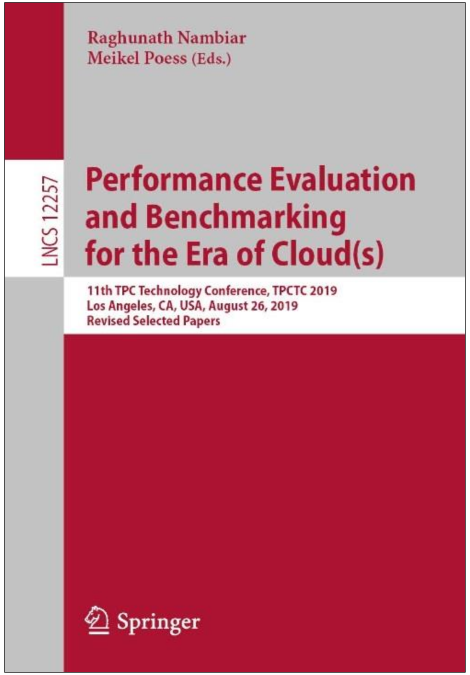
peakmarks

peakmarks® Performance Study on Exadata Storage Tiering for Online Transaction Processing

November 2019



Raghunath Nambiar
Meikel Poess (Eds.)

LNCS 12257

Performance Evaluation and Benchmarking for the Era of Cloud(s)

11th TPC Technology Conference, TPCTC 2019
Los Angeles, CA, USA, August 26, 2019
Revised Selected Papers

Springer

peakmarks® presented its software at the
11th Technology Conference of the Transaction
Processing Performance Council (TPC)
2019 in Los Angeles.

Agenda

1 Platform Description

2 peakmarks® Workload Overview

3 Workloads to determine Online Transaction Processing Performance

4 Performance Results of Online Transaction Processing

5 Conclusion and Recommendations

Copyright Information

Safe Harbor Statement

All performance data in this presentation were determined with the peakmarks® Software under certain conditions and do not necessarily correspond to the manufacturer's specifications. All information in this presentation is current as of November 2019.

Abbreviations and Metrics

[MBps]	megabyte per second
[GBps]	gigabyte per second
[dbps]	database blocks per second
[rbps]	redo blocks per second
[dbpt]	database blocks per transaction
[kBpt]	kilobyte per transaction
[s]	seconds
[ms]	milliseconds
[µs]	microseconds

[IOPS]	I/O operations per second
[qps]	queries per second
[rps]	rows per second
[tps]	transactions per second
[eps]	executions (SQL) per second
[Mops]	million operations per second
Nodes	number of cluster nodes
Jobs	number of workload processes
BuCache	Database Buffer Cache
FlCache	Database or Exadata Flash Cache

peakmarks

Performance is not everything.
But without performance, everything is worth nothing.

Platform Description

Platform

Server

	Oracle Exadata X5-2 Database Server	Oracle Exadata X5-2 2-node RAC Cluster
Launch date	2015	2015
Processor	Intel Xeon E5-2699 v3 (2.3 – 3.6 GHz)	Intel Xeon E5-2699 v3 (2.3 – 3.6 GHz)
#cpus, total	2	4
#cores, total	36	72
#threads, total	72	144
PCI Express	Gen 3	Gen 3
Memory type	DDR4	DDR4
DRAM capacity, total	768 GByte	1,536 GByte
DRAM capacity, per core	21 GByte	21 GByte
Operating System	Bare metal, OEL	Bare metal, OEL
Connectivity	InfiniBand, 2 x 40 Gbit/sec	InfiniBand 2 x 40 Gbit/sec per database server

Platform

Storage

	Oracle Exadata X5-2 Storage Server High Capacity	Oracle Exadata X5-2 Quarter Rack with 3 Storage Server HC
Launch date	2015	2015
DRAM capacity, total		
Flash capacity, total raw	6.4 TByte	19.2 TByte
Disk capacity, total raw	48 TByte	144 TByte
Connectivity	InfiniBand 2 x 40 Gbit/sec	InfiniBand 2 x 40 Gbit/sec per storage server
File system	ASM normal redundancy ASM allocation unit 4 MByte	ASM normal redundancy ASM allocation unit 4 MByte
Compression	No	No
Deduplication	No	No

Platform

Database

		Oracle Exadata X5-2 Quarter Rack with 3 Storage Server HC
Oracle version		19.3 Enterprise Edition
Database block size		8 kByte
Log Modus		NOARCHIVELOG
DataGuard		No
REDO Log Files, per instance		4 x 4 GByte, non-multiplexed
SGA size		384 GByte
peakmarks® Software		Version 9.4, Build 191130
peakmarks® Database size		1 , 2, 4, 8, 12, 16, 24 TByte

Notes:

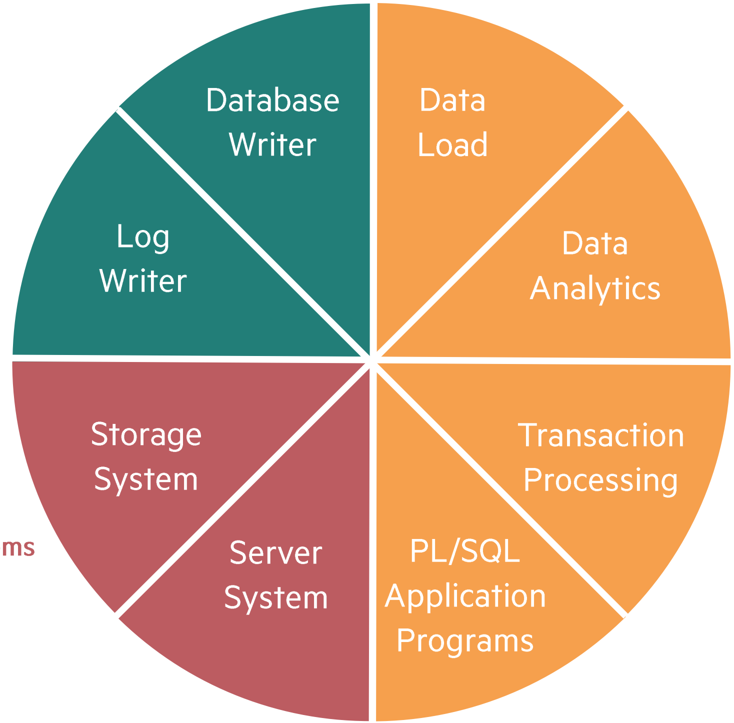
- To ensure full transparency, the peakmarks® Software generates individual Oracle AWR reports for each single performance test. In Oracle AWR reports, the idle wait event "enq: UL - contention" indicates process synchronization by the peakmarks® control process and does not cause wait states of workload execution processes.
- peakmarks® shows slightly better performance results than AWR because peakmarks® is the inner snapshot around tests while AWR is the outer snapshot for performance statistics.

peakmarks

Simple. Representative. Fast.

peakmarks® Workload Overview

More than 30 micro-benchmarks in 8 workload groups



Workload for Database Writer

- DBWR-THR

Workloads for Log Writer

- LGWR-LAT1
- LGWR-LAT25
- LGWR-LAT125
- LGWR-THR

Workloads for Storage Systems

- STO-READ
- STO-OFFLOAD
- STO-RANDOM
- STO-SCATTER

Workloads for Server Systems

- SRV-MIXED
- SRV-QUERY1
- SRV-QUERY25
- SRV-REPORT
- SRV-SCAN

Workloads for Data Load

- DL-BUFFER
- DL-DIRECT
- DL-STREAM

Workloads for Data Analytics

- DA-STORAGE
- DA-OFFLOAD
- DA-ROWSTORE
- DA-COLSTORE

Workloads for Transaction Processing

- TP-MIXED1
- TP-MIXED2
- TP-CUSTOM
- TP-REPORT
- TP-LOOKUP
- TP-LIGHT
- TP-MEDIUM
- TP-HEAVY

Workloads for PL/SQL Application Programs

- PLS-MIXED
- PLS-ADD
- PLS-BUILTIN
- PLS-FIBO
- PLS-PRIME

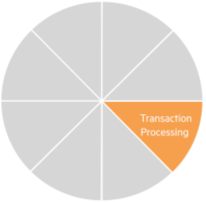
Simple and understandable Performance Metrics

Scope	Key Performance Metric	Measurement Unit	Workloads
Server Performance	▪ Query throughput ▪ Query response time ▪ Buffer cache scan rate	[qps] [ms] [MBps]	Look-up queries, more complex queries, reports, scans, mixed queries, and scans on cached tables in the Oracle buffer cache
Storage Performance	▪ SQL sequential read throughput ▪ SQL random I/O throughput ▪ SQL random I/O service time	[MBps] [iops] [µs]	Conventional storage, intelligent storage with offload technology
LGWR Performance	▪ Log writer throughput ▪ Log writer latency	[tps], [MBps] [ms]	Transactions with different REDO sizes
DBWR Performance	▪ Database writer throughput	[dbps]	
Data Load Performance	▪ Data load rate	[MBps] [rps]	Buffered data load (transactional systems), direct data load (data warehouse and analytic systems), streamed data load (IOT applications)
Data Analytics Performance	▪ Data scan rate	[MBps] [rps]	Conventional storage, intelligent storage with offload technology, row store, column store
Online Transaction Processing Performance	▪ Transaction throughput ▪ Transaction response time	[tps] [ms]	Transactions of different complexity; read-intensive transaction mix with data load, write-intensive transaction mix with heavy updates and data load
Processor Performance	▪ PL/SQL operation throughput ▪ PL/SQL algorithm processing time	[Mops] [s]	Arithmetic operations on different numeric data types, mixed built-in operations on different data types, recursive Fibonacci number algorithm, prime number algorithm

peakmarks

Swiss precision in performance measurement.

Workloads to determine the Online Transaction Processing Performance



Online Transaction Processing Workloads

Motivation

For capacity planning reasons, it is necessary to know the performance characteristics of a platform for transactions of varying complexity. Transaction Processing is the most complex database operation.

The goal is to

- Optimize the transaction throughput and transaction response time
- Validate the impact of several factors on transaction throughput and response time:
  - Ratio of database size and buffer cache size
  - transaction size
  - I/O random read service time
  - log writer latency
- Identify the limiting resource

Online Transaction Processing Workloads

Key Performance Metrics

- SQL transaction throughput in transactions per second [tps]
- SQL transaction response time in milliseconds [ms]

peakmarks® KPM Reports

- kpm_tp.sql
- kpm_tpplus.sql

Online Transaction Processing Workloads

Description

Workload	Measurement Unit	Action
TP-REPORT	[tps] [ms]	Online report of transaction processing application. SELECT Ø 25 rows via index.
TP-LOOKUP	[tps] [ms]	Fast lookup query. SELECT single row via index, e.g., SELECT an account, product. If configured, this workload uses tables in the memory-optimized row store introduced in 18c for fast look-up. Otherwise, it uses conventional tables.

Note

These transaction processing workloads are generic to all applications in all industries.

Online Transaction Processing Workloads

Description

Workload	Measurement Unit	Action
TP-LIGHT	[tps] [ms]	Light transaction type. SELECT/UPDATE single row via index, e.g., SELECT/UPDATE an account, product, or order with different SELECT/UPDATE ratios using SELECT FOR UPDATE locking. The workload parameter specifies the update ratio in %; the following values are supported {0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100}. This workload shows maximum transaction throughput and minimum transaction response time.
TP-MEDIUM	[tps] [ms]	Medium transaction type. SELECT/UPDATE Ø 25 rows via index, e.g., SELECT/UPDATE last month's bank account bookings with different SELECT/UPDATE ratios using SELECT FOR UPDATE locking. The workload parameter specifies the update ratio in %; the following values are supported {0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100}.
TP-HEAVY	[tps] [ms]	Heavy transaction type. SELECT/UPDATE Ø 125 rows via index, e.g., SELECT/UPDATE last month’s cell phone call records with different SELECT/UPDATE ratios using SELECT FOR UPDATE locking. The workload parameter specifies the update ratio in %; the following values are supported {0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100}.

Note
These transaction processing workloads are generic to all applications in all industries.

Online Transaction Processing Workloads

Description

Workload	Measurement Unit	Action
TP-MIXED1	[tps] [ms]	A read-intensive mix of different transaction types. Logical reads: 83% read, 17% write; avg 256-byte REDO per transaction. This workload is a complex workload that is composed of the equally weighted simple workloads TP-REPORT and TP-LOOKUP, TP-MEDIUM (with 40% UPDATE) and DL-BUFFER (with 2 rpt).
TP-MIXED2	[tps] [ms]	A write-intensive mix of different transaction types. Logical reads: 65% read, 35% write; avg 1,725-byte REDO per transaction. This workload is a complex workload that is composed of the equally weighted simple workloads TP-LIGHT (with 40% UPDATE), TP-MEDIUM (with 30% UPDATE), TP-HEAVY (with 20% UPDATE), and DL-BUFFER (with 3 rpt).

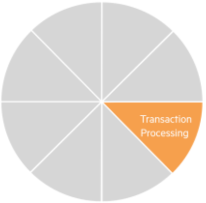
Notes

- TP-MIXED1 and TP-MIXED2 are the most representative peakmarks® workloads for determining Oracle online transaction processing performance capabilities on a specific platform.
- TP-MIXED1 achieves much higher transaction rates and CPU utilization than TP-MIXED2.
- These kinds of transaction processing workloads are generic to all industry applications.
- Peakmarks provides several performance reports for TP workloads: kpm_tp.sql (used in this presentation) shows overall transaction performance, and kpm_tpplus.sql provides more detailed information.

peakmarks

Stop guessing. Start measuring.

Online Transaction Processing Performance



Online Transaction Processing Performance

Workload TP-REPORT – online report, avg 25 rows per query

1 TByte database size

```
                                          CPU  CPU  CPU  CPU  CPU Transactions Response  IO time      REDO  LogFile BuCache FlCache Elapsed
                          Upd                busy user  sys idle  iow        total     time     read      data     sync    read    read    time
 Run Test Workload        [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]     [ms]    [kBpt]     [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- -------- -------- -------- ------- ------- -------
   6    2 TP-REPORT    N/A     1     1    2    1    1   98    0          371    2.697    0.307    0.871    0.594   52.34  100.00     302
        4 TP-REPORT    N/A     1     8    9    7    1   92    0        4,550    1.749    0.307    0.421    0.825   74.87  100.00     303
        6 TP-REPORT    N/A     1    16   17   14    2   83    0        9,208    1.723    0.291    0.440    0.661   73.43  100.00     304
        8 TP-REPORT    N/A     1    24   25   20    3   75    0       12,931    1.845    0.300    0.458    0.732   72.26  100.00     304
       10 TP-REPORT    N/A     1    32   33   27    4   67    0       16,963    1.873    0.299    0.470    0.668   71.29  100.00     305
       12 TP-REPORT    N/A     1    40   41   33    5   59    0       20,150    1.972    0.309    0.474    0.595   71.07  100.00     304
       14 TP-REPORT    N/A     1    48   49   40    6   51    0       22,905    2.078    0.318    0.482    0.733   70.57  100.00     304
```

2 TByte database size

```
                                          CPU  CPU  CPU  CPU  CPU Transactions Response  IO time      REDO  LogFile BuCache FlCache Elapsed
                          Upd                busy user  sys idle  iow        total     time     read      data     sync    read    read    time
 Run Test Workload        [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]     [ms]    [kBpt]     [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- -------- -------- -------- ------- ------- -------
   8    2 TP-REPORT    N/A     1     1    2    1    0   98    0        3,267    0.306    0.314    0.055    0.442   96.94   99.99     300
        4 TP-REPORT    N/A     1     8   10    8    1   90    0        2,663    2.989    0.299    0.913    1.105   42.71  100.00     303
        6 TP-REPORT    N/A     1    16   19   15    2   81    0        5,078    3.121    0.308    0.917    0.664   41.44  100.00     305
        8 TP-REPORT    N/A     1    24   27   22    3   73    0        7,329    3.247    0.315    0.944    0.682   39.64  100.00     305
       10 TP-REPORT    N/A     1    32   36   29    5   64    0        9,419    3.380    0.329    0.941    0.778   39.79  100.00     304
       12 TP-REPORT    N/A     1    40   44   36    6   56    0       11,027    3.606    0.346    0.952    0.687   39.10  100.00     304
       14 TP-REPORT    N/A     1    48   52   42    7   48    0       12,572    3.794    0.361    0.943    0.689   39.58  100.00     304
```

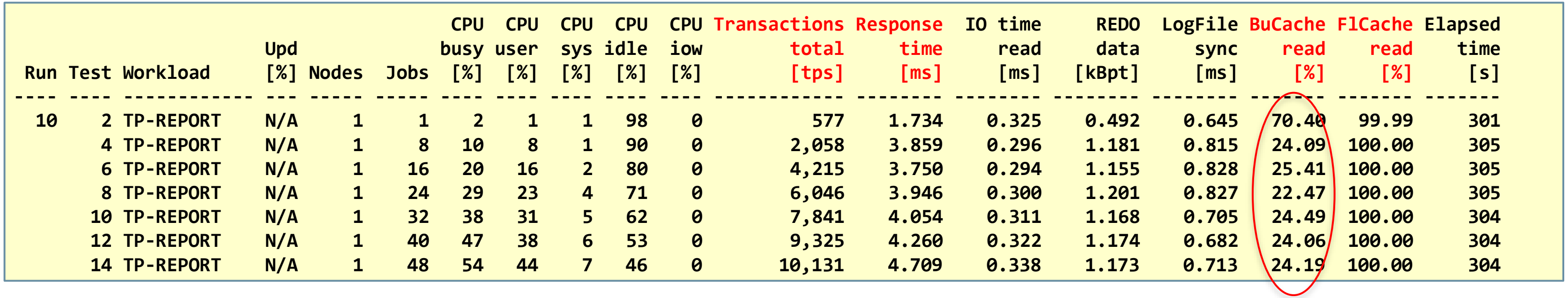
Notes:

- As expected, the buffer cache hit rate decreases with increasing database size.
- Flash cache hit rate 100%, I/O read service time still at flash level.

Online Transaction Processing Performance

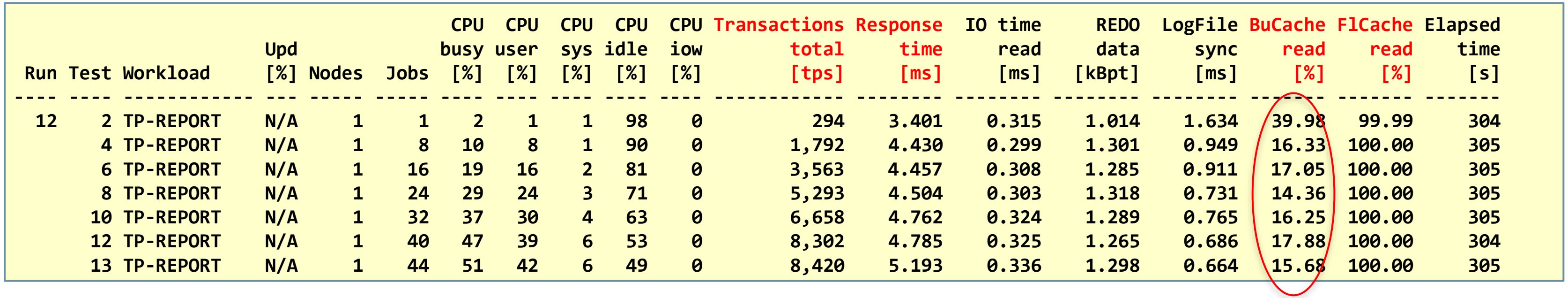
Workload TP-REPORT – online report, avg 25 rows per query

4 TByte database size



```
                                          CPU  CPU  CPU  CPU  CPU Transactions Response   IO time      REDO   LogFile BuCache FlCache Elapsed
                       Upd               busy user  sys idle  iow        total     time      read      data      sync    read    read    time
Run  Test Workload     [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]      [ms]    [kBpt]      [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- --------- --------- --------- ------- ------- -------
  10    2 TP-REPORT    N/A     1     1    2    1    1   98    0          577    1.734     0.325     0.492     0.645   70.40   99.99     301
        4 TP-REPORT    N/A     1     8   10    8    1   90    0        2,058    3.859     0.296     1.181     0.815   24.09  100.00     305
        6 TP-REPORT    N/A     1    16   20   16    2   80    0        4,215    3.750     0.294     1.155     0.828   25.41  100.00     305
        8 TP-REPORT    N/A     1    24   29   23    4   71    0        6,046    3.946     0.300     1.201     0.827   22.47  100.00     305
       10 TP-REPORT    N/A     1    32   38   31    5   62    0        7,841    4.054     0.311     1.168     0.705   24.49  100.00     304
       12 TP-REPORT    N/A     1    40   47   38    6   53    0        9,325    4.260     0.322     1.174     0.682   24.06  100.00     304
       14 TP-REPORT    N/A     1    48   54   44    7   46    0       10,131    4.709     0.338     1.173     0.713   24.19  100.00     304
```

8 TByte database size



```
                                          CPU  CPU  CPU  CPU  CPU Transactions Response   IO time      REDO   LogFile BuCache FlCache Elapsed
                       Upd               busy user  sys idle  iow        total     time      read      data      sync    read    read    time
Run  Test Workload     [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]      [ms]    [kBpt]      [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- --------- --------- --------- ------- ------- -------
  12    2 TP-REPORT    N/A     1     1    2    1    1   98    0          294    3.401     0.315     1.014     1.634   39.98   99.99     304
        4 TP-REPORT    N/A     1     8   10    8    1   90    0        1,792    4.430     0.299     1.301     0.949   16.33  100.00     305
        6 TP-REPORT    N/A     1    16   19   16    2   81    0        3,563    4.457     0.308     1.285     0.911   17.05  100.00     305
        8 TP-REPORT    N/A     1    24   29   24    3   71    0        5,293    4.504     0.303     1.318     0.731   14.36  100.00     305
       10 TP-REPORT    N/A     1    32   37   30    4   63    0        6,658    4.762     0.324     1.289     0.765   16.25  100.00     305
       12 TP-REPORT    N/A     1    40   47   39    6   53    0        8,302    4.785     0.325     1.265     0.686   17.88  100.00     304
       13 TP-REPORT    N/A     1    44   51   42    6   49    0        8,420    5.193     0.336     1.298     0.664   15.68  100.00     305
```

Notes:

- As expected, buffer cache hit rate decreases with increasing database size.
- Flash cache hit rate 100%, I/O read service time still at flash level.

Online Transaction Processing Performance

Workload TP-REPORT – online report, avg 25 rows per query

12 TByte database size

```
                                        CPU  CPU  CPU  CPU  CPU Transactions Response  IO time     REDO  LogFile BuCache FlCache Elapsed
                       Upd             busy user  sys idle  iow        total     time     read     data     sync    read    read    time
 Run Test Workload     [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]     [ms]   [kBpt]     [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- -------- -------- -------- ------- ------- -------
  14    2 TP-REPORT    N/A     1     1    2    1    0   98    0          235    4.257    0.461    0.777    0.715   54.90   99.85     315
        4 TP-REPORT    N/A     1     8    8    7    1   92    0        1,268    6.279    0.400    1.354    0.710   12.92   99.98     304
        6 TP-REPORT    N/A     1    16   15   12    2   85    0        2,485    6.384    0.406    1.353    0.655   13.04   99.98     305
        8 TP-REPORT    N/A     1    24   23   19    3   77    0        3,872    6.138    0.396    1.331    0.950   13.70   99.98     306
       10 TP-REPORT    N/A     1    32   32   26    4   68    0        5,277    6.013    0.385    1.343    0.688   12.72  100.00     305
       12 TP-REPORT    N/A     1    40   41   34    5   59    0        6,688    5.924    0.376    1.332    0.695   13.41  100.00     305
       14 TP-REPORT    N/A     1    48   49   41    6   51    0        7,562    6.304    0.384    1.322    0.683   14.28  100.00     305
```

16 TByte database size

```
                                        CPU  CPU  CPU  CPU  CPU Transactions Response  IO time     REDO  LogFile BuCache FlCache Elapsed
                       Upd             busy user  sys idle  iow        total     time     read     data     sync    read    read    time
 Run Test Workload     [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]     [ms]   [kBpt]     [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- -------- -------- -------- ------- ------- -------
  16    2 TP-REPORT    N/A     1     1    2    1    0   98    0           79   12.647    0.668    1.582    0.467   24.94   99.72     301
        4 TP-REPORT    N/A     1     8    8    6    1   92    0        1,209    6.563    0.386    1.396    0.639   12.67   99.92     305
        6 TP-REPORT    N/A     1    16   14   11    2   86    0        2,240    7.097    0.398    1.371    0.686   11.29   99.89     304
        8 TP-REPORT    N/A     1    24   23   19    3   77    0        3,855    6.173    0.369    1.348    0.810   12.57   99.95     306
       10 TP-REPORT    N/A     1    32   32   26    4   68    0        5,335    5.946    0.366    1.352    0.701   12.11   99.95     305
       12 TP-REPORT    N/A     1    40   42   35    5   58    0        6,880    5.765    0.365    1.362    0.698   11.62   99.99     305
       14 TP-REPORT    N/A     1    48   50   41    6   50    0        7,720    6.178    0.374    1.358    0.746   11.68   99.99     305
```

Note

With 16 TByte database size flash cache hit rate no longer achieves 100%.

Online Transaction Processing Performance

Workload TP-REPORT – online report, avg 25 rows per query

24 TByte database size

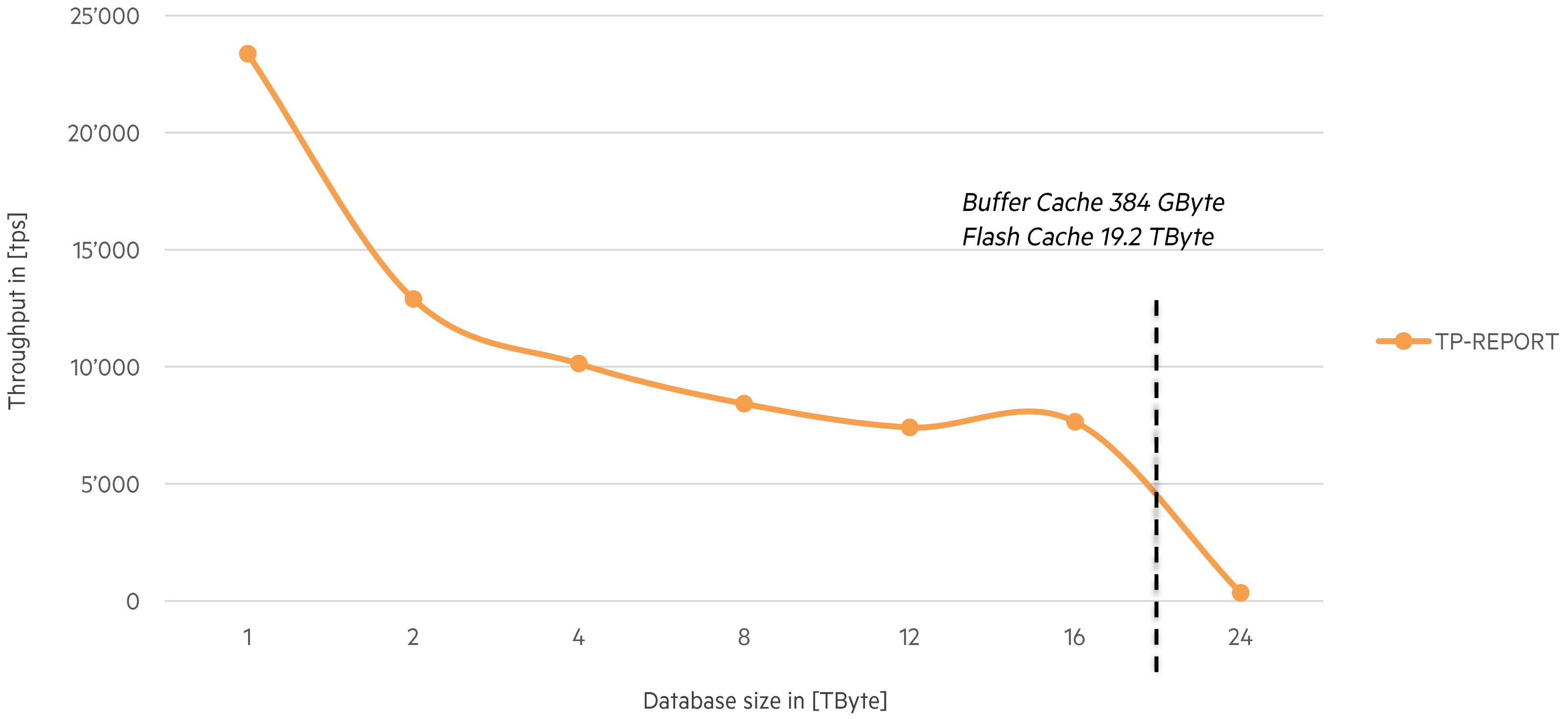
```
                                              CPU  CPU  CPU  CPU  CPU Transactions Response  IO time     REDO  LogFile BuCache FlCache Elapsed
                         Upd                 busy user  sys idle  iow        total     time     read     data     sync    read    read    time
 Run Test Workload       [%] Nodes  Jobs  [%]  [%]  [%]  [%]  [%]        [tps]     [ms]     [ms]   [kBpt]     [ms]     [%]     [%]     [s]
---- ---- ------------ --- ----- ----- ---- ---- ---- ---- ---- ------------ -------- -------- -------- -------- ------- ------- -------
  18    2 TP-REPORT    N/A     1     1    1    1    0   99    0           42   23.693    1.185    1.527    0.685   23.09   87.85     300
        3 TP-REPORT    N/A     1     4    2    1    0   98    0          172   23.024    1.447    1.519    0.489   15.30   89.54     305
        4 TP-REPORT    N/A     1     8    2    1    0   98    0          172   41.141    3.043    1.322    0.922   19.10   93.75     466
        5 TP-REPORT    N/A     1    12    3    2    1   97    0          347   31.526    1.972    1.350    0.816   17.18   91.65     353
        6 TP-REPORT    N/A     1    16    2    2    1   98    0          248   41.182    2.788    1.330    0.680   17.29   90.80     653
```

Notes:

- Database (24 TByte) no longer fits into flash cache (19.2 TByte).
- For some tests, flash cache hit rate falls below 90%.
- Transaction throughput drops by factor, and response time increases by factor.

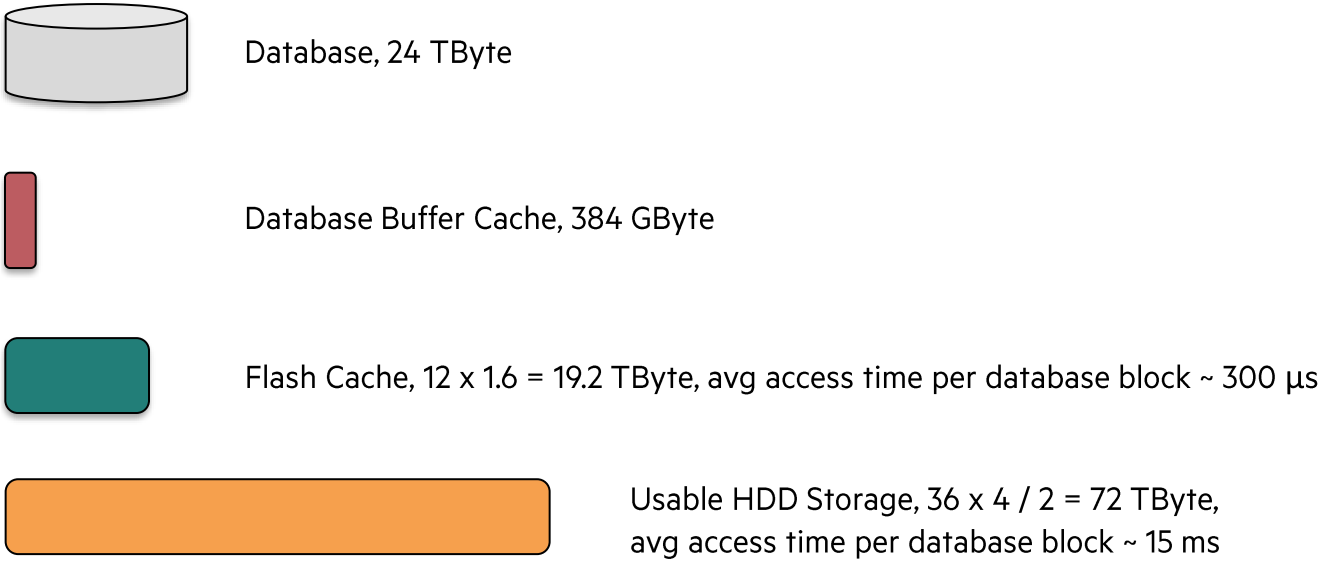
Online Transaction Processing Performance

Impact of ratio buffer cache size / database size – Impact of storage tiering



Online Transaction Processing Performance

Exadata X5-2 QRHC Storage Tiering for 24 TByte database



Database, 24 TByte

Database Buffer Cache, 384 GByte

Flash Cache, 12 x 1.6 = 19.2 TByte, avg access time per database block ~ 300 µs

Usable HDD Storage, 36 x 4 / 2 = 72 TByte,
avg access time per database block ~ 15 ms

Conclusion & Recommendation

The transaction processing performance is good as long as the active data fits into the first tier.

If the amount of active data exceeds the capacity of the flash cache, performance drops sharply.

When planning the capacity of Exadata Storage Servers, it is essential to ensure that all active data fits into the flash cache.

peakmarks Mission

Identify Key Performance Metrics for Oracle Database Platforms.

On-Premises and in the Cloud.

For Quality Assurance, Evaluations, and Capacity Planning.