Journal of Network and Computer Applications 84 (2017) 118–130

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

A green framework for DBMS based on energy-aware query optimization and energy-efficient query processing

Binglei Guo[a,*], Jiong Yu[a], Bin Liao[b], Dexian Yang[a], Liang Lu[a]

[a] School of Information Science and Engineering, Xinjiang University, Urumqi 830046, PR China
[b] College of Statistics and Information, Xinjiang University of Finance and Economics, Urumqi 830012, PR China

ARTICLE INFO

Keywords:
Green database
Query optimization
Energy saving
Energy efficiency
Query-plan evaluation
Energy-aware

ABSTRACT

Traditional database systems result in high energy consumption and low energy efficiency due to the lack of consideration of energy issues and environmental adaptation in the design process. In this study, we report our recent efforts on this issue, with a focus on energy-aware query optimization and energy-efficient query processing. Firstly, a method of modeling energy cost of query plans during query processing based on their resource consumption patterns is proposed, which helps predict energy cost of queries before execution. Secondly, as the traditional query optimizer focuses on solely optimizing for performance and ignores energy-efficient query plans, a query-plan evaluation model is proposed after a comprehensive study of plan evaluation principles. Using the cost model as a basis, the evaluation model can utilizes the trade-offs between power and performance of plans, and helps the query optimizer select plans that meet performance requirements but result in lower energy cost. Finally, a green database framework integrated with the two above models is proposed to enhance a commercial DBMS. Experimental results reveal that, with reliable and accurate statistical data, the proposed framework in this study can achieve significant energy savings and improve energy efficiency.

1. Introduction

After the Copenhagen conference, constructing low-carbon society has become a global consensus. As databases are one of the major energy-consuming components in a data center, constructing green database systems to save energy and improve energy efficiency has become research hotspots in recent years. Energy management has been a significant aspect in designing and implementing green database systems. The arising of this new paradigm of taking energy consumption as a first-class optimization goal is driven by the following facts: 1) The proportion of electricity cost in today's large-scale data management systems (typically, a data center) is increasing year by year (Kushagra, 2010). In the United States, electricity consumption of data centers in 2011 accounted for 2% of the total national power grid (Tianwen, 2013). In the same year, energy consumption of data center in China accounted for 5% of the total national energy consumption, reached 70 billion kW h (China IDC, 2012); 2) Environmental problems caused by increasing energy consumption around the world has prompted the government to supervise the IT capacity of enterprises in the world (Webb, 2008). The emission of CO_2 in the field of IT industry accounted for 2% of the world and this proportion will be doubled by the year of 2020 (Global action plan, 2007); 3) Data centers cannot control the use of energy, and the cost of electricity has begun to exceed the cost of hardware (Rasmussen, 2011), which has a negative impact on the density, scalability and related environmental design of data centers (Peiquan et al., 2014; Ran et al., 2016); 4) Energy efficiency improvement of servers is significantly below expectations (Barroso and Hölzle, 2007; Graefe, 2008; Peiquan et al., 2014). The number of cores of processors is expected to double every 18 months while the case for performance per watt is at a slower rate of every two years (Belady, 2007).

Triggered by these facts, energy issues of database servers have recently draw attention from some well-known journals and conferences in the database field, such as EDBT 2008 (Graefe, 2008), CIDR 2009 (Harizopoulos et al., 2009; Lang and Patel, 2009; Hamilton, 2009), ICDE 2010 (Xu et al., 2010), SIGMOD 2010 (Xu, 2010; Tsirogiannis et al., 2010; Bjørling et al., 2010), VLDB 2010 (Canim et al., 2010), VLDB 2011 (Pelley et al., 2011), SSDBM 2011 (Tu et al., 2011), EDBT 2011 (Kim and Rotem, 2011), SIGMOD Record (Wang et al., 2011), IEEE Data Engineering Bulletin 2011 (Lang et al., 2011), VLDB 2012 (Xu et al., 2012; Lang et al., 2012), EDBT 2012 (Kunjir et al., 2012), SIGMOD 2013 (Do et al., 2013), ICDCS 2013 (Xu et al.,

* Corresponding author.
E-mail addresses: xinggbl@163.com (B. Guo), yujiong@xju.edu.cn (J. Yu), liaobin@xju.edu.cn, liaobin665@163.com (B. Liao), dhds250elf@sina.cn (D. Yang), luliang19891108@gmail.com (L. Lu).

http://dx.doi.org/10.1016/j.jnca.2017.02.015
Received 26 October 2016; Received in revised form 7 January 2017; Accepted 23 February 2017
Available online 27 February 2017
1084-8045/

2013a, 2013b), VLDB 2014 (He, 2014), and IEEE Transactions on Computers (Xu et al., 2015).

This study aims to propose an energy-aware DBMS with energy efficiency. It is clear that with these growing energy cost and energy-inefficient servers, the challenge here is to reduce energy cost and improve energy efficiency of DBMSs while maintain the performance levels that are accepted by users. A crucial aspect of our work that distinguishes it from previous work (Lang et al., 2011; Xu et al., 2015) is our focus on the trade-off available between performance and power during query optimization that have received little attention so far, and leveraging this trade-off to select plans that save energy for the overall system. Recent studies (Xu, 2010; Lang et al., 2011; Kunjir et al., 2012) have shown that, in a traditional database, there are energy-efficient query plans with little performance degradation. However these plans are very likely and mostly indeed ignored by traditional optimizers. As a consequence, DBMS can be energy-efficient by selecting those query plans. We verified this viewpoint by our experimental results shown in Section 6.3.

This study proposed a new way of optimizing and processing queries. Our contributions can be summarized as follows.

(1) We provide an extensive study of the effect of memory size and cache structures (i.e. cached data) on various costs of query processing. As a consequence, three main cache structures (Database Buffer Cache, Dictionary Cache, and Library Cache) in memory are associated with I/O and CPU resource consumption, which helps improve prediction accuracy of the energy cost model.
(2) We propose an accurate and portable energy cost model for DBMS in a static environment to predict energy cost of query processing and hence the DBMS can be energy-aware.
(3) We propose a simple but practical query-plan evaluation model based on an in-depth analysis of query optimization mechanisms. The evaluation model can be used to evaluate the superiority of alternative query plans towards a specific optimization goal. With the help of an accurate cost model, the query optimizer (embedded with the evaluation model) can take both power and performance into consideration and select plans with lower energy cost and higher energy efficiency.
(4) We propose a green database framework that integrated with the two above models. We utilize our framework to enhance a commercial DBMS with query workloads generated from TPC benchmarks. Numerical analysis and experimental results verified the effectiveness of the proposed framework.

The remainder of this study is organized as follows. Related works are surveyed in Section 2. We explore the relationship between main cache structures and query processing in Section 3. We describe the technical details of our energy cost model in Section 4. Next, we present the query-plan evaluation model in Section 5. Section 6 presents the numerical analysis and experimental results of our models. Finally, conclusions are drawn in Section 7.

2. Related work

The primary goal of traditional database is performance, which means to run the query faster without taking energy cost into consideration. Therefore, the goal of green databases has been shifted to energy-aware query processing and query optimization. To realize the goal, the query optimizer should be energy-aware of all the plans of a query and be able to select the desired plan that saves energy during query processing.

Current energy-saving technologies for database systems are generally focused on the hardware and software. On the hardware side, existing work focused on utilizing new abilities of modern hardware, such as the modern processors, that process at a lower voltage and frequency. Study (Chen and Kuo, 2007) achieved energy efficiency by leveraging the Dynamic Voltage Scaling (DVS) technique to trade performance with energy. Study (Lang and Patel, 2009) used the Processor Voltage/Frequency Control (PVC) technique that lets DBMS change the processor voltage and frequency. Study (Meza et al., 2009) achieved 45% power saving while sacrificing 5% performance degradation by reallocating the database across fewer disks and turning off the unused ones. Studies (Schall et al., 2010; Canim et al., 2010; Pelley et al., 2011; Beckmann et al., 2011; Bausch et al., 2012; Do et al., 2013; He, 2014) reduced the total energy consumption of database system and improve energy efficiency by using energy-efficient hardware devices (e.g., SSDs).

On the software side, most work concentrated on building energy cost model for database query workload. Study (Xu et al., 2013a, 2013b) built a series of physical models of individual relational operators in a query plan based on their time and power cost per operation. By taking CPU and disk as the main energy-consuming components in a database system, study (Xu et al., 2015) utilized the traditional performance driven optimizer to build an energy cost model for the query plans. Similar with (Xu et al., 2013a, 2013b, 2015), study (Liu et al., 2013) built a power cost model by utilizing the product of the number of tuples indexed by CPU and the number of columns retrieved. By extracting common features of queries (such as selectivity factors and tuple sizes), study (Rodriguez-Martinez et al., 2011) built a peak power cost model in a clustered environment. Study (Lianghuai and Hongyan, 2014) proposed a component-level power modeling method that investigated into individual cores to cope with the multi-core architecture by exploiting the utilization of the main components (CPU and disk). Study (Dokeroglu et al., 2015) utilized common sub-expressions among queries and proposed a set of heuristic algorithms to improve the total execution time. Study (Lang and Patel, 2009) proposed the QED mechanism (Improved Query Energy-efficiency by Introducing Explicit Delays) that utilizes the common components of queries in a workload. And (Lang et al., 2011) proposed an energy-efficient framework that considers performance requirements specified in SLAs. Study (Sarda and Haritsa, 2004) presented a tool called PLASTIC which amortizes optimization overheads of queries with a technique of plan recycling. Study (Leverich, 2014) presented two hardware proposals and two software proposal to reduce power cost of servers and improve power efficiency of servers operating as a cluster. Study (Lefevre, 2014) proposed new physical design tuning methods to address traditional DBMSs' inadequacies for emerging system architectures and analytics. By mapping the main hardware resource consumption to power consumption, study (Binglei et al., 2015a, 2015b) constructed a power consumption model based on the smallest units of resource consumption of queries.

While our work is similar with (Rodriguez-Martinez et al., 2011; Lang et al., 2011; Xu et al., 2013a, 2013b, 2015), we differ from these approaches in at least four major aspects. First, we build a comprehensive energy cost model by defining and using fine-grained parameters that reflect system resources (CPU, disk, and memory) consumed by queries. Second, our work differentiates data sources (data accessed from disk or data cached in memory) to improve accuracy when estimating the energy cost of reading data during query processing. And we also take energy cost of using memory resource into consideration. Third, we do not only just simply save energy for queries; we improve energy efficiency of queries by utilizing the trade-offs between performance and power. And the last, existing work mostly focused on testing and analyzing energy efficiency of database servers. However, the relationship between energy consumption and performance of the DBMS is controversial (Tsirogiannis et al., 2010; Wang et al., 2011; Peiquan et al., 2014). We present a detailed profile of the relationship between energy consumption and performance using our framework.