

ACOD-BI

Création Automatique de Datamarts Optimisés

Manuel d'utilisation

Table des matières

1	Introduction.....	3
1.1	Présentation de l'outil	3
1.2	Objectif du document.....	3
2	Prérequis	4
3	Vue d'ensemble de la démarche	5
4	Démarche évolutive.....	6
4.1	Etape 1 – Utilisation d'ACOD-BI sans statistiques d'utilisation des données dans les restitutions ..	6
4.2	Etape 2 – Utilisation d'ACOD-BI avec statistiques d'utilisation des données dans les restitutions ..	6
5	Concepts manipulés lors des saisies	7
6	Les fichiers CSV	7
7	Détail des étapes.....	8
7.1	Etape 1 (Administrateur) - Initialisation	8
7.1.1	(Obligatoire) DUMP de la source avec statistiques et sans données	8
7.1.2	(Obligatoire) Saisie des paramètres globaux.....	8
7.2	Etape 2 (ACOD-BI)	12
7.3	Etape 3 (Administrateur) – Sélection des attributs et des indicateurs.....	12
7.3.1	(Obligatoire) Fichier des informations	12
7.3.2	(Facultatif) Fichier des entités	12
7.3.3	(Facultatif) Fichier des liens hiérarchiques.....	12
7.4	Etape 4 (ACOD-BI)	12
7.5	Etape 5 (Administrateur) – Saisie des fréquences d'utilisation	13
7.5.1	(Obligatoire) Fichier des entités	13
7.5.2	(Obligatoire) Fichier des cas d'utilisation.....	13
7.6	Etape 6 (ACOD-BI)	13
7.7	Etape 7 (Administrateur) – Création et chargement du datamart.....	13
8	Statistiques d'utilisation des données	14
8.1	Fréquence d'utilisation des indicateurs.....	14
8.2	Fréquence d'utilisation des attributs	15
8.2.1	Saisie au niveau attribut.....	15
8.2.2	(Facultatif) Saisie au niveau branche	15
9	Codes du datamart.....	18
10	Annexe A – Structure des fichiers CSV	19
10.1	Fichier des paramètres globaux	19
10.2	Fichier des informations	20
10.3	Fichier des entités	22
10.4	Fichier des liens hiérarchiques.....	23
10.5	Fichier des cas d'utilisation	24
11	Annexe B – Création et chargement du datamart	26
11.1	Paramètres de l'instance	26
11.2	Droits du propriétaire du datamart.....	26
11.3	Package d'outils d'ACOD-BI	26
11.4	Création du datamart	26
11.5	Chargement du datamart	26

1 Introduction

1.1 Présentation de l'outil

Nature de l'outil

ACOD-BI génère automatiquement des datamarts optimisés dans un volume contraint. Il identifie les agrégats optimums à partir de statistiques de restitution.

Dans sa version actuelle, ACOD-BI vous permet de générer un datamart en quelques heures au lieu de quelques semaines pour une création manuelle.

Le datamart généré est fourni sous la forme de deux scripts SQL, lisibles et modifiables. Une fois le datamart généré l'utilisateur n'est donc plus dépendant de l'outil.

Accès à ACOD-BI

ACOD-BI est pour le moment accessible gratuitement en version bêta. Pour créer votre datamart contactez contact@acod-bi.com.

Données en entrée

ACOD-BI détermine la structure du datamart essentiellement en fonction des informations suivantes :

- Structure logique des données,
- Statistiques sur les données,
- Statistiques sur l'utilisation des données dans les restitutions,
- Volume accordé au datamart.

Caractéristiques des datamarts créés par ACOD-BI

Les datamarts créés par ACOD-BI sont modélisés en étoile.

Les agrégats les plus pertinents sont ajoutés aux tables de faits de détail.

Pour chaque table de faits de détail, une vue unique rassemble les indicateurs¹ et les attributs² de la table de faits et de ses tables de dimension.

L'utilisateur construit ses requêtes sur ces vues sans se préoccuper de la manière dont sont modélisées les tables de dimension et sans être impacté en cas d'évolution du modèle du datamart.

Les structures nécessaires sont créées pour permettre au mécanisme de QUERY_REWRITE d'Oracle d'utiliser les agrégats de manière transparente.

1.2 Objectif du document

Ce document explique la démarche à suivre pour :

1. Fournir à l'outil les informations qui lui sont nécessaires,
2. Générer les scripts de création et d'alimentation du datamart.

¹ Un indicateur est une information sur laquelle, dans les restitutions, est appliquée une fonction d'agrégation (SUM, MIN et MAX).

² Un attribut est une information utilisée, dans les restitutions, pour des regroupements et/ou des filtres.

2 Prérequis

La modélisation de la source³ est proche de la 3ème forme normale.

Plus la modélisation est proche de la 3ème forme normale, mieux l'outil connaît la structure logique de la source et plus le datamart généré est pertinent.

Les statistiques de la source sont à jour.

Un dump de la source sans les données mais avec les statistiques peut être fourni.

L'administrateur du système peut fournir une estimation de la fréquence d'utilisation dans les restitutions des données les plus utilisées.

Plus ces estimations sont complètes et justes, et plus le datamart généré est pertinent.

Pour un début de projet pour lequel ces statistiques sont difficiles à estimer, voir le chapitre « Démarche évolutive ».

La source est contenue dans un seul schéma.

Si ce point est bloquant une évolution de l'outil est envisageable.

Le schéma du datamart dispose d'un database link vers la source.

Si ce point ne peut pas être respecté, des modifications manuelles du script d'alimentation du datamart seront nécessaires.

Les clés étrangères formant des boucles ne sont pas prises en compte.

Si ce point est bloquant une évolution de l'outil est envisageable.

Le seul SGBD supporté pour le moment est le SGBDR Oracle.

Si nécessaire un portage sur un autre SGBD est envisageable.

L'outil a été testé avec des datamarts sur la version 12.2.0.1.0 mais devrait fonctionner sur des versions antérieures.

³ Dans les cas les plus classiques la source est un data warehouse.

Un chargement direct depuis un système opérationnel est envisageable dans les conditions suivantes :

- la modélisation du système opérationnel est proche d'une 3ème forme normale,
- on dispose d'une fenêtre de temps, pendant laquelle le système opérationnel n'est pas utilisé, suffisamment longue pour charger le datamart.

3 Vue d'ensemble de la démarche

Les échanges entre l'administrateur et le système sont réalisés via des fichiers CSV⁴.
Pour plus de détails voir le chapitre « Les fichiers CSV ».

Etape 1 (Administrateur)

L'administrateur du système fournit un dump de la source sans les données mais avec les statistiques.
L'administrateur du système fournit diverses informations générales. Certaines sont facultatives d'autres obligatoires.

Etape 2 (ACOD-BI)

Le dump de la source est monté dans un des environnements d'ACOD-BI.
ACOD-BI génère un fichier (dit fichier des informations) permettant à l'administrateur du système de choisir les informations correspondant à des attributs et/ou des indicateurs ainsi qu'un fichier (dit fichier des entités) listant les entités.

Etape 3 (Administrateur)

L'administrateur du système indique dans le fichier des informations, les colonnes de la source correspondant à des attributs et/ou des indicateurs.

Etape 4 (ACOD-BI)

ACOD-BI charge le fichier des informations et génère le fichier CSV de saisie des statistiques d'utilisation des attributs dans les restitutions (dit fichier de cas d'utilisation).

Etape 5 (Administrateur)

L'administrateur du système indique dans le fichier de cas d'utilisation les fréquences d'utilisation des attributs et dans le fichier des entités le poids relatif de chaque groupe d'indicateurs.

Etape 6 (ACOD-BI)

ACOD-BI charge le fichier de cas d'utilisation et le fichier des entités.
Il génère le meilleur datamart possible au vu des informations reçues (modèle logique de la source, statistiques des données, statistiques des utilisations de données et volume alloué au datamart).
Les scripts de création et de chargement du datamart sont mis à disposition de l'administrateur du système.

Etape 7 (Administrateur)

L'administrateur du système intègre le datamart à son système.

⁴ ACOD-BI génère et charge tous les fichiers CSV à chaque étape.

Pour faciliter sa compréhension, le chapitre « vue d'ensemble de la démarche » ne décrit que les générations et les chargements de fichier indispensables de chaque étape.

4 Démarche évolutive

Il est parfois difficile d'estimer les statistiques d'utilisation des données en début de projet. Pour pallier à ce problème, il est possible d'utiliser ACOD-BI en deux temps.

4.1 Etape 1 – Utilisation d'ACOD-BI sans statistiques d'utilisation des données dans les restitutions

ACOD-BI est exécuté sans statistiques d'utilisation des données et sans agrégats pré calculés. Les restitutions sont développées de manière normale en s'appuyant sur les vues fournies par ACOD-BI.

4.2 Etape 2 – Utilisation d'ACOD-BI avec statistiques d'utilisation des données dans les restitutions

Les statistiques d'utilisation des données sont déduites des restitutions (dans la version actuelle d'ACOD-BI ceci est une opération manuelle).

ACOD-BI est exécuté avec ces statistiques et avec un volume cible permettant de créer des agrégats.

Le datamart est recréé et ré alimenté.

Les développements des restitutions ne sont pas impactés.
Les performances des restitutions sont améliorées.

5 Concepts manipulés lors des saisies

En plus des notions d'attribut et d'indicateur mentionnées précédemment, trois autres concepts sont manipulés dans les fichiers.

Les entités

Les entités correspondent aux tables de la source⁵.

Les informations

Les informations correspondent aux attributs et aux indicateurs sélectionnés.

Pour une colonne de la source utilisée à la fois en tant qu'attribut et en tant qu'indicateur, il est créé deux informations.

Les liens hiérarchiques

Les liens hiérarchiques correspondent aux clés étrangères de la source.

6 Les fichiers CSV

Les échanges entre l'administrateur et ACOD-BI sont réalisés via des fichiers CSV.

Pour simplifier la saisie par l'administrateur, ACOD-BI génère des fichiers CSV pré-remplis.

ACOD-BI génère et charge tous les fichiers CSV à chaque étape.

En cas d'itérations successives, ACOD-BI re génère les fichiers CSV en intégrant les saisies déjà effectuées par l'administrateur dans l'itération précédente.

Le format des nombres est sans séparateur de milliers, avec séparateur décimal ".".

Les fichiers CSV peuvent être édités, par exemple dans Excel ou dans OpenOffice Calc.

L'administrateur du système ne doit pas créer de nouvelles lignes et ne doit modifier que les colonnes dont le titre est préfixé par « NEW ».

Les fichiers CSV sont échangés entre ACOD-BI et l'administrateur du système par FTP-S.

La structure des fichiers CSV est décrite dans l'annexe A « Structure des fichiers CSV ».

⁵ Il existe de rares exceptions : si, pour une clé étrangère de la source, les colonnes de la clé dans la table fille forment une clé unique obligatoire alors les deux tables partagent la même entité dans ACOD-BI.

7 Détail des étapes

7.1 Etape 1 (Administrateur) - Initialisation

7.1.1 (Obligatoire) DUMP de la source avec statistiques et sans données

La commande pour obtenir le dump est :

```
exp <username>/<password>@<sid> OWNER=<username> FILE=<username>.dmp compress=N  
LOG=exp_<username>.log rows=N
```

<username> étant le propriétaire de la source.

7.1.2 (Obligatoire) Saisie des paramètres globaux

L'administrateur saisit dans le fichier des paramètres globaux d'ACOD-BI les valeurs spécifiques à son système.

7.1.2.1 (Obligatoire) Code identifiant la source

L'administrateur du système doit indiquer un code permettant d'identifier la source (paramètre CODE_DWH).

Ce code permet de faciliter la gestion des différents environnements.

7.1.2.2 (Obligatoire) Database link du datamart vers la source

L'administrateur du système doit indiquer le nom du database link du datamart permettant de lire la source (paramètre DBL_DTM2DWH).

7.1.2.3 (Obligatoire) Noms des tablespaces du datamart

L'administrateur du système doit indiquer les noms des tablespaces du datamart à savoir :

- Le tablespace des données des tables de faits
(Paramètre TBS_DTM_FACT_DATA),
- Le tablespace des indexes des tables de faits
(Paramètre TBS_DTM_FACT_INDEX),
- Le tablespace des données des tables de référence
(Paramètre TBS_DTM_REF_DATA),
- Le tablespace des indexes des tables de référence
(Paramètre TBS_DTM_REF_INDEX).

Si le découpage par tablespace proposé est insuffisant des modifications manuelles du script de création du datamart seront nécessaires.

7.1.2.4 (Obligatoire) Espace disque cible pour le datamart

L'administrateur du système doit indiquer quel espace disque il souhaiterait que le datamart généré occupe.

Cela est fait avec le paramètre EXPLICIT_DTM_ADS qui indique, en octets, la cible souhaitée en termes d'espace occupé par le datamart dans les tablespaces.

7.1.2.5 (Facultatif) Fonction de détermination du modèle de code entité

Le modèle de code entité⁶ sert de base à ACOD-BI pour construire les noms des tables et des objets associés dans le datamart.

Par défaut le modèle de code entité est identique au nom de la table correspondante de la source.

Dans certain cas il peut être intéressant de transformer le nom de la table, par exemple pour supprimer un préfixe ou pour transcoder des noms de table non parlant en noms plus compréhensibles.

Cette opération est faite par la fonction de détermination du code entité.

Le nom de la fonction est indiqué par le paramètre FUNC_NTAB_CODE2ENTI_CODE.

La fonction doit être définie dans la source et sa signature doit être :

```
<nom_de_fonction>(nom_de_table IN VARCHAR2) RETURN VARCHAR2
```

7.1.2.6 (Facultatif) Fonction de détermination du modèle de code information

Le modèle de code information⁷ sert de base à ACOD-BI pour construire les noms des colonnes du datamart ainsi que les noms de certaines tables.

Les noms des colonnes de la source peuvent être utilisés comme modèle de code information mais dans certain cas il peut être intéressant de transformer le nom de la colonne, par exemple pour supprimer un préfixe ou pour transcoder des noms de colonne non parlants en noms plus compréhensibles.

Cette opération est faite par la fonction de détermination du code information.

Le nom de la fonction de détermination est indiqué par le paramètre FUNC_NCOL_CODE2INFO_CODE.

La fonction de détermination peut être une fonction prédéfinie d'ACOD-BI ou une fonction définie dans la source.

Pour les fonctions prédéfinies d'ACOD-BI la valeur du paramètre doit être

```
<nom_de_fonction>@DBL_STARP.
```

La signature de la fonction doit être :

```
<nom_de_fonction>(code_entite IN VARCHAR2, nom_de_table IN VARCHAR2, nom_de_colonne IN VARCHAR2) RETURN VARCHAR2
```

Par défaut, si le nom de la colonne dans la source est préfixé par le code entité, le modèle de code information est le nom de la colonne sans le préfixe, sinon le modèle de code information est identique au nom de la colonne.

La valeur de paramètre correspondante est NCOL_CODE2INFO_CODE_V1@STARP.

A la date du 07/03/2018 il n'existe pas d'autre fonction de détermination du code information prédéfinie dans ACOD-BI. D'autres fonctions de détermination prédéfinies seront ajoutées au fil de l'eau en fonction des besoins.

⁶ Voir le chapitre « Codes du datamart » pour voir la construction d'un code à partir d'un modèle de code.

⁷ Voir le chapitre « Codes du datamart » pour voir la construction d'un code à partir d'un modèle de code.

7.1.2.7 (Facultatif) Fonction de détermination du code différenciant des liens hiérarchiques

Lorsque, dans la source, deux clés étrangères ont la même table fille et la même table mère, le système est susceptible de créer plusieurs colonnes dans les vues du datamart, pour une seule colonne de la table mère de la source.

Il est préférable de différencier les colonnes par des codes significatifs plutôt que par des numéros. Avec certaines normes de nommage, ces codes significatifs peuvent être extraits des noms de clés étrangères.

Cette opération est faite par la fonction de détermination du modèle de code différenciant⁸ des liens hiérarchiques.

Le nom de la fonction de détermination est indiqué par le paramètre FUNC_NFK_CODE2HILI_DIFF_CODE.

La fonction de détermination peut être une fonction prédéfinie d'ACOD-BI ou une fonction définie dans la source.

Pour les fonctions prédéfinies d'ACOD-BI la valeur du paramètre doit être <nom_de_fonction>@DBL_STARP.

La signature de la fonction doit être :

```
<nom_de_fonction>(nom_de_la_fk IN VARCHAR2) RETURN VARCHAR2
```

Par défaut, pour déterminer le modèle de code différenciant d'une clé étrangère, le système teste si le nom de cette clé étrangère est de la forme :

```
[[FK_]<table fille>_]<table mere>_<code complémentaire>.
```

Si c'est le cas le modèle de code différenciant est <code complémentaire>.

Si ce n'est pas le cas le modèle de code différenciant est nul.

La valeur de paramètre correspondante est NFK_CODE2HILI_DIFF_CODE_V1@STARP.

A la date du 07/03/2018 il n'existe pas d'autre fonction de détermination du code différenciant des liens hiérarchiques prédéfinie dans ACOD-BI. D'autres fonctions de détermination prédéfinies seront ajoutées au fil de l'eau en fonction des besoins.

⁸ Voir le chapitre « Codes du datamart » pour voir la construction d'un code à partir d'un modèle de code.

Exemple d'utilisation de NFK_CODE2HILI_DIFF_CODE_V1@STARP

Source

La source contient une table SALES, une table CUSTOMER et une table COUNTRY.

La table COUNTRY contient un COUNTRY_CODE (identifiant) et un COUNTRY_LABEL.

La table CUSTOMER contient un COUNTRY_CODE_BIRTH et un COUNTRY_CODE_HOME avec des clés étrangères vers COUNTRY nommées respectivement FK_CUSTOMER_COUNTRY_BIRTH et FK_CUSTOMER_COUNTRY_HOME.

La table SALES est une table de faits. Elle permet entre autre de suivre le nombre de produits vendus et les montants. Elle contient une clé étrangère vers la table CUSTOMER.

ACOD-BI

L'outil regroupe dans une vue unique toutes les informations des ventes.

Cette vue contient (entre autres) :

- une colonne pour le country label home du client,
- une colonne pour le country label birth du client.

Il est préférable de différencier ces deux colonnes par des codes significatifs (BIRTH et HOME) plutôt que par des numéros.

La norme de nommage des clés étrangères de la source est FK_<table fille>_<table mère>[<code complémentaire>] (comme pré supposé par défaut par le système).

Les deux codes différenciant sont extraits des noms des clés étrangères et l'on obtient comme noms de colonnes dans la vue des ventes :

COUNTRY_LABEL_HOME
COUNTRY_LABEL_BIRTH

7.2 Etape 2 (ACOD-BI)

Le dump de la source est monté dans un des environnements d'ACOD-BI.
ACOD-BI charge tous les fichiers CSV en entrée du système.
ACOD-BI génère l'ensemble des fichiers CSV en sortie du système.

7.3 Etape 3 (Administrateur) – Sélection des attributs et des indicateurs

7.3.1 (Obligatoire) Fichier des informations

(Obligatoire) L'administrateur du système indique dans le fichier des informations, les colonnes de la source correspondant à des attributs et/ou des indicateurs⁹.

(Facultatif) L'administrateur du système peut modifier la codification des attributs et des indicateurs^{10 11}.

7.3.2 (Facultatif) Fichier des entités

(Facultatif) L'administrateur du système peut modifier les modèles de code des entités du datamart^{12 13}.

Le poids relatif des indicateurs de l'entité, qui est dans le même fichier, sera traité à l'étape 5.

7.3.3 (Facultatif) Fichier des liens hiérarchiques

(Facultatif) L'administrateur du système peut modifier les modèles de code des codes différenciant des liens hiérarchiques^{14 15}.

7.4 Etape 4 (ACOD-BI)

ACOD-BI charge tous les fichiers CSV en entrée du système.
ACOD-BI génère l'ensemble des fichiers CSV en sortie du système.

⁹ Colonnes "NEW_ATTR_ACTIVATION" et "NEW_INDI_ACTIVATION" du fichier des informations.

¹⁰ Voir le chapitre « Codes du datamart » pour plus de détails.

¹¹ Colonnes "NEW_EXPLICIT_ATTR_CODE_MODEL" et "NEW_EXPLICIT_INDI_CODE_MODEL" du fichier des informations.

¹² Voir le chapitre « Codes du datamart » pour plus de détails.

¹³ Colonne "NEW_EXPLICIT_CODE_MODEL" du fichier des entités.

¹⁴ Voir le chapitre « Codes du datamart » pour plus de détails.

¹⁵ Colonne "NEW_EXPLICIT_DIFF_CODE_MODEL" du fichier des liens hiérarchiques.

7.5 Etape 5 (Administrateur) – Saisie des fréquences d'utilisation

7.5.1 (Obligatoire) Fichier des entités

(Obligatoire) L'administrateur du système indique dans le fichier des entités le poids relatif de chaque groupe d'indicateurs^{16 17}.

7.5.2 (Obligatoire) Fichier des cas d'utilisation

(Obligatoire) L'administrateur du système indique dans le fichier de cas d'utilisation les fréquences d'utilisation des attributs et/ou des branches d'attributs^{18 19}.

7.6 Etape 6 (ACOD-BI)

ACOD-BI charge tous les fichiers CSV en entrée du système.

ACOD-BI génère l'ensemble des fichiers CSV en sortie du système.

Les scripts de création et de chargement du datamart sont générés et déposés par ACOD-BI dans le répertoire de sortie du système.

7.7 Etape 7 (Administrateur) – Création et chargement du datamart

L'administrateur du système télécharge par FTP-S les scripts de création et de chargement du datamart depuis le répertoire de sortie d'ACOD-BI.

Il utilise ces scripts pour créer et charger le datamart dans son propre système²⁰.

Pour plus de détails voir « Annexe B – Création et chargement du datamart ».

¹⁶ Voir le chapitre « Fréquence d'utilisation des indicateurs » pour plus de détails.

¹⁷ Colonne "NEW EXPLICIT INDICATOR(S) WEIGHT" du fichier des entités.

¹⁸ Voir le chapitre « Fréquence d'utilisation des attributs » pour plus de détails.

¹⁹ Colonnes "NEW_EXPLICIT_ATTRIBUT_UF" et "NEW_EXPLICIT_BRANCH_UF" du fichier des cas d'utilisation.

²⁰ Les scripts générés sont lisibles et modifiables manuellement.

Par exemple, le script de chargement fonctionne en mode annule et remplace global. Il peut être modifié pour fonctionner dans un autre mode, au moins sur les tables de faits les plus volumineuses.

8 Statistiques d'utilisation des données

8.1 Fréquence d'utilisation des indicateurs

ACOD-BI n'a pas besoin de connaître la fréquence d'utilisation de chaque indicateur dans les restitutions. Chaque attribut et chaque indicateur appartient à une entité. On appelle « Groupe d'indicateurs » les entités porteuses d'au moins un indicateur.

ACOD-BI a seulement besoin de connaître le poids relatif de chaque groupe d'indicateurs dans les restitutions.

L'administrateur du système peut donner un poids relatif à chaque groupe d'indicateurs dans le fichier des entités²¹.

Exemple

Source

La source contient deux tables de faits, les commandes et les factures, chacune porteuse d'un montant et d'une quantité.

Les valeurs analysées sont les montants et les quantités.

Restitutions

Au cours d'une journée standard les restitutions suivantes sont exécutées :

- 3 restitutions sur le montant de facture,
- 1 restitution sur la quantité de facture,
- 5 restitutions sur le montant et la quantité de facture,
- 4 restitutions sur le montant et la quantité de commande.

ACOD-BI

Le référentiel contient les éléments suivants :

- Une entité (Facture) avec deux indicateurs (montant de facture et quantité de facture).
L'entité Facture est un groupe d'indicateurs.
- Une entité (Commande) avec deux indicateurs (montant de commande et quantité de commande).
L'entité Commande est un groupe d'indicateurs.

Fréquence d'utilisation

La fréquence d'utilisation du groupe d'indicateurs Facture est :

$$(3+1+5)/(3+1+5+4)$$

La fréquence d'utilisation du groupe d'indicateurs Commande est :

$$4/(3+1+5+4)$$

Pour traduire cela, il suffit de donner à Facture le poids 9 (3+1+5) et à Commande le poids 4.

²¹ Par défaut le poids d'un groupe d'indicateurs est égal au nombre d'indicateurs du groupe. Cependant, cela ne correspond généralement pas à la réalité.

8.2 *Fréquence d'utilisation des attributs*

8.2.1 Saisie au niveau attribut

ACOD-BI a besoin d'avoir une estimation de la fréquence à laquelle chaque attribut est utilisé avec chaque groupe d'indicateurs.

Grace aux poids des groupes d'indicateurs, on connaît déjà la fréquence d'utilisation des groupes d'indicateurs.

L'administrateur du système peut saisir pour chaque couple <attribut, groupe d'indicateurs> la fréquence d'utilisation dans les restitutions de l'attribut, pour une utilisation avérée du groupe d'indicateurs.

8.2.2 (Facultatif) Saisie au niveau branche

Pour des modèles complexes la saisie de toutes les fréquences d'utilisation <attribut, groupe d'indicateurs> peut s'avérer à la fois laborieuse et difficile à estimer.

C'est pourquoi ACOD-BI permet à l'administrateur du système de saisir la fréquence d'utilisation dans les restitutions d'une branche d'une hiérarchie d'attributs pour un groupe d'indicateurs.

Lorsque l'administrateur du système saisit une fréquence d'utilisation au niveau branche, les attributs de la branche sont impactés comme suit :

- Attributs pour lesquels l'administrateur A saisi une fréquence d'utilisation.
 - ➔ Aucun impact.
- Attributs pour lesquels l'administrateur N'A PAS saisi de fréquence d'utilisation.
 - ➔ Les fréquences d'utilisation sont ajustées automatiquement avec les règles suivantes :
 - la combinaison des fréquences d'utilisation de tous les attributs de la branche correspond à la fréquence d'utilisation saisie au niveau branche.
 - A chaque attribut est affectée une fréquence d'utilisation par défaut. Celle-ci sert de critère de pondération pour répartir la fréquence d'utilisation de la branche sur les attributs.

Exemple

Source

La source contient deux tables de faits, les commandes et les factures, chacune porteuse d'un montant et d'une quantité.

Les valeurs analysées sont les montants et les quantités.

Chaque facture est associée à une date, un produit et un client.

Chaque client a un grand nombre d'attributs. Les plus utilisés, de très loin, dans les restitutions sont sa tranche d'âge et sa région.

Chaque produit a comme attribut son libellé et sa couleur.

Restitutions

Au cours d'une journée standard 130 restitutions sont exécutées, réparties comme suit :

- 40 restitutions sur le montant et/ou la quantité de commande,
- 100 restitutions sur le montant et/ou la quantité de facture.
(10 restitutions portent à la fois sur les factures et les commandes)

Parmi les 100 restitutions portant sur les factures :

- 100 restitutions utilisent la date comme filtre ou critère de regroupement,
- 50 restitutions utilisent des informations du produit comme filtre ou critère de regroupement,
- 50 restitutions utilisent des informations du client comme filtre ou critère de regroupement.

Parmi les 50 restitutions utilisant des informations du client :

- 20 utilisent la tranche d'âge
- 30 utilisent la région
(Parmi elles certaines utilisent les 2)

Fréquences d'utilisation

Pour traduire ce cas, l'administrateur du système saisit les informations qui suivent.

Poids des groupes d'indicateurs :

- Commande : 40
- Facture : 100

Fréquence d'utilisation de l'attribut date pour les factures : 1.00

Fréquence d'utilisation de la branche produit pour les factures : $50/100=0.5$

Cette fréquence va être distribuée sur tous les attributs de la branche en utilisant leur fréquence d'utilisation par défaut comme critère de pondération.

Fréquence d'utilisation de la branche client pour les factures : $50/100=0.5$

Fréquence d'utilisation de l'attribut tranche d'âge pour les factures : $20/100=0.2$

Fréquence d'utilisation de l'attribut région pour les factures : $30/100=0.3$

Les fréquences d'utilisation pour les factures des autres attributs de la branche client vont être ajustées de manière à ce que la combinaison des fréquences d'utilisation de tous les attributs de la branche donne 0.5. Leur fréquence d'utilisation par défaut est utilisée comme critère de pondération.

Remarque

ACOD-BI traite l'utilisation de deux attributs comme deux évènements dé corrélés.

Ceci permet d'avoir une relation entre la fréquence d'utilisation d'une branche et la fréquence d'utilisation de chacun des attributs²².

Attention aux incohérences

En cas de saisie de fréquence d'utilisation d'une branche, il peut y avoir une incohérence dans les deux cas suivants :

- 1) L'administrateur du système saisit une fréquence d'utilisation sur tous les attributs et la fréquence d'utilisation de la branche est différente de la combinaison des fréquences d'utilisation des attributs.
C'est pourquoi il est déconseillé de saisir une fréquence d'utilisation sur une branche et sur tous ses attributs.
- 2) La fréquence d'utilisation de la branche est inférieure à la combinaison des fréquences d'utilisation d'attributs saisies.

En cas d'incohérence, une erreur est levée.

La ligne du fichier des cas d'utilisation introduisant l'incohérence n'est pas chargée et l'erreur est enregistrée dans la table des logs.

Exemple

Ligne N du fichier des cas d'utilisation.

Fréquence d'utilisation de la branche client pour les factures : 0.7

Ligne N+1 du fichier des cas d'utilisation.

Fréquence d'utilisation de l'attribut tranche d'âge du client pour les factures : 0.5

Ligne N+2 du fichier des cas d'utilisation.

Fréquence d'utilisation de l'attribut région pour les factures : 0.5

Lors de la prise en compte de la ligne N+2 du fichier des cas d'utilisation, la combinaison des fréquences d'utilisation d'attributs saisies sur la branche client = $1 - (1 - 0.5) \times (1 - 0.5) = 0.75 > 0.7$

Le chargement de la ligne N+2 du fichier des cas d'utilisation introduirait une incohérence.

La ligne N+2 du fichier des cas d'utilisation n'est pas prise en compte.

La fréquence d'utilisation de l'attribut région pour les factures reste libre.

Un message d'erreur est ajouté à la table de log.

²² Fréquence d'utilisation d'une branche = 1 – Produit sur tous les attributs(1-fréquence d'utilisation).

9 Codes du datamart

3 types de codes permettent de construire les noms des objets du datamart :

- Les codes entité
Ils permettent de construire les noms de la plupart des tables.
- Les codes information
Ils permettent de construire les noms de la plupart des colonnes ainsi que de certaines tables. Ils correspondent à un attribut ou un indicateur.
- Les codes différenciants de lien hiérarchique.
Ils permettent de distinguer des liens hiérarchiques ayant la même entité mère et la même entité fille. Ils servent de complément aux codes informations pour nommer certaines colonnes.

Ces codes doivent respecter les règles suivantes :

- Unicité des codes entité,
- Unicité des codes information,
- Unicité des codes différenciants pour deux liens hiérarchiques ayant les mêmes entités mère et fille.
- Taille des codes entité ≤ 4
- Taille des codes information ≤ 19
- Taille des codes différenciants de lien hiérarchique ≤ 4

Le principe de construction des codes est le même pour chacun des 3 types.

Une fonction permet de déterminer une valeur par défaut pour le modèle de code à partir de noms d'objets dans la source (voir les paramètres globaux FUNC_NTAB_CODE2ENTI_CODE, FUNC_NCOL_CODE2INFO_CODE et FUNC_NFK_CODE2HILI_DIFF_CODE).

Cette valeur par défaut peut être écrasée par la saisie explicite d'un modèle de code (fichier « exc_enti.csv » colonne NEW_EXPLICIT_CODE_MODEL, fichier « exc_info.csv » colonnes NEW_EXPLICIT_ATTR_CODE_MODEL et NEW_EXPLICIT_INDI_CODE_MODEL, fichier « exc_hili.csv » colonne NEW_EXPLICIT_DIFF_CODE_MODEL).

Le modèle de code obtenu est transformé (si nécessaire) en un code qui respecte les règles d'unicité et de taille.

10 Annexe A – Structure des fichiers CSV

Le format des nombres est sans séparateur de milliers, avec séparateur décimal ".".

10.1 Fichier des paramètres globaux

Le nom du fichier est « exc_param.csv ».

Les colonnes du fichier des paramètres globaux sont :

CODE

Code permettant d'identifier le paramètre global.

VALUE_TYPE

Type de données (String ou Float)

OLD_VALUE

Valeur du paramètre avant chargement du fichier dans le système.

DEFAULT_VALUE

Valeur par défaut du paramètre.

OLD_VALUE_IS_DEFAULT

Indique si la valeur dans le système avant chargement du fichier est la valeur par défaut.

NEW_VALUE

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

DESCRIPTION

Description du paramètre.

10.2 Fichier des informations

Le nom du fichier est « exc_info.csv ».

Les colonnes du fichier sont :

DWH_TABLE

Nom de la table de la source contenant la colonne correspondant à l'information.

DWH_COLUMN

Nom de la colonne de la source correspondant à l'information.

OLD_ENTITY_DTM_CODE

Nom de l'entité du datamart contenant l'information.

ATTRIBUT_ENABLED

Indique si l'information peut être utilisée comme attribut.

OLD_ATTR_ACTIVATION

Indique si l'information est un attribut.

Valeur avant chargement du fichier dans le système.

(Obligatoire) NEW_ATTR_ACTIVATION

Indique si l'information est un attribut.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

OLD_ATTR_DEFAULT_UF

Valeur par défaut de la fréquence d'utilisation pour cet attribut²³.

Valeur avant chargement du fichier dans le système.

(Facultatif) NEW_ATTR_DEFAULT_UF

Valeur par défaut de la fréquence d'utilisation pour cet attribut²⁴.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

OLD_ATTR_DTM_CODE

Code information de l'attribut.

Valeur avant chargement du fichier dans le système.

OLD_ATTR_CODE_MODEL

Modèle ayant permis de construire le code information de l'attribut.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_ATTR_CODE_MODEL

Modèle permettant de construire le code information de l'attribut saisi par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

²³ Pour plus de détails voir « Fréquence d'utilisation des attributs et des indicateurs ».

²⁴ Pour plus de détails voir « Fréquence d'utilisation des attributs et des indicateurs ».

(Facultatif) NEW_EXPLICIT_ATTR_CODE_MODEL

**Modèle permettant de construire le code information de l'attribut saisi par l'administrateur²⁵.
Remplace la valeur trouvée par la fonction de détermination du modèle de code information.
Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.**

INDICATOR_ENABLED

Indique si l'information peut être utilisée comme indicateur.

OLD_INDI_ACTIVATION

Indique si l'information est un indicateur.

Valeur avant chargement du fichier dans le système.

(Obligatoire) NEW_INDI_ACTIVATION

Indique si l'information est un indicateur.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

OLD_INDI_AGREG_RULE

Règle d'agrégation de l'indicateur.

Valeur avant chargement du fichier dans le système.

(Obligatoire) NEW_INDI_AGREG_RULE

Règle d'agrégation de l'indicateur.

Les valeurs possibles sont MIN, MAX, SUM.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

OLD_INDI_DTM_CODE

Code information de l'indicateur.

Valeur avant chargement du fichier dans le système.

OLD_INDI_CODE_MODEL

Modèle ayant permis de construire le code information de l'indicateur.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_INDI_CODE_MODEL

Modèle permettant de construire le code information de l'indicateur saisi par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

(Facultatif) NEW_EXPLICIT_INDI_CODE_MODEL

Modèle permettant de construire le code information de l'indicateur saisi par l'administrateur²⁶.

Remplace la valeur trouvée par la fonction de détermination du modèle de code information.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

²⁵ Voir le chapitre « Codes du datamart » pour voir la construction d'un code à partir d'un modèle de code.

²⁶ Voir le chapitre « Codes du datamart » pour plus de détails.

10.3 Fichier des entités

Le nom du fichier est « exc_enti.csv ».

Les colonnes du fichier sont :

DWH_MAIN_TABLE

Table principale de la source correspondant à l'entité.

DWH_TABLE(S)

Liste des tables de la source correspondant à l'entité.

OLD_DTM_CODE

Code de l'entité.

Valeur avant chargement du fichier dans le système.

OLD_CODE_MODEL

Modèle ayant permis de construire le code de l'entité.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_CODE_MODEL

Modèle permettant de construire le code de l'entité saisi par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

(Facultatif) NEW_EXPLICIT_CODE_MODEL

Modèle permettant de construire le code de l'entité saisi par l'administrateur²⁷.

Remplace la valeur trouvée par la fonction de détermination du modèle de code entité.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

OLD INDICATOR(S) WEIGHT

Poids relatif des indicateurs de l'entité dans les restitutions.

Valeur avant chargement du fichier dans le système.

OLD EXPLICIT INDICATOR(S) WEIGHT

Poids relatif des indicateurs de l'entité dans les restitutions saisi par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

(Obligatoire) NEW EXPLICIT INDICATOR(S) WEIGHT

Poids relatif des indicateurs de l'entité dans les restitutions saisi par l'administrateur²⁸.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

²⁷ Voir le chapitre « Codes du datamart » pour plus de détails.

²⁸ Voir le chapitre « Fréquence d'utilisation des indicateurs » pour plus de détails.

10.4 Fichier des liens hiérarchiques

Le nom du fichier est « exc_hili.csv ».

Les colonnes du fichier sont :

DWH_FK

Clé étrangère de la source correspondant au lien hiérarchique.

CONFLICT

Indique s'il existe au moins un autre lien hiérarchique avec les mêmes entités fille et mère.

CONFLICT_LIST

Liste des liens hiérarchiques avec les mêmes entités fille et mère.

OLD_DTM_DIFF_CODE

Code différenciant.

Valeur avant chargement du fichier dans le système.

OLD_DIFF_CODE_MODEL

Modèle ayant permis de construire le code différenciant du lien hiérarchique.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_DIFF_CODE_MODEL

Modèle permettant de construire le code différenciant du lien hiérarchique saisi par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

(Facultatif) NEW_EXPLICIT_DIFF_CODE_MODEL

Modèle permettant de construire le code différenciant du lien hiérarchique saisi par l'administrateur²⁹.

Remplace la valeur trouvée par la fonction de détermination du modèle de code différenciant du lien hiérarchique.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

DWH_TABLE

Table fille de la clé étrangère dans la source.

DWH_COLUMN(S)

Liste des colonnes de la clé étrangère dans la table fille dans la source.

DWH_R_UK

Nom de la contrainte unique référencée par la clé étrangère dans la source.

DWH_R_TABLE

Nom de la table référencée par la clé étrangère dans la source.

DWH_R_COLUMN(S)

Nom des colonnes référencées par la clé étrangère dans la source.

²⁹ Voir le chapitre « Codes du datamart » pour plus de détails.

10.5 Fichier des cas d'utilisation

Le nom du fichier est « exc_uc.csv ».

Dans les descriptions qui suivent :

- « branche d'attributs pour le groupe d'indicateurs » est abrégé en « branche ».
- « fréquence d'utilisation pour le groupe d'indicateurs » est abrégée en « fréquence d'utilisation ».

Les colonnes du fichier sont :

INDICATOR_ENTITY

Code de l'entité contenant les indicateurs (appelé aussi groupe d'indicateurs)

ATTRIBUT_ENTITY

Code de l'entité contenant les attributs.

ATTRIBUT_CODE

Code de l'attribut.

HIERARCHY

Hiérarchie permettant de naviguer depuis le groupe d'indicateurs vers l'entité des attributs.

BRANCH_UF_ENABLED

Indique si la saisie de la fréquence d'utilisation de la branche est autorisée.

OLD_BRANCH_UF

Fréquence d'utilisation de la branche.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_BRANCH_UF

Fréquence d'utilisation de la branche saisie par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

(Obligatoire) NEW_EXPLICIT_BRANCH_UF

Fréquence d'utilisation de la branche saisie par l'administrateur³⁰.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

ATTRIBUT_UF_ENABLED

Indique si la saisie de la fréquence d'utilisation de l'attribut est autorisée.

OLD_ATTRIBUT_UF

Fréquence d'utilisation de l'attribut.

Valeur avant chargement du fichier dans le système.

OLD_EXPLICIT_ATTRIBUT_UF

Fréquence d'utilisation de l'attribut saisie par l'administrateur lors d'une itération précédente.

Valeur avant chargement du fichier dans le système.

³⁰ Voir le chapitre « Fréquence d'utilisation des attributs » pour plus de détails.

(Obligatoire) NEW_EXPLICIT_ATTRIBUT_UF

Fréquence d'utilisation de l'attribut saisie par l'administrateur³¹.

Nouvelle valeur à prendre en compte par ACOD-BI au chargement du fichier.

INDI_DWH_MAIN_TABLE

Table principale de la source associée au groupe d'indicateurs.

INDI_DWH_TABLE(S)

Liste des tables de la source associées au groupe d'indicateurs.

FK_HIERARCHY_LIST

Liste des clés étrangères de la source permettant de naviguer des indicateurs vers les attributs.

ATTR_ENTI_DWH_MAIN_TABLE

Table principale de la source associée à la branche d'attributs ou à l'attribut.

ATTR_ENTI_DWH_TABLE(S)

Liste des tables de la source associées à la branche d'attributs ou à l'attribut.

ATTR_DWH_TABLE

Table de la source contenant la colonne correspondant à l'attribut.

ATTR_DWH_COLUMN

Colonne de la source correspondant à l'attribut.

³¹ Voir le chapitre « Fréquence d'utilisation des attributs » pour plus de détails.

11 Annexe B – Création et chargement du datamart

11.1 Paramètres de l'instance

L'instance Oracle sur laquelle le datamart va être installé doit avoir les paramètres suivants :

- QUERY_REWRITE_ENABLED=TRUE
- QUERY_REWRITE_INTEGRITY=STALE_TOLERATED
- SKIP_UNUSABLE_INDEXES=TRUE
- STAR_TRANSFORMATION_ENABLED=TRUE

11.2 Droits du propriétaire du datamart

L'utilisateur Oracle, propriétaire du schéma du datamart doit avoir les droits système suivants :

- CREATE SESSION,
- ALTER SESSION
- CREATE CLUSTER,
- CREATE DATABASE LINK,
- CREATE SEQUENCE,
- CREATE SYNONYM,
- CREATE TABLE,
- CREATE VIEW,
- CREATE PROCEDURE,
- CREATE TRIGGER,
- CREATE MATERIALIZED VIEW,
- CREATE DIMENSION.

Il doit aussi disposer de quotas suffisants sur les tablespaces³² du datamart.

11.3 Package d'outils d'ACOD-BI

L'administrateur doit créer le package d'outils PKG_ACOD_TOOLS dans le schéma du datamart.

La méthode la plus simple est de se connecter avec le propriétaire du datamart et d'exécuter les scripts suivants :

- PKG_ACOD_TOOLS_decl.sql
- PKG_ACOD_TOOLS_body.sql

11.4 Création du datamart

L'administrateur doit créer les structures du datamart dans son schéma.

La méthode la plus simple est de se connecter avec le propriétaire du datamart et d'exécuter le script cre_dtm_<code identifiant du datamart>.sql.

11.5 Chargement du datamart

L'administrateur doit charger les données du datamart.

La méthode la plus simple est de se connecter avec le propriétaire du datamart et d'exécuter le script load_dtm_<code identifiant du datamart>.sql.

³² Pour plus de détails voir les chapitres « Noms des tablespaces du datamart » et « Espace disque cible pour le datamart ».