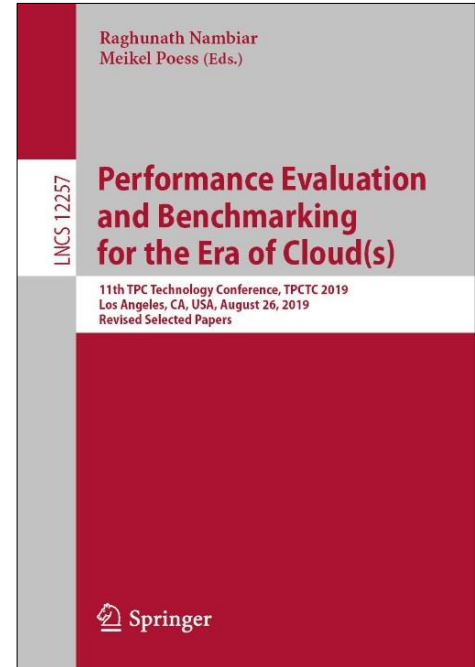


# peakmarks® Performance Study on Exadata Storage Tiering

November 2019



peakmarks® showcased its software at the  
11<sup>th</sup> TPC Technology Conference 2019.



All performance data were determined with the peakmarks® software under certain conditions and do not necessarily correspond to the manufacturer's specifications.



[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

[ $\mu$ s] microseconds

[IOPS] I/O operations per second

[qps] queries per second

[rps] rows per second

[tps] transactions 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



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

## Platform description



## Server

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



## Storage

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



## Database

		Oracle Exadata X5-2 Quarter Rack High Capacity, single instance
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

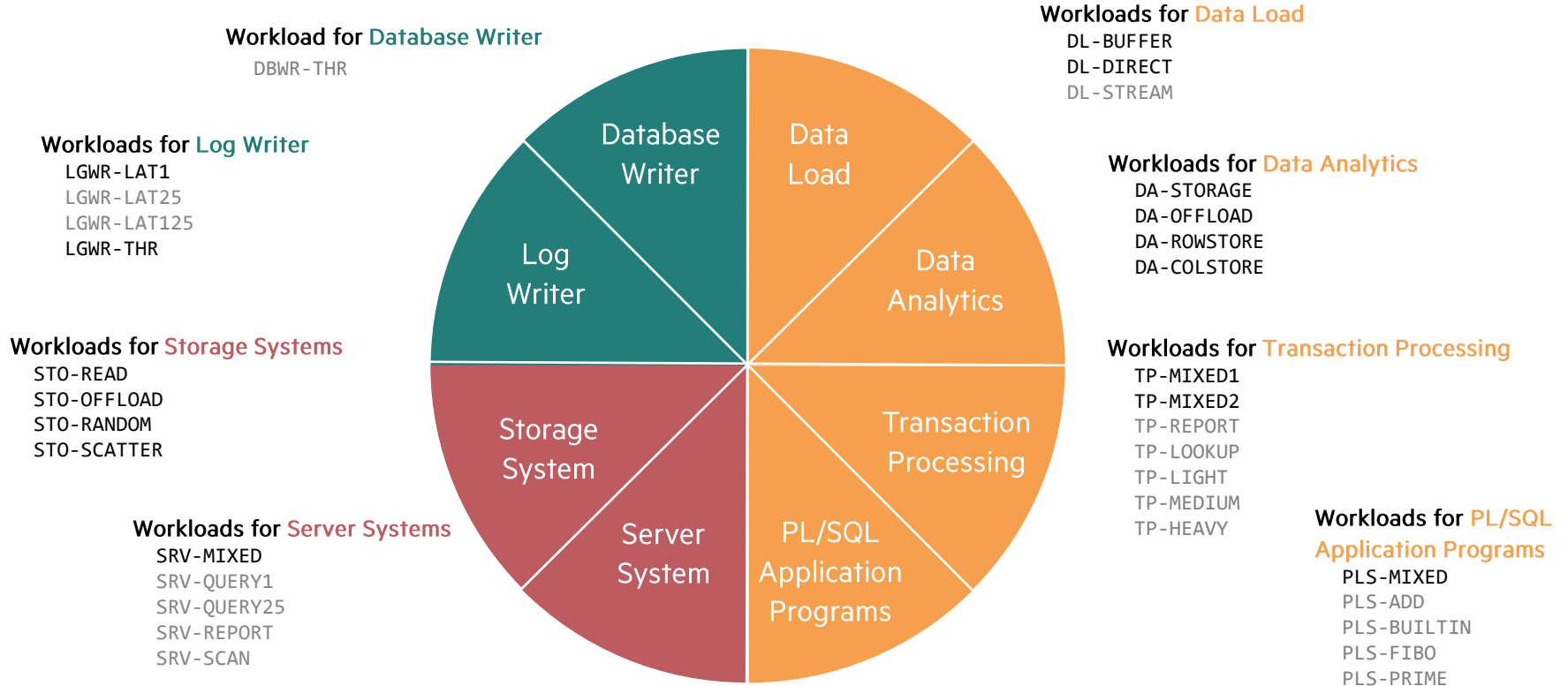




Simple. Representative. Fast.

## peakmarks<sup>®</sup> Workload Overview

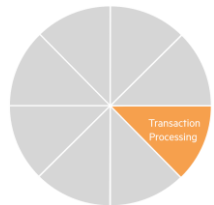
More than 30 micro-benchmarks in 8 workload groups





Swiss precision in performance measurement.

## Workloads to determine the Transaction Processing Performance





## 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:
  - » database size
  - » buffer cache size
  - » transaction size
  - » I/O random read service time
  - » log writer latency



## Key Performance Metrics

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



## 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. <i>If configured, this workload uses tables in the memory-optimized row store for fast look-up introduced in 18c. Otherwise, this workload uses conventional tables.</i>

### Note

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





## Description

Workload	Measurement Unit	Action
TP-LIGHT	[tps]	Light transaction type.
	[ms]	SELECT/UPDATE <b>single row</b> 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]	Medium transaction type.
	[ms]	SELECT/UPDATE <b>Ø 25 rows</b> 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]	Heavy transaction type.
	[ms]	SELECT/UPDATE <b>Ø 125 rows</b> 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.



## Description

Workload	Measurement Unit	Action
TP-MIXED1	[tps] [ms]	<p>A <b>read-intensive mix</b> of different transaction types.</p> <p>Logical reads: 83% read, 17% write; avg 256-byte REDO per transaction.</p> <p>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).</p>
TP-MIXED2	[tps] [ms]	<p>A <b>write-intensive mix</b> of different transaction types.</p> <p>Logical reads: 65% read, 35% write; avg 1,725-byte REDO per transaction.</p> <p>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).</p>
TP-CUSTOM	[tps] [ms]	<p>A <b>configurable and customized</b> mix of different transaction types.</p> <p>Default configuration compatible to peakmarks Version 9 TP-MIXED.</p>

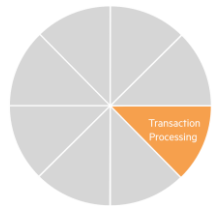
### Notes

- TP-MIXED1 and TP-MIXED2 are the most representative peakmarks workloads for determining Oracle transaction processing performance capabilities on a specific platform.
- TP-MIXED1 achieves much higher transaction rates and CPU utilization than TP-MIXED2.
- This kind of transaction processing workload is generic to all industry applications.
- Peakmarks provides several performance reports for TP workloads: kpm\_tp.sql (used in this presentation) shows overall transaction performance, kpm\_tpplus.sql provides more detailed information, and kpm\_tpio.sql provides I/O information.



Stop guessing. Start measuring.

## Transaction Processing Performance





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

1 TByte database size

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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, buffer cache hit rate decreases with increasing database size.
- Flash cache hit rate 100%, I/O read service time still at flash level.



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

4. TByte database size

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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.



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

12 TByte database size

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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%.



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

24-TByte database size

Run	Test	Workload	Upd [%]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	IO time read [ms]	REDO data [kBpt]	LogFile sync [ms]	BuCache read [%]	FlCache read [%]	Elapsed time [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

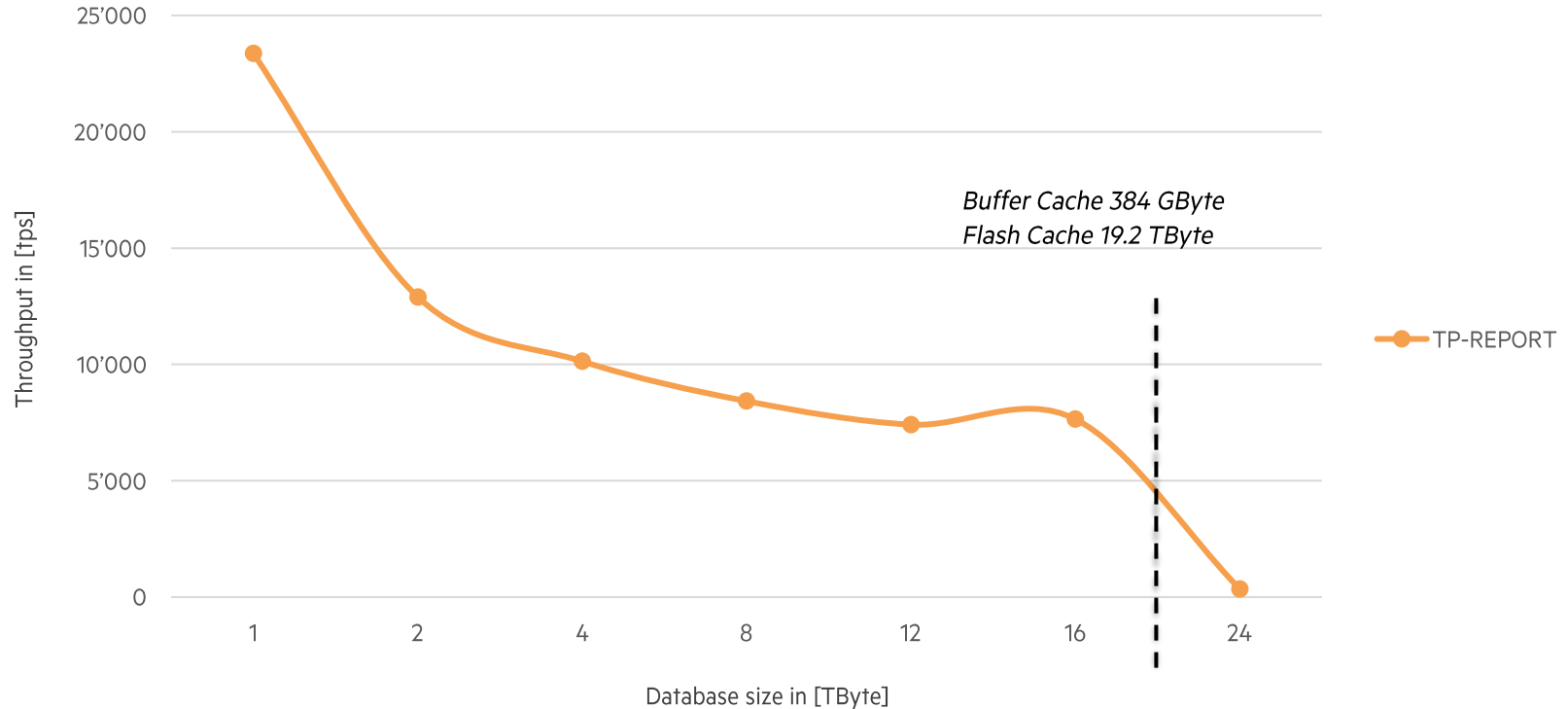
### Notes:

- 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.





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





## Exadata X5-2 QRHC Storage Tiering for 24 TByte database



Database, 24 TByte



Database Buffer Cache, 384 GByte



Flash Cache,  $12 \times 1.6 = 19.2$  TByte, avg access time per database block  $\sim 300 \mu\text{s}$



Usable HDD Storage,  $36 \times 4 / 2 = 72$  TByte,  
avg access time per database block  $\sim 15 \text{ ms}$



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.