

# Modèle multidimensionnel des données complexes

## Cas des activités d'apprentissage dans un dispositif de formation en ligne

Midouni Djallal\*

Chikh Azzeddine\*\*

\* Département d'informatique, Faculté des sciences de l'ingénieur, Université de Tlemcen  
djmidouni@hotmail.com

\*\* Department of Information Systems, King Saud University, Riyadh, Kingdom of Saudi Arabia  
az\_chikh@ccis.ksu.edu.sa

### Résumé

La vocation d'un entrepôt de données est l'analyse de données pour l'aide à la décision dans les entreprises. La modélisation multidimensionnelle est la base des entrepôts de données et l'analyse en ligne (OLAP). Ces techniques sont efficaces pour traiter les données simples numériques, mais elles ne sont pas adaptées aux données variées et hétérogènes provenant de différentes sources, appelés communément données complexes. Dans cet article, nous abordons le problème de la modélisation multidimensionnelle des données complexes à travers le cas des données technopédagogiques liées aux activités d'apprentissage d'un cours en ligne CIST (Cours sur l'Ingénierie des Systèmes Techniques) pour les étudiants de post-graduation en informatique. Nous proposons un métamodèle multidimensionnel étendu pour ces données technopédagogiques.

**Mots Clés :** Modélisation des entrepôts, modèle multidimensionnel, analyse en ligne, données complexes technopédagogiques.

### 1. Introduction

L'intérêt pour l'analyse de données s'est développé énormément ces dernières années. Les entreprises se sont rendues compte de l'efficacité de la technologie OLAP (*OnLine Analytical Processing*) dans l'analyse et l'exploration des données. Cette technologie est utilisée dans les systèmes d'aide à la décision. Le plus souvent, ces systèmes sont basés sur des techniques d'entrepôt de données pour exploiter la grande masse d'informations disponibles dans les entreprises à des fins d'analyse et d'aide à la décision.

La manière la plus appropriée pour faciliter cette analyse OLAP est la modélisation multidimensionnelle des données. Cette dernière représente les données comme des points dans un espace multidimensionnel [10, 21]. Les données sont vues comme des sujets d'analyse (les faits) étudiés selon plusieurs axes (les dimensions). Chaque dimension est liée à un ou plusieurs points de vues définissant ainsi le degré de granularité des données (hiérarchies).

Contrairement aux modèles relationnels ou orienté-objet, les modèles multidimensionnels sont les plus appropriés pour faire l'analyse et faciliter la prise de décision dans les entreprises. Ils permettent d'observer des faits à travers des indicateurs (mesures) et des dimensions. Autrement dit, le modèle multidimensionnel se compose de **faits** contenant les *mesures* à analyser et de **dimensions** contenant les *paramètres* de l'analyse.

La modélisation multidimensionnelle est donc une technique qui vise à organiser les données de telle sorte que les applications OLAP soient performantes et efficaces. Cependant, cette technique n'est pas adaptée à un certain type de données, dites complexes. La nécessité de gérer et de traiter ce type de données n'a cessé de s'accroître à cause de la variété de ces données, cette variété de données met clairement en évidence la nécessité de créer de nouveaux modèles multidimensionnels.

Les modèles existants, tel que le schéma en étoile, le schéma en constellation ou le schéma en flocon de neige, ont été conçus afin de rendre les données d'un entrepôt prêtes à l'analyse. Ces modèles offrent un cadre agréable pour faire la modélisation multidimensionnelle des données simples, mais ils ne sont pas adaptés aux données complexes. En effet, les données complexes comportent des mesures non additives, non agrégables et qui ont des niveaux de granularités différents. Ce qui rend leur intégration dans des structures multidimensionnelles plus difficile.

Le présent travail vise à apporter des solutions aux problèmes de la modélisation multidimensionnelle de données complexes, en l'occurrence les données technopédagogiques liées aux activités d'apprentissage d'un cours en ligne pour les étudiants de post-graduation en informatique. Ces données sont qualifiées de complexes étant donné qu'elles sont de nature à la fois qualitative et quantitative, en plus du caractère multimédia de certaines d'entre elles. L'entrepôt est formé de plusieurs magasins interconnectés partageant les mêmes données sur les apprenants, les ressources, les enseignants, etc. Chaque magasin contient également d'autres données sur des cours différents

(« aide à la décision », « recherche d'information », « Intelligence artificielle », « Ingénierie des systèmes techniques », etc.).

Notre objectif est de proposer un modèle multidimensionnel pour ces données technopédagogiques, plus particulièrement pour les données du cours d'ingénierie des systèmes techniques, et de généraliser ce modèle vers un métamodèle pour l'entrepôt de données technopédagogiques. Le rôle de cet entrepôt est d'intégrer et de stocker toute information utile à l'équipe pédagogique composée d'enseignants, de tuteurs, de coordonnateurs pédagogiques, d'administrateurs de LMS, etc. et de conserver l'historique des données technopédagogiques pour supporter les analyses pédagogiques effectuées nécessaires aux prises de décision stratégiques.

Outre cette introduction, nous présentons dans la section 2 une définition de données complexes. La section 3 étudie les principaux travaux traitant la modélisation multidimensionnelle des données et plus précisément les données complexes. La section 4 est relative à notre contribution par la proposition du modèle multidimensionnel du module d'ingénierie des systèmes techniques qui sera généralisé par la suite vers un métamodèle permettant de prendre en charge tous les types de données technopédagogiques liées aux activités d'apprentissage d'un cours en ligne pour les étudiants de post-graduation en informatique. La section 5 décrit une implémentation possible de ce métamodèle dans une base de données relationnelle ainsi que la manière de l'instancier pour définir les autres magasins de données. La dernière section conclut ce travail et présente quelques perspectives d'utilisation et de recherche ouvertes par ce méta modèle.

## 2. Données complexes

La description des données complexes nécessite une certaine précision et un espace de représentation adapté. A ce jour, il n'existe pas de modèle universel pour toutes les formes de données complexes. Les données sont qualifiées de complexes si elles sont [6]:

- *multiformats* : l'information est représentée sous différents formats (BD, données numériques, images, sons, vidéos...); et/ou
- *multisttructures* : les données peuvent être structurées, non structurées ou semi-structurées (BD relationnelles, documents XML...); et/ou
- *multisources* : les données proviennent de différents origines (BD réparties, web...); et/ou
- *multimodales* : un même phénomène est décrit par plusieurs canaux ou points de vue (exp: données exprimées dans des échelles ou des langues différentes...); et/ou
- *multiversions* : les données sont évolutives en termes de définition ou de valeur (BD

temporelles, recensements périodiques...).

Nous montrons maintenant pourquoi les données technopédagogiques sont complexes en nous basant sur la définition précitée.

- *multiformats* : Les données relatives aux scores et notes obtenues sont représentées sous forme de BD, les données relatives aux ressources pédagogiques sont multimédia, etc.
- *multisttructures* : Les données sur les profils des apprenants sont considérées comme fortement structurés alors que celles relatives aux scénarios pédagogiques sont semistruées (documents XML).
- *multisources*: Certaines ressources pédagogiques sont développés localement et peuvent par conséquent être récupérés à partir d'un entrepôt indexé avec LOM. Pendant que d'autres ressources peuvent émaner du Web. Les données sur les profils des apprenants peuvent être extraites à partir d'une base de données de la scolarité.
- *multimodales* : les exemples fournis dans un cours peuvent être dans des langues différentes (Anglais, Français, Arabe). Les résultats d'évaluation peuvent utiliser des échelles différentes en fonction de l'enseignant et de l'activité d'évaluation.
- *multiversions* : par exemple les données d'évaluation sont par nature évolutives.

## 3. Modélisation multidimensionnelle des données complexes

Le modèle de données multidimensionnel est le cœur d'un système décisionnel. Toutes les expériences ont montré que la modélisation d'un système décisionnel nécessite des approches spécifiques différentes des approches utilisées dans les systèmes transactionnels. L'une des différences importantes entre les systèmes classiques (systèmes transactionnels) et les systèmes décisionnels (entrepôt de données) est l'organisation des données dans le système, ou plus simplement, le modèle de données. Un modèle dimensionnel contient les mêmes informations qu'un modèle relationnel classique, mais présente les données dans un format plus approprié pour faire l'analyse de données.

De nombreux travaux ont étudié la modélisation multidimensionnelle. Certains proposent des langages algébriques pour faciliter l'interrogation et la manipulation des données de l'entrepôt [4, 2, 20, 16, 13, 14]. Ces modèles peuvent être classés en trois niveaux [7, 3]. Au niveau conceptuel, on trouve des modèles proches de l'utilisateur et indépendants de l'implémentation. Au niveau logique, les modèles dépendent du SGBD utilisé dans l'implémentation, mais restent compréhensibles pour les utilisateurs finaux. Au

niveau physique, les modèles dépendant du SGBD spécifique utilisé sont conçus pour décrire la manière dont les données seront stockées.

Ces différentes propositions sont parfaitement adaptées aux applications de données classiques, mais ne répondent pas complètement aux exigences des applications à base de données complexes telles que les applications pédagogiques. La majorité de ces travaux ne prennent pas en compte les objets à structure complexes. Cependant, Olivier Teste a spécifié des modèles de représentation et des langages de manipulation qui sont dédiés aux entrepôts et magasins de données complexes et évolutives et qui sont basés sur le paradigme objet [20, 19, 15]. Il a intégré par ailleurs dans son modèle la dimension temporelle afin de conserver l'évolution des données de manière pertinente.

Wan et Zeitouni proposent un modèle multidimensionnel, pour un autre type de données complexes (objets mobiles), qui considère le temps et l'espace comme des dimensions importantes dans leur analyse multidimensionnelle [22]. Tanasescu et al. ont conçu un modèle UML générique basé sur un modèle général pour mieux identifier et représenter tous les types des données complexes afin qu'elles soient prêtes au processus de modélisation multidimensionnelle [18, 5].

Les efforts de modélisation des données spatiales, considérées comme un autre type de données complexes, se concentrent sur la représentation arbitraire des objets géométriques (points, lignes, polygones, etc.) dans un espace multidimensionnel [8]. La technologie SOLAP est basée sur une structure multidimensionnelle pour supporter l'analyse spatio-temporelle [17]. Miquel et al. proposent des solutions pour concevoir ces structures lorsque les sources de données sont hétérogènes des points de vue temporel, spatial et sémantique [11]. D'autres auteurs, comme Zghal et al. se sont intéressés aux problèmes de la modélisation multidimensionnelle des données spatiales en se basant sur le développement de l'entrepôt spatial [23].

Dans le domaine de l'éducation et plus particulièrement celui du e-learning, à notre connaissance il n'y a pas de travaux significatifs qui proposent des modèles multidimensionnels de données pédagogiques.

Les travaux existant dans les autres domaines se révèlent inadaptés à notre contexte de travail car ils ne prennent pas en compte le problème de l'hétérogénéité des données technopédagogiques.

#### **4. Modélisation multidimensionnelle des données technopédagogiques**

Ce travail concerne la modélisation multidimensionnelle des données

technopédagogiques liées aux activités d'apprentissage d'un cours en ligne pour les étudiants de post-graduation en informatique. Les modèles existants tels que les modèles en étoile ou en constellation sont inadaptés aux données technopédagogiques. Pour cela, le but de ce travail est de proposer un modèle multidimensionnel qui permettra de traiter et d'analyser ce type de données complexes.

L'entrepôt de données technopédagogiques liées aux activités d'apprentissage d'un cours en ligne pour les étudiants de post-graduation en informatique est organisé sous forme d'une collection de magasins de données (DataMarts). Chaque magasin contient les données spécifiques, à un cours particulier (« aide à la décision », « recherche d'information », « Intelligence artificielle », « Ingénierie des systèmes techniques », etc.). Il est défini par un ensemble de faits et de dimensions partagées avec d'autres magasins de données. Nous nous intéressons à la modélisation multidimensionnelle du cours sur « l'ingénierie des systèmes techniques » que nous tentons ensuite de généraliser aux autres magasins de l'entrepôt. Le choix de ce cours a été motivé par le fait qu'il contient une plus grande variété d'activités et par conséquent il est le plus représentatif.

Notre démarche de modélisation est incrémentale. Nous procédons à construire un nouveau modèle multidimensionnel, **IST**. Ensuite, nous créons un métamodèle qui généralise le modèle IST pour pouvoir modéliser les autres magasins de données. En d'autres termes, l'idée derrière cette démarche est de modéliser le module le plus représentatif dans l'entrepôt pédagogique afin d'extraire les différents concepts qui vont permettre de créer un métamodèle générique pour générer les autres modules de l'entrepôt.

##### **4.1. Le modèle multidimensionnel du magasin du cours CIST**

L'analyse multidimensionnelle du cours d'ingénierie des systèmes techniques (CIST) a permis d'observer deux sujets d'analyses importants extraits de la spécification IMS-LD<sup>1</sup>, (1) « Activity » qui désigne les activités pédagogiques composant le cours et (2) « Outcome » qui désigne les résultats obtenus dans chaque activité. Ils sont étudiés selon plusieurs axes d'analyse :

- « Learner » : désigne les apprenants de la population cible. Chaque apprenant peut être par exemple soit apprenant normal, modérateur, ou secrétaire.

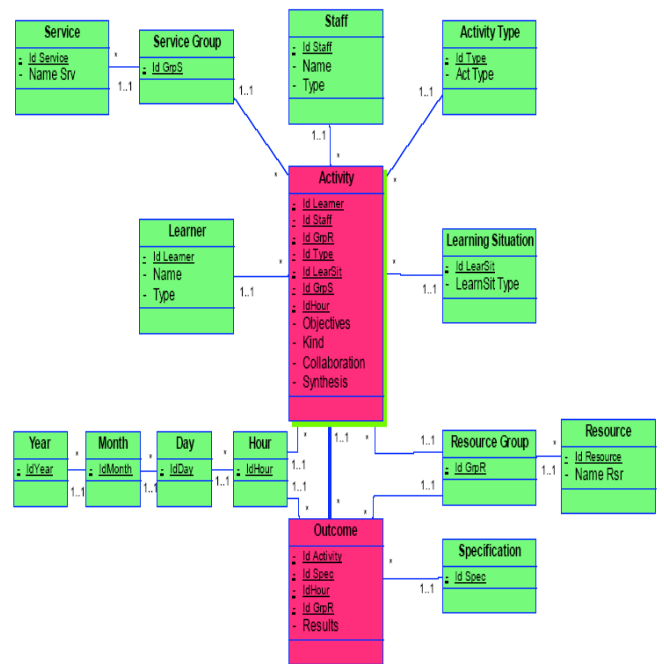
---

<sup>1</sup> IMS-LD : Instructional Management System- Learning Design, [http://www.imsglobal.org/learningdesign/ldv1p0/imsld\\_bestv1p0.html](http://www.imsglobal.org/learningdesign/ldv1p0/imsld_bestv1p0.html)

- « Staff » : désigne les membres de l'équipe pédagogique qui intervient dans le cours d'une manière générale et l'activité en particulier. Chaque membre peut être soit enseignant normal ou tuteur.
- « Activity Type » désigne si l'activité est une activité d'apprentissage ou de support.
- « Learning Situation » désigne les types de situation d'apprentissage utilisés dans l'activité : étude de cas, débat, simulation, par problème, etc.
- « Resource » désigne les ressources pédagogiques multimédia nécessaires à la réalisation des activités ou obtenus comme résultats de ces dernières. Un exemple peut être une séquence vidéo d'une visioconférence, un manuel d'utilisation d'un logiciel, une image satellite, etc.
- « Service » désigne les service Web ou d'autres outils nécessaires à la réalisation des activités tels que l'email, le chat, les forums, le wiki, etc.
- « Time » est une autre dimension commune avec les autres magasins de données, elle sert à stocker la date de l'enregistrement concerné, elle est exprimée par une hiérarchie à quatre niveaux : Hour→Day→Month→Year.
- « Specification » désigne les résultats attendus de l'activité qui seront comparés aux résultats réels obtenus dans l'activité. Ces derniers sont contenus dans la table de fait « Outcome ».

Pour modéliser les données du cours CIST, nous avons utilisé, dans un premier temps, un schéma en étoile. Nous avons mis les mesures relatives aux deux sujets d'analyse dans une seule table de fait, cette table est liée à toutes les dimensions mentionnées ci-dessus. Cependant, ce modèle pose un problème majeur. Nous avons deux mesures qui ne dépendent pas totalement des mêmes dimensions, ce qui nous a amené à proposer un modèle en constellation. Par conséquent, nous avons mis les mesures dans deux tables de faits séparées et entourées chacune de ces dernières par les dimensions appropriées.

Cette solution n'est pas tout à fait adaptée à nos faits « Activity » et « Outcome », qui dépendent l'un de l'autre puisqu'on ne peut pas analyser l'un sans avoir l'autre. Ceci explique nos motivations pour établir le lien entre les deux tables de faits. Les deux mesures ont un degré de granularité différent qui est exprimé par le lien hiérarchique existant entre les tables de faits.



**Figure 1** : Le modèle Multidimensionnel « IST »

La table de fait « Activity » contient quatre mesures « Objectives », « Kind », « Collaboration » et « Synthesis ».

- La mesure « Objectives » définit les objectifs attendus de l'accomplissement de l'activité en terme de compétence. Un exemple d'objectif serait : à la fin de l'activité de lecture et débat, l'apprenant sera en mesure de faire des comparaisons de structures rhétoriques différentes.
- La mesure « Kind » permet de connaître la nature de l'activité : *individuelle* ou *collaborative*.
- La mesure « Collaboration » permet de savoir quel type de collaboration : entre étudiants ; entre étudiants et tuteurs ; entre étudiants et évaluateur ; entre étudiants et enseignant.
- La mesure « Synthesis » permet de résumer les résultats obtenus dans l'activité contenus dans le fait « Outcome ».

La table de fait « Outcome », qui contient la mesure « Results » qui désigne les résultats réels obtenus dans les activités, est liée à la première table de fait « Activity » par un lien hiérarchique.

Dans la première table, on trouve généralement les valeurs globales et génériques, et dans la deuxième on a le détail de ces valeurs.

Pour prendre en compte le lien, qui existe entre la mesure *Synthesis* de la table de faits « Activity » et la mesure *Results* de la table de faits « Outcome »,

nous avons étendu le modèle en constellation classique de telle façon que ce lien soit représenté. Ainsi, nous avons défini un **lien hiérarchique** entre la table de faits « *Activity* » et la table de faits « *Outcome* ».

La table de faits « *Activity* » joue un double rôle dans ce modèle, elle est considérée comme une table de faits par rapport aux dimensions qui sont autour d'elle et elle joue le rôle de dimension par rapport à la table de fait de niveau de granularité plus bas « *Outcome* ».

Passons maintenant aux dimensions du modèle IST. Pendant la modélisation du module du cours CIST, nous avons constaté qu'il y a un autre type de dimension, en plus des dimensions classiques et temporelles qui représentent des axes d'analyse, ce sont les **dimensions multimédia** qui vont contenir tous les ressources nécessaires à la réalisation des activités ou obtenues par ces dernières. On ne peut pas appliquer l'analyse OLAP sur ce type dimension mais elles vont servir comme un axe de vérification et de révision pour l'équipe pédagogique.

#### 4.2. Le métamodèle

Afin de modéliser l'entrepôt de données technopédagogiques, c'est à dire la collection de magasins de données, nous proposons un métamodèle orienté objet (figure 2) permettant de créer les différents magasins de données de l'entrepôt. Ce métamodèle est une généralisation du modèle IST décrit auparavant avec la prise en compte des faits complexes et les nouveaux concepts définis lors de la modélisation multidimensionnelle du magasin CIST.

Peu de travaux proposent des métamodèles multidimensionnels, l'objectif de ces travaux est de spécifier des métamodèles pour représenter les bases de données multidimensionnelles. Dans [23] les auteurs ont spécifié un métamodèle pour la construction d'un entrepôt de données spatiales. Aussi un autre métamodèle multidimensionnel proposé par Abelló [1], il s'est basé sur le langage UML pour donner un plus de sémantique en profitant des concepts objets tel que les relations de généralisation et de composition. Cependant ces travaux ne sont pas suffisants et ne sont pas adaptés pour représenter les concepts multidimensionnels comme nous souhaitons le faire.

Notamment, le standard de l'OMG (Object Management Group), CWM (Common Warehouse MetaModel) qui propose un ensemble de métamodèles pour les techniques d'entrepôt de données [12]. Cet ensemble CWM est assez complet pour modéliser un entrepôt de données dans son ensemble. Mais le métamodèle multidimensionnel proposé par CWM représente les aspects multidimensionnels d'une façon

générale, il ne prend pas en compte de tous les objets d'une base de données multidimensionnelle. Il faut le combiner avec d'autres métamodèles du même standard pour avoir une représentation plus complète. En plus ce métamodèle ne permet pas de spécifier et de représenter les nouveaux concepts multidimensionnels que nous avons proposé.

Notre métamodèle présenté dans la suite constitue une extension de ces trois derniers travaux qui soit applicable aux données technopédagogiques. La figure 2 montre une représentation UML du métamodèle, ce qui va nous permettre de mieux représenter les concepts multidimensionnels génériques (Dimensions, Faits, Mesures) et les autres concepts extraits de notre étude du magasin CIST (tables de faits multiples et hiérarchisées, dimensions multimédia).

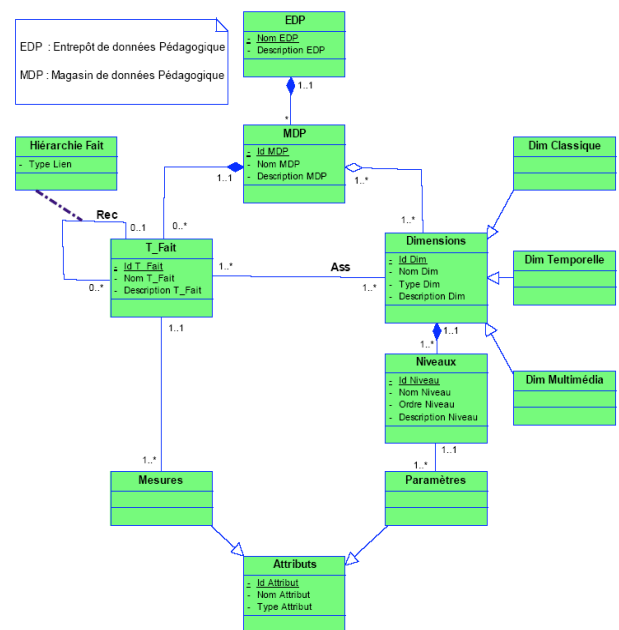


Figure 2 : Le MétaModèle

L'instanciation de ce métamodèle va nous permettre de créer les différents magasins de données dans l'entrepôt pédagogique. Ce dernier, qui est représenté par la classe « *EDP* », est composé d'un ensemble de magasins de données représenté par la classe « *MDP* ». Chaque magasin (MDP) est caractérisé par un ensemble de faits, qui représentent les sujets d'analyses, et un ensemble de dimensions, qui représentent les axes d'analyse. A chaque fait correspond une ou plusieurs mesures et à chaque dimension correspond un ensemble de paramètres. Ces deux derniers concepts (« *Mesures* » et « *Paramètres* ») héritent de la même classe « *Attribut* » mais ils ont une sémantique différente dans les bases de données multidimensionnelles. Les faits complexes sont caractérisés par la relation récursive **Rec** qui permet d'associer à chaque table de fait une autre table de

faits, cette relation permet d'exprimer l'hierarchie des tables de faits.

L'hierarchie de dimensions est matérialisée par les deux classes « *Dimension* » et « *Niveaux* », on associe à chaque niveau son ordre hiérarchique dans la dimension. Par exemple, l'ordre du niveau *Heure* de la dimension *Temps* est égal à zéro et l'ordre de *Jour* égal à un (figure 1).

Nous définissons les principales classes de notre métamodèle comme suit :

- **Entrepôt de données**

Notre entrepôt de données est matérialisé par la classe EDP (Entrepôt de Données Pédagogiques), cette classe est le conteneur global de toutes les classes du métamodèle. La classe EDP est définie par  $(N^{EDP}, Des^{EDP})$  où

- $N^{EDP}$  est le nom de l'entrepôt,
- $Des^{EDP}$  est la description de l'entrepôt.

- **Magasin de données**

C'est un entrepôt de données spécialisé, destiné à ne contenir que les informations élaborées pour un objectif particulier. Par exemple le magasin CIST contient seulement les données du cours CIST. Le magasin de données est représenté par la classe MDP (Magasin de Données Pédagogiques), qui est définie par le quadruplet  $(N^{MDP}, F^{MDP}, D^{MDP}, Ass)$  où

- $N^{MDP}$  est le nom du magasin de données,
- $F^{MDP} = \{F1, F2, \dots\}$  est l'ensemble des faits,
- $D^{MDP} = \{D1, D2, \dots\}$  est l'ensemble des dimensions,
- **Ass** est la fonction qui va associer chaque fait avec ses dimensions.

- **La table de faits**

Une table de fait est la table centrale du modèle multidimensionnel. Elle contient les différentes mesures de l'activité à analyser, ces mesures peuvent être observées selon différentes dimensions. Cette classe est définie par le triplet  $(N^F, M^F, Rec)$  où

- $N^F$  est le nom du fait,
- $M^F = \{m1, m2, \dots\}$  est l'ensemble des mesures,
- **Rec** est la fonction récursive qui va associer un fait avec un autre fait pour permettre l'hierarchisation des faits.

- **La dimension**

Une dimension est un axe d'analyse au sein d'une structure multidimensionnelle. Elle est composée d'une liste ordonnée de paramètres (attributs) qui partagent une signification sémantique commune dans le domaine modélisé. Elle est définie par le quadruplet  $(N^{Dim}, P^{Dim}, H^{Dim}, T^{Dim})$  où

- $N^{Dim}$  est le nom de la dimension,
- $P^{Dim} = \{p1, p2, \dots\}$  est l'ensemble des paramètres,
- $H^{Dim} = \{H1, H2, \dots\}$  est l'ensemble des niveaux formant les hiérarchies de cette dimension,
- $T^{Dim} = \{\text{Classique, Temporelle, Multimédia}\}$  est le type de la dimension.

Nous distinguons trois types de dimensions dans notre modèle : les dimensions classiques, les dimensions temporelles et les dimensions multimédia. La dimension **classique** sert à enregistrer les valeurs pour lesquelles sont analysées les mesures de l'activité. La dimension **temporelle** s'ajoute à l'entrepôt pour maintenir l'historique de l'évolution des données pédagogiques dans le temps. La dimension **multimédia** contient les différents types de données multimédias contenues dans notre entrepôt pédagogique. Les données de ce type de dimension sont difficiles à manipuler par les outils d'analyses actuels, leur but dans notre entrepôt pédagogique est l'archivage de ces données multimédia pour faire la vérification et le contrôle des résultats d'analyses.

## 5. Prototype et Evaluation

Afin de valider notre métamodèle (décrit dans la section 4.2), nous avons développé un prototype d'aide à la conception et à la modélisation de notre entrepôt, intitulé **GEDP (Générateur d'Entrepôt de Données Pédagogique)**.

Notre outil facilite la tâche de l'administrateur de l'entrepôt pour créer et générer des magasins de données à fin de construire l'entrepôt global, tout en respectant nos nouveaux concepts définis dans le métamodèle des données pédagogiques. En effet, l'élaboration de l'entrepôt pédagogique suit un processus de développement à trois niveaux : conceptuel, logique et physique.

La figure 3 décrit ce processus de modélisation, la génération des magasins de données passe généralement par ces trois étapes :

- premièrement, on génère une instance du métamodèle (figure 2), cette instance représente le modèle multidimensionnel du magasin de données en cours de modélisation,
- deuxièmement, on fait la transformation de ce modèle soit vers un fichier XML ou vers une base de données relationnelle,
- ce dernier choix va nous permettre dans la troisième étape de choisir soit un entrepôt XML ou un entrepôt relationnel.

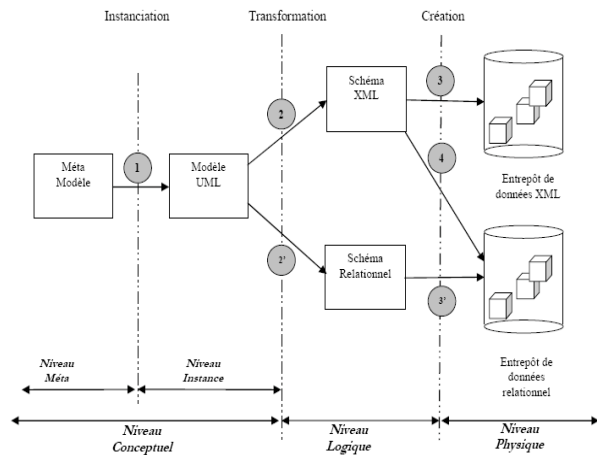


Figure 3 : Le processus de modélisation

Nous avons opté pour ce travail la solution relationnelle. En effet, GEDP est un prototype implanté au-dessus du SGBD Oracle version 10g. Le choix d'un SGBD relationnel est motivé par la grande capacité de stockage ainsi la performance lors de la manipulation des données. En effet, les systèmes de gestion de bases de données relationnelles offrent d'excellentes performances en terme de rapidité d'accès, de volume de stockage et de stabilité des données.

GEDP se base sur une approche incrémentale. L'administrateur élabore l'entrepôt étape par étape en construisant les différents magasins de données de l'entrepôt pédagogique. L'architecture de ce prototype, comme le montre la figure 4, est composée essentiellement d'une interface utilisateur et un générateur de script.

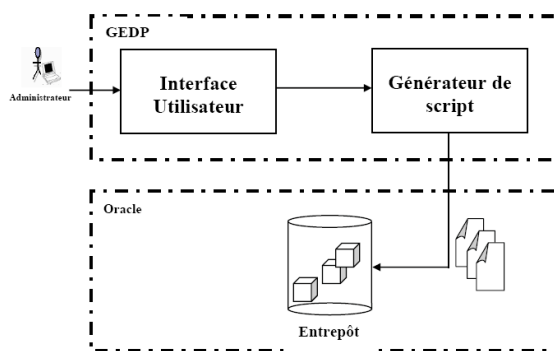


Figure 4 : Architecture du prototype GEDP

L'interface utilisateur permet de définir les différents magasins de données en instanciant notre métamodèle par l'introduction des différents éléments (dimensions, faits,...) du schéma dimensionnel. Le générateur de script est le module responsable de la génération des scripts, ces derniers scripts permettent la création du schéma de l'entrepôt de données technopédagogiques dans une base de données relationnelle, en s'appuyant sur notre métamodèle défini dans la section 4.2.

## 6. Conclusion et perspectives

Le travail présenté dans ce mémoire traite la modélisation multidimensionnelle des données complexes. Notre objectif est d'intégrer les données technopédagogiques liées aux activités d'apprentissage d'un cours en ligne pour les étudiants de post-graduation en informatique dans une structure multidimensionnelle pour apporter une aide au processus décisionnel. Pour répondre à cet objectif, nous avons proposé une approche de modélisation et d'implémentation de l'entrepôt pédagogique en se basant sur un métamodèle que nous avons conçu et développé.

Dans un premier temps, nous avons modélisé le magasin de données le plus complet et le plus représentatif de l'entrepôt pédagogique, le magasin du cours CIST. Pendant cette modélisation, nous avons constaté la difficulté de modéliser et d'intégrer les données technopédagogiques dans une structure multidimensionnelle. Par conséquent, nous avons senti le besoin de proposer de nouveaux concepts qui étendent les modèles existants vers un nouveau type de modèle.

Dans un second temps, nous avons proposé un métamodèle en généralisant le modèle multidimensionnel du cours CIST. L'apport de notre métamodèle se situe au niveau de la prise en compte des tables de faits multiples et hiérarchisées et les dimensions multimédia. L'instanciation de ce métamodèle permet de spécifier et de définir les différents magasins de données de l'entrepôt pédagogique indépendamment des plate-formes techniques.

En fin, nous avons développé ce métamodèle en implémentant un prototype GEDP, acronyme de **G**énérateur d'**E**ntrepôt de **D**onnées **P**édagogiques. Il comporte une interface utilisateur et un module générateur de script permettant de créer automatiquement les différents composants de l'entrepôt de données.

Les perspectives que nous envisageons de conduire sont les suivantes :

- Notre approche étant incrémentale, à partir des retours d'usage nous essayons de faire évoluer le prototype que nous avons réalisé afin de lui permettre une meilleure manipulation de tous les éléments de notre entrepôt.
- La généralisation progressive du méta modèle défini en ajoutant des nouveaux concepts afin de lui permettre de prendre en compte d'autres types de données complexes (par exemple, définir des nouveaux types de mesures et d'autres types de dimensions). Ainsi l'objectif sera l'élaboration des nouveaux modèles de plus haut niveau d'abstraction.
- La définition d'une méthodologie de conception et de construction pour les

entrepôts de données pédagogiques. A l'heure actuelle, il existe des méthodes de conception des entrepôts de données mais elles ne sont pas adaptées aux données complexes technopédagogiques. Nous essayons de faire une extension des méthodes existantes afin d'offrir un cadre complet pour concevoir des systèmes décisionnels comportant des données complexes.

- A travers ce travail, nous pensons définir de nouveaux opérateurs OLAP : il est nécessaire de prévoir l'extension des opérateurs OLAP existants pour prendre en compte les nouveaux concepts définis dans ce travail (les faits multiples hiérarchisés et les dimensions multimédia). Une voie possible, pour les faits multiples, est de s'inspirer des opérateurs OLAP traitant les hiérarchies de dimensions pour définir des nouveaux opérateurs permettant le traitement des faits hiérarchisés.

## Références

- [1] Abelló A., YAM<sup>2</sup>: A Multidimensional Conceptual Model, PhD Thesis, Politècnica de Catalunya. Barcelona, April 2002.
- [2] Agrawal R., Gupta A., Sarawagi S., Modeling Multidimensional Databases, Research Report, SanJose (California), 1995.
- [3] Batini C., Ceri S., Navethe S.B., Conceptual Database Design: An Entity-Relationship Approach, Benjamin-Cummings Publishing, 1992.
- [4] Cabibbo L., Torlone R., A Logical Approach to Multidimensional Databases. EDBT 1998:183-197.
- [5] Darmont J., Boussaid O., Bentayeb F., Rabaseda S., Zellouf Y., Web multiform data structuring for warehousing, In C. Djeraba.; Multimedia Systems and Applications, Vol. 22, Kluwer, 2002, 179-194.
- [6] Darmont J., Boussaid O., Ralaivao J., Aouiche K., An Architecture Framework for Complex Data Warehouses, 7th International Conference on Enterprise Information Systems, Miami, USA, May 2005.
- [7] Elmasri R., Navethe S.B., Fundamentals of database systems, Benjamin-Cummings Publishing. 3ième edition, 2000.
- [8] Gutting R. H., An introduction to spatial database systems, VLDB Journal, 1994.
- [9] Jensen M., Moller T., Pedersen TB., Specifying OLAP cubes on XML data, Journal Of Intelligent Information Systems, 17(2/3):255--280, 2001.
- [10] Kimball R., The Data Warehouse Toolkit: Practical techniques for building dimensional data warehouses. John Wiley. 1996.
- [11] Miquel M., Bédard Y., Brisebois A., Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, ISI-NIS, volume X, 2001.
- [12] OMG, Common Warehouse Metamodel (CWM) Specification, March 2003, Vers 1.1.
- [13] Pedersen TB., Jensen CS., Multidimensional Data Modeling for Complex Data, In Proceedings of ICDE, pp. 336--345, 1999.
- [14] Pokorny J., Sokolowsky P., A Conceptual Modeling Perspective for Data Warehouses, Electronic Business Engineering 1999.
- [15] Ravat F., Teste O., Zurfluh G., Modélisation et extraction de données pour un entrepôt objet, BDA'2000, Oct 2000, Blois (France).
- [16] Ravat F., Teste O., Zurfluh G., Modélisation multidimensionnelle systèmes décisionnels, EGC, Jan 2001, Nantes (France).
- [17] Rivest, S., Bédard, Y. & Marchand P., 2001, Towards better support for spatial decision-making: Defining the characteris SOLAP, Geomatica, 2001
- [18] Tanasescu A., Modélisation multidimensionnelle de données complexes, EDIIS DEA ECD, Univ Lyon2, 2003.
- [19] Teste O., Elaboration d'entrepôts de données complexes, Actes du XVIIIème Congrès INFORSID'00, ISBN 2-906855-16-2, p229-245, 16-19 mai 2000, Lyon (Rhône, France).
- [20] Teste O., Modélisation et manipulation d'entrepôts de données complexes et historisées, Thèse de Doctorat - Université Paul Sabatier de Toulouse (France), 2000.
- [21] Vassiliadis P., Sellis T., A Survey on Logical Models for OLAP Databases, SIGMOD 99.
- [22] Wan T., Zeitouni K., Modélisation d'objets mobiles dans un entrepôt de données, 5èmes journées EGC, janvier 2005.
- [23] Zghal H., Faiz S., Ben Ghézala H., CASME: A CASE Tool for Spatial Data Marts Design and Generation, DMDW 2003.