

Cours de base de données avancés - ENSG -3A -2017



Introduction – Plan du cours

- ✚ Suite du cours de BDD initiation
- ✚ Système de Gestion de Base de Données
- ✚ Base De Données relationnelles - TD
- ✚ Base de données NoSQL – Applications
- ✚ Mini-projets en équipe



Ressources / Références

- ✚ Livre : Des bases de Données à l'internet (Philippe Mathieu)

disponible: http://www.bbs-consultant.net/IMG/pdf/bdd_mathieu.pdf (1999)

- ✚ Livre : de UML à SQL (Soutou, Editeur: Eyrolles)

- ✚ Quelques cours en ligne :

- ✚ ENST-B, Lyon2, UFR-IMA Grenoble

- ✚ Quelques 'tutorials' BD sur l'internet

- ✚ Attention! Tout ce que vous lisez n'est pas vrai...





Ressources / Références

- ❖ Cours Systèmes de Gestion de Bases de Données.
Jean-Pierre CHEINEY, Philippe PICOUEY, Jean-Marc SAGLIO d'ENST (Ecole Nationale Supérieure des Télécommunications)

<http://www.bd.enst.fr/polyv7/chap0.htm>

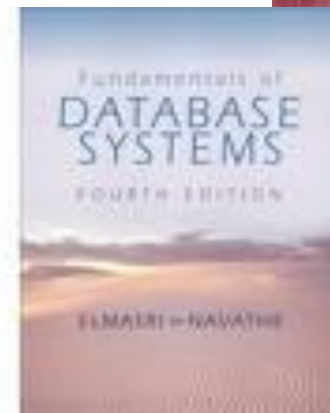
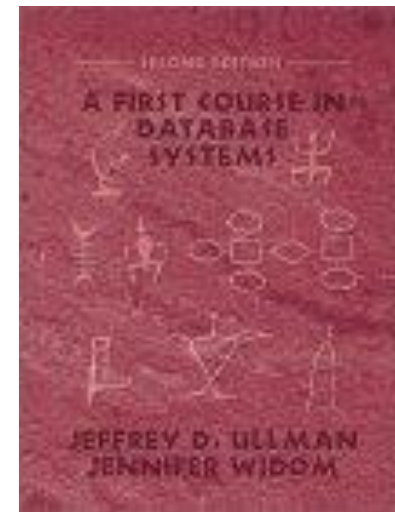
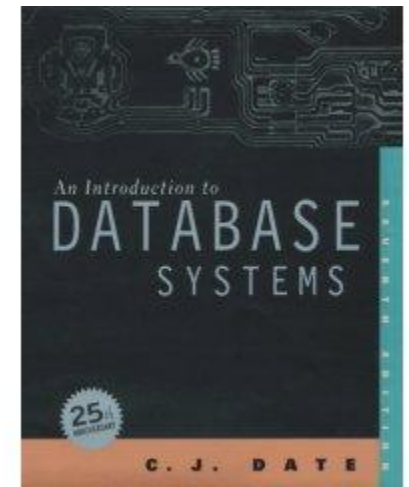
- ❖ MODÉLISATION DES DONNÉES : Approche pour la conception des bases des données – Par Philippe Guézélou

<http://philippe.guezelou.free.fr/mcd/mcd.htm>



Livres classiques

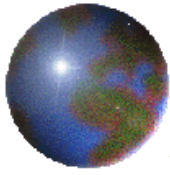
- ✚ *An Introduction to Database Systems*, C.J. DATE, ADDISSON-WESLEY
- ✚ A First Course in Database Systems, par Ullman and Widom, (Prentice-Hall, 1999)
- ✚ *Fundamentals of Database systems* R. ELMASRI et S.B. NAVATHE, ADDISSON-WESLEY, 1994





Sommaire

- ✚ Introduction: Historique
- ✚ Système d'information
- ✚ Le système de gestion de bases de données (SGBD)
- ✚ Objectifs des SGBD
- ✚ Types de base de données
 - ✚ Bases hiérarchiques
 - ✚ Bases réseaux
 - ✚ Bases relationnelles
 - ✚ Bases déductives
 - ✚ Bases objet
- ✚ Conception de Bases de Données relationnelles
 - ✚ Etapes de la conception d'une base de données
 - ✚ Conception de la base de données: le modèle entité-association
 - ✚ Le modèle relationnel
 - ✚ Passage du modèle entité-association à un modèle relationnel



Introduction: Historique



Introduction: Historique

Avant les années 60 développement des SGF
Système de Gestion de Fichiers

Base de données = Fichiers

- ✚ Travailler directement sur un fichier présente plusieurs inconvénients :
- 1. Manipulation de données **lourde et compliquée**. Il faut être expert en programmation
- 2. Le programmeur doit connaître la **localisation physique** des fichiers, la **structure physique** des enregistrements, **le mode d'accès** à ces fichiers
- 3. Toute modification de la structure des enregistrements (ajout d'un champ par exemple) entraîne la **réécriture de tous les programmes** qui manipulent ces fichiers



Introduction: Historique

Durant les années 60 naissance de la première génération de SGBD

- ✚ Séparation de la description des données et des programmes d'application
- ✚ Avènement des langages d'accès navigationnels.
- ✚ Nous parlerons de modèles d'accès ou navigationnels



Introduction: Historique

✿ Deux modèles clés :

1. Groupe **CODASYL** (appelé ensuite DBTG) publia ses premières spécifications de langage pour le **modèle de base de données de réseau** qui est devenu généralement connu comme le Modèle de Données **Codasyl**. Basé sur les idées de Bachman
2. Le **modèle hiérarchique** utilisait l'IMS (Information Management System) d'IBM.

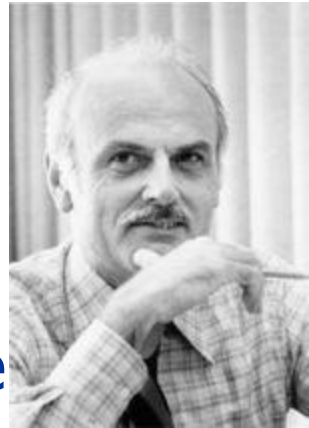




Introduction: Historique

Durant les années 70 : deuxième génération de SGBD

- ✚ **Modèle Relationnel** (Edgar Codd)
- ✚ Codd critiqua les modèles existant parce qu'ils mélangeaient la **description abstraite de la structure de l'information** et les **descriptions des mécanismes physiques d'accès**.
- ✚ **Modèle Entité-Association** (Chen)
- ✚ L'article de Chen est le 35ième article le plus cité en Informatique !





Introduction: Historique

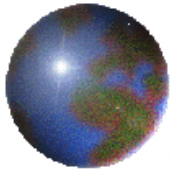
Années 80-90...

- ✚ Troisième génération Modèle Orienté Objet

L'information est représentée sous la forme d'**objets**

- ✚ Base de Données Déductive

Un système de base de données déductive peut faire des **déductions** (c.a.d. déduire des règles ou des faits additionnel) basé sur des **règles** et des **faits** stockés dans la base de données (déductive).



Systeme d'information



Introduction : Système d 'information

- ❖ Généralement, un **système d'information** (SI) est le système des **personnes**, des **enregistrements de données** et des **activités** qui transforment les données et l'information à l'intérieur d'une organisation. Cela inclut les processus **manuels** ou **automatiques**.
- ❖ Terme habituellement utilisé pour signifier : un **système d'information informatique** (concernant le composant informatique)
- ❖ Organisation de l'information pour l'entreprise, au sein de l'entreprise.
 - ❑ Représentation de l'information
 - ❑ Traitement de l'information
 - ❑ Présentation de l'information



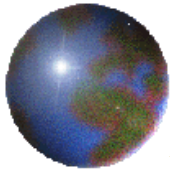


- ✿ Les bases de données font aujourd'hui parties de la vie de tous les jours :
 - ✦ *Si nous allons à la banque to retirer ou déposer de l'argent,*
 - ✦ *Si nous réservons un hôtel ou un vol,*
 - ✦ *Si nous utilisons le système d'une bibliothèque,*
 - ✦ *Si nous achetons des produits dans un supermarché (système de contrôle de stock)*
 - ✦ *Si nous achetons des produits à un vendeur internet sur le web.*
- ✿ Ce sont des **applications traditionnelles des bases de données**



✚ Plus récemment, il y a eu de **nouvelles application des bases de données** :

- ✚ ***Bases de données multimédia*** (stockage de vidéo clips, sons et images)
- ✚ ***Systèmes d'information géographique (SIG –stockage et analyse de cartes, données météo, images satellites)***
- ✚ ***Entrepôt de données ou Datawarehouses, et online analytical processing (OLAP)*** pour extraire et analyser des informations de très grandes bases de données, pour des prises de décision
- ✚ ***Technologie de bases de données temps réel et actives*** (pour contrôler les processus industriels et manufacturiers)
- ✚ ***Techniques de recherche dans les bases de données,*** utilisées pour l'amélioration des recherches internet.



SGBD



Le système de gestion de bases de données (SGBD)

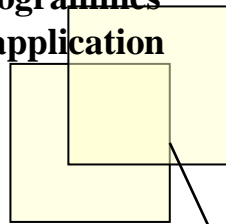
- ✚ Organisation des données, c'est à dire mise en place de l'organisation en vue d'un traitement.
- ✚ Un outil permettant:
 - ✚ insérer, modifier et rechercher efficacement des données spécifiques dans un grand nombre d'informations
- ✚ Une interface entre les utilisateurs et la mémoire secondaire donnant l'illusion que l'information est comme les utilisateurs le souhaitent



SGBD

Composé de trois couches :

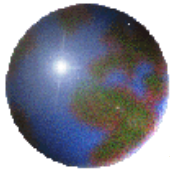
Programmes
d'application



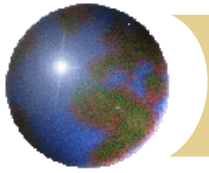
Le système de gestion de fichiers: Il gère le **stockage physique de l'information**. (dépendant du matériel utilisé)

Le SGBD interne: Il s'occupe du **placement et de l'assemblage des données**, gère les liens et l'accès rapide.

Le SGBD externe Il s'occupe de la **présentation et de la manipulation des données** aux concepteurs et utilisateurs (langages de requêtes, outils de présentation)



Objectifs des SGBD



Pourquoi un SGBD?

Quels sont les **objectifs** et **avantages** de l'approche **SGBD** par rapport aux **fichiers** classiques ?



Pourquoi un SGBD ? Objectifs ?

❖ 1. Indépendance physique

- ❖ disques, machine, méthodes d'accès, modes de placement, tri, codage des données...ne sont pas apparents
- ❖ Le SGBD offre une structure canonique permettant la représentation des données réelles sans se soucier de l'aspect matériel

❖ 2. Indépendance logique

- ❖ différentes vues possibles
- ❖ chaque groupe de travail se concentre sur les données qui l'intéressent.
 - Ex.: Enseignants et étudiants, aspect administratif, salles, cours....



Pourquoi un SGBD ?

✚ 3. Manipulation possible par des non-informaticiens

- ✚ Faciliter l'utilisation
- ✚ Obtention de données par des langages (langage déclaratif standardisé et non procédural)

✚ 4. Accès efficace aux données

- ✚ Les accès disque sont lents relativement à l'accès à la mémoire centrale
 - algorithmes efficaces de recherche de données



Pourquoi un SGBD ?

- ❖ 5. Administration centralisée des données
 - ❖ Pour l'administrateur : outils de vérification de cohérence des données, de restructuration, de sauvegarde, de réplication.
 - ❖ L'administration est **centralisée** et est **réservée** a un **très petit groupe** de personnes pour des raisons de **sécurité**.



Pourquoi un SGBD ?

❖ 6. Non-redondance des données

❖ Le SGBD doit permettre d'éviter la duplication d'informations

- la perte de place mémoire
- demande des moyens humains importants pour saisir et maintenir à jour plusieurs fois les mêmes données

❖ EX. si on a les adresses des enseignants dans un système administratif et dans un système de gestion de cours, si l'adresse est changée...



Pourquoi un SGBD ?

7. Cohérence des données

- Obtenue par la vérification des **contraintes d'intégrité**.
- Une contrainte d'intégrité est une contrainte sur les données de la base, qui doit toujours être vérifiée pour assurer la cohérence
- Quelques contraintes :
 - la date d'entrée à l'UPMF doit être < que la date de sortie
 - Un enseignant peut enseigner au maximum 20 heures par semaine ...
 - l'emprunteur d'un livre doit être abonné à la bibliothèque
 - Le salaire doit être compris entre 400 et 10000 Euros.
- Les contraintes doivent être exprimées et gérées **dans la base** et **non** dans les applications



Pourquoi un SGBD ?

✚ 8. Concurrence d'accès aux données

✚ plusieurs personnes (ou applications) en même temps. Problème d'intégrité des données:

- Si deux personnes resservent une salle de cours, pour une même période, en même temps...
- Réservation des places dans le train...



Pourquoi un SGBD ?

✚ 9. Sécurité des données.

▣ Protection d'accès non autorisés ou mal intentionnés. Mécanismes permettant d'autoriser, contrôler et enlever des droits d'accès à certaines informations à n'importe quel usager:

- La personne qu'organise l'emploi du temps des enseignants ne peut pas avoir accès aux salaires.

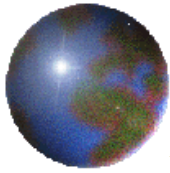
▣ Tolérer les pannes. Capacité de rendre cohérente la base:

- Coupure de courant..
- Une application qui plante...

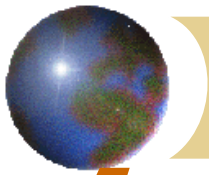


Pourquoi un SGBD ?

- ✚ Ces 9 points sont **difficiles de trouver** dans une **même SGBD**.
- ✚ Il faut choisir selon en donnant priorité à certains critères :
 - ✚ efficacité d'accès
 - ✚ économie de disque
 - ✚ sécurité
 - ✚



Types de base de données



5 grands types de BDD

✚ Bases hiérarchiques :

- ✚ les premiers, avec gestion de pointeurs arborescente

✚ Bases réseaux:

- ✚ les plus rapides, la navigation est aussi entre pointeurs

✚ Bases relationnelles:

- ✚ la plus utilisée. Données représentées sur forme des tables. Basé sur l'algèbre relationnelle et un langage déclaratif (généralement SQL).

✚ Bases déductives:

- ✚ Le langage d'interrogation se base sur le calcul des prédicats et logique 1er ordre

✚ Bases objet

- ✚ Chaque champ est un objet, chaque donnée est active et possède ses propres méthodes d'interrogation et d'affectation. L'héritage est utilisé comme mécanisme de factorisation de la connaissance



Types de base de données

- ✚ Nous rappellerons rapidement les caractéristiques essentielles des **modèles navigationnels**.

C'est-à-dire :

le Modèle hiérarchique,

le Modèle réseau



Types de base de données

- ⊗ Bases hiérarchiques :
 - ⊠ les premiers, avec gestion de pointeurs arborescente
- ⊗ Bases réseaux:
 - ⊠ les plus rapides, la navigation est aussi entre pointeurs, mais plus ouvertes.
- ⊗ Bases relationnelles:
 - ⊠ la plus utilisée. Basé sur l'algèbre relationnelle. Données représentées sur forme des tables
- ⊗ Bases déductives
 - ⊠ langage en calcul de prédicats et logique 1er ordre
- ⊗ Bases objet
 - ⊠ instances de classes hiérarchisées.



le Modèle hiérarchique

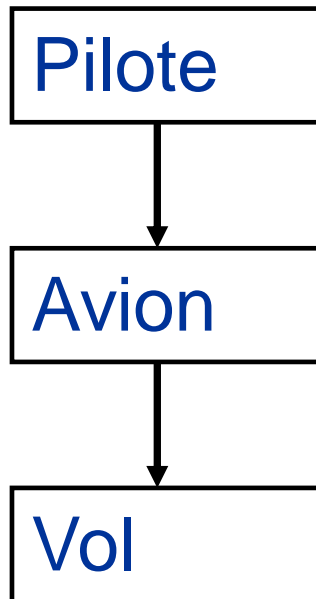
- ✚ Employé par des systèmes très répandus comme IMS (IBM)
- ✚ Dans le modèle hiérarchique, la structure des données est décrite par **deux concepts**:
 1. L'enregistrement logique ou **article** qui regroupe l'ensemble des propriétés ou **champs** constituant les caractéristiques d'une entité
 2. Le **lien orienté** associant les articles selon une arborescence: *un article père est relié à N articles fils*
- ✚ La BD est constituée d'une **collection d'arbres** (forêt)



le Modèle hiérarchique

Arbre TRANSPORT_AERIEN

✚ Exemple:



Article PILOTE Racine

PNOM caractère (20)

ADDRESSE caractère (30)

...

Article AVION Parent = PILOTE

ANOM caractère (10)

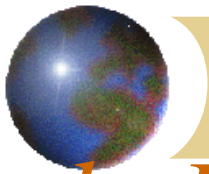
CAPACITE entier

Article VOL Parent = AVION

VNOM caractère (5)

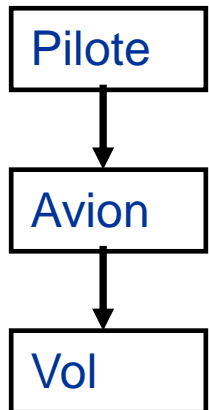
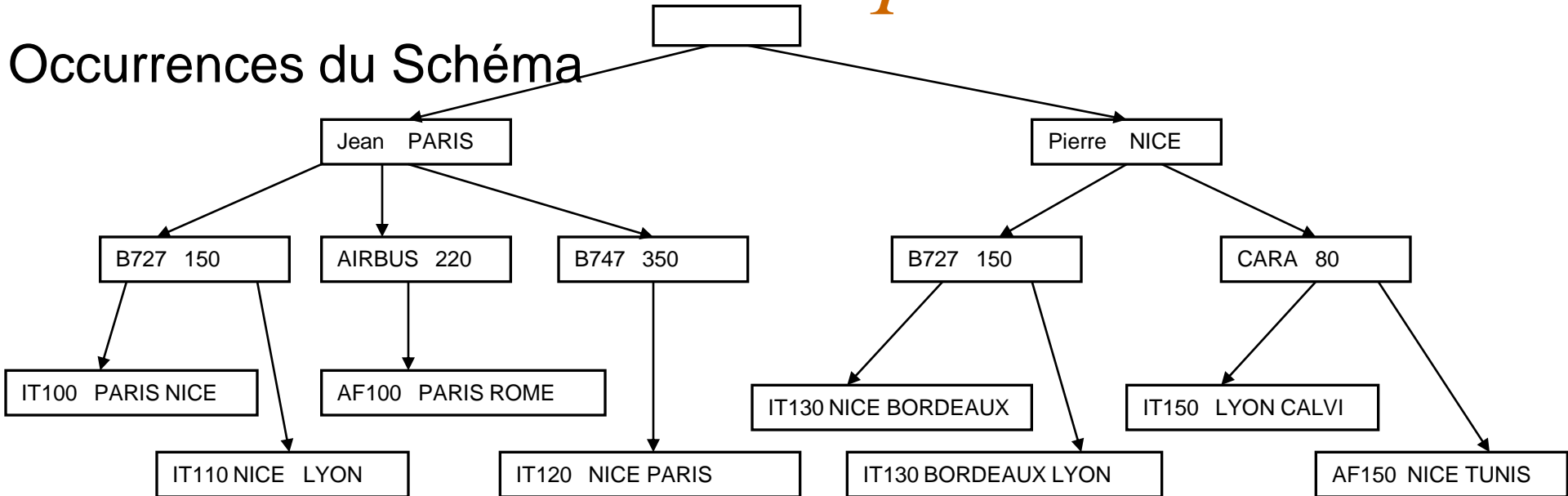
ORIGINE caractère (20)

DESTINATION caractère (20)



le Modèle hiérarchique

Occurrences du Schéma



Arbre TRANSPORT_AERIEN

Article **PILOTE** Parent = Racine
PNOM caractère (20)
ADDRESSE caractère (30)
...

Article **AVION** Parent = PILOTE
ANOM caractère (10)
CAPACITE entier

Article **VOL** Parent = AVION
VNOM caractère (5)
ORIGINE caractère (20)
DESTINATION caractère (20)



le Modèle hiérarchique

- ✚ Le modèle est simple mais présente quelques faiblesses due à la structure arborescente:
 1. Seule une association [1-N] sera naturellement représentée
 2. Il y a duplication d'enregistrements communs à deux arborescences



Types de base de données

- ⊗ Bases hiérarchiques :
 - ⊠ les premiers, avec gestion de pointeurs arborescente
- ⊗ Bases réseaux:
 - ⊠ les plus rapides, la navigation est aussi entre pointeurs, mais plus ouvertes.
- ⊗ Bases relationnelles:
 - ⊠ la plus utilisée. Basé sur l'algèbre relationnelle. Données représentées sur forme des tables
- ⊗ Bases déductives
 - ⊠ langage en calcul de prédicats et logique 1er ordre
- ⊗ Bases objet
 - ⊠ instances de classes hiérarchisées.



le Modèle réseau

- ✚ Proposé par le groupe DBTG (Data Base Task Group): auparavant CODASYL Consortium
- ✚ Son inventeur original était **Charles Bachman**.
- ✚ Idée : permettre une modélisation **plus naturelle** des relations entre entités.



le Modèle réseau

- ✚ Différence entre un modèle hiérarchique et un modèle de réseau :

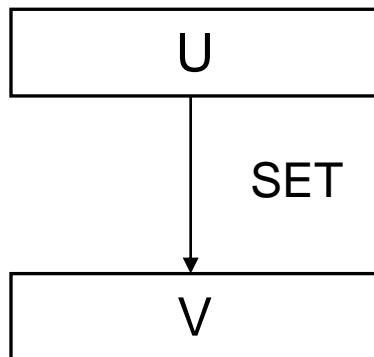
Le modèle hiérarchique structure les données comme un arbre d'articles (chaque enregistrement ayant un article parent et de nombreux enfants)

Le modèle de réseau permet à chaque enregistrement d'avoir de multiples enregistrements parents et enfants, formant une structure en treillis.



le Modèle réseau

- ✚ Les liens entre articles appelés SET matérialise une association entre deux types d'articles distincts.



u est dit **propriétaire**, tout article v relié à u est dit **membre**

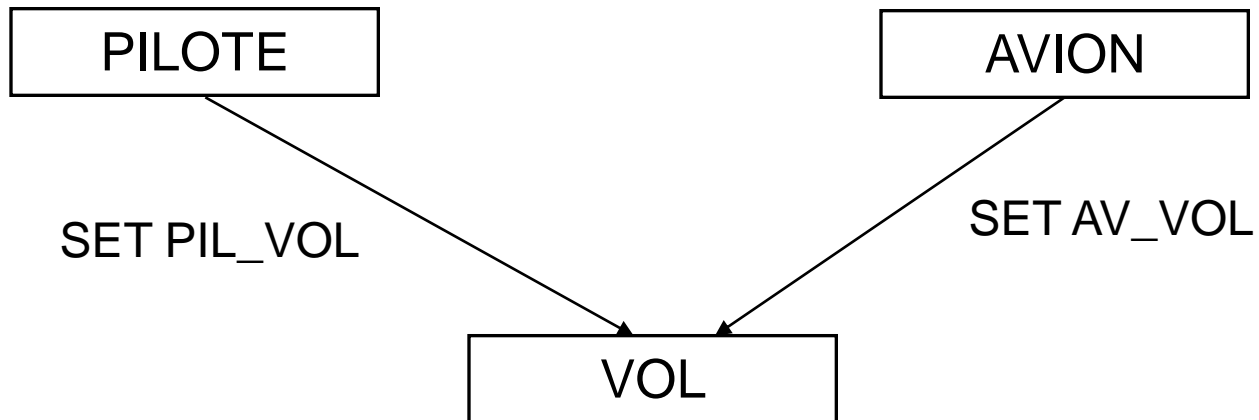
u et l'ensemble des v constitue une occurrence du SET

Un SET est une relation 1-n, mais le type V peut aussi être membre pour une autre relation SET : le graphe des occurrences permet donc des liens n-n entre articles.



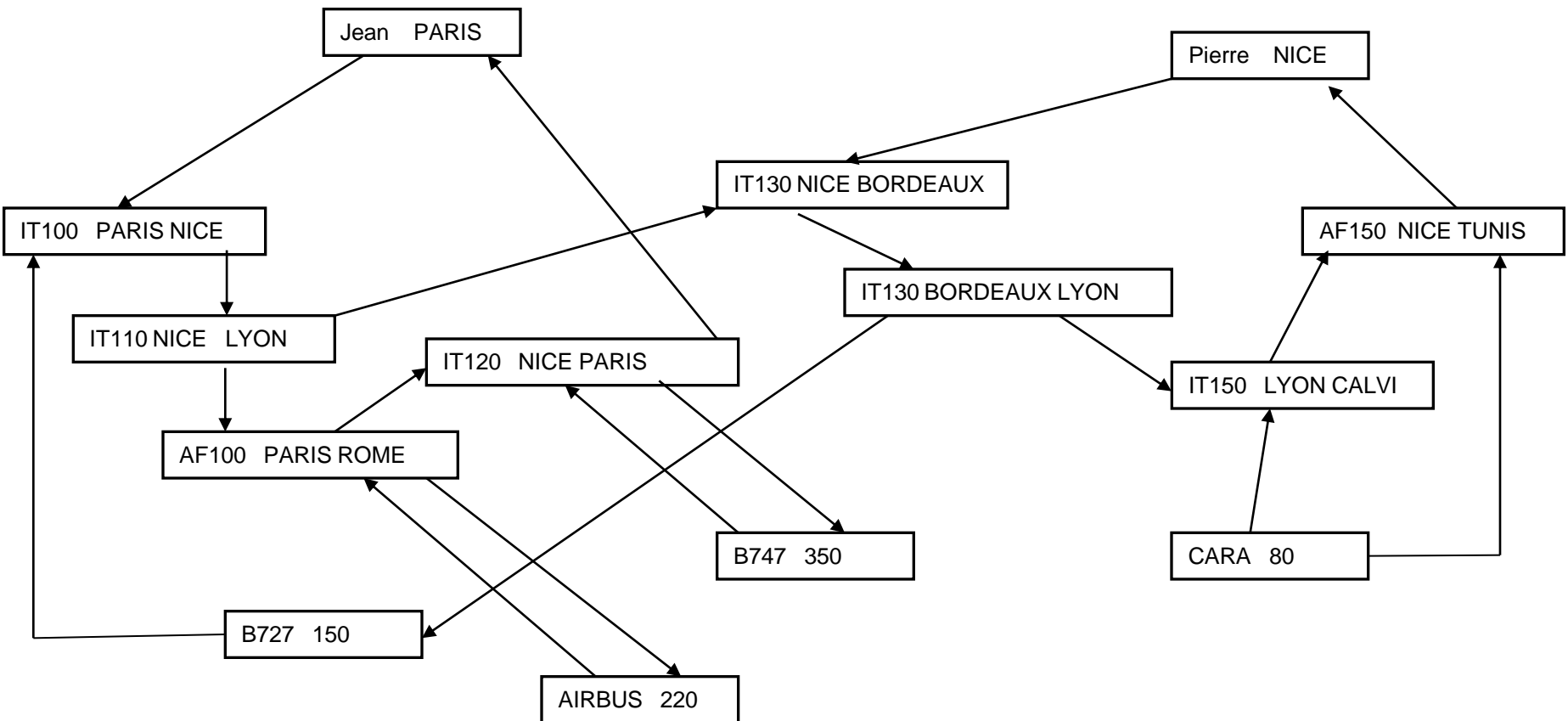
le Modèle réseau

✚ Exemple:





le Modèle réseau



Occurrences du Schéma



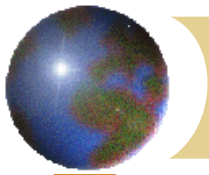
le Modèle réseau

- ⊕ Le modèle était largement implémenté et utilisé.
- ⊕ Mais il a échoué à devenir dominant pour deux raisons principales :
 1. IBM a choisi de **garder le modèle hiérarchique** dans leurs produits établis.
 2. Il fut surpassé par **le modèle relationnel**, qui offrait une interface de plus haut niveau et plus déclarative.



Types de base de données

- ⊗ Bases hiérarchiques :
 - ⊠ les premiers, avec gestion de pointeurs arborescente
- ⊗ Bases réseaux:
 - ⊠ les plus rapides, la navigation est aussi entre pointeurs, mais plus ouvertes.
- ⊗ Bases relationnelles:
 - ⊠ la plus utilisée. Basé sur l'algèbre relationnelle. Données représentées sur forme des tables
- ⊗ Bases déductives
 - ⊠ langage en calcul de prédicats et logique 1er ordre
- ⊗ Bases objet
 - ⊠ instances de classes hiérarchisées.



Types de base de données

✚ Bases relationnelles :

✚ Oracle	www.oracle.com
✚ DB2	www.software.ibm.com
✚ Sybase	www.sybase.com
✚ SQL Server	www.microsoft.com
✚ Ingres	s2k-ftp.CS.Berkeley.EDU/pub/Ingres
✚ Informix	www.informix.com
✚ Acces	www.microsoft.com
✚ Paradox	www.corel.com
✚ Visual Dbase	www.borland.com
✚ FoxPro	www.microsoft.com
✚ File Maker	www.claris.fr
✚ MySQL	www.mysql.net (ou .com)
✚ MSQL	Hughes.com.au
✚ Postgres	www.postgresql.org
✚ InstantDB	www.instantdb.co.uk (100% Java)



Types de base de données

- ⊗ Bases hiérarchiques :
 - ⊠ les premiers, avec gestion de pointeurs arborescente
- ⊗ Bases réseaux:
 - ⊠ les plus rapides, la navigation est aussi entre pointeurs, mais plus ouvertes.
- ⊗ Bases relationnelles:
 - ⊠ la plus utilisée. Basé sur l'algèbre relationnelle. Données représentées sur forme des tables
- ⊗ Bases déductives
 - ⊠ langage en calcul de prédicats et logique 1er ordre
- ⊗ Bases objet
 - ⊠ instances de classes hiérarchisées.



Bases déductives

- ✚ Un système de **base de données déductive** peut générer de nouveaux faits et règles, basés sur des règles et des faits stockés dans la base de donnée (déductive).
- ✚ De tels systèmes :
 - ✚ Gèrent principalement des règles et des faits.
 - ✚ Utilisent un langage déclaratif (prolog ou datalog) pour spécifier ces règles et faits.
 - ✚ Utilisent un moteur d'inférence qui peut déduire de nouveaux faits et règles à partir de ceux données.
- ✚ Aussi connues comme : bases de données logiques, systèmes de connaissances, et bases de données inférentielles.



Bases déductives

- ⊗ Similaires aux systèmes experts.
- ⊗ Différences:
 - ⊠ Les systèmes experts “traditionnels” supposent que tous les faits et règles dont ils ont besoin (*base de connaissance*) seront chargés dans la *mémoire principale*.
 - ⊠ Une base de données déductive utilise une base de donnée *stockée sur disque* comme sa base de connaissance.
 - ⊠ Les systèmes experts traditionnels habituellement prennent leur faits et règles d'un *expert réel*
 - ⊠ Les bases de données déductives puisent leur connaissance *dans les données*.



Types de base de données

- ⊗ Bases hiérarchiques :
 - ⊠ les premiers, avec gestion de pointeurs arborescente
- ⊗ Bases réseaux:
 - ⊠ les plus rapides, la navigation est aussi entre pointeurs, mais plus ouvertes.
- ⊗ Bases relationnelles:
 - ⊠ la plus utilisée. Basé sur l'algèbre relationnelle. Données représentées sur forme des tables
- ⊗ Bases déductives
 - ⊠ langage en calcul de prédicats et logique 1er ordre
- ⊗ Bases objet
 - ⊠ instances de classes hiérarchisées.



Bases objet

- ✚ L'information est représentée sous la forme d'objets
- ✚ Utilisation plutôt rare.
 - ✚ Cependant, elles sont utilisées dans les domaines d'application tels que l'ingénierie (spatiale notamment), les télécommunications, la physique des hautes énergies et la biologie moléculaire, et certains domaines des services financiers.
- ✚ Bases d'objet :
 - ✚ O2 www.o2tech.fr
 - ✚ Gemstone www.gemstone.com
 - ✚ ObjectStore www.objectdesign.com
 - ✚ Jasminecai.com/jasmine



Schémas

- ✚ Dans tout modèle de données (modèle de données objet, modèle relationnel, etc.), il est important de distinguer entre:
 - la description de la base de données
 - et
 - la base elle-même.
- ✚ La description de la base de données est appelée le *schéma de base de données*:
 - ▣ Cela concerne la structure et pas le contenu de l'information.
 - ▣ Spécifié pendant la conception de la base



Trois métiers

✚ L'administrateur de bases de données (ABD or DBA en Anglais)

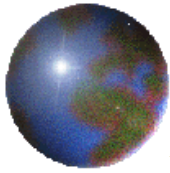
- ✚ Gère les droits d'accès,
- ✚ Coordonne et surveille l'utilisation de la BD
- ✚ Procure le software et hardware si nécessaire
- ✚ L'ABD est responsable pour la sécurité de la BD, la lenteur de réponse du système, etc.

✚ Le concepteur de BD

- ✚ Identifie les données à stocker
- ✚ Choisit la structure appropriée pour la BD
- ✚ Communique avec les utilisateurs finaux pour comprendre les besoins

✚ Les utilisateurs finaux

- ✚ Font des requêtes, mettent la BD à jour, et génèrent des rapports.
- ✚ Plusieurs catégories d'utilisateurs finaux : sophistiqués (analystes, scientifiques); naïfs (employés de banques, réceptionnistes); etc.



Conception de Bases de Données relationnelles



Conception de BDD relationnelles

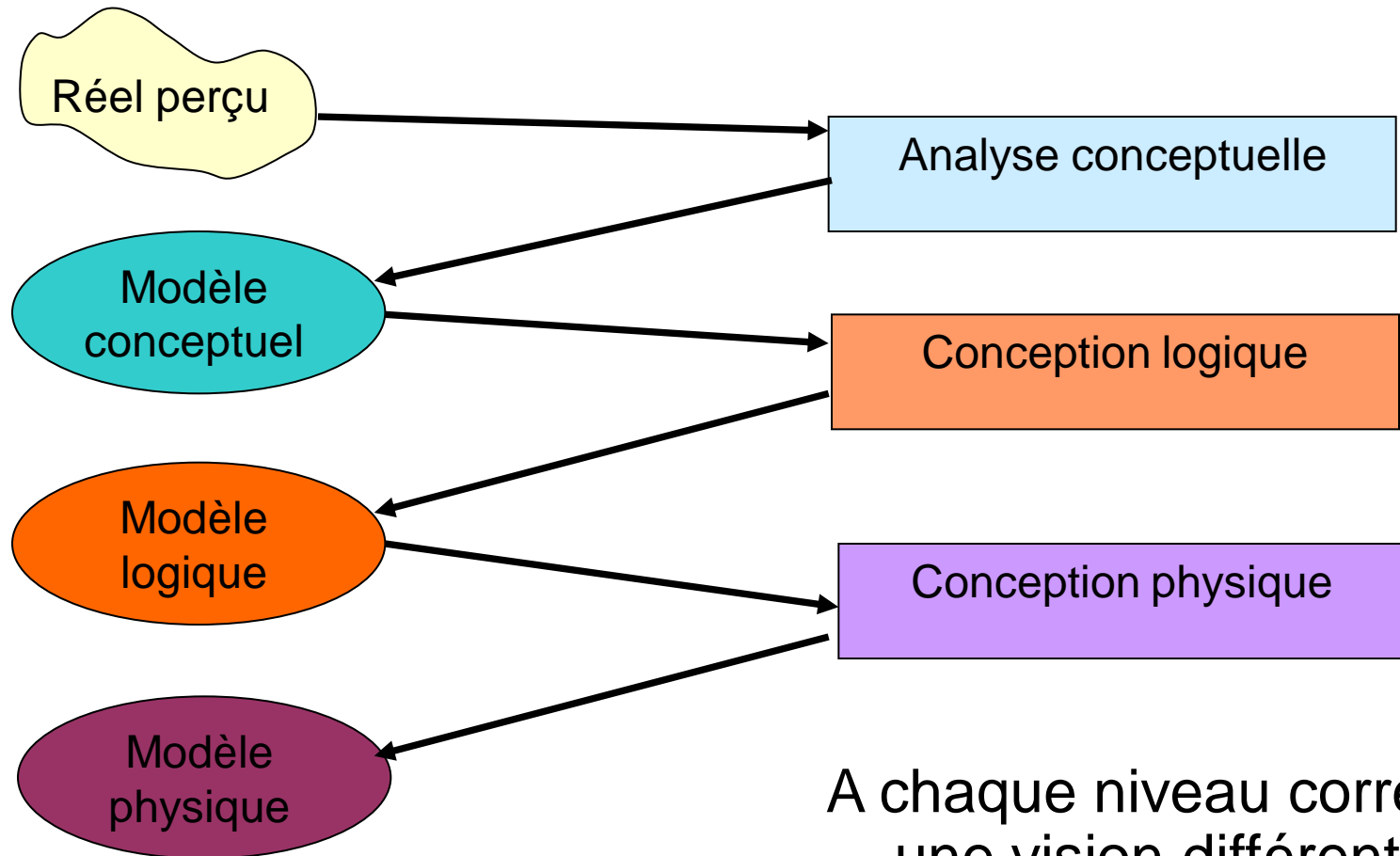
Généralement:

- ✚ Etapes de la conception d'une base de données
- ✚ Conception de la base de données: le modèle entité-association
- ✚ Passage du modèle entité-association à un modèle relationnel
- ✚ Réalisation de la base de données: le modèle relationnel

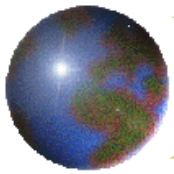


Étapes de la conception d'une BD

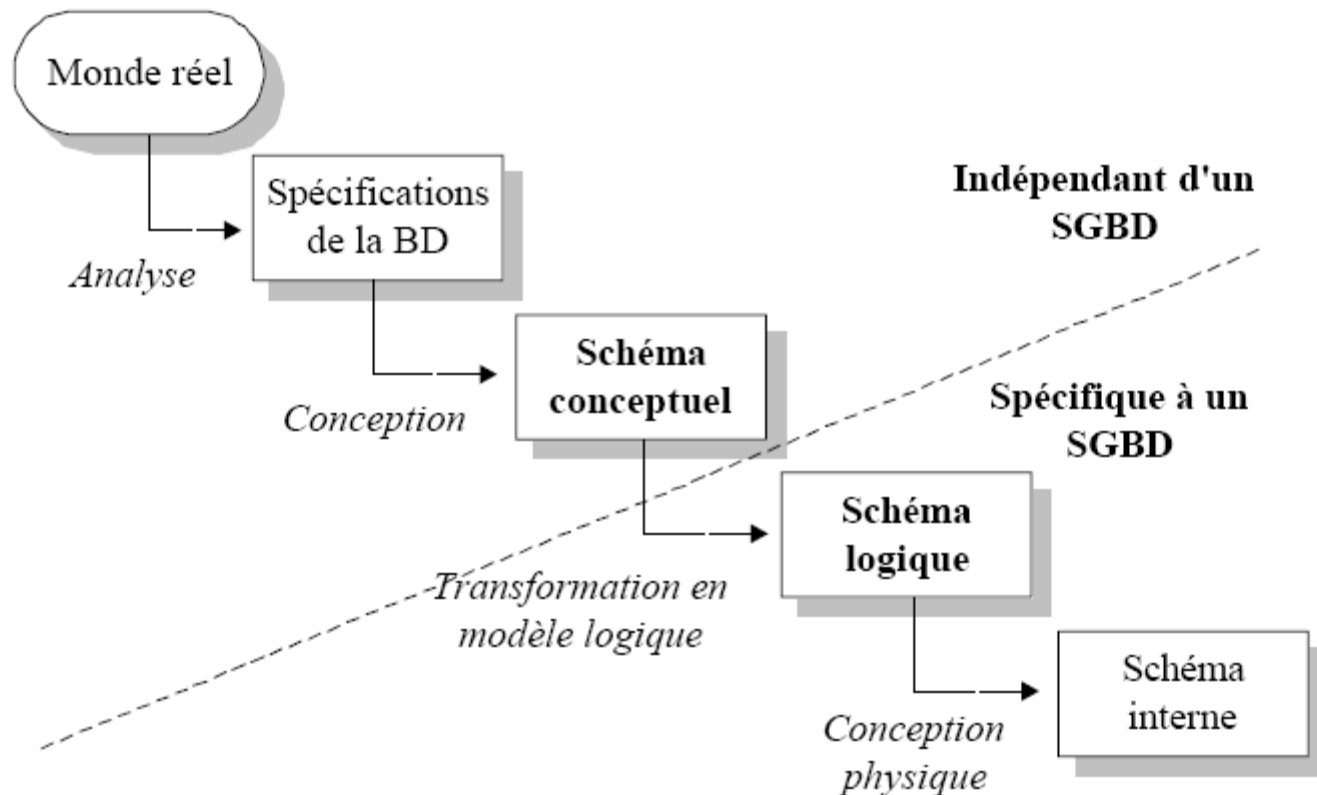
3 étapes:

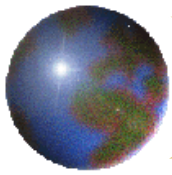


A chaque niveau correspond
une vision différente des
données

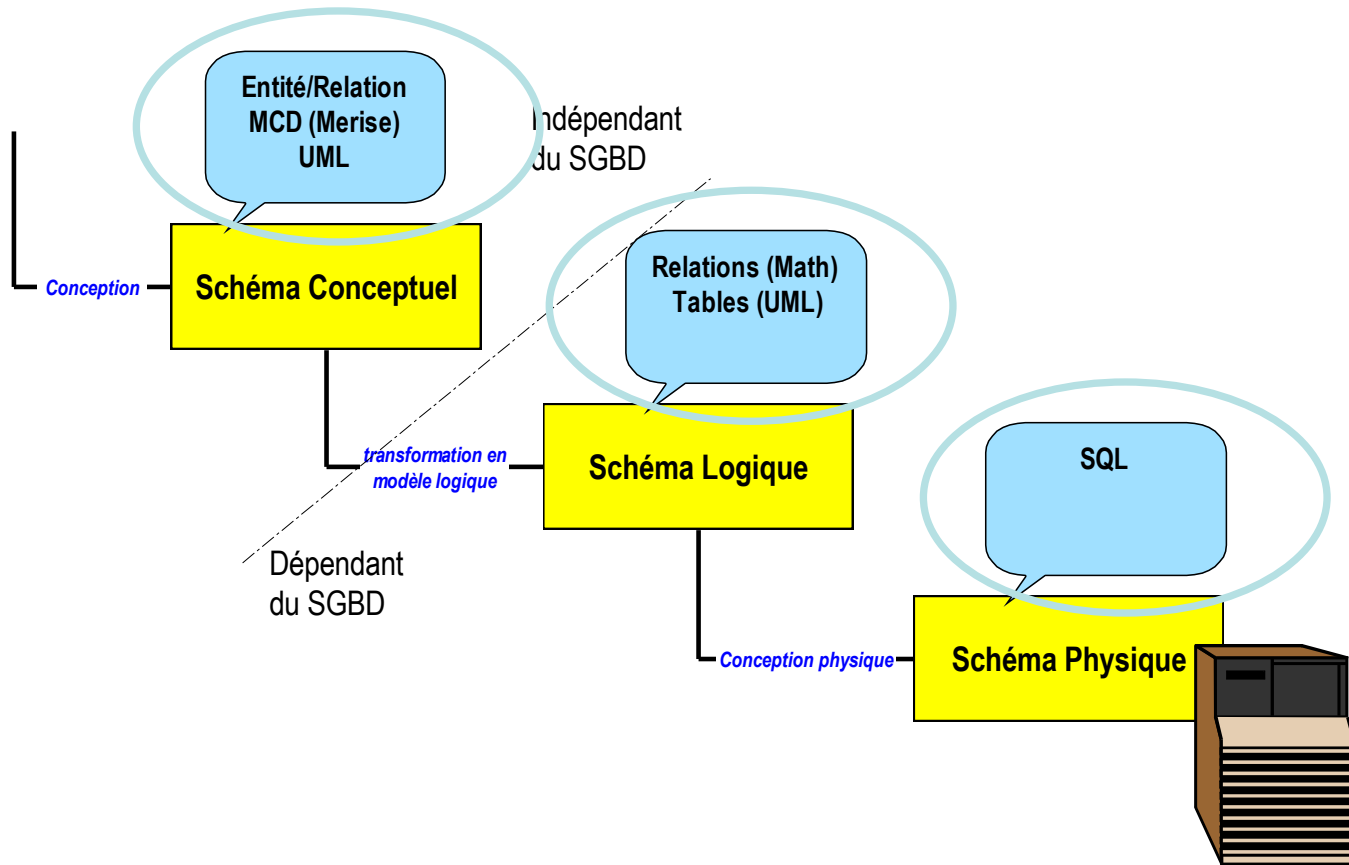


Processus de conception





Processus de conception





MERISE

- ✚ Merise est une **méthode d'analyse**, de **conception** et de **gestion de projet** complètement intégrée
- ✚ Merise: méthode *spécifiquement française*
D'autres: SDM, Axial, OMT, etc.
- ✚ Sépare les données et les traitements à effectuer avec le système d'information en différents modèles conceptuels et physiques.
- ✚ Pour la **conception** d'un BD le plus intéressant est le MCD (**Modèle Conceptuel des Données**).



UML un langage pour

- ✚ Visualiser

- ✚ chaque symbole graphique a une sémantique

- ✚ Spécifier

- ✚ de manière précise et complète, sans ambiguïté,

- ✚ Construire

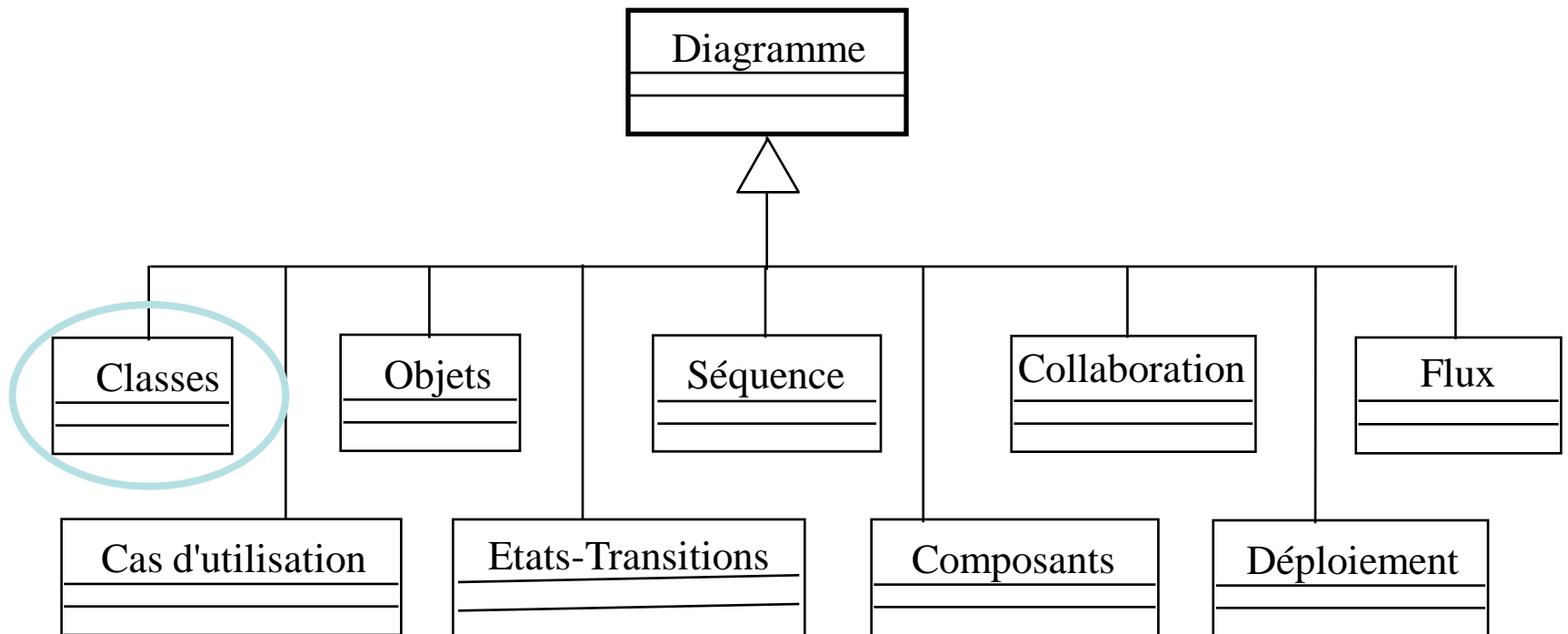
- ✚ les classes, les relations SQL peuvent être générées automatiquement

- ✚ Documenter

- ✚ les différents diagrammes, notes, contraintes, exigences seront présentés dans un document.



UML





Méthodologie

✚ 3 étapes d'évolution du travail :

- 📖 analyse : découverte, interrogation, compréhension,
- 📖 étude/conception : réflexion
- 📖 réalisation : concrétisation

Données → Traitement → Information → Action/Décision



Méthodologie

A chacune des étapes, il faut faire une distinction entre :

- les données (thème du cours)
- les traitements (organisation de l'interface homme/machine non abordée dans ce cours).



Méthodologie

- ✚ Valider le travail lors d'un passage d'une étape à l'autre.
- ✚ Progression linéaire et continue.

Il est plus sage et plus responsable de savoir recommencer l'intégralité du processus face à la rencontre d'une erreur plutôt que de continuer après avoir bricolé une correction.