

---

**DISEÑO DE UN WORKLOAD REAL PARA UN BENCHMARK DE  
BASES DE DATOS ORIENTADAS A GRAFOS**

**MAURICIO ORELLANA GALAZ  
INGENIERÍA CIVIL EN COMPUTACIÓN**

**RESUMEN**

Junto al surgimiento de aplicaciones que requieren almacenar y analizar grandes cantidades de información no estructurada y compleja, surgen también los sistemas de gestión de bases de datos orientadas a grafos (SGBDG). Estos sistemas se caracterizan porque su enfoque no está en gestionar datos de forma individual y aislada, sino que ponen énfasis en las relaciones existentes entre dichos datos. El surgimiento de los SGBDG trae consigo la necesidad de realizar benchmarking sobre ellos a modo de poder establecer comparaciones. En este contexto, GDBench es un benchmark enfocado a evaluar el rendimiento de sistemas de bases de datos considerando como caso de uso una red social. Sin embargo, la carga de trabajo (workload) de GDBench no representa un escenario real. En efecto, el workload utilizado por GDBench era aleatorio, no considerando una lógica desde el punto de vista del orden en el que eran ejecutadas las operaciones. En esta memoria se presenta el desarrollo de un workload real, basado en el análisis de una red social existente (Facebook), para GDBench. El nuevo workload simula el comportamiento de un conjunto de usuarios interactuando en la red social detenida por el esquema de datos de GDBench, combinando un conjunto de 6 consultas interactivas y 6 consultas analíticas. El benchmark resultante fue evaluado sobre cuatro sistemas de bases de datos, considerando aquellas orientadas a grafos (Sparksee y Neo4j), RDF (Apache Jena) y relacional (PostgreSQL). Sobre cada uno de los sistemas de bases de datos se ejecutó el workload aleatorio y el workload real. Al realizar la comparación de resultados, se obtuvo que si bien se produjeron diferencias en los tiempos de ejecución de las consultas que conformaban el workload, estas no fueron lo suficientemente significativas como para alterar los resultados del proceso de benchmarking. Como trabajo futuro se plantea un uso más intensivo del benchmark para evaluar la influencia del workload en el proceso de benchmarking.

Palabras claves: SGBDG, BDOG, Benchmarking, Workow, Workload.

## ABSTRACT

Along with the emergence of applications that require storing and analyzing large amounts of unstructured and complex information, graph database management systems (GDBMS) also emerge. These systems are characterized because their approach is not to manage data in an individual and isolated way, but rather they emphasize the relations existing between the data. The emergence of GDBMS brings with it the need to perform benchmarking on these systems so that comparisons can be made. In this context, GDBench is a benchmark oriented to evaluate the performance of database systems considering social network as a use case. However, GDBench's workload does not represent a real scenario.

In fact, the workload used by GDBench was random, not considering logic from the point of view of the order in which the operations were executed. In this document we present the development of a real workload, based on the analysis of an existing social network (Facebook). The new workload simulates the behavior of a set of users interacting in the social network defined by the GDBench data schema, combining a set of 6 interactive queries and 6 analytic queries. The resulting benchmark was evaluated on four database systems, considering those oriented to graphs (Sparksee and Neo4j), RDF (Apache Jena) and relational (PostgreSQL). Over each database system was executed either the random workload and the real workload. When we compared the results we found that there were differences in the execution times of the queries, but the differences were not significant enough to alter the results of the benchmarking process. As future work, we propose a more intensive use of the benchmark to evaluate the influence of workload in the

benchmarking process.