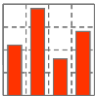


Performance Baseline of Hitachi Data Systems UCP for Oracle

Part III: Storage Performance with PCI attached Flash

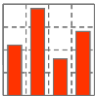
Benchmark Performance Suite Release 8.4 (Build 130830)

August 2013



- 1 Introduction to Storage Performance Tests**
- 2 Storage Configuration
- 3 Benchmark Results – Sequential I/O
- 4 Benchmark Results – Random I/O
- 5 Reviewing Storage Benchmark Results

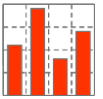
Storage Performance



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 I/O pattern, but different o/s calls dependent upon the
 - operating system
 - system load (Oracle changes system call dependent on load)

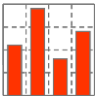
Storage Performance



Why measure Storage performance?

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

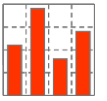
Storage Performance



What is measured?

- Throughput
 - Data transfer in mega byte per second [MBps]
 - Data transfer in blocks per second [bps]
 - I/O Operations in O/S system calls [IOPS]
- Service Time
 - For random I/O operation in [ms] or [μ s]
- Efficiency of
 - Auto-Tiering
 - RAID-level
 - Striping
 - Remote mirroring
 - Virtualization

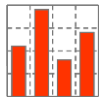
Storage Performance



How is Storage performance measured?

- Benchware developed a specific Oracle scenario for each test case

Storage Performance

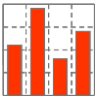


Overview of Storage performance tests with Benchware test codes

Oracle Storage Performance Sequential I/O	Test Code for Data in-memory	Test Code for Data on default Storage	Test Code for Data in Flash Cache	Test Code for Data in Cell Flash Cache
▪ Sequential read	-	STO-12	STO-13	STO-14
▪ Sequential write	-	STO-22	-	-
▪ Sequential write (LGWR stress test)	DBL-11	-	-	-

Oracle Storage Performance Random I/O	Test Code for Data in-memory	Test Code for Data on default Storage	Test Code for Data in Flash Cache	Test Code for Data in Cell Flash Cache
▪ Random read 25 I/O's per SQL	-	STO-32	STO-33	STO-34
▪ Random write (DBWR stress test)	STO-41	-	-	-
▪ Mixed random read write	-	STO-52	STO-53	STO-54
▪ More aggressive random read 25 I/O's per 125'000 rows per SQL	-	STO-62	STO-63	STO-64

Storage Performance

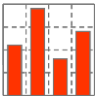


Remarks on other benchmark tools . . .

- SAP, TPC, Swingbench, Hammerora, ...
 - No specific storage performance metrics
 - SAP even does not document storage configuration for its benchmarks

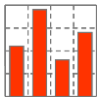
- I/O load generators like IOzone, IOgen, IOmeter, vdbench
 - Do not reflect database reality – database I/O is much more complex
 - Measure throughput and service times on operating system level, but not on database level

- `dbms_resource_manager.calibrate_io`
 - Does not recognize hybrid flash/disk architectures



- 1 Introduction to Storage Performance Tests
- 2 Storage Configuration**
- 3 Benchmark Results – Sequential I/O
- 4 Benchmark Results – Random I/O
- 5 Reviewing Storage Benchmark Results

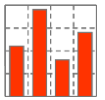
Storage Performance



Storage System Architecture and Configuration

Internal Storage (PCI attached flash within Server)	HDS UCP
PCI flash storage <ul style="list-style-type: none">#modulesCapacity	16 16 x 1.2 = 19.2 TByte
FC channels	2 x 8 Gbps

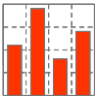
Storage Performance



Volume and File Management

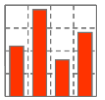
Volume Manager	HDS UCP
#LUN's	32
Concatenated oder Striped	striped
Queue depth	-
Special parameters	Preferred read from ASM failure group on FusionIO cards

File System	HDS UCP
Block size [KByte]	-
Direct I/O	Yes
Concurrent I/O	Yes
Asynchronous I/O	Yes
Special parameters	-

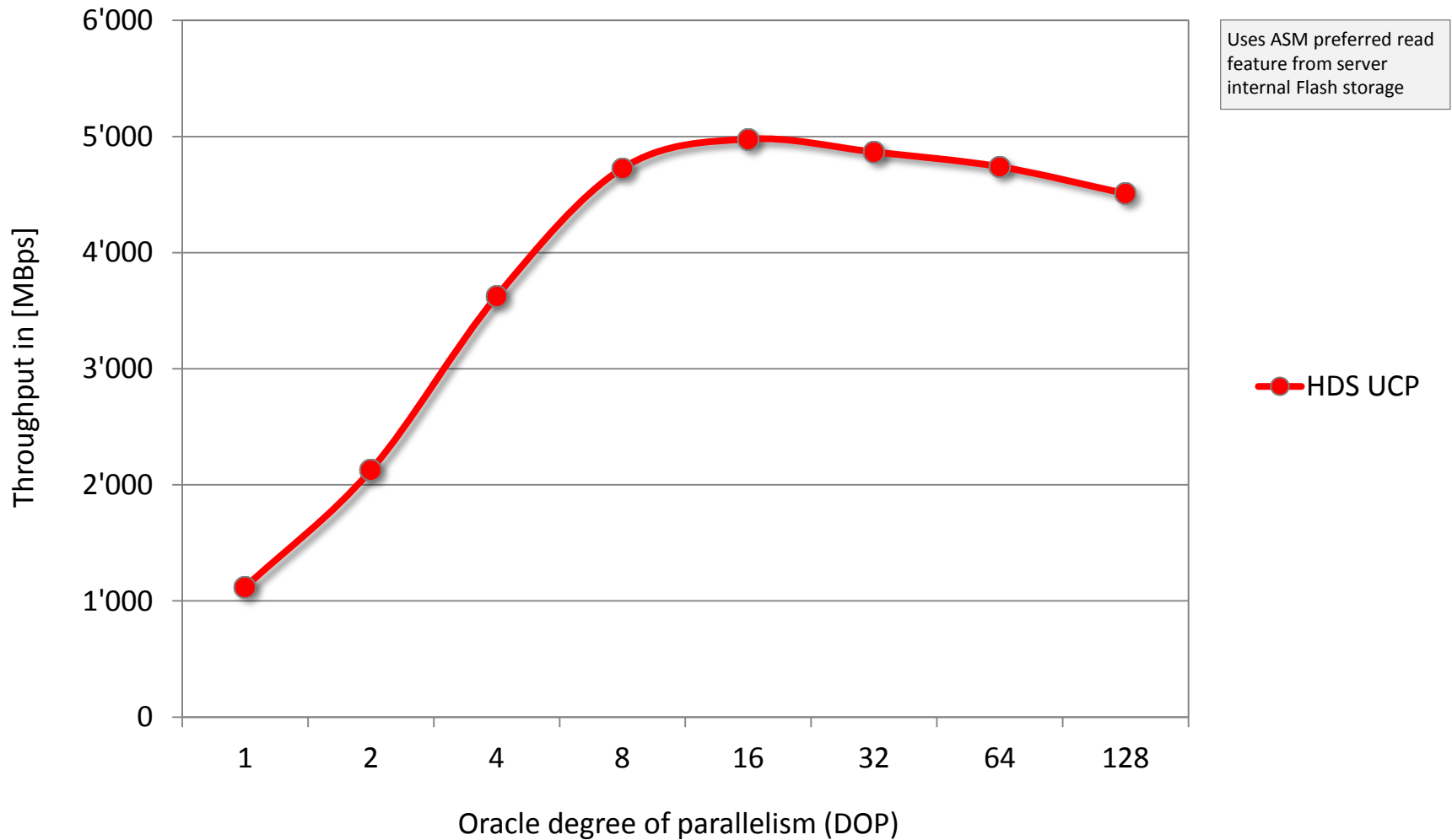


- 1 Introduction to Storage Performance Tests
- 2 Storage Configuration
- 3 Benchmark Results – Sequential I/O**
- 4 Benchmark Results – Random I/O
- 5 Reviewing Storage Benchmark Results

Storage Performance



Sequential read, single process



Storage Performance



Sequential read, single process

HDS UCP with
Flash

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]
13	1 STO-12	1	1	1	4	1	1125	143097	1118	3	2	0	0	0	0	316
	2 STO-12	1	1	2	6	2	2136	272113	2126	2	1	0	0	0	0	294
	3 STO-12	1	1	4	10	2	3639	463775	3623	3	1	0	0	0	0	285
	4 STO-12	1	1	8	14	3	4743	604921	4726	2	1	0	0	0	0	276
	5 STO-12	1	1	16	16	4	4996	637244	4978	2	1	0	0	0	0	262
	6 STO-12	1	1	32	17	4	5289	622978	4867	2	1	0	0	0	0	268
	7 STO-12	1	1	64	19	5	5943	606679	4740	2	1	0	0	0	0	258
	8 STO-12	1	1	128	21	6	8666	577129	4509	3	2	0	0	0	0	223

Legend:

#N number of RAC nodes

#J number of jobs

#T number of threads (PX)

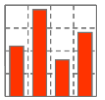
[s] elapsed time in seconds

[iops] i/o operations per second

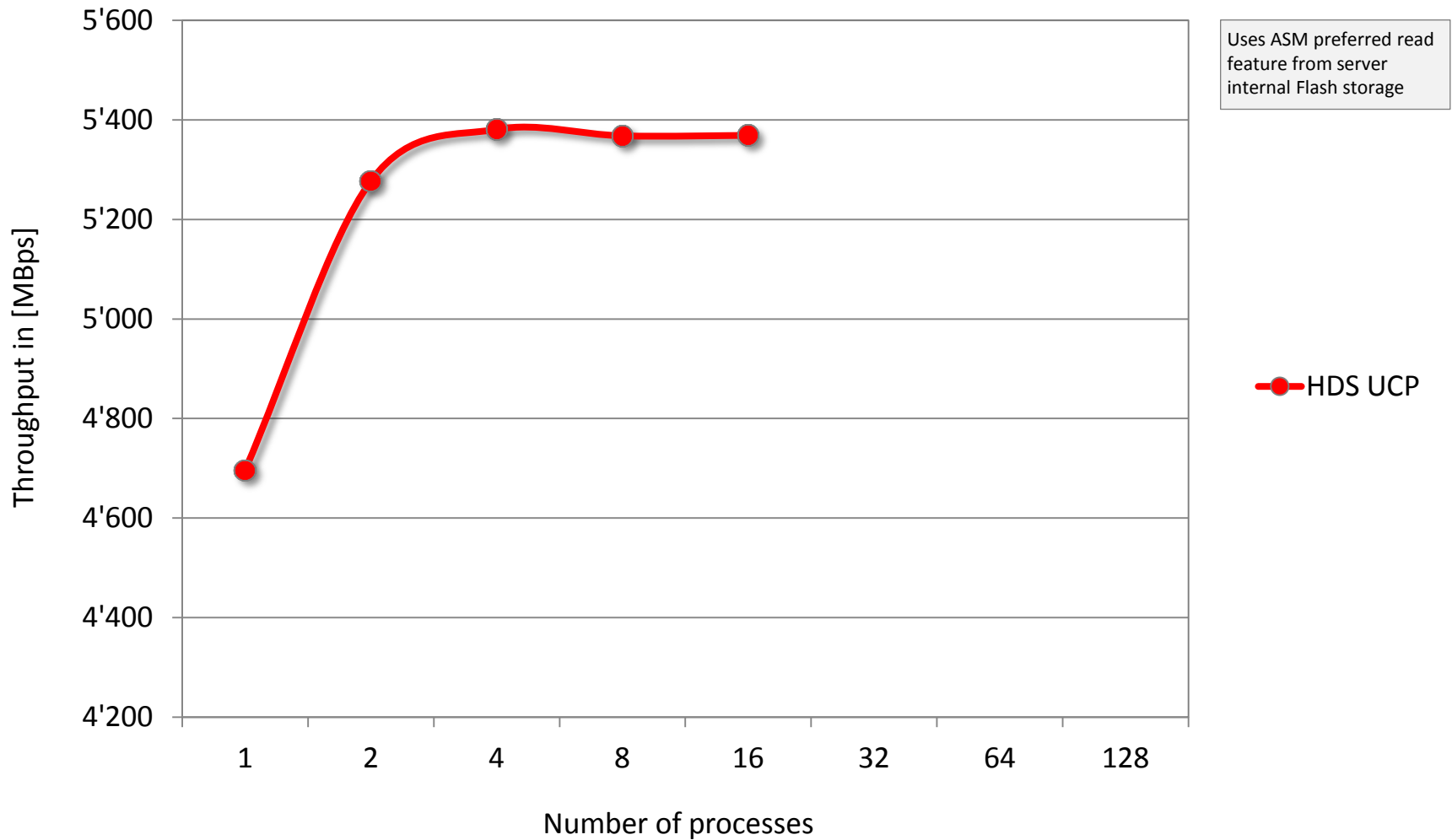
[bps] blocks per second

[MBps] mega byte per second

Storage Performance



Sequential read, multiple processes



Storage Performance



Sequential read, multiple processes

HDS UCP with
Flash

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]
14	1 STO-12	1	1	8	14	3	4716	601032	4696	2	1	0	0	0	0	272
	2 STO-12	1	2	8	17	4	5294	675396	5277	5	5	0	2	0	0	309
	3 STO-12	1	4	8	18	5	5401	688771	5381	3	4	0	0	0	0	303
	4 STO-12	1	8	8	18	5	5387	687070	5368	3	3	0	0	0	0	324
	5 STO-12	1	16	8	19	5	5386	687199	5369	3	2	0	2	0	0	329

Legend:

#N number of RAC nodes

#J number of jobs

#T number of threads (PX)

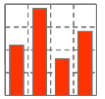
[s] elapsed time in seconds

[iops] i/o operations per second

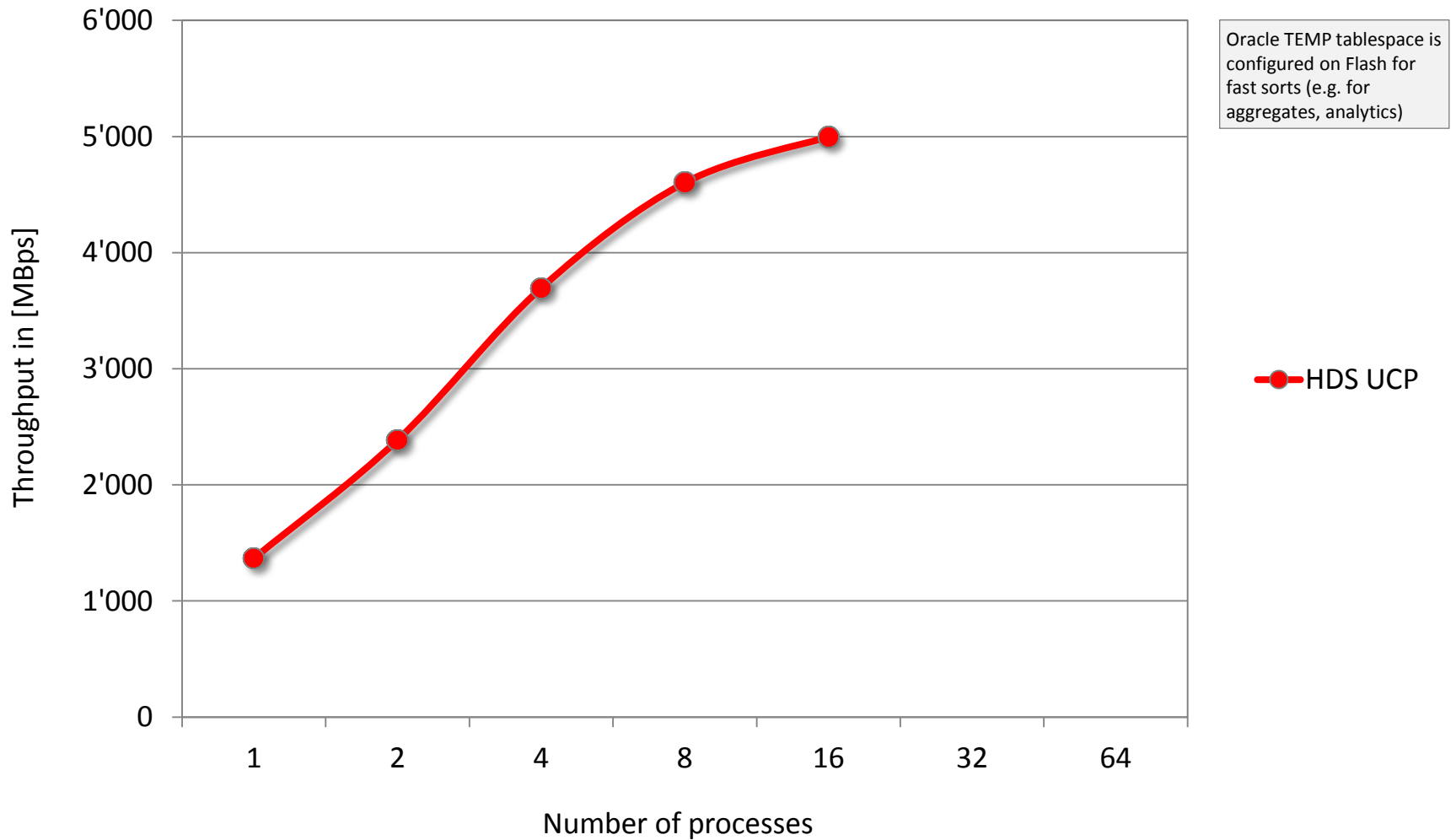
[bps] blocks per second

[MBps] mega byte per second

Storage Performance



Sequential write



Storage Performance



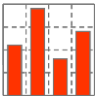
Sequential write

HDS UCP with
Flash

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]
19	1 STO-22	1	1	1	3	2	15	0	0	1550	88	1366	86	0	0	48
	2 STO-22	1	2	1	4	3	25	0	0	2702	156	2385	150	0	0	55
	3 STO-22	1	4	1	6	4	63	0	1	4137	243	3694	163	0	0	71
	4 STO-22	1	8	1	8	6	75	3	1	5134	312	4602	178	0	0	114
	5 STO-22	1	16	1	9	6	101	4	2	5571	319	4996	202	0	0	210

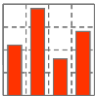
Legend:

- #N number of RAC nodes
- #J number of jobs
- #T number of threads (PX)
- [s] elapsed time in seconds
- [iops] i/o operations per second
- [bps] blocks per second
- [MBps] mega byte per second

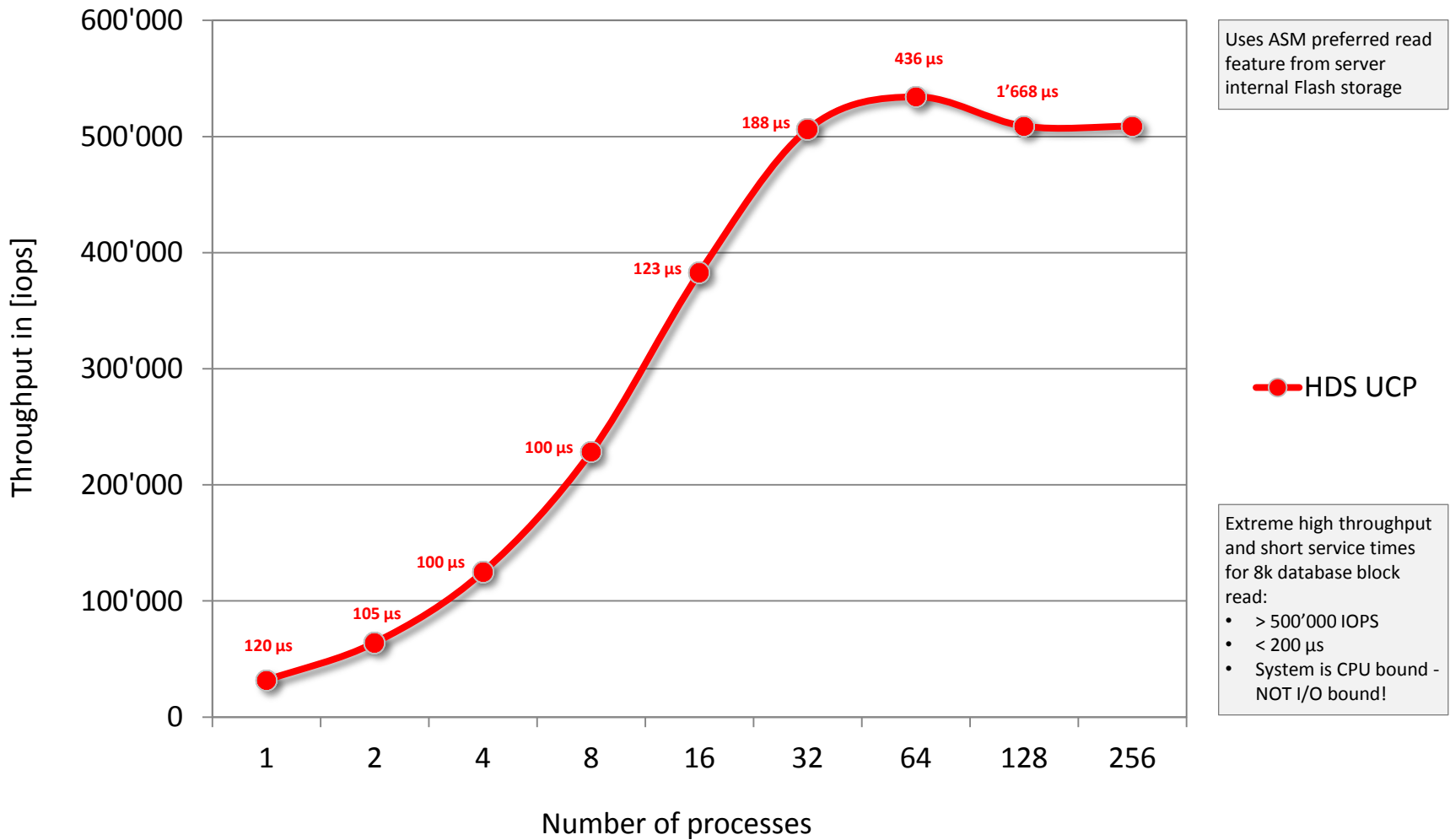


- 1 Introduction to Storage Performance Tests
- 2 Storage Configuration
- 3 Benchmark Results – Sequential I/O
- 4 Benchmark Results – Random I/O**
- 5 Reviewing Storage Benchmark Results

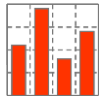
Storage Performance



Random read



Storage Performance



Random read

HDS UCP with
Flash

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
	18 STO-62	1	256	1	98	48	508889	508888	3976	793	730	6	61	0	0	310

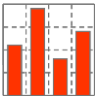
Excerpt from AWR
report Run 15 Test 16

Top 5 Timed Foreground Events					
~~~~~					
Event	Waits	Time (s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		4,451		20.9	
db file parallel read	2,686,643	2,033	1	9.6	User I/O
db file sequential read	3,079,002	1,343	0	6.3	User I/O

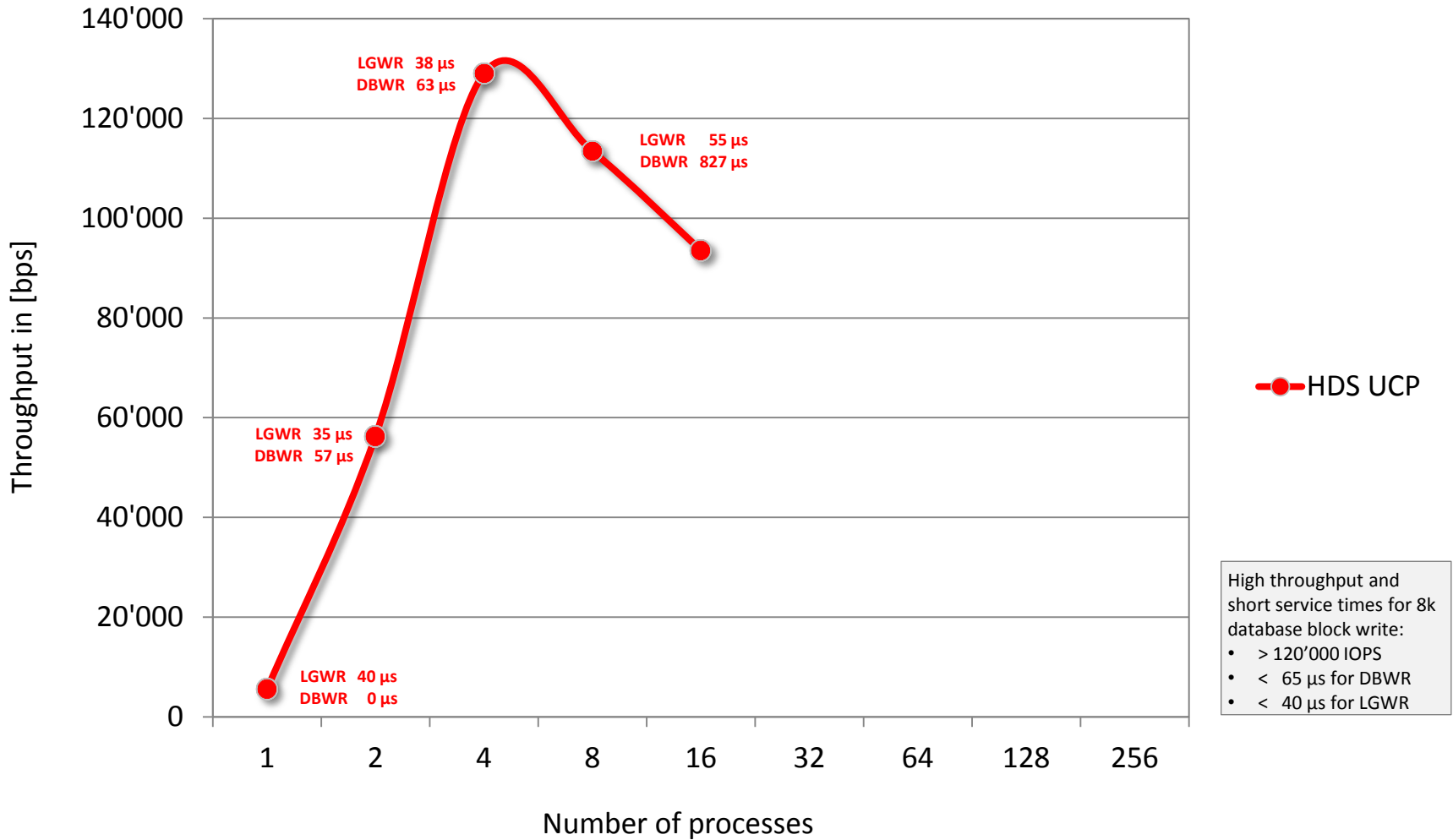
### Remarks:

- In this test case, Oracle optimizes random I/O operations
- "db file sequential read" is a true single block read operation (one system call per database block read)
- "db file parallel read" is an optimized non-contiguous multi block random read operation (one system call for at least 2 or more database block read)

# Storage Performance



## Random write



# Storage Performance



## Random write

HDS UCP with Flash

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]
20	1 STO-41	1	1	1	5	1	22	19	0	5158	5549	62	1714	0	0	59
	2 STO-41	1	2	1	12	3	35	115	1	27756	56255	478	3148	0	0	57
	3 STO-41	1	4	1	29	9	42	190	3	111034	129013	1080	4645	0	0	62
	4 STO-41	1	8	1	42	10	181	360	122	112373	115463	1140	6292	0	0	64
	5 STO-41	1	16	1	64	11	290	579	211	93257	93564	1155	4815	0	0	78
	6 STO-41	1	32	1	96	11	441	886	289	65643	68580	1128	1582	0	0	108
	7 STO-41	1	64	1	96	8	513	949	320	41714	44981	985	166	0	0	208
	8 STO-41	1	128	1	94	7	651	1299	343	32661	35187	928	40	0	0	316

Excerpt from AWR report Run 20 Test 3

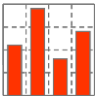
Load Profile	Per Second	Per Transaction	Per Exec	Per Call
Block changes:	326,418.7	52.1		
Physical reads:	183.0	0.0		
Physical writes:	123,972.5	19.8		

Background Wait Events		DB/Inst: XXXXX/XXXXX		Snaps: 9569-9570		
Event	Waits	%Time -outs	Total Wait Time (s)	Avg wait (ms)	Waits /txn	% bg time
log file parallel write	288,039	0	11	0	0.7	4.2
db file parallel write	112,121	0	7	0	0.3	2.7

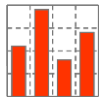
Statistic	Total	per Second	per Trans
physical write total IO requests	6,884,147	106,696.2	17.0
physical writes	7,998,828	123,972.5	19.8
physical writes from cache	7,998,483	123,967.1	19.8
redo writes	288,039	4,464.3	0.7



- 1 Introduction to Storage Performance Tests
- 2 Storage Configuration
- 3 Benchmark Results – Sequential I/O
- 4 Benchmark Results – Random I/O
- 5 Reviewing Storage Benchmark Results**



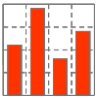
# Benchmark Results



## Summary Storage Performance

	Metric		HDS UCP With Server internal Flash
Sequential I/O			
▪ Read	[MBps]		5'381
▪ Write	[MBps]		4'996
Random I/O read			
▪ Throughput	[iops]		534'207
▪ Service time	[ms]		0.436
▪ Optimal throughput with good service time	[iops@ms]		506'047@0.188
Random I/O write			
▪ Throughput	[iops]		111'034
▪ Throughput	[bps]		129'013
▪ Service time, DBWR multi block	[ms]		0.063
▪ Service time, LGWR multi block	[ms]		0.038

# Benchmark Results



## Reviewing Storage Performance

- Extreme high performance on server internal Flash storage
  - Sequential read > 5.3 GBps
  - Random read > 500'000 IOPS for 8 kByte db block size
  - Random read < 200  $\mu$ s service time for 8 kByte db block size
- Extreme high random read rate with 8k Oracle database blocks on a single database instance (!)

**BENCHWARE**

*swiss precision in performance measurement*

*[www.benchmarkware.ch](http://www.benchmarkware.ch)*

*[info@benchmarkware.ch](mailto:info@benchmarkware.ch)*