ACOD-BI

Automatic Creation of Optimized Datamarts

Presentation

Table of Contents

1 Introduction

Tool nature

ACOD-BI automatically generates optimized datamarts in a limited space storage.
It identifies optimum aggregates from reporting statistics.

Specificity

ACOD-BI differs from other tools of its category by its capability to identify the most optimum aggregates on complex models within a limited space storage.
This capability allows it, once the necessary data are collected, to generate the target datamart from end to end, without any action from an administrator to guide its choices.

**Ultimately, the integration of this feature into a DBMS optimizer or into a reporting tool will make the creation and use of datamarts invisible for users.
The tool will drastically improve performance for summary queries with no administrative work for a relatively low disk space cost.**

Generation

In its current version, ACOD-BI allows you to generate a datamart in a few hours instead of a few weeks for manual creation.

The generated datamart is provided in the form of two readable and editable SQL scripts. Once the datamart is generated you are no longer dependent on the tool.

Input data

ACOD-BI defines the target datamart structure mainly from the following data:

- Logical data model,
- Data statistics,
- Statistics on data utilization in reporting,
- Space storage granted to the datamart.

Structure of datamarts created by ACOD-BI

In its current version, ACOD-BI runs exclusively on Oracle.

Generated datamarts are modeled in Star Schema.
The most pertinent aggregates are added to the detail fact tables.

For every detail fact table, a unique view gathers indicators[1] and attributes[2] from the fact table and its associated dimensions.
The user builds queries on these views without worrying about how the dimension tables are modeled and without being impacted in case of datamart model evolution.

The system creates the additional required structures to allow the Oracle QUERY_REWRITE mechanism to automatically use aggregates.

[1] An indicator is an information on which reporting applies an aggregation function (SUM, MIN and MAX).

[2] An attribute is an information used by reporting in filters or in group by clause.

2 Architecture

ACOD-BI is mainly composed of 3 independent modules communicating via a repository.
This split into modules makes easier the tool evolutions.

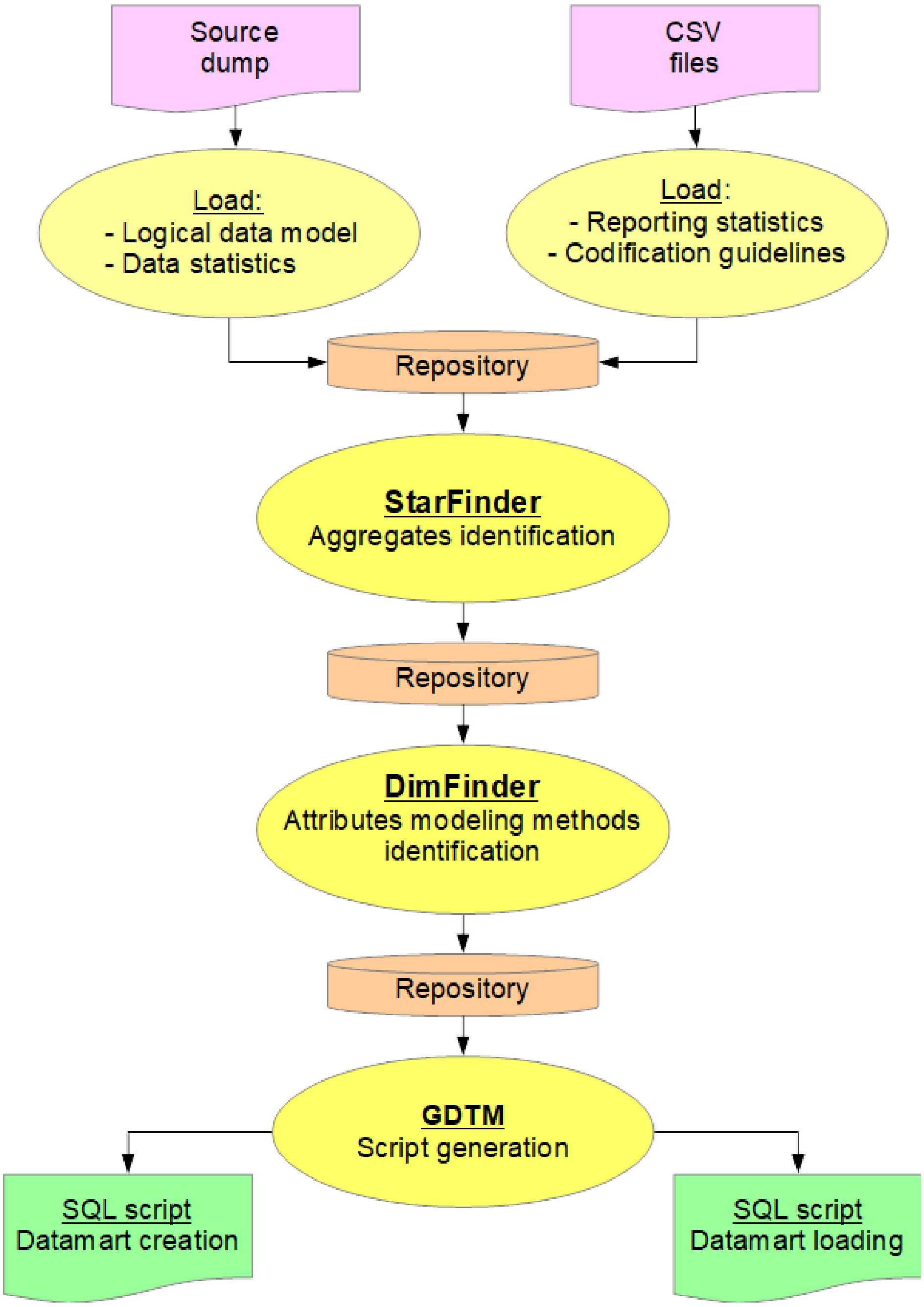

Figure 1 - ACOD-BI logical architecture

StarFinder

The StarFinder module determines which aggregate levels on fact tables are the most pertinent to obtain the best performance in reporting.

StarFinder makes an optimum aggregates selection: it is impossible to do better (More precisely, the difference between the theoretically ideal selection and StarFinder selection is negligible).

The StarFinder main algorithm is independent of the technology used.

DimFinder

The DimFinder module determines how the attributes must be physically modeled in the generated datamart.

DimFinder is weakly dependent on the RDBMS.

GDTM (Generate datamart)

The GDTM module generates the datamart creation script and the load script.

GDTM is dependent on the RDBMS.

Repository loading

In addition to this 3 modules, programs load the following data in the repository:

- Logical data model, deduced from source[3] meta-data,
- Data statistics, deduced from source meta-data,
- Statistics on data utilization in reporting, provided by csv files,
- Instructions about tables and column names, provided by csv files[4].

[3] In the most standard cases the source is a data warehouse.
Loading data directly from the operational system can be done if the following conditions are satisfied:

- The operational system data model is close to the third normal form,
- There is a long enough time window during which the operational system is unused to load the datamart.

[4] For more details refer to the « User manual » chapter « Datamart codes ».

3 Short-term evolutions

3.1 *Query set input*

ACOD-BI could accept as input a query set to optimize.
This query set could, at choice:

- Be added to the statistics on data utilization in reporting,
- Replace the statistics on data utilization in reporting.

3.2 *Use of other technologies*

ACOD-BI can be ported to other RDBMS or other types of databases than relational databases.

Porting to technologies other than Oracle requires a mechanism to redirect detail level queries to optimum aggregate.

This mechanism can be implemented by integrating or interfacing ACOD-BI with the RDBMS engine or with the reporting tool.

Note: ACOD-BI already has a mechanism for selecting the most relevant aggregate(s) for a simple query on the datamart. This mechanism is not used in the current version.

4 Medium-term evolution - Creation and use of datamarts invisible for users

The goal is to make creation and use of datamarts virtually invisible.

The system will work as follows:

- User queries are made directly on data warehouse or operating system.
- Statistics on data utilization in reporting are deduced from query analysis.
- The DBA activates the creation of datamart. He enters the desired datamart disk space.
- User queries are redirected to datamart automatically.

5 Example database

Oracle 12c
Execution time of a set of summary queries is **divided by 20** for an **additional disk space usage of 42%** using **ACOD-BI compared to a basic datamart**.
It is **divided by 3000** for an **additional disk space usage of 68%** using **ACOD-BI compared to** a direct use of the **data warehouse**.

Oracle Autonomous Data Warehouse
Execution time is **divided by 15** using **ACOD-BI compared to a basic datamart**.
It is **divided by 45 compared to** a direct use of the **data warehouse**.

For more details refer to http://acod-bi.net/example.