

PARTIE 3 : CONCEPTION DE LA SOLUTION

La solution proposée va être conçue en plusieurs étapes. La solution étant un système décisionnel s'articulant autour d'un entrepôt de données, on va donc concevoir cet entrepôt de données, sa zone d'alimentation, sa zone de restitution en passant par la conception des cubes.

Chapitre V

Conception de la zone d'entreposage

Introduction

Nous allons dans ce chapitre présenter la conception de la zone d'entreposage qui constitue la première étape de notre conception. À l'issue de cette étape, nous aurons les mesures avec les tables de faits qui les contiennent, ainsi que les dimensions selon lesquelles les mesures vont être analysées. La démarche de **Ralph Kimball**, que nous avons suivie dans cette conception, comprend les étapes suivantes :

- sélectionner le processus d'activité à modéliser (étape effectuée lors de l'étude des besoins),
- déclarer le grain de ce processus (niveau de détail des données),
- choisir les dimensions participantes à ce processus (les axes selon lesquels le processus sera analysé),
- identifier les faits numériques (ce que nous voulons mesurer dans l'analyse).

Nous utiliserons dans cette démarche la modélisation multidimensionnelle.

V.1. Activité de production

V.1.1. Présentation de l'activité

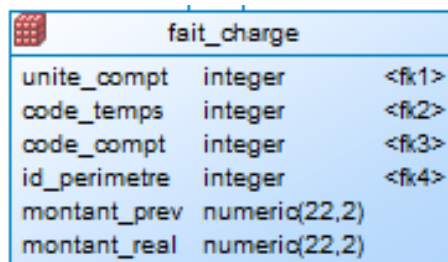
La production fait partie de la vie de l'entreprise depuis sa création. Pour maintenir ou augmenter sa production, Sonatrach doit nécessairement assurer une bonne gestion de son capital et mettre en place de nouveaux moyens. Pour faire face à la demande future en hydrocarbures, il est donc important de connaître les besoins des décideurs à évaluer, qui s'articulent autour de trois volets : volet « suivi des charges », volet « suivi des amortissements de développement », et volet « suivi de production ».

V.1.2. Les faits

- **Le grain** : La déclaration du grain permet de spécifier ce qu'une ligne de la table de faits représente. Le grain doit être le plus fin possible, ce qui veut dire qu'il serait avantageux de retenir les données les plus atomiques possibles.

Dans notre cas, le grain correspond à : connaître le montant des charges prévues, le montant des charges réalisées, et le coût opératoire par Mois, par Association, par Périmètre, et par Nature de charge.

- **Les mesures** :
 - Le montant prévu accordé aux charges de production : <montant_prev>
 - Le montant réalisé accordé aux charges de production : <montant_real>
- **Détails du fait** :



fait_charge		
unite_compt	integer	<fk1>
code_temps	integer	<fk2>
code_compt	integer	<fk3>
id_perimetre	integer	<fk4>
montant_prev	numeric(22,2)	
montant_real	numeric(22,2)	

Figure 21 : Le fait « Charge »²

² Toutes les figures de ce chapitre correspondent à des captures d'écran réalisées à partir du logiciel de conception Power Designer 16.5

Charge	
Attributs	Description
unite_compt	Unité de comptable, concaténation d'un code et d'un numéro séquentiel
code_temps	Code de la date, concaténation du numéro du mois et de l'année
code_compt	Code comptable, concaténation du numéro de la classe comptable et d'un numéro séquentiel
id_perimetre	Numéro séquentiel
montant_prev	Montant prévisionnel accordé aux charges
montant_real	Montant réalisé accordé aux charges

Tableau 11 : Tableau descriptif du fait « Charge »

V.1.2.2. Suivi des amortissements de développement

- **Le grain :** Le grain correspond à : connaître l'amortissement de développement d'une association par Mois, Périmètre et Association.
- **Les mesures :**
 - Valeur de l'amortissement d'une association : <amortissement>
- **Détail du fait :**

fait_amortissement		
unite_compt	integer	<fk1>
code_temps	integer	<fk2>
id_perimetre	integer	<fk3>
amortissement	numeric(22,2)	

Figure 22 : Le fait « Amortissement de développement »

Amortissement de développement	
Attributs	Description
unite_compt	Unité de comptable, concaténation d'un code et d'un numéro séquentiel
code_compt	Code comptable, concaténation du numéro de la classe comptable et d'un numéro séquentiel
id_perimetre	Numéro séquentiel
amortissement	Valeur de l'amortissement

Tableau 12 : Tableau descriptif du fait « Amortissement de développement »

V.1.2.3. Suivi de production :

- **Le grain :** Le grain correspond à : connaître la quantité de production prévue et réalisée de chaque produit par mois, par Périmètre, Produit, et Association.
- **Les mesures :**
 - Production prévue : <production_prev>
 - Production réalisée : <production_real>
- **Détail du fait :**

fait_production		
id_perimetre	integer	<fk1>
unite_compt	integer	<fk2>
code_temps	integer	<fk3>
id_produit	integer	<fk4>
production_real	numeric(22,2)	
production_prev	numeric(22,2)	

Figure 23 : Le fait « Production »

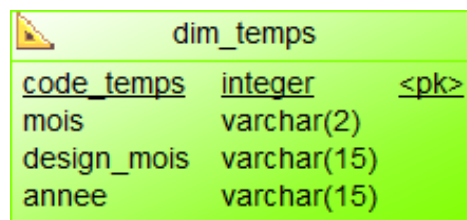
Production	
Attributs	Description
id_perimetre	Numéro séquentiel
unite_compt	Unité de comptable, concaténation d'un code et d'un numéro séquentiel
code_temps	Code de la date, concaténation du numéro du mois et de l'année
id_produit	Numéro séquentiel
production_prev	Production prévue
production_real	Production réalisée

Tableau 13 : Tableau descriptif du fait « Production »

V.1.3. Les dimensions

V.1.3.1. Dimension Temps :

C'est une dimension triviale dans un système décisionnel, elle permet de faire des regroupements temporels selon le mois et l'année, les utilisateurs ont besoin de suivre leur activité d'un mois à un autre et d'en garder l'historique



dim_temps		
<u>code_temps</u>	<u>integer</u>	<pk>
mois	varchar(2)	
design_mois	varchar(15)	
annee	varchar(15)	

Figure 24 : La dimension « Temps »

Temps	
Attributs	Description
code_temps	Code de la date, concaténation du numéro du mois et de l'année.
mois	Numéro du mois
design_mois	Désignation du mois
annee	Année

Tableau 14 : Tableau descriptif de la dimension « Temps »

V.1.3.2. Dimension Association :

Cette dimension décrit toutes les associations de la Sonatrach, dans le domaine de production des hydrocarbures, elle contient l'unité comptable de l'association ainsi que son nom.

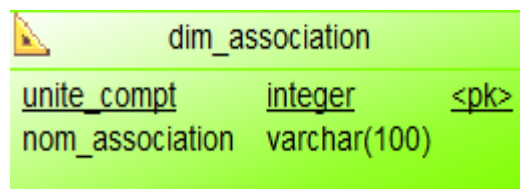


Figure 25 : La dimension « Association »

Association	
Attributs	Description
unite_compt	Unité de comptable, concaténation d'un code et d'un numéro séquentiel
nom_association	Nom de l'association

Tableau 15 : Tableau descriptif de la dimension « Association »

V.1.3.3. Dimension Nature-charge :

Cette dimension permet de regrouper les différentes charges de production qui sont enregistrées au compte comptable de la classe 6 (60 marchandise consommées, 61 matière et fournitures consommées, 62 services...).

Le compte comptable est un compte dans lequel une entreprise enregistre ses actions financières. Il est composé de plusieurs chiffres, le premier chiffre du compte représente le numéro de la classe dans lequel doit s'inscrire l'opération concernée.

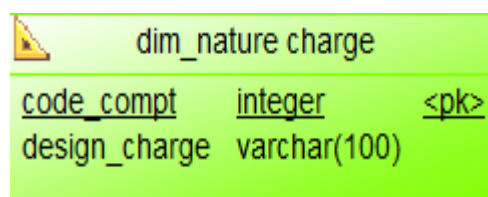


Figure 26 : La dimension « Nature-charge »

Nature-charge	
Attributs	Description
code_compt	Code comptable, concaténation du numéro de la classe comptable et d'un numéro séquentiel
design_charge	Désignation de la nature de charge

Tableau 16 : Tableau descriptif de la dimension « Nature-charge »

V.1.3.4. Dimension Périmètre :

Elle permet d'assurer un suivi par périmètre ou gisement. En effet, chaque gisement (In Saleh, In Amenas, Hassi Berkine...), est décomposé en bloc (403a, 403d...).

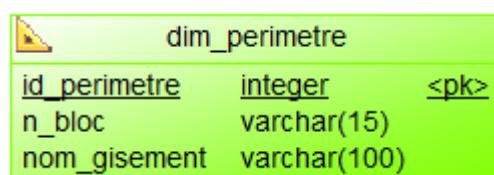


Figure 27 : La dimension « Périmètre »

Périmètre	
Attributs	Description
id_perimetre	Numéro séquentiel
n_bloc	Numéro du bloc
nom_gisement	Nom du gisement

Tableau 17 : Tableau descriptif de la dimension « Périmètre »

V.1.3.5. Dimension Produit

Cette dimension contient les produits des hydrocarbures, on distingue quatre types de produit : Pétrole brute, Gaz naturel, Condensat, GPL (gaz de pétrole liquifié).

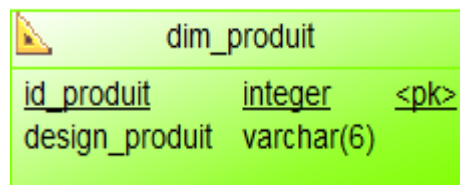


Figure 28 : La dimension « Produit »

Produit	
Attributs	Description
id_produit	Numéro séquentiel
design_produit	Désignation de produit

Tableau 18 : Tableau descriptif de la dimension « Produit »

V.1.4. Dimensions participantes pour chaque fait

Dimension Fait	Temps	Association	Nature charge	Périmètre	Produit
Charge	✓	✓	✓	✓	
Amortissement	✓	✓		✓	
Production	✓	✓		✓	✓

Tableau 19 : Tableau descriptif des dimensions participantes pour chaque fait

V.1.5. Schéma globale de l'activité « Production »

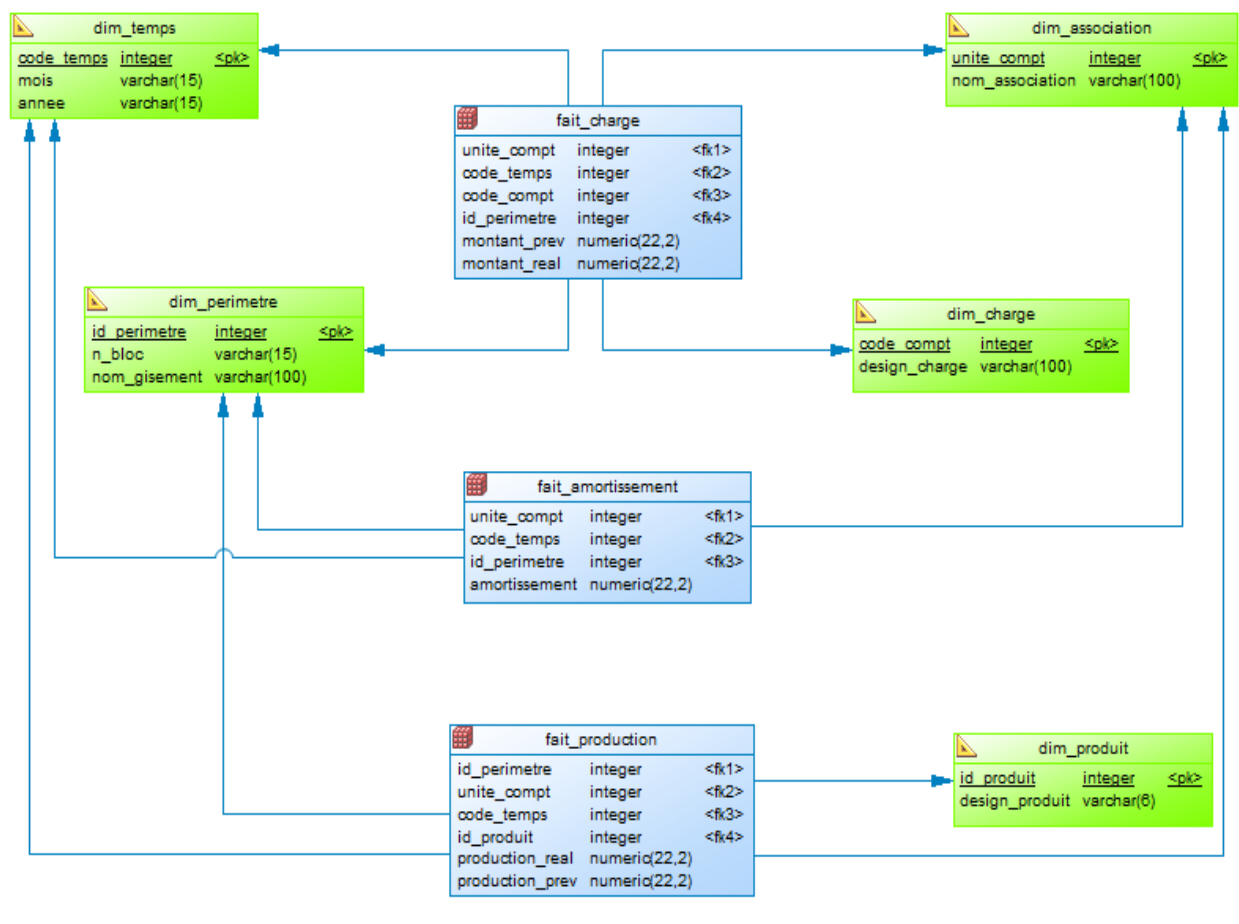


Figure 29 : schéma en constellation de l'activité « Production »

Conclusion

Au terme de cette première étape, nous avons identifié les mesures, les faits qui leur sont associés et les dimensions selon lesquelles seront analysés ces faits. Et cela en tenant compte des besoins des utilisateurs définis précédemment. Conformément à la définition des besoins, nous avons trois volets : suivi des charges, suivi des amortissements, et suivi de production. Ces deux volets correspondent un schéma en constellation qui va constituer l'entrepôt.

L'entrepôt de données que nous avons conçu va être exploité par les utilisateurs, mais avant cela il est nécessaire de l'alimenter.

Nous passons donc à l'étape suivante qui est la construction de la zone d'alimentation de l'entrepôt, abordé dans le prochain chapitre.

Chapitre VI

Conception de la zone d'alimentation

Introduction :

L'alimentation de l'entrepôt de données est une étape importante dans le projet décisionnel, car elle garantira la pertinence et la qualité des données que contiendra l'entrepôt et assurera au décideur l'accès à la bonne information. L'alimentation se fait grâce à l'opération d'ETL, il s'agira d'extraire les données pertinentes ce qui nécessite une bonne connaissance des sources de données, puis de transformer les données au format voulu, pour enfin les charger selon une périodicité précise et de la manière adaptée. Nous présentons dans ce chapitre les différentes étapes que nous avons suivi pour l'alimentation de notre entrepôt.

VI.1. Identification des sources de données

Lors de l'étape de l'identification des besoins, nous avons adopté une démarche qui assure que les besoins définis correspondent à ceux des utilisateurs et que les données existantes permettent de répondre à de tels besoins. De plus l'étape de conception de l'entrepôt a abouti un schéma en constellation qui indique les données nécessaires.

Comme vu précédemment, nos sources de données sont les fichiers Excel issues de département gestion activité « Amont ». Chaque fichier correspond aux réalisations et prévisions des charges et à la production vendue des hydrocarbures.

L'outil utilisé dans le processus d'ETL est Talend Open Studio qui est à la fois gratuit et puissant pouvant supporter une charge importante de données.

VI.2. Extraction des données

Une fois les données identifiées, nous procédons à leur extraction, nous allons récupérer les données jugées pertinentes, destinées à être exploitées et en accord avec le résultat à obtenir. Il est à noter que les fichiers sources ont subis plusieurs modifications pour les rendre plus structurés et organisés sous forme de colonnes avec des champs d'attributs :

- à partir du fichier Excel récapitulatif des charges on a pu extraire les fichiers suivant :
 - fichier associations (UC, association),
 - fichier gisements (id_gisement, gisement, UC),
 - fichier blocs (bloc, id_gisement, UC),
 - fichier charges (UC, association, id_perimetre, gisement, bloc, date, num_compte, design_charge, objectif, réalisation),
 - fichier amortissement (UC, association, id _perimetre, gisement, bloc, date, amortissement).
- à partir du fichier Excel récapitulatif de production vendue on a pu extraire les fichiers suivant :
 - fichier produit (id_produit, produit),
 - fichier production (UC, association, id_perimetre, gisement, bloc, date, id_produit, produit, objectif, réalisation).

VI.3. Transformation

Les données extraites subissent des transformations avant d'être chargées dans l'entrepôt, nous distinguerons les opérations de :

- **Filtrage** : le filtrage nous a permet de faire des restrictions sur les données nulles, les champs vides...etc., dans notre cas l'utilisation de filtrage est d'éliminer le chargement :
 - des associations ou le montant des charges non présent
 - des produits non fournis par certaines associations.
- **Conversion** : la conversion consiste à changer le type de données de certains attributs, les unités de mesures...etc., par exemple :
 - **Date**: date (dd-mm-yyyy) \longrightarrow String (ddmmyyyy).
 - **Production vendue** :
KDA (kilo dinars) \longrightarrow TEP (tonne équivalent pétrole).
 - **Montant charges** : Integer \longrightarrow BigDecimal.
- **Changement des libellés des champs** : le changement des libellés permet d'éviter des conflits qui se produisent lorsqu'on utilise des noms différents pour le même attribut, ainsi pour rendre les champs plus significatifs. Par exemple l'unité comptable des associations se trouve avec des différents libellés dans plusieurs fichiers sources :

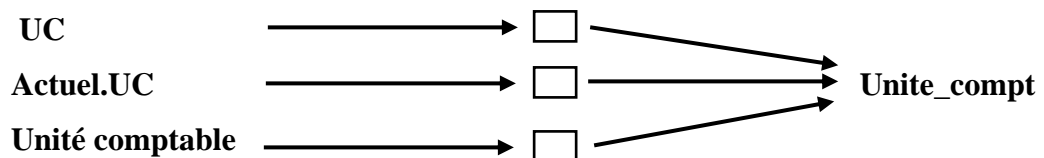


Figure 30 : Exemple de changement des libellés

De plus de ces opérations on a utilisé les opérations de sélection (sélectionner le mois ou l'année d'une date donnée), de concaténation (concaténer le mois et l'année pour avoir le code date), et de regroupement des données.

Ces transformations se font généralement au moment du mapping (correspondance entre les données).

VI.4. Chargement

Nous distinguerons le chargement des dimensions et le chargement des tables de faits, mais aussi un chargement initial et des chargements périodiques. Le chargement initial ne comporte pas de contraintes particulières, cependant les chargements périodiques doivent prendre en compte différents critères :

- La fréquence de mise à jour des fichiers sources,
- Le volume de données à charger, le temps nécessaire et le temps disponible pour le chargement.
- Niveau d'actualisation des données demandé par les utilisateurs.

Le chargement des tables de faits doit venir après le chargement des dimensions car les premières référencent les secondes à l'aide des clés étrangères.

L'opération de chargement de la source vers l'entrepôt sera exécutée à chaque fin du mois, car le plus détaillé en termes de période d'analyse est le mois.

La figure suivante illustre le processus de chargement.

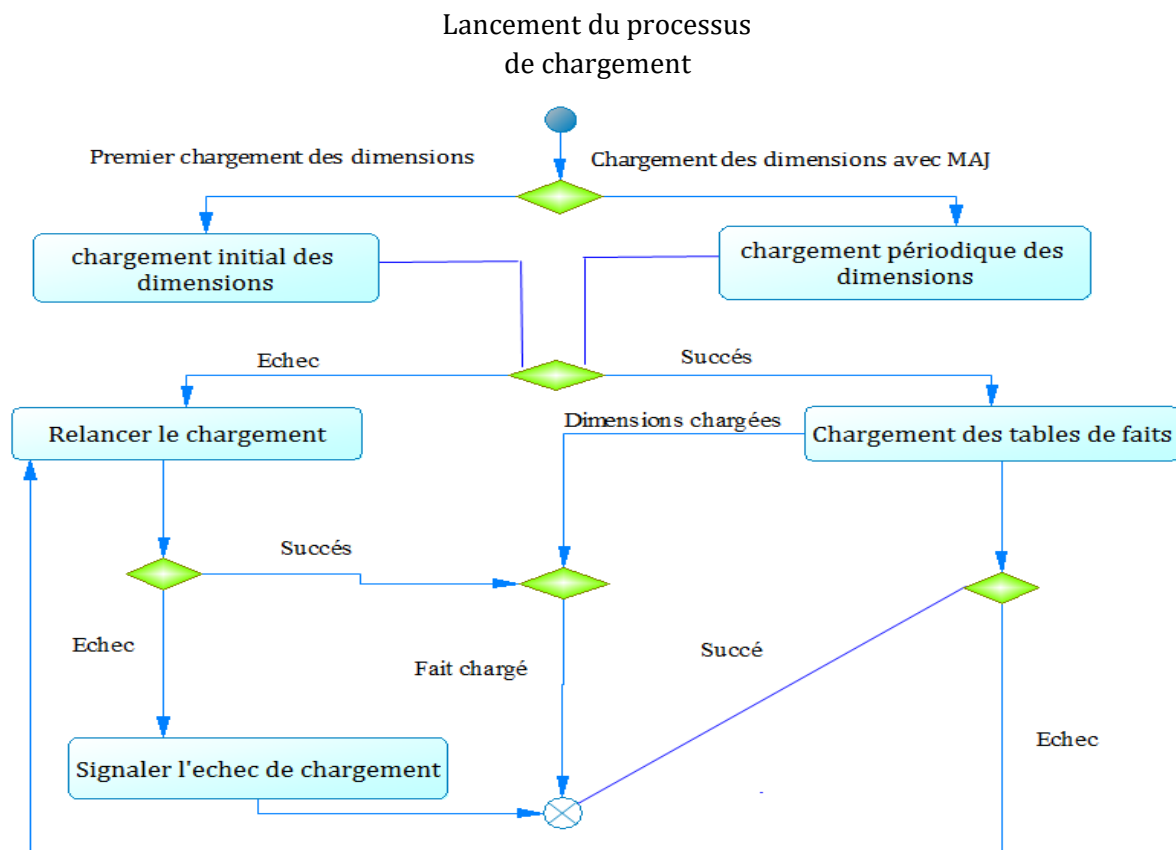


Figure 31 : Diagramme d'activité de processus de chargement.

Conclusion

Dans ce chapitre nous avons abordé la conception de la zone d'alimentation qui va nous assurer des données fiables et de qualité dans l'entrepôt de données. Nous avons vu les différentes étapes : identification des sources de données, extraction, transformation et enfin chargement. Pour la périodicité des chargements, elle a été déterminée en fonction de la fréquence de mise à jour des tables sources et des besoins des utilisateurs (chaque fin du mois).

Cette étape achevée nous passons dans le chapitre suivant, à la conception des cubes de données.

Chapitre VII

Conception des cubes dimensionnels

Introduction

Dans ce chapitre nous abordons la conception des cubes dimensionnels, une structure très souple à l'interrogation des données offrant une navigation adaptée et une exploitation. Un cube correspond à un sujet d'analyse : un fait décrit par les besoins des utilisateurs, il contient un ensemble d'indicateurs en hiérarchies, les hiérarchies étant décomposées en niveaux. La navigation peut se faire de manière croisée (interroger la valeur d'un indicateur selon trois axes par exemple), de manière ascendante ou descendante en jouant sur le niveau de granularité des axes d'analyse et de manière spécifique à l'aide de filtres.

L'expression des besoins des utilisateurs a conduit à la conception entre autres un schéma en constellation concernant les trois volets : suivi des charges et suivi des amortissements, et suivi de production. Ce schéma correspond (dimensions, mesures) aux cubes qui seront conçus. Nous détaillons, dans ce qui suit la conception de chaque cube.

VII.1. définition des niveaux et hiérarchies des dimensions

Les hiérarchies de chaque dimension ainsi que les niveaux correspondant à chaque hiérarchie vont définir une structure hiérarchique. Cette structure permettra la navigation à l'intérieur de la dimension à travers les opérations de « forage vers le bas » (accès à un niveau de détail supérieur) ou de « forage vers le haut » (accès à un niveau de détail inférieur).

Par exemple, l'utilisateur peut analyser la production par année (sommet de la hiérarchie) puis par mois (opération de forage vers le bas).

Le tableau suivant regroupe les dimensions de l'ensemble des cubes, avec la définition des hiérarchies, niveaux et colonnes correspondantes.

Dimension	Hiérarchies	Niveaux	Colonnes
Dim_temps	h_temps ALL → N1 → N2	Niveau 1 = N1	annee
		Niveau 2 = N2	code_date
			mois
			design_mois
Dim_association	h_association ALL → N1	Niveau 1 = N1	unit_compt
			nom_association
Dim_nature charge	h_nature_charge ALL → N1	Niveau 1 = N1	code_compt
			design_chareg
Dim_perimetre	h_perimetre ALL → N1 → N2	Niveau 1 = N1	nom_gisement
		Niveau 2 = N2	id_perimetre
			n_bloc
Dim_produit	h_produit ALL → N1	Niveau 1 = N1	id_produit
			design_produit

Tableau 20 : Tableau donnant les niveaux et les hiérarchies de chaque dimension

VII.2. Listes des cubes

Le tableau suivant liste tous les cubes à réaliser avec, pour chacun, les mesures correspondantes et les dimensions participantes :

Nom de cube	Mesures	Dimensions
Volet suivi des charges :		
Charge	<ul style="list-style-type: none">▪ Montant_prev▪ Montant_real	Dim_temps
		Dim_association
		Dim_nature_charge
		Dim_peimetre
Volet suivi des amortissements :		
Amortissement de développement	<ul style="list-style-type: none">▪ Amortissement	Dim_temps
		Dim_association
		Dim_peimetre
Volet suivi de production :		
Production	<ul style="list-style-type: none">▪ Production_prev▪ Production_real	Dim_temps
		Dim_produit
		Dim_association
		Dim_perimetre

Tableau 21 : Liste des cubes

VII.3. schémas des cubes

VII.3.1. volet suivi des charges

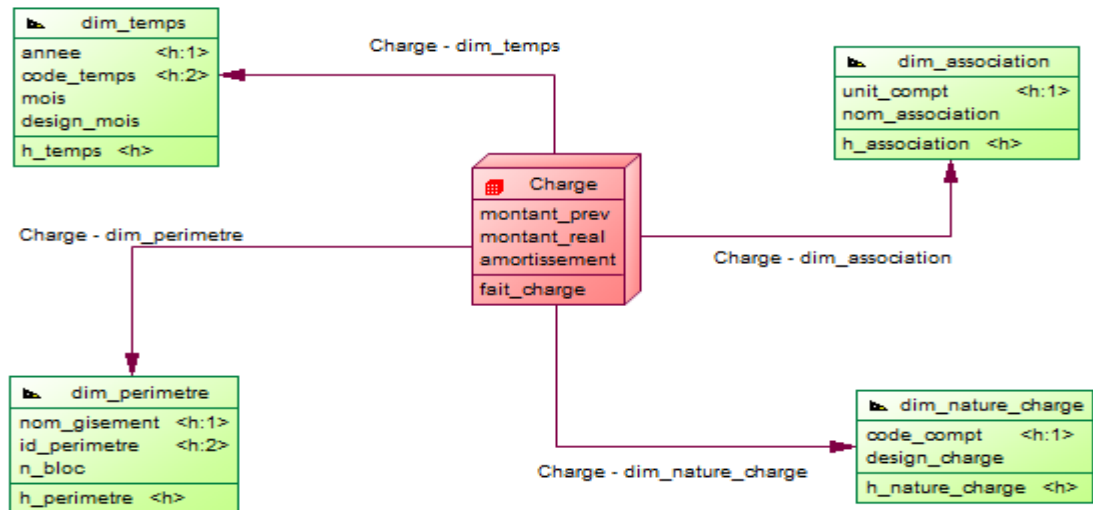


Figure 32 : Cube dimensionnel « Charge »

VII.3.2. volet suivi des amortissements de développement

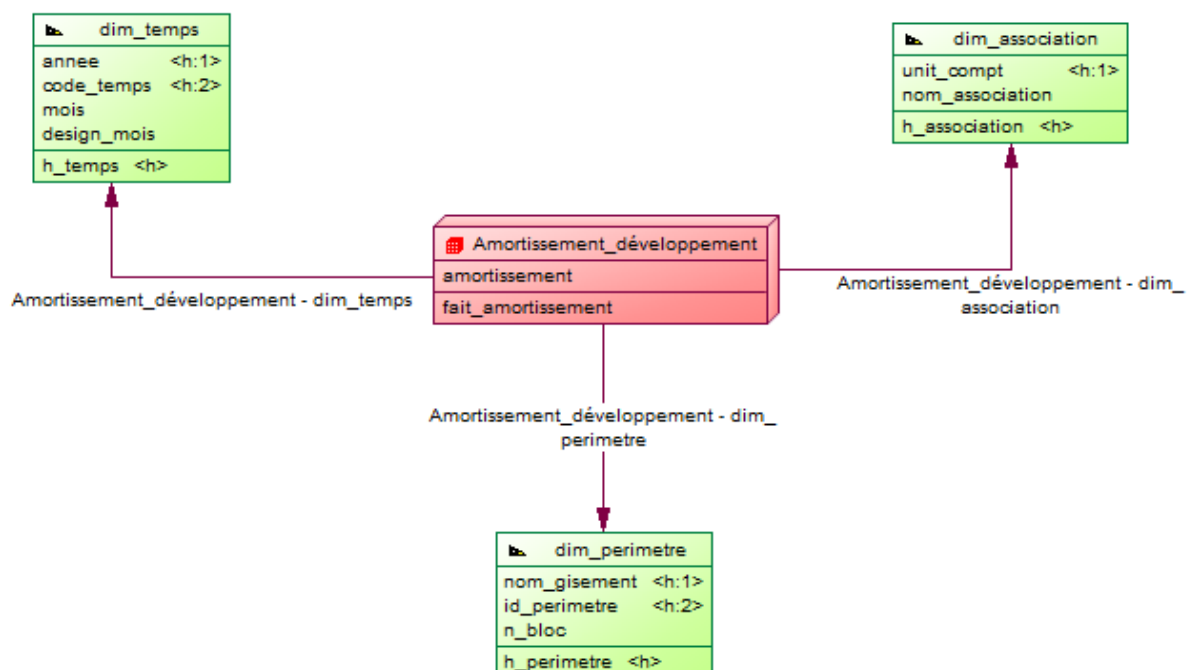


Figure 33 : Cube dimensionnel « Amortissement Développement »

VII.3.3. volet suivi de production

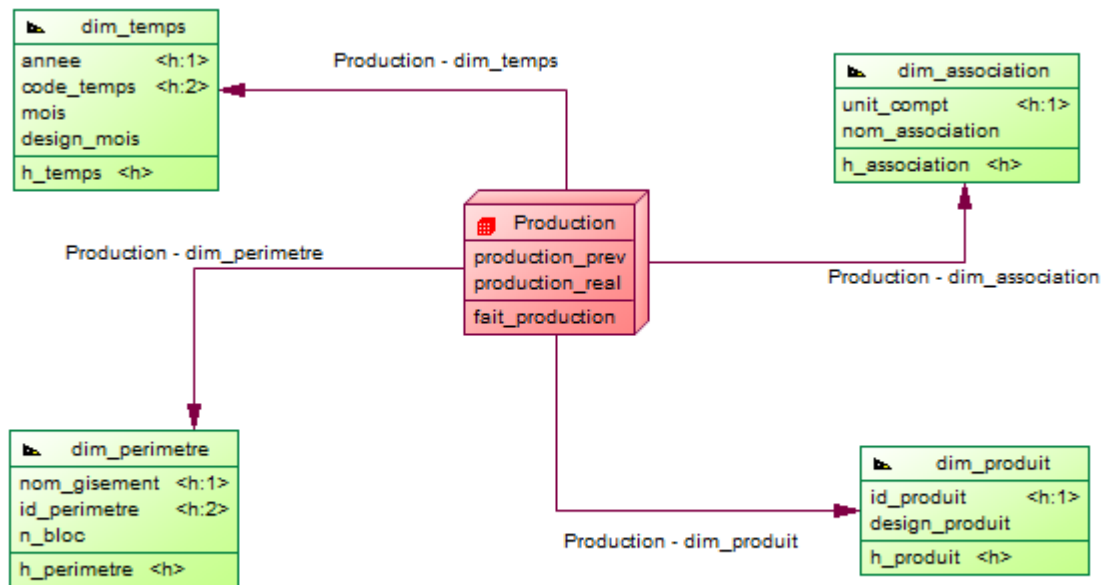


Figure 34 : Cube dimensionnel « Production »³

³ Toutes les figures de ce chapitre correspondent à des captures d'écran réalisées à partir du logiciel de conception Power Designer 16.5

Conclusion

Dans ce chapitre nous avons conçu les cubes dimensionnels qui vont permettre d'effectuer des analyses multidimensionnels sur les données. Nous avons défini les mesures, les dimensions participantes ainsi que la structure hiérarchique pour chaque dimension. Ces cubes sont relatifs aux trois volets : suivi des charges, suivi des amortissements de développements et suivi de production. Ils offrent la possibilité aux décideurs d'exploiter les informations relatives à ces deux volets, grâce à une navigation rapide, intuitive et personnalisable (choix des axes d'analyse, du niveau du détail...).

La conception des cubes achevée, nous devons les implémenter. Cette implémentation est une des étapes de la réalisation et du déploiement, que nous aborderons dans la partie suivante