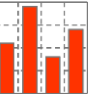


Introduction to Storage (STO) Performance Tests with Oracle

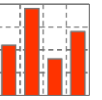
Technical Presentation

September 2014



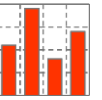
- 1 Why measure STO performance?
- 2 What is measured?
- 3 Overview of STO performance tests
- 4 Monitoring STO performance tests
- 5 Example
- 6 Summary

Why measure Storage performance?



- Storage performance is essential not only for overall Oracle database performance, but also for system management tasks like backup, recovery and archiving
- Oracle uses all kinds of I/O patterns, but different o/s calls dependent upon
 - operating system
 - system load (Oracle changes system call dependent on load)

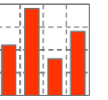
Why measure Storage performance?



Oracle I/O access pattern

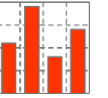
- Oracle sequential read
 - User processes: full table scan, full index scan
 - Temp segment
 - Backup, restore, recovery RMAN, Export, Data Pump
 - ARCH: reading online REDO logfile
- Oracle random read
 - User processes
- Oracle sequential write
 - Temp segment
 - Backup, restore RMAN, Export, Data Pump
 - LWGR process: small block size
 - ARCH processes: writing archived REDO logfile
 - RVWR process: flashback log file writer
 - CTWR process: block change tracking file
- Oracle random write
 - DBWR processes

What is measured?



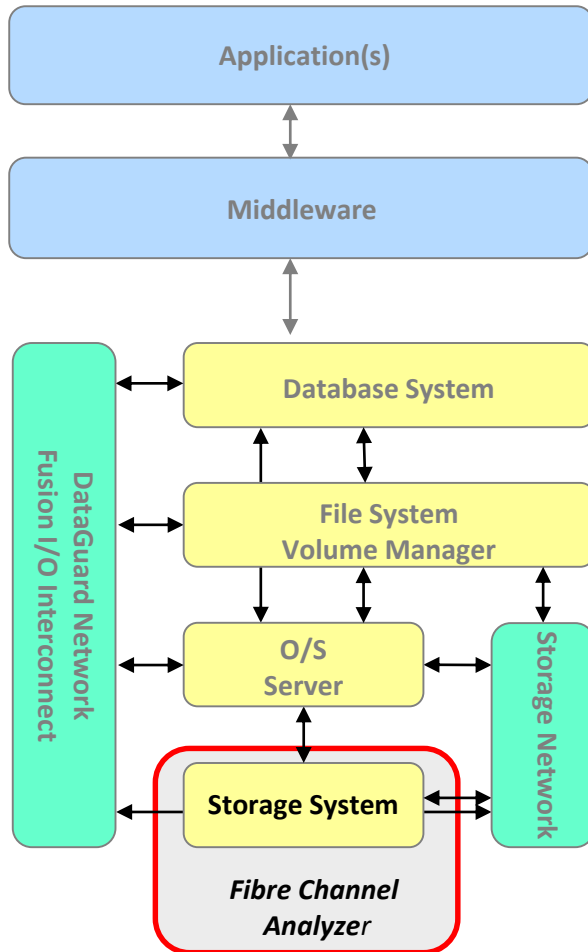
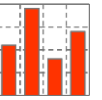
- Throughput of system
 - Data transfer in mega byte per second [MBps]
 - Data transfer in database blocks per second [dbps]
 - I/O Operations in O/S system calls per second [IOPS]
- Service Time
 - For I/O operations in [ms] or [μ s]

What is measured?



- Efficiency of
 - Auto-Tiering
 - RAID-level
 - Striping
 - Remote mirroring
 - Virtualization
- Storage performance from the Oracle point of view
 - Using Oracle database as load generator for the storage system
 - I/O throughput and response time measured within Oracle

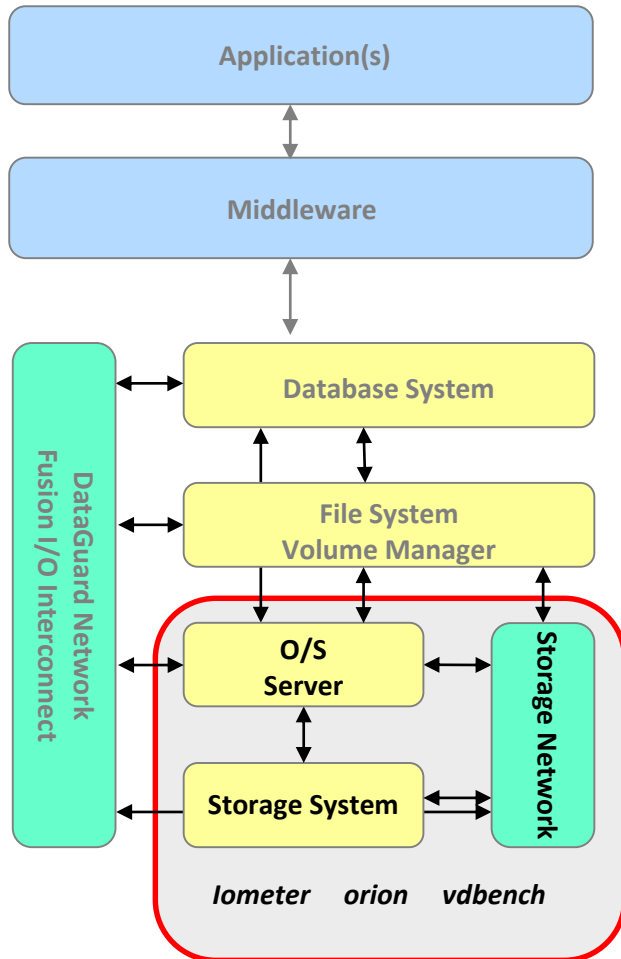
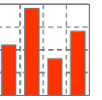
Comparison of I/O performance tests



■ Storage System I/O Performance

- Useful to test storage system performance at port level
- Vendors data sheet numbers

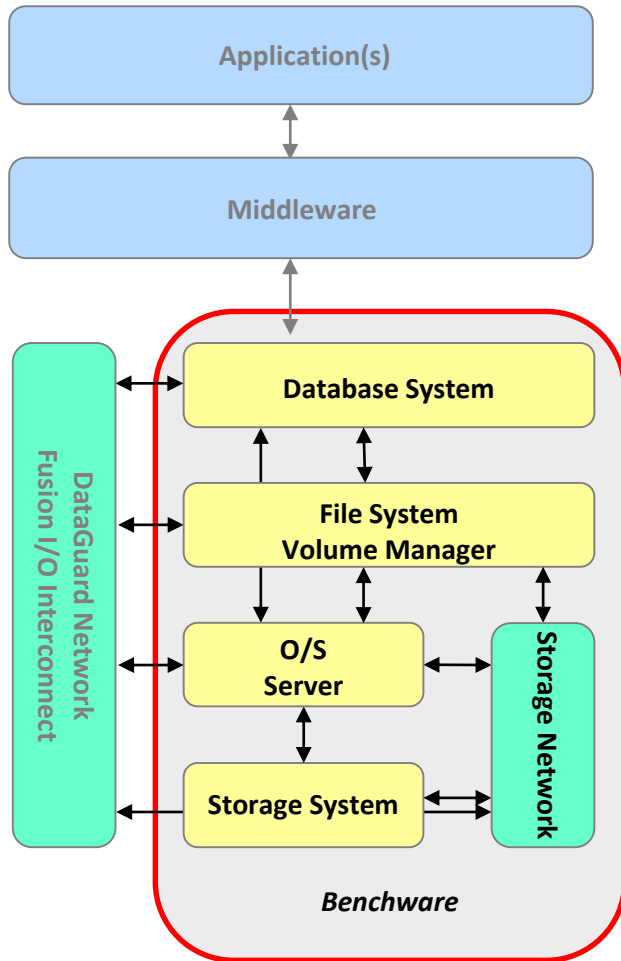
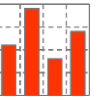
Comparison of I/O performance tests



■ Server System I/O Performance

- Tools like vdbench, lometer, Orion, etc. just generate I/O system calls, but no further I/O processing
- Useful to analyze transfer performance between storage system and server system
- Unable to test storage grids, e.g. Exadata Storage Cell
- Difficult to test Oracle ASM infrastructure

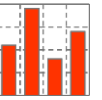
Comparison of I/O performance tests



■ Database System I/O Performance

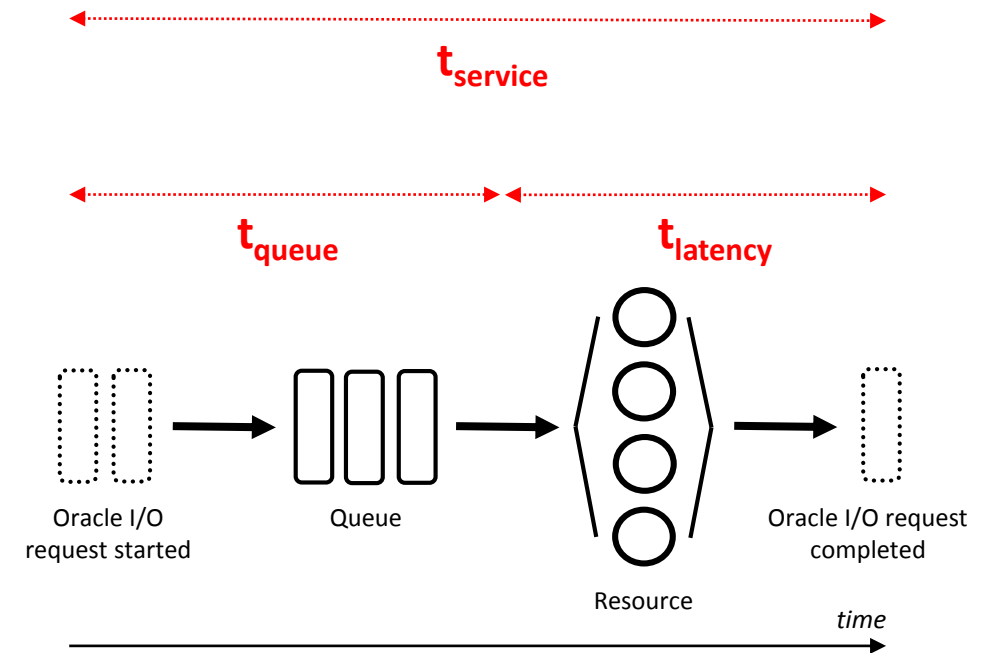
- Most complex I/O operation
- Database buffer cache management
 - » find a free slot
 - » replace older blocks
 - » synchronize access to buffer cache
 - » database block consistency checks
- Database I/O needs much more cpu resources than simple I/O generator
 - » Rule of thumb: 15'000 - 25'000 IOPS per x86 core
 - » Throughput does not scale linear
- `dbms_resource_manager.calibrate_io` does not recognize hybrid storage systems and delivers wrong results

I/O Response Time

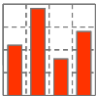


- Oracle wait events measure complete I/O service time including
 - Queue or wait time t_{queue}
 - » I/O scheduler queue
 - » Device driver queue
 - » Host bus adapter queue
 - Latency of storage t_{latency}
 - Transit time through all software layers of I/O stack
 - Non default queue configuration is essential when using flash technology

- Oracle provides histogram of service times



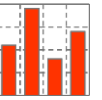
Storage hierarchy



Capacity and access times (order of magnitude is important)

Server System	Capacity	Average Access Time	Throughput
CPU register	1 KByte	< 1 ns	
Level 1 Cache on-chip	128 KByte	1 ns	
Level 2 Cache	8 MByte	5 ns	
Level 3 Cache	32 MByte	15 ns	
Main memory	4 TByte	100 ns	Seq/Ran: > 10 GBps
PCI attached flash	2.4 TByte	< 100'000 ns	Seq: > 1 GBps Ran: > 500'000 IOPS
Storage System			
Cache	1 TByte	500'000 ns	
Solid state disks (SSD)	1.4 TByte	1'000'000 ns	Seq: > 1 GBps Ran: > 200'000 IOPS
Hard drive disks (HDD)	600 GByte, 15k rpm	8'000'000 ns	Ran: 250 IOPS
Hard drive disks (HDD)	3.0 TByte, 7.2k rpm	15'000'000 ns	Ran: 120 IOPS
Tape drive	2.5 TByte pro LTO Cartridge	50'000'000'000 ns	Seq: 160 MBps

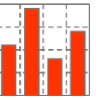
Overview of STO performance tests with Benchware test codes



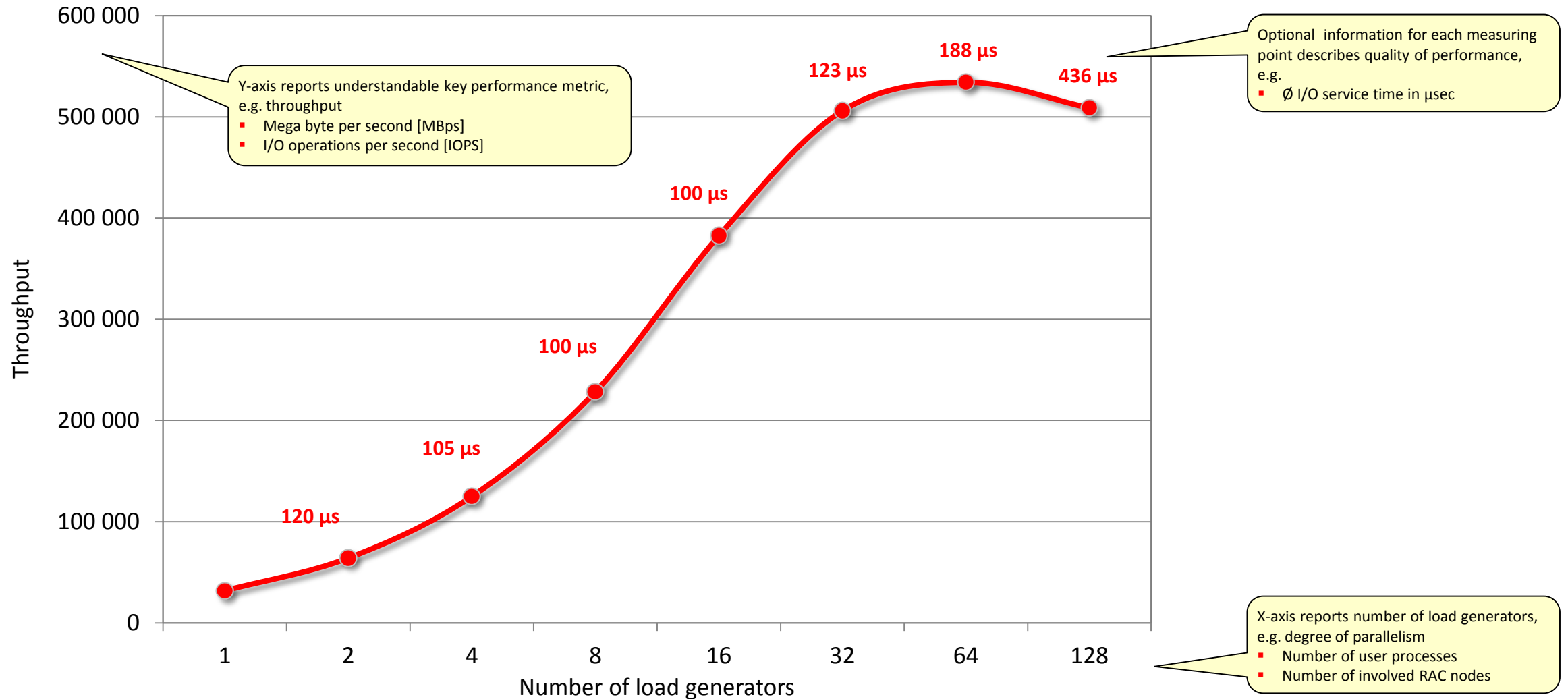
Oracle Storage Performance Sequential I/O	Test Code using DEFAULT Storage	Test Code using Database Flash Cache	Test Code using Cell Flash Cache
▪ Sequential read	STO-12	STO-13	STO-14
▪ Sequential write	STO-22	-	-

Oracle Storage Performance Random I/O	Test Code for using DEFAULT Storage	Test Code for using Database Flash Cache	Test Code for using Cell Flash Cache
▪ Random read, 25 I/O's per 125'000 rows per SQL	STO-62	STO-63	STO-64
▪ Random write (DBWR stress test)	STO-42	-	-
▪ Mixed random read write (50:50)	STO-52	STO-53	STO-54

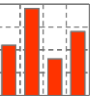
Monitoring STO performance tests



All load profiles from single process to saturation



Monitoring STO performance tests



■ Key Performance Metrics

- I/O throughput in [MBps]
- I/O throughput in [IOPS]
- I/O service time in [s]

Run	Tst Code	#N	#J	#T	CPU busy [%]	CPU sys [%]	Physical read [iops]	Physical read [bps]	Physical read [MBps]	Physical write [iops]	Physical write [bps]	Physical write [MBps]	REDO write [iops]	Hitrate db flash [%]	Hitrate exa flash [%]	Elap time [s]
15	10 STO-62	1	1	1	3	2	31747	31745	248	11	23	0	2	0	0	51
	11 STO-62	1	2	1	6	4	64039	64037	500	59	68	1	5	0	0	51
	12 STO-62	1	4	1	12	7	124897	124895	976	164	168	1	10	0	0	52
	13 STO-62	1	8	1	23	13	228361	228359	1784	356	351	3	19	0	0	57
	14 STO-62	1	16	1	45	27	382520	382518	2988	629	607	5	31	0	0	68
	15 STO-62	1	32	1	77	45	506050	506047	3954	854	812	6	41	0	0	103
	16 STO-62	1	64	1	95	51	534207	534206	4174	879	832	7	41	0	0	196
	17 STO-62	1	128	1	98	51	508957	508962	3976	810	764	6	40	0	0	305

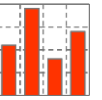
Legend:

Run	benchmark run id	#N	number of RAC nodes	[rps]	rows per second	[iops]	i/o operations per second	[s]	time in seconds
Tst	benchmark test id	#J	number of load generators (jobs)	[tps]	transactions per second	[dbps]	database blocks per second	[ms]	time in milli seconds
Code	benchmark test code	#T	number of threads (PX)	[ops]	operations per second	[MBps]	mega byte per second	[μs]	time in micro seconds

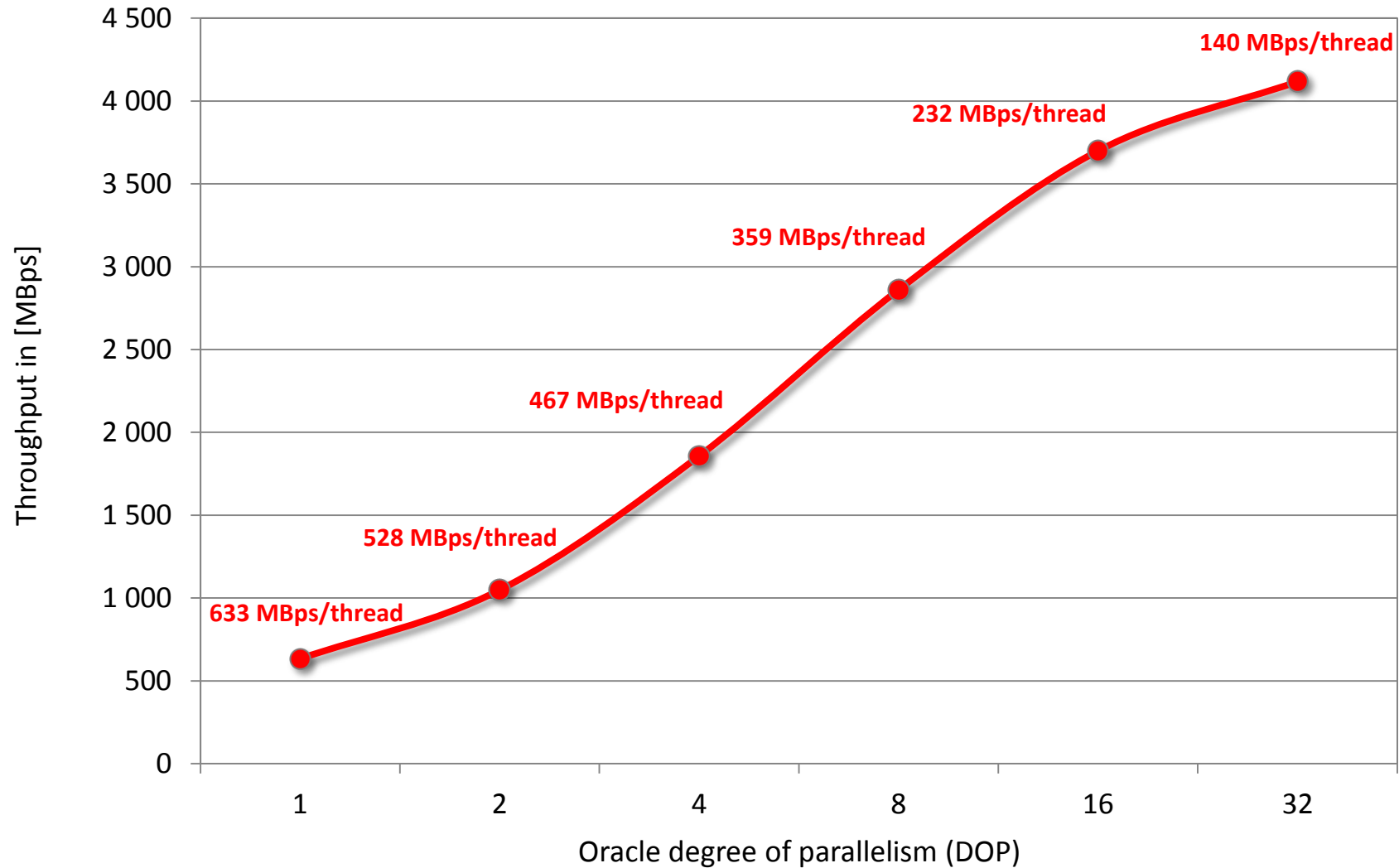
Max throughput:

- In this case CPU saturated system

Example of STO performance test results



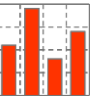
Oracle storage performance: sequential read, single process



This test is to determine the optimal degree of parallelism (DOP) for a single Oracle process

The overhead of PQ increases with increasing DOP

Summary



Reviewing Storage Performance

- Benchware proves real world storage performance for Oracle databases
- Storage vendors data sheets do not reflect database I/O performance
- Database I/O is much more complex than native storage system I/O
 - Buffer cache management
 - Concurrency
 - Block consistency

BENCHWARE

swiss precision in performance measurement

www.benchmarkware.ch

info@benchmarkware.ch