

ACOD-BI

Création Automatique de Datamarts Optimisés

Présentation

Table des matières

1	Introduction.....	3
2	Architecture.....	4
3	Evolutions à court terme	6
3.1	Jeu de requêtes en entrée.....	6
3.2	Utilisation d'autres technologies.....	6
4	Evolution à moyen terme - Création et utilisation transparente du datamart.....	6
5	Base d'exemple.....	6

1 Introduction

Nature de l'outil

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

Spécificités

ACOD-BI se distingue des autres outils de sa catégorie par sa capacité à identifier les agrégats optimaux sur des modèles complexes avec en cible, un espace de stockage limité.

Cette capacité lui permet, une fois les informations nécessaires récoltées, de générer le datamart cible de bout en bout, sans que l'intervention d'un administrateur soit nécessaire pour orienter ses choix.

A terme, l'intégration de cette fonctionnalité dans l'optimiseur d'un SGBD ou dans un outil de restitution permettra de rendre la création et l'utilisation des datamarts transparentes pour les utilisateurs. L'outil améliorera les performances de manière drastique pour les requêtes d'agrégation, sans nécessiter de travail d'administration, pour un coût d'espace disque relativement faible.

Génération

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.

Données en entrée

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

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

Structure des datamarts créés par ACOD-BI

Dans sa version actuelle, ACOD-BI travaille uniquement sur Oracle.

Les datamarts créés 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 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 Architecture

ACOD-BI est composé principalement de 3 modules indépendants communiquant entre eux via un référentiel.

Ce découpage en modules est destiné à faciliter les évolutions de l'outil.

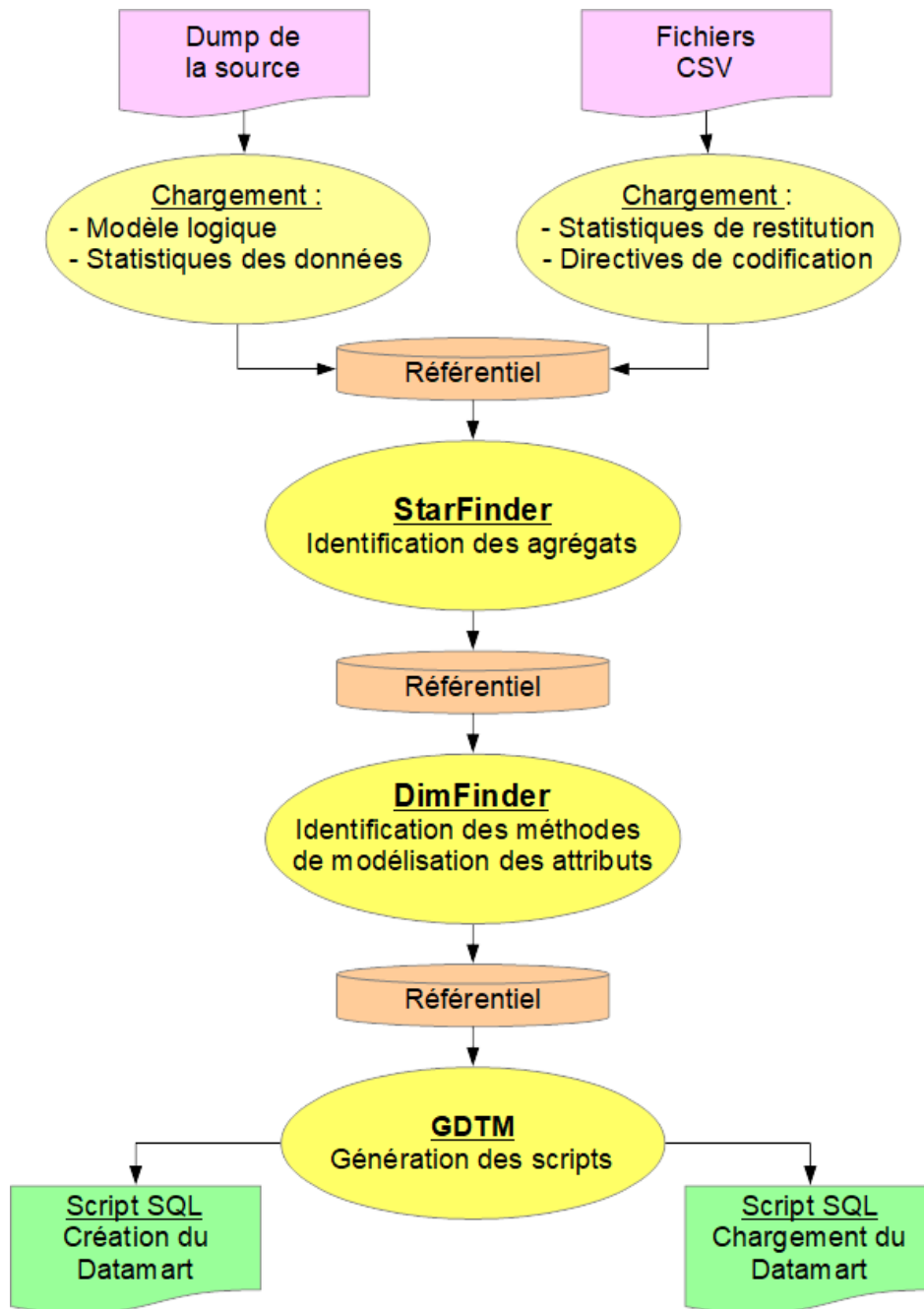


Figure 1 - Architecture logique de ACOD-BI

StarFinder

Le module StarFinder détermine quels niveaux d'agrégation sur les tables de faits sont les plus pertinents pour obtenir les meilleures performances en restitution.

En termes de choix des agrégats, la sélection effectuée par ACOD-BI est optimale. On ne peut pas faire mieux (Plus précisément, l'écart entre la sélection théoriquement idéale et la sélection du moteur est négligeable).

L'algorithme principal de StarFinder est indépendant de la technologie utilisée.

DimFinder

Le module DimFinder détermine comment les attributs doivent être physiquement modélisés dans le datamart généré.

DimFinder est faiblement dépendant du SGBD-R.

GDTM (Generate datamart)

Le module GDTM génère le script de création et le script d'alimentation du datamart.

GDTM est dépendant du SGBD-R.

Chargement du référentiel

En complément de ces 3 modules, des programmes de chargement permettent d'alimenter le référentiel.

Les données référentielles chargées sont :

- le modèle logique des données, déduit des métadonnées de la source³,
- les statistiques sur les données, déduites des métadonnées de la source,
- les statistiques sur l'utilisation des données dans les restitutions, fournies par fichier csv.
- les directives sur les noms de tables et de colonnes, fournies par fichier csv⁴.

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.

⁴ Pour plus de détails voir le « Manuel d'utilisation » chapitre « Codes du datamart ».

3 Evolutions à court terme

3.1 Jeu de requêtes en entrée

Avec cette évolution, ACOD-BI peut prendre en entrée un jeu de requêtes à optimiser.

Ce jeu de requêtes peut, au choix :

- s'ajouter aux statistiques sur l'utilisation des données dans les restitutions,
- remplacer les statistiques sur l'utilisation des données dans les restitutions.

3.2 Utilisation d'autres technologies

ACOD-BI peut être adapté à d'autres SGBDR ou à d'autres types de bases que les bases relationnelles.

Le portage sur d'autres technologies qu'Oracle nécessite d'avoir un mécanisme permettant de rediriger les requêtes de niveau détail vers l'agrégat optimum.

Ce mécanisme peut être mis en place par une intégration ou un interfaçage avec le moteur du SGBDR ou avec l'outil de restitution.

Remarque : ACOD-BI dispose déjà d'un mécanisme de sélection du (des) agrégat(s) les plus pertinents pour une requête simple sur le datamart. Ce mécanisme n'est pas utilisé dans la version actuelle.

4 Evolution à moyen terme - Création et utilisation transparente du datamart

L'objectif est de rendre la création et l'utilisation des datamarts quasi-transparentes.

Le mode de fonctionnement sera le suivant :

- Les requêtes utilisateur sont faites directement sur le datawarehouse ou le système opérationnel.
- Les statistiques d'utilisation des données sont déduites de l'analyse des requêtes sur le datawarehouse ou le système opérationnel.
- Le DBA active la création du datamart. Il indique le volume souhaité.
- Les requêtes utilisateur sont redirigées vers le datamart de manière transparente.

5 Base d'exemple

Oracle 12c

Le **temps d'exécution** d'un jeu de requêtes d'agrégation est **divisé par 20** pour un **surcoût d'espace disque de 42%** en utilisant **ACOD-BI** par rapport à **datamart basique**.

Il est **divisé par 3000** pour un **surcoût d'espace disque de 68%** par rapport à une utilisation directe du **datawarehouse**.

Oracle Autonomous Data Warehouse

Le **temps d'exécution** est **divisé par 15** en utilisant **ACOD-BI** par rapport à un **datamart basique**.

Il est **divisé par 45** par rapport à une utilisation directe du **datawarehouse**.

Pour plus de détails voir <http://acod-bi.com/exemple>.