

Soporte a la Selección Sostenible de Sistemas Gestores de Bases de Datos Relacionales*

Vanessa Rodríguez-Horcajo¹ and Jennifer Pérez¹[0000-0003-3192-7995]

E.T.S. Ingeniería de Sistemas Informáticos, Universidad Politécnica de Madrid,
Madrid, Spain

vanessa.rodriguez.horcajo@alumnos.upm.es, jennifer.perez@upm.es

Resumen Las bases de datos y cómo éstas persistan su información juegan un papel crucial en la sostenibilidad del desarrollo software, ya que su acceso a memoria tiene un impacto inmediato en el consumo energético. Sin embargo, a día de hoy no hay directrices para elegir el sistema gestor de bases de datos adecuado en términos de eficiencia energética. Este trabajo presenta los primeros resultados de esta elección y el estudio de la variabilidad existente de sistemas gestores de bases de datos relacionales para su soporte.

Keywords: variabilidad · bases de datos · líneas de producto software · consumo energético · sostenibilidad · green

1. Introducción

Actualmente se están consiguiendo grandes avances en el diseño de hardware y protocolos de comunicaciones que reducen el consumo energético [1]. Sin embargo, todavía hay mucho trabajo por hacer para conseguir la reducción del consumo energético del software que se ejecuta sobre dicho hardware [2], [3]. Un aspecto crítico en el consumo energético de las aplicaciones software es la gestión que realizan del gran volumen de datos que manejan. Este elevado volumen de datos, que implica a su vez el incremento de centros de datos y la capacidad de los mismos, ocasiona también el aumento del consumo energético de las aplicaciones puesto que se requiere mayor acceso a memoria. Por este motivo, es necesario conocer la criticidad energética de utilizar un sistema gestor de bases de datos (SGBD) u otro, y si, dependiendo de las características de la base de datos (BD), los SGBD se comportan de forma diferente en términos de eficiencia energética. Este trabajo es un primer paso para abordar esta necesidad proporcionando un mecanismo que permita al ingeniero software configurar las características de una BD relacional y obtener la información energética del SGBD a utilizar, asistiendo la toma de decisiones sostenibles en el diseño de la persistencia de sistemas software.

* Este trabajo esta parcialmente financiado por el Ministerio de Ciencia e Innovación (MICINN) a través del proyecto “SIoTCom: Sustainability-Aware IoT Systems Driven by Social Communities” (PID2020-118969RB-I00).



2. Caracterización de SGBD relacionales

Dada la variabilidad de características de las BBDD relacional, se ha definido un árbol de características para su representación. Este árbol de características permite al ingeniero de software, a través de la selección de las características requeridas de la BBDD relacional a diseñar, obtener como producto resultante una recomendación de SGBD relacional en términos de eficiencia energética. El tamaño de las aplicaciones, su complejidad o calidad influye en el consumo energético debido a las instrucciones de entrada/salida que generan accesos a memoria o disco [4], [5]. Por ello, estos mismos factores son importantes a la hora de caracterizar un SGBD con el propósito de analizar su consumo energético. Existen trabajos en la literatura sobre el consumo energético de BBDD relacionales sobre aspectos concretos como el uso de operaciones concretas como INSERT o JOINS [6]. Sin embargo, no existe una caracterización del dominio y su estudio energético para el soporte a la toma de decisiones de persistencia sostenibles.

En este trabajo se ha realizado la caracterización del dominio de los SGBD relacionales para su análisis energético teniendo en cuenta las métricas de tamaño, estructurales y operacionales, establecidas por la literatura para el análisis de otros factores de las BBDD como el mantenimiento o el diseño [7], [8] [9]. La Figura 1 muestra el modelo de características en el que se establecen como características el tamaño, tanto en términos estructurales (número de tablas) como de población (número de tuplas), las relaciones entre tablas (1:1, 1:N, N:M), y las operaciones más importantes que se pueden realizar (CREATE, INSERT, UPDATE, DROP, ALTER, DELETE Y SELECT) con sus diferentes variantes. Si bien es cierto, ha sido necesario establecer unas métricas que definan el tamaño de la BBDD en terminos estructurales y de población, ya que no se han encontrado trabajos que establezcan esta categorización y sus rangos. Este trabajo establece el tamaño mediante los valores XS, S, M, L y XL, como se realiza con las aplicaciones software. El tamaño de la base datos en términos de número de tablas se ha establecido a partir de las evidencias encontradas en el ámbito académico, para los tamaños XS y S, y profesional, para M, L y XL. El tamaño XS se establece en $XS < 5$ considerando el tamaño medio de los ejemplos para explicar conceptos de BBDD. El tamaño S ($5 \leq S \leq 15$) se ha determinado en base a los ejemplos que aparecen en libros de BBDD [10] y ejercicios y exámenes del ámbito universitario. Finalmente, para establecer los tamaños M, L y XL se ha realizado una encuesta a los equipos de desarrollo software de 3 grandes compañías españolas de las áreas de telefonía y la banca, preguntándoles por los SGBD utilizados y por el volumen de tablas que manejaban en bases de datos pequeñas, medianas y grandes. A partir de los resultados de las encuestas, se han establecido los siguientes valores: M ($15 < M \leq 60$), L ($60 < L \leq 100$) y XL ($100 < XL$). Por otro lado, también se ha considerado el tamaño de la población de la base datos, es decir, el número de tuplas que tiene almacenadas, siendo las variantes: vacía, XS: $< 1K$, S: $1K-10K$, M: $10K-100K$, L: $100k-500k$, XL: $> 500k$, siguiendo la misma métrica que SonarQube para determinar el tamaño de una aplicación en función del número de líneas de código.

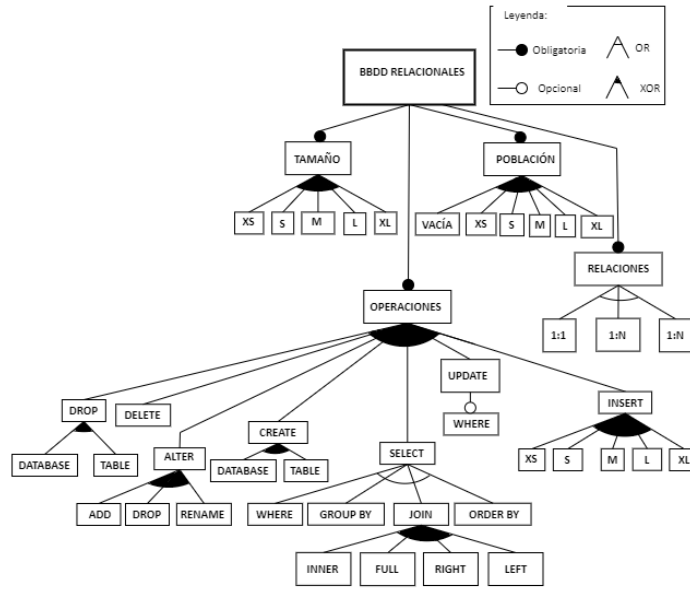


Figura 1. Modelo de Características de SGBD relacionales

3. Recomendación Sostenible para la creación de BBDD Relacionales

La recomendación sostenible se ha de realizar en base a la medición controlada y sistemática de consumo energético. En este trabajo se presentan las mediciones realizadas para los SGBD de MySQL, PostgreSQL y Sql Server para la creación de BBDD S y XS aplicando los comandos CREATE DATABASE y CREATE TABLE. Estas mediciones se han realizado con la herramienta Joulemeter en un ordenador Intel(R) Core(TM) i7-10750H CPU, @2.60GHz (15.8 GB usable), con sistema operativo Windows 10 Home versión 21H2 de 64 bits. Debido a la variación que puede existir en el tiempo de ejecución de una misma operación SQL contra una base de datos empleando SQL, todas las operaciones se ejecutaron tres veces en las mismas condiciones, y se tomó como resultado la media aritmética del consumo energético de las tres ejecuciones. Además, para realizar las mediciones y poder determinar la recomendación una vez seleccionadas las características del árbol (ver Figura 1), se han realizado 63 mediciones energéticas para las XS y 54 para las S, diseñando los siguientes experimentos para cada SGBD: XS: Creación en SQL de 6 BBDD compuestas por 3 tablas con 6 variantes en sus relaciones: 1:1, 1:N, N:M, 1:1 y 1:N, 1:N y N:M, y 1:1 y N:M, y una variante adicional con 4 tablas con los cuatro tipos de relación. S: Creación en SQL de 6 BBDD compuestas por 6 tablas con 6 variantes en sus relaciones: 1:N, N:M, 1:1 y 1:N, 1:N y N:M, 1:1 y N:M, y los tres tipos de relaciones.

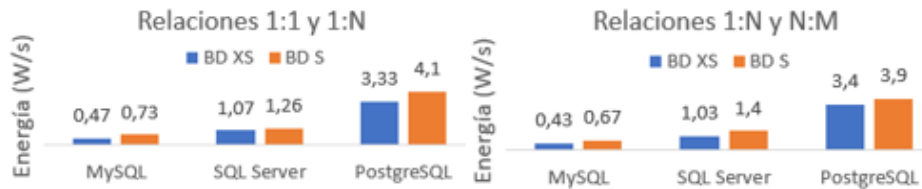


Figura 2. Comparativa consumo energético Creación BBDD XS y S con relaciones (1:1 y 1:N, 1:N y N:M)

4. Conclusiones

Este trabajo evidencia la importancia de los SGBD relacionales en términos de consumo energético y que el tamaño influye en su incremento, siendo MySQL el que menos consume en la creación de BBDD XS y S (ver Figura 2). Como trabajo futuro inmediato se pretende completar la experimentación para BBDD M y las operaciones de INSERT, SELECT, UPDATE y DROP para generar la recomendación al ingeniero de software de estas configuraciones. Por otro lado, a posteriori se dará cobertura al árbol completo.

Referencias

1. Madhura, S.: A review on low power vlsi design models in various circuits. *J Electron* 4(2), 74–81 (2022)
2. Group TC: Smart 2020: enabling the low carbon economy in the information age. GeSi, The Climate Group (2008)
3. Capra E, Formenti G, Francalanci C, Gallazzi S: The impact of MIS software on IT energy consumption. In: *European Conference on Information Systems* (2010)
4. Calero C, Piattini M: Puzzling out software sustainability. *Sustain Comput Inform Syst* 16:117–124 (2017)
5. Guamán, D., Pérez, J. & Valdiviezo-Díaz, P. : Estimating the energy consumption of model-view-controller applications. *J Supercomput* (2023).
6. M. Rodríguez et al.: Analyzing power and energy consumption of large join queries in database systems. *IEEE Symposium on Industrial Electronics & Applications*, Kuching, Malaysia, pp. 148-153, (2013) <https://doi.org/10.1109/ISIEA.2013.6738985>
7. Mishra, S., Aisuryalaxmi, E., Mall, R.: Estimating Database Size and Its Development Effort at Conceptual Design Stage. In: *Global Trends in Information Systems and Software Applications*. ObCom. Communications in Computer and Information Science, vol 270. Springer, Berlin, Heidelberg. (2012)
8. B. Jamil and A. Batool, "SMARTS: Software Metric Analyzer for Relational Database Systems," 2010 International Conference on Information and Emerging Technologies, Karachi, Pakistan, pp. 1-6 (2010)
9. Piattini, M., Calero, C. & Genero, M. Table Oriented Metrics for Relational Databases. *Software Quality Journal* 9, 79–97 (2001).
10. Elmasri, R., Navathe, S.B.: *Fundamentals of Database Systems*, 7th edition, Pearson (2016)