Science & Technology Facilities Council

Technical Report

DL-TR-2010-004

Working Notes of Distributed Computing Group: Benchmarking the RamSam-20 Flash Storage PCIe Card

DJ Cable

June 2010

Enquiries about copyright, reproduction and requests for additional copies of this report should be addressed to:

Chadwick Library
Science and Technology Facilities Council
Daresbury Laboratory
Daresbury Science and Innovation Campus
Warrington
WA4 4AD

Tel: +44(0)1925 603397
Fax: +44(0)1925 603779
email: library@dl.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk/

ISSN 1362-0207

Working Notes of Distributed Computing Group

Benchmarking the RamSan-20 Flash Storage PCIe card

Introduction

Regular solid state drives (such as the Intel and MemoRight drives previously tested[1]) emulate hard disk devices, complete with disk controllers and SATA interface. The RamSan-20[2], from Texas Memory Systems Inc.[3] of Houston, Texas takes a different approach, providing 450GB (useable – 640GB raw) of SLC NAND flash storage on a full-size PCIe card. Applications for this technology are claimed to include video editing, financial modelling, simulations and rendering. The Distributed Computing Group recently had the opportunity to briefly evaluate the RamSan-20. This document summarises our findings.

The RamSan-20 card

The card features four Xilinx FPGAs, each acting as controller for 20 flash chips, with ECC, in a RAID-5 configuration. The total amount of storage is 450GB. There is a PowerPC processor, which runs the wear-levelling algorithm. A PCIe (x4) interface is provided. The product data sheet claims the card draws only 15 watts.

The UK supplier, NAS (UK) Ltd.[4], quoted a list price of $18,000 (£11,613 at an exchange rate of £1:$1.55) giving a cost per gigabyte of $40 (£25.81).

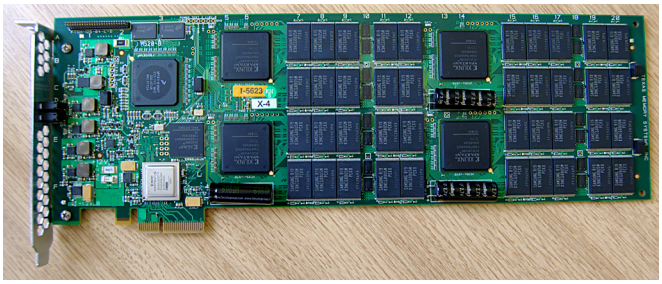

Figure 1: Front view of the RamSan-20. Note the flash memory chips (middle and right), the four FPGA controllers (next to the memory), and the PowerPC chip (on the left).

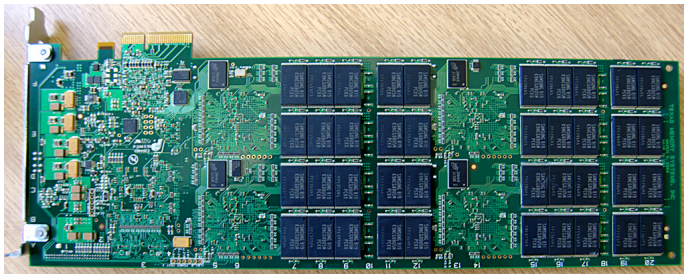

Figure 2: Rear view of the RamSan-20.

Specifications (from product datasheet):

IOPS:	Random reads	120,000
	Random writes	50,000
	Random read/write, 70% reads	70,000
Bandwidth:	Reads	700 MB/s
	Writes	500 MB/s
Read latency:		50 µsec
CPU:	PowerPC	@ 333 Mhz
Life:	With 25% writes	12 years
Power required:		15 watts

[1] http://www.cse.scitech.ac.uk/disco/publications/WorkingNotes.SSD.pdf
[2] http://www.ramsan.com/products/ramsan-20.htm
[3] http://www.ramsan.com/
[4] http://www.nas.uk.com/index.php

Host hardware

The card was installed in a Viglen mini-tower PC, which had 1 free PCIe slot. The PC has an Intel Core 2 Duo CPU, model E8400, running at 3.00GHz; 4GB of RAM; on-board Intel graphics; and a 250GB SATA hard drive.

Host environment

The operating system was Scientific Linux 5.4, with kernel release 2.6.18-164.11.1.el5.

Installation

Physical installation of the RamSan-20 is a simple matter of inserting the card into a free PCIe slot. On the test host this required temporary removal of the internal hard drive to provide enough clearance. After the card had been installed the hard drive was replaced. Following installation it was found that the DVI output from the motherboard was no longer working. VGA output was, however, still available. As the integrated Intel graphics engine feeds both outputs, this effect was unexpected.

Software drivers are provided with the card, including a kernel module, udm.ko. A minor problem was discovered when the default Scientific Linux kernel, 2.6.18-164.11.1.el5PAE, was not recognised by this module. Booting the system into 2.6.18-164.11.1.el5 resolved this issue.

Regular SSDs are treated by the OS as SCSI drives, and thus are given devices names of the form `/dev/sd<a>`. The RamSan-20 is given a device name by its driver of `/dev/tms-rs/<n>.0.0`, where `<n>` represents the card id. In this case, the device name was `/dev/tms-rs/1.0.0`.

The device cannot be partitioned. It was found that, although `fdisk` will recognise the device, it cannot write a partition table. However `mkfs` can be used to make a filesystem in the regular way. `xfs`, `ext2` and `ext3` filesystems were all tried, although `ext2` was used for benchmarking. On the test system the RamSan-20 provided a useable 414GB of disk space with `ext2`. Block size is fixed on the device at 4KB.

Benchmarking with iozone

To get an idea of the bandwidth obtainable from the RamSan-20, the iozone synthetic benchmark[1] was used in multi-threaded throughput mode, at 4, 8, 16 and 24 threads, with a command similar to this one:

```
./iozone -t 4 -F io0 io1 io2 io3 -I -s 32g -r 64k -i 0 -i 1 -i 2 -b <outputfile>.xls
```

The file size used (`-s`) varied with number of processes – 32GB at 4 and 8 threads, 24GB at 16 threads, and 16GB at 24 threads. The record size used (`-r`) was 64KB. Tests (`-I 0 –I 1 –I 2`) were seq. write/re-write, seq. read/re-read and random read/write. The `–I` flag ensured that direct io was used, bypassing the system cache.

For each test, three runs were made and the results averaged.

A direct comparison was made with the Intel X25-E 64GB SSD by installing that drive in the same test host. Three runs, each of 8 threads writing 5GB files (due to capacity limitations), were made using the Intel drive and the results averaged.

[1] http://www.iozone.org/

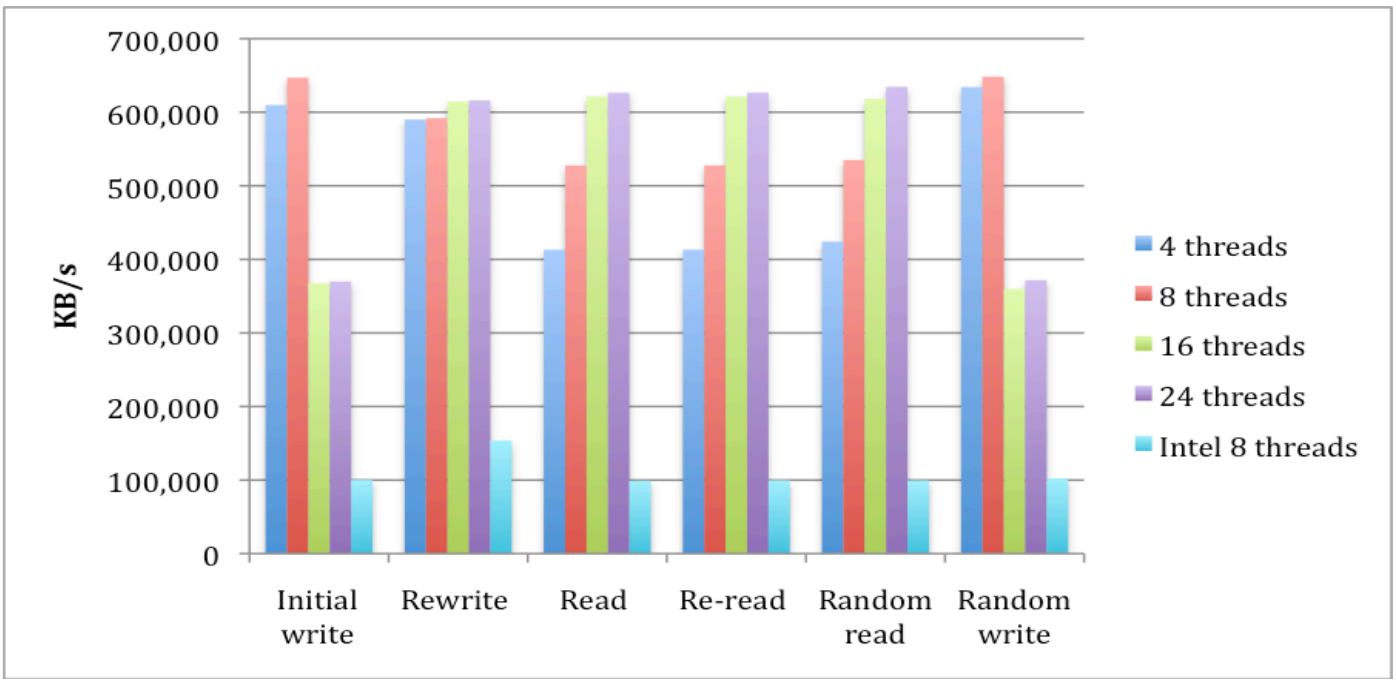

Figure 3: iozone throughput test results.

Sequential write bandwidth peaked at 646.9MB/s, almost 30% faster than claimed. Sequential read bandwidth peaked at 626.5MB/s, 89.5% of claimed performance. Read bandwidth increased with number of threads; write bandwidth decreased. Read and re-read figures were almost exactly the same, suggesting that cache effects were completely absent, as expected.

Sequential write performance showed variations by thread count when compared with the sequential re-write test. At lower thread counts (4, 8) the initial write bandwidth was higher than the re-write. However, with 16 and 24 threads the re-write bandwidth was considerably higher than the initial write. This may have been the result of i/o requests queuing on the FPGA controllers due to the higher number of threads. Unlike conventional hard drives, random reads and writes were very similar to sequential reads and writes. This is due to the solid-state nature of the drive; because there are no disk heads to move there is no seek overhead.

The Intel drive consistently achieved 100MB/s on almost every type of load, except for the re-write test, where it improved to 153MB/s. More informal tests at 2 and 4 threads suggested that the conventional SATA interface and disk controller are restricting the Intel drive's performance – or, to put it another way, the RamSan's PCIe interface and multiple FPGA controllers allow it to deliver significantly better bandwidth.

Conclusion

In terms of cost, the Intel drive can be found online for £530 (see http://bit.ly/avOAG4, for instance). This equates to £8.28 per gigabyte, which makes the RamSan-20 approximately three times as expensive (at list price – discounts may apply). In this test it delivered six times the performance (at eight threads) and seven times the capacity. We do not have enough information at this time to assess the relative reliability and useful life of either drive.

Whilst unlikely to find favour in HPC environments, the RamSan-20 could conceivably be useful in supporting visualisation of model outputs.

Acknowledgements

Thanks to NAS (UK) Ltd., in particular Richard Goss, for the opportunity to evaluate the RamSan-20. Also, thanks to the support team at Texas Memory Systems Inc.

22/2/2010 version 1.00 (final)

D.J. Cable, Distributed Computing Group, Computational Science and Engineering Department, STFC Daresbury Laboratory, Daresbury, Warrington, Cheshire, WA4 4AD, UK