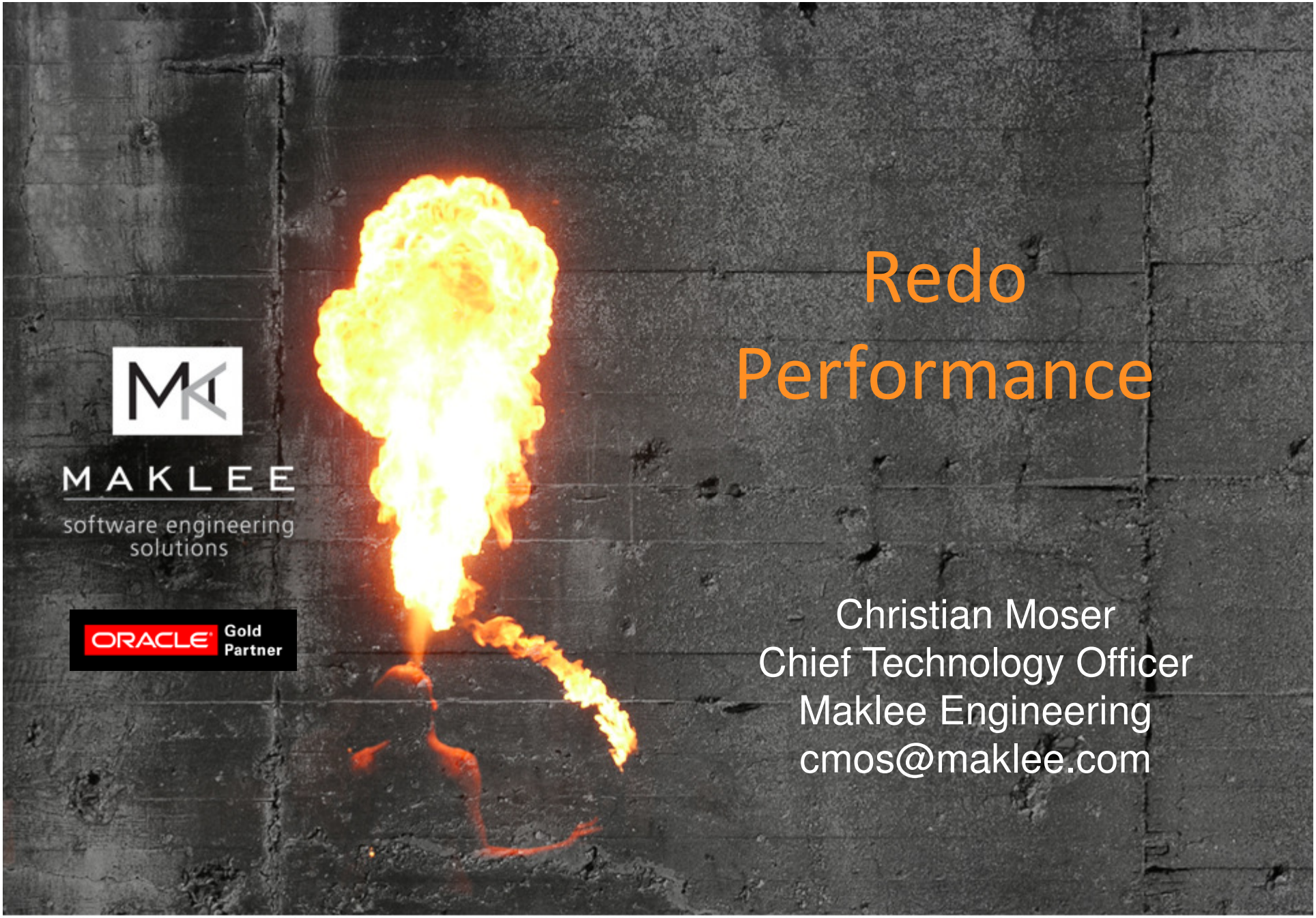
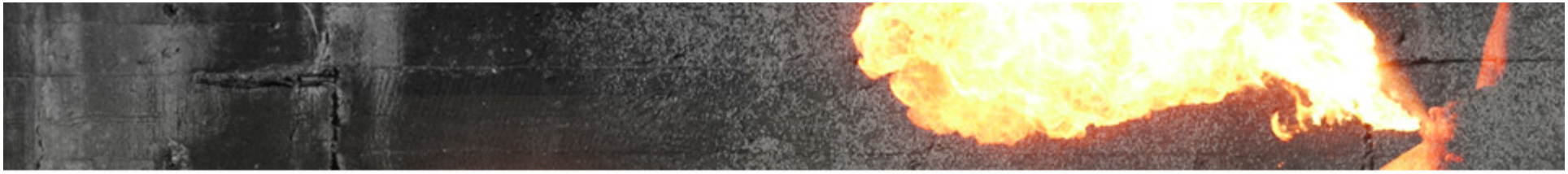


Redo Performance

Christian Moser
Chief Technology Officer
Maklee Engineering
cmos@maklee.com

Benchmark Goals

- Benchmark Redo performance
  - Perform test to measure maximum bandwidth of redo subsystem
- Determine what is the optimal configuration for the Redo sub system.
  - Single or multi member logfile group
  - How many redo log files?
  - When talking with customers about how to setup and configure redo log files these are the 2 battles we normally loose



Maximum Bandwidth of Redo Subsystem

MAKLEE

Redo Test

- The first test stressed the logwriter (LGWR)
- Database is in ”no archivelog” as we are not interested in the archiving performance at this point
- Large redo log files
- All tables and indexes are cached in SGA
  - No need for the DBWR to flush buffers
- LGWR is the only active process doing physical IOs
- 80 sessions updating random rows in separate tables
  - No waits and no row lock contention
- Measured throughput of redo subsystem

Redo Size per Second

- Over 1.1 GB redo generated per second and written by LGWR

Cache Sizes

	Begin	End		
Buffer Cache:	256,000M	256,000M	Std Block Size:	8K
Shared Pool Size:	51,200M	51,200M	Log Buffer:	746,764K

Load Profile

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	55.3	0.1	0.06	1.39
DB CPU(s):	18.9	0.0	0.02	0.48
Redo size:	1,170,702,693.4	1,387,310.9		
Logical reads:	390,154.7	462.3		
Block changes:	532,617.9	631.2		
Physical reads:	1.6	0.0		
Physical writes:	0.2	0.0		

Function Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
LGWR	0M	0.46	0M	53.2G	1225.71	1.1G	1308	19.70
Others	6M	6.03	.126544	1M	1.05	.021090	310	0.30
Buffer Cache Reads	1M	1.48	.021090	0M	0.00	0M	70	0.03
Direct Writes	0M	0.00	0M	0M	0.02	0M	0	
TOTAL:	7M	7.97	.147635	53.2G	1226.79	1.1G	1688	15.32

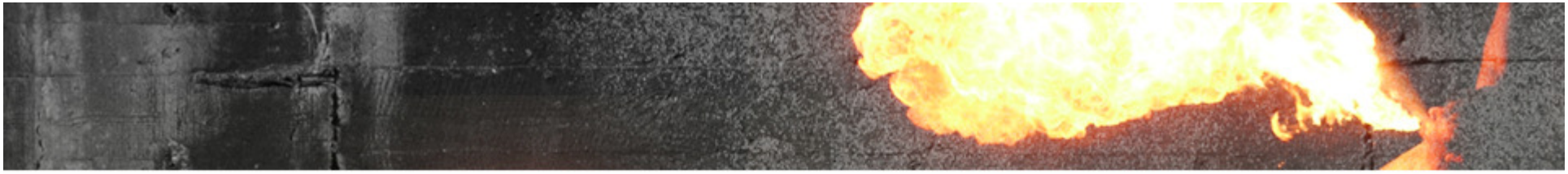
MAKLEE

log buffer space

- Due to the high rate of redo traffic generated per second, we quickly exhaust our log buffer space
- With a larger log buffer we believe the performance would be even better
- Eventhough LOG_BUFFER is set much higher, Oracle caps the size of the log buffer
  - Increasing or decreasing the size of the SGA makes the log buffer size shrink
  - We are still working on the magic formula...
  - But 1 GB/sec is impressive enough

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
log buffer space	29,363	933	32	35.57	Configuration
DB CPU		897		34.19	
latch: redo allocation	674,285	730	1	27.81	Other
latch: undo global data	27,932	25	1	0.95	Other
buffer busy waits	7,869	16	2	0.62	Concurrency



Single or Multi-Member Redo Log Group

MAKLEE

Single or Multi-Member Redo Log Group

- Maklee Engineering recommends single member Redo log groups
  - If the LUN with the redo log file is already mirrored at the storage layer, there is no need to add another member for redundancy
  - Performance is better with single-member redo log files
  - The LGWR has less work to do
  - Less burden on the overall I/O subsystem
  - The logfile write only completes when the IO to all members have been completed

- Most customers use multi-member redo log groups
  - Not trusting the storage layer
  - Paranoia wins

Redo Log Test

- The following test will stress the logwriter (LGWR)
- Database is in ”no archivelog” as we are not interested in the archiving performance at this point
- 40 sessions updating random rows in separate tables
  - No waits and no row lock contention
- Both LGWR and DBWR are active
  - Logfile switches occur
- Decent update activity without pushing the limits

Redo Size per Second

Single-Member

Cache Sizes

	Begin	End		
Buffer Cache:	256,000M	256,000M	Std Block Size:	8K
Shared Pool Size:	51,200M	51,200M	Log Buffer:	746,764K

Load Profile

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	29.5	0.1	0.06	1.41
DB CPU(s):	29.1	0.1	0.06	1.39
Redo size:	609,413,318.5	1,351,301.2		
Logical reads:	194,326.6	430.9		
Block changes:	269,631.1	597.9		
Physical reads:	3.2	0.0		
Physical writes:	53,184.9	117.9		

2-Member

Cache Sizes

	Begin	End		
Buffer Cache:	256,000M	256,000M	Std Block Size:	8K
Shared Pool Size:	51,200M	51,200M	Log Buffer:	746,764K

Load Profile

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	30.2	0.1	0.06	1.51
DB CPU(s):	29.3	0.1	0.06	1.47
Redo size:	639,516,110.9	1,493,248.2		
Logical reads:	232,229.1	542.3		
Block changes:	320,281.2	747.9		
Physical reads:	1.3	0.0		
Physical writes:	57,320.0	133.8		

Roughly same redo size generated

MAKLEE

Log Writer Process (LGWR)

- Single-Member

Function Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
LGWR	1M	2.03	.022536	26G	763.45	600.040	2961	9.34
DBWR	0M	0.00	0M	17.9G	32866.09	412.895	0	
Others	3M	4.80	.067610	1M	1.76	.022536	252	0.65
Buffer Cache Reads	1M	1.80	.022536	0M	0.00	0M	80	0.05
Direct Reads	0M	0.72	0M	0M	0.25	0M	0	
Direct Writes	0M	0.00	0M	0M	0.47	0M	0	
TOTAL:	5M	9.35	.112683	43.9G	33632.02	1012.95	3293	8.45

- 2-Member

Function Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
LGWR	1M	2.65	.021401	55.9G	1352.32	1.2G	1364	22.13
DBWR	0M	0.00	0M	20.3G	36443.98	444.772	0	
Others	3M	5.05	.064205	2M	2.31	.042803	290	1.20
Direct Reads	1M	1.03	.021401	0M	0.36	0M	0	
Buffer Cache Reads	0M	0.24	0M	0M	0.00	0M	11	0.27
Direct Writes	0M	0.00	0M	0M	0.66	0M	0	
TOTAL:	5M	8.97	.107009	76.2G	37799.64	1.6G	1665	18.34

2-Member writes twice as much data
Less is better

MAKLEE

Logfile Switch Duration

- Single-Member

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		1,293		98.63	
log file switch completion	245	12	50	0.94	Configuration
library cache: mutex X	1,171	4	4	0.34	Concurrency
log file sync	80	3	34	0.21	Commit
latch: redo allocation	1,629	0	0	0.03	Other

- 2-Member

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		1,369		97.16	
log file switch completion	239	23	98	1.66	Configuration
log file sync	81	10	125	0.72	Commit
undo segment extension	353	4	12	0.30	Configuration
log file switch (private strand flush incomplete)	42	3	76	0.23	Configuration

2-Member twice as long per logfile switch
Less is better

MAKLEE

Update Statement

- Single-Member
  - Elapsed time per execution 64.2 msec

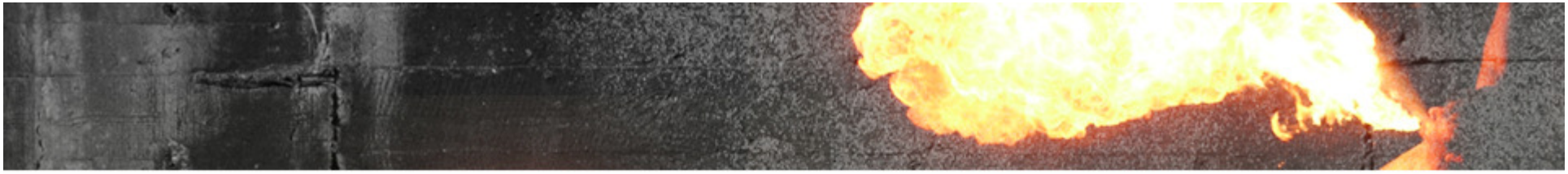
Elapsed Time (s)	Executions	Elapsed Time per Exec (s)	%Total	%CPU	%IO	SQL Id	SQL Module	SQL Text
1,302.13	40	32.55	99.34	98.65	0.00	4mssqv00uhy0c	SQL*Plus	DECLARE x NUMBER := 0; v_r PLS...
1,285.22	20,000	0.06	98.05	98.97	0.00	bz5nnm9wc3wba	SQL*Plus	UPDATE CF1 SET C2 = 'AAAAAAAAB...

- 2-Member
  - Elapsed time per execution 69.1 msec

Elapsed Time (s)	Executions	Elapsed Time per Exec (s)	%Total	%CPU	%IO	SQL Id	SQL Module	SQL Text
1,397.28	40	34.93	99.20	97.44	0.00	4mssqv00uhy0c	SQL*Plus	DECLARE x NUMBER := 0; v_r PLS...
1,381.79	20,000	0.07	98.10	97.49	0.00	bz5nnm9wc3wba	SQL*Plus	UPDATE CF1 SET C2 = 'AAAAAAAAB...

2-Member 7% longer elapsed time
Less is better

MAKLEE



How Many Redo Log Files

MAKLEE

How Many Redo Log Files

- Maklee Engineering's position
  - Determine how much redo space you generate per day or per hour on average
  - Decide how often you want to perform a logfile switch
  - Divide by three and create 3 decent size redo logfiles

- Common customer position
  - Most customers believe that many small redo logfiles are better than a few large ones
  - Think that 12 x 1 GB is better than 3 x 4 GB

Redo Log Test

- The following test will stress the logwriter (LGWR)
- Database is in ”no archivelog” as we are not interested in the archiving performance at this point
- 40 sessions updating random rows in separate tables
  - No waits and no row lock contention
- Both LGWR and DBWR are active
  - Logfile switches occur
- Decent update activity without pushing the limits
- Redo log files used for test
  - Large 3 x 4 GB
  - Small 12 x 1 GB

Redo Size per Second

Large Redo Logs

Cache Sizes

	Begin	End		
Buffer Cache:	256,000M	256,000M	Std Block Size:	8K
Shared Pool Size:	51,200M	51,200M	Log Buffer:	746,764K

Load Profile

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	29.5	0.1	0.06	1.41
DB CPU(s):	29.1	0.1	0.06	1.39
Redo size:	609,413,318.5	1,351,301.2		
Logical reads:	194,326.6	430.9		
Block changes:	269,631.1	597.9		
Physical reads:	3.2	0.0		
Physical writes:	53,184.9	117.9		

Small Redo Logs

Cache Sizes

	Begin	End		
Buffer Cache:	256,000M	256,000M	Std Block Size:	8K
Shared Pool Size:	51,200M	51,200M	Log Buffer:	746,764K

Load Profile

	Per Second	Per Transaction	Per Exec	Per Call
DB Time(s):	31.1	0.1	0.07	1.72
DB CPU(s):	29.4	0.1	0.06	1.63
Redo size:	520,586,544.5	1,409,386.5		
Logical reads:	177,880.3	481.6		
Block changes:	246,470.2	667.3		
Physical reads:	7,494.3	20.3		
Physical writes:	50,213.9	135.9		

Large Redo Logs 15 % higher redo size
More work done
More is better

MAKLEE

Log Writer Process (LGWR)

- Large Redo Logs

Function Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
LGWR	1M	2.03	.022536	26G	763.45	600.040	2961	9.34
DBWR	0M	0.00	0M	17.9G	32866.09	412.895	0	
Others	3M	4.80	.067610	1M	1.76	.022536	252	0.65
Buffer Cache Reads	1M	1.80	.022536	0M	0.00	0M	80	0.05
Direct Reads	0M	0.72	0M	0M	0.25	0M	0	
Direct Writes	0M	0.00	0M	0M	0.47	0M	0	
TOTAL:	5M	9.35	.112683	43.9G	33632.02	1012.95	3293	8.45

- Small Redo Logs

Function Name	Reads: Data	Reqs per sec	Data per sec	Writes: Data	Reqs per sec	Data per sec	Waits: Count	Avg Tm(ms)
LGWR	5M	7.75	.092278	27.1G	782.54	512.697	5858	4.46
DBWR	0M	0.00	0M	20.3G	42402.04	384.541	0	
Buffer Cache Reads	3.1G	1171.86	58.5228	0M	0.00	0M	63.5K	4.24
Others	6M	7.25	.110733	4M	5.04	.073822	529	0.86
Direct Reads	2M	3.25	.036911	1M	1.14	.018455	0	
Direct Writes	0M	0.00	0M	1M	2.10	.018455	0	
TOTAL:	3.1G	1190.11	58.7627	47.5G	43192.86	897.349	69.9K	4.23

MAKLEE

Large Redo Logs 15% more data written
More is better

Number of Logfile Switches

Large Redo Logs

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		1,293		98.63	
log file switch completion	245	12	50	0.94	Configuration
library cache: mutex X	1,171	4	4	0.34	Concurrency
log file sync	80	3	34	0.21	Commit
latch: redo allocation	1,629	0	0	0.03	Other

Statistic	Total	per Hour
log switches (derived)	6	486.79

Small Redo Logs

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		1,594		94.60	
log file switch completion	1,105	51	46	3.01	Configuration
db file sequential read	25,481	25	1	1.49	User I/O
db file scattered read	38,019	14	0	0.81	User I/O
cursor: pin S wait on X	598	7	12	0.41	Concurrency

Statistic	Total	per Hour
log switches (derived)	28	1,860.33

Small Redo Logs 4X log switches
Same wait time per log switch
Less is better

MAKLEE

Update Statement

- Large Redo Logs
  - Elapsed time per execution 64.2 msec

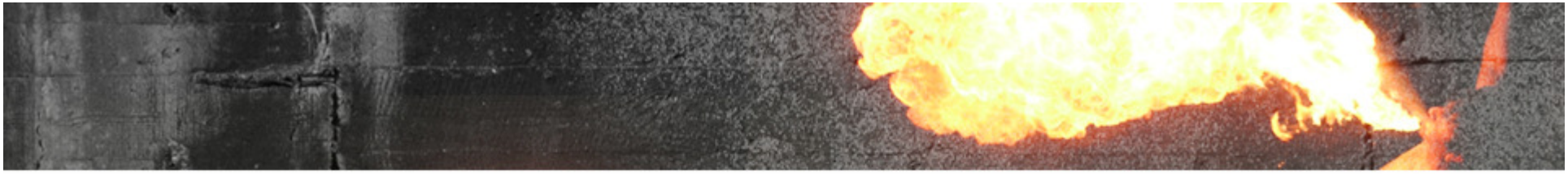
Elapsed Time (s)	Executions	Elapsed Time per Exec (s)	%Total	%CPU	%IO	SQL Id	SQL Module	SQL Text
1,302.13	40	32.55	99.34	98.65	0.00	4mssqv00uhy0c	SQL*Plus	DECLARE x NUMBER := 0; v_r PLS...
1,285.22	20,000	0.06	98.05	98.97	0.00	bz5nnm9wc3wba	SQL*Plus	UPDATE CF1 SET C2 = 'AAAAAAAAB...

- Small Redo Logs
  - Elapsed time per execution 82.2 msec

Elapsed Time (s)	Executions	Elapsed Time per Exec (s)	%Total	%CPU	%IO	SQL Id	SQL Module	SQL Text
1,665.99	40	41.65	98.87	95.00	2.32	4mssqv00uhy0c	SQL*Plus	DECLARE x NUMBER := 0; v_r PLS...
1,644.87	20,000	0.08	97.62	95.43	2.35	bz5nnm9wc3wba	SQL*Plus	UPDATE CF1 SET C2 = 'AAAAAAAAB...

Small Redo Logs 28% longer elapsed time
Less is better

MAKLEE



Technical Slides

MAKLEE

System Configuration

- Hardware
  - HP ProLiant DL980 G7 Server
  - Intel Xeon E7-4870 processors @2.4 GHz
  - 8 deca-core, hyperthreads disabled, 80 logical CPUs
  - 1 TB physical memory
- Operating System
  - Red Hat Enterprise Linux Server release 6.1 (Santiago)
  - Version 2.6.32-131.17.1.el6.x86_64
  - 64-bit kernel (x86), 4 KB pagesize
- Storage Sub-System
  - Fusion ioDrive
  - 8 x HP 1.28TB MLC PCIe IO Accelerator

MAKLEE