Chapter LXXVI
GR-OLAP:
Online Analytical Processing of Grid Monitoring Information

Julien Gossa
LIRIS–INSA Lyon, France

Sandro Bimonte
LIRIS–INSA Lyon, France

INTRODUCTION

The Grid is an emerging solution for sharing resources through a network. It is meant to manage heterogeneous resources in world-scale multi-institutional networks. Grid resources monitoring and network monitoring are very active research areas with actually efficient solutions. Unfortunately, these solutions are limited in terms of analysis of the gathered data. Our proposition is to use data warehouse (DW) and online analytical processing (OLAP) technologies on Grid monitoring information. This allows new complex analyses of crucial importance for Grid users' everyday tasks. Unfortunately, the implementation raises several challenging issues.

This article is organized as follows. First, we introduce concepts of DW, OLAP, and Grids, and we discuss recent advances in Grid monitoring as well as the needs and usage of the Grid users. Then we present our conceptual and implementation solutions. Finally, we discuss our main contribution and point out the future works.

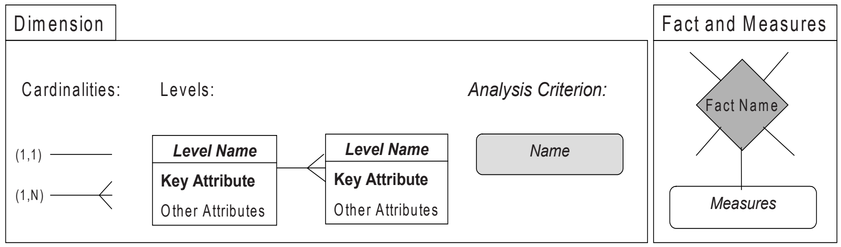
BACKGROUND: DATA WAREHOUSE & OLAP

Definition and Usage

DW in addition to OLAP technologies intend to be an innovative decision support for business intelligence and knowledge discovery. It has now become a leading topic in business organizations as well as in the research community. The main motivation is to take benefits from the enormous amount of data in distributed and heterogeneous databases to enhance data analysis and decision making (Kimball & Ross, 2002).

A DW is a subject-oriented, integrated, nonvolatile, and time-variant collection of data stored in a single site repository and collected from multiple

Figure 1. MultiDimEr conceptual multidimensional model



sources (Inmon, 1996). Information in the DW is organized following a multidimensional model in order to allow precomputation and fast access to summarized data in support of management's decisions. This multidimensional model organizes data in analysis axes called dimensions. Analyzed subjects or facts are characterized by metrics called measures. Dimensions can be organized following hierarchy schemas, thus allowing navigation through different levels of detail of analysis.

An OLAP server calculates and optimizes the hypercube, that is, the set of fact values for all combinations of dimension instances (called members). In order to optimize accesses to the data, query results are precalculated in the form of aggregates. This allows the decision makers to explore the different dimensions at different granularities. This analysis process is conducted by navigating into the multidimensional cube through some OLAP operators (roll-up, drill-down, slice, rotate, etc.).

Finally, interactive user interfaces (OLAP clients) have been developed to support knowledge discovery, promoting the iterative nature of the analysis process. OLAP clients visually represent the multidimensional structure of the hypercube and formulize multidimensional queries. The most adopted data presentation paradigm is the pivot table, a 2-D spreadsheet with associated subtotals and totals. It supports complex data by nesting several dimensions on the *x*- or *y*-axis and displaying data on multiple pages.

The three-tier architecture composed of a DW, an OLAP server, and an OLAP client effectively allows multidimensional analysis.

In the following, we will use the conceptual multidimensional model MultiDimEr (Malinowski & Zimányi, 2006) in order to describe our proposal. The details of the model are presented in Figure 1.

BACKGROUND: THE GRID AND ITS MONITORING

The term *the Grid* was coined in the mid 1990s to denote a paradigm of distributed computing infrastructure for advanced science and engineering (Foster & Kesselman, 1994).

A Glimpse at the Grid Computing

The real and specific problem that underlies the Grid concept can be summarized as "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations" (Foster, Kesselman, & Tuecke, 2001). This sharing concerns direct access to resources (e.g., computers or hosts, software, data, etc.) as required by a range of collaborative problem-solving and resource brokering strategies emerging in industry, science, and engineering. It needs high control: Resource providers and consumers must define clearly and carefully what is shared, who is allowed to share what, and the conditions of the sharing. A set of individuals and/or institutions defined by such sharing rules form what is called a virtual organization (VO). An example of VO is the application service providers, storage service providers, cycle providers, and consultants engaged by a car manufacturer to perform scenario evaluation during planning for a new factory.

In such an information system, the monitoring definitively takes a crucial place.