

Bases de données

Un aperçu du modèle relationnel par l'exemple

TMR_01

v130d

2022-01-16



Christina.Khnaisser@USherbrooke.ca

Luc.Lavoie@USherbrooke.ca

© 2018-2021, Μητίς (<http://info.usherbrooke.ca/llavoie>)
CC BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

Plan

- Un problème : l'évaluation à l'UdeS
 - Analyse
 - Modèle logique de données (prédicats, entêtes, clés)
 - Exemple de données
- Des requêtes
 - R1 à R5
- Un retour sur le modèle de données
 - Améliorations nécessaires et souhaitables
- Synthèse de la démarche
 - Quelle démarche avons-nous suivie et pourquoi ?
 - Vers des règles de pratique ?
- Les colles du prof



Un exemple : l'évaluation à l'UdeS

**Un petit exercice
s'inspirant très
librement des
informations requises
aux fins d'évaluation à
l'université de
Samarcande**

- Analyse
- Modèle logique (prédicats, entêtes, clés)
- Base de données

Description du contexte et énoncé des objectifs

- L'Université de Samarcande (UdeS), fondée en 1927, propose différentes activités pédagogiques dans plusieurs domaines.
- Depuis deux ans, le nombre de personnes étudiantes ne cesse d'augmenter.
- Afin de mieux répondre à leurs besoins, il est nécessaire de pouvoir ajuster la description de l'offre de formation aux parcours effectifs de formation afin d'en faciliter l'évaluation.

Description du problème

- L'UdeS désire :
 - constituer un répertoire des activités proposées ;
 - consigner les inscriptions et les résultats (notes) par étudiant, par activité et par type d'évaluations.
- *La formulation précédente est volontairement inappropriée. Pourrons-nous découvrir en quoi ?*

Solution — Quelles sont les relvars ? (1/2)

○ Activité

- Dès lors que l'UdeS définit une activité, elle peut être offerte.
- L'activité est identifiée par un sigle et caractérisée par un titre.
- *Note : une activité comporte des évaluations.*

○ Étudiant

- Dès lors que l'UdeS admet une personne étudiante, celle-ci peut s'inscrire à des activités, y participer et se présenter aux évaluations de celles-ci.
- L'étudiant est identifié par un matricule et caractérisé par un nom et une adresse (*que nous réduirons, au fin de l'exemple, au seul nom de la ville, de la municipalité, de la commune ou du village*).

Solution — Quelles sont les relvars ? (2/2)

○ Type d'évaluation

- Dès lors que l'UdeS autorise certains types d'évaluation (TE), une activité peut comporter une évaluation de ce type.
- Un TE est identifié par un code et caractérisé par une description.

○ Résultat

- Une note est obtenue par un étudiant dans le cadre d'une activité lors d'une évaluation au cours d'un trimestre.
- Un résultat est identifié par le matricule de l'étudiant, le sigle de l'activité, le trimestre et le code de TE ; il est caractérisé par une note.

Solution — Quels sont les prédicats ?

- L'*activité* de sigle « sigle », décrite par le titre « titre », **est offerte par l'UdeS.**
- L'*étudiant* dont le matricule est « matricule », le nom est « nom » et l'adresse est « adresse », **est inscrit à l'UdeS.**
- Le *type d'évaluation* de code « code », décrit par la description « description », **est autorisé à l'UdeS.**
- Le *résultat* pour l'évaluation « TE » dans le cadre de l'activité « activite » au trimestre « trimestre », décrit par la note « note », **a été obtenu par l'étudiant** dont le matricule est « matricule ».

Solution — Quels sont les prédicats ? (bis)

- L'*activité* ~~de~~ *identifiée par le* sigle « sigle », décrite par le titre « titre », **est offerte par l'UdeS**.
- L'*étudiant* ~~dont~~ *identifié par* le matricule ~~est~~ « matricule », *décrit par* le nom ~~est~~ « nom » et l'adresse ~~est~~ « adresse », **est inscrit à l'UdeS**.
- Le *type d'évaluation* ~~de~~ *identifié par le* code « code », décrit par la description « description », **est autorisé à l'UdeS**.
- Le *résultat* ~~pour~~ *identifié par* l'évaluation « TE » dans le cadre de l'activité « activite » au trimestre « trimestre », décrit par la note « note », **a été obtenu par l'étudiant** dont le matricule est « matricule ».

Faits et propositions

- En substituant les valeurs des attributs d'un tuple aux variables correspondantes du prédicat de la relation, on obtient la proposition (le fait).
- Dans l'exemple précédent, les variables sont les identifiants entre guillemets.

Faits et propositions — un exemple

L'*étudiant* identifié par le matricule « matricule », décrit par le nom « nom » et l'adresse « adresse » **est inscrit à l'UdeS.**

+

matricule	nom	adresse
15110132	Sergeï	Chandler

=

L'*étudiant* identifié par le matricule « 15110132 », décrit par le nom « Sergeï » et l'adresse « Chandler » **est inscrit à l'UdeS.**

MLD (itération 0) : définition des relvars

Activité {sigle : Texte ; titre : Texte}

Étudiant {matricule : Texte ; nom : Texte ; adresse : Texte}

TypeÉvaluation {code : Texte ; description : Texte}

Résultat {matricule : Texte ; TE : Texte ; activité : Texte ;
trimestre : Texte ; note : Entier}

MLD (itération 1) : définition des types

SigleCours { $x \in \text{Texte} \mid x \sim [A-Z]\{3\}[0-9]\{3\}$ }

Activité { sigle : **SigleCours** ; titre : Texte }

Matricule { $x \in \text{Texte} \mid x \sim [0-9]\{8\}$ }

Étudiant { matricule : **Matricule** ; nom : Texte ; adresse : Texte }

TypeEval { $x \in \text{Texte} \mid x \in \{TP, PR, IN, FI\}$ }

TypeÉvaluation { code : **TypeEval** ; description: Texte }

Note { $x \in \text{Entier} \mid 0 \leq x \leq 100$ }

Trimestre { $t \in \text{Texte} \mid t \sim [0-9]\{4\}[1-3]$ }

Résultat { matricule : **Matricule** ; TE : **TypeEval** ; activité : **SigleCours** ;
trimestre: **Trimestre** ; note : **Note** }

MLD (itération 2) : définition des clés

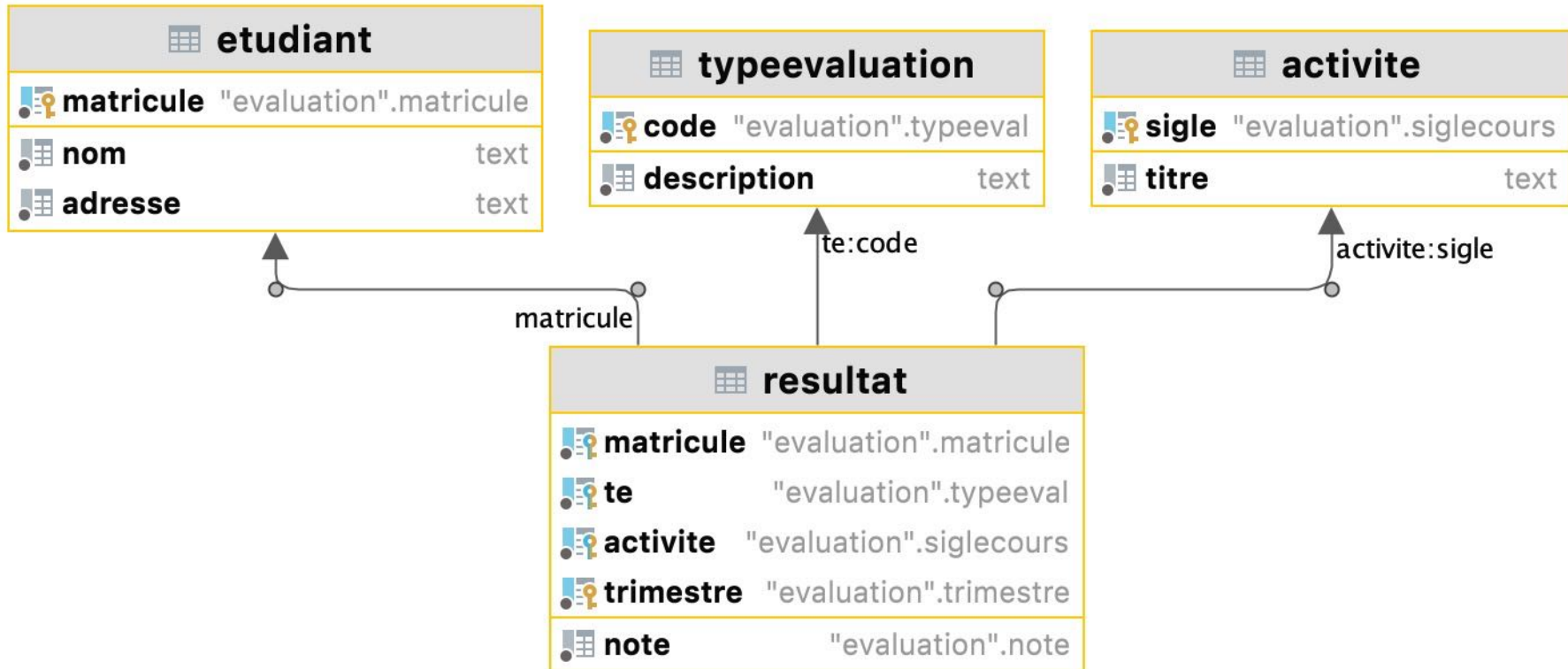
○ Clés candidates

- Clé Activité {sigle}
- Clé TypeEvaluation {code}
- Clé Etudiant {matricule}
- Clé Résultat {matricule, TE, activité, trimestre}

○ Clés référentielles

- Résultat.matricule → Étudiant.matricule
- Résultat.TE → TypeÉvaluation.code
- Résultat.activité → Activité.sigle

Diagramme relationnel de la première solution



Exemple de données

Étudiant

<u>matricule</u>	nom	adresse
15113150	Paul	>Δ ^ρ σ ^ρ ▷ ^b
15112354	Éliane	Blanc-Sablon
15113870	Mohamed	Tadoussac
15110132	Sergeï	Chandler

Activité

<u>sigle</u>	titre
IFT159	Analyse et programmation
IFT187	Éléments de bases de données
IMN117	Acquisition des médias numériques
IGE401	Gestion de projets
GMQ103	Géopositionnement

TypeÉvaluation

<u>code</u>	description
IN	Examen intra
FI	Examen final
TP	Travail pratique
PR	Projet

Résultat

<u>matricule</u>	<u>TE</u>	<u>activité</u>	<u>trimestre</u>	note
15113150	TP	IFT187	20133	80
15112354	FI	IFT187	20123	78
15113150	TP	IFT159	20133	75
15112354	FI	GMQ103	20123	85
15110132	IN	IMN117	20123	90
15110132	IN	IFT187	20133	45
15112354	FI	IFT159	20123	52

Des requêtes

Vérifions notre MLD à l'aide de quelques requêtes représentatives

1. Quels sont les étudiants inscrits en IFT 187 ?
2. Quels sont les étudiants inscrits à une activité en d'informatique à l'automne 2013 ?
3. Quels étaient les étudiants en situation d'échec au final à l'automne 2012 ?
4. Produire le relevé de notes d'Éliane.
5. Quels étudiants ne sont inscrits à aucune activité ?

R1 — Quels sont les étudiants inscrits en IFT 187 ?

R1 — Analyse

○ Clarification

- La formulation utilisée pourrait indiquer qu'on s'intéresse au seul trimestre courant. Le requérant nous précise cependant qu'il vise toutes les inscriptions depuis la première offre de l'activité.

○ Entête

- InscritsIFT187 {matricule : Matricule}

R1 — Solution

○ Clarification

- La formulation utilisée pourrait indiquer qu'on s'intéresse au seul trimestre courant. Le requérant nous précise cependant qu'il vise toutes les inscriptions depuis la première offre de l'activité.

○ Entête

- Inscrits IFT187 {matricule : Matricule}

○ Requête

- (Résultat σ (activité='IFT187')) π {matricule}

R1 — Notes

- La solution précédente suppose qu'un étudiant n'est (considéré) inscrit à une activité qu'à partir du moment où il a été évalué au moins une fois au sein de cette activité.
- Pourquoi ?
- Si l'inscription doit être dissociée de l'évaluation et pouvoir être établie avant la première évaluation, comment devrait-on changer le MLD ?
- Cette dernière question relève de la modélisation, et non de la conception.

R2 — Quels sont les étudiants inscrits à une activité d'informatique à l'automne 2013 ?

R2 — Analyse

○ Clarification

- Une activité d'informatique est définie comme étant toute activité dont le sigle débute par le préfixe 'IFT'. Supposons une fonction $\text{préfixe}(\text{texte}, n)$ qui retourne un texte à partir d'un texte et d'un nombre de lettres.
- Le trimestre « Automne 2013 » est codé par « 20133 ».

○ Entête

- `InscritsIFT {matricule : Matricule}`

R2 — Solution

○ Clarification

- Une activité d'informatique est définie comme étant toute activité dont le sigle débute par le préfixe 'IFT'. Supposons une fonction $\text{préfixe}(\text{texte}, n)$ qui retourne un texte à partir d'un texte et d'un nombre de lettres.
- Le trimestre « Automne 2013 » est codé par « 20133 ».

○ Entête

- $\text{InscritsIFT} \{ \text{matricule} : \text{Matricule} \}$

○ Requête

- $(\text{Résultat } \sigma (\text{préfixe}(\text{activité}, 3) = \text{'IFT'} \wedge \text{trimestre} = \text{'20133'}))$
 $\pi \{ \text{matricule} \}$

R2 — Notes

- Les prédicats sont-ils exacts ?
- Comment distinguer l'admission de l'inscription ?

R2 — Solution : un deuxième essai par la reformulation des prédicats

- L'*activité* identifiée par le sigle « sigle », décrite par le titre « titre », **est offerte par l'UdeS.**
- L'*étudiant* identifié par le matricule « matricule », décrit par le nom est « nom » et l'adresse est « adresse » **est inscrit admis à l'UdeS.**
- Le *type d'évaluation* identifié par le code « code », décrit par la description « description » **est autorisée à l'UdeS.**
- **Résultat :**
 - L'*étudiant* dont le matricule est « matricule » **est inscrit à l'activité « activite » au trimestre « trimestre » à l'UdeS.**
 - Le *résultat* identifiée par l'évaluation « TE » dans le cadre de l'activité « activite » au trimestre « trimestre », décrit par la note « note », **a été obtenu par un étudiant inscrit à l'activité** dont le matricule est « matricule ».

R2 — Revue introspective des prédicats

- Le prédicat de la relvar Résultat est en fait la conjonction de deux prédicats plus simples, l'un (le résultat de l'étudiant) dépendant de l'autre (l'inscription de l'étudiant).
- Nous verrons, lorsque nous traiterons de la normalisation des relations, qu'une telle situation n'est pas souhaitable.

R3 — Quels étaient les étudiants en échec au final à l'automne 2012 ?

R3 — Analyse

○ Clarification

- Une situation d'échec est une note inférieure à 60.
- Un final est un 'Examen final' représenté par le code 'FI'.
- « Automne 2012 » est codé par « 20123 ».

○ Entête

- Échecs20123 {matricule : Matricule}

R3 — Solution

○ Clarification

- Une situation d'échec est une note inférieure à 60.
- Un final est un 'Examen final' représenté par le code 'FI'.
- « Automne 2012 » est codé par « 20123 ».

○ Entête

- Échecs20123 {matricule : Matricule}

○ Requête

- (Résultat σ (note < 60 \wedge TE='FI' \wedge trimestre='20123'))
 π {matricule}

R4 — Produire le relevé de notes d'Éliane

R4 — Analyse : variante 0

○ Clarification

- Nous supposons qu'il n'y a pas d'homonymie.

○ Entête

- RelevéÉliane

{ TE : TypeEval ; activité : SigleCours ;
trimestre : Trimestre ; note : Note }

R4 — Solution : variante 0

○ Clarification

- Nous supposons qu'il n'y a pas d'homonymie.

○ Entête

- RelevéÉliane

{ TE : TypeEval ; activité : SigleCours ;
trimestre : Trimestre ; note : Note }

○ Requête

- ((Résultat \bowtie Étudiant) σ (nom='Éliane')) π
{TE, activité, trimestre, note}

R4 — Analyse : variante 1

○ Clarification

- Le matricule a été introduit pour différencier les homonymes. Un relevé produit sur la seule base du nom est donc susceptible d'être inexact. En conséquence, nous demanderons à Éliane son matricule.

○ Entête

- RelevéÉliane

{ TE : TypeEval ; activité : SigleCours ;
trimestre : Trimestre ; note : Note }

R4 — Solution : variante 1

○ Clarification

- Le matricule a été introduit pour différencier les homonymes. Un relevé produit sur la seule base du nom est donc susceptible d'être inexact. En conséquence, nous demanderons à Éliane son matricule.

○ Entête

- RelevéÉliane
 { TE : TypeEval ; activité : SigleCours ;
 trimestre : Trimestre ; note : Note }

○ Requête

- (Étudiant σ (matricule='15112354')) π {TE, activité, trimestre, note}

R4 — Analyse : variante 2

○ Clarification

- On demande d'ajouter le titre de l'activité dans le relevé.

○ Entête

- Relevé2Éliane

{ TE : TypeEval ; sigle : SigleCours ; titre : Texte ;
trimestre : Trimestre ; note : Note }

R4 — Solution : variante 2

○ Clarification

- On demande d'ajouter le titre de l'activité dans le relevé.

○ Entête

- Relevé2Éliane

{ TE : TypeEval ; sigle : SigleCours ; titre : Texte ;
trimestre : Trimestre ; note : Note }

○ Requête

- Remarquons qu'il est nécessaire de joindre la relation Activité pour obtenir le titre et que l'attribut de jointure n'y porte pas le même nom que dans la relation Résultat.
- (((Étudiant σ (matricule='15112354'))
 ρ {activité \rightarrow sigle})
 \bowtie Activité)
 π {TE, sigle, titre, trimestre, note}

R5 — Quels étudiants ne sont inscrits à aucune activité ?

R5 — Analyse

○ Clarification

- Encore une fois, il est nécessaire de faire préciser la période à couvrir, en termes de trimestres. Supposons que ce soit les trois trimestres de l'année 2013, la question devient donc :
 - Quels étudiants ne sont inscrits à aucune activité **en 2013** ?
- Supposons également qu'on désire avoir un maximum d'information sur ces étudiants et pas seulement leur matricule (à savoir tous les attributs disponibles dans la relation Étudiant).

○ Entête

- NonInscrits2013
{matricule : Matricule ; nom : Texte ; adresse : Texte}

R5 — Analyse (suite)

- Nous réalisons cette requête en calculant d'abord les étudiants **inscrits** à au moins une activité en 2013.
- Inscrits2013 :
(Résultat $\sigma ('20131' \leq \text{trimestre} \leq '20133')$) π {matricule}
- Puis en les **soustrayant** de l'ensemble des étudiants
- NonInscrits2013 :
Étudiant — (Étudiant \bowtie Inscrits2013)

R5 — Solution

Étudiant –

(Étudiant \bowtie

((Résultat σ ('20131' \leq trimestre \leq '20133')) π {matricule}))

ou mieux :

soit

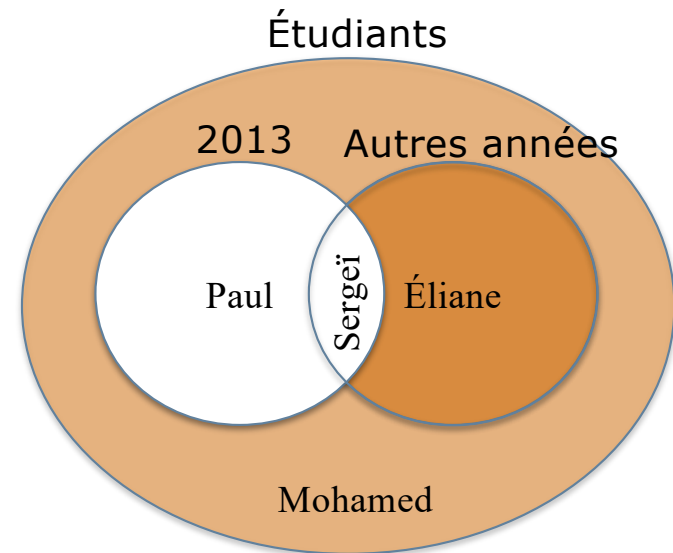
Inscrits2013 := (Résultat σ ('20131' \leq trimestre \leq '20133')) π {matricule} :

Étudiant — (Étudiant \bowtie Inscrits2013)

R5 — Résultats

Étudiant –
(Étudiant ⋈
((Résultat σ ('20131' \leq trimestre \leq '20133')) π {matricule}))

matricule	nom	adresse
15112354	Éliane	Blanc-Sablon
15113870	Mohamed	Tadoussac



R5 — Solution 2

- Justin remarque qu'il semble possible de simplifier l'expression précédente si on admet que la **négation** de la proposition
 $(\text{'20131'} \leq \text{trimestre} \leq \text{'20133'})$
s'exprime par
 $(\text{trimestre} < \text{'20131'} \vee \text{'20133'} < \text{trimestre})$
- Il obtient ainsi
(Étudiant
 $\bowtie (\text{Résultat } \sigma (\text{trimestre} < \text{'20131'} \vee \text{'20133'} < \text{trimestre})))$
 $\pi \{\text{matricule, nom, adresse}\}$

R5 — Solution 2 (revue)

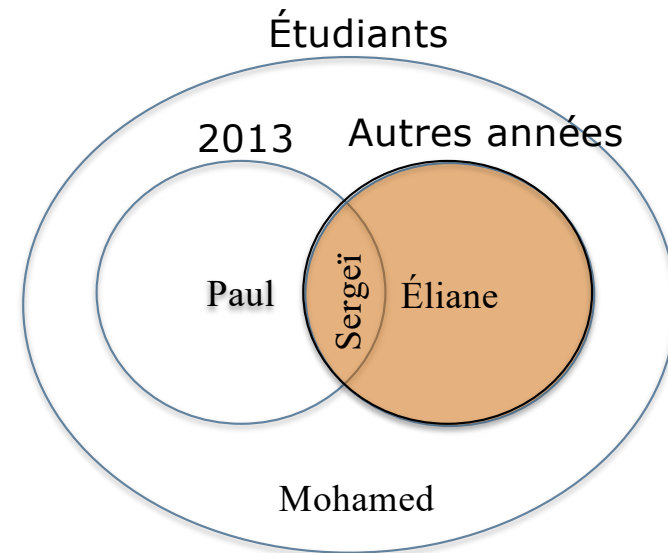
- Justin a-t-il raison ?
 - **NON !**

R5 — Solution 2 (résultats)

(Étudiant

$\bowtie(\text{Résultat } \sigma (\text{trimestre} < '20131' \vee '20133' < \text{trimestre}))$
 $\pi \{\text{matricule, nom, adresse}\}$

matricule	nom	adresse
15112354	Éliane	Blanc-Sablon
15110132	Sergeï	Chandler



R5 — Pourquoi cette incohérence ?

- Parce qu'un étudiant peut s'inscrire lors de plusieurs trimestres, on ne peut pas logiquement conclure qu'un étudiant inscrit en 2013 ne s'est inscrit à aucune autre année :
 - Par exemple, Sergeï s'est inscrit en 2012 et en 2013.
- On ne peut pas logiquement conclure qu'un étudiant qui ne s'est pas inscrit en 2013 s'est inscrit lors d'une autre année :
 - Par exemple, Mohamed est admis, mais il ne s'est pas encore inscrit à une activité.

Un retour sur le modèle de données

Nos requêtes ont permis de diagnostiquer des lacunes et des inadéquations dans le modèle.

Comment pouvons-nous faire mieux ?

- Améliorations nécessaires
- Améliorations souhaitables

Améliorations nécessaires et souhaitables

- La description initiale était la suivante :
 - constituer un répertoire des activités proposées ;
 - consigner les inscriptions et les résultats (notes) par étudiant, par activité et par type d'évaluations.
- *Pourquoi cette description est-elle inappropriée ?*
- *Quelles sont les modifications au MLD qui en découlent ?*
 - *les nécessaires*
 - *les souhaitables*

La démarche

Proposition d'une démarche systématique dans l'élaboration d'une solution

Un avant-gout de règles de pratique

- Élaboration du MLD (entêtes, prédicats, clés)
- Élaboration d'une requête
- Élaboration des cas de test

La démarche (1/3) : Élaboration du MLD

1. Lire l'énoncé.
2. Comprendre les besoins des utilisateurs.
3. Clarifier les concepts du domaine métier.
4. Définir les prédicats.
5. Identifier les contraintes :
 - contraintes de type,
 - contraintes de relation (multi-attributs),
 - contraintes de clés candidates,
 - contraintes référentielles,
 - contraintes générales (multi-relations).

La démarche (2/3) : Élaboration d'une requête

1. Clarifier la question.
2. Expliciter les hypothèses.
3. S'assurer que la question a un sens sur la base des prédicats des relations du MLD.
4. En déduire l'entête de la relation qui sera le résultat et en formuler le prédicat (qui devient alors le libellé de la requête).
5. Élaborer des cas de tests.
6. Diviser la requête pour en traiter plus facilement les parties (diviser pour régner).

La démarche (3/3) : Élaboration des cas de test

○ Les cas

- le cas général : vérifier l'utilisabilité pratique
 - un cas représentatif de ce qui sera le plus souvent requis
 - d'envergure aussi petite que possible, tout en étant réaliste
- les cas particuliers : vérifier la couverture
 - des cas exceptionnels et non compris dans le cas général
- les cas non couverts : vérifier la fiabilité
 - des cas non requis et non couverts
 - ils doivent cependant être signalés ne pas provoquer de dysfonctionnements
- le « gros » cas : vérifier les performances
 - un cas semblable au cas général
 - d'envergure aussi importante que possible, tout en étant réaliste

Règles de pratique

Discussion à propos des règles de pratique

- Pourquoi des règles de pratique ?
- La démarche proposée peut-elle servir de point de départ à l'élaboration de règles de pratique ?

Règles de pratique

- Discussion en classe
- Lien avec les cours d'éthique et de déontologie

Les colles du prof

- Questions de revue
- Connaissez-vous Ramac ?
- Perspectives



Les colles du prof

- Pourquoi l'étape de clarification est-elle essentielle ?
- Pourquoi est-il utile, sinon nécessaire, de fixer l'entête du résultat a priori ?
- Une même requête peut-elle être formulée de différentes façons (par plusieurs expressions différentes) ?
- Comment décririez le rôle de la jointure en regard de la formulation des requêtes ?
- Quel est le lien entre la jointure et la clé référentielle ?

Le disque RAMAC 305 d'IBM Stockage de masse, version 1956

- Capacité
 - 5 millions de caractères de 6 bits, soit environ 3,67 Mo
- Débit moyen
 - un peu moins de 9 ko/s
- Dimensions (cabinet)
 - 1,52 m x 0,74 m x 1,72 m
- Poids
 - environ 1 tonne
- Cout
 - 50 000 USD (1956), soit environ
 - 486 000 USD (2021)
- Production
 - un millier d'unités (1956-1961)
 - fin de vie 1969

https://en.wikipedia.org/wiki/IBM_305_RAMAC

<http://www.bhmag.fr/actualites/nostalgie-premier-disque-dur-ibm-ramac-305-1-tonne-15324>



Perspectives

- Combien coûte un disque domestique de 4 To ?
 - Quel est son débit ?
 - Quelles sont ses dimensions ?
 - Quelle est sa consommation énergétique ?
- Combien coûte un dispositif de stockage haute performance de 4 To ?
 - Quel est son débit ?
 - Quelles sont ses dimensions ?
 - Quelle est sa consommation énergétique ?
- Quels sont les ratios de coût (USD/Mo) de ces dispositifs par rapport au RAMAC 305 ?
- Quels sont les ratios de volume (cm³/Mo) ?
- Quels sont les ratios de consommation d'énergie (watt/Mo) ?
- Combien coûte le stockage d'une photo de 12 mégapixels sur chacun de ces dispositifs (utiliser un format RAW à 24 bits/pixel) ?

