston-Tessier, 75166 Paris

Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Modèle physique de données

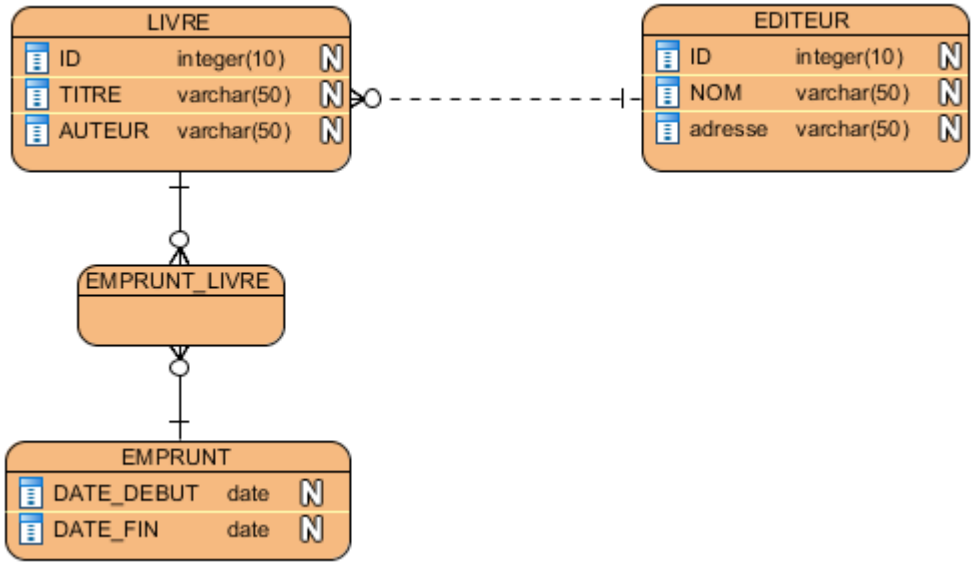
- **Création des tables à partir du diagramme de classes**
  - 1 classe = 1 table
  - 1 attribut d'instance = 1 colonne de table
  - 1 association entre classe = 1 relation entre tables
    - Relation 1..N
    - Relation N..1
    - Relation N..N
    - Relation 1..1



- **Passage du Diagramme de classes au Modèle Physique de données**
  - Relation 1..N => clé étrangère du "côté du N"
  - Relation N..1 => clé étrangère du "côté du N"
  - Relation N..N => table d'association avec 2 clés étrangères
  - Relation 1..1 => clé étrangère d'un côté ou de l'autre

- **Différences entre Diagramme de classes et MPD**
  - Il n'y a pas de clé étrangère dans un diagramme de classes mais il y en a dans un MPD (voir diapos suivantes)
  - Il y a des relations N..N dans un digramme de classes qui sont remplacées par des tables de jointures dans un MPD (voir diapos suivantes)

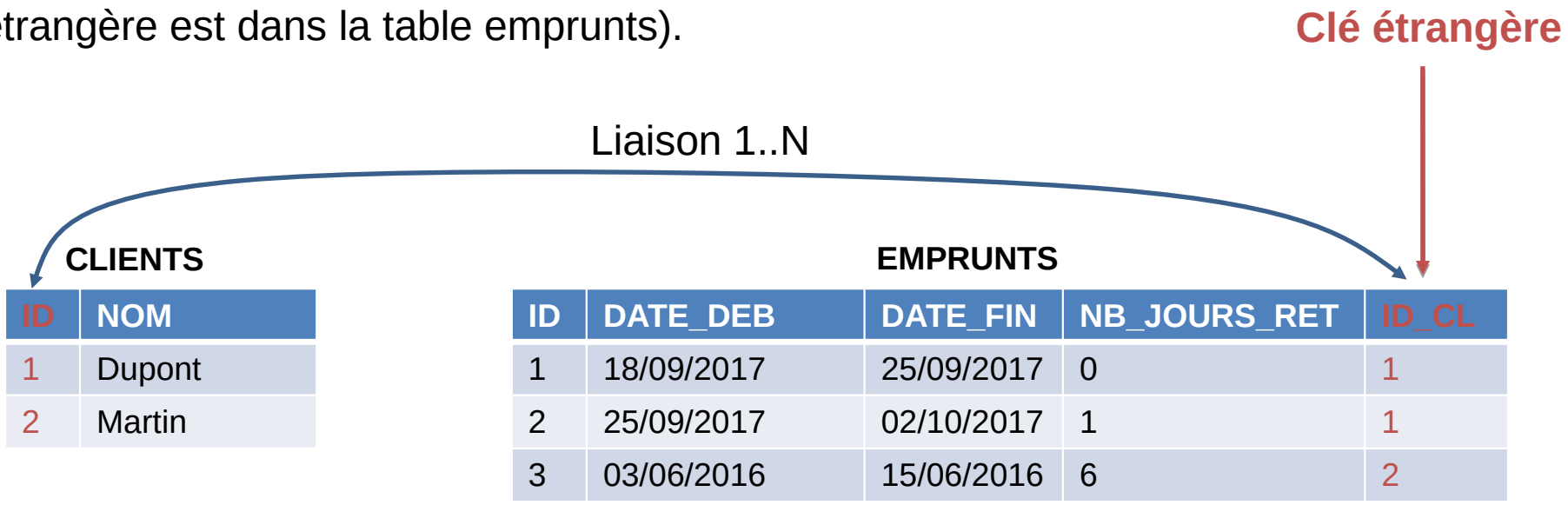
- **Modèle physique de données**
  - Relations entre entités
  - Les relations N..N sont modélisées sous forme d'une table de jointure



DEFINITIONS

- **Modèle physique de données**
  - Relation 1..N, N..1 ou 1..1
    - Attribut de jointure ou **clé étrangère (Foreign Key)**
  - Exemple: entre Client et Emprunt

- **Astuce :** dans une relation 1..N la clé étrangère est toujours "du côté du **N**", ici dans la table EMPRUNTS (c'est le client qui a de 1 à N emprunts donc la clé étrangère est dans la table emprunts).



CLIENTS

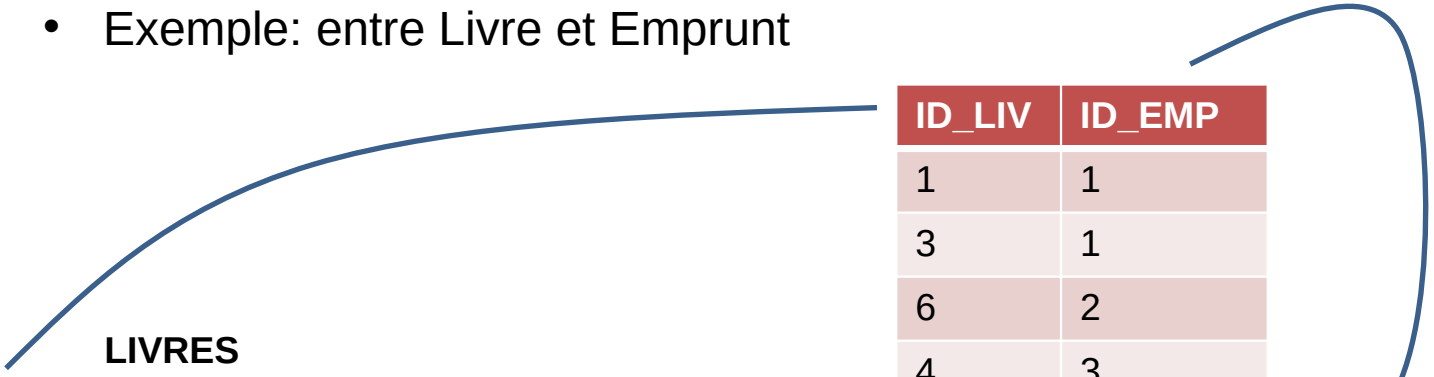
ID	NOM
1	Dupont
2	Martin

EMPRUNTS

ID	DATE_DEB	DATE_FIN	NB_JOURS_RET	ID_CL
1	18/09/2017	25/09/2017	0	1
2	25/09/2017	02/10/2017	1	1
3	03/06/2016	15/06/2016	6	2

DIAGRAMMES

- **Modèle physique de données**
  - Relation N..N
    - **Table de jointure**
  - Exemple: entre Livre et Emprunt



ID_LIV	ID_EMP
1	1
3	1
6	2
4	3

LIVRES

ID	TITRE
1	Germinal
2	Madame Bovary
3	L'assommoir
4	Le Rouge et Le Noir
5	Guerre et Paix
6	Le Petit Prince

EMPRUNTS

ID	DATE_DEB	DATE_FIN	NB_JOURS_RET	ID_CL
2	18/09/2017	25/09/2017	0	1
3	25/09/2017	02/10/2017	1	1
1	03/06/2016	15/06/2015	6	2

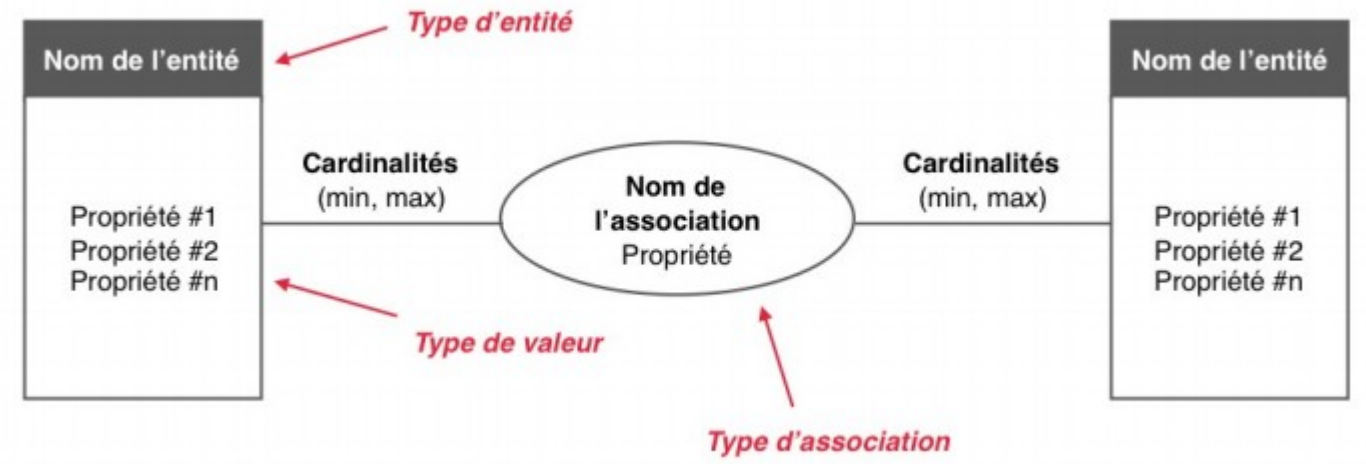
Diginamic
Formation

« Dynamisez votre carrière digitale ! »

MERISE

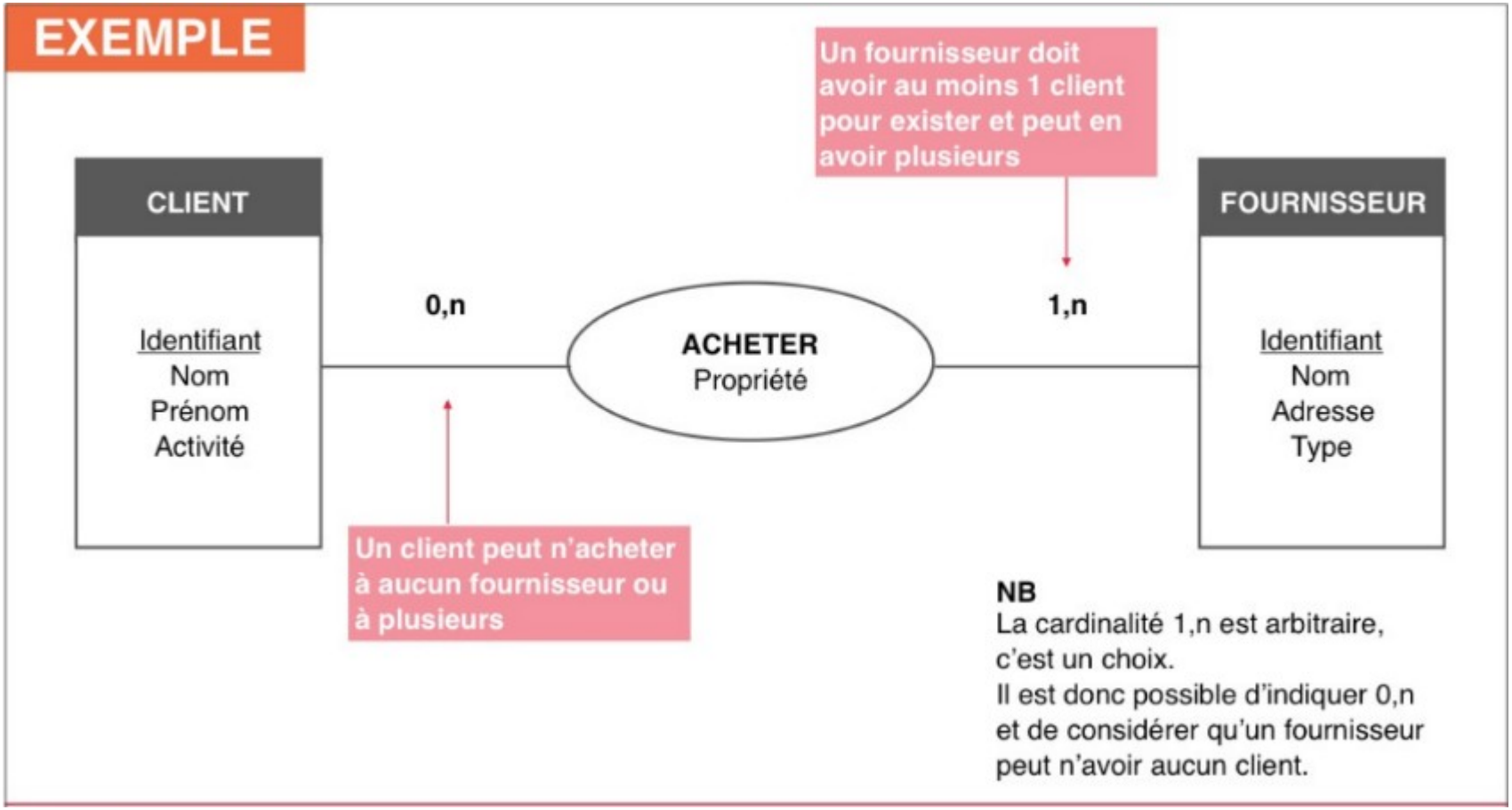
Méthode MERISE

- Méthode de conception française de développement et de réalisation de projets informatiques
- Basée sur la séparation du modèle de données en 2 modèles:
  - Modèle conceptuel de données
  - Modèle physique de données
- Complémentaire à UML.



Méthode MERISE

- Attention les cardinalités sont inversées / UML



Diginamic
Formation
« Dynamisez votre carrière digitale ! »

Normalisation

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

FORMES NORMALES

- 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.

NORMALISATION

- 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

NORMALISATION

- 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

NORMALISATION

- **Exemple de table non normalisée:**
  - Données non atomiques
  - SQL peu performant pour analyser des chaines 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

PREMIÈRE FORME NORMALE

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

PREMIÈRE FORME NORMALE

- **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**.

DEUXIÈME FORME NORMALE

- **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

ARTICLE	ID_FOUR
Boulon de 12	1
Boulon de 14	1
Perceuse XT2000	2

ID	LIBELLE	CODE	SIRET
1	SARL Dupont	2187	AF102
2	Générale de Mécanique	3128	GM227

TROISIÈME FORME NORMALE

- **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	FOURNISSEUR	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

Diginamic
Formation

« Dynamisez votre carrière digitale ! »

Bases de données SQL vs NoSQL

Système de gestion de base de données relationnelles:

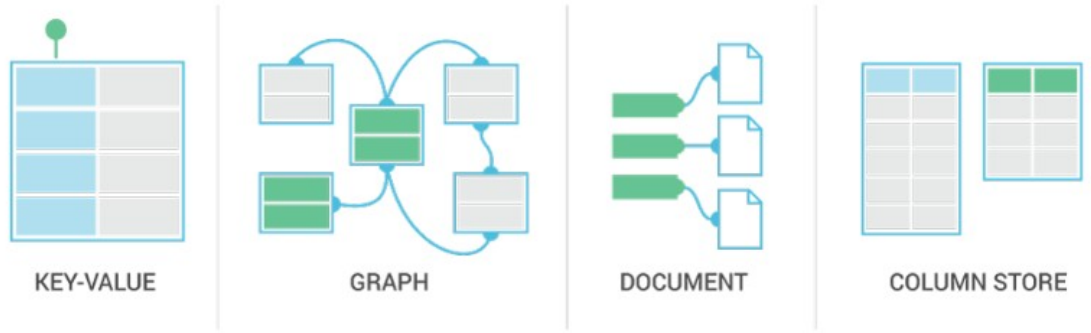
- Les données sont stockées dans des tables, appelées aussi relations

	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 – LES TYPES (1/2)

- **Bases NoSQL, plusieurs types:**

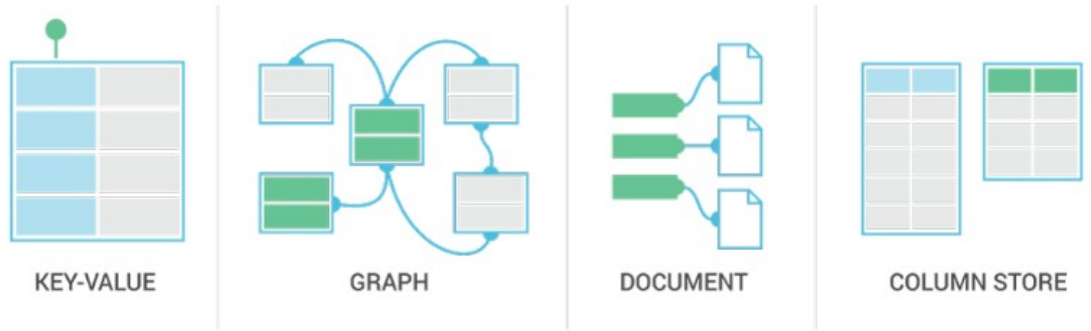


- **Key-Value**: les données sont stockées avec une clé unique. Bases les plus simples.
  - Bases de ce type: Redis, Amazon DynamoDB

- **Document**: comme la Key-Value mais la valeur peut-être structurée. Plusieurs formats reconnus pour la value: 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 – LES TYPES (2/2)

- **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 oriéntée objet dont le fonctionnement repose sur la théorie des graphes**
  - Bases de ce type: DataStax, InfiniteGraph, MarkLogic, etc..

ACIDITE D'UNE BASE

- **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.

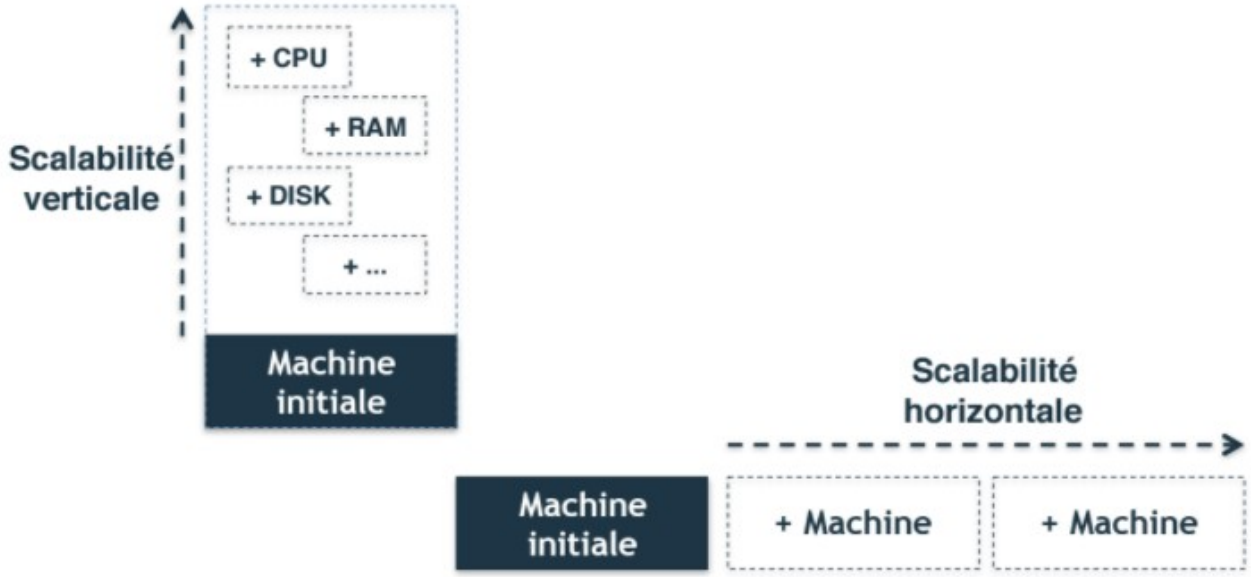
- **Durabilité:**
  - Une fois une transaction validée, la base de données est dans un état stable, quel que soit l'évènement qui peut se produire, comme une panne d'électricité.

BASES ET ACIDITÉ

- **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 **ww.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…**

SCALABILITÉ

- **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 ET SCALABILITE

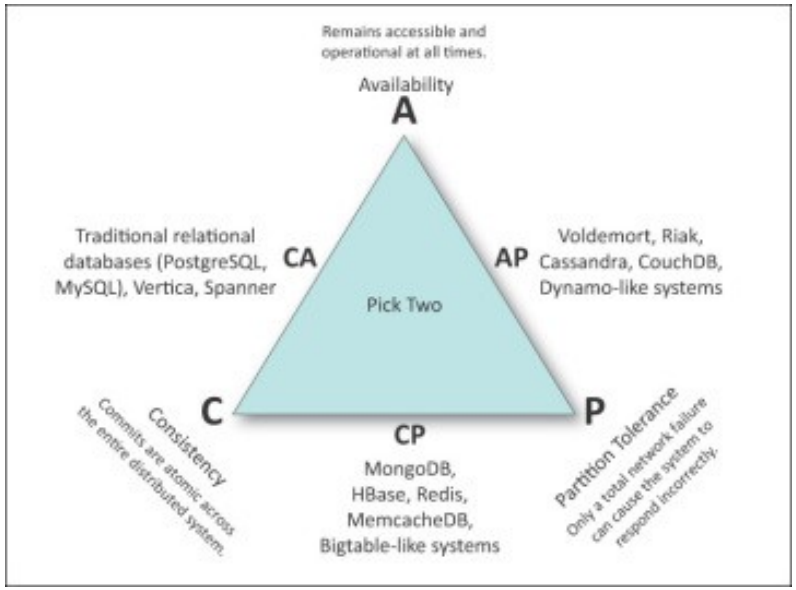
- **Bases de données relationnelles.**
  - Exige une scalabilité verticale.
  - Pas compatible avec la scalabilité horizontale car on ne peut pas garantir la cohérence des données. En effet si une base de données est répartie sur plusieurs machines (ie. nœud) il y a un temps de réplication des informations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

- **Bases NoSQL.**
  - sont conçues pour la scalabilité horizontale.

BASES ET THÉORÈME CAP

- **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 horizontal (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)



NOSQL ET DATASCIENCES (1/2)

Le Big Data porte essentiellement sur des données échues (historiques) qui ne nécessite pas la cohérence.

L'explosion des volumes des données ont nécessité des innovations :

- Bases NoSQL en terme d'architecture de stockage (scalabilité horizontale).
- Modèle Map-Reduce en terme d'algorithmie, permettant de répartir les traitements sur plusieurs machines

NOSQL ET DATASCIENCES (2/2)

Ces approches d'architecture et d'algorithmie permettent de répartir les données sur plusieurs machines distinctes.

- Map Reduce : modèle de traitement d'un ensemble de données.
- Hadoop / Cassandra / MongoDB : framework de développement d’application utilisant le modèle Map Reduce.

Nouvelle discipline est apparue: les data sciences.

TP n°1: Modélisation