Semantic Publishing Benchmark Full Disclosure Report
Irini Fundulaki
10/06/2015

Table of Contents

Preface

This is the Full Disclosure Report for the Semantic Publishing Benchmark and for Virtuoso and for Scale Factor SF3 (256M 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	[DATABASE] DATABASEFILE = VIRTUOSO.DB TRANSACTIONFILE = /1S2/DBS/SPB256.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 = 1111 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 = 32 ASYNCQUEUEMAXTHREADS = 64 MAXQUERYMEM = 30G HASHJOINSPACE = 30G INDEXTREEMAPS = 64 [CLIENT] SQL_QUERY_TIMEOUT = 0

```
SQL_TXN_TIMEOUT    = 0
SQL_ROWSET_SIZE    = 10000
SQL_PREFETCH_BYTES = 120000
[HTTPServer]
ServerPort            = 8890
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, /1S1/DBS/SPB256.DB = Q1, /1S2/DBS/SPB256.DB =
Q2
[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
```

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	AMAZON EC2 R3.8XLARGE

Table 3: System Configuration

2. Data Generation & Loading

Generated 256 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	SPB VERSION 2.0 (SPB_BM_2.0)
Time Compression ratio	

Table 3: 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 4: External Datasets & Ontologies (Listing)

2.1.3 Dataset Characteristics
Statistics for the datasets are produced and stored in file

./dataset.info

No Generated Triples	255709208
No Creative Works	8521474
No Inferred Triples	N/A

Table 5: Dataset Characteristics

2.1.4 Data Generator Parameters

Data Generator Parameters can be found in files ./definitions.properties and ./spb256.properties

./SPB256.PROPERTIES
```
DATAGENERATORWORKERS=4
DATASETSIZE=  231000000
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 6: Data Generator Parameters

2.2 Data Generation Times

Start Time (in hh:mm:ss)	09:38:45
End Time (in hh:mm:ss)	10:00:50
Generation Time (in s)	1325

Table 7: Data Generation Times

2.3. Bulk Loading

Start Time (in hh:mm:ss)	10:00:50
End Time (in hh:mm:ss)	11:42:52
Bulk Load Time (in s)	6120 (OUT OF 2144 WHICH FOR FULL TEXT INDEXING)

Table 8: Bulk Loading Time

3. Benchmark Test Driver

3.1. Basic test driver configuration details

Basic Test driver configuration details can be found in ./spb256.properties file.

No of read threads	30
No of write threads	3

Table 9: Basic Test Driver Configuration Details

3.2. Configuration Parameters for Driver Warmup

Test driver configuration details for Driver Warmup can be found in ./spb256.properties file.

```
WARMUPPERIODSECONDS=600
AGGREGATIONAGENTS=30
QUERYTIMEOUTSECONDS=300
WARMUP=TRUE
```

Table 10: Configuration Parameters for Driver Warmup

3.3. Configuration Parameters for Driver Execution

Test driver configuration details for Driver Execution can be found in ./spb256.properties file.


```
AGGREGATIONAGENTS=30
EDITORIALAGENTS=3
BENCHMARKRUNPERIODSECONDS=1200
QUERYTIMEOUTSECONDS=300
BENCHMARKBYQUERYMIXRUNS=0
MINUPDATERATETHRESHOLDOPS=0.0
MINUPDATERATETHRESHOLDREACHTIMEPERCENT=0
MAXUPDATERATETHRESHOLDOPS=0.0
ENDPOINTURL=HTTP://LOCALHOST:8890/SPARQL
ENDPOINTUPDATEURL=HTTP://LOCALHOST:8890/SPARQL
QUERIESPATH=./DATA/SPARQL
RUNBENCHMARK=TRUE
```

Table 11: 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)
115.3838	NA	109.8517

Table 12: 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
6120	20MIN	NA	11238	131822

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	304	114	745	304	11249
Q2	116	61	325	116	11252
Q3	328	165	590	329	11251
Q4	95	31	299	95	11254
Q5	355	44	4344	356	11253
Q6	290	21	3985	289	11249
Q7	160	27	7526	160	11249
Q8	661	418	8401	660	11244
Q9	118	36	376	118	11251
Q10	508	16	3794	505	11246
Q11	101	25	443	101	11250
Q12	53	5	250	53	11251

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 13: Analytical Mix Report

Query	Arithmetic Mean of Execution Time	Min Exec Time	Max Exec Time	90th percentile of Exec Time	Count of Executions
inserts	18	4	198	18	105653
updates	26	10	240	26	12028
deletes	14	3	165	14	13141

Table 14: Update Operation Report

5. Recovery

Execution Duration	NA
Time to Recover	NA

Table 14: Recovery

6. Pricing Summary

	Item	Price
TOTAL	AMAZON EC2 R3.8XLARGE 3 YEARS MAINTENANCE, 24/7	24528 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 15: Checklist