



*The graph & RDF
benchmark reference*

***Semantic Publishing Benchmark
Full Disclosure Report
Irina Fundulaki
08/06/2015***

Table of Contents

Preface	3
General Terms	4
1. System Description	5
1.1 Database System	5
1.2 Database Engine Configuration	6
1.3 Platform Description	6
1.4 Network Infrastructure Information	7
2. Data Generation & Loading	8
2.1 Dataset Information	8
2.1.1 Description	8
2.1.2 External Datasets	8
2.1.3 Dataset Characteristics	9
2.1.4 Data Generator Parameters	9
2.2 Data Generation Times	10
2.3. Bulk Loading	10
3. Benchmark Test Driver	10
3.1. Basic test driver configuration details	10
3.2. Configuration Parameters for Driver Warmup	10
3.3. Configuration Parameters for Driver Execution	10
3.3. Test Driver Reference Data	11
4. Performance Metrics	11
4.1. SPB Primary Metrics	11
4.2. Query Execution Report	12
5. Recovery	14
6. Pricing Summary	15
7. Attachment's CheckList	16

Preface

This is the Full Disclosure Report for the Semantic Publishing Benchmark and for Virtuoso and for Scale Factor SF1 (64M Triples).

General Terms

Test Sponsor

Test sponsor of this benchmark is OpenLink Software.

1. System Description

Entry level server hardware.

1.1 Database System

Vendor Name	OPENLINK SOFTWARE
Database Name	VIRTUOSO OPENSOURCE
Version No	7.50.3213
Database Engine Configuration	<pre>[DATABASE] DATABASEFILE = VIRTUOSO.DB TRANSACTIONFILE = VIRTUOSO.TRX ERRORLOGFILE = VIRTUOSO.LOG ERRORLOGLEVEL = 7 SYSLOG = 0 TEMPSTORAGE = TEMPDATABASE FILEEXTEND = 200 STRIPING = 1 [TEMPDATABASE] DATABASEFILE = VIRTUOSO.TDB TRANSACTIONFILE = VIRTUOSO.TTR FILEEXTEND = 200 ; ; SERVER PARAMETERS ; [PARAMETERS] SERVERPORT = 1607 SERVERTHREADS = 100 CHECKPOINTSYNCMODE = 0 CHECKPOINTINTERVAL = 0 NUMBEROFBUFFERS = 2000000 MAXDIRTYBUFFERS = 1700000 MAXCHECKPOINTREMAP = 2500000 DEFAULTISOLATION = 2 UNREMAPQUOTA = 0 CASEMODE = 2 DIRSALLOWED = / PLDEBUG = 0 TESTCOVERAGE = COV.XML RESOURCESCLEANUPINTERVAL = 1 THREADCLEANUPINTERVAL = 1 TRANSACTIONAFTERIMAGELIMIT = 1500000000 FDSPERFILE = 4 MAXMEMPOOLSIZE = 40000000 ADJUSTVECTORSIZE = 1 THREADSPERQUERY = 24 ASYNCQUEUEMAXTHREADS = 24 MAXQUERYMEM = 30G HASHJOinspace = 30G INDEXTREEMAPS = 64 [CLIENT] SQL_QUERY_TIMEOUT = 0</pre>

<pre> SQL_TXN_TIMEOUT = 0 SQL_ROWSET_SIZE = 10000 SQL_PREFETCH_BYTES = 120000 [HTTPSERVER] SERVERPORT = 8897 SERVERROOT = VSP SERVERTHREADS = 40 MAXKEEPALIVES = 10 KEEPALIVETIMEOUT = 10 MAXCACHEDPROXYCONNECTIONS = 10 PROXYCONNECTIONCACHETIMEOUT = 10 DAVROOT = DAV HTTPLOGFILE = LOGS/HTTP01042015.LOG [SPARQL] RESULTSETMAXROWS = 2000000 LABELINFERENCENAME = FACETS IMMUTABLEGRAPHS = INFERENCE-GRAPHS, * SHORTENLONGURIS = 1 [STRIPING] SEGMENT1 = 1024, /1D1/DBS/SPB64.DB = Q1, /1D2/DBS/SPB64.DB = Q2, /1D3/DBS/SPB64.DB = Q3, /1D4/DBS/SPB64.DB = Q4, /1D5/DBS/SPB64.DB = Q5 [FLAGS] ENABLE_MT_TXN = 1 ENABLE_MT_TRANSACT = 1 QP_THREAD_MIN_USEC = 100 MP_LOCAL_RC_SZ = 0 DBF_EXPLAIN_LEVEL = 0 ENABLE_EXACT_P_STAT = 1 HASH_JOIN_ENABLE = 2 ENABLE_G_IN_SEC = 1 </pre>

Table 1: DBMS Characteristics

1.2 Database Engine Configuration

Cache Configuration	MAXCACHEDPROXYCONNECTIONS = 10 PROXYCONNECTIONCACHETIMEOUT = 10
Transaction Isolation Level/Model	READ COMMITTED
Reasoning Strategy/configuration	BACKWARD REASONING
triple/quadruple indices	PSOG, POSG, GS, SP, OP
indices	FULL TEXT INDICES
Optimisations	SAMEAS

Table 2: Database Engine Configuration

1.3 Platform Description

Model	CUSTOM BUILT
Processors	2 X INTEL XEON E5-2630, 6X 2.30GHZ, SOCKEL 2011, BOXED

Memory	3 X CORSAIR XMS3 DIMM KIT 64GB, DDR3-1333, CL9-9-9-24
No Disks/Type/Storage Configuration	8 DISKS STORAGE INTERFACES: INTEL CORPORATION PATSBURG 6-PORT SATA AHCI CONTROLLER (REV 06) 6 X SEAGATE CONSTELLATION ES 2TB, SATA 3GB/S 2 X CRUCIAL MX100 512GB, SATA 6GB/S
Network Adapters	N/A
Operating System	CENTOS RELEASE 6.2 (FINAL)
File System	EXT4 (ON HDDS), EXT2 (ON SSDS)
No Threads	24
No Cores	12
Memory	192 GB
Total Disks Capacity	13TB

Table 3: System Configuration

No disks	8
Storage interfaces	INTEL CORPORATION PATSBURG 6-PORT SATA AHCI CONTROLLER (REV 06)
Storage Technologies	
RAID/HBA controller	N/A

Table 3: No Disks/Type/Storage/Configuration

1.4 Network Infrastructure Information

Model	N/A
Network Switches	N/A
Wiring Information	N/A

Table 4: Network Infrastructure Information

--	--

2. Data Generation & Loading

Generated 64 M triples, using SPB's data generator. Loaded using SPB's loading process.

2.1 Dataset Information

2.1.1 Description

Scale Factor	SF1, 64M
Reference Data Size	25M
Data Format	NQUADS
Data Generator Version	SPB VERSION 2.0 (SPB_BM_2.0)
Time Compression ratio	

Table 5: 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	CREATIVELWORK.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 6: External Datasets & Ontologies (Listing)

2.1.3 Dataset Characteristics

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

NO GENERATED TRIPLES	63709210
NO CREATIVE WORKS	1489675
NO INFERRED TRIPLES	N/A

Table 7: Dataset Characteristics

2.1.4 Data Generator Parameters

Data Generator Parameters can be found in files ./definitions.properties and
./spb64.properties

./SPB64.PROPERTIES DATAGENERATORWORKERS=4 DATASETSIZE= 39000000 CREATIVEWORKSPATH=./GENERATED 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 8: Data Generator Parameters

--	--

2.2 Data Generation Times

Start Time (in hh:mm:ss)	20:41:57
End Time (in ms)	20:48:36
Generation Time	399

Table 9: Data Generation Times

2.3. Bulk Loading

Start Time	20:48:36
End Time	21:09:46
Bulk Load Time	1270 (OUT OF WHICH 1188 FOR FULL TEXT INDEXING)

Table 10: Bulk Loading Time

3. Benchmark Test Driver

3.1. Basic test driver configuration details

Basic Test driver configuration details can be found in ./spb64.properties file.

No of read threads	22
No of write threads	2

Table 11: Basic Test Driver Configuration Details

3.2. Configuration Parameters for Driver Warmup

Test driver configuration details for Driver Warmup can be found in ./spb64.properties file.

WARMUPPERIODSECONDS=600 AGGREGATIONAGENTS=22 QUERYTIMEOUTSECONDS=300 WARMUP=TRUE

Table 12: Configuration Parameters for Driver Warmup

3.3. Configuration Parameters for Driver Execution

Test driver configuration details for Driver Execution can be found in ./spb64.properties file.

```

AGGREGATIONAGENTS=22
EDITORIALAGENTS=2
BENCHMARKRUNPERIODSECONDS=1200
QUERYTIMEOUTSECONDS=300
BENCHMARKBYQUERYMIXRUNS=0
MINUPDATERATETHRESHOLDOPS=0.0
MINUPDATERATETHRESHOLDREACHTIMEPERCENT=0
MAXUPDATERATETHRESHOLDOPS=0.0
ENDPOINTURL=HTTP://LOCALHOST:8897/SPARQL
ENDPOINTUPDATEURL=HTTP://LOCALHOST:8897/SPARQL
QUERIESPATH=./DATA/SPARQL
RUNBENCHMARK=TRUE

```

Table 13: 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	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	

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)
149.0385	NA	156.8325

Table 14: Primary Metrics

4.2. Query Execution Report

Duration of Bulk Load (in ms)	Duration of Measurement Window (in minutes)	No complete analytical mixes per second	No complete interactive mixes	No complete update operations performed
1270	20MIN	NA	14510	188199

Table 10: 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	280	143	491	280	14516
Q2	110	60	190	110	14518
Q3	183	90	425	183	14517
Q4	43	12	1601	43	14522
Q5	120	16	756	120	14518
Q6	104	24	712	103	14522
Q7	99	33	795	99	14521
Q8	377	232	1922	377	14518
Q9	99	27	192	99	14519
Q10	227	14	1607	227	14516
Q11	94	3	238	94	14519
Q12	30	5	104	30	14520

Table 11: Interactive Mix Report

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Exec Time	Count of Executions
Q1	NA	NA	NA	NA	NA
Q2	NA	NA	NA	NA	NA
Q3	NA	NA	NA	NA	NA
Q4	NA	NA	NA	NA	NA
Q5	NA	NA	NA	NA	NA
Q6	NA	NA	NA	NA	NA
Q7	NA	NA	NA	NA	NA
Q8	NA	NA	NA	NA	NA
Q9	NA	NA	NA	NA	NA

Q10	NA	NA	NA	NA	NA
Q11	NA	NA	NA	NA	NA
Q12	NA	NA	NA	NA	NA

Table 15: Analytical Mix Report

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Exec Time	Count of Executions
inserts	10	4	972	10	150589
updates	20	7	449	20	18805
deletes	11	2	156	11	18805

Table 16: Update Operation Report

--	--

5. Recovery

Execution Duration	NA
Time to Recover	NA

Table 14: Recovery

6. Pricing Summary

	<i>Item</i>	<i>Price</i>
	2 X INTEL XEON E5-2630, 6X 2.30GHZ, SOCKET 2011, BOXED	
	3 X CORSAIR XMS3 DIMM KIT 64GB, DDR3-1333, CL9-9-9-24	
	6 X SEAGATE CONSTELLATION ES 2TB, SATA 3GB/S 2 X CRUCIAL MX100 512GB, SATA 6GB/S	
	MEMORY: 192 GB	
	TOTAL DISK CAPACITY: 13 TB	
	INTEL CORPORATION PATSBURG 6-PORT SATA AHCI CONTROLLER (REV 06)	
TOTAL		5212,59 USD
	VIRTUOSO OPEN SOURCE 7.50.3213 , 3 YEARS OF MAINTENANCE	15000 USD
TOTAL		20212,59 USD

Table 15: 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 17: Checklist