Mike Revitt

Tuesday, 22 December 2020

DATABASE FLASH DISK PERFORMANCE

Introduction

There are any number of ways to address database performance from straight forward database tuning to extremely complex high-performance SAN arrays, this study was carried out to determine if the Sun Storage F5100 Flash Array can deliver cost efficient performance for I/O constrained databases.

When I started this case study, I naively expected it to be relatively straight forward to compare the performance of the Sun Storage F5100 Flash Array with an alternative SAN storage array. How wrong I was, whilst it was very easy to make the SAN storage the performance bottle neck, no matter how I setup my environment I could not configure a situation where the Sun Storage F5100 Flash Array became the bottle neck. This amply demonstrated the performance superiority of the F5100, but made measuring the gains a challenge. Eventually I used a variant of the TPC-H benchmark to deliver sufficient I/O for the tests.

High Performance at Commodity Pricing

In my write intensive tests, I was able to achieve the following improvements

- 662% improvement in elapsed run time
- 382% improvement in DB I/O
- 278% improvement in DB IOPS
- 400% increase in CPU utilisation

In my read intensive tests, I was able to achieve the following improvements

- 588% improvement in peak DB I/O
- 762% improvement in average DB I/O
- 714% improvement in peak Server IOPS
- 786% improvement in average Server IOPS

The Sun Storage F5100 Flash Array offers significant benefits to the DB performance problems at commodity pricing, with a starting list price of just £37,300 for the smallest appliance rising to £129,800 for the full-size appliance. Performance scales linearly when increasing the size of the appliance and multiple appliances can be added to the database storage grid to facilitate scale out requirements.

Better business agility and continuity

All of the performance improvements, which are noted above, were achieved without incurring any database downtime, this was done by utilising standard Oracle features allowing the Sun Storage F5100 Flash Array to become a standard component of the DBAs tuning toolkit.

In the write tests, we identified the following areas of I/O contention.

- Redo Logs
- Archive Logs
- Recovery Area
- Hot Tables

One of the main benefits of the Sun Storage F5100 Flash Array is that, as an appliance there is no complex architecture required prior to use; simply attach the unit to a spare SAS port, preferably two for redundancy, and it is ready to go.

If there are no spare SAS ports, a SAS card will have to be added to the server into a PCI-X 2.0 slot; in most cases this will require a minor outage to implement.

Higher server utilisation

To measure the gains of the system the following write intensive tests were performed, as detailed below.

- Test 1 – Traditional layout, Archive, Redo and Data all on a RAID 10 SAN
- Test 3 – Move the REDO to the F5100
- Test 2 – Move the Recovery Area (Archive Logs and Flashback) to the F5100
- Test 4 – Move the REDO and Recovery Area to the F5100
- Test 5 – Move the REDO, Recovery Area and hot tables to the F5100

Tested Value	Test 1	Test 3	Test 2	Test 4	Test 5
Test Start Time	17:03	21:08	18:19	22:15	23:21
Test End Time	17:56	21:52	18:34	22.27	23:29
Test Duration	00:53	00:44	00:15	00:13	00:08
DB Throughput Peak	55 MB/s	68 MB/s	85 MB/s	130 MB/s	150 MB/s
DB Throughput Avg.	34 MB/s	38 MB/s	70 MB/s	90 MB/s	130 MB/s
DB IOPS Peak	1,100	1,250	1,800	2,300	2,150
DB IOPS Avg.	720	950	1,350	1,950	2,000
Server IOPS Peak	2,018	2,043	3,608	4,208	4,563
Server IPOS Avg.	1,600	1,900	3,200	4,000	4,500
CPU Utilisation Peak	31.74%	30.28%	63.93%	65.70%	94.13%
CPU Utilisation Avg.	23%	24%	53%	60%	92%
CPU I/O Wait Peak	50.08%	58.19%	27.07%	30.47	2.06%
CPU I/O Wait Avg.	50.0%	48.7%	23.4%	22.1%	1.5%
Data Generator CPU Load[1]	14.87%	16.56%	40.50%	46.43%	64.62%
SQL Response Time[2]	106.61%	90.49%	17.73%	13.28%	3.76%

[1] The load of the client generating the load, taken at 40% completion

[2] Taken at 40% completion mark against a baseline set at SQL Response time of 8,801 seconds

The tests are shown in order of performance gain, and you will note that the relocation of the Recovery Area realised a greater performance gain than the relocation of the REDO logs against my expectation.

The other tests I ran was to measure the gains of the system when running read intensive operations, as these tests generated very little REDO or Archive the only measureable benefits occurred when I moved the hot table onto the F5100 as shown below.

- Test 1 – Traditional layout, Archive, Redo and Data all on a RAID 10 SAN
- Test 2 – Traditional layout, Archive, Redo and Data all on the F5100

Test Value	Test 1	Test 2
DB Throughput Peak	340 MB/s	2,000 MB/s
DB Throughput Avg.	260 MB/s	1,980 MB/s
Server IOPS Peak	1,400	10,000
Server IPOS Avg.	1,050	8,250
CPU Utilisation Peak	25%	65%
CPU Utilisation Avg.	25%	65%
CPU I/O Wait Peak	70%	21.1%

It is clear from these results that it is possible to realise significant performance gains by simply relocating high I/O activities onto the Sun Storage F5100 Flash Array.

Test Harness

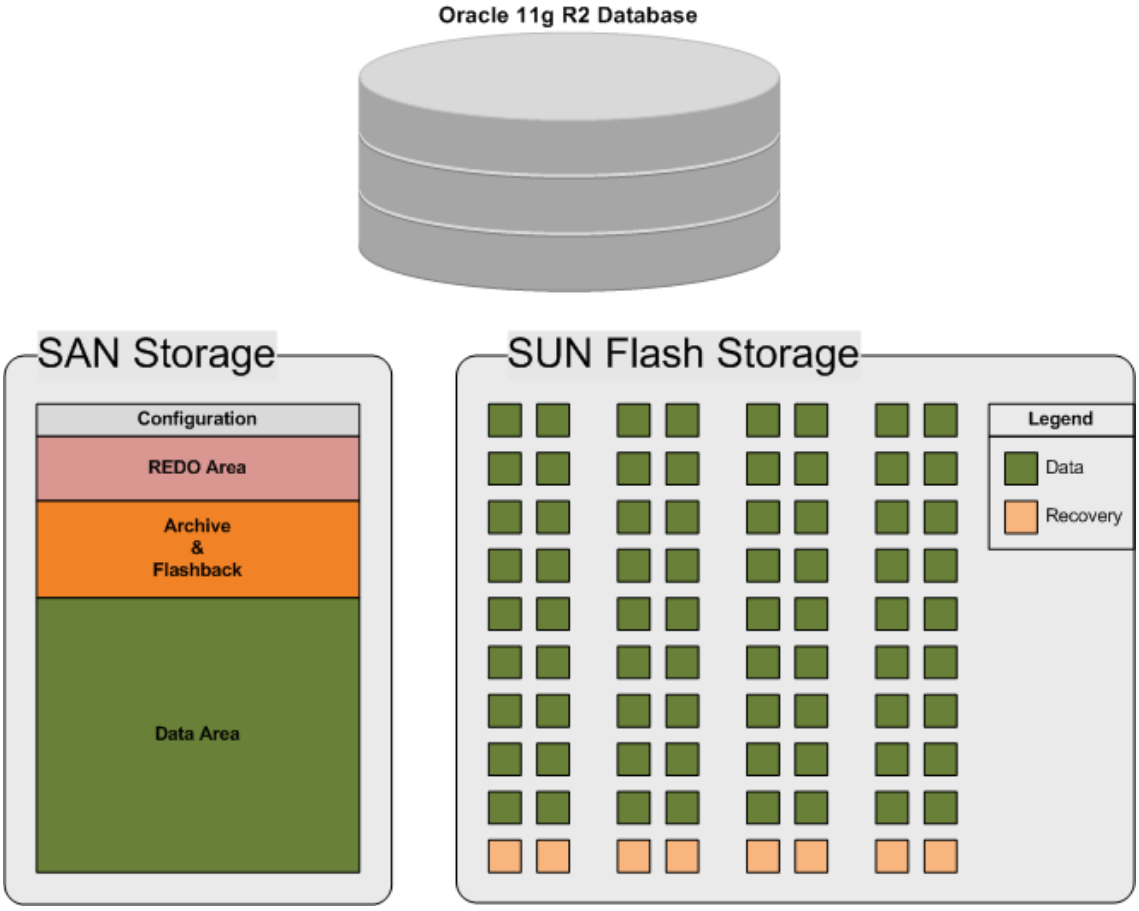
All tests were carried out on the following hardware

- Sun X4600 M2 server with 8 dual core CPUs
- Sun Storage F5100 Flash Array with 80 Flash Disks
- Dual Port, Sun SAS PCIe Cards
- Oracle Enterprise Linux R5 U4 64bit
- Oracle 11g R2 Database with a 10GB SGA
- Oracle Automatic Storage Manager for disk management.

The SAN was configured with three RAID-10 LUNs that were presented to the database server; all disk management on the SAN was managed by the SAN array.

The Sun Storage F5100 Flash Array presents each flash module as an individual storage device, in the case of a fully size appliance this results in eighty storage devices being presented to the server. This necessitates the need for a volume manager and we used Oracle ASM to manage all database storage.

My calculations showed that there was no performance benefit from separating the REDO from the Archive areas on the Sun Storage F5100 Flash Array and so we simply balanced the recovery area across all controllers in the appliance to maximise throughput as shown in the diagram below.



Test Summary

In a lot of benchmark tests, multiple tricks are used to obtain the desired results and I nearly had to resort to this trick myself as I will explain later. But the purpose of these tests was not to skew the results in favour of a desired outcome, but to represent the Sun Storage F5100 Flash Array as it could be deployed into any environment. Therefore, I did not perform any database tuning or SQL tuning in any of the environments. Every test used the same start and end point against the same data set and the database was deployed with default settings throughout, in addition to this the database was shutdown and restarted between tests to ensure consistency.

Initially I planned to use a TPC-C benchmark to demonstrate the benefits of the Sun Storage F5100 Flash Array and all was going well during the tests with the SAN storage as it was relatively easy to load up the servers such that storage was the bottle neck. However, when I connected Sun Storage F5100 Flash Array the CPU became the immediate bottleneck and I started resorting to tricks to force additional I/O, primarily by reducing the size of the SGA to prevent any caching from taking place. This quickly lead me to running tests against a database that was not viable and definitely not representative of those in common usage so I abandoned this approach and changed strategy to a variant of the TPC-H benchmark using the settings as shown above.

In our test harness X4600 M2 has a PCI-e bandwidth of 20GB/s and we used four SAS HBA cards, each of which was connected to an eight lane slot giving a total theoretical slot bandwidth of 16GB/s. Each of the HBA cards is capable of 12Gb/s giving a total theoretical bandwidth of 48Gb/s or 6GB/s. In our read intensive tests we achieved a peak throughput of 2GB/s by simply running a test out of the box against a default database and without performing any tuning or optimisation. For me this is an impressive result that shows the true potential of this appliance.