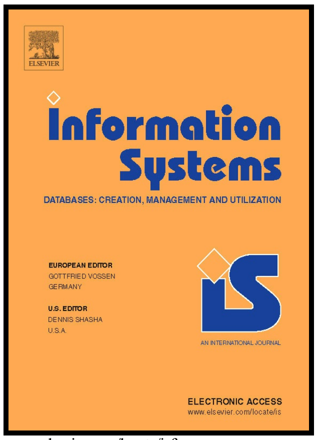

www.elsevier.com/locate/infosys

Eco-Physic: Eco-Physical Design Initiative for Very Large Databases

Amine Roukh, Ladjel Bellatreche, Selma Bouarar, Ahcene Boukorca

PII: S0306-4379(16)30512-9
DOI: http://dx.doi.org/10.1016/j.is.2017.01.003
Reference: IS1196

To appear in: *Information Systems*

Received date: 24 October 2016
Accepted date: 17 January 2017

Cite this article as: Amine Roukh, Ladjel Bellatreche, Selma Bouarar and Ahcen
Boukorca, Eco-Physic: Eco-Physical Design Initiative for Very Large Databases
Information Systems, http://dx.doi.org/10.1016/j.is.2017.01.003

Eco-Physic: Eco-Physical Design Initiative for Very Large Databases☆

Amine Roukh[a], Ladjel Bellatreche[b,*], Selma Bouarar[b], Ahcene Boukorca[b]

[a] *University of Mostaganem, Mostaganem, Algeria*
[b] *LIAS/ISAE-ENSMA, University of Poitiers, Poitiers, France*

Abstract

In the Big Data Era, the management of energy consumption by servers and data centers has become a challenging issue for companies, institutions, and countries. In data-centric applications, Database Management Systems are one of the major energy consumers when executing complex queries involving very large databases. Several initiatives have been proposed to deal with this issue, covering both the hardware and software dimensions. They can be classified in two main approaches assuming that either **(a)** the database is *already deployed* on a given platform, or **(b)** it is *not yet deployed*. In this study, we focus on the first set of initiatives with a particular interest in physical design, where optimization structures (e.g., indexes, materialized views) are selected to satisfy a given set of non-functional requirements such as query performance for a given workload. In this paper, we first propose an initiative, called *Eco-Physic*, which integrates the energy dimension into the physical design when selecting materialized views, one of the redundant optimization structures. Secondly, we provide a multi-objective formalization of the materialized view selection problem, considering two non-functional requirements: query performance and energy consumption while executing a given workload. Thirdly, an evolutionary algorithm is developed to solve the problem. This algorithm differs from the existing ones by being interactive, so that database administrators can adjust some energy sensitive parameters at the final stage of the algorithm execution according to their specifications. Finally, intensive experiments are conducted using our mathematical cost model and a real device for energy measurements. Results underscore the value of our approach as an effective way to save energy while optimizing queries through materialized views structures.

Keywords: Physical Design, Energy Efficiency, Power Management
2010 MSC: 00-01, 99-00

1. Introduction

Global energy consumption has not ceased to increase. Three key factors are behind this steady rise: **(1)** the ubiquity of the Worldwide Internet and the pervasiveness of mobile handsets, netbooks and laptop PCs. According to the Natural Resources Defense Council report, if the World-Wide Internet was a country, it would be the 12th-largest consumer of electricity in the world, between Spain and Italy [2]. **(2)** The high demand of users and decision makers to efficiently store, access and manage huge volumes of data emanating from sensors, social networks, experimentations, simulations, etc. of both everyday and analytical applications. The management of these data is usually performed by Database Management Systems (DBMSs). **(3)** The extensive usage of data centers as a *deployment platform* for such applications motivates companies owning these data to invest in expanding their use of such centers. Recently, Google announced details about the expansion of its data center by 500,000 square feet to provide a new mega data center over 1.3m square feet[1].

These three factors significantly increase the consumed energy and hence the carbon footprint of data centers. The augmentation of the energy consumption is mainly caused by the high power requirements of Information Technology (IT) equipment and cooling infrastructures. In a traditional data center, for every kilowatt of power used to drive a server, another kilowatt is needed to cool it [3, 4]. Industry experts, such as the SMARTer 2020, report that the global data center emissions will grow 7 percent every year through 2020 [5]. DBMSs are today one of the major challenges toward energy efficiency [6]. Figure 1 depicts the energy consumption distribution per component in a data center. It reveals that half of the total energy particularly results from *IT* equipment (CPU, memory, disk, etc.), which is a very significant part. However, identifying the power consumption behavior of these main components is the first step toward improving their energy efficiency.

Faced with this situation, energy saving becomes a

☆This work is an extension of our paper [1]

*Corresponding author.
Email addresses: roukh.amine@univ-mosta.dz (Amine Roukh), ladjel.bellatreche@ensma.fr (Ladjel Bellatreche), selma.bouarar@ensma.fr (Selma Bouarar), ahcene.boukorca@ensma.fr (Ahcene Boukorca)

[1] http://www.datacenterdynamics.com/design-build/