

Índice general

Dedicatoria	I
Agradecimientos	II
Índice de Figuras	v
Índice de Tablas	VI
Resumen	VII
1. Introducción	8
1.1. Descripción del problema	8
1.2. Hipótesis	9
1.3. Solución	9
1.4. Contribuciones	10
2. Contexto	11
2.1. Introducción a los grafos	11
2.2. Sistemas de gestión de bases de datos orientadas a grafos	14
2.3. Benchmarking de BDOG	16
2.3.1. Benchmarking	16
2.3.2. Benchmarks en general	17
2.3.3. Benchmarks para BDOG	18
2.4. Graph Data Benchmark (GDBench)	19
2.4.1. Generador de datos	20
2.4.2. Generador de operaciones	21
2.4.3. Evaluador de operaciones	23

3. Diseño y construcción del workload	25
3.1. Análisis de un workflow real (Facebook)	25
3.2. Definición de un workflow real	28
3.3. Diseño del workload real para GDBench	29
3.4. Construcción del TestDriver	34
3.5. Reportes de Benchmarking	38
4. Experimentos	42
4.1. Metodología de evaluación	42
4.2. Configuración y ejecución de los experimentos	43
4.3. Rendimiento de los sistemas evaluados según workload	44
4.4. Influencia del tipo de workload	46
5. Conclusiones	53
Glosario	55
Bibliografía	56
Anexos	
A: Comparación de tiempos de consulta	60
A.1. Comparación usando gráficos de barra	60
A.2. Comparación usando gráficos de caja	72

ÍNDICE DE FIGURAS

	página
2.1. Representación gráfica de un grafo.	12
2.2. Tipos de grafo	13
2.3. Elementos y características deseables de un benchmark [12].	19
2.4. Modelo de datos de GDBench.	20
3.1. Workflow basado en Facebook	27
3.2. Autómata compuesto por consultas interactivas de GDBench.	31
3.3. Ejemplo de grafo social.	32
3.4. Diagrama de clases.	37
3.5. Ubicación de mediana y cuartiles.	39
3.6. Gráfico de caja	40
3.7. Ejemplo de gráfico de caja.	40
3.8. Ejemplo de registro de consultas en archivo XML.	41
4.1. Rendimiento de sistemas ante workload aleatorio.	45
4.2. Rendimiento de sistemas ante workload real.	46
4.3. Gráfico de caja consulta Q1	52
A.1. Ejecución consulta Q1 según workload.	61
A.2. Ejecución consulta Q2 según workload.	62
A.3. Ejecución consulta Q3 según workload.	63
A.4. Ejecución consulta Q4 según workload.	64
A.5. Ejecución consulta Q5 según workload.	65
A.6. Ejecución consulta Q6 según workload.	66
A.7. Ejecución consulta Q7 según workload.	67
A.8. Ejecución consulta Q8 según workload.	68
A.9. Ejecución consulta Q9 según workload.	69
A.10. Ejecución consulta Q10 según workload.	70
A.11. Ejecución consulta Q11 según workload.	71
A.12. Ejecución consulta Q12 según workload.	72

ÍNDICE DE TABLAS

	página
2.1. Consultas de GDBench.	22
3.1. Consultas de GDBench de tipo interactivo.	29
3.2. Consultas de GDBench de tipo analítico.	30
3.3. Ejemplo de ejecución un micro-workflow.	33
3.4. Consultas de GDBench redefinidas.	35
4.1. Variación de tiempo del workload real respecto al workload aleatorio.	47
4.2. Tiempo promedio para consultas según tipo de workload en Sparksee.	48
4.3. Tiempo promedio para consultas según tipo de workload en Neo4j.	49
4.4. Tiempo promedio para consultas según tipo de workload en JenaTDB.	50
4.5. Tiempo promedio para consultas según tipo de workload en PostgreSQL.	51
A.1. Ejecución consulta Q1 según workload.	60
A.2. Ejecución consulta Q2 según workload.	61
A.3. Ejecución consulta Q3 según workload.	62
A.4. Ejecución consulta Q4 según workload.	63
A.5. Ejecución consulta Q5 según workload.	64
A.6. Ejecución consulta Q6 según workload.	65
A.7. Ejecución consulta Q7 según workload.	66
A.8. Ejecución consulta Q8 según workload.	67
A.9. Ejecución consulta Q9 según workload.	68
A.10. Ejecución consulta Q10 según workload.	69
A.11. Ejecución consulta Q11 según workload.	70
A.12. Ejecución consulta Q12 según workload.	71