Professeur : M. Alphonse Diombo THIAKANE

Contact : 77 639 99 88 / 76 661 36 36

Mail : [adthiakane@convergence.sn](mailto:adthiakane@convergence.sn)

1. Problématique de l’informatique décisionnelle

On dispose d’une masse de données très importante issues de sources différentes et de formats variés (base de données, tableur, fichiers, cloud…). Cependant on est confronté à un certain nombre de problèmes telles que :

* La dispersion des données physiquement🡺centralisation
* La nomenclature sur les variables, modalités, unités, formats… 🡺harmonisation
* Des utilisateurs ni statisticiens, ni informaticiens🡺outils d’interrogation autonome
* Evolutions des données🡺système périodique de mise à jour | système global

1. Mise en place du système global

Les phases de la mise en place d’un système global se déclinent comme suit :

* Les interviews : recueillir les besoins en information
* La revue documentaire

Il est conseillé de faire de la revue documentaire avant de faire une interview (ou après) pour se familiariser avec certains termes techniques du domaine sur lequel on fait du décisionnel.

* Annuaire des concepts quantitatifs

On recense l’ensemble des concepts (variables) quantitatifs

* Annuaire des concepts qualitatifs

On recense l’ensemble des concepts (variables) qualitatifs

* Matrice de compatibilité

Dans cette matrice on fait un croisement entre les concepts qualitatifs et ceux quantitatifs et repérer les croisements qui ont un sens.

**Etude de cas :**

**Mise en place d’un système décisionnel de suivi et d’évaluation des performances de la politique nationale dans le secteur de l’éducation et de l’enseignement**

**Relevé des variables :**

|  |  |
| --- | --- |
| **Quantitatives** | **Qualitatives** |
| * *Nombre d’apprenants* * *Nombre de réussite* * *Nombre de redoublement* * *Nombre d’abandon* * *Population* * *Superficie* | * *Niveau d’étude (maternelle, primaire, secondaires, …, formation professionnelle)* * *Classes d’âge (0-6ans, 7-12ans,…)* * *Sexe de l’élève (Masculin, Féminin)* * *Types d’établissements (écoles, instituts, universités)* * *Cycles (CFEE, BFEM, BAC, …)* * *Etablissements (lycée X, Ecoles Y, université Z, …)* * *Région (Dakar, Kaolack…)* * *Départements (Pikine, Rufisque…)* * *Communes (HLM, Colobane…)* * *Pays propriétaires (Gambie, Canada…)* * *Pays d’étude (Sénégal, hors Sénégal)* * *Milieu (urbain, rural)* * *Existence de corrélation (oui, non)* * *Sexe\_maire (Homme, Femme)* * *Sexe\_prési\_dep (Homme, Femme)* * *Sexe\_gouverneur (Homme, Femme)* * *Année* |

NB : Il faut éviter en décisionnel des variables qui ne sont pas sommables. Par exemple taux à la place d’effectifs.

**Exercice :**

Mettre en ligne l’ensemble des concepts quantitatifs et en colonne l’ensemble des concepts qualitatifs. Ensuite identifier les croisements qui ont un sens. S’il a un sens, voir s’il a été demandé ou non.

0 = n’a pas de sens

1 = non demandé

2 = demandé

**Matrice de compatibilité**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Nombre d’apprenants | Nombre de réussite | Nombre de redoublement | Nombre d’abandon | Population | Superficie |
| Niveau d’étude | 2 | 2 | 2 | 2 | 0 | 0 |
| Classes d’âge | 2 | 1 | 1 | 1 | 1 | 0 |
| Sexe de l’élève | 2 | 2 | 1 | 1 | 1 | 0 |
| Types d’établissement | 2 | 2 | 1 | 1 | 0 | 0 |
| Cycles | 2 | 2 | 2 | 2 | 0 | 0 |
| Etablissements | 2 | 2 | 2 | 2 | 0 | 0 |
| Région | 2 | 2 | 2 | 2 | 2 | 2 |
| Départements | 2 | 2 | 2 | 2 | 2 | 2 |
| Communes | 2 | 2 | 2 | 2 | 2 | 2 |
| Pays propriétaires | 2 | 2 | 1 | 1 | 0 | 0 |
| Pays d’étude | 2 | 2 | 2 | 2 | 0 | 0 |
| Milieu | 2 | 2 | 2 | 2 | 0 | 1 |
| Existence de corrélation | 2 | 2 | 1 | 1 | 2 | 2 |
| Sexe\_maire | 2 | 2 | 1 | 1 | 0 | 0 |
| Sexe\_prési\_dep | 2 | 2 | 1 | 1 | 0 | 0 |
| Sexe\_gouverneur | 2 | 2 | 1 | 1 | 0 | 0 |
| Année | 1 | 1 | 1 | 1 | 1 | 1 |

1. Modélisation de la solution décisionnelle

Pour rappel, l’exercice a débuté par le recueil des besoins des décideurs (interviews et revue documentaires). Ensuite par la détermination des concepts quantitatifs, d’une part, et des concepts qualitatifs, d’autre part ; enfin par l’élaboration de la matrice de compatibilité.

* **Gestion des concepts qualitatifs**

Pour réaliser notre base de données décisionnelle, chaque concept qualitatif est transformé en une entité. Celle-ci aura deux attributs : un code qui sera son identifiant et un libellé.

* **Gestion des concepts quantitatifs**

Chaque concept quantitatif devient une association. Conceptuellement, cette association sera reliée à l’ensemble des entités représentant les concepts qualitatifs qui leur sont compatibles.

* **Optimisation des concepts quantitatifs**

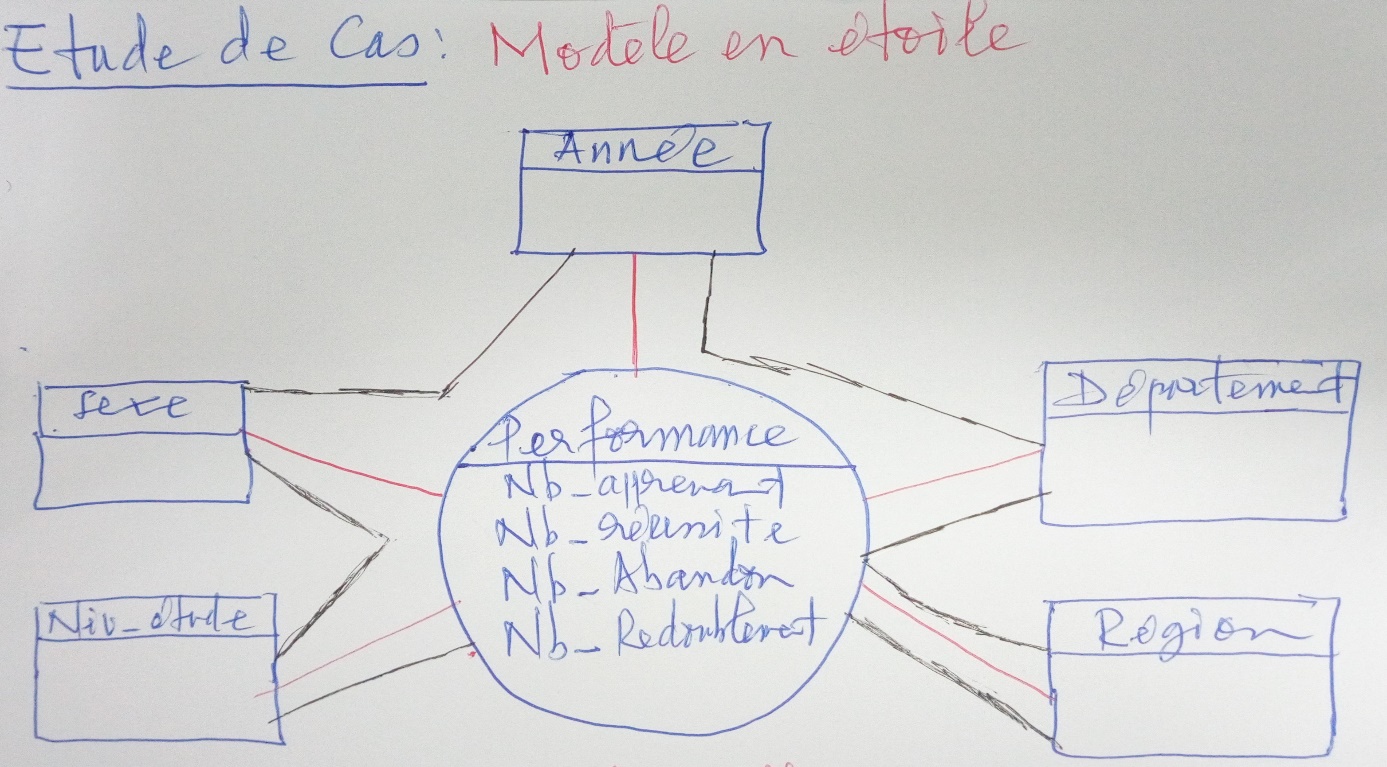
Lorsque deux ou plusieurs concepts quantitatifs partagent **EXACTEMENT** les mêmes concepts qualitatifs compatibles, nous pouvons les fusionner dans la même association.

* **Optimisation des concepts qualitatifs**

Lorsqu’il existe une dépendance fonctionnelle (ou hiérarchie) entre deux ou plusieurs concepts qualitatifs, nous pouvons optimiser notre modèle. Seul le concept qualitatif le plus fin (ou le plus détaillé) reste lié à l’association. Les autres concepts qualitatifs vont être reliés entre eux sous la forme suivante :

1. Quelques définitions

* **Modèle en étoile**



Conceptuellement, notre modèle est composé d’associations autour desquelles gravitent des entités compatibles. Le modèle ainsi obtenu est appelé ***modèle en étoile.***

Les tables correspondant aux entités, et donc aux concepts qualitatifs sont appelées tables de dimension. On parlera donc de dimension année, de dimension département ou de dimension niveau d’étude.

Les tables correspondant aux concepts quantitatifs sont appelées table des faits. On parlera ainsi du fait nombre d’apprenants, du fait nombre d’abandons.

N-B : Dans une table de fait, il faut toujours mettre des concepts indiquant des aspects temporels.

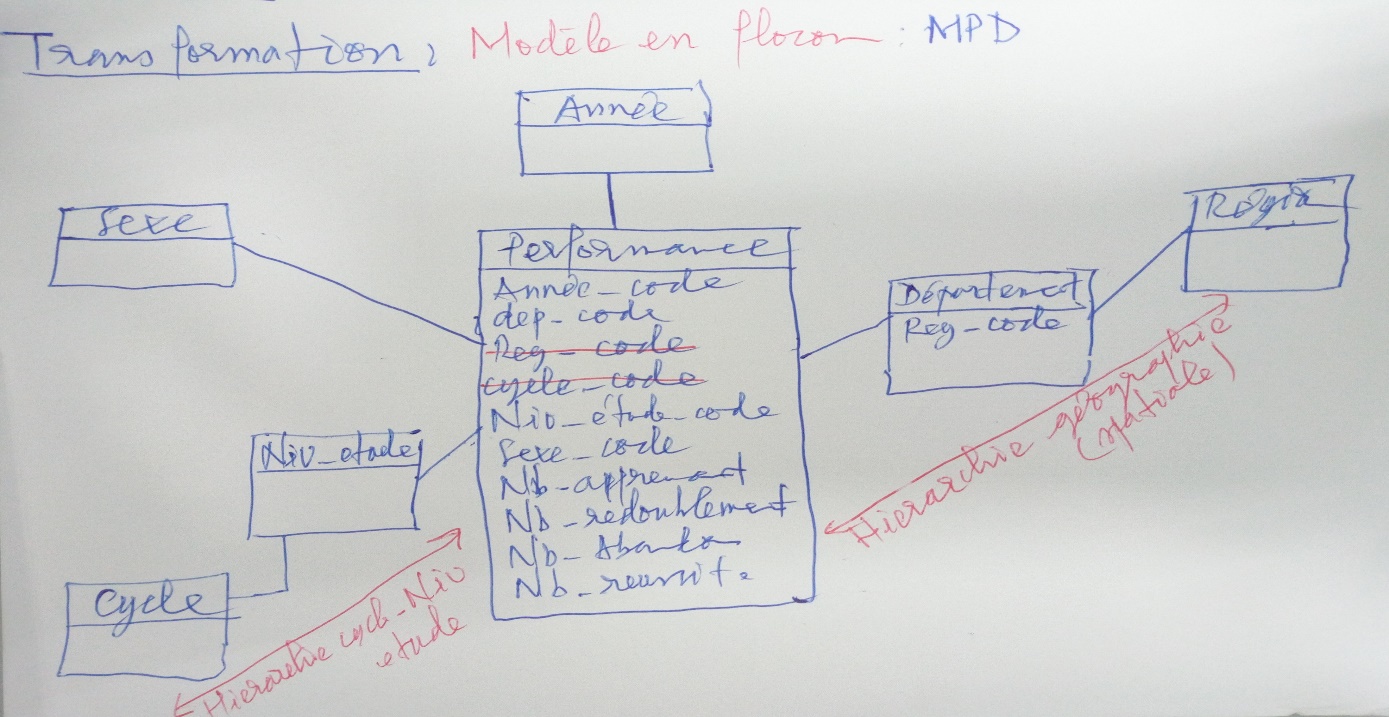
Une dimension est

Un fait est la valeur d’un attribut quantitatif mesuré par rapport à une modalité/occurrence d’une dimension.

Le modèle en étoile fait partie des modèles dits multidimensionnels.

* **Modèle en flocon de neige**

Lorsqu’on utilise l’optimisation des dimensions en exploitant les hiérarchies, le modèle obtenu est le suivant :

On remarquera que dans le modèle en étoile, il n’y a pas de lien entre dimensions alors que dans le modèle ci-dessus les dimensions d’une même hiérarchie sont reliées entre elles. Ce modèle n’est donc pas un modèle en étoile mais un modèle ***en flocon de neige***.

Ce modèle est aussi composé de tables de faits et de tables de dimensions. Le modèle en flocon de neige est également un modèle multidimensionnel.

1. Alimentation du DataWareHouse

* **Outils d’ETL**

Une fois le DW conçu avec ses tables de faits et ses tables de dimensions, il s’agit maintenant d’alimenter ces tables de faits et ces tables de dimensions à partir des données issues des différentes sources.

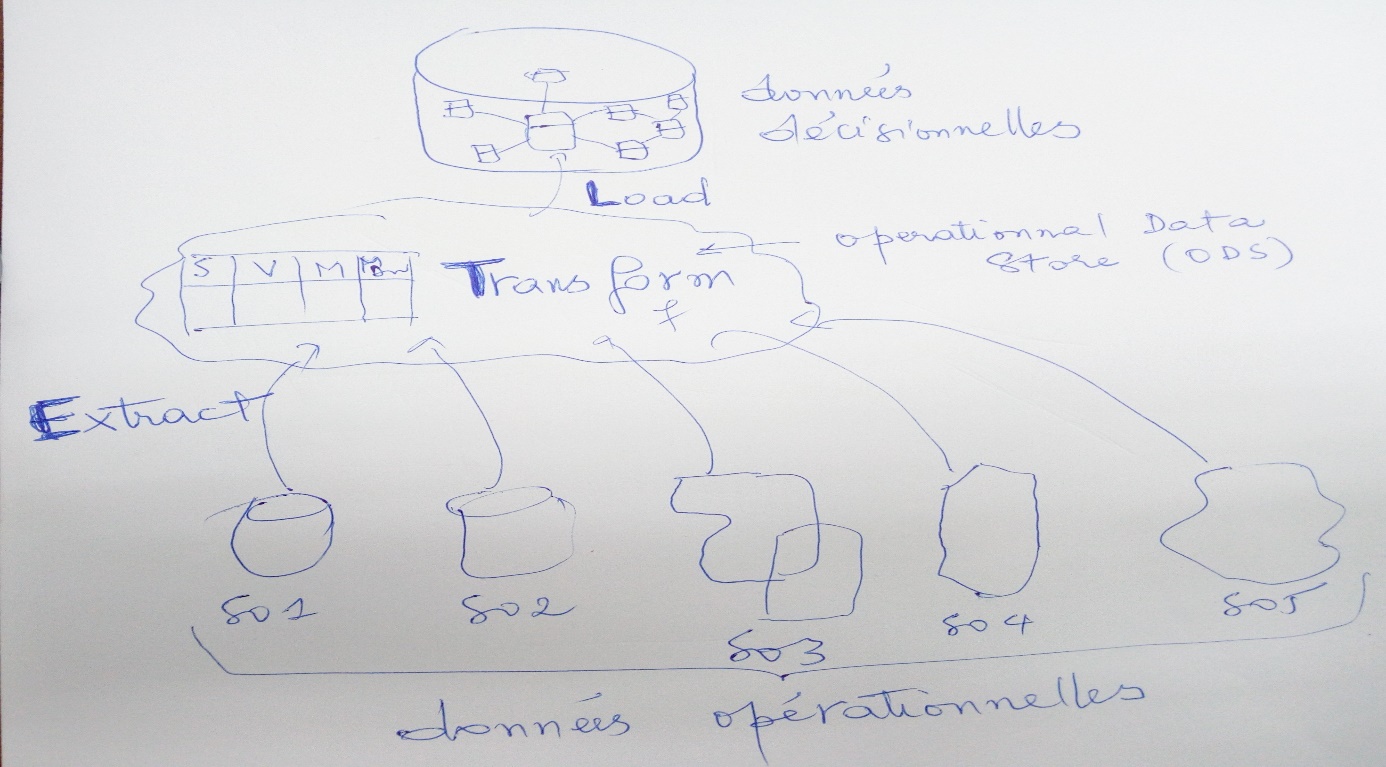
Pour rappel, les données des différentes sources ne sont pas harmonisées ni pour les nomenclatures des dimensions, ni pour les unités des faits.

Avant d’alimenter le DW, il est donc nécessaire d’élaborer une zone tampon où seront provisoirement stockées les données sources avant leur harmonisation et leur transfert dans le DW. Cette zone tampon est appelée « Operationnal Data Store » (ODS).

L’alimentation du DW sera faite par un programme en trois étapes :

* La première étape consiste à extraire (**E**xtract) les données opérationnelles des sources et à les stocker dans l’ODS ;
* La deuxième étape consiste à harmoniser les données extraites via des transformations (**T**ransform) ;
* La troisième étape consiste à charger (**L**oad) les données ainsi harmonisées dans le DW.

Ces outils d’alimentation du DW sont appelés des **ETL** (Extract, Transform, Load).



* **Processus de l’ETL**
* **Phase d’extraction**

La phase d’extraction consiste à se connecter aux différentes sources et à utiliser des requêtes de type SELECT pour récupérer des données dont on a besoin.

* **Phase de transformation**

La phase de transformation est réalisée par la programmation. Il peut s’agir de codes de type if…then ou de programme plus élaboré. Par exemple, nous pouvons concevoir une table de correspondance qui pour chaque :

- source de données S ;

- variable V ;

- Modalité MS

Fournit : la modalité MDW correspondante à charger dans le DW.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source S | Variable V | Modalité MS | Libellé Modalité | Modalité MDW |
| S01 | Region | DK | Dakar | DAKAR |
| S02 | Reg | 01 | Region de Dakar | DAKAR |
| … |  |  |  |  |
| S01 | Niv-etud | 14 | Tle | TERMINAL |

Cette table de correspondance sera stockée dans l’ODS.

* **Phase de chargement**

Le chargement (Load) de données dans le DW sera réalisé grâce à des requêtes de type INSERT.

1. Notion de Datamart

Le DW couvre l’ensemble des thématiques de l’entreprise sur l’ensemble des périodes étudiées. Il peut arriver que nous ayons besoin d’un DW limité à un secteur de notre entreprise (DW du secteur santé au Sénégal) limité à une période (DW 2000-2010 du Sénégal), limité à une zone géographique (DW de la région de Ziguinchor).

Ce sous DW est appelé **Datamart** ou magasin de données.

Notons que techniquement, il est possible de créer d’abord le DW, et d’en tirer ensuite les Datamarts. Inversement, nous pouvons aussi d’abord créer des Datamarts puis de les consolider en un DataWareHouse.

