

LDBC

*The graph & RDF
benchmark reference*

***Semantic Publishing Benchmark
Full Disclosure Report
Irina Fundulaki
10/06/2015***

Table of Contents

Preface	3
General Terms	4
1. System Description	5
1.1 Database System	5
1.2 Database Engine Configuration	5
1.3 Platform Description	5
1.4 Network Infrastructure Information	6
2. Data Generation & Loading	7
2.1 Dataset Information	7
2.1.1 Description	7
2.1.2 External Datasets	7
2.1.3 Dataset Characteristics	8
2.1.4 Data Generator Parameters	8
2.2 Data Generation Times	9
2.3 Bulk Loading	9
3. Benchmark Test Driver	9
3.1. Basic test driver configuration details	9
3.2. Configuration Parameters for Driver Warmup	9
3.3. Configuration Parameters for Driver Execution	10
3.3. Test Driver Reference Data	10
4. Performance Metrics	11
4.1. SPB Primary Metrics	11
4.2. Query Execution Report	11
5. Recovery	13
6. Pricing Summary	14
7. Attachment's CheckList	15

Preface

This is the Full Disclosure Report for the benchmark results produced by Semantic Publishing Benchmark v2.0 for GraphDB SE 6.3 alpha, Scale Factor SF3 (256M Triples), single CPU server, with 64GB of RAM and 2xSSD drives.

General Terms

Test Sponsor

Test sponsor of this benchmark is Ontotext AD, Bulgaria. It is a leading provider of core semantic technology distinctive for its performance, scale, robustness and compliance with open standards. Ontotext is unique as a developer that provides state of the art semantic technology in two distinct but complementary areas – semantic graph database engines and text analytics. Ontotext solutions have been applied for business critical projects in the areas of publishing (BBC, Financial Times, Oxford University Press), life sciences (Astra Zeneca), cultural heritage (British Museum, Getty Trust), telecommunications (Korea Telecom), government organizations (UK Parliament) and others.

Ontotext is the developer of GraphDB semantic graph database.

1. System Description

1.1 Database System

Vendor Name	ONTOTEXT AD
Database Name	GRAPHDB STANDARD EDITION
Version Number	6.3 ALPHA, BUILD 4888f632
Database Engine Configuration	SINGLE NODE; Java HotSpot(TM) 64-bit server VM (build 24.75-b04, mixed mode)

Table 1: DBMS Characteristics

1.2 Database Engine Configuration

Cache Configuration	18G CACHE MEMORY OF 54G TOTAL RESERVED MEMORY (JAVA -XMX54G)
Transaction Isolation Level/Model	READ COMMITTED
Reasoning Strategy/configuration	FORWARD-CHAINING
triple/quadruple indices	POS, PSO, PCSO, SPOC, PRED. LISTS
indices	GEO-SPATIAL, TEXT INDEXES (LUCENE CONNECTOR)
Optimisations	OWL:SAMEAS

Table 2: Database Engine Configuration

1.3 Platform Description

Model	SUPERMICRO X10SRI-F, SOCKET R3, INTEL C612
Processors	CPU INTEL XEON E5-1650 V3 3.5GHZ, 15MB L3 CACHE, S2011
Memory	MEMORY 4 X 16GB DDR4-2133 2Rx4 ECC REG DIMM
No. Disks/Type/Storage Configuration	1 X HDD WESTERN DIGITAL HDD 500GB SATAIII RAID EDITION RE4 - 7200RPM 64MB CACHE – FOR THE OPERATING SYSTEM ONLY 2 X SSD DRIVES 400 GB SSD , SAMSUNG 845DC PRO
Network Adapters	I350 DUAL PORT GIGABIT LAN
Operating System	14.04 UBUNTU 64 BIT
File System	EXT4, SOFTWARE RAID-0 (SSD)
CPU Type & Count	INTEL XEON, 1 CPU
No. Threads	12
No. Cores	6

Memory	64 GB
Total Disks Capacity	734 GB (SSD), 1 X 413 GB (HDD)

Table 3: System Configuration

No. disks	2 SSD, 1 HDD
Storage interfaces	SATA3
Storage Technologies	SSD, HDD
RAID/HBA controller	SOFTWARE INTEL CORPORATION PATSBURG 6-PORT SATA AHCI CONTROLLER (REV 06), RAID-0

Table 4: No. Disks/Type/Storage/Configuration

1.4 Network Infrastructure Information

Model	NO EXTERNAL NETWORK
Network Switches	N/A
Wiring Information	N/A

Table 5: Network Infrastructure Information

2. Data Generation & Loading

Generated 231 M triples, using SPB's data generator. Loaded using SPB's loading process.

2.1 Dataset Information

2.1.1 Description

Scale Factor	SF3, 256M
Reference Data Size	25M
Data Format	NQUADS
Data Generator Version	2.0.1369E4A
Time Compression ratio	N/A

Table 6: Dataset characteristics

2.1.2 External Datasets

External datasets and ontologies can be found in the directories provided with the SPB distribution. The directories where this information is located are:

./DATA/DATASETS	
INTERNATIONAL-FOOTBALL-COMPETITIONS-3.TTL	ENTITIES_RANKS.TTL
INTERNATIONAL-FOOTBALL-TEAMS-2.TTL	GEONAMES_EUROPE.TTL
SCOTTISH-FOOTBALL-COMPETITIONS-1.TTL	DBPEDIA_COMPANIES.TTL
SCOTTISH-FOOTBALL-TEAMS-2.TTL	DBPEDIA_EVENTS.TTL
FORMULA1-COMPETITIONS-8.TTL	ENTITIES_PREFLABELS.TTL
FORMULA1-TEAMS-3.TTL	GEONAMES_SAMEAS_LINKS_EN_EN.TTL
ENGLISH-FOOTBALL-COMPETITIONS-1.TTL	DBPEDIA_PERSONDATA.TTL
ENGLISH-FOOTBALL-TEAMS-2.TTL	UK-PARLIAMENT-IDENTIFIERS-PEOPLE-8.TTL
./DATA/ONTOLOGIES/CORE	
CMS.1.2.TTL	CREATIVework.0.9.TTL
COMPANY.1.4.TTL	GEONAMES_ONTOLOGY_V3.1-SPB.TTL
CORECONCEPTS.0.6.TTL	PERSON.0.2.TTL
PROVENANCE.1.1.TTL	SYSTEMLOGIC.TTL
TAGGING.1.0.TTL	
./DATA/ONTOLOGIES/DOMAIN	
CNEWS-1.2.TTL	CURRICULUM-ONTOLOGY-4.TTL
SPORT.2.3.TTL	

Table 7: External Datasets & Ontologies (Listing)

2.1.3 Dataset Characteristics

Statistics for the datasets are produced and stored in file
./generated.256/dataset.info

NO. GENERATED TRIPLES	231000000
NO. CREATIVE WORKS	8821390
NO. TOTAL TRIPLES	481740264

Table 8: Dataset Characteristics

2.1.4 Data Generator Parameters

Data Generator Parameters can be found in files definitions.properties and test.256.properties.

TEST.256.PROPERTIES
DATAGENERATORWORKERS=8 DATASIZE=231000000 CREATIVEWORKSPATH=./GENERATED.256 QUERYSUBSTITUTIONPARAMETERS=100000 ENABLECOMPRESSIONONGENERATEDDATA=FALSE
DEFINITIONS.PROPERTIES
ABOUTSALLOCATIONS=0.1006, 0.2313, 0.3088, 0.2278, 0.1035, 0.028 MENTIONSALLOCATIONS=0.9477, 0.0382, 0.0093, 0.0031, 0.0012, 0.0005 ENTITYPOPULARITY=0.05, 0.95 USEPOPULARENTITIES=0.3, 0.7 CREATIVEWORKTYPESALLOCATION=0.45, 0.35, 0.2 ABOUTANDMENTIONSALLOCATION=0.85, 0.15 EDITORIALOPERATIONSALLOCATION=0.8, 0.1, 0.1 AGGREGATIONOPERATIONSALLOCATION=0.087, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083 EXPONENTIALDECAYUPPERLIMITOFCWS=5000 EXPONENTIALDECAYRATE=0.1 EXPONENTIALDECAYTHRESHOLDPERCENT=0.05 MAJOREVENTS=5 MINOREVENTS=100 SEEDYEAR=2011 DATAGENERATIONPERIODYEARS=1 CORRELATIONSAMOUNT=50 CORRELATIONSMAGNITUDE=60 CORRELATIONDURATION=0.2 CORRELATIONENTITYLIFESPAN=0.5 MINLAT = 36.5 MAXLAT = 70.0 MINLONG = -9.5 MAXLONG = 31.0

Table 9: Data Generator Parameters

2.2 Data Generation Times

Start Time (in hh:mm:ss)	16:31:09
End Time (in hh:mm:ss)	16:46:00
Generation Time (in s)	870.984

Table 10: Data Generation Times

2.3. Bulk Loading

Start Time (in hh:mm:ss)	16:46:00
End Time (in hh:mm:ss)	22:32:00
Bulk Load Time (in s)	19654.705

Table 11: Bulk Loading Time

3. Benchmark Test Driver

3.1. Basic test driver configuration details

Basic Test driver configuration details can be found in test.256.properties file.

No. read threads	8
No. write threads	2

Table 12: Basic Test Driver Configuration Details

3.2. Configuration Parameters for Driver Warmup

Test driver configuration details for Driver Warmup can be found in ./test.256.properties file.

WARMUPPERIODSECONDS=600 AGGREGATIONAGENTS=8 QUERYTIMEOUTSECONDS=300 WARMUP=TRUE
--

Table 13: Configuration Parameters for Driver Warmup

--	--

3.3. Configuration Parameters for Driver Execution

Test driver configuration details for Driver Execution can be found in test.256.properties file.

```

AGGREGATIONAGENTS=8
EDITORIALAGENTS=2
BENCHMARKRUNPERIODSECONDS=1800
QUERYTIMEOUTSECONDS=300
BENCHMARKBYQUERYMIXRUNS=0
MINUPDATERATETHRESHOLDOPS=0.0
MINUPDATERATETHRESHOLDREACHTIMEPERCENT=0
MAXUPDATERATETHRESHOLDOPS=0.0
ENDPOINTURL=HTTP://LOCALHOST:4082/OPENRDF-SESAME/REPOSITORIES/LDBC256
ENDPOINTUPDATEURL=HTTP://LOCALHOST:4082/OPENRDF-
SESAME/REPOSITORIES/LDBC256/STATEMENTS
QUERIESPATH=./DATA/SPARQL
RUNBENCHMARK=TRUE
    
```

Table 14: Configuration Parameters for Driver Execution

3.3. Test Driver Reference Data

The reference data for the test driver are located at:

./DATA/DATASETS	
INTERNATIONAL-FOOTBALL-COMPETITIONS-3.TTL	ENTITIES_RANKS.TTL
INTERNATIONAL-FOOTBALL-TEAMS-2.TTL	GEONAMES_EUROPE.TTL
SCOTTISH-FOOTBALL-COMPETITIONS-1.TTL	DBPEDIA_COMPANIES.TTL
SCOTTISH-FOOTBALL-TEAMS-2.TTL	DBPEDIA_EVENTS.TTL
FORMULA1-COMPETITIONS-8.TTL	ENTITIES_PREFLABELS.TTL
FORMULA1-TEAMS-3.TTL	GEONAMES_SAMEAS_LINKS_EN_EN.TTL
ENGLISH-FOOTBALL-COMPETITIONS-1.TTL	DBPEDIA_PERSONDATA.TTL
ENGLISH-FOOTBALL-TEAMS-2.TTL	UK-PARLIAMENT-IDENTIFIERS-PEOPLE-8.TTL
./DATA/ONTOLOGIES/CORE	
CMS.1.2.TTL	CREATIVWORK.0.9.TTL
COMPANY.1.4.TTL	GEONAMES_ONTOLOGY_V3.1-SPB.TTL
CORECONCEPTS.0.6.TTL	PERSON.0.2.TTL
PROVENANCE.1.1.TTL	SYSTEMLOGIC.TTL
TAGGING.1.0.TTL	
./DATA/ONTOLOGIES/DOMAIN	
CNEWS-1.2.TTL	CURRICULUM-ONTOLOGY-4.TTL
SPORT.2.3.TTL	

Table 15: Reference Data and Ontologies

4. Performance Metrics

4.1. SPB Primary Metrics

Query Rate Interactive mix (queries/second)	Query Rate Analytical Mix (queries per second)	Update Rate (Operations per second)
54.6364	N/A	9.4967

Table 16: Primary Metrics

4.2. Query Execution Report

Duration of Bulk Load (in s)	Duration of Measurement Window (in minutes)	No. complete analytical mixes	No. complete interactive mixes	No. complete update operations performed
19654.705	30MIN	N/A	8009	17094

Table 17: Query Execution Report

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Avg Exec Time	Count of Executions
Q1	203	4	2522	204	8013
Q2	7	2	1414	7	8014
Q3	48	8	3912	48	8012
Q4	23	2	2991	22	8014
Q5	93	3	5873	93	8013
Q6	171	48	4965	170	8014
Q7	87	5	5568	87	8011
Q8	155	88	5562	154	8014
Q9	223	4	5703	223	8015
Q10	583	1	6392	585	8014
Q11	26	3	3277	25	8013
Q12	122	4	5532	123	8013

Table 18: Interactive Mix Report

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Exec Time	Count of Executions
Q1-Q13	N/A	N/A	N/A	N/A	N/A

Table 19: Analytical Mix Report

--	--

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Exec Time	Count of Executions
inserts	55	25	5493	55	13696
updates	120	65	1194	120	1707
deletes	64	33	1843	62	1691

Table 20: Update Operation Report

5. Recovery

Execution Duration	N/A
Time to Recover	N/A

Table 21: Recovery

--	--

6. Pricing Summary

	<i>Item</i>	<i>Price</i>
	SERVER CASE SUPERMICRO CSE-825TQ-R740LPB, 2X740W REDUNDANT 80 PLUS PLATINUM	
	MOTHERBOARD SUPERMICRO X10SRI-F, SOCKET R3, INTEL C612, UP TO 256GB DDR4, 1 PCI-E 3.0 X16, 2 PCI- 1	
	CPU INTEL XEON E5-1650 v3 (6 CORES) 3.5GHZ,15MB L3 CACHE, S2011	
	COOLER SNK-P0048AP4, SUPERMICRO ,2U ACTIVE CPU HS FOR 2U ,SQUARE&NARROW	
	MEMORY 16GB DDR4-2133 2RX4 ECC REG DIMM	
	SSD DRIVES 400 GB SSD , SAMSUNG 845DC PRO	
	HDD WESTERN DIGITAL HDD 500GB SATAIII RAID EDITION RE4 - 7200RPM	
TOTAL		3 223 EUR
	GRAPHDB STANDARD EDITION, SINGLE NODE LICENSE (7 200 EUR), 3 YEARS OF MAINTENANCE (5400 EUR), 24X7 SUPPORT PACKAGE (19 500 EUR)	32 100 EUR
TOTAL		35 323 EUR

Table 22: Pricing Information

7. Attachment's CheckList

CLAUSE 1	RDF DATABASE LOAD SCRIPTS IF ANY	RDF_DATABASE_SCRIPTS.ZIP
CLAUSE 2	RDF DATABASE CONFIGURATION OF MORE DETAILS ARE AVAILABLE	RDF_DATABASE_CONFIGURATION.ZIP
CLAUSE 3	OPERATING SYSTEM AND HARDWARE SETTINGS	OPERATING_SYSTEM_HARDWARE_SETTINGS.ZIP
CLAUSE 4	BENCHMARK TEST DRIVER CONFIGURATION AND ANY MODIFIED QUERY TEMPLATES AND (OR) ONTOLOGIES / REFERENCE DATASETS	BENCHMARK_TEST_DRIVER_CONFIG.ZIP
CLAUSE 5	ALL OF THE LOG FILES GENERATED BY THE BENCHMARK DRIVER	LOG_FILES.ZIP

Table 23: Checklist