

BASE DE DONNÉES

JOURNÉE 1 - MATIN

1

Karine BRIFAUT

BASE DE DONNÉES DÉFINITIONS

- Définition - Une base de données (**BD**) :
 - **collection d'informations** ou de données qui existent sur une longue période de temps et qui décrivent les activités d'une ou plusieurs organisations,
 - ensemble de données modélisant les objets d'une partie du monde réel et servant de support à une application informatique

BASE DE DONNÉES DÉFINITIONS

- Définition - Système de gestion de bases de données (**SGBD**) = DataBase Management Systems – DBMS :
 - **ensemble de logiciels** systèmes permettant aux utilisateurs d'insérer, de modifier et de rechercher efficacement des données spécifiques dans une grande masse d'informations (pouvant atteindre plusieurs Terra) partagée par de multiples utilisateurs.

BASE DE DONNÉES DÉFINITIONS

- Exemples de SGBD:
 - Orienté Relationnel :
 - Oracle, SQL Server, Sybase, DB2, Informix, ...,
 - Access, Paradox, 4D, ...,
 - MySQL, MariaDB, PostGres, ...
 - Orientés-objets :
 - Versant, Objectstore, O2, ...
 - Big Data
 - Teradata, Oracle, EMC, MongoDB, Cassandra

BASE DE DONNÉES

OBJECTIF D'UN SGBD

- Principaux composants d'un SGBD :
 - Système de gestion de fichiers
 - Gestionnaire de requêtes
 - Gestionnaire de transactions
- Principales fonctionnalités d'un SGBD :
 - Contrôle de la redondance d'information
 - Partage des données
 - Gestion des autorisations d'accès
 - Vérifications des contraintes d'intégrité
 - Sécurité et reprise sur panne

BASE DE DONNÉES

TYPE DE BASE DE DONNÉES

- Il existe différentes techniques pour stocker les données
- Chaque approche a ses avantages et ses inconvénients
 - Cout
 - License, besoin physique, apprentissage
 - Performances
 - Charge

BASE DE DONNÉES

BASE DE DONNÉES OBJET

- Stock d'informations groupées sous formes **de collections d'objets persistants**.
- Dans une base de données objets les informations sont regroupées sous forme d'objets ⇔ un conteneur logique qui englobe des informations et des traitements relatifs à une chose du monde réel.
- Les bases de données objets sont mises en œuvre par
 - un système de gestion de base de données objet — logiciel qui manipule le contenu de la base de données
 - un programme écrit dans un langage de programmation orientée objet.

BASE DE DONNÉES

BASE DE DONNÉES STRUCTUREES

- Fait référence à des bases de données de type **Big Data**
- Les big data, datamasse, ou mégadonnées, parfois appelées données massives, désignent des ensembles de données qui deviennent tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données.
- Volume qui se compte en zettaoctets (10^{21} Octets = 1 Giga à la puissance 3)

BASE DE DONNÉES

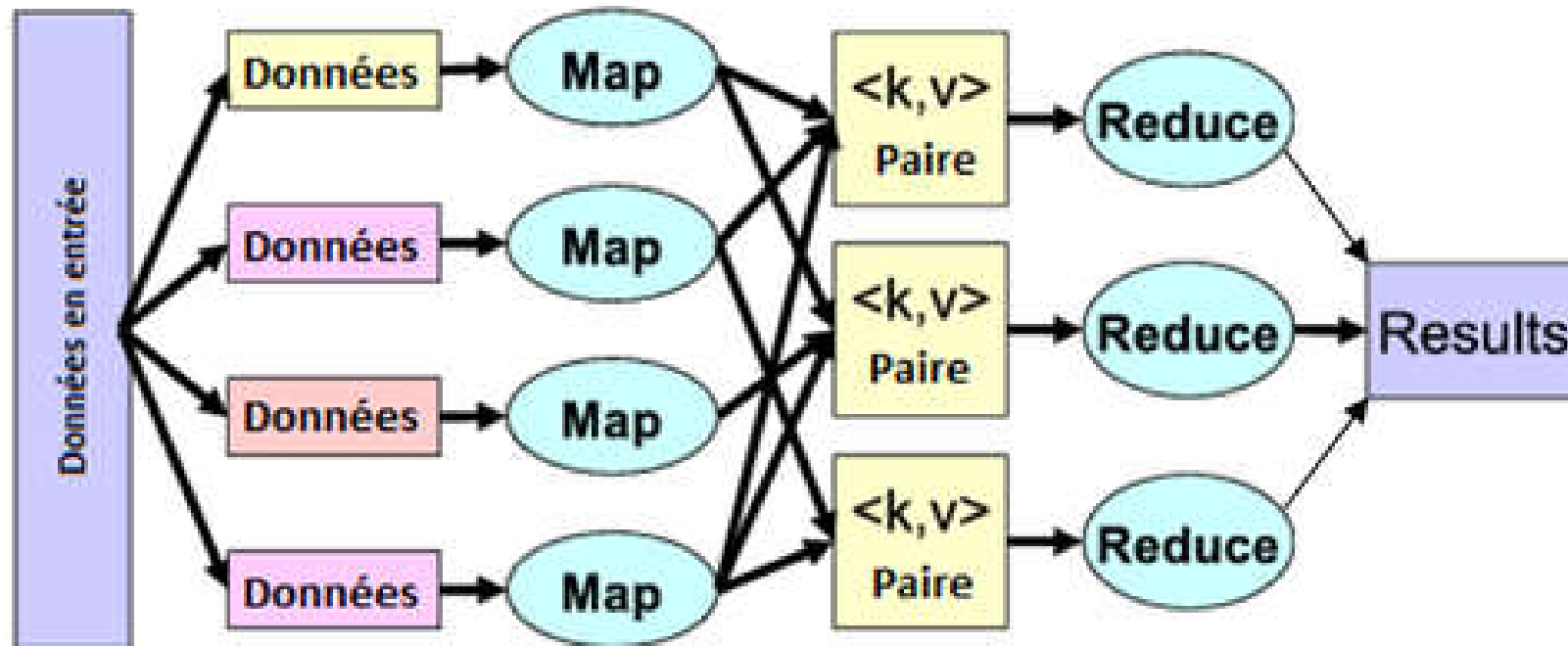
BASE DE DONNÉES STRUCTUREES

- Les bases de données relationnelles classiques ne permettent pas de gérer les volumes de données du Big Data.
- De nouveaux modèles de représentation permettent de garantir les performances sur les volumétries en jeu. Ces technologies, dites de Business Analytics & Optimization (**BAO**) permettent de gérer des bases massivement parallèles.
- Des patrons d'architecture "Big Data Architecture framework (BDAF)" sont proposés par les acteurs de ce marché comme MapReduce développé par Google et utilisé dans le framework Hadoop.
- Avec ce système les requêtes sont séparées et distribuées à des nœuds parallélisés, puis exécutées en parallèles (map). Les résultats sont ensuite rassemblés et récupérés.

BASE DE DONNÉES

BASE DE DONNÉES STRUCTUREES

- Exemple de fonctionnement Map Reduce



BASE DE DONNÉES

BASE DE DONNÉES NOSQL

- Héritières des Big Data.
- Bases à très gros volume, pas de structuration (ni relationnel, ni objet)
- Exemple de consommateur : Google, Amazon, Facebook
- Exemple de produit : Hbase, MongoDB, Cassandra

BASE DE DONNÉES

BASE DE DONNÉES NOSQL

- Concrètement, il faut imaginer une seule table avec plein de colonne.
- Pas de jointure mais une représentation des données qui colle aux besoins

```
{  
  'id': 10,  
  'nom': 'dupont',  
  'prenom': 'david',  
  'email': 'me@palo-it.com',  
  'adresse':  
  {  
    'libelle': '10 rue du test',  
    'ville': 'paris',  
    'pays': 'France',  
    'code postal': '75009'  
  }  
}
```

BASE DE DONNÉES

LE MODÈLE RELATIONNEL

- Objectif
 - Permettre un haut degré d'indépendance des programmes et des activités
 - Fournir une base solide pour traiter les problèmes de cohérences et de redondance des données
 - Permettre la manipulation des données via un langage simple
- Résultat
 - Des tables : matrice à deux dimensions
 - Des contraintes
 - Le langage SQL

BASE DE DONNÉES

LE MODÈLE RELATIONNEL

- Il est fondé sur la théorie mathématique des relations.
 - Héritière des ensembles
- Quatre notions importantes
 - Domaine
 - Attribut
 - Relation
 - Tuple
- Elle ne définit pas
 - D'ordre dans la définition d'une structure
 - D'information physique sur comment est stockée la donnée

BASE DE DONNÉES

DOMAINE - ATTRIBUT - RELATIONS

- Domaines
 - Sont des ensembles de valeurs caractérisés par un nom
 - Les données y prennent valeur.
- Relation
 - Sous-ensemble du produit cartésien d'une liste de domaines caractérisé par un nom
- Attribut
 - Colonne d'une relation caractérisée par un nom
- Tuple
 - Ligne d'une relation correspondant à un enregistrement

BASE DE DONNÉES CONTRAINTES

- Il est possible d'ajouter des contraintes dans la représentation des données relationnelles
- Ont les appelle aussi des règles d'intégrité structurelle
- On y retrouve les notions
 - De clefs
 - De Contraintes référentielles
 - De Contraintes valeur nulle
 - De Contraintes d'entité
 - De Contraintes de domaine

BASE DE DONNÉES CONSTRAINTES

- Unicité de clefs
 - Par définition, une relation est un ensemble de tuples. Un ensemble n'ayant pas de doublons, il faut pouvoir identifier les tuples de manière unique.
- Une clef
 - Ensemble minimal d'attributs dont la connaissance des valeurs permet d'identifier un tuple unique de la relation considéré.

BASE DE DONNÉES

CONSTRAINTES

- Contraintes de références
 - Contrainte d'intégrité portant sur une relation R1, consistant à imposer que la valeur d'un groupe d'attributs apparaisse comme valeur de clef dans une autre relation R2.
- Valeurs nulles
 - Valeur conventionnelle introduite dans une relation pour représenter une information inconnue ou inapplicable
- Contrainte d'entité
 - Contrainte d'intégrité imposant que toute relation possède une clef primaire et que tout attribut participant à cette clef primaire soit non nul.
- Contrainte de domaine
 - Contrainte d'intégrité imposant qu'une colonne d'une relation doit comporter des valeurs vérifiant une assertion logique

BASE DE DONNÉES

BASE DE DONNÉES RELATIONNELLE

- Ensemble d'informations **décomposées et organisées** dans des matrices appelées **relations ou tables** conformément au modèle de données relationnel.
- Le contenu de la base de données peut ainsi être synthétisé par des opérations d'algèbre relationnelle telles que l'intersection, la jointure et le produit cartésien.
- Le modèle de données relationnel permet d'utiliser les **opérateurs de l'algèbre relationnelle** pour retrouver une information quelconque stockée dans la base de données, ainsi que les informations connexes.

BASE DE DONNÉES

BASE DE DONNÉES RELATIONNELLE

- Opérateurs ensemblistes :
 - UNION, INTERSECTION, DIFFERENCE, PRODUIT UNION, PRODUIT
- Opérateurs relationnels spécifiques :
 - SELECTION, PROJECTION, JOINTURE, DIVISION
- Opérateurs dérivés :
 - JOINTURE EXTERNE, SEMI-JOINTURE, ...

BASE DE DONNÉES

OPERATEUR - UNION

- $T1 \cup T2$ est la relation contenant les tuples appartenant à $T1$ ou à $T2$

T1	C1	C2	C3		T2	C1	C2	C3
	x1	x2	x3			x1	x2	x3
	y1	y2	y3			w1	w2	w3
	z1	z2	z3			z1	z2	z3
	a1	b2	b3					
$T1 \cup T2$	C1	C2	C3					
	x1	x2	x3					
	y1	y2	y3					
	z1	z2	z3					
	a1	b2	b3					
	w1	w2	w3					

- Commutatif et associatif

BASE DE DONNÉES

OPÉRATEUR - INTERSECTION

- $T1 \cap T2$ est la relation contenant les tuples appartenant à $T1$ **et** à $T2$

T1	C1	C2	C3		T2	C1	C2	C3
	x1	x2	x3			x1	x2	x3
	y1	y2	y3			w1	w2	w3
	z1	z2	z3			z1	z2	z3
	a1	b2	b3					
$T1 \cap T2$	C1	C2	C3					
	x1	x2	x3					
	z1	z2	z3					

- Commutatif et associatif

BASE DE DONNÉES

OPERATEUR – DIFFERENCE

- $T1 - T2$ est la relation contenant les tuples appartenant à $T1$ **n'appartenant pas** à $T2$

T1	C1	C2	C3		T2	C1	C2	C3
	x1	x2	x3			x1	x2	x3
	y1	y2	y3			w1	w2	w3
	z1	z2	z3			z1	z2	z3
	a1	b2	b3					
$T1 - T2$	C1	C2	C3					
	y1	y2	y3					
	a1	b2	b3					

- Non commutatif et non associatif

BASE DE DONNÉES

OPERATEUR – PRODUIT CARTÉSIEN

- $T1 \times T2$ est défini par la relation $Q(A1, \dots, An, B1, \dots, Bp)$ telle que : $(a1, \dots, an, b1, \dots, bp) \in Q$ ssi $(a1, \dots, an) \in T1$ et $(b1, \dots, bp) \in T2$

T1	A	B	C		T2	X	Y	
	a1	b1	c1			x1	y1	
	a2	b2	c2			x2	y2	
	a3	b3	c3					
T1 x T2	A	B	C	X	Y			
	a1	b1	c1	x1	y1			
	a2	b2	c2	x1	y1			
	a3	b3	c3	x1	y1			
	a1	b1	c1	x2	y2			
	a2	b2	c2	x2	y2			
	a3	b3	c3	x2	y2			

- Commutatif et associatif

BASE DE DONNÉES

OPERATEUR - SELECTION

- Selection

- à partir d'un ensemble A, obtenir un ensemble B dont les n-uplets contiennent certaines composantes des n-uplets de A.

Selon une approche de matrice, le nombre de colonne est modifié (**seules les colonnes pertinentes sont conservées**) et éventuellement le nombre de ligne. En effet, si une ou plusieurs colonnes retenues permettent d'identifier une ligne de manière unique (par exemple, une colonne contient les numéros d'employé ou de sécurité sociale), alors il n'y aura pas de modification du nombre de lignes. Mais dans le cas contraire, en vertu du principe de la théorie des ensembles excluant la présence de doublons, le nombre de ligne sera éventuellement inférieur ;

BASE DE DONNÉES

OPERATEUR – SELECTION

- La sélection : opérateur SELECT - sélection d'un sous-ensemble de tuples d'une relation qui vérifient une condition

NV	Marque	Couleur	Puissance
1234 GH 75	Renault	verte	7
678 AZ 23	Citroen	blanche	6
345 AZ 34	Renault	rouge	5
Puissance = 6(Vehicule)			
678 AZ 23	Citroen	blanche	6
Marque=Citroen(Vehicule)			
678 AZ 23	Citroen	blanche	6

BASE DE DONNÉES

OPERATEUR - PROJECTION

- Projection

- à partir d'un ensemble A, obtenir un ensemble B dont les n-uplets contiennent certaines composantes des n-uplets de A.

Selon une approche de matrice, le nombre de colonne est modifié (**seules les colonnes pertinentes sont conservées**) et éventuellement le nombre de ligne. En effet, si une ou plusieurs colonnes retenues permettent d'identifier une ligne de manière unique (par exemple, une colonne contient les numéros d'employé ou de sécurité sociale), alors il n'y aura pas de modification du nombre de lignes. Mais dans le cas contraire, en vertu du principe de la théorie des ensembles excluant la présence de doublons, le nombre de ligne sera éventuellement inférieur ;

BASE DE DONNÉES OPERATEUR - PROJECTION

NV	Marque	Couleur	Puissance
1234 GH 75	Renault	verte	7
678 AZ 23	Citroen	blanche	6
345 AZ 34	Renault	rouge	5
NV,Couleur(Vehicule)			
1234 GH 75		verte	
678 AZ 23		blanche	
345 AZ 34		rouge	
Marque(Vehicule)			
	Renault		
	Citroen		

BASE DE DONNÉES

OPERATEUR – JOINTURE

- La jointure : noté \bowtie combiner une paire de tuples de deux relations en un seul tuple.

Client					Vente				
	numero	nom	adresse	tel		numero	id_produit	id_client	date
	1	Jhon	Paris	0154689789		1	6	1	01/2014
	2	Durant	Nice	NULL		2	8	1	02/2014
	3	Loarq	Nice	0654689789		3	233	2	01/2014
	4	Dupont	Vanves	NULL		4	5489	4	01/2013

Client					Vente				
	numero	nom	adresse	tel		numero	id_produit	id_client	date
	1	Jhon	Paris	0154689789		1	6	1	01/2014
						2	8	1	02/2014

BASE DE DONNÉES

OPERATEUR – DIVISION

- La Division permet de rechercher dans une relation les sous tuples qui sont complétés par tous ceux d'une autre relation. Elle permet ainsi d'élaborer la réponse à des questions de la forme « quel que soit x, trouver y »

Vin					Qualité		
	Cru	Mill	Qualité			Mil	Qualité
	Volnay	1983	A			1983	A
	Chablis	1979	A			1979	A
	Chablis	1983	A				
	Julienas	1986	B				
				Cru			
				Chablis			

BASE DE DONNÉES MANIPULATION

1. Installation de la base de données (le server)
2. Gestion des droits, paramétrages, ...
3. Modélisation des données
4. Installation d'un client
5. Ajustement des modèles en fonction des performances
6. Backup / load balancing

BASE DE DONNÉES INSTALLATION

1. Contraintes d'architectures physiques
 - Mémoire, processeurs
 - Espace disque
2. Choix de l'OS
 - Windows, Linux, Unix, ... : Pourquoi ?
3. Bien respecter les contraintes associés à l'OS
 - Version, patch, droits

BASE DE DONNÉES INSTALLATION

4. Installation de l'OS

5. Installation de la base

- Téléchargement du binaire associé à la plateforme cible (Windows, Unix, 32b, 64b)
- Lecture du readme associé
- Lancement du/des scripts d'installation
- Choix des répertoires physiques
- Choix de sécurisation initial

BASE DE DONNÉES

INSTANCES ET SCHÉMA

- Schéma de base de données :
 - description de la structure des données
 - ensemble de définitions exprimées en langage de description de données (DDL – Data Definition Language) (LDD en français)
 - create, alter, drop, rename
- Instances de base de données :
 - données de la base à un instant T
 - manipulées par un langage de manipulation de données (DML - Data Manipulation Language) (LMD en français)
 - select, insert, update, delete