ORACLE

ORACLE WEBCENTER CONTENT 12C PERFORMANCE RESULTS

PERFORMANCE TESTING ON WEBCENTER CONTENT WITH ORACLE TEXT SEARCH AND ELASTIC SEARCH

An Oracle White Paper, 1st Edition

April 20, 2021 | Version 1.0

TABLE OF CONTENTS

PURPOSE STATEMENT

This document covers scalability and sizing recommendations for Web Center Content (WCC) sizing effort on Oracle Text Search (OTS) and Elastic Search (ES) as part of Web Center Content Suite Release 12.2.1.4.0Suite Release 12.2.1.4.0

DISCLAIMER

INTRODUCTION

Oracle WebCenter Content, an Oracle Fusion Middleware component, is an integrated suite of applications designed for managing content. It consolidates unstructured content from across diverse systems so it can be centrally managed and then exposes it from within desktop productivity tools, business applications, and mobile devices to fit the needs of today's users.

Currently, Web Center Content supports a variety of search indexer engines.
Among all the engines, Oracle Text Search (OTS) provides a rich searching capability including full text searches with relevancy ranking, complex query structures, and improved performance compared to the existing full-text search engine DATABASE.FULLTEXT. However, in a large enterprise setup with millions of content items and a high ingestion use-case, rebuilding of OTS index is very slow and is often a multi-day activity.
Integrating Elastic search with WebCenter Content can reduce the rebuild time and enable an improved near real-time search response. In this integration, all the existing search features are retained and the APIs and user interfaces remain untouched, too.

HIGH VOLUME CONTENT & IMAGING APPLICATION CHARACTERISTICS

In many large organizations, critical business applications depend on handling large volumes of content and manage millions of new documents every day. Accounts payable, accounts receivable, and human resources deal with extremely large volumes of paper documents that need to be scanned and contributed to a repository where they can be managed.
In certain industries, such as insurance, midsized organizations also deal with high volumes of documents although the number of consumers may be low (perhaps only a few hundred). For a large enterprise setup where content items run into millions, Elastic search integration with WebCenter Content provides better performance compared to Oracle Text Search in terms of index rebuild time. Furthermore, content search functionality is streamlined so that content can be quickly retrieved using well known Oracle Text Search (e.g. content number, or customer ID). WebCenter Content communicates with Elastic search through Elastic search REST API.

OBJECTIVE

- Validate performance results of Single node OTS vs ES with 250k documents as a baseline.
- Validate performance results for Single node OTS vs. ES Cluster with 5M documents.

TEST METHODOLOGY

- This sizing activity is split it into following 3 phases

    1. Data Seeding
    2. Index Building
        a) 250k documents base (Single Node)
        b) 5M documents base (Cluster)

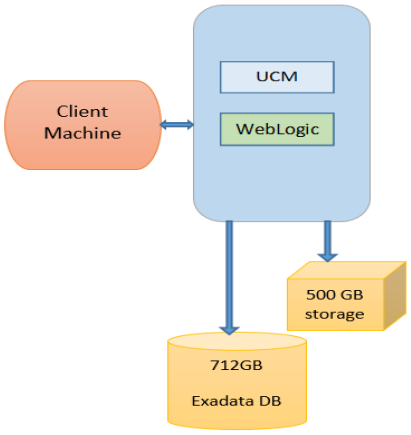
3. Documents Search Operation
    a) 250k documents base (Single Node)
    b) 5M documents base (Cluster)

1. DATA SEEDING

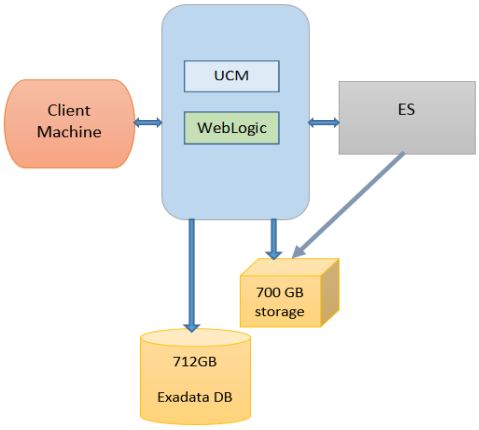
- Document types
    - Documents with formats such as .pdf,.docx,.txt,.jpeg, etc. are seeded to the system
- Seeding Methodology
    - Used CS (Content Server) Loader utility (Java swing app), Oats, Python and Selenium scripts to seed multiple documents which was done in a multithreaded manner
- Document shape
    - 250 K for baseline, further seeded to 5 M documents
- Document base for 5 M consists of
    - 800k records txt and doc type of 100 KB each
    - 2500k records txt and doc type of 10 KB each
    - 600K records .jpeg type of 11 KB each and 600k .png type of 5 KB to 10KB each
    - 500k records of .pdf type of 87 KB each
- Observation
    - Ingesting 5M documents into the system, it took almost 1 week of time.
    - During this process CPU and memory are moderate.

TOPOLOGY FOR SINGLE NODE

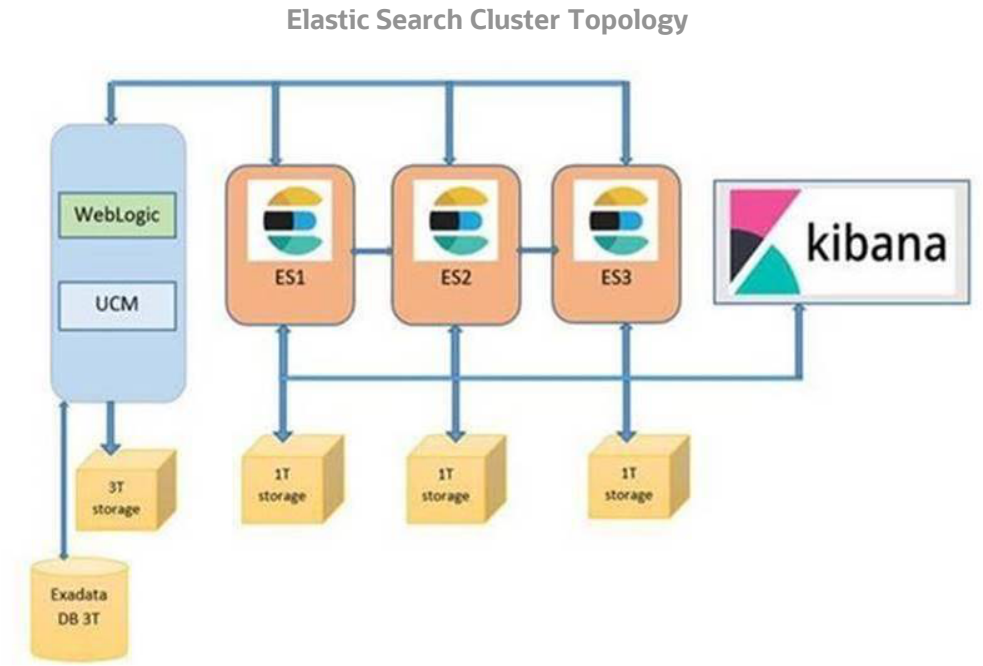
Oracle Text Search Topology



Elastic Search Topology



TOPOLOGY FOR CLUSTER



HARDWARE CONFIGURATION

This section covers about hardware configuration for WebCenter Content, where UCM is running on one server and Elastic server is running on three different application servers, and the database server is running on an Exadata node having below configuration. Refer to the following link for more details on OCI compute shapes:
https://docs.oracle.com/en-us/iaas/Content/Compute/References/computeshapes.htm

Table 1-1: WebCenter Content and Elastic Server Configurations

H/W DETAILS FOR APP SERVER	WCC and ES
Shape	VM.Standard1.8
CPU Count	8
Network Bandwidth (Gbps)	2.4
Memory (GB)	56
Local Disk	Block Storage Only

Table 1-2: Database Configurations

H/W DETAILS FOR DB SERVER	DB
Shape	VM.Standard2.24 - total 48 VCPU
CPU Count	48 VCPU
Available Data Storage	1070 GB
Total Storage Size	2048 GB
Memory (GB)	314 GB

WEBCENTER CONTENT WITH OTS AND ES BENCHMARK

This chapter provides the results of WebCenter Content on OTS and ES performed on single node w/ 250k documents and on cluster w/ 5 million documents using index operation and search operation scenario.

2. INDEX BUILDING

2.1 Index Building on 250k Documents Base Benchmark (Single Node)

Results

WCC Single Node OTS vs ES w/ 250k Doc Baseline									
Flow	#Users	Time Taken in hh:mm		Throughput* docs/Sec		CPU %			
		OTS	ES	OTS	ES	OTS	ES	DB on OTS	DB on ES
Ingest and Searchable	1	13:00	9:00	5	8	10	60	6	5
Fast Rebuild	1	1:14		56		16		5	
Fast Rebuild - New Metadata	1	1:15	0:40	56	104	11	70	7	6
Full Index Rebuild	1	12:30	8:00	6	9	18	55	6	5
Elastic search Reindex	1		0:30		139		>80		6
OTS to ES Migration	1		1:15		46				7
* --> throughput is the unit measured to show the no of documents processed per second									

Figure 1-1: WebCenter Content Single Node OTS vs ES 250k Doc Baseline Indexing Results

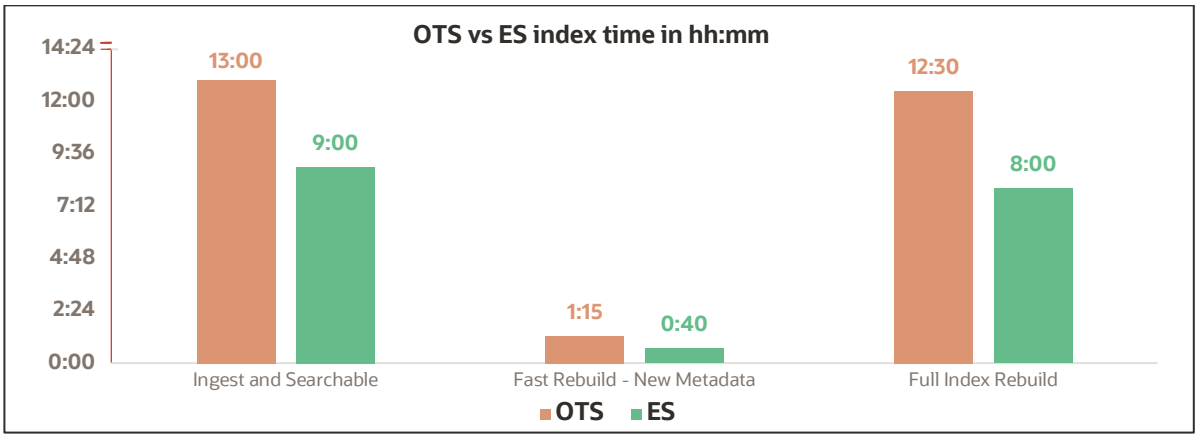
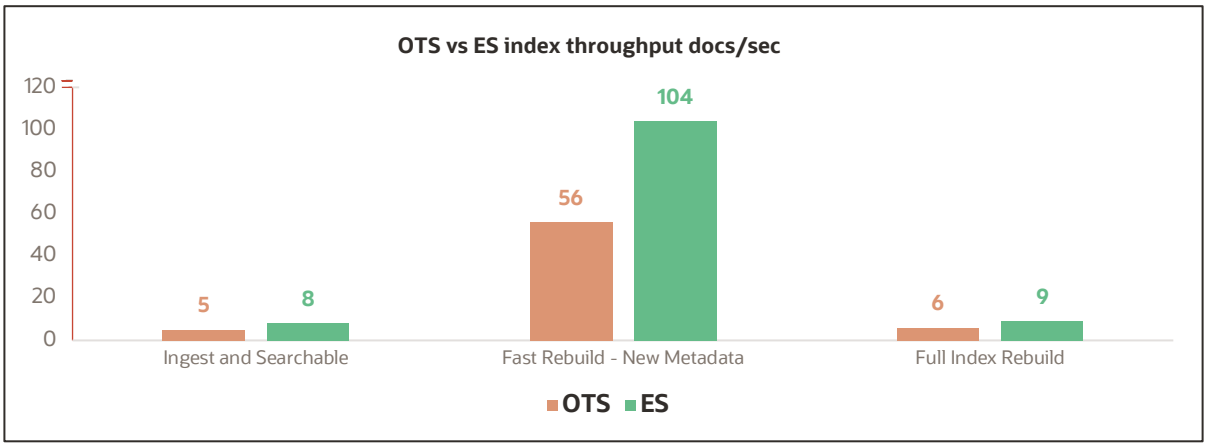


Figure 1-2: WebCenter Content Single Node OTS vs ES 250k Doc Baseline Throughput



Observation

- In indexing activity, Elastic search performs almost 30-50% faster compared to Oracle Text Search engine
- Throughput and time taken to complete the activity on Elastic search is faster compared to Oracle Text Search
- CPU and memory usage are slightly higher on a single-node Elastic search compared to Cluster Elastic search

2.2 Index Building on five million Documents Base Benchmark (Cluster)

Results

WCC 1 Node OTS vs 3 Node ES Cluster w/ 5M Docs Indexing Operation												
Flow	#Users	Time Taken in hh:mm:ss		Throughput docs/Sec		CPU%						
		OTS	ES	OTS	ES	OTS	UCM on ES	ES 1st node	2nd Node	3rd Node	OTS-DB	ES-DB
Ingest and Searchable	1		156:00:00		11		17	25	27	20		7
Fast Rebuild	1	20:00:00		70		20	18				5	
Fast Rebuild - New Metadata	1	20:30:00	11:15:00	68	124	35	37	67	69	70	5	8
Full Index Rebuild	1	155:00:00	110:00:00	9	13	11	10	35	38	40	4	3
Elastic search Reindex	1		14:15:00		98		15	40	43	45		7
OTS to ES Migration	1		9:00:00		154		25	70	72	73		6

Figure 1-3: WebCenter Content 1 Node OTS vs ES w/ 5M Doc Indexing Results

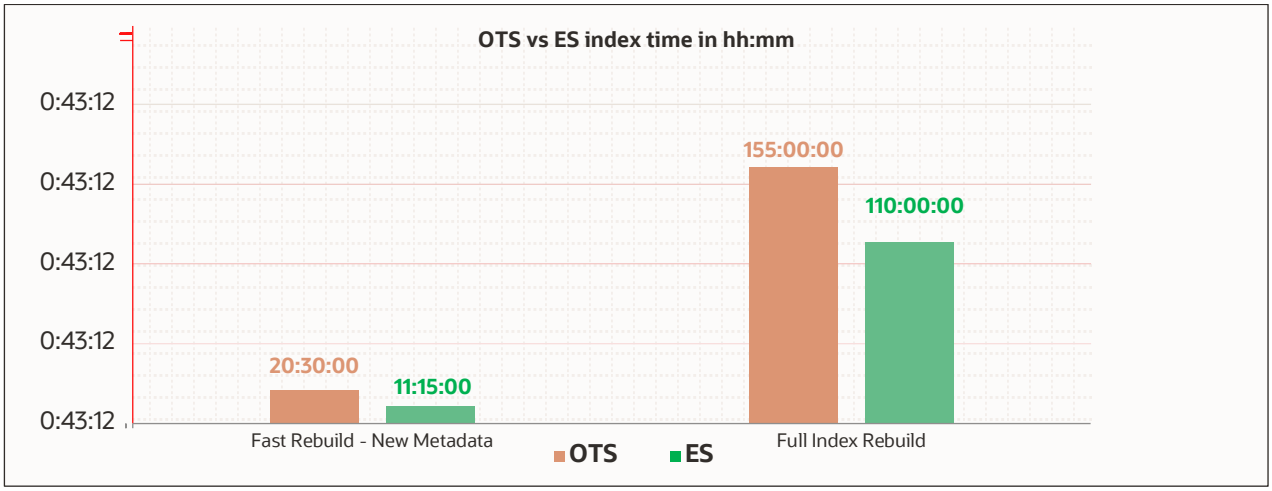
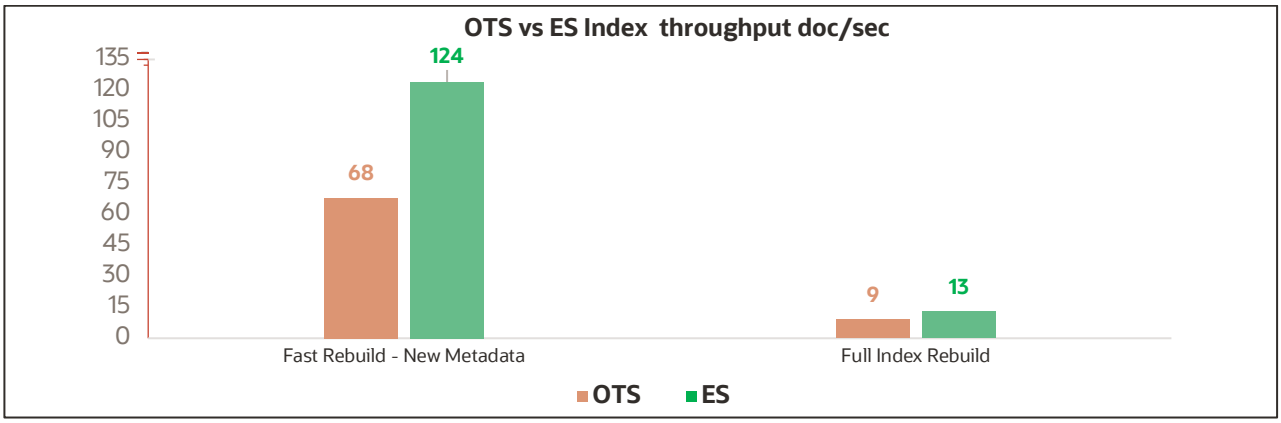


Figure 1-4: WebCenter Content 1 Node OTS vs ES w/ 5M Doc Indexing Throughput



Observation

- Throughput increased for indexing operations when per indexer batch cycle content settings on WebCenter Content has increased to 500.
- Observed 40-50% improvement on indexing flows when content items per indexer batch settings was changed from 25 to 500
- CPU and memory utilization is moderate when "per indexer batch cycle" set to 500. Also, performance has improved compared to that of default setting of 'per indexer batch cycle" set to 25
- Observed CPU and memory are distributing evenly across all the 3 nodes. Max heap utilized 26GB/40 GB for migration activity but for other flows heap was utilized almost 20GB/25 GB.
- Kibana was configured for monitoring Elastic search 3 node indexes. There were almost 160+ shards present in the system and all 5M docs are equally distributed across 4 security groups excluding additional 4 default groups present in WebCenter Content such as public, secure, IPMsys, IPMannotate.
- Max Shards size occupied was ~30 GB and almost 1.2 T block storage volume has consumed in all three nodes during the indexing and migration process. The total segment count has reached almost ~700
- Overall performance of indexing is faster in Elastic search 3 nodes cluster compare to Oracle Text Search

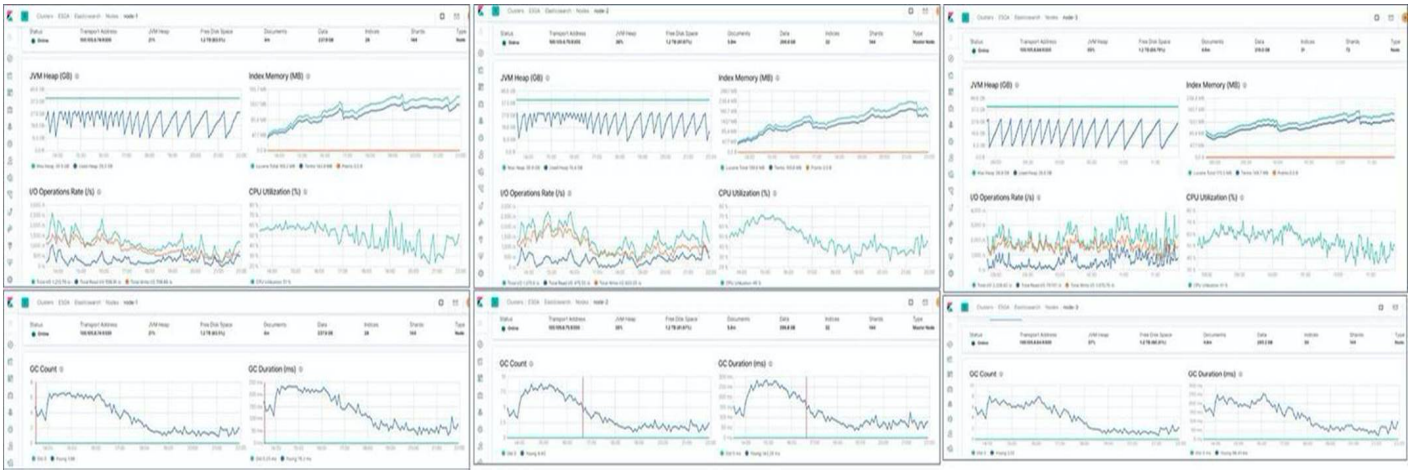
PERFORMANCE MONITORING TOOLS USED:

Kibana: - is a data visualization dashboard for Elastic search. It provides visualization capabilities on top of the content indexed on an Elastic search cluster. Kibana 7.6.0 version was installed to make compatible with ES 7.6.0.

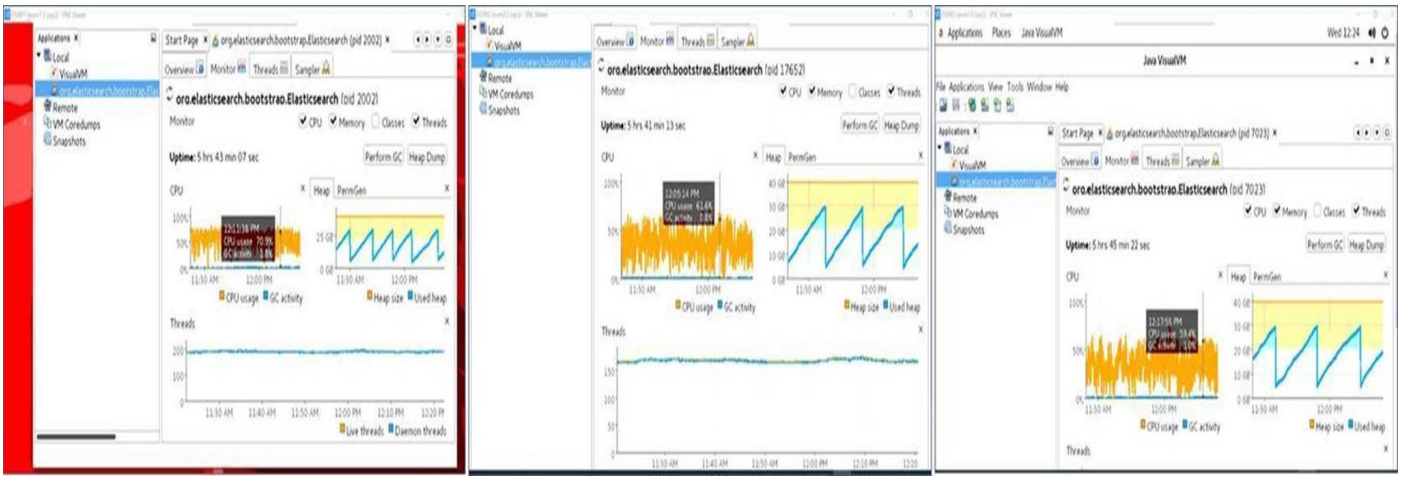
JVisualVM: - is a graphical user interface that provides detailed information about Java technology-based applications (Java applications) while they are running on a given Java Virtual Machine. This was configured on WebCenter Content to monitor UCM app server.

KIBANA Graphs During OTS to ES Migration on 3 ES Nodes w/ 5M Docs over 4 hrs.

GC and CPU do not have any issue during the migration process



JVisualVM Graphs During 5M Migration Process Over 4 hrs.



Shards Specific Index Space from Kibana for Migration Flow

GET /_cat/shards/es1_*

```
es1_development-authoring 4 r STARTED 254435 16.1gb VMIP1 node-2
es1_development-authoring 4 p STARTED 254435 16.2gb VMIP2 node-1
es1_development-authoring 0 p STARTED 253920 16.2gb VMIP3 node-3
es1_development-authoring 0 r STARTED 253920 16.2gb VMIP1 node-2
es1_development-review    3 p STARTED 246507 17.8gb VMIP3 node-3
es1_development-review    3 r STARTED 246507 17.8gb VMIP1 node-2
es1_development-review    4 r STARTED 246592 17.8gb VMIP1 node-2
es1_development-review    4 p STARTED 246592 17.8gb VMIP2 node-1
es1_development-review    0 p STARTED 246243 17.8gb VMIP3 node-3
es1_development-review    0 r STARTED 246243 17.8gb VMIP1 node-2
es1_development-published 2 p STARTED 256291 28.9gb VMIP1 node-2
es1_development-published 2 r STARTED 256291 28.9gb VMIP2 node-1
es1_development-published 1 r STARTED 255997 28.9gb VMIP3 node-3
es1_development-published 1 p STARTED 255997 28.9gb VMIP2 node-1
es1_development-published 3 p STARTED 256939   30gb VMIP3 node-3
es1_development-published 3 r STARTED 256939   29gb VMIP1 node-2
es1_development-published 4 r STARTED 256025 29.9gb VMIP3 node-3
es1_development-published 4 p STARTED 256025 28.9gb VMIP2 node-1
es1_development-published 0 p STARTED 256558   29gb VMIP3 node-3
es1_development-published 0 r STARTED 256558 28.9gb VMIP1 node-2
es1_development-approval  2 r STARTED 247120 10.7gb VMIP3 node-3
es1_development-approval  2 p STARTED 247120 10.8gb VMIP2 node-1
```

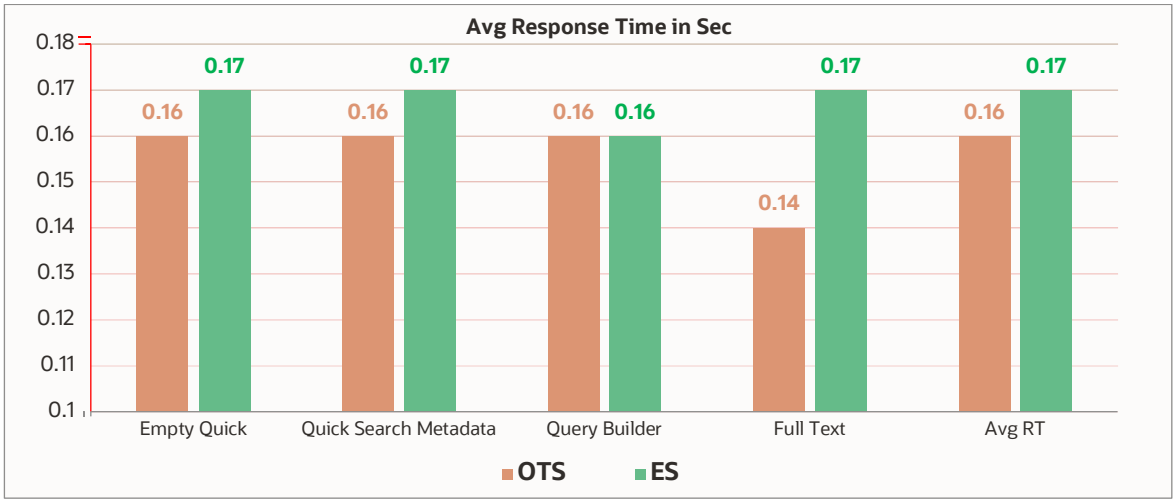
3. DOCUMENTS SEARCH OPERATION

3.1 Document Search Operation on 250K documents Base Benchmark (Single Node)

Results

WCC Single Node OTS vs ES w/ 250k Doc Baseline for Search Operation									
Flow	#Users	Avg RT		Throughput docs/Sec		CPU %			
		OTS	ES	OTS	ES	OTS	ES	OTS-DB	ES-DB
WCC Empty Quick Search	310	0.16	0.17	52	51	10	25	5	4
WCC Quick Search Metadata	300	0.16	0.17						
WCC QueryBuilder Search	80	0.16	0.16						
WCC Full Text Search	310	0.14	0.17						
Avg RT	**1000**	**0.16**	**0.17**						

Figure 1-5: WCC 1 node OTS vs ES w/ 250k doc Search Results



Observation

- The load test carried out with 1000 concurrent users performing the Search operation on both OTS and ES that contains 250k documents on a single node.
- The results observed on both the ES and OTS are almost similar.
- Max heap usage utilization is 14/40 GB on OTS whereas max heap usage on a single node ES is 29 GB/40 GB.

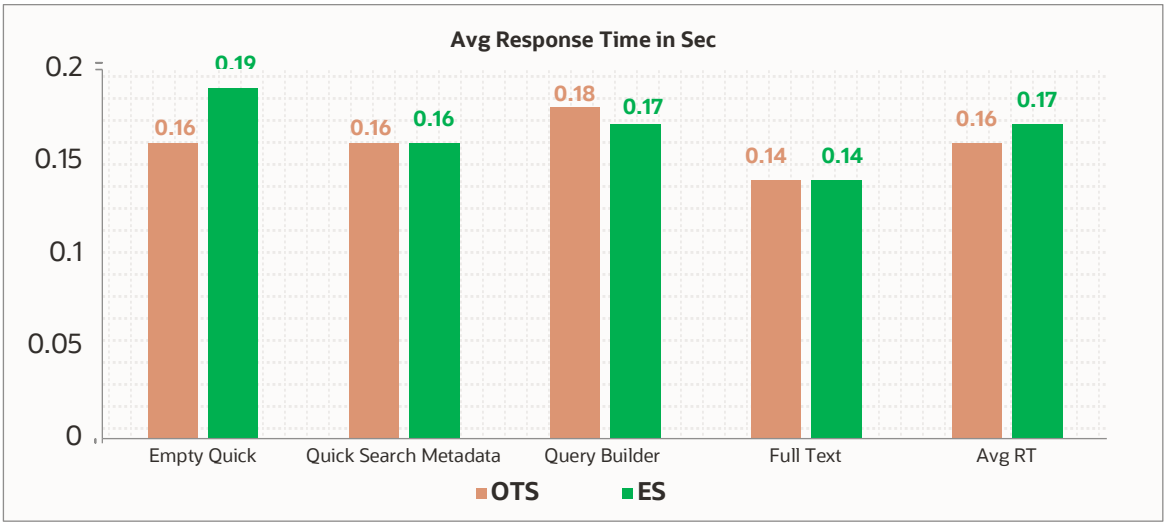
Hits/Sec Graph from OATS for 1000 Concurrent Users Combo Load:

3.2 Document Search Operation on 5 million documents Base Benchmark (Cluster)

Results

WCC 1 node OTS vs 3 Node ES Cluster w/ 5M docs Search Operation									
Flow	#Users	Avg RT in sec		Throughput docs/Sec		CPU %			
		OTS	ES	OTS	ES	OTS	ES	OTS-DB	ES-DB
WCC Empty Quick Search	310	0.16	0.19	52	52	16	26	6	5
WCC Quick Search Metadata	300	0.16	0.16						
WCC QueryBuilder Search	80	0.18	0.17						
WCC Full Text Search	310	0.14	0.14						
Avg RT	**1000**	**0.16**	**0.17**						

Figure 1-6: WCC 1 Node OTS vs ES w/ 5M Doc Search Results



Observation

- Search Performance of Elastic search and Oracle Text Search are almost similar.

PERFORMANCE TUNING

Test Settings and Configurations:

The performance tuning recommendations are common for all configurations. Shape specific recommendations are referenced explicitly throughout the section.

- **UCM JVM tunings:**
  - **UCM:** -Xmx2G –Xms40G
- **UCM Layer tuning (Config.cfg)**

  - SearchIndexerEngineName=OracleTextSearch(In Case of OTS)
  - SearchIndexerEngineName=ELASTICSEARCH (In Case of ES)
  - FsAutoConfigure=true
  - IDC_ID=IpmNode
  - WebServer=javaAppServer
  - FileEncoding=UTF8
  - NumConnections=50
  - MaxThreads=250
  - MaxRequestThreadCount=400
  - DisableTotalItemsSearchQuery=true
  - SkipAnalyzeServerConfigClassesDir=true
  - DisableReportErrorStack=true
  - FsCleanUpCacheDuringIndexing=1

- **Elastic Search Layer JVM tuning :**

  -Des.networkaddress.cache.ttl=60 -Des.networkaddress.cache.negative.ttl=10 -XX:+AlwaysPreTouch -Xss1m -Djava.awt.headless=true -Dfile.encoding=UTF-8 -Djna.nosys=true -XX:-OmitStackTraceInFastThrow -Dio.netty.noUnsafe=true -Dio.netty.noKeySetOptimization=true -Dio.netty.recycler.maxCapacityPerThread=0 -Dio.netty.allocator.numDirectArenas=0 -Dlog4j.shutdownHookEnabled=false -Dlog4j2.disable.jmx=true -Djava.locale.providers=COMPAT -Xms25g -Xmx25g -XX:+UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFraction=75 -XX:+UseCMSInitiatingOccupancyOnly -Djava.io.tmpdir=/tmp/elasticsearch-15003267407758163319 -XX:+HeapDumpOnOutOfMemoryError -XX:HeapDumpPath=data

- **Elasticsearch.yml configuration:**

  - cluster.name: ESGA
  - node.name: <node1, node2,node3>
  - http.port: 9200
  - xpack.security.enabled: true
  - xpack.security.transport.ssl.enabled: true
  - xpack.security.transport.ssl.verification_mode: certificate
  - xpack.security.transport.ssl.keystore.path: elastic-certificates.p12
  - xpack.security.http.ssl.enabled: true
  - xpack.security.http.ssl.verification_mode: certificate

- xpack.security.http.ssl.keystore.path: elastic-certificates.p12
- xpack.security.http.ssl.truststore.path: elastic-certificates.p12
- discovery.seed_hosts: ["node1 IP"," node2 IP "," node3 IP "]
- cluster.initial_master_nodes: ["node-1","node-2","node-3"]
- http.host: ["_lo:ipv4_", "_ens3:ipv4_"]
- transport.host: ["_lo:ipv4_", "_ens3:ipv4_"]
- transport.publish_port: 9300

- **WCC WLS Tunings:**

  - Production mode enabled
  - All log levels set to WARNING level
  - HTTP access log disabled.
  - Change initial and max capacity of data source connection pool to 100.

- OLT VU Pacing Think Time = 30 secs
  - Lower : 50%
  - Upper : 185%
  - Enabled auto indexing

CONCLUSION

This document consolidates the best practices and recommendations for sizing of WCC using OTS and ES 12.2.1.4.0 to ensure best performance and scalability.

GLOSSARY

WCC – Web Center Content
OTS – Oracle Text Search
ES – Elastic Search
OCI – Oracle Cloud Infrastructure
RT – Response Time
Throughput - Unit measured to show the number of documents processed per second

CONNECT WITH US

Call +1.800.ORACLE1 or visit oracle.com.
Outside North America, find your local office at oracle.com/contact.

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

ORACLE WEBCENTER CONTENT 12C Performance RESULTS

Performance TESTING ON Webcenter content WITH Oracle Text Search AND ELASTIC SEARCH
June 2121