

Lexique des bases de données

Par Dimitri PIANETA

Mars 2021

Sommaire

A-	3
B-	5
C-	8
D-	15
E-	20
F-	22
G-	23
H-	25
I-	26
J-	28
K-	29
L-	30
M-	31
N-	35
O-	37
P-	39
R-	43
S-	45
T-	49
U-	50
V-	52
X-	55
Z-	56

A-

algèbre relationnel : L'algèbre relationnelle est la théorie mathématique inventée par Edgar Franck Codd sur laquelle repose les bases de données relationnelles. Dérivée de la théorie des ensembles, cette algèbre utilise des opérateurs ensemblistes et des opérateurs relationnels qui permettent de définir comment les informations sont liées entre elles et comment on peut y accéder. À partir d'un nombre basique d'opérations, il est ainsi possible d'accéder des données qui font sens entre elles (relations).

Opérateurs ensemblistes de l'algèbre relationnelle

- **Union** : sélectionne les données de tous les ensembles
- **Intersection** : sélectionne uniquement les données communes
- **Différence** : sélectionne uniquement les données faisant partie d'un ensemble mais pas de l'autre
- **Produit cartésien** : sélectionne tous les couples possibles entre les éléments des différents ensembles

Opérateurs relationnels

- Sélection
- Projection
- Jointure
- Division

analyse multidimensionnelle : ensemble des techniques OLAP ou OLTP permettant d'analyser statistiquement des données à partir de valeurs appelées « dimensions » que l'on agrège pour les besoins de l'analyse.

anomalie (transactionnelle) : obtention d'information(s) erronée(s) au cours d'un traitement informatique portant sur des données. Les anomalies se combattent par l'utilisation judicieuse d'un niveau d'isolation des transactions.

API (*Application Programming Interface* ou interface de programmation) : définit la manière dont un composant informatique communique avec un autre. Le plus souvent, une API consiste en une bibliothèque de composants logiciels, c'est-à-dire d'une liste de fonctions ou procédures considérées comme utiles pour d'autres programmes et exploitables à l'aide de n'importe quel langage de développement informatique.

arbre équilibré (*B-Tree* ou *Balanced Tree*) : structure de données arborescente dont la profondeur est constante, c'est-à-dire que toutes les feuilles sont situées au même niveau.

ASCII (*American Standard Code for Information Interchange*) : norme américaine employée pour le codage des caractères alphanumériques. Elle a été inventée par l'Américain Bob Bemer en 1961, pour l'utilisation des télétypes et pour le codage des littéraux dans les systèmes informatiques.

assertion (affirmation) : dans les bases de données, règle globale (contrainte multitable) qui doit toujours être vérifiée.

association : Les bases de données sont constituées de différentes données homogènes qui discutent entre elles (les entités). Lorsque l'on représente les liaisons et les interactions entre ces différentes données cohérentes, on parle d'association.

En conception de base de données, on dit qu'une association précise et explicite le lien entre plusieurs entités. Une association se fait entre 2 entités minimum mais il n'y a pas de maximum.

asynchrone (contraire de synchrone) : se dit d'un mécanisme ou d'un traitement découplé d'un autre alors qu'il en est dépendant.

atomique (insécable) : se dit d'un programme qui doit être exécuté en tout ou rien garanti par le biais de transactions. Se dit d'une donnée qui ne peut être subdivisée. Par exemple, un numéro de sécurité sociale n'est pas une donnée atomique. On parle aussi d'atomicité.

attribut : Quand on découvre la conception de bases de données, on utilise beaucoup de termes sur mesure. Pas de chance, certains sont synonymes. C'est le cas d'Attribut qui peut être remplacé par Propriété (voir ce lien, la définition y est complète).

Un attribut est un identificateur (idéalement explicite – un nom parlant) et correspond à une donnée que l'on va informatiser en base et qui se caractérise par un nom, un emplacement, un type de données ainsi que d'autres caractéristiques (identifiant par exemple...).

authentification : opération consistant à vérifier les informations d'identification d'un processus tentant d'accéder au serveur. Il est autorisé s'il figure sur la liste d'acceptation.

Azure : solution globale de cloud computing de Microsoft qui propose, entre autres, des services de bases de données SQL Server.

autocommit : validation automatique d'une transaction.

Auto-incrément : donnée numérique dont la particularité est de fournir une nouvelle valeur résultat de la précédente augmentée d'un pas de croissance généralement de 1. La graine d'un auto-incrément est la valeur initiale et peut aussi être définie à la valeur que l'on souhaite. Dans SQL Server, les autoincrément sont définis soit par la propriété IDENTITY associée à une colonne d'une table, soit par le biais de séquence.

autojointure : jointure effectuée d'une table ou d'une vue sur elle-même.

autorisation : concerne la sécurité d'accès des utilisateurs aux bases de données relationnelles.

B-

Balanced Tree, voir arbre équilibré.

balayage (ou *scan*) : technique qui consiste à accéder à une donnée d'une table ou d'un index par lecture de toutes les lignes l'une après l'autre.

balise : élément du langage XML définie par une chaîne de caractères constituant un identificateur et placé entre chevrons, par exemple <uneBalise>.

Banque de données : Lorsque plusieurs bases de données fonctionnent au sein d'un système homogène (entreprise, recherche médicale par exemple), on parle de banque de données ou data bank en anglais.

Les informations peuvent alors être physiquement être présentes au même endroit ou pas (lieu physique ou site internet regroupant différentes sources). L'idée principale est de faciliter la recherche en proposant un lieu commun.

base de données : ensemble d'informations (données structurées) enregistrées sur des supports persistants, accessibles par un ordinateur, et qui sont destinées à satisfaire simultanément plusieurs utilisateurs de manière sélective et sécurisée et dans un temps opportun. L'interrogation et la mise à jour s'effectuent par l'intermédiaire d'un logiciel, le SGBD.

Base de données objet : Dans la grande famille des bases de données existe le vilain petit canard : les bases de données objet. Apparues en même tant que les langages de programmation orientés objet dans les années 1990, les bases de données objet se distingue des bases de données classiques (les bases de donnée relationnelles).

En effet, une base de données objet manipule des objets. En informatique, un objet est une notion théorique qui représente une entité complexe et évolutive. Plutôt que de traiter tous les aspects en direct d'un objet, on y accède via un mécanisme simplifié qui masque la complexité interne. On a donc pas accès à une grande matrice de données (colonnes et lignes comme dans une base de données relationnelles) mais plutôt des portes d'entrées spécifiques que l'on peut ouvrir (ou pas) afin d'y ajouter, modifier et supprimer des contenus particuliers. Classes, héritage, encapsulation, polymorphisme... tout ce qui existe en POO s'applique aux bases de données Objet.

Les bases de données objet prennent tout leur sens lorsqu'il s'agit de modéliser des données très riches et complexes avec de nombreuses variations : documents multimédias, données géographiques, conception par ordinateur...

De par sa conception initiale, les bases de données sont rapides. Rarement besoin de jointures. De plus, les SGBDO permettent de réaliser des opérations sur mesure pour chaque objet au delà de Ajout/Modification/Suppression... Cerise sur le gâteau, une base de données objet et un logiciel réalisé en programmation orientée objet ne nécessitent qu'une seule et même modélisation informatique. Les objets utilisés en programmation se retrouvent directement en base de données.

Base de données en colonnes : Aujourd'hui, les bases de données rangent généralement les données en ligne : visuellement, on dispose de vues avec des colonnes qui contiennent le nom des champs et en ligne les valeurs.

Pour les bases de données en colonne, c'est différent car les données sont stockées en colonne. Pour les utilisateurs de haut niveau, ce changement ne se ressent pas du tout. Mais pour l'organisation et la manipulation des données par les outils informatiques, tout change. Ainsi, au lieu de manipuler les données ligne par ligne une fois sérialisée (c'est à dire mises bout à bout pour être traitées plus rapidement), les données sont utilisées colonne par colonne.

Les données en ligne :

1,Jean,DUPONT,7400;2,Julie,VAUQUIER,01100;3,Estelle,MALIA,75400;4,Théo,CONTIN,89130;

Puis en colonne :

1,2,3,4;Jean,Julie,Estelle,Théo;DUPONT,VAUQUIER,MALIA,CONTIN;74000,01100,75400,89130;

L'idée de travailler avec les colonnes n'est pas nouveau (prémices en 1969) mais ce n'est que depuis 2005 avec l'apparition de tels outils distribués sous licence Open Source que les bases de données en colonnes ont vraiment décollé.

En fonction des volumes de données à traiter et des usages, les bases de données sont plutôt orientées lignes ou colonnes. Les bases de données OLTP sont généralement en lignes et les bases de données OLAP sont plutôt en colonnes.

Base de données relationnelle : **Relationnelle** se dit d'une base de données qui manipule des informations hiérarchisées qui sont liées entre elle par des contraintes de relations. À la différence d'autres types de bases de données, les bases de données relationnelles fonctionnent avec des tableaux à deux dimensions nommés tables qui sont liés à d'autres tables par un mécanisme de relations reposant sur l'algèbre relationnelle. Ces relations définissent et précisent les liens entre les tables.

Conceptualisé par Edgar Frank Codd en 1970, la base de données relationnelle est aujourd'hui la famille la plus répandue des bases de données.

batch : ensemble de commande SQL ou Transact SQL généralement exécutées en une seule fois. Contrairement à une procédure, un batch n'est pas stocké dans la base et n'admet pas d'arguments. Il peut cependant renvoyer des messages et des jeux de données et encapsuler des transactions. Il est généralement utilisé pour des traitements à usage unique, comme la création des objets d'une base.

BCM (Bulk Changed Map) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer par un bit à 1 parmi les 63 904 extensions suivantes lesquelles ont été visées par une opération BULK INSERT depuis la dernière sauvegarde complète.

benchmark : banc d'essai destiné à étalonner et comparer la performance des systèmes.

BI (Business Intelligence) : ensemble des techniques de l'informatique décisionnelle (OLAP) destinée aux décideurs afin de piloter l'entreprise par la prise de décision découlant de l'analyse et du forage des données.

BIOS (*Basic Input Output System*) : système élémentaire pour les entrées/sorties. Composant microprogrammé d'un ordinateur destiné à assurer les tâches d'entrée et de sortie de bas niveau pour les périphériques.

Big data : Existant depuis les années 2000, c'est le mot à la mode depuis 2010. Tout le monde parle BIG DATA (on voit plus rarement le terme recommandé mégadonnées).

Avec l'augmentation des capacités de stockage et la réduction des coûts liés, les entreprises collectent de plus en plus d'informations. Ces informations structurées sont rangées dans des bases de données (parmi d'autres outils) et sont parfois tellement volumineuses que leur exploitation nécessite des outils sur mesure. On parle alors de big data.

Ce qui est nouveau et très intéressant c'est que l'on peut exploiter ces données qui, prises une à une, n'ont pas vraiment d'intérêt mais qui, cumulées, recèlent une vraie valeur :

- Prévisions, tendances, prospective, recherche sur les attentes, les usages, les besoins des utilisateurs;
- Usage en intelligence artificielle ;
- Exploitation statistiques, visualisation de données.

En big data, ce sont les données qui permettent de trouver des modèles. C'est l'inverse en informatique décisionnelle classique. Et ça c'est tout à fait nouveau : nous sommes ici dans un champ de recherche inédit en informatique avec des efforts de R&D importants.

Les secteurs d'activités qui sont déjà très consommateurs d'informations voient ici de nouvelles opportunités, c'est le cas en météorologie, en médecine, en gestion des risques... avec l'apparition de stratégies d'entreprises pilotées par les données (data-centric).

Cependant, le challenge avec les big data est multiple :

- S'assurer que les données collectées sont pertinentes : le plus n'est pas le mieux et le volume ne sert à rien si les données sont fausses.
- Savoir exploiter efficacement les données : vaut-il mieux des statistiques précises ou des statistiques justes ?
- Gérer les aspects liés à l'éthique, à la sécurité de l'information collectée et à la vie privée : qui peut collecter ? sous quelles conditions ? que deviennent les données ?

bitmap (matrice de bit) : un index *bitmap* recense toutes les valeurs possibles de la clé d'index par une combinaison de bits, ce qui compresse les entrées de l'index.

BLOB (*Binary Large Object*), voir LOB.

blocage : empêchement temporaire d'effectuer une action. Les blocages dans les bases de données sont le résultat de l'application des verrous permettant de protéger des accès concurrents intempestifs.

Boot (page Boot) : page particulière dans les fichiers de données de la base contenant des métadonnées essentielles de celle-ci.

B-Tree, voir arbre équilibré.

C-

cache (ou mémoire cache) : mémoire qui contient temporairement des copies de données provenant d'une autre source de données, afin de diminuer le temps d'accès d'un matériel informatique. Dans SQL Server, le cache est assuré par l'utilisation de la RAM (voir ce mot).

Caractéristique : En conception de base de données avec Merise, une Caractéristique est un synonyme de Propriété. Très simplement, c'est une donnée élémentaire (on ne peut pas la scinder / décomposer / deviner / calculer). Les caractéristiques vont se retrouver un peu partout en base mais toujours bien rangées dans des entités ou au sein d'associations.

cardinalité : nombre de lignes d'une table.

Cardinalités : Lorsque l'on conçoit une base de données avec le MCD de Merise, on obtient un schéma avec des entités et des associations. Pour préciser au mieux les associations, on utilise des cardinalités. Les cardinalités sont des caractères (0,1, n) qui fonctionnent par couple et qui sont présents de chaque côté d'une association (sur chaque « patte »). Ils donnent des indications très intéressantes et permettent par la suite de construire la base de données :

- avec la création de clés étrangères dans le cas d'une CIF
- avec la création d'une table intermédiaire dans le cas d'une CIM

Les cardinalités possibles sont :

- 0,1 : au minimum 0, au maximum 1 seule valeur (CIF) ;
- 1,1 : au minimum 1, au maximum 1 seule valeur (CIF) ;
- 0,n : au minimum 0, au maximum plusieurs valeurs ;
- 1,n : au minimum 1, au maximum plusieurs valeurs.

Les cardinalités maximales (le côté droit ou le dernier caractère) définissent si la relation est une CIM ou une CIF.

Dans l'image, on a « 1,1 » en cardinalité du côté gauche de l'association et une cardinalité « 0,n » du côté droit. Concrètement, il faut lire les cardinalités ainsi :

Une structure se situe dans 1 et 1 seule ville ;

Dans une ville peut se situer 0 ou plusieurs (n) structures.

Les cardinalités sont importantes car elles traduisent les règles de gestion, elles doivent être validées avec l'utilisateur final.

casse (en informatique) : forme majuscule (haut de casse) ou minuscule (bas de casse) des lettres.

catalogue (ou catalogue) : collection de schémas SQL. En réalité, ce terme ancien désignait la base de données. Par extension, il définit aussi les informations de métadonnées des bases.

certificat : document électronique destiné à s'assurer de l'identité d'un utilisateur. Un certificat encapsule une clé de cryptage.

Champ : En base de données, un champ peut signifier plusieurs choses différentes et pourtant proches suivant le degré d'informatisation et d'abstraction :

- À l'étape du MCD (Merise), le champ est la valeur contenue à l'intérieur d'une propriété au sein d'une entité ou d'une relation. Si on exportait le contenu d'une entité dans un tableur, on peut imaginer qu'un champ est l'intersection d'une ligne (ensemble cohérent de données) et d'une colonne (propriété).
- Aux étapes ultérieures (MPD, MLD, base de données en production), le champ est simplement équivalent à la propriété ou l'attribut d'une table.
- Aux étapes de conception d'interface, un champ est une zone dans laquelle l'utilisateur pourra lire ou renseigner une valeur (zone de texte, bouton radio, zone de liste...).

Il y a pas mal de confusions avec ce terme. Dans tous les cas, il est préférable lorsqu'il est évoqué, de bien préciser ce qui est sous-entendu.

CHECKPOINT: commande forçant l'écriture physique des données modifiées de la base.

cherchable : néologisme signifiant qu'un prédicat peut utiliser un index vu que l'accès aux données de l'index s'effectue par une technique dichotomique et non par un balayage de toutes les valeurs.

CHECK : contrainte de ligne de table permettant de valider une ou plusieurs valeurs de colonnes.

chiffrement : cryptage par calcul mathématique.

CIF : Une **CIF** ou **Contrainte d'Intégrité Fonctionnelle** est un type d'association entre 2 entités.

Elle se caractérise par un 1 en cardinalité supérieure (0,1 ou 1,1) sur une des pattes de la relation. On dit alors que la relation est **porteuse d'une dépendance fonctionnelle**. Une CIF indique donc une dépendance. Une des entités de l'association est déterminée par la connaissance d'une (ou plusieurs) autre entité présente dans l'association. Cette association est forte et hiérarchique. Sans entité parent, il ne peut pas y avoir d'entité enfant. À ne pas confondre avec une CIM.

Une CIF se transforme ensuite (lorsque l'on passe au MPD / MLD) en une clé étrangère. Cette clé étrangère est un champ ajouté à la table située du côté du 1 en cardinalité maximale qui reprend le champ clé primaire de la table située de l'autre côté de la relation.

Clé primaire : Voilà un terme qui revient très souvent et qui est souvent confondu avec le terme identifiant qui est très proche.

Lorsque l'on conçoit une base de données, on utilise souvent une méthode afin de ne rien oublier et de structurer son travail. Merise en est une. Dans la méthode Merise, le concepteur réalise d'abord un MCD qu'il va transformer en MPD et en MLD.

À l'étape du MCD, on a des identifiants dans les entités. À l'étape du MPD et MLD on a des **clés primaires** dans les tables. Identifiants et clés primaires sont la même chose mais à des stades différents de la conception d'une base de données.

CIM : Une **CIM** ou **Contrainte d'Intégrité Multiple** est un type d'association entre 2 entités minimum.

Sur un MCD, elle se caractérise par l'absence de 1 en cardinalité maximale (sur aucune « patte » de la relation). Par déduction, on ne peut avoir que 0,n ou 1,n en cardinalité. À ne pas confondre avec la CIF.

Concrètement, une CIM représente un couple unique auquel on peut affecter des propriétés particulières. On dit alors que la CIM est porteuse de propriétés (ou de données). Non hiérarchique, la CIM est construite à partir des identifiants des entités qui lui sont liées.

clause : partie syntaxiquement distincte d'un ordre SQL précisant un fonctionnement particulier (membre d'une commande SQL). Par exemple : clause ORDER BY de l'ordre SELECT.

clé : ensemble de colonnes dont les valeurs permettent de retrouver au plus une seule occurrence des informations d'une table. Dans cette dernière, elle assure l'unicité d'accès à la ligne. Une clé peut être primaire ou alternative.

clé alternative (ou subrogée) : clé pouvant se substituer à l'usage à une clé primaire d'une table. Celle-ci peut contenir plusieurs clés alternatives. Elle se traduit dans la table par la contrainte UNIQUE.

clé candidate : clé pouvant figurer en tant que clé primaire ou alternative.

clé d'index : par extension du concept de clé, ensemble de colonnes qui assurent la recherche dans un index, pas forcément unique.

clé de cryptage (ou de chiffrement) : donnée binaire permettant par le biais d'un algorithme de chiffrement de crypter ou décrypter de l'information en toute confidentialité (chaque clé étant en principe différente).

clé de hachage : donnée sur laquelle est calculée une valeur de hachage.

clé étrangère : Une **clé étrangère** en bases de données est un mécanisme simple qui permet de garantir l'intégrité référentielle entre les données de différentes tables. Concrètement, la clé étrangère oblige une table à être liée aux données d'une autre table. Ainsi dans l'exemple suivant, une structure est obligatoirement liée à une ville par la clé de la ville.

Ces 2 clés sont la traduction concrète de la relation entre les tables « structure » et « ville ».

Par convention, les clés étrangères sont représentées préfixées du symbole dièse #.

Dans une table (au niveau du MPD et du MLD), une clé étrangère est une clé particulière qui n'est pas la clé principale de la table. La clé principale est appelée clé primaire. Lorsqu'une ou plusieurs autres clés apparaissent dans une table, ce sont donc des clés étrangères. À la différence des clés primaires, on peut retrouver plusieurs fois la même valeur dans les clés étrangères d'une table (et c'est logique, dans l'exemple ci-dessus, il est normal que la même ville puisse être utilisée par plusieurs structures).

Les clés étrangères sont la conséquence directe de l'application de règles qui permettent le passage du MCD au MPD / MLD. Les clés sont les identifiants du MCD. Elles deviennent clés primaires ou clés étrangères suivant les cas de figures :

Dans une CIF (une relation avec une cardinalité supérieure à 1), on obtient une clé étrangère. L'identifiant de l'entité qui n'est pas du côté du 1 est ajouté en tant que clé étrangère à la table du côté du 1. Au passage, cette clé récupère le #.

Dans une CIM (une relation sans cardinalité supérieure à 1), il n'y a pas de clé étrangère mais une seule clé primaire composée des identifiants de chaque table liée.

clé primaire : clé composée d'une ou plusieurs colonnes dont les valeurs sont toujours renseignées (non NULL). Une table ne peut avoir qu'une seule colonne clé primaire. Elle se traduit dans la table par la contrainte PRIMARY KEY.

clé subrogée : synonyme de clé alternative, voir ce mot.

cliché (*snapshot*) : copie d'un ensemble de données.

cliché de base de données (database snapshot) : captation des données de l'intégralité de la base à un instant donné.

client/serveur (ou C/S) : mode d'interaction en réseau permettant à deux processus, appelés client et serveur, de communiquer entre eux. Dans les SGBDR de type C/S, le client génère les requêtes et les envoie au serveur. Ce dernier exécute les requêtes et renvoie la réponse au client. Le C/S est généralement opposé à la méthode d'accès par partage de fichiers.

CLOB, voir LOB.

cloud computing : solution informatique hébergée proposant des services dont l'emplacement physique est indéfini et « mouvant », afin de rendre le système très hautement disponible (redondance occulte).

cluster (littéralement, emplacement) : pour une instance de serveur SQL, un cluster est un ensemble de machines redondées destinées à assurer un seul et même service par le relais d'une machine de secours en cas de défaillance de celle principale. Les applications clientes ne voient qu'un seul serveur de bases de données.

clustered : pour un index, cette notion indique qu'il est créé « sur place », c'est-à-dire sur la table elle-même. Les index dits « secondaires » (non clustered) sont, quant à eux, des copies d'une partie des données de la table.

clustering : technique de haute disponibilité consistant à générer un serveur « fantôme » qui représente une grappe de serveurs redondants de manière à ce que l'un des participants (appelé « noeud ») puisse être remplacé en cas de défaillance. Dans SQL Server, le clustering peut se faire avec de nombreux noeuds, mais il nécessite de partager les bases de données via un SAN.

coeur (*core*) : sous-ensemble d'un microprocesseur moderne, constitué d'une UAL (Unité arithmétique et logique) indépendante. Il est présent en plusieurs exemplaires dans le même processeur et destiné à augmenter la puissance de calcul global en permettant le parallélisme.

cohérence : ensemble des liaisons logiques des différents éléments d'une base de données n'offrant aucune contradiction.

collation : littéralement, c'est la comparaison de deux informations afin de vérifier si elles sont identiques. Dans SQL Server, il s'agit d'objets définissant les propriétés de sensibilité à la casse (majuscule/minuscule), aux accents (et autres caractères diacritiques), de kanatypes (japonais) ou encore de la largeur de trait des caractères.

colonne : élément vertical dans une table, présentant un ensemble de valeurs de même type.

columnstore (index) : technique d'indexation de différentes colonnes d'une même table basée sur une structure dite « verticale », rendant uniforme l'accès à l'une des colonnes cherchées.

COMMIT: commande de validation de transaction.

Common Table Expression (CTE) : forme particulière d'écriture de sous-requêtes permettant de définir un ensemble de pseudo-tables utilisées par une requête finale et introduite par le mot-clé WITH.

compression : technique consistant à organiser différemment les données pour en diminuer le volume physique tout en conservant l'usage logique. SQL Server dispose de nombreuses possibilités de compression, pour certaines colonnes (vardecimal, sparse), comme pour les lignes, pages et fichiers de sauvegarde.

concaténation : opération consistant à mettre bout à bout des chaînes de caractères afin d'en former une plus complète.

concurrence : Dans l'univers des **bases de données**, la notion de **concurrence** a une signification particulière bien différente du mot concurrent que nous connaissons tous et qui s'applique à l'économie.

En base de données, la concurrence signifie « **accès simultanés** ». Il est très fréquent que plusieurs personnes (ou programmes) souhaitent accéder « en même temps » aux mêmes données. C'est le cas lorsque l'on consulte un site web. C'est aussi le cas lorsque plusieurs utilisateurs souhaitent accéder à la même fiche client dans un logiciel de gestion commerciale.

Le cas est plus problématique lorsque plusieurs utilisateurs souhaitent modifier « en même temps » les mêmes données ou qu'une donnée consultée par une personne est parallèlement modifiée par une autre.

Pour gérer ces situations problématiques d'accès concurrentiels aux données, les bases de données utilise des règles de gestion et d'intégrité des données nommées ACID. Il en découle des traitements par lots (les transactions) et la mise en place de mécanismes d'isolation (pose de verrous pour interdire l'accès à certaines données pendant la durée d'un traitement). Les verrous posent aussi d'autres problèmes et il faut parfois décaler dans le temps les opérations conflictuelles. Ce qui entraîne encore d'autres situations problématiques (données non à jour, doublons, synchronisation...).

Derrière le terme de concurrence d'accès aux données, il y a aussi un enjeu de performance : en fonction du volume de données à traiter, tous les SGBD ne se valent pas. Certains peuvent absorber des dizaines de demandes simultanées alors que d'autres sont capables d'en traiter beaucoup plus. Enfin, l'aspect matériel et l'architecture technique déployée devient souvent un goulot d'étranglement lorsque les problématiques de volume arrivent.

connexion : phase durant laquelle un processus encore inconnu tente de pénétrer dans le serveur SQL en essayant de s'authentifier.

conteneur : objet destiné à en contenir d'autres. Les principaux conteneurs dans SQL Server sont : l'instance qui contient les bases, la base qui contient des schémas SQL et le schéma SQL qui contient des objets relationnels.

contention : phénomène généralement transitoire par lequel, du fait de blocage, les temps de réponse moyens d'un système commence à s'allonger de façon problématique. Les embouteillages sur la route sont un exemple de contention.

courbe de Lebesgue (ou courbe en Z) : courbe cheminant par un tracé en forme de Z et destinée à assurer le remplissage d'un espace en parcourant tous les éléments qui le pavent.

contrainte : objet d'une base ayant pour objectif de programmer des règles de gestion au niveau des données. Les contraintes agissent soit par barrage, soit par mécanisme de modification automatique. Il existe trois niveaux de contraintes : celle de domaine qui porte sur des valeurs (réalisée par des RULE avec SQL Server), celle de table qui porte sur les données (clé primaire, alternative, étrangère, validation et obligation de valeur : PRIMARY KEY, UNIQUE, FOREIGN KEY, CHECK, NOT NULL) et celle générale (dite « d'assertion ») qui porte sur de multiples tables et qui est réalisée généralement par des déclencheurs sinon par des contraintes CHECK appelant des UDF.

contrainte d'intégrité : Toutes les contraintes que doivent respecter les données sont appelées contraintes d'intégrité référentiel, ou plus simplement contrainte d'intégrité.

couvrant : se dit d'un index dont la composition (listes des colonnes composant la clé d'index et colonnes incluses) suffit à répondre pleinement aux données d'une table pour la requête considérée. Un index couvrant évite la double lecture : recherche dans l'index plus lecture d'informations complémentaires dans la table.

CPU (*Central Processing Unit*, unité centrale de traitement) : composant essentiel de l'ordinateur qui exécute les instructions machine. Les CPU modernes possèdent plusieurs coeurs.

CTE, voir Common Table Expression.

CRUD : Abréviation en anglais de **Create, Read, Update, Delete**. En français : Écrire, Lire, Mettre à jour, Supprimer.

Il s'agit des opérations de base attendues par un SGBD (mais pas uniquement, on s'en sert aussi dans d'autres types de logiciels). Le terme CRUD date de 1983, c'est James Martin qui a popularisé l'expression dans son livre *Managing the Database Environment*.

On rencontre parfois le terme SCRUD ou le S initial signifie Search (Rechercher) ou encore SCRUDL avec le L final signifiant Large pour les bases de données volumineuses. Les abréviations DRULAB (Delete, Read, Update, Lock, Add, Browse) et BREAD (Browse, Read, Edit, Add, Delete) sont très similaires.

CU (*Cumulative Update*) : ensemble de mises à jour, corrections, améliorations cumulatif et testé de correctifs logiciels (voir aussi *Service Pack*) visant généralement une thématique spécifique (sécurité,

composant, fonctionnalité d'un produit...) et, par conséquent, destiné à résoudre un problème particulier.

cryptage (ou crypter) : action consistant à transformer une information naturellement interprétable (claire) en une autre impossible à déchiffrer (opaque) pour celui qui ne connaît pas la méthode employée ou bien ne possède pas les paramètres de celle-ci.

cylindre (disque) : ensemble des pistes des différents plateaux d'un disque dur informatique ne nécessitant aucun déplacement de la tête de lecture pour en lire les données (celles-ci étant lues par la simple rotation synchrone des plateaux).

D-

Database Snapshot, voir cliché de base de données.

DATALINK : type particulier de données que la norme SQL a prévu pour le stockage des fichiers sous le contrôle du SGBDR. Il est représenté dans SQL Server par FILESTREAM.

Data mining : Le data mining ou exploration de données en français est l'ensemble des techniques qui permet d'extraire des informations au sein de données et plus précisément de découvrir des structures (modèles et comportements) au sein de vastes ensemble de données. Le data-mining utilise les apports des statistiques et de l'intelligence artificielle pour isoler l'information valorisable et la présenter. Les informations extraites doivent être :

- justes, précises, fiables ;
- sourcées (on doit pouvoir savoir d'où elles viennent précisément) ;
- intéressantes voire nouvelles.

En big-data et informatique décisionnelle, extraire de l'information cachée au sein de masses d'informations est fondamental. Le data mining est ainsi cette tâche clé qui consiste à fouiller, prospecter, forer dans des bases de données afin d'**en faire ressortir une connaissance** (information qui a de la valeur et qui est exploitable).

Cette idée d'exploiter l'information déjà présente (analyse secondaire – on ne cherche pas à collecter l'information mais simplement à l'analyser) est vieille de plusieurs siècles et très tôt en informatique (dès les années 60 avec les statistiques exploratoires), des outils ont été produits dans ce but. Les données textuelles sont les plus simples à traiter mais les éléments multimédias (sons, images) sont aujourd'hui bien supportés par les outils disponibles.

Les outils logiciels les plus utilisés pour le data mining sont :

- Knime, R, Weka, Orange, RapidMiner pour les logiciels libres ;
- KXEN, Excel et SQL Server (avec les modules supplémentaires Business intelligence de Microsoft), Oracle DM, SAS Entreprise Miner, STATISTICA, IBM SPSS, BrainCube, CORICO pour les logiciels commerciaux.

On peut aujourd'hui retenir trois méthodes majeures pour fouiller les données :

- Méthode **SEMMA** : créée par l'éditeur de logiciels SAS Institute, cette méthode s'applique à tous les logiciels (même ceux qui ne sont pas édités par l'entreprise SAS Institute). Elle décrit un ensemble d'outils et de méthodes logiques à mettre en place afin de fouiller au mieux une grande masse d'informations. C'est une approche technique.
- Méthode **CRISP-DM** : il s'agit de 6 étapes qui permettent d'industrialiser la fouille de données (compréhension métier, recherche de données, préparation des données, exploration et modélisation de l'information, segmentation (cœur de la méthode), évaluation et livraison/restitution).
- Méthode **DMAIC Six Sigma** : empruntée à l'industrie, la méthode met l'accent sur la recherche de défauts et l'amélioration de la qualité. Cela s'applique très bien à la fouille de données.

Allant plus loin que les simples calculs statistiques (intelligence artificielle, réseaux de neurones, calculs bayesiens, machine learning), l'exploration de données se base sur deux grandes familles d'algorithmes :

- **Méthodes descriptives** : cette façon de faire consiste à analyser les données disponibles et à les classer sans a-priori de façon à mettre en évidence des normes de comportement et à repérer des aberrations.
- **Méthodes prédictives** : la tâche consiste ici à regarder les données actuelles et à prévoir le futur. En se basant sur les éléments connus, on peut imaginer une certaine reproductibilité.

Les écueils les plus fréquents à l'usage du data-mining sont essentiellement liés à l'humain :

- Poser les bonnes questions ;
- Savoir écarter les données fausses, pouvoir constituer un ensemble de données qui a du sens et qui est vrai ;
- Faire la différence entre corrélation et causalité ;
- Ne pas confondre justesse et précision ;
- Savoir exploiter les résultats et savoir les réfuter parfois ;
- Résoudre les problèmes d'éthique et les respecter les données privées.

Déclencheur : Les **déclencheurs** ou **triggers** (en anglais) sont une fonctionnalité avancée très pratique des bases de données. Ils permettent de **déclencher une action lorsque certains paramètres sont réunis** dans une table.

Un déclencheur peut se produire lorsque :

- des données sont insérées (INSERT) ;
- des données sont modifiées (UPDATE) ;
- des données sont supprimées (DELETE) ;
- Juste avant (BEFORE) ou juste après (AFTER) l'action sur les données.

On se sert des triggers pour :

- **mettre à jour** d'autres données en déplaçant une partie de la logique applicative dans la base de données plutôt que dans l'outil utilisé (site web) ;
- **vérifier** les données à rentrer dans la base (contrôle avant ou après la requête) ;
- **remplir** un fichier de log afin de savoir qui a fait quoi ;
- **pallier à un défaut** d'intégrité référentielle (sur certaines bases de données qui ne le gèrent pas ou lorsque le concepteur ne l'a pas mis en place)

Spécificités des déclencheurs

- Le déclencheur s'exécute pour toutes les lignes. Si 10 lignes doivent être affectées par un traitement, le déclencheur se lancera 10 fois.
- Il ne peut y avoir qu'un seul déclencheur par action et moment. On ne peut pas empiler les triggers. On peut par contre créer des instructions qui se suivent dans le même trigger pour réaliser des opérations multiples (Avant insertion / Après modification...).
- Les déclencheurs ne peuvent pas lancer toutes les sortes d'actions (transaction, sélection de données). Les déclencheurs ne peuvent pas non plus agir sur la table qui l'a déclenché.
- Les triggers peuvent avoir des erreurs. En cas d'erreur et lorsque la base de données le permet, les actions sont déconstruites et la requête n'est pas réalisée (principe du rollback).

DBA (DataBase Administrator) : administrateur de bases de données.

DBCC (*Database Console Command*) : outil interne permettant de piloter le moteur de stockage de SQL Server.

DCL (*Data Control Language*) : partie du SQL qui traite de la sécurité par mise en place de privilèges sur des objets, accordés aux utilisateurs. Ordre Transact-SQL GRANT, DENY et REVOKE.

DCM(*Differential Changed Map*) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer par un bit à 1 parmi les 63 904 extensions qui la suivent, lesquelles ont été mises à jour depuis la dernière sauvegarde complète.

DDL (*Data Definition Language*) : partie du SQL qui traite de la structuration de la base et permet donc de modifier le schéma de celle-ci. Comprend principalement des ordres SQL CREATE, ALTER et DROP.

DEFAULT: contrainte de colonne de table indiquant qu'une valeur sera donnée à cette colonne en cas d'insertion de ligne ignorant la colonne.

déclencheur (*trigger*) : routine Transact-SQL attachée à une table ou une vue (déclencheur DML), ou encore une base ou un serveur (déclencheur DDL), exécutée lors de la survenue d'un événement (INSERT, UPDATE, DELETE pour les déclencheurs DML et CREATE, ALTER, DROP, etc., pour les déclencheurs DDL).

déférable : possibilité de dévolution d'une action. Une contrainte déférable possède une évaluation qui peut être reportée en fin de transaction. On parle aussi de « déférabilité ».

degré : nombre de colonnes d'une table.

déni (de privilège) : interdiction de l'octroi d'un privilège.

densité : indice inverse de la sélectivité (voir ce terme).

Dépendance fonctionnelle : Toute ligne d'une table est identifiable à la valeur de sa clef primaire. La valeur des autres champs de la ligne est en quelque sorte rattachée à la valeur de la clef primaire. On dit que ces autres champs sont en dépendance fonctionnelle avec la clef primaire, c'est-à-dire qu'ils lui sont rattachés.

dépersonnalisation : état permettant à un utilisateur de se faire temporairement passer pour un autre.

diacritique : signe typographique (accent, cédille...) ou forme (ligature...) affecté à une lettre afin de lui conférer une graphie ou une phonie différente. Par extension, on parle de « caractère diacritique » lorsqu'une lettre est, par exemple, accentuée.

diadique : caractérise une fonction ou un opérateur prenant en compte deux éléments en entrée.

Dictionnaire des données : Le dictionnaire des données est une représentation tabulaire de la structure d'une base de données. Il consiste à énumérer tous les champs de toutes les tables d'une base de données. Il prend par exemple la forme suivante :

Champ	Type	Longueur	Vide ?	Par défaut	Description
uneChaine	Chaîne	10	Oui		Chaîne de car. A 10 car. maximum
unEntier	Entier	1	non	0	Entier compris entre -127 et 128
UnDecimal	décimal	1,2	non	0.00	Décimal à 5 chiffres, 2 après la virgule
...

Champ : nom du champ, à savoir dénomination de la donnée à stocker

Type : type de donnée du champ (chaîne de caractères, entier, décimal, date, heure, etc) ;

Longueur : pour une chaîne de caractères, longueur maximale de la chaîne ; pour entier, en général, le nombre d'octets sur lequel est codé l'entier ; pour un nombre décimal, le nombre de chiffres total suivi du nombre de chiffres après la virgule. ;

Vide : Indique si le champ peut être vide, c'est-à-dire avoir la valeur null ;

Par défaut : valeur que le champ prend par défaut ;

Description : précise la signification du champ à stocker.

différence : opération relationnelle consistant à rechercher les lignes qui n'existent pas d'une table par rapport à une autre. Avec l'opérateur EXCEPT.

DML (Data Manipulation Language) : partie du SQL qui traite de la manipulation des données, en lecture (SELECT) comme en écriture (INSERT, UPDATE, DELETE et MERGE).

DMV (Data Management View, vue de gestion des données) : ensemble de vues système (présentes dans le schéma sys) destinées à fournir des informations sur l'état de l'instance SQL à différents niveaux (serveur, bases, objets...). Ces données ne figurant qu'en mémoire, elles sont perdues en cas d'arrêt du serveur.

domaine : au niveau conceptuel, ensemble des valeurs possibles d'un attribut. Au niveau SQL : objet d'une base de données permettant de définir un type de données associable à des règles afin de représenter un domaine conceptuel. Les domaines SQL Server sont réalisés par le biais des TYPE.

données géodésiques : données de cartographie prenant en compte la courbure spatiale de l'écorce terrestre, permettant la mesure et la représentation de la surface de la terre.

données relationnelle : Relationnelle se dit d'une base de données qui manipule des informations hiérarchisées qui sont liées entre elle par des contraintes de relations. À la différence d'autres types de bases de données, les bases de données relationnelles fonctionnent avec des tableaux à deux dimensions nommés tables qui sont liés à d'autres tables par un mécanisme de relations reposant sur l'algèbre relationnelle. Ces relations définissent et précisent les liens entre les tables.

doublon : ligne d'une table contenant des données strictement identiques à une autre.

DTD (Document Type Definition, définition de type de document) : document décrivant un modèle permettant de contrôler le contenu de documents XML notamment. Actuellement en voie de disparition du fait de la prééminence de XML Schema pour assurer cette même fonction.

durabilité : propriété du SGBDR qui assure qu'une transaction informatique validée survit de façon permanente. Elle se traduit généralement par l'écriture des données sur un support physique et durable.

dump (littéralement, « vidage ») : copie de la mémoire vive et des registres d'un processeur, permettant d'avoir un instantané de l'état d'un système. Terme trop souvent confondu avec la notion de sauvegarde. Pour les bases de données, un dump est un fichier contenant une extraction des commandes SQL permettant de reconstituer la base.

E-

EBCDIC (*Extended Binary Coded Decimal Interchange Code*) : mode de codage des caractères sur 8 bits créé par IBM à l'époque des cartes perforées. EBCDIC existe en six versions différentes incompatibles entre elles, et certains caractères de ponctuation ne sont pas disponibles dans toutes les versions. EBCDIC est néanmoins encore utilisé dans les systèmes AS/400 d'IBM ainsi que sur les mainframes sous MVS, VM ou DOS/VSE. Cet encodage de caractères étant très décrié du fait de sa complexité et de son aspect propriétaire, il a tendance à disparaître au profit d'encodages plus standards comme l'ASCII, Unicode ou encore UTF-16.

équijointure : opération de jointure dont le prédicat utilise exclusivement des opérateurs d'égalité entre des colonnes de table.

ensembliste : manipulation d'ensembles de données utilisant le schéma relationnel.

Entité : Dans la méthode Merise au niveau du MCD, une entité correspond à un **ensemble homogène d'informations** qui correspondent au même « objet » à informatiser. Cette **entité** a un nom unique afin de la manipuler facilement. Plus tard dans l'analyse, l'entité se transforme en table et devient concrètement une table lors de la réalisation effective de la base de données.

Cet ensemble d'informations, l'**entité**, partage les mêmes caractéristiques et peut être manipulé au sein du système d'information mais aussi en discutant entre informaticiens et personnes du métier. L'entité « *contrat* » et l'entité « *client* » sont typiques des bases de données de gestion. Isoler, comprendre le rôle et les informations qui se trouvent dans ces coquilles est une des premières étapes de l'analyse.

Une **entité** reste cependant peu utile si elle n'est pas remplie avec des informations calibrées et si elle ne discute pas avec les autres entités de la base de données : on parle alors de propriétés et d'associations ce qui a donné le nom de modèle entités-associations que l'on croise parfois.

Dans un MCD, les entités sont représentées par des rectangles et dans le cartouche supérieur de chaque rectangle est écrit au singulier le nom de l'entité.

Entrepôts de données : Plus souvent rencontré dans sa version anglais data **warehouse**, les **entrepôts de données** sont des bases de données qui agrègent l'information éparpillée au sein du système d'information, les rend compatibles et exploitables dans le cadre de l'informatique décisionnelle.

Ainsi une entreprise peut avoir des données qui proviennent de tableur Excel, d'une base de données Access pour les commerciaux, de la base de données marketing au format MySQL ou encore de catalogues produits venant directement des fabricants. L'entrepôt de données a pour tâche de fournir une vision homogène et cohérente afin de pouvoir croiser les données et ressortir des informations à plus haute valeur ajoutée.

Les entrepôts de données ont aussi un rôle de mémoire. Ils doivent permettre d'historiser les données et de continuer à les interroger même si les données initiales viennent à disparaître ou être transformées. Les données rentrées dans l'entrepôt de données ne sont ainsi plus modifiables.

- Un entrepôt de données réalise donc plusieurs tâches :
- Extraire les données de production

- Transformer, épurer, homogénéiser les données
- Stocker les données
- Permettre l'interrogation

ES (ou E/S : Entrée/Sortie, en anglais IO : Input/Output) : opération de lecture ou d'écriture d'une page.

espace de nom (*namespace*) : lieu abstrait destiné à accueillir des ensembles de termes appartenant à une même famille. Dans les langages informatiques, ce sont souvent des fichiers particuliers, regroupant des définitions d'objets et de méthodes, partagés et utilisés par d'autres programmes. Anciennement « bibliothèque » (*library*).

exception : condition particulière, généralement non prévue, pouvant survenir du fait d'une défaillance matérielle, d'un bogue logiciel, d'une contrainte violée ou d'une condition la générant.

expansion : mots ou expressions dérivés d'un autre mot ou d'un acronyme (utilisé dans l'indexation textuelle), par exemple Société nationale des chemins de fer au lieu de SNCF.

expression rationnelle : modèle créé via les caractères ASCII, permettant de manipuler des chaînes de caractères à l'aide de motifs (notamment, extraction de portions de la chaîne correspondant au modèle). Ce système est emprunté au système POSIX (système d'exploitation). De nombreux scripts sous Unix l'utilisent.

extension (*extent*) : ensemble consécutif de 8 pages physiques de données (table ou index) figurant sur le disque.

F-

famine : situation dans laquelle un système informatique est engagé dans un traitement où une phase critique ne peut être finalisée avec certitude.

file group, voir groupe de fichiers.

Filestream : technique destinée à stocker des fichiers de toutes natures, sous le contrôle exclusif ou partagé du serveur SQL en faisant croire que les données sont présentes dans une table.

Filetable : type de table particulier permettant de représenter les fichiers et répertoires d'un point d'entrée du système de fichier et d'y effectuer des opérations de fichier (création, modification, suppression...) sous le contrôle exclusif ou partagé du serveur SQL. C'est une extension du concept de Filestream.

flou : élément dont le dessin est incertain. L'indexation textuelle permet des recherches floues.

floutage : rendre flou tout ou partie d'un élément. Dans l'indexation textuelle, permet de réaliser des recherches par approximation.

fonction : routine de code générant une information de sortie constituée d'une seule valeur atomique (scalaire) ou d'un ensemble de données (table). Une fonction admet généralement des paramètres en entrée.

fonction scalaire : fonction renvoyant une seule valeur atomique.

fonction table : fonction renvoyant un ensemble de données. Elle peut être constituée d'une seule instruction SELECT contenant d'éventuels paramètres (fonction table en ligne) ou bien d'une table de sortie dont les données sont alimentées par le biais de commandes INSERT, UPDATE, DELETE (fonction table multi-instruction).

forme fléchie : ensemble des mots dérivés de la racine d'un mot pour ses différents modes (pluriel, singulier), genres (masculin, féminin, neutre) ou temps (conjugaison).

FOREIGN KEY : contrainte de clé étrangère applicable à une ligne de table. Oblige à définir les valeurs de référence par lien à une autre table.

fragmentation (d'index) : présence de « trous », c'est-à-dire d'espaces inaffectés dans les structures d'index. La fragmentation résulte des opérations de mise à jour de données (INSERT, UPDATE et DELETE) affectant les index.

Full Text Search, voir *plain text*.

G-

GAM (*Global Allocation Map*) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer par un bit à 1 parmi les 63 904 extensions lesquelles sont vides de toutes données utilisables.

Géodésique, voir données géodésiques.

granularité (de l'information) : ensemble plus ou moins grand de données. Ligne, ensemble de lignes, page, extensions, tables, base et serveur sont les différentes granularités au sein d'une instance SQL Server.

graphe : ensemble de points connexes (« puits ») reliés les uns aux autres par des chemins (« noeuds »), formant un réseau orienté ou non.

groupe de fichiers (*file group*) : emplacement logique de stockage des données des tables et index. Correspond aux notions de *tablespace* d'Oracle. Un groupe de fichiers peut contenir plusieurs fichiers. Une table ou un index peut être placé sur un groupe de fichiers particulier. En cas de partitionnement des données (tables et/ou index), il faut créer autant de groupes de fichiers qu'il y a de partitions.

Groupement : existe une clause bien pratique en SQL qui permet d'opérer des regroupements. Il s'agit de GROUP BY. On s'en sert pour grouper des éléments identiques.

Table Ventes

identifiant	Date	Nom Client	Montant Vente
1	2017-01-15	ImmoPlus	1600
2	2017-01-15	CanalImmo	800
3	2017-01-15	RégieRE	750
4	2017-01-16	RégieRE	2050
5	2017-01-16	RégieRE	380
6	2017-01-17	ImmoPlus	520
7	2017-01-18	CanalImmo	990
8	2017-01-18	CanalImmo	1050

Par exemple, pour un même client dont on connaît le détail de chaque vente, on peut réaliser un groupement par client et effectuer un calcul statistique sur les ventes (somme des ventes, moyenne des ventes, nombre cumulé de ventes, vente maximum ou minimum)...

Concrètement, la requête va **regrouper les éléments identiques** et **réaliser une fonction d'agrégation sur les autres champs** choisis.

En réalisant la requête suivante, on affiche la somme des ventes par client :

```
SELECT "Nom Client", SUM("Montant Vente") FROM "Table Ventes"
```

```
GROUP BY ("Nom Client");
```

En réalisant la requête suivante, on affiche la moyenne des ventes par jour :

```
SELECT "Date", AVG("Montant Vente") FROM "Table Ventas"  
  
GROUP BY ("Date");
```

Très pratique, le GROUP BY fait partie des outils de base lorsque l'on interroge les bases de données SQL et que l'on souhaite faire parler les données. Il évite d'afficher des doublons et permet d'avoir un premier niveau de données statistiques.

GUID (*Globally Unique Identifier*) : suite de 16 octets de valeurs arbitraire, générée automatiquement à partir de l'identifiant unique d'un périphérique (adresse MAC d'une carte réseau), de la date et de l'heure en cours et d'une séquence numérique, et dont la valeur s'avère en principe unique dans tout l'univers. En réalité, la collision de deux GUID, c'est-à-dire la possibilité d'obtenir un GUID identique en provenance de deux sources différentes, n'est pas impossible mais très peu probable, même si les adresses MAC ont été recyclées du fait de l'épuisement des valeurs.

H-

hachage : mécanisme de conversion d'une donnée d'un type quelconque en une donnée condensée, (entier ou chaîne binaire), par le biais d'une fonction.

Hadoop : outil open source basé sur le langage Java destiné à faciliter la création d'applications distribuées et « échelonnables » (scalables), notamment par le biais d'un algorithme de type MapReduce pour un accès à des données stockées sur une grille de serveur et destiné principalement au Big Data.

haute disponibilité : ensemble de techniques permettant de continuer l'activité du serveur même en cas de défaillance sévère. SQL Server intègre des mécanismes natifs de haute disponibilité comme le log shipping, le clustering, le mirroring ou la technologie AlwaysOn (combinant clustering et mirroring). Certains de ces mécanismes peuvent pallier une perte de données totale d'un des serveurs, sans discontinuité de fonctionnement ni perte de données, et sans que les applications aient à se reconnecter.

hash, voir hachage.

hash join, voir jointure par hachage.

hash aggregate, voir agrégation par hachage.

HTTP (*HyperText Transfer Protocol*) : protocole de communication client-serveur développé pour le *World Wide Web* et basé sur des requêtes littérales envoyées sur des contenus web, dont l'adresse est formée d'une URL.

HTTP endpoint, voir point de terminaison

HSM (*Hardware Security Module*) : dispositif physique (boîtier électronique) réputé inviolable, placé sur le réseau informatique et stockant les clés et algorithmes de cryptage.

hyperthreading : implémentation dans les processeurs Intel du *Simultaneous Multithreading* à deux voies qui consiste à simuler dans un seul coeur deux processeurs logiques, par une technique de partage des ressources, en alternance pour certaines, en redondance pour d'autres.

I-

IAM(*Index Allocation Map*) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer, pour un objet particulier stockant des données (table, index, partition...) dans les 4 Go suivant du fichier, que les extensions utilisées.

identifiant : terme pouvant désigner une colonne servant de clé dans une table, comme un nom d'objet (identifiant SQL).

identifiant SQL : nom d'un objet SQL (table, vue, colonne, contrainte...).

IDENTITY : propriété d'auto-incrémentation d'une colonne d'une table. Ne peut être affectée qu'à une seule colonne au plus par table. À défaut, commence à 1 avec un pas d'incrément de 1. Peut être paramétrée au niveau graine (valeur initiale) et pas. Peut être réinitialisée à tout moment.

impersonnalisation, voir dépersonnalisation.

In Memory : technologie récente destinée à supporter des tables figurant essentiellement en mémoire vive par opposition aux tables habituellement destinées à du stockage physique. Ceci inclut en sus, l'utilisation de procédures stockées compilées en code natif (C).

inclus (index) : se dit lorsque les composantes d'un index (clé plus éventuelle clause INCLUDE) sont incluses dans un autre index.

index : structure de stockage redondante, dotée d'une organisation particulière dont l'agencement permet d'accélérer certaines recherches. Les techniques en jeu dans l'indexation sont toutes basées sur la famille d'algorithmes « diviser pour régner ».

indexation : recherche et mise en place des index (voir ce mot) les plus pertinents en vue de diminuer sensiblement les temps de réponse. L'indexation vise à réduire drastiquement le coût des recherches dans une base de données, dès lors que la volumétrie des données pèse sur les temps de réponse.

indexation textuelle : mécanisme d'indexation de littéraux permettant des recherches *plain text*.

instance : installation d'un serveur (SQL Server) logique. Un même serveur système (Windows) peut accueillir différentes instances (jusqu'à 50) dans différentes éditions/versions. Chaque instance peut avoir des réglages différents, mais chacune entre en concurrence pour les ressources physiques du serveur (CPU, RAM, disque, réseau...).

intersection : opération relationnelle consistant à rechercher les tuples communs à deux relations. Se fait avec l'opérateur ensembliste INTERSECT dans le langage SQL.

intégrité : l'intégrité des données au sein d'une base procède par le respect des contraintes.

Intégrité référentielle : 1. Ce mot fait parfois peur aux débutants en base de données et il est fondamental en base de données. Lorsque j'étais étudiant, c'était un incontournable (j'ai eu cette question au BAC notamment).

L'intégrité référentielle est un mécanisme qui protège le contenu et les relations d'une base de données. Correctement construite, une base de données fait appel à des données situées dans

différents emplacements (les tables). Pour que les données restent utilisables et cohérentes, il ne faut pas que l'on puisse détruire des données qui dépendent les unes des autres. C'est le rôle de l'intégrité référentielle de protéger ces relations.

Dit autrement, l'intégrité référentielle est une contrainte que le concepteur de bases de données s'impose afin de garantir l'intégrité des données stockées.

Techniquement dans une base de données relationnelle, l'intégrité référentielle vérifie qu'une valeur de clé étrangère existe bien en tant que valeur de clé primaire dans une autre table. Quand une clé étrangère est utilisée, elle doit obligatoirement exister en tant que clé primaire dans la table liée.

2. mécanisme permettant d'assurer dans une table fille que les valeurs d'une clé étrangère soient toujours synchronisées avec celle de la table mère dite de référence. L'intégrité référentielle peut être assurée de manière déclarative (DRI par contrainte FOREIGN KEY) ou fonctionnelle par le biais de déclencheurs.

IO (I/O, *Input/Output*), voir ES.

ISO (*International Standardization Organization*) : organisme international de normalisation, composé de représentants nationaux. L'ISO produit des normes internationales dans les domaines industriels et commerciaux. Le langage SQL est fortement normalisé. Une norme ne devient ISO, qu'après avoir été norme nationale (le langage SQL a été normalisé pour la première fois par l'ANSI en 1986).

isolation : mécanisme qui prévoit que les modifications de données effectuées au cours d'une transaction ne soient pas atteignables par d'autres utilisateurs. L'isolation est assurée par la pose de verrous plus ou moins forts ou étendus en fonction de force d'isolement demandée.

isolation snapshot : mode d'isolation des transactions agissant par verrouillage optimiste.

J-

JDBC(*Java DataBase Connectivity*) : interface de programmation (API, en anglais) destinée à mettre en relation des bases de données (le plus souvent relationnelles) et des programmes informatiques de gestion (*middleware*) écrits avec le langage Java. JDBC reprend la spécification ODBC comme référence de son implémentation.

jointure : opération permettant d'associer plusieurs objets de la base par le biais d'un lien logique de données entre les différentes tables ou vues.

jointure par fusion : technique de jointure faisant appel à l'algorithme de tri fusion, c'est-à-dire par rapprochement des valeurs lors de la lecture descendante dans deux listes triées.

jointure par hachage : technique de jointure qui consiste à transformer les données à joindre en données « hachées » (calculées à l'aide d'une fonction de hachage) afin de travailler sur un plus petit ensemble de données mieux ventilé.

jointure par boucle imbriquée : technique de jointure qui consiste à utiliser une boucle externe itérant sur les lignes d'une table, encapsulant une boucle interne itérant sur les lignes de l'autre table.

journal de transaction : outil traçant l'ensemble des opérations de mise à jour des données destiné à la finalisation des transactions, soit par validation (COMMIT), soit par retour arrière (ROLLBACK). Pour ce faire, sont inscrites dans le journal les valeurs des données avant de procéder à la mise à jour. L'annulation entraînant la relecture du journal.

joker : caractère particulier servant à définir un élément de recherche générique, par exemple une lettre quelconque ou une chaîne de caractères de longueur quelconque. Essentiellement utilisé dans l'opérateur LIKE de SQL ou dans la recherche textuelle.

K-

kanatype : forme d'écriture des idéogrammes japonais, déclinée en deux graphies : le katakana (à la graphie très anguleuse) et l'hiragana (à la graphie cursive). Le katakana étant plutôt destiné au titrage des documents tandis que l'hiragana est utilisé pour les textes « littéraires ».

L-

lecture logique : opération de lecture d'une page de données en mémoire.

Lebesgue, voir courbe de Lebesgue.

lecture physique : opération de lecture d'une page de données sur le disque.

ligne : élément horizontal d'une table, constitué d'une énumération de données de différents types. Unité de stockage logique minimale d'une table.

ligne fantôme : dans les index, ligne ayant été supprimée logiquement, mais dont l'emplacement n'a pas été réutilisé, ce qui conduit à de la fragmentation. Dans les transactions, ligne indésirable apparaissant dans le cadre d'une opération concurrente d'insertion dans une table et nécessitant d'augmenter le niveau d'isolation pour l'interdire afin d'éviter des résultats faux.

LOB (*Large Object*) : objet de grandes dimensions stocké dans la base de données. LOB est un terme générique pour désigner les CLOB (*Character LOB*, grand texte encodé en ASCII), les NCLOB (*National Character LOB*, grand texte encodé en UNICODE), ou encore BLOB (*Binary LOB*). Ces types de données génériques doivent être réalisés avec respectivement les types de données VARCHAR(MAX), NVARCHAR(MAX) et VARBINARY(MAX) dans SQL Server.

lot, voir batch.

logiciels de base de données : Pour concevoir une base de données, les professionnels plusieurs types d'outils logiciels :

- Des SGBD pour construire concrètement les bases de données ;
- Des outils de modélisation pour représenter au mieux le fonctionnement de la base de données ;
- Des outils d'administration destinés à optimiser les performances par exemple ou à réaliser des options de maintenance ;
- Des outils d'extraction et de recherche de données (data-mining).

LLD (Langage de définition de données), voir DDL.

LMD (Langage de manipulation de données), voir DML.

LOOKUP, voir recherche éparse.

LUN (*Logical Unit Number*) : identifiant d'une unité de stockage logique réalisée à partir de disques physiques originellement dans un système disque en SCSI. Aujourd'hui fréquemment employé pour désigner un espace de stockage vu comme un disque, taillé dans un SAN.

M-

masquage : (aussi appelé obfuscation) technique consistant à interdire l'accès à certaines informations. Permet, dans SQL Server, de masquer la définition des vues ou le code des routines.

MCD : Dans la méthodologie Merise destinée à créer des bases de données, il y a des outils dédiés aux traitements et aux données. Le **MCD (Modèle Conceptuel des Données)** est un des outils majeurs concernant les données.

Le **MCD** est une représentation graphique de haut niveau qui permet facilement et simplement de comprendre comment les différents éléments sont liés entre eux à l'aide de diagrammes codifiés dont les éléments suivants font partie :

- Les entités (1 rectangle = 1 objet) ;
- Les propriétés (la liste des données de l'entité) ;
- Les relations qui expliquent et précisent comment les entités sont reliées entre elles (les ovales avec leurs « pattes » qui se rattachent aux entités) ;
- Les cardinalités (les petits chiffres au dessus des « pattes »).

Utilisé assez tôt en conception de base de données, le **MCD** évoluera ensuite vers les autres outils de Merise, à savoir le MPD et le MLD.

Cet outil **permet d'échanger entre informaticiens et non-informaticiens** sur l'outil à informatiser. On peut ainsi à partir d'un MCD **valider et préciser des règles** qui s'appliqueront à la future base de données (d'après le MCD en image ci-dessus) :

- Une fiche est remplie par un et un seul chargé d'appui ;
- Un contact peut avoir différentes qualités (dans ce modèle « utilisateur, salarié, porteur de projet ») en fonction des fiches.

merge join, voir jointure par fusion.

Merise : Merise est une méthode informatique dédiée à la modélisation qui analyse la structure à informatiser en termes de systèmes. Le gros avantage de cette méthode est qu'elle permet de cadrer le projet informatique et de « discuter » en se comprenant entre utilisateurs et informaticiens.

Créée dans les années 70 sur commande de l'État français et destinée aux gros projets informatiques de l'époque, la méthode a perduré jusqu'à aujourd'hui. Son utilisation très répandue en Europe constitue un socle difficilement contournable lorsque l'on s'attache à la création de bases de données.

Merise est en fait un outil analytique qui facilite la création de base de données et de projets informatique. Le principal auteur de la méthode est Hubert Tardieu qui se basa sur les travaux autour du modèle relationnel de Codd. Concrètement Merise (que l'on prononce Meurise) permet de :

- hiérarchiser les préoccupations du gestionnaire de projet informatique
- décrire le fonctionnement du système à informatiser et notamment :
 - Les données (MCD) : quelles sont les relations et les dépendances entre les différents acteurs (*client – commande – produit – fournisseur par exemple*)
 - Les traitements (MCT) : comment les acteurs travaillent-ils ensemble (*comment se passe une commande concrètement par exemple*)

- proposer une implémentation logique (MLD, MLT) du point précédent
- Proposer une construction concrète et utilisable du point précédent (MPD, MOT)

mémoire virtuelle : extension de la mémoire vive physique représentée par un ou plusieurs fichiers disque sur lequel on effectue des opérations de déchargement ou de chargement au fur et à mesure de l'utilisation des données qui y sont stockées. La mémoire étant organisée en pages, le processus de montage et descente des pages de la RAM vers les disques est appelée pagination, et procède généralement pas échange (swap disque).

métadonnées : données permettant de décrire les données de la base, constituée par des tables dites « système » et accessibles depuis SQL Server 2005 uniquement par des vues représentant la vision logique (schéma SQL de nom INFORMATION_SCHEMA - norme SQL) ou physique (schéma SQL sys -vues spécifiques SQL Server).

mise à jour : toute opération conduisant à une écriture des données : INSERT, UPDATE, DELETE ou MERGE.

mirroring : technique de haute disponibilité consistant à reproduire en permanence, de façon synchrone ou asynchrone, une base de données sur un autre serveur distant ou sur plusieurs autres, par le biais de la couche HTTP.

MLD : Le **MLD** ou **Modèle Logique des Données** est simplement la représentation textuelle du MPD. Il s'agit juste de la représentation en ligne du schéma représentant la structure de la base de données. Il n'y a pas de travail poussé à réaliser à cette étape, il s'agit juste d'appliquer quelques règles toutes simples. Cette étape est parfois omise.

On représente ainsi les données issues de la modélisation Merise sous la forme suivante :

- Chaque ligne représente une table ;
- C'est toujours le nom de la table qui est écrit en premier ;
- Les champs sont listés entre parenthèses et séparés par des virgules ;
- Les clés primaires sont soulignées et placées au début de la liste des champs ;
- Les clés étrangères sont préfixées par un dièse.

À noter que le MLD prend parfois un R et devient MLDR : le R signifiant simplement Relationnel.

modèle conceptuel : représentation sémantique d'une base de données à l'aide de structures de données et de règles de description basées sur des « idées » (concepts), donnant lieu à un schéma dit « conceptuel » découlant de l'analyse d'un système. L'acronyme MCD signifie « modèle conceptuel de données ».

modèle externe : représentation « visuelle » d'une base de données (vues, fonctions utilisateur, procédures...), destinée à faciliter la manipulation des données pour les applications externes. L'acronyme MED signifie « modèle externe de données ».

modèle logique : représentation mathématique d'une base de données (relations) découlant de l'application de l'algèbre relationnelle. L'acronyme MLD signifie « modèle Logique de données ».

modèle physique : représentation physique d'une base de données (tables, vues, liens d'intégrité, déclencheurs, stockage, index...), prenant en compte les spécificités du SGBD. L'acronyme MPD signifie « modèle physique de données ».

modèle relationnel : nom commun donné à une représentation de base de données spécifique à un SGBDR. Voir modèle physique.

monadique : se dit d'une fonction ou d'un opérateur qui n'admet qu'un seul argument. Par exemple, la fonction SQL UPPERCASE, ou encore l'opérateur – (négation).

mot noir : mot vide de sens que l'on évite de considérer dans les requêtes de recherches *plain text* (par exemple, articles, pronoms, conjonctions de coordination...).

Moteur de base de données : Dans un logiciel de base de données il y a plusieurs composants qui discutent entre eux. Certains composants discutent aussi avec l'extérieur et un des composants les plus importants est le **moteur de base de données**.

Le moteur de base de données a un rôle central. Il est le garant des opérations CRUD et vérifie, autorise et exécute les actions suivantes :

- Création de données ;
- Lecture de données ;
- Mise à jour de données ;
- Suppression de données.
- Mais le moteur de base de données a d'autres tâches qui lui sont dévolues :
- La gestion des fichiers et du stockage sous-jacent ;
- L'optimisation des accès et des performances (mémoire cache, antémémoire);
- La bonne application des règles de gestion (certaines découlant de CRUD, d'autres non) comme le suivi des droits d'accès, les mécanismes de verrous, les transactions ou l'intégrité référentielle ;
- La journalisation des opérations (fichier log de traçabilité des opérations) ;
- L'automatisation de certaines opérations traitées ensemble (les procédures stockées).

MTBF (*Mean Time Between Failures*) : moyenne de temps de bon fonctionnement entre mise en route initiale et première panne ou entre deux pannes d'un système.

MPD : Dans la méthodologie Merise, le **MPD (Modèle Physique des Données)** fait suite au MCD. Ensuite viendra le MLD.

L'étape de création du MPD est presque une formalité comparée à la création du MCD. En s'appuyant sur des règles simples (et qui fonctionnent à tous les coups), l'analyste fait évoluer sa modélisation de haut niveau pour la transformer en un schéma plus proche des contraintes des logiciels de bases de données. Il s'agit de préparer l'implémentation dans un SGBDR.

Concrètement, cette étape permet de **construire la structure finale de la base de données** avec les différents liens entre les éléments qui la composent. Pour la peine, on change aussi de vocabulaire :

Les entités se transforment en tables ;

Les propriétés se transforment en champs (ou attributs) ;

Les propriétés se trouvant au milieu d'une relation génèrent une nouvelle table ou glissent vers la table adéquate en fonction des cardinalités de la relation ;

Les identifiants se transforment en clés et se retrouvent soulignés. Chaque table dispose d'au minimum 1 clé dite primaire ;

Les relations et les cardinalités se transforment en champs parfois soulignés : il s'agit de créer des « clés étrangères » reliées à une « clé primaire » dans une autre table.

C'est aussi à cette étape que les aspects performances sont pris en compte. La construction d'index ou des opérations de « dénormalisations » faciliteront la rapidité de la base de données mais pourront aussi entraîner des incohérences. Ce dernier point est à réserver aux professionnels aguerris donc.

Modélisation : En conception de base de données, la **modélisation** est l'étape indispensable (pour le professionnel compétent) qui permet de comprendre les processus métiers et de les transcrire informatiquement en flux de données et de traitements.

Derrière cette définition un peu rude se trouve un besoin simple : comprendre comment font les utilisateurs actuellement (avec ou sans informatique), isoler les suites d'actions, les règles de gestion du métier et traduire le tout dans un ensemble de documents à la fois compréhensible par l'utilisateur et par l'informaticien.

Afin d'être réutilisable et compréhensible par des tiers n'ayant pas participé au projet initial, la modélisation repose sur des méthodes (comme Merise par exemple. À ne pas confondre avec UML).

La modélisation est la brique indispensable qui permet de construire une base de données solide et conforme aux attentes des utilisateurs. Cette étape est préalable à la construction concrète de la base de données et la construction des différents écrans que manipuleront les utilisateurs.

MPD, voir modèle physique.

N-

NAS (*Network Attached Storage*) : boîtier de stockage en réseau, généralement partagé entre différents ordinateurs. Ne convient généralement pas pour le stockage des bases de données de SQL Server sauf attachement particulier (mais performances moins bonnes).

NCLOB, voir LOB.

niladique : se dit d'une fonction qui n'admet aucun argument (nombre d'arguments nuls). Par exemple, la fonction SQL `CURRENT_DATE`.

nonclustered : se dit d'un index qui n'est pas clustered. Voir ce dernier mot.

n-uplet, voir tuple.

Nested Loop Join, voir jointure par boucle imbriquées.

nested triggers : déclencheurs imbriqués.

NULL : marqueur d'absence de valeur. Dans SQL Server, ce marqueur est représenté par un bit dans le « slot » représentant la ligne de la table ou de l'index.

NoSQL : **NoSQL** (ou **Not Only SQL** ou encore **Non-relationnel SQL**) est une façon de concevoir les bases de données qui repose non plus sur la notion relationnelle des SGBD (des tables avec des liens entre elles) mais sur une architecture plus simple à manipuler.

Initialement, le besoin est apparu avec les très très grosses bases de données qui avaient des problèmes de rapidité et de cohérence (voir tests ACID). En effet, lorsqu'il y a des milliards de données à traiter, les performances peuvent se dégrader de manière significative. Et lorsqu'il faut s'assurer que les données sont bien synchronisés en temps réel sur différents lieux aux quatre coins du globe, des problèmes insurmontables apparaissent (connu par exemple sous le nom de théorème CAP).

Le principe du NoSQL est donc de **simplifier au maximum pour améliorer les performances**. Cette simplification entraîne des effets de bords indésirables. Les relations très structurées et permettant de représenter des interactions complexes sont délaissées au profit de modèles plus simples de représentation : les informaticiens nomment cela un **tableau associatif** que l'on peut comparer à un immense tableur composé d'une clé (la première colonne) et de valeurs qui en découlent (la ou les autres colonnes).

NoSQL ne s'oppose donc pas à SQL et il est fréquent de retrouver les deux dans les systèmes informatiques d'une certaine ampleur. On utilise alors SQL et les bases de données relationnelles pour modéliser finement et NoSQL pour des besoins plus simples mais où la rapidité est cruciale.

Effet de mode à causes de leur utilisation par les ténors du web, le NoSQL est encore très jeune et le marché est loin d'être mature : les prémices de NoSQL date de 1998 mais le coup d'envoi de cette technologie ne date réellement que de 2009. Pourtant, on trouve déjà sur le marché différents outils spécifiques ou compatibles NoSQL : Cassandra, Voldemort, MongoDB, SimpleDB et CouchDB notamment.

NOT NULL : contrainte de colonne indiquant que la colonne ne peut rester vide (valeur obligatoire).

NUMA(*Non Uniform Memory Access*) : accès mémoire non uniforme - système d'adressage multiprocesseur dans lequel les zones mémoire sont séparées et accessibles par certains processeurs en fonction des différents bus alloués. Le système NUMA a été conçu pour pallier les limites de l'architecture SMP dans laquelle tout l'espace mémoire est accessible par tous les processeurs, ce qui induit de la contention du fait de la concurrence d'accès.

O-

occurrence : une ligne particulière évaluée dans une table.

En base de données, on distingue le modèle d'une côté (la structure avec les entités et les relations) et d'un autre côté le contenu (il faut bien les remplir avec des valeurs ces beaux modèles).

Et bien, une **occurrence**, c'est tout simplement une « **ligne** » de valeurs. Dans une entité, une occurrence correspond à l'ensemble des valeurs des propriétés rattachées à un seul identifiant. Dans une relation, une occurrence correspond à l'ensemble des valeurs des propriétés de la relation (représenté par les clés de chaque entité liée) : on l'appelle alors une **occurrence de relation**.

ODBC (*Open DataBase Connectivity*) : interface de programmation (API, en anglais) destinée à mettre en relation des bases de données (le plus souvent relationnelles) et des programmes d'informatique de gestion (middleware). Initié par un comité indépendant (le SQL Access Group), ODBC est devenu populaire grâce à Windows de Microsoft. Il est basé sur l'API CLI de SQL.

octroi (de privilège) : autorisation donnée à un utilisateur de lancer certaines commandes SQL sur certains objets.

offset (décalage) : par extension, désigne une adresse accessible de manière relative.

OLAP (*On-Line Analytical Processing*) : catégorie de logiciels permettant une analyse d'informations multidimensionnelles stockées dans une base de données dont la structure particulière (décisionnelle) est prévue à cet effet.

OLTP (*On-Line Transaction Processing*) : type de traitement informatique dans lequel une base de données effectue des transactions (écriture par ajout, modification ou suppression) à la demande de l'utilisateur et par opposition au mode batch.

optimiseur : organe d'un SGBD relationnel destiné à élaborer le plan d'exécution d'une requête le plus adéquat (donc le moins coûteux).

ORM : Le terme **ORM** vient de l'anglais **Object-Relational Mapping**. En français, on croise parfois (rarement) le terme **mapping objet-relationnel**.

Il s'agit d'une technique de programmation informatique qui permet de simplifier l'accès à une base de données en proposant à l'informaticien des « objets » plutôt que d'accéder directement à des données relationnelles. Ce niveau d'abstraction supplémentaire fait correspondre le monde objet (programmation orientée objet) et le monde relationnel (les bases de données relationnelles classiques et massivement utilisées aujourd'hui).

L'intérêt principal est d'éviter beaucoup de code très similaire pour l'informaticien ce qui est source d'erreurs et de temps perdu. Ainsi plutôt que de devoir recoder à la main les fonctions de base (CRUD – création, lecture, modification, suppression), les informaticiens ajoutent une couche logicielle intermédiaire qui se charge de la « traduction ».

La difficulté principale consiste à trouver la « bonne traduction » pour toutes les opérations possibles. Souvent, ce n'est pas possible et il faut faire des concessions.

ordre SQL (ou commande SQL) : construction littérale dont la syntaxe répond aux spécificités du SQL et qui intime au moteur de la base de données d'effectuer tel ou tel travail (SELECT, UPDATE, CREATE, ALTER, GRANT, etc.).

OS (*Operating System*) : système d'exploitation permettant à un ou plusieurs programmes d'utiliser au mieux des périphériques de différentes natures et d'en partager les ressources entre les programmes. Bien qu'installé sur l'OS Windows, SQL Server possède son propre OS de nom SOS pour SQL Server Operating System).

overclocking : surcadencement. Manipulation destinée à faire fonctionner un processeur à une vitesse supérieure à sa vitesse nominale. Ceci diminuant en général le MTBF.

P-

page (de données) : unité de base de structuration physique des données (table, index) d'une base de données. Les pages de données sont identiques sur le disque (physique) et en mémoire (logique) et constituent la plus petite unité de lecture/écriture du SGBDR.

partitionnement : technique consistant à scinder les données d'objets volumineux (tables ou index) afin de répartir la charge et diminuer le coût d'accès.

parseur : module d'analyse syntaxique d'une commande SQL permettant d'une part d'en vérifier la syntaxe et d'autre part de la traduire en algèbre relationnelle.

PCA (Plan de continuité d'activité) : ensemble des moyens à mettre en oeuvre afin de pallier une panne en diminuant le plus possible, ou en éradiquant les temps d'indisponibilité (haute disponibilité). La plupart du temps assurée par des outils hardware (redondance) et logiciels (réplication). SQL Server dispose de nombreux moyens internes pour assurer cette tâche (log shipping, clustering, mirroring, AlwaysOn, réplication de données...). Voir aussi PRA.

plain text, voir recherches plain text.

persistance : mécanismes assurant qu'un programme puisse se terminer sans que ses données ou son état d'exécution ne soient perdus. Dans un SGBDR, fonction assurée par la journalisation des transactions.

PFS (*Page Free Space*) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer la quantité d'espace résiduel parmi les 8 088 pages qui suivent.

PITR (*Point In Time Recovery*) : restauration avec arrêt à un point dans le temps. Technique consistant à arrêter une restauration effectuée à partir d'une sauvegarde transactionnelle à une date/heure précise afin de récupérer une base dont l'état des données est à un instant du passé.

plan d'exécution (voir plan de requête)

plan de requête : ensemble des opérations unitaires (lecture des données, opérations d'algèbre relationnelle telles que jointures, restriction, union... calcul divers tels que agrégation, tri...) que le SGBDR a prévu d'effectuer en réponse à une requête.

plateau (disque) : composant d'un disque magnétique destiné au support de l'information et constitué d'un disque métallique en rotation à vitesse constante dont la surface est une couche ferromagnétique destinée à recevoir l'information par polarisation magnétique. La plupart des disques possèdent plusieurs plateaux (généralement 3).

point de terminaison : point d'accès permettant l'utilisation du protocole HTTP pour communiquer avec le serveur SQL.

position ordinale : rang d'un élément au sein d'une série, d'une collection. L'emplacement par défaut des colonnes d'une table SQL est défini par leur position ordinale dans le dictionnaire des données (colonne `ORDINAL_POSITION` de la vue `INFORMATION_SCHEMA.COLUMNS`).

PRA (Plan de reprise d'activité) : ensemble des moyens à mettre en oeuvre afin de pallier un sinistre majeur en diminuant le plus possible les temps d'indisponibilité (haute disponibilité). La plupart du temps assurée par des outils hardware (redondance) et logiciels (réplication) répartis sur différents sites d'exploitation. SQL Server dispose de nombreux moyens internes pour assurer cette tâche (log shipping, clustering, mirroring, AlwaysOn, réplication de données...). Voir aussi PRA.

prédicat : expression booléenne dont l'évaluation donne Vrai ou Faux. En SQL, du fait de la logique trois états, un prédicat peut prendre trois valeurs : TRUE, FALSE ou UNKNOWN.

PRIMARY KEY : contrainte de clé primaire applicable à une ligne de table.

privilège : possibilité accordée à un utilisateur de lancer telle ou telle commande sur tel ou tel objet.

procédure stockée : routine Transact SQL ou SQL CLR stockée au sein de la base dans une table système, composée de code que l'on peut lancer depuis une application cliente. Une procédure peut lancer tout type de requêtes (DDL, DML, DCL, TCL...), gérer des transactions et exécuter du SQL dynamique. Elle admet des paramètres « scalaires » en entrées (y compris variable table, XML, spatial...) et peut renvoyer des valeurs scalaires et des jeux de données.

Une **procédure stockée** (*stored procedure*) est une suite d'instructions SQL qui s'exécutent les unes à la suite des autres et qui est conservée de façon durable (1 jour, 1 mois, 1 an, il n'y a pas de limite). Pour faire une analogie avec Excel, on pourrait dire qu'il s'agit d'une macro qui répète un ensemble de demandes simples préalablement enregistrées.

Une fois créée, on retrouve donc la procédure stockée parmi les autres objets de la base de données (comme une table) et on peut l'appeler pour l'exécuter.

À l'intérieur d'une procédure stockée, on peut utiliser des conditions et des boucles (IF, WHILE, LOOP, REPEAT, CASE, ITERATE...) ce qui permet d'effectuer des opérations beaucoup plus riches sur les données directement dans la base.

Tous les SGBD ne permettent pas l'utilisation de procédures stockées (MySQL ne le permet que depuis la version 5).

```
1 DELIMITER
2 CREATE PROCEDURE afficher_client_selon_region (IN region_id INT)
3 BEGIN
4     SELECT numauto, nom, region
5     FROM Client
6     WHERE region = region_id;
7 END |
8 DELIMITER ;
```


processeur : composant électronique central de tout ordinateur, qui exécute les instructions machine des programmes informatiques.

processus : code informatique devant effectuer un traitement exécuté sur un processeur.

profilier (SQL) : outil de trace des activités opérationnelles, notamment transactionnelles, de l'instance dont la vocation est de diagnostiquer par le biais de mesure, les problèmes potentiels, par exemple de performances des requêtes et routines.

projection : opération de l'algèbre relationnelle qui consiste à ne prendre que certains attributs d'une relation en précisant leur nom. Dans le langage SQL la projection se fait dans la clause SELECT de l'ordre SELECT.

propriétaire : par rapport aux objets relationnels, utilisateur SQL particulier ayant créé l'objet considéré ou l'ayant dernièrement modifié.

Propriété : En conception de base de données avec Merise, il y a des termes un peu barbares. **Propriété** est un de ces termes.

- Une **propriété** est un « champ » de l'entité. Dans l'exemple ci-dessus, par exemple, on a l'entité « séjour » qui contient des propriétés qui commence par « sej_ ». Chaque propriété donne une information précise les données stockées liées à l'entité. Le contenu de chaque propriété varie pour chaque occurrence. Ainsi « sej_datedebut » et « sej_datefin » varieront d'un séjour à l'autre ;
- Chaque entité possède au moins une propriété ;
- La première propriété d'une entité est un numéro unique, l'identifiant ;
- Ces propriétés sont dites « **atomiques** » c'est à dire que c'est la donnée la plus fine que l'on trouve dans la base (il n'y a pas de sous-données aux propriétés) ;
- Les propriétés se retrouvent aussi dans certaines associations (les CIM). Ici, c'est le cas avec « passecommande » qui contient 3 propriétés et avec « estreferent » qui contient 1 propriété ;
- On dit que les entités et les CIM sont « **porteuses** » de propriétés.
- On ne doit pas avoir de propriétés déductibles (c'est à dire qui se calculent). L'exemple « sej_duree » est un faux ami ici : il ne calcule pas la durée d'un séjour mais correspond à un libellé propre à l'entreprise (en fait un arrondi sur mesure de la durée réel afin d'être simple et parlant) ;
- Par convention et pour simplification, on utilise des préfixes aux propriétés. Ça permet de s'y retrouver.

Une propriété est parfois nommée « attribut » ou « caractéristique ».

Propriétés ACID : Abréviation pour Atomicity, Consistency, Isolation et Durability. En français Atomicité, Cohérence, Isolation et Durabilité.

Il s'agit des propriétés nécessaires pour garantir une transaction informatique. Très utiles dans les bases de données lors d'opérations critiques (par exemple lors de transferts d'argent), les **propriétés ACID** se matérialisent par les fonctions de transactions. Toutes les bases de données ne mettent pas en place ces propriétés.

Conceptualisés en 1983 par Jim Gray dans les années 1970, les propriétés ACID doivent leur nom à Reuter et Härder.

- **Atomicité** : Tout ou rien. Soit l'opération se fait en entier, soit elle ne se fait pas du tout. La notion d'atomicité sous-entend la possibilité de défaire une opération avortée.
- **Cohérence** : L'opération doit assurer que la base de données sera dans un état valide après l'opération.
- **Isolation** : L'opération doit se faire en toute autonomie sans dépendance à une autre opération.
- **Durabilité** : En cas de problème important (coupure d'électricité), les modifications apportées sont bien enregistrées.

Proxy : objet non relationnel spécifiquement utilisé par l'agent SQL pour définir un contexte de sécurité dans lequel une étape de travail peut être exécutée lorsqu'elle doit utiliser des ressources externes à l'instance SQL

pseudo-table : table volatile non persistante n'ayant pas d'existence propre en dehors de certains contextes. Les pseudo-tables *inserted* et *deleted* ne sont disponibles que dans les déclencheurs ou la clause OUTPUT d'un ordre SQL de mise à jour. Elles empruntent leur structure à la table cible. On peut les considérer comme des vues « volatiles ».

R-

RAID (*Redundant Array of Independant Disk*) : techniques permettant de répartir des données sur plusieurs disques durs afin d'améliorer soit les performances, soit la sécurité ou la tolérance aux pannes de l'ensemble du stockage.

RAM (*Random Access Memory*) : mémoire à accès aléatoire. C'est la mémoire vive de l'ordinateur, qui sert de cache à SQL Server pour les données, les plans de requête et de nombreuses métadonnées.

RBS (*Remote BLOB Storage*) : outil permettant de stocker les LOB par le biais de dispositifs de stockages externes aux serveurs SQL

recherche (ou SEEK en anglais) : technique qui consiste à accéder à une donnée par dichotomie dans un index afin d'accélérer la quête de ladite donnée.

recherche éparsé (ou LOOKUP en anglais) : technique qui consiste à accéder à diverses données éparsées par lancement en parallèle d'autant d'opérations de dichotomie dans l'index afin d'accélérer la quête desdites données.

recherches plain text : possibilité de recherches de mots ou combinaisons de différents mots à l'aide d'opérateurs particuliers opérants à travers des données littérales.

RECOVERY, voir récupération.

récupération : phase pendant laquelle la base ayant redémarrée, rejoue les dernières transactions dont les données n'ont pas fait l'objet d'une écriture physique sur disque.

récurtivité : se dit d'un processus qui fait appel à une nouvelle instance de lui-même (réentrant). La notion de récurtivité a été introduite par le mathématicien américain Stephen Kleene. En informatique, une fonction qui contient un appel à elle-même est dite réursive. Il est nécessaire qu'elle contienne une condition d'arrêt, sous peine de boucler indéfiniment.

Recursive Triggers : déclencheurs réursifs.

redondance : pour les données, duplication inutile, contre-performante et dangereuse des informations. Pour le hardware, moyen utile destiné à améliorer la fiabilité des systèmes par doublement des équipements afin de prévenir la défaillance de l'un d'entre eux.

relation : objet mathématique porteur de données (voir algèbre relationnelle) contenant des attributs et affecté d'une clé. Toutes les occurrences d'une relation doivent être évaluées (pas de "NULL") et il n'y a pas de doublons (du fait de l'obligation de clé). Une table peut être relationnelle ou non. Elle sera relationnelle à deux conditions : qu'elle possède une clé primaire et que toutes ses colonnes soient NOT NULL.

relationnel[le[s]] : qui se rapporte aux relations. Base de données relationnelle : base de données structurée, contenant des relations (voir ce mot).

requête : demande formulée en vue d'obtenir une faveur, une mesure, un résultat. Dans les bases de données, ordre lancé au serveur afin d'exécuter une tâche ou d'obtenir un résultat.

révocation (de privilège) : retrait d'un privilège précédemment octroyé.

restauration : mécanisme permettant de reconstituer une base de données à partir d'une sauvegarde préalablement effectuée de ladite base. La commande restauration de SQL Server est RESTORE.

restriction : opération de l'algèbre relationnelle qui consiste à ne prendre que certains tuples d'une relation en fonction d'un critère de filtrage. Dans le langage SQL, la restriction se fait principalement dans la clause WHERE.

rôle : collection de privilèges que l'on peut attribuer à un utilisateur.

ROLLBACK : commande d'annulation de transaction.

routine : désigne tout type de programme Transact SQL qu'il soit UDF (fonction utilisateur), procédure, déclencheur ou encore « batch ».

S-

SAN (*Storage Area Network*) : réseau de stockage permettant de mutualiser des ressources disques. Contrairement aux NAS (voir ce mot), les SAN sont directement attachés aux serveurs et ne passent ni par les hubs ni par les switches d'un réseau IP, mais utilise des protocoles spécifiques tel Fibre Channel (norme ANSI X3T11) permettant une connexion haut débit (de l'ordre du gigabit par seconde). SQL Server ne fonctionne de façon optimale que lorsque le serveur utilise un SAN dédié.

Sargable, voir searchable.

sauvegarde : mécanisme produisant un ensemble de données permettant de reconstituer l'intégralité de la base à un instant donné. SQL Server permet la sauvegarde à chaud de toute base à l'aide de la commande BACKUP et dans différentes modes (complète, différentielle ou transactionnelle) de l'intégralité de la base, d'un fichier ou encore d'un groupe de fichier.

SCAN, voir balayage.

schéma de la base : structure et articulation de l'ensemble des objets relationnels de la base (tables, vues, procédures, contraintes, déclencheurs...). Une opération de modification de schéma impose de lancer une commande du DDL SQL.

Schéma relationnel : Le schéma relationnel est représentation qui permet de mettre en évidence les relations entre les données. Ce schéma a une syntaxe, un formalisme, qui lui est propre. Et vous devez connaître. Dans un schéma relationnel, les tables sont appelées « relations » et les champs appelés « attributs ».

Exemple :

Facture(numFacture, #client, dateFacture, montantHT, montantTVA,...)

On note que la clé primaire est numFacture doit toujours être souligné et la clef étrangère est #client toujours précédés d'un #. Facture est le nom de la table.

Les relations entre tables sont dictées par les clefs primaires et clefs étrangères.

- a) **Relation « un à plusieurs »** : entre tables, il peut exister un ou plusieurs liens dits « de un à plusieurs » (one to many). Dans l'exemple ci-après, une facture est par exemple associée à un client. A une facture correspond par conséquent un et un seul client. Inversement, à un client peuvent correspondre aucune à plusieurs factures.

Table « Facture »					Table « Client »		
numFacture	client	dateFacture	montantHT	...	numClient	nomClient	...
1	1	05/01/2020	500,00		1	Nestle	
2	2	05/01/2020	750,00		2	Vinci	
3	2	06/01/2020	25000,00	
4	1	06/01/2020	600,00				
...				

On comprend que le « 2 » figurant le champ « client (clef étrangère) » de la table champ renvoie vers le client de la table Client, c'est-à-dire Vinci.

Une clef étrangère fait toujours référence à une clef primaire.

- b) Relation « plusieurs à plusieurs » :** Entre tables, il peut également exister un ou plusieurs liens dits « de plusieurs à plusieurs (many to many) ». Par exemple, un salarié peut travailler sur plusieurs projets. Inversement, plusieurs salariés peuvent travailler sur le même projet. On obtiendra un schéma relationnel de la forme suivante :

Salarie(numSalarie, prenomSalarie, nomSalarie,...)

Clef primaire : numSalarie

Equipe(#salarie, #projet)

Clef primaire : salarie, projet

Clefs étrangères :

- *Salarie référence au champ numSalarie de la relation Salarie*
- *Projet en référence au champ numProjet de la relation Prijet*

Projet(numProjet, dureeProjet, prixProjet,..)

Clef primaire : numProjet

schéma SQL : conteneur situé au niveau de la base et destiné à stocker les différents objets relationnels de la base de données. L'idée étant d'organiser différents espaces de stockages logiques afin de ventiler ces objets, par exemples dans des « catégories » fonctionnelles. Dans SQL Server, le schéma par défaut est dbo. Il en existe d'autres comme INFORMATION_SCHEMA et sys.

SEEK, voir recherche.

sécurisable : objet de la base de données relationnelle pouvant être sécurisé (table, vue, procédure...).

sélectivité : indice permettant de déterminer approximativement le nombre de lignes résultant d'une opération de recherche portant sur une valeur précise. Calculée par division du nombre de valeurs sur le nombre de lignes.

séquence : objet de la base, indépendant de toute table et destiné à fournir une numérotation monotone (auto-incrémentation).

serveur : définit un système qui peut indifféremment être une machine physique ou virtuelle (PS, machine Windows) ou bien logicielle (serveur SQL, Web...).

service pack (SP) : ensemble de mises à jour, corrections, améliorations de logiciels livré sous forme d'un package installable. Microsoft fournit régulièrement des « services packs » pour Windows comme pour SQL Server qu'il est nécessaire d'installer la plus rapidement possible. (Voir aussi CU - Cumulative Update)

session : laps de temps pendant lequel un utilisateur ayant réussi à s'authentifier navigue dans le serveur SQL et dans les bases relationnelles pour y effectuer différentes opérations.

SGAM (*Shared Global Allocation Map*) : page technique particulière d'un fichier de données d'une base SQL Server permettant d'indiquer par un bit à 1 parmi les 63 904 extensions lesquelles sont mixtes et ont au moins une page vides de toutes données utilisable.

SGBD : logiciel permettant à différents utilisateurs de visualiser, créer, maintenir ou supprimer de l'information et des structures de données dans une base de données. Le SGBD peut être relationnel; dans ce cas, on lui donne l'acronyme SGBDR.

SIG (Système d'information géographique, en anglais GIS pour Geographical Information System) : système de gestion de données planaires et spatiales permettant d'élaborer des objets géométriques et géographiques.

signature : donnée ajoutée à un élément informatique (par exemple programme) et vérifiable par le biais d'un certificat, afin d'obtenir une assurance sur l'identité de la source de l'objet.

slammer : célèbre « vers » ayant contaminé de nombreux serveurs SQL 2000 suite à une faille de sécurité qui avait été détectée et corrigée six mois plus tôt, mais dont les administrateurs des serveurs SQL n'avaient pas pris la peine mettre en oeuvre le correctif.

snapshot (voir également database snapshot, isolation snapshot) : cliché, c'est-à-dire captation de données à un instant précis pour copie de travail.

snapshot isolation, voir Isolation snapshot.

SQL (*Structured Query Language*) : littéralement langage structuré de requêtes. Langage informatique à base d'ordres permettant de créer des structures de données relationnelles, d'y stocker l'information et de la maintenir, d'extraire les données et de gérer des utilisateurs et leurs autorisations. Le langage SQL n'est pas le seul langage opérant sur les bases de données. On peut noter le langage QBE de Zloof ou QUEL à l'origine dans Ingres.

SRID (*Spatial Reference IDentifier*) : identificateurs de référence spatiale. Code correspondant à un système de référence spatial (paramètres) basé sur un ellipsoïde spécifique destinée à fournir une approximation du mappage de la terre sur une zone considérée.

SSAS (*SQL Server Analysis Services*) : solution Microsoft pour BI incluant un moteur décisionnel pour la création et l'analyse des cubes, y compris avec le data mining et fournit avec les versions Developer, Standard et BI.

SSD (*Solid State Drive*) : disque électronique dont le support de stockage est constitué de mémoire flash.

SSIS (*SQL Server Integration Services*) : ETL de Microsoft fournit avec les versions Developer, Standard et BI.

SSMS (*SQL Server Management Studio*) : outil client pour la gestion et le développement des applications utilisant une instance SQL Server. Existe en deux versions : Express (allégée) et professionnelle (complète) venant avec les éditions Developer, Standard, BI et Enterprise.

SSRS (*SQL Server Reporting Services*) : outil de reporting de Microsoft fournit avec les versions Developer, Standard, BI et Express with Advanced Services.

slot : emplacement physique d'une ligne (par réservation d'espace) dans une page de données relatif à une table ou un index.

socket : emplacement de CPU physique.

spatial : ensemble des techniques mises en oeuvre pour traiter les données du SIG (voir ce mot), notamment cartographie géodésique.

split (de page) : opération consistant à couper une page d'index en deux lorsqu'elle est pleine, afin d'y ajouter une ligne. Conduit à de la fragmentation.

SMTP (*Simple Mail Transfer Protocol*) : protocole standardisé de communication utilisé pour le transfert de courrier électronique vers les serveurs de messagerie électronique.

statistiques (sous-jacentes aux index et à certaines colonnes de table) : métadonnées composées d'un échantillon des valeurs de l'index ou de la colonne obtenu par une méthode statistique (sondage ou analyse complète) permettant d'évaluer la distribution et donc la sélectivité d'une recherche dans l'index ou hors index (table).

storage, voir groupe de fichiers.

synchrone : se dit d'un traitement ou d'un mécanisme déclenché systématiquement par un autre de manière simultanée ou séquentielle. Les déclencheurs SQL sont activés de manière synchrone aux mises à jour des tables ou des vues. La mise à jour des index d'une base est synchrone avec les modifications de données, sauf pour les index textuel qui peuvent être mis à jour au fil de l'eau ou par batch.

synonymisation : technique consistant à indiquer les synonymes d'un mot.

T-

table : structure de données basique contenant une collection de données organisée sous forme de colonnes et formant des lignes.

TDE (*Transparent Data Encryption*) : technique consistant à crypter le stockage (au niveau des fichiers physiques) et non les données dans les tables.

thesaurus : document d'un index textuel contenant une liste définie de termes et leurs équivalents pour activer la recherche par synonyme ou équivalences.

thêta jointure : opération de jointure des tables qui n'est pas basée sur un critère d'égalité entre colonnes.

thread : littéralement fil d'exécution. Processus allégé ayant sa propre pile d'appel, mais partageant la mémoire. Dans SQL Server, les requêtes pouvant être parallélisées par le lancement de multiples threads, cela permet d'obtenir plus rapidement une réponse à une requête.

TPC (*Transaction processing Performance Council*) : organisme indépendant des éditeurs mettant en place des benchmarks destinés au SGBDR et recensant les performances de solutions globales.

transaction : ensemble d'actions qui modifient les données d'une base de données, d'un état cohérent à un nouvel état cohérent.

transaction explicite : transaction dont le démarrage (BEGIN TRANSACTION) et la finalisation (par COMMIT ou ROLLBACK) ont été initiés par l'utilisateur ou le code applicatif.

transaction implicite : transaction dont les frontières (démarrage et finalisation) ont été initiées automatiquement par le moteur relationnel.

transact SQL : langage spécifique à Microsoft SQL Server pour la programmation des routines (UDF, procédures stockées, routines ou batch...). À l'origine inventé par Sybase.

trigger, voir déclencheur.

tuple ou n-uplet : occurrences des attributs d'une entité ou d'une association dans le cadre d'un modèle conceptuel. Équivalent à la ligne d'une table ou d'une vue au niveau du modèle physique ou externe (ligne comprenant ou non des valeurs pour chaque colonne).

turbo boost : technologie présente sur les CPU Intel, permettant de désactiver à la volée certains coeurs tout en augmentant la fréquence des autres.

U-

UDF (*User Defined Function*) : fonction utilisateur développée par écriture d'un programme en langage Transact SQL et disponible dans la base pour être utilisée au sein de requêtes comme de routines SQL (UDF, procédure ou déclencheur).

UML : Abréviation d'**United Modeling Language**. Langage de Modélisation unifié en français.

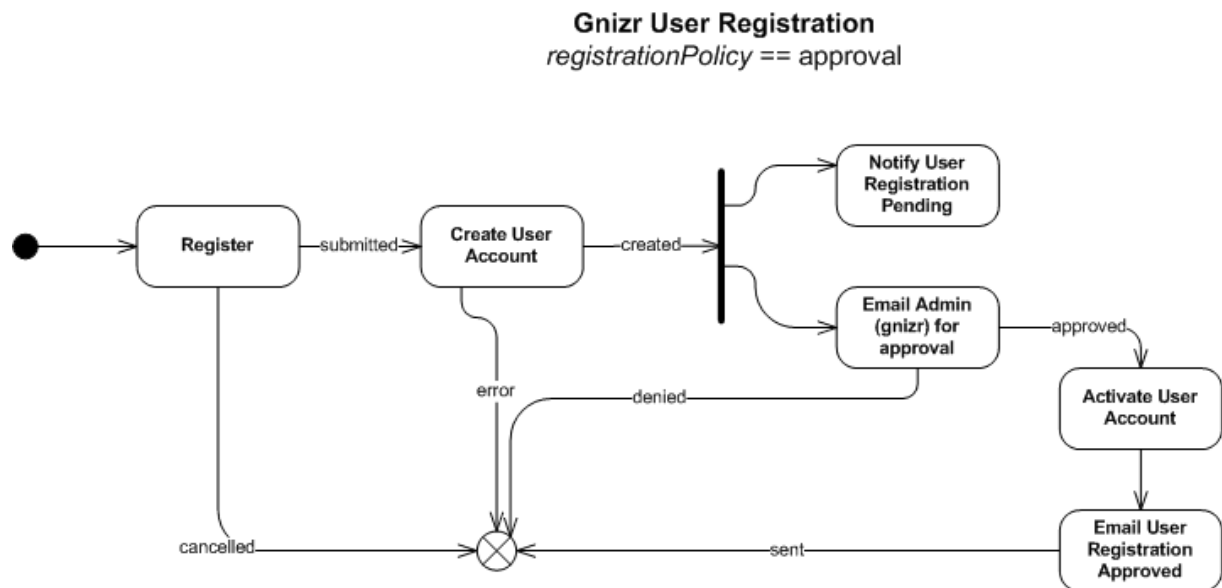
UML est un outil souvent utilisé en développement informatique. Il permet via des outils graphiques de représenter le fonctionnement d'un outil informatique (c'est particulièrement vrai en Programmation Orientée Objet ou UML permet de modéliser directement les objets qui seront ensuite manipulés par le programme informatique).

Outil plus récent que Merise, UML propose une petite quinzaine de diagrammes qui ne remplacent pas Merise. Les 2 outils ont des objectifs différents :

Merise est un ensemble cohérent pour construire des bases de données. C'est une méthode idéale de modélisation pour construire une base de données relationnelle.

UML ne propose pas de démarche ni d'organisation (ce n'est pas une méthode). C'est un outil idéal pour la conception logiciel dans un langage de type objet.

Les 2 peuvent fonctionner indépendamment ou de concert.



UNIQUE : contrainte (dite d'unicité) représentant une clé alternative.

unicité : données vérifiant une contrainte UNIQUE.

unicode : standard informatique développé par le Consortium Unicode, qui vise à donner à tout caractère de n'importe quel système d'écriture de langue (alphabets, idéogrammes, etc.) un

identifiant numérique unique, et ce quels que soient la plate-forme informatique ou le logiciel utilisé. Le codage Unicode est basé sur la norme ISO/CEI 10646 et représente un caractère sur 2 octets.

union : opération relationnelle consistant à concaténer les tuples de deux tables relations différentes. Se fait avec l'opérateur ensembliste UNION dans le langage SQL.

unité lexicale : élément de contenance de termes littéraux : mot, phrase, paragraphe, chapitre...

URL (*Uniform Resource Locator*) : chaîne de caractères utilisée pour accéder à des ressources web (serveur, pages...).

UUID (*Universal Unique Identifier*), voir GUID.

V-

verrou : mécanisme destiné à protéger certains modes d'accès à une ressource par d'autres processus que celui qui l'a posé.

verrou mortel (interblocage, étreinte fatale) : problème rencontré lorsque différents processus transactionnels ont posé des verrous sur des objets distincts, et que chacun de ces processus tente d'acquiescer un nouveau verrou sur un objet que l'autre a déjà verrouillé. La seule méthode de résolution connue à ce jour consiste à annuler la transaction.

verrou optimiste : technique consistant à vérifier en fin de traitement que la modification repose bien toujours sur le même jeu de données, c'est-à-dire n'ayant subi aucune modification. Pour ce faire, il est nécessaire d'obtenir une version d'entrée des informations et de voir en sortie si elle est toujours d'actualité. Dans le cas contraire, la transaction doit être abandonnée.

verrou pessimiste : technique consistant à verrouiller de manière préventive, c'est-à-dire à placer les verrous avant de commencer le traitement et à les libérer après finalisation de la tâche.

Verrou SQL : En SQL, un **verrou** de base de données est un **mécanisme de sécurité**. Il permet d'interdire ou de limiter l'accès à un élément d'une base de données. Ça peut être une ligne d'une table ou une table entière. L'utilité des verrous est simple : conserver l'ACIDité de la base.

Verrous SQL : 2 exemples concrets

Imaginons le cas d'un site de vente en ligne de billets d'avions avec un internaute qui souhaite réserver un vol à bord d'un avion. La réservation de la place se fera au dernier moment : c'est à dire lorsque la personne a validé sa commande et est en train de saisir sa CB et d'attendre que la banque veuille bien donner son feu vert. À ce moment, pour protéger l'utilisateur et lui permettre de finaliser sa transaction, il faut que le site web empêche les autres utilisateurs de réserver le même siège en même temps.

Pour cela, un **verrou en écriture** peut être posé sur le siège sélectionné pendant 3 minutes. Si au bout de 3 minutes, l'internaute n'a pas réalisé l'achat, le verrou peut être libéré et la vente peut être faite avec un autre internaute.

Autre exemple avec des statistiques. Imaginons qu'une base de données soit utilisée en temps réel par des utilisateurs tout autour du globe et qu'en même temps, il faille réaliser des calculs un peu lourd (15 minutes) de statistiques. Si pendant le calcul des utilisateurs viennent ajouter / modifier / supprimer des données, le programme de statistiques se retrouvera avec des données différentes entre le début de son calcul et sa fin.

Dans ce cas, on peut imaginer **verrouiller les données d'une table en écriture et ne conserver que l'accès en lecture**.

Verrous en base de données

Derrière cette **notion de verrou** se cache une autre notion, importante pour les bases de données : la **concurrence**. Autrement dit, qui et combien de personnes peuvent-elles accéder en même temps aux mêmes données sans compromettre les données de la base.

Les verrous SQL sont utilisés de façons très différentes d'un [SGBD](#) à l'autre et même entre différentes versions du même SGBD. Lorsqu'il y a des [transactions SQL](#) (une suite logique d'opérations qui doivent être réalisées d'une seule traite), les verrous sont nécessairement utilisés. On utilise alors les instructions COMMIT et ROLLBACK pour libérer les verrous après avoir validé/invalidé la suite d'opérations SQL.

- On utilise l'instruction LOCK TABLES en SQL pour verrouiller une ou plusieurs tables ;
- On utilise UNLOCK TABLES pour déverrouiller une ou plusieurs tables ;
- On peut réaliser des verrous exclusifs (EXCLUSIVE), partagés (SHARE) que ce soit sur une table ou sur une ligne.

vue : table virtuelle car elle ne nécessite aucune allocation en mémoire pour contenir les données (sauf vue indexée). Seule sa structure est stockée dans le dictionnaire de données. Une vue est créée à l'aide d'une instruction SELECT appelée « requête de définition ». Cette requête interroge une ou plusieurs tables ou vues. Une vue se « recharge » chaque fois qu'elle est interrogée.

Une base de données n'a pas besoin d'interface graphique pour fonctionner. Et c'est très perturbant lorsque l'on débute avec ces outils. Alors il existe des outils variés qui permettent de représenter la structure d'une base de données (le MCD de Merise) et les données qui se trouvent à l'intérieur (PhpMyAdmin pour MySQL par exemple). Mais cela ne suffit pas et ne doit pas être confondu avec les vues.

Pour des raisons de rapidité d'accès et de simplification, il existe les **vues**. Les vues sont des requêtes préparées à l'avance dont les paramètres (mais pas les résultats) sont stockés dans la base de données. Les interrogations complexes sont ainsi optimisées : en enregistrant un modèle plutôt qu'en le préparant de A à Z à chaque fois, l'utilisateur va **plus vite** et fait **moins d'erreurs**.

Concrètement, la base de données stocke la requête qui peut ensuite être appelée (une vue a un nom) et exécutée. Grâce à ce nom, on peut ensuite les utiliser tels des tables dans les requêtes. C'est pratique, rapide et génère des requêtes plus lisibles tout en masquant la complexité sous-jacente.

Les vues ont d'autres caractéristiques :

- Une vue peut masquer certaines données. Étant une requête, la vue n'affiche que les informations souhaitées. Elle permet ainsi d'extraire des données choisies et de masquer des données non souhaitées.
- Certains SGBD permettent de stocker les résultats des vues et peuvent faire office de mémoire tampon (cache).
- Les vues ne prennent pas de place. C'est la requête qui est enregistrée (pas le résultat).
- On peut réaliser des regroupements de données (somme, moyenne...) qui peuvent être réutilisés très facilement.
- Les vues sont habituellement en lecture seule (d'où le terme de « vue ») mais dans certains systèmes et sous certaines conditions, on peut manipuler les données d'une vue

Création d'une vue SQL

```
CREATE VIEW ToutesLesFactures AS
SELECT fac_id, fac_ref, cli_nom
FROM facture, client
WHERE facture.cli_id = client.cli_id ;
```

Utilisation d'une vue SQL

Plutôt que de retaper la requête ci-dessus avec sa jointure, on peut directement appeler la vue :

```
SELECT * FROM ToutesLesFactures ;
```

vue indexée : se dit d'une vue contenant des données synchrones provenant des tables qui définissent la vue.

X-

XML (*eXtensible Markup Language*) : métalangage de balisage « extensible » destiné à décrire des structures de données, spécialement utilisé sur le Web et dans l'échange de données informatisé. XML est un standard du World Wide Web Consortium (W3C) et sert de base pour créer des langages balisés spécialisés. Les bases de données relationnelles utilisent couramment XML Schéma, un dialecte XML spécialisé dans la gestion des données des bases relationnelles.

XML Schema : sous-ensemble du langage XML déterminant des spécifications de balises destinées à formaliser le contenu de documents XML. XML Schema est donc un métalangage de description de documents XML. Il est devenu un standard, après avoir été initié par Microsoft.

XPath : langage, non XML, destiné à localiser un fragment de document XML. Il est souvent utilisé conjointement avec XQuery.

XQuery : langage de requête informatique permettant la manipulation de documents XML : extraction, calculs ou encore construction de nouveaux documents ou fragments XML.

XSLT (*Extensible Stylesheet Language Transformations*) : langage de transformation de document XML permettant de générer d'après un document XML, un nouveau document dans un autre format, notamment pour la génération de pages web.

Z-

Z (courbe en Z), voir courbe de Lebesgue.