2005 Performance / Price Sort and PennySort

Robert Ramey
April 15, 2005

Abstract: Sort Benchmark[1] is a set of world's records catalogued by Microsoft Research that evaluate the progress that computer technology has been making on transaction processing. In the various sort benchmarks, the PennySort and Performance / Price Sort were defined to test the maximum cost efficiency of sort machines. There are two divisions: Daytona for off the shelf configurations and Indy for one of-a-kind custom setups. This paper recounts our experience with our commercial program – the Postman's Sort in the PennySort/Daytona competition.

Robert Ramey Software Development

830 Cathedral Vista Lane
Santa Barbara, CA 93110
www.rrsd.com

Pennysort Benchmark: The earliest description I can find for the Pennysort contest is and article published by Jim Gray of Microsoft Research. The idea is to scale sorting benchmarks according to the cost of the hardware required. The procedure is:

1) Calculate the total cost of the hardware
2) Amortize this cost over 3 years to calculate the number of seconds one can purchase for a penny. This is one's sorting "budget"
3) Through testing, determine the largest file of 100 byte records that one can sort within the sorting "budget". The resulting number is GB sorted / penny which is the pennysort "score"

Postman's sort: The postman's sort is a commercial sorting utility that has been marketed since 1993. It evolved from an article originally published in the C User's Journal in 1992. It has been a winner in past sorting benchmark contests. As file sizes started to get larger and CPUs got faster, the comparative advantage of postman's sort started to diminish in importance while i/o time started to dominate the time. By 2000 the Postman's sort wasn't much faster than its competitors for the types of files typically sorted. In 2000, I invested significant effort in optimizing the i/o of the postman's sort. This included exploitation of asynchronous i/o on all platforms and a fair amount of tuning for windows NT (which is what I had at the time.) More importantly, I reorganized the data structure of the work file at the algorithmic level in order to greatly diminish seek times on the work file. I didn't really have access to the equipment I felt was necessary to really compete in the contest. I also had some bugs to work out regarding access to files greater than 2 GB on windows platforms. Given the new machines that are starting to become available, I think the postman's sort will once again have significant comparative advantage.

Hardware/OS Configuration: I went to my nearest computer store - Channel Data Systems (www.channeldata.com) which is located about 1/4 mile from my house. I asked them to configure for me their most popular model. This was a D865GBFL Intel mother board with 512 Memory, 3.2 GHz Pentium 4 processor, one 80 GB SATA drive and one parallel 80GB ATA drive. The operating system was Windows XP/Professional

I ran my own tests for an afternoon testing different buffer sizes, assignments of disks to input and output and work files etc. I was surprised to discover that placing the work files on a separate drive made little difference in the sorting time. I presume this was due in large part to the fact that my new i/o significantly reduced number of seeks and that the newer drives have on board cache memory. In fact, the time to sort a give file hardly varied at all for all combinations. We ran Sandra (http://www.sisoftware.net) which showed that the SATA drive was capable of throughput of 48MB/sec.

I asked channel data to reconfigure the system with two 80 GB SATA drives configured as RAID/0 (excluding a 5GB partition for windows). I came back the next day and re-ran Sandra which confirmed almost a doubling of sequential i/o throughput. I re-ran my sorting benchmarks and found that the total elapsed times were reduced by about 1/3. . Attached (Quote.pdf) is the retail quote for this final configuration. Total system cost was $950.28.

The two SATA drives were configured with two partitions each - 5GB and 75GB. The 75 GB were configured as RAID/0 to appear as one volume. Assuming 3 year system lifetime - 94,608,000 seconds and dividing by 95028 pennies we get a time budget of 996 seconds. That is 16 min 36 seconds.

Results: We generated various sized files using the standard SortGen[2] program (Compiled in Linux, the SortGen generates exactly 100-byte records.), and sorted them to find the maximum amount that could be sorted under the time budget. The results are recorded in the following table.

Product	Time Budget	Best Time	Sorted GB	Category
Postman's Sort	996	16 min 19 sec	16.3	Daytona

Subsequent Tests: These results were not unsatisfactory, but we wanted to experiment with more drives. I also wanted to verify buffer sizes and other tuning parameters in psort which hadn't been adjusted in several years. These tests resulted in changing some key psort performance parameters. Unfortunately, I wasn't able to run the official tests on the original configuration.

Historical Comparison: Its illuminating to review the last 20 years of computing technology through the following table. The procedure used to calculate each line of the table is:

(1) Sort the largest file you can in a minute.
(2) Compute the system price in $ per minutes (3-year depreciation => system price divided by 1,576,800)
(3) Compute the GB/$ sorted by dividing item 1 by item 2.

Table 3: Historical Performance/Price results.						
year	MB/sec	GB/$	System	Sys price (M$)	CPUs	
1985	0.02	0.05	M6800 Bitton et al	0.03	1	Datamation
1986	0.03	0.01	Tandem Tsukerman	0.3	3	Datamation
1987	3.85	0.05	Cray YMP, Weinberger	7.0	1	Datamation
1991	14.29	0.54	IBM 3090, DFsort/Saber	2.5	1	Datamation
1990	0.31	0.15	Kitsuregawa	0.2	1	Datamation
1993	1.20	0.11	Sequent, Graefe	1.0	32	Datamation
1994	1.72	0.16	IPSC/Wisc DeWitt	1.0	32	Datamation
1994	11.11	5.25	Alpha, Nyberg	0.2	1	Datamation
1995	28.57	2.70	SGI/Ordinal, Nyberg	1.0	16	Minute/Daytona
1995	19.61	37.10	IBM, Agarwal	0.05	1	Minute/Indy
1996	100.00	15.76	NOW, Arpaci-Dusseau	0.6	32	Minute/Indy
1997	140.17	8.41	Now 95 , Arpaci-Dusseau	2.0	95	Minute/Indy
1997	86.21	6.27	SGI/Ordinal, Nyberg	1.3	14	Minute/Daytona
1998	1.74	125.00	PostmanSort	0.0013	1	Penny/Daytona
1998	1.74	144.00	NTSort	0.0012	1	Penny/Indy
1999	2.23	174.99	Postman Sort	0.0012	1	Penny/Daytona
1999	2.46	220.59	NTSort	0.0010	1	Penny/Indy
1999	3.51	314.51	HMSort	0.0010	1	Penny/Indy
1999	3.78	338.17	HMSort Post-April 1st	0.0010	1	Penny/Indy
2000	6.50	608.86	HMSort	0.0010	1	Penny/Indy
2001	6.50	608.86	HMSort	0.0010	1	Penny/Indy
2002	8.64	1165.70	DMSort	0.000672	1	Penny/Indy
2002	10.00	1079.50	THSort	0.000857	1	Penny/Daytona
2005	**19.16**	**1660.00**	**Postman's Sort**	**0.000950**	**1**	**Penny/Daytona**

1 Sort Benchmark homepage at http://research.microsoft.com/barc/SortBenchmark/
2 SortGen at http://research.microsoft.com/barc/SortBenchmark

Channel Data Systems
4141 State Street, Suite A-2
Santa Barbara, CA 93110
(805) 964 6695   (805) 964 7519
tammy@channeldata.com

CDS QUOTE

DATE	BID #
2/28/2005	2005-395

NAME / ADDRESS

Robert Ramey
830 Cathedral Vista Ln
Santa Barbara, CA 93110

REQUESTED	TERMS	DUE DATE	FOB	PROJECT/JOB	PHONE
	Due on receipt	2/28/2005		050228TR	

QTY	ITEM	DESCRIPTION	CDS COST	Total
1	~ ANTEC MID 3...	Antec Mid-Tower Case, 300 Watt Power Supply, Black	95.75	95.75T
1	~ D865GBFL	HyperThreaded P4 w/ 800Mhz FSB, 8x AGP, SATA, On-board Video, Audio, & 10/100 LAN	118.25	118.25T
1	P4 3.2GHZ, 800...	Intel Pentium 4 3.2Ghz, 800FSB Processor	282.50	282.50T
2	80 GB SATA -150	Maxtor DiamondMax Plus 9 - Hard drive - 80 GB - standard - 3.5' - SATA-150 - 7200 rpm	80.75	161.50T
1	~ 512MB DDR P...	512MB PC3200 DDR 400MHZ Memory	110.00	110.00T
1	~ 1.44MB BLACK	1.44MB Floppy Drive, Black	9.50	9.50T
1	~ WIN XP PRO ...	Microsoft Windows XP, Professional OEM	173.75	173.75T
	SALES TAX NO...	Note: This Quote Does Not Include California Sales Tax. Sales Tax Would Apply.	0.00	
			0.00	0.00

Total $951.25

*Channel Data Systems Parts and Labor Warranty: HARDWARE-Covers all hardware in house for one year (unless otherwise stated) from the date of purchase. Limited to hardware purchased from and installed by CDS. The Customer shall be responsible for transport of the machine to and from CDS for warranty repair. SOFTWARE: CDS guarantees proper software operation for the first 30 days from the date of purchase, limited to the software purchased from and installed by CDS. Any subsequent re-configuration or re-installation of software is not warranted after the first 30 days from the date of purchase. All repairs carry a 30-day warranty.**

*Quote is valid 2-weeks from date issued!

**Please ask about our extended service contracts.   Signature _____