

基于数据库负载的SQL能耗预测模型

国冰磊¹ 于炯¹ 廖彬² 杨德先¹

(新疆大学软件学院 乌鲁木齐 830008)¹ (新疆财经大学统计与信息学院 乌鲁木齐 830012)²

摘要 传统数据库以性能(吞吐量、响应时间)为首要优化目标,忽略了数据库系统的能量消耗。在一味追求性能的同时,高能耗问题日益突出,为数据库负载构建能耗模型是构建绿色数据库的基础。通过量化查询负载执行过程中对系统资源(CPU与磁盘)的消耗,将资源消耗产生的时间代价和功耗代价转化为时间代价预测模型和功率代价预测模型,在单站点数据库服务器上实现了为数据库系统构建资源单位代价统一的能耗预测模型。采用多元线性回归工具拟合模型的重要参数,实验结果验证了能耗预测模型的可行性;并分别在静态与动态的系统环境下对系统不同类型查询负载的能耗进行预测与评价,验证了该模型的准确性,使得提出的能耗模型适合于构建能耗感知的绿色数据库。

关键词 绿色计算,SQL执行能耗,查询处理,绿色数据库

中图法分类号 TP315 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.01.040

SQL Energy Consumption Forecasting Model Based on Database Load Status

GUO Bing-lei¹ YU Jiong¹ LIAO Bin² YANG De-xian¹

(School of Software, Xinjiang University, Urumqi 830008, China)¹

(College of Statistics and Information, Xinjiang University of Finance and Economics, Urumqi 830012, China)²

Abstract In a typical database server, performance (throughput or response time) is the first-class optimization goal. However, energy consumption of database systems is ignored by the service providers and users, which makes the high energy consumption be a serious problem in data centers in the processing of chasing performance. Building energy consumption model for query workload is the first step to create a green database. By quantifying the system resources (CPU and Disk) consumed by query workload and transforming the time cost and energy cost into two independent models (time estimation model and power estimation model), the energy estimation model with uniform resource unit was implemented in a single-site database server. Using the multiple linear regression method to compute the key parameters of the above models, experimental results prove the feasibility of our model. To further prove the accuracy and efficiency of the above model, we also made it work under two different system settings (static system and dynamic system), making it more suitable for building the energy-aware green database.

Keywords Green computing, SQL execution energy consumption, Query processing, Green database

1 引言

在传统数据中心,数据库负载占用着大量的服务器资源(包括CPU、磁盘I/O等),使得数据库系统成为数据中心能量消耗的主要组成部分^[1]。为此,研究并解析数据库系统的能耗问题,为DBMS的高能耗问题找到优化方法或解决途径,并在满足时间及性能约束的前提下构建功率感知的绿色数据库,是解决大型数据中心中数据库系统高能耗问题的关键。文献[2]指出,2006年美国数据中心一年的耗电量高达610亿千瓦时(KWh),约占整个国家一年耗电量的1.5%。此外,数据中心的用电需求总量以每年12%的比率增长,导致仅2011年数据中心的耗电量就达到1000亿千瓦时(KWh),相当于74亿美元的电费^[3]。文献[4]指出服务器工作3年的用电成本将超过服务器最初的采购成本,同时增加

的能耗导致的环境问题促使政府在全世界各地开始监管企业的能耗控制能力。所以无论从降低数据中心的运营成本还是从降低能耗保护环境的角度出发,研究数据库节能技术都具有重要的现实意义及广泛的应用前景。

为了节省数据中心的运营成本并减少对环境的影响,学术界与工业界分别从硬件、操作系统、虚拟机、分布存储系统^[5-7]等层次去提高数据中心的能源使用效率。而本文关注DBMS中的查询优化模块,从查询处理的方式以及查询优化(能耗优化)的本质考虑数据库系统的节能问题,提出基于数据库I/O访问类型的SQL执行能耗预测模型。从解析后的执行计划中提取资源消耗的相关信息,并将其映射为系统CPU资源及磁盘资源的消耗,最终转化为系统各部件的能量消耗。本文的主要工作是在对DBMS查询优化器优化机制进行分析的基础上,为数据库系统构建单位代价统一的能耗

到稿日期:2015-12-01 返修日期:2016-04-09 本文受国家自然科学基金项目(61462079,61262088,61562086,61363083)资助。

国冰磊(1991-),女,硕士生,CCF学生会会员,主要研究方向为云计算、绿色计算,E-mail: xingbl@163.com;于炯(1964-),男,博士,教授,博士生导师,主要研究方向为云计算与大数据;廖彬(1986-),男,博士,讲师,主要研究方向为数据库理论与技术、绿色计算、数据挖掘;杨德先(1991-),男,硕士生,主要研究方向为云计算、绿色计算。

模型,并通过实验验证能耗模型的有效性,证明该模型能够应用于构建功率感知的节能数据库系统。实验结果表明,能耗模型的平均相对误差为5.73%,绝对误差不超过9.6%。并且,该模型将CPU功耗和磁盘功耗统一转换为单数据块读的单位代价,将模型嵌入到数据库查询优化器中,不需要计算出实际的功耗值,通过比较数据块的数量就能了解数据库在当前的动态资源消耗及能耗情况。

本文第2节介绍了构建绿色数据库的相关工作;第3节详述了能耗模型的构建基础与理论;第4节论述了能耗模型的实现并对结果进行分析,同时考察了MBRC(一次多数据块读操作从磁盘读取数据块的数量)的设置对模型准确度的影响;最后对全文进行了总结。

2 相关研究

在学术界与工业界,功率感知的绿色数据库领域已有的研究成果可以分为如下两大类:基于硬件的数据库节能与基于软件的数据库节能。早期的研究成果以硬件节能最为常见,如设计耗电比(Performance-per-watt)更优的硬件,调整并适应硬件在不同系统状态下的功率状态等。主要方法是利用耗电比更优的硬件设备替换现有系统中的高能耗设备,如利用固态硬盘(Solid State Drives, SSD)替换机械磁盘,在提高数据库能耗效率的同时还提高了系统的性能。由于SSD相对于传统机械磁盘的成本偏高,因此构建基于闪存与硬盘的混合存储系统是当前的研究热点^[8,9]。

在过去的十年里,大多数工作集中在硬件及系统层,目的在于提高数据库系统的能源效率。然而在软件层面的研究却相对较少,其主要集中在研究功率感知的查询优化器,核心思想是构建功耗/能耗模型,将其无缝地嵌入数据库原有的查询优化器中,以达到为数据库负载估算实时功耗/能耗的目的^[10-19]。构建功率感知的查询优化器的目的是在保证满足时间性能约束的前提下,为SQL语句选择低功耗/低能耗的执行计划,最终提高系统的能源效率。

Rodriguez等^[10]利用多元线性回归实验,通过训练负载,提取不同查询计划的共同特征(如元组大小、基数、列数等)为数据库构建了较准确的SQL语句级功耗模型。模型的重要参数来自内部传感器或功率计上测得的功率以及负载统计数据。其构建的功耗模型准确性较高,但仅限于选择查询,并不能对所有的SQL语句类型进行完全覆盖。

Xu等^[11-13]针对运算符级别的功耗模型展开了广泛与深入的研究。其基本思想是通过总结运算符在执行过程中的共同特征(如CPU处理的普通元组数、索引元组数、从磁盘读取的页面数等)为通用运算符构建功耗模型,并利用一系列的简单查询校准模型重要参数;最后,通过构建一个查询计划评估引擎,按照性能约束和功率消耗来评估一个查询计划的好坏,该模型能够达到为数据库节能的目的,但存在精度较差的缺点。

Tsirogiannis等^[14]利用5种不同的数据库并改变查询计划、压缩算法、访问方法、并行类型和级别等来研究影响数据库能耗的因素。研究发现,对传统数据库进行能耗优化是无用的,因为能源效率总是与性能正相关。其提出的关于能耗与性能之间的关系具有较大的争议,学术界对该问题还处于讨论与探索阶段,并没有形成一致的观点^[15]。

杨良怀等^[16]采用线性回归方法来拟合CPU各个核的利用率、执行频率、磁盘利用率以预测系统功耗。实验从执行核粒度考察CPU的执行频率与利用率对系统功耗的综合影响,避免了在多核架构下整体CPU的活动信息掩盖单个核使用情况的问题。实验结果表明,系统功率与执行频率呈正相关。

国冰磊等^[17]在对SQL查询处理机制研究的基础上,对一系列查询优化原则进行实验,表明不同优化原则对性能提升及能耗降低的有效性。实验结果表明,CPU利用率是影响单站点数据库系统功耗的最关键因素,SQL能耗优化方法可忽略内存优化且应该均衡考虑性能优化及功耗优化两方面。

本文工作与文献^[12]最为接近,但是不同之处在于:本文在对CPU的能耗估算中选取CPU指令数而不是CPU处理的元组数来反映CPU的工作量。CPU指令数较元组能更好地反映CPU的能量消耗情况(因为以元组数为最小计量单位的误差较大,而CPU指令数总数则能更好地反应CPU的工作量)。将本文提出的能耗模型嵌入到数据库查询优化器中,能够方便用户实时地掌握数据库能耗,是构建功率感知的绿色数据库的基础。

3 实时能耗预测模型

SQL语句是操作关系型数据库的标准接口,所有上层应用程序对数据库的操作最终都会转化为SQL语句对数据库的操作。SQL语句的执行效率和产生的能耗将直接影响数据库的性能和能源消耗。本文对SQL语句的执行进行能耗建模,达到了为数据库进行能耗建模的目的。

为了能够方便且有效地对数据库系统的总能耗进行预测,并面向数据库的时间性能与功率消耗进行均衡的控制与调节,将SQL语句的能耗预测模型分为两个相互独立的简单模型:SQL语句的时间代价预测模型与SQL语句的功率代价预测模型。

SQL语句在执行过程中会消耗CPU、磁盘、内存、网络等资源,SQL语句的执行代价由使用相关资源产生的代价累加而成(即时间代价与功率代价)。对于SQL语句,由于CPU、磁盘是主要的资源消耗组件,因此本文的能耗模型主要考虑CPU与磁盘的能源消耗。其他能耗部件,以内存为例,通过优化内存功耗带来的系统总功率优化的升值空间不大,内存的动态功耗波动对系统动态功耗几乎无影响^[17],因此本文在模型中未显式地把内存部分的功耗加入到模型构造中,而将其作为系统的静态功率进行处理。

CPU与磁盘资源的消耗会产生两种代价:1)时间代价;2)功率代价。本文将分别针对总时间代价和总功率代价进行代价建模,最终得到系统总能耗模型。代价模型中都包括CPU代价和IO代价。最后将这些资源消耗进行量化,统一为单位相同的资源代价模型。

(1)CPU代价:对于CPU产生的时间代价与功率代价,选取能更好反应CPU工作量的CPU指令数总数(设为CPU_sum)对代价进行估算^[19],这也是本文模型相对于其他模型的优点。单纯地利用CPU处理的元组总数对CPU的工作量进行估算会造成较大的误差^[12],因为在SQL语句执行的过程中,CPU除了要处理元组,还要负责解析SQL语句、构建查询树(计划选择空间)、挑选最优的查询计划等,这些操作都需消耗CPU指令;不同于元组总数,CPU指令总数能真实反应CPU的工作量。

(2) 磁盘代价: 磁盘产生的时间代价与功率代价与磁盘读写数据量密切相关。在 SQL 执行过程中最常见的 I/O 类型为: 单数据块读 (Single Read Data Block)、多数据块读 (Multiple Read Data Block)。以 Oracle 为例, 其读取磁盘数据的最小 I/O 单位是数据块 (DataBlock)。单数据块读是指一次 I/O 仅从磁盘读取单个数据块, 并把它读入共享缓存 (Buffer Cache) 中, 设其代价为 $single_R_BCOST$ 。多数据块读是一次 I/O 操作读取系统规定数量的数据块, 设其代价为 $multi_R_BCOST$ 。以读取数据块的块数作为磁盘代价的计量单位, 能良好地反应磁盘的工作量。

3.1 SQL 时间代价预测模型

SQL 执行的时间代价由 CPU 时间代价与 I/O 时间代价累加而成。在 Oracle 中, 有一个参数 $Cost(\%CPU)$, 用于表示 CPU 时间代价在总时间代价中所占的比例 (可由系统统计数据得到)。由于实验中涉及的都是执行时间较短且结构典型的 SQL 语句, 因此 CPU 的时间代价 (相对于 I/O 时间代价) 往往忽略不计, 即总的时间代价就是 I/O 时间代价。

$$COST_time = IO_cost_time \quad (1)$$

根据 I/O 类型, 利用单数据块读的单次时间代价将这两种类型的 I/O 时间代价转换为统一的代价单位, 为了避免在不同数据库配置下数据块读操作有不同的操作时间, 所有代价都会转换为单数据块读的时间代价, 并最终求解出数据库总时间代价。利用单数据块读的时间代价作为最小代价单位, I/O 类型为单数据块读的总时间代价即为单数据块读的总次数。

I/O 类型为单数据块读的时间代价:

$$IO_cost_time = single_R_BCOST = single_R_B_sum \quad (2)$$

I/O 类型为多数块读的时间代价:

$$IO_cost_time = multi_R_BCOST = multi_R_B_sum \quad (3)$$

统一代价单位的多块读时间代价:

$$multi_R_BCOST = multi_R_B_sum * MREDTIM / SREADTIM \quad (4)$$

其中, $single_R_B_sum$ 是单数据块读的总次数, $multi_R_B_sum$ 是多数数据块读的总次数。MREDTIM 是一次多数数据块读的平均时间 (单位为毫秒), SREADTIM 是单数据块读的平均时间 (单位为毫秒)。利用式 (4) 可将多块读的时间代价转换为单块读的时间代价, 同时实际的时间代价可由式 (2) 或式 (4) 与 SREADTIM 相乘得到。

3.2 SQL 功率代价预测模型

功耗预测模型所估算的是系统动态总功率, 是负载情况下 CPU 与 I/O 操作产生的动态功率之和, 其中不包括系统空闲时的功耗。同样, 利用单数据块读的功率代价将 CPU 功率代价与 I/O 功率代价统一, 因此所有功率代价都会转换为单数据块读的功率代价, 并最终求解出数据库的总功耗。

$$COST_power = IO_cost_power + CPU_cost_power \quad (5)$$

3.2.1 I/O 功耗 (IO_cost_power)

I/O 操作类型为单块读与多块读, 因此仅计算 I/O 次数而不先考虑每次 I/O 的功率消耗。

单数据块读的功率代价即为单数据块读的总次数, 得到 I/O 类型为单数据块读的功率代价:

$$IO_cost_power = IO_single_cost = single_R_B_sum \quad (6)$$

I/O 类型为多数块读的功率代价:

$$IO_cost_power = IO_multi_cost = multi_R_B_sum \quad (7)$$

统一代价单位的多块读功率代价:

$$IO_multiple_cost = multi_R_B_sum * multi_R_B_power / single_R_B_power \quad (8)$$

其中, $multi_R_B_power$ 是一次多数数据块读的功率, $single_R_B_power$ 是一次单数据块读的功率。可利用式 (8) 将多块读功率代价转换为单块读功率代价。

3.2.2 CPU 功耗 (CPU_cost_power)

在 SQL 语句执行的过程中, 每一种对数据的处理都会依据处理量与处理方式的不同请求 CPU 执行一定数量的指令, 将 SQL 执行过程中消耗的 CPU 指令称为 CPU 总指令数 (CPU_sum), 为方便计算, CPU 总指令数 (CPU_sum) 以万为计量单位。利用转换因子 (设为 Transfer1) 把 CPU 功率代价转换为 I/O 功率代价, 实现 CPU 功耗与 I/O 功耗单位统一。Transfer1 即为计算一次单数据块读的能耗与等价的 CPU 指令数能耗的参数。

$$CPU_cost = CPU_sum / Transfer1 \quad (9)$$

转换因子的计算公式如下:

$$Transfer1 = CPU_watt * single_R_B_power \quad (10)$$

对于不同型号的 CPU, 其处理指令的能力存在较大差异, 这里的处理能力是指 CPU 的指令功率能力 (CPU_watt, 即每瓦特可以完成的 CPU 指令数)。实验通过对数据库进行初步的功耗建模, 训练出 CPU_watt 的值。

统一代价单位的 CPU 功率代价:

$$CPU_cost_power = CPU_sum / (CPU_watt * single_R_B_power) \quad (11)$$

由此可得, I/O 类型为单数据块读的系统总动态功率代价:

$$COST_power = IO_cost_power + CPU_cost_power = single_R_B_sum + CPU_sum / (CPU_watt * single_R_B_power) \quad (12)$$

I/O 类型为多数块读的系统总动态功率代价:

$$COST_power = IO_cost_power + CPU_cost_power = multi_R_B_sum * multi_R_B_power / single_R_B_power + CPU_sum / (CPU_watt * single_R_B_power) \quad (13)$$

设数据库系统总功耗为 Total_power, 总能耗为 Total_energy, 系统的空闲功率 (系统在无负载情况下的功耗) 为 Idle_power, 则系统总功耗与总能耗如下:

$$Total_power = COST_power + Idle_power \quad (14)$$

$$Total_energy = (COST_power + Idle_power) * IO_cost_time \quad (15)$$

4 模型设计与实现

将时间代价预测模型与功率代价预测模型合并即得到系统总能耗预测模型。实验需求出 CPU 指令功率能力、一次单数据块读的功率、一次多数数据块读的功率。以 I/O 类型为单数据块读的情况为例求导模型, 利用式 (12)、式 (14) 得到 I/O 类型为单数据块读的系统总功耗如式 (16) 所示:

$$Total_power = (1 / CPU_watt) * CPU_sum + single_R_B_power * single_R_sum + idle_power \quad (16)$$

设系统的总功率为 y , CPU_sum 为 X_1 , single_R_B_sum 为 X_2 , 得到如下多元线性回归方程式:

$$y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 \quad (17)$$

其中, X_1, X_2 为自变量(回归变量), 参数 $\beta_0, \beta_1, \beta_2$ 称为回归系数, 分别代表系统空闲功率、每万条 CPU 指令消耗的功率及一次单数据块读的功率。

4.1 实验环境及数据采集

实验通过双机通信的方式排除功率监控软件对采样数据准确度的干扰。采用 HOPI 数字功耗仪, 搭建能耗建模的实验平台, 实验功耗数据采样频率设置为每秒 1 次, 如图 1 所示。实验总体环境描述如表 1 所列。实验测试所用表为 EMP, DEPT, ADDRESS, SALARY, JOBTITLE, 数据量分别为 500 万, 100 万, 100 万, 100 万, 100 万。

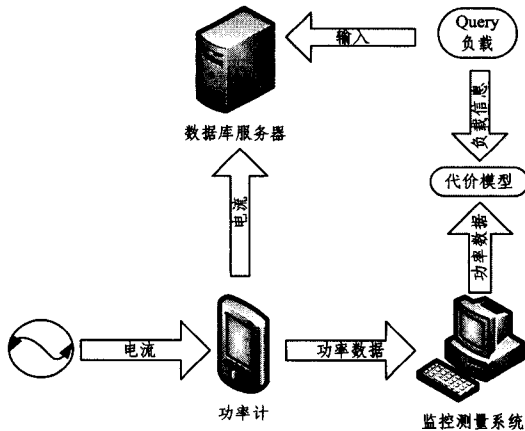


图 1 能耗测试环境示意图

表 1 实验环境描述

项目	描述
操作系统	Windows 7 64 位 SP1 (DirectX 11)
数据库管理系统	Oracle 11g
功耗数据测量	HOPI 功率测试仪 (USB 智能版), 功率误差值 ±0.01W, 采样频率为 1~3s 之间, 单位为 kWh
功耗数据采集	用电监测仪数据分析系统 V1.0.1
实验机 CPU	Intel(R) Core i7-4790 @ 3.60GHz
实验机内存	4 GB (金士顿 DDR3L 1600MHz)
实验机硬盘	西数 WDC WD10EZEX-21M2NA0(997GB/7200 转/分)

Oracle 系统中, 系统处理能力是影响执行计划中操作代价参数的重要因素, DBMS_STATS 中有相应的存储过程 (GATHER_SYSTEM_STATS) 来收集相关数据。这些系统统计数据包括: CPU 指令总数 (CPU_sum)、单块读次数 (single_R_B_sum)、多块读次数 (multi_R_B_sum)、单块读 I/O 时间 (SREADTIM)、多块读 I/O 时间 (MREDTIM) 以及 MBRC (db_file_multiblock_read_count) 即多数据块读时一次 I/O 读取的数据块数量。其中 SREADTIM 和 MREDTIM 需要在负载模式下多次收集。

为了避免驻留在共享池 (shared_pool) 及缓存 (BUFFER_CACHE) 中的数据影响其他 SQL 语句的执行, 每提交一次 SQL 语句到客户端 (SQL Plus) 都需清空缓存 (BUFFER_CACHE) 及共享池 (shared_pool)。同时为了减小误差, 每条 SQL 要在清空共享池和缓存的情况下连续执行 100 次。采用 PL/SQL+动态 SQL 使 SQL 语句无缝嵌套在 PL/SQL 代码段。

实验中, 系统需绕过缓存从磁盘直接读入数据块。因为从磁盘读取数据块和从缓存读取数据块是两个不同的操作, 所以需要定期清除缓存, 以保证与缓存相关的数据操作的统计数据 (db_block_gets_from_cache, physical_reads_cache 及 consistent_gets_from_cache) 为零。

为了保证 SQL 执行引擎按照实验设计的执行方式处理数据, 需要使用 Oracle Hints 让查询优化器改变执行计划并让 SQL 执行引擎按照指定的执行计划执行 SQL 语句。但是需要注意 Hints 失效的情况, 实验中建立的索引必须为唯一非空索引才能保证所有关于索引的操作有效。当存在索引范围扫描、索引唯一扫描、索引完全扫描、由 ROWID 访问表等操作时会出现单数据块读操作, 当存在全表扫描和快速完全索引扫描时会出现多数据块读。实验设计了两组数据集, 一组用于训练出回归模型中回归系数的值, 一组用于验证模型的有效性 with 准确度。实验中训练集及测试集分别为 100 条和 10 条。训练集实例与测试集实例分别如表 2、表 3 所列。

表 2 实验所用训练集实例

序号	数据操作类型	I/O 类型	SQL 例句
1	全表扫描	多块读	select/* + FULL(PART) */ * from SCOTT.PART;
2	索引范围扫描	单块读	Select/* + INDEX (CUSTOMER SYS_C0011165) */ /C_CUSTKEY from SCOTT.CUSTOMER where C_CUSTKEY between 6000 and 1300000;
3	ROWID 访问表	单块读	select/* + ROWID(PART) */ * from SCOTT.PART;

表 3 实验所用测试集实例

序号	数据操作类型	I/O 类型	SQL 例句
1	快速完全索引扫描	多块读	Select/* + INDEX_FFS (SUPPLIER SYS_C0011162) */ /S_SUPPKEY from SCOTT.SUPPLIER;
2	索引完全扫描	单块读	select /* + INDEX (PART SYS_C0011161) */ / P_PARTKEY from SCOTT.PART;
3	索引唯一扫描	单块读	select/* + INDEX (REGION SYS_C0011112) */ / R_REGIONKEY from SCOTT.REGION where R_REGIONKEY =43;

4.2 模型求解与验证

数据库系统的总功率由空闲功率与动态功率组成。空闲功率是系统在无负载情况下的功率, 动态功率是执行 SQL 语句所需的额外功率。系统的动态功率反映了系统当前的资源消耗情况。当系统处于空闲状态时, 测得其平均功率为 30.92W。在 DBMS 独占系统资源的情况下, 执行训练集 (见表 2), 收集统计数据, 在 Matlab 中用多元线性回归方法进行线性拟合, 可得到如下公式:

$$y = 38.60 + 0.0386 \times X_1 + 0.3733 \times X_2 \quad (18)$$

式(18)中的参数含义与式(17)中一致。理想情况下, 常数项应该等于空闲功率(30.92W), 但拟合结果为 38.60W, 与真实值的误差为 19.9%。因测试时无法预知系统的活动, 且除 CPU 和磁盘外其他组件也存在功率消耗, 其功率应高于纯粹的空闲功率。在负载模式下收集的 SREADTIM 和 MREDTIM 分别为 3.59 和 11.15(单位为毫秒)。

由此可得, 在 I/O 类型为单块读时系统总能耗预测模型如下:

$$Total_energy = (38.60 + 0.0386 \times X_1 + 0.3733 \times X_2) \times (X_2 \times 3.59 \times 10^{-3}) \quad (19)$$

将所得的参数值带入式(13), 设 multi_R_B_sum 为 X_3 , 同理可得在 I/O 类型为多块读时系统总能耗预测模型为:

$$Total_energy = (38.60 + X_3 \times 1.05 + X_1 \times 0.0386) \times (X_3 \times 11.15 \times 10^{-3}) \quad (20)$$

为了简化数据库中不同种类查询负载能耗的预测过程,在不区分 I/O 操作类型的情况下,利用式(19)、式(20)得到通用的系统总能耗预测模型:

$$Total_energy = \{138.574X_2 + X_1(0.1386X_2 + 0.4304X_3) + 430.39X_3 + 11.7075X_3^2 + 1.3401X_3^3\} * 10^{-3} \quad (21)$$

为验证模型的准确性,实验设计的测试集分别实现了单数据块读和多数据块读的 I/O 操作,其 SQL 语句实例如表 3 所列。在两种不同的系统环境下对 SQL 语句的执行能耗进行测量,分别运行了 10 条测试语句。

(1)静态环境:系统上只运行 DBMS,数据库系统独占系统资源,SQL 语句依次单个执行。此时多处理级(Multi-Processing Level, MPL)为 1。

(2)动态环境:将 MPL 设置为 2,即存在一个其他应用程序与数据库系统共享系统资源。实验中运行.NET 程序进行模拟。测试结果如图 2—图 5 所示。

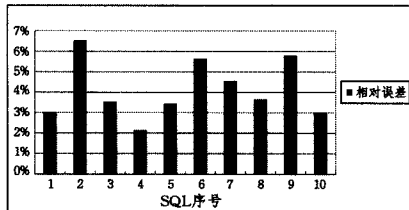


图 2 静态环境下能耗预测图(I/O 类型为单数据块读)

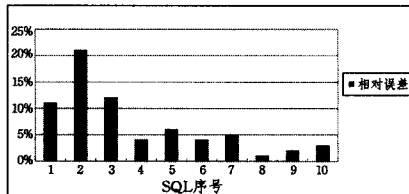


图 3 动态环境下能耗预测图(I/O 类型为单数据块读)

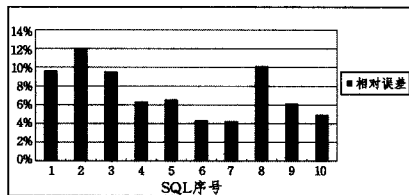


图 4 静态环境下能耗预测图(I/O 类型为多数据块读)

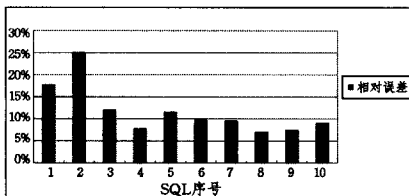


图 5 动态环境下能耗预测图(I/O 类型为多数据块读)

由图 2 和图 3 可知,在静态环境下($MPL=1$),本文模型比文献[10,11]中的模型对能耗预测的准确性更高。通过分析实验数据可知,本文模型的优越性在于在估算 CPU 能耗时所选取的 CPU 指令数为最小计量单位,其能更好地反应 CPU 的能量消耗。在静态环境下的平均相对误差为 5.73%,绝对误差不超过 9.6%。由图 4 和图 5 可知,在动态环境下($MPL=2$),由于系统活动情况无法预知,数据库系统与其他程序竞争系统资源,在有些地方的绝对误差会超过 20%,平

均相对误差为 9.27%。实验结果表明,本文模型能够较准确地预测数据库系统的能耗以及系统重要组件(CPU、磁盘)的能耗。

4.3 MBRC 设置与模型稳定性

MBRC(db_file_multiblock_read_count)即执行一次多数数据块读操作从磁盘读取数据块的数量。通常情况下操作系统的最大 I/O 会影响 Oracle 的 MBRC 设置,在最大 I/O 为 1MB 时,MBRC 值为 $1MB/db_block_size$ 。实验中,数据库服务器设定的 db_block_size (数据块大小,在数据库创建时被指定,数据库一旦创建,其值就不可更改)的值为 8kB,即上述实验中系统允许的 MBRC 最大值为 128。对比图 2、图 3 与图 4、图 5,发现多块读的模型精确度要略低于单块读的模型精确度。为了找到其原因,利用 10046 事件,分析跟踪文件后发现在设置完 MBRC 后系统在执行负载时并不一定能保证其值的有效性,即实际执行过程中并不是每次都读取规定数量的数据块,详细信息如表 4 所列。在不同数据库配置下,表 4 中的数据存在不同程度的差异。为了考察 MBRC 的设置对模型的影响,设实验中设置的 MBRC 值与实际一次读取数据块总量之间的差值为 d_value ($[0, 1MB/db_block_size)$)。实验中改变 MBRC 的值,分别在 MBRC 为 4,5,10,20 的情况下对 5 张不同的表进行全表扫描,并计算不同 MBRC 设置下的能耗误差,如图 6 所示。

表 4 MBRC 与 d_value 值

MBRC	一次 I/O 实际读取数据块数量(平均值)	d_value (平均值)
4	3.5	0.5
5	4	1
10	8.6	1.4
20	18	2

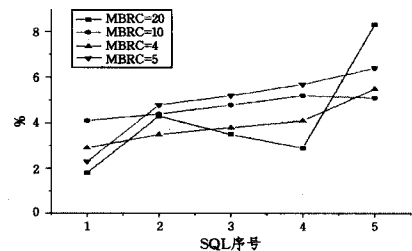


图 6 MBRC 设置与能耗误差

结合表 4 与图 6 发现随着 MBRC 值的增大,理论上与实际上一次 I/O 读取的数据块数之间的差值变大(即 d_value 变大),能耗误差与 MBRC 值的设定之间没有明显的关系。但是仔细观察图 6 发现在 MBRC 为 10 时,误差范围比较稳定。由此推测,针对不同的数据库配置环境,存在一个对应的 MBRC 值符合能耗预测模型稳定性的要求。但这一初步的结果需要在不同数据库配置环境下进行大量的实验验证,本文主要关注静态环境下的能耗建模,是否存在一个最优的 MBRC 值也是我们未来研究的重点。

结束语 本文提出的能耗预测模型不需单独测量系统各组件的能耗,只需提取 SQL 执行计划中的相关资源消耗信息及利用功率计测量出系统的整机功耗就可以实现对系统整体能量消耗的预测。将复杂的能耗建模过程分解为两个相对简单的独立预测模型(时间代价预测模型和功率代价预测模型),降低了模型构建的复杂度。相比于传统数据库,能耗感知的绿色数据库更能适应以节能为中心的全球低碳化趋势。

基于本文提出的能耗预测模型,可以达到为数据库负载估算能量消耗的目的,但是如何挑选节能的查询计划以及如何在满足系统性能需求的同时最小化能耗,需要充分考虑性能与功率之间的折中问题。

本文下一步工作要点:

(1)以关系型数据库为基础,进一步研究 MBRC 值与模型稳定性的关系及在不同的数据库配置环境下是否存在一个最优的 MBRC 值。

(2)将本文提出的能耗预测模型扩展到动态环境下(MPL>2),研究如何进一步提高模型的精确度与健壮性。

(3)以本文提出的能耗预测模型为基础,研究如何捕获系统在动态情况下对性能退化所能容忍的最大限度,进而指导查询优化器挑选性能较优且功耗低的节能查询计划。

参 考 文 献

- [1] POESS M, NAMBIAR R O. Energy cost, the key challenge of today's data centers; a power consumption analysis of TPC-C results[J]. Proceedings of the VLDB Endowment, 2008, 1(2): 1229-1240.
- [2] KURP P. Green Computing[J]. Communications of the ACM, 2008, 51(10): 11-13.
- [3] HARIZOPOULOS S, SHAH M, MEZA J, et al. Energy efficiency: The new holy grail of data management systems research [J]. arXiv preprint arXiv:0909.1784, 2009.
- [4] BRILL K G. Data center energy efficiency and productivity[J]. the Uptime Institute-White Paper, 2007(5): 176-184.
- [5] LIAO Bin, YU Jiong, ZHANG Tao, et al. Energy-Efficient Algorithms for Distributed File System HDFS[J]. Chinese Journal of Computers, 2013, 36(5): 1047-1064. (in Chinese)
廖彬, 于炯, 张陶, 等. 基于分布式文件系统 HDFS 的节能算法研究[J]. 计算机学报, 2013, 36(5): 1047-1064.
- [6] LIAO Bin, YU Jiong, SHUN Hua, et al. Energy-Efficient Algorithms for Distributed Storage System Based on Data Storage Structure Reconfiguration [J]. Journal of Computer Research and Development, 2013, 50(1): 3-18. (in Chinese)
廖彬, 于炯, 孙华, 等. 基于存储结构重配置的分布式存储系统节能算法[J]. 计算机研究与发展, 2013, 50(1): 3-18.
- [7] LIAO Bin, YU Jiong, ZHANG Tao, et al. Novel Energy-efficient Metadata Dynamic Modeling and Management Approach for Cloud Storage System[J]. Journal of Chinese Computer Systems, 2013, 10(34): 2407-2412. (in Chinese)
廖彬, 于炯, 张陶, 等. 一种适应节能的云存储系统元数据动态建模与管理方法[J]. 小型微型计算机系统, 2013, 10(34): 2407-2412.
- [8] GRAY J. Tape is dead, disk is tape, flash is disk, RAM locality is king[OL]. http://signallate.com/signallake.com/innovation/Flash_is_Good.pdf.
- [9] WANG Jiang-tao, LAI Wen-yu, MENG Xiao-feng. Flash-Based Database; Studies, Techniques and Forecasts [J]. Chinese Journal of Computers, 2013, 36(8): 1549-1567. (in Chinese)
王江涛, 赖文豫, 孟小峰. 闪存数据库: 现状, 技术与展望[J]. 计算机学报, 2013, 36(8): 1549-1567.
- [10] RODRIGUEZ-MARTINEZ M, VALDIVIA H, SEGUEL J, et al. Estimating Power/Energy consumption in Database Servers[J]. Procedia Computer Science, 2011, 6(1): 112-117.
- [11] XU Z. Building a power-aware database management system [C]// Proceedings of the Fourth SIGMOD PhD Workshop on Innovative Database Research. ACM, 2010: 1-6.
- [12] XU Z, TU Y C, WANG X. Exploring power-performance tradeoffs in database systems[C]// 2010 IEEE 26th International Conference on Data Engineering (ICDE). IEEE, 2010: 485-496.
- [13] XU Z, TU Y C, WANG X. PET: reducing database energy cost via query optimization[J]. Proceedings of the VLDB Endowment, 2012, 5(12): 1954-1957.
- [14] TSIROGIANNIS D, HARIZOPOULOS S, SHAH M A. Analyzing the energy efficiency of a database server [C] // Proc. of SIGMOD '10 Indianapolis. IN, USA, 2010: 231-242.
- [15] JIN Pei-quan, XING Bao-ping, JIN Yong, et al. Survey on energy-aware green databases[J]. Journal of Computer Application, 2014, 34(1): 46-53. (in Chinese)
金培权, 邢宝平, 金勇, 等. 能耗感知的绿色数据库研究综述[J]. 计算机应用, 2014, 34(1): 46-53.
- [16] YANG Liang-huai, ZHU Hong-yan. Whole system realtime power profiling and modeling[J]. Journal of Computer Science, 2014, 41(9): 32-37. (in Chinese)
杨良怀, 朱红燕. 整机系统实时功率剖析与建模[J]. 计算机科学, 2014, 41(9): 32-37.
- [17] GUO Bing-lei, YU Jiong, LIAO Bin, et al. SQL Energy Consumption Modeling and Optimization Research [J]. Journal of Computer Science, 2015, 42(10): 202-207. (in Chinese)
国冰磊, 于炯, 廖彬, 等. SQL 能耗建模及优化研究[J]. 计算机科学, 2015, 42(10): 202-207.
- [18] LANG W, PATEL J. Towards eco - friendly database management systems[J]. arXiv preprint arXiv:0909.1767, 2009.
- [19] ZHU Yi, XIAO Fang-xiong, ZHOU Hang, et al. Method for Modeling and Analyzing Software Energy Consumption of Embedded Real-Time System[J]. Journal of Computer Research and Development, 2014, 51(4): 848-855. (in Chinese)
祝义, 肖芳雄, 周航, 等. 一种嵌入式实时系统软件能耗建模与分析的方法[J]. 计算机研究与发展, 2015, 51(4): 848-855.
- [18] LI K, SHAN Z G, WANG J, et al. Overview on Major Research Plan of Trustworthy Software[J]. Bulletin of National Natural Science Foundation of China, 2008, 22(3): 145-151. (in Chinese)
李克, 单志广, 王戟, 等. 可信软件基础研究重大研究计划综述[J]. 中国科学基金, 2008, 22(3): 145-151.
- [19] YANG L, ZHANG J Z, HOU X R. Nonlinear Systems of Algebraic Equations and Automated Theorem Proving[M]. Beijing: Shanghai Scientific and Technological Education Publishing House, 1996. (in Chinese)
杨路, 张景中, 侯晓荣. 非线性代数方程组与定理机器证明[M]. 北京: 上海科技教育出版社, 1996.
- [20] YANG L, XIA B C. Inequality Machine Proving and Automated Discovery[M]. Beijing: Science Press, 2008. (in Chinese)
杨路, 夏壁灿. 不等式机器证明与自动发现[M]. 北京: 科学出版社, 2008.

(上接第 198 页)