Orange Bytes

Award winning newsmagazine of the North Orange County Computer Club

Vol 27 No 5 **May 2002** **$4.00**

NOCCC MEETINGS - May 5

9:30 a.m. *Irvine Hall*
NEW TECHNOLOGIES
Surprise speaker!

1:00 p.m. *Irvine Hall*
GENERAL MEETING
Frances Caldwell from NuWays about internet advertising

8:30 a.m

Visual Programming I . . . . Science 109
Visual Basic and Visual Basic Script for beginners

9:00 a.m.

Auto cad . . . . . Wilkinson 111
Visual Programming II . . . . Science 109
Visual C++ and Visual J++ for beginners
Windows . . . . . Science 111
Windows ME & Windows 9x related questions
BeginningLinux . . . . . Wilkinson 210
Linux on the desktop

9:30 a.m.

Java . . . . . Science 203
Computer Aided Investing . Wilkinson 221
mutual funds and other investmet options
Hardware Essentials . . . . . Science 306
system power and basic over clocking
Intermediate Linux . . . . . Wilkinson 210
Linux for programmers and system administrators

10:00 a.m.

Visual Programming III . . . . Science 109
intermediate and advanced Visual Basic

11:15 a.m.

Office Suites!, Etc. . . . . Science 111
modifying outlook's "contacts" form.
PC Q&A. . . . . Irvine Hall
speech recognition and USB problems
Quicken . . . . . Science 203
Quicken and QuickBooks
Visual Programming IV . . . . Science 109
Office program development using VB
Macintosh . . . . . Wilkinson 210
exploring OS X
Understanding O.S.s . . . . Wilkinson 111
get help with DOS, Windows 3.1/9x/ME, OS/2,...

12:00 noon

PIG SIG . . . . . Argyros Hall Cafeteria

1:00 noon

General meeting . . . . . Irvine Hall

2:30 p.m.

Geneology . . . . . Irvine Hall
discover your family history with modern tools
Hardware Essentials . . . . . Science 109
system power and basic over clocking
OS/2 . . . . . Science 203
OS/2 news and installation

http://www.noccc.org

future meetings
June 2, July 14, August 11

news magazine staff

Publication Chairman
Dave Keays. . . . . editor@noccc.org

Editor *— selects and edits the articles in the Bytes.*
Dave Keays. . . . . editor@noccc.org

Associate Editor/Production *— lays out, formats, and submits the Bytes to the printer.*
Dave Keays. . . . . editor@noccc.org

Editor/Reviews *— communicates with the vendors and the members who review products.*
Joe Mizer . . . . . Jmmizer@juno.com
Lloyd Mizer . . . . . lmizer@juno.com

Web Reporters *— searches user groups websites for articles .*
Ted Williams. . . . . TedWilliams@alum.mit.edu

Copy Editor*— proofs Bytes for mistakes.*
Dave Keays. . . . . editor@noccc.org

Circulation*— Picks up Bytes for the printer and oversees the mailing and distribution.*
Alan Pearlman . . . . . president@noccc.org

Commercial Swap Meets*— Distributes Orange Bytes and NOCCC material at swap meets.*
O.T.Stoll . . . . . ot@stollcomputers.com

Commercial Advertising*— Gets ads.*
DJ Jennings . . . . . dj97@hal-pc.org

Classified Advertising*— Obtains members' computer- related non-commercial ads.*
Jim Sanders. . . . . jsanders@wdc.net

Help Line*— Keeps a list of volunteers and the subjects they will answer questions about.*
Ted Williams. . . . . TedWilliams@alum.mit.edu

Contributing Editor
Timothy Everingham . . . . . teveringham@earthlink.net

apcug
Member of
The Association of
Personal Computer User Groups

Change of Address?

Don't miss a single issue. Also, if we have your e-mail address, we can notify you of meetings and special events. Send address or e-mail changes to Alan Pearlman at membership@noccc.org Or Box 3616, Orange, CA 92857

President's Message

Now that everyone has filed their taxes and applied for those big refunds, it's time to start planning for next year. Your membership dues and any other items you care to donate are tax deductible. So if you just happen to need a tax deduction, keep us in mind.

We have decided to replace the Internet SIG with a Genealogy SIG, since many members have asked for this SIG. We are always looking for new ways to improve the club. We will also be discontinuing the job networking SIG as no one was showing up.

Please give me a call if you want to help with our booth at the ACP Swap Meet on Sunday, May 26, 2002.

Our new membership drive is still going on. The more new members we have, the bigger the drawing prize will be. We are still having our great membership raffle in June 2002. The rules to enter this drawing are simple, just get someone to join the club and you and the new member are automatically entered. The more members we get, the BIGGER the June raffle prize. If you know of any former members of NOCCC whose membership expired before 1995, they are also eligible for the raffle when they renew. Also, please remember that when any new members join, they get 12 free raffle tickets for the regular drawing at the main meeting, and the referring member gets 6.

Our membership Raffle is for everyone who joined since JULY 2001. Our current list of eligible members is below. If we have omitted anyone from the list below, please let me know so I can make sure you're included in the June drawing. In order to participate in the raffle, you must be at the main meeting in June.

Ronald Petty
Williamson Uy
Betty Wilson
Pete Madrid
Larry Farkas
Kerry K Mullins
James C. Wagner
Willi Pfeifroth
Maurice Jacques
Roy Rosenberger
Tim Quinn
Gerry Resch
Ron Daigh
Joseph Vezeau
John Mosby
Phillip Woolston
Michael Iwaki
Joe Spellman
Kaz Adachi
Chris Quinn
Steven M. Sloan
Kathryn Kane
Chuck Varanay
Jay Rodriguez
John Forgy, Jr
Michael Gordon
Frances Caldwell
John Maierhofer
Carla Reisch

Board Nominations

July 1, 2002 through June 30, 2003

Officers:

President: Alan Pearlman
Vice-President: James C. Smith
Secretary: John Johnson
Treasurer: Elise Edgel
Past President: Terry Warren
Webmaster: Jim Peugh
Editor: Dave Keays

Board Members:

Cathy Grammer-Margolin
George Margolin
Else Olavsson
Gerry Resch
Linda Russell
Jim Sanders
Cathy Shimozono
Herb Wong

Hard Disk Drive Optimization

By Herbert Wong, Jr., NOCCC (ocug@singularitytechnology.com)

Hard disk drives have evolved into highly reliable and extremely inexpensive mass storage devices. A few simple ideas may help improve your daily computing experiences.

Size Matters

Computer science has a standard set of definitions for quantities. A thousand bytes is 1,000 bytes (ten to the third power). A kilobyte is 1,024 bytes (two to the tenth power). A million bytes is 1,000,000 (ten to the sixth power; 1,000 times 1,000) and megabyte is 1,048,576 bytes (two to the twentieth power; 1,024 times 1,024). Finally, a billion bytes is 1,000,000,000 (ten to the ninth power; 1,000 times1,000 times 1,000), at least in the United States, and a gigabyte is 1,073,741,824 bytes (two to the thirtieth power; 1,024 times 1,024 times 1,024).

Computer marketing sells you short. Advertising redefines many standard computer science terms. A kilobyte becomes 1,000 bytes, a megabyte becomes 1,000,000 bytes, and a gigabyte becomes 1,000,000,000 bytes.

In other words, 1,000 equals 1,024, 1,000,000 equals 1,048,576, and 1,000,000 equals 1,073,741,824. Of course, this works in the favor of the manufacturer and shortchanges you. An advertised “twenty gigabyte” drive should be almost one and a half billion bytes larger.

To be clear and accurate, distinguish between one thousand bytes and one kilobyte, one million bytes and one megabyte; and one billion bytes and one gigabyte whenever possible.

Partitions

The foundation of a hard disk drive’s file system is a partition. The framework is the formatting of logical drives (C:, D:, E:, etc.). There are two types of partitions under Windows, primary and extended.

A primary partition can contain the boot drive (which can load the operating system at startup) after formatting (with the FORMAT utility). An extended partition can contain one or more logical drives that cannot be boot drives. The partitions and logical drives are created using the FDISK utility. The logical drives must be formatted with the FORMAT utility before they can contain data.

Under older versions of Windows, a physical hard disk drive can contain a single primary partition, a single extended partition, or both a primary partition and an extended partition. Technically, those are the only things that Windows can recognize. The rest of the world is decades more advanced.

Under older versions of Windows (and DOS), logical drive letters are assigned (at least initially during OS installation) to hard disk drives according to a few simple rules. The physical hard drives are inspected in sequence (port 0’s master, port 0’s slave, then port 1’s master, port 1’s slave, etc.) and logical drive letters are first assigned to each primary partition that is found.

Next, the first physical hard drive is inspected again and logical drive letters are assigned to any logical drive contained in an extended partition (if it exists). Each of the remaining physical hard disk drives is inspected in sequence (port 0’s master, port 0’s slave, then port 1’s master, port 1’s slave, etc.).

I suggest labeling logical drives with a naming convention to facilitate having a collection of drives in a computer or network. I abbreviate the manufacturer’s name, drive capacity, number of the drive (first, second, third, etc.), and logical drive number (a primary partition’s logical drive number is given 0 and an extended partition’s logical drive starts at 1).

Suppose a computer contains a Seagate 80 gigabyte drive and a Maxtor 40 gigabyte drive. Both contain a single primary partition and an extended partition with two logical drives. The drive letters C: through H: would go to: SEA80GB#1P0, MX40GB#1P0, SEA80GB#1P1, SEA80GB#1P2, MX40GB#1P1, and MX40GB#1P2, respectively.

Suppose the same computer contains a Seagate 80 gigabyte drive and a Maxtor 40 gigabyte drive. Now the Seagate contains a single primary partition and an extended partition with two logical drives; and the Maxtor contains only an extended partition with three logical drives. The drive letters C: through H: would go to: SEA80GB#1P0, SEA80GB#1P1, SEA80GB#1P2, MX40GB#1P0, MX40GB#1P1, and MX40GB#1P2, respectively.

This (or similar) naming convention helps to easily identify where data resides. There has been many an occasion where the label told me that I was in the wrong folder on the wrong drive.

Change Drive Letter And Path

The first thing that I do after installing Windows is to change the drive letter of all the CD/DVD-ROM (read only memory) and CD-R/RW drives. I change the first CD-ROM to drive R:, the DVD to drive V: (or to drive R: if there is no CD-ROM), and the CD-R/RW to drive W:. Now I can stick in an optical disc into every machine and know what drive letter it is.

If you don't change the drive letters as above, bad things can happen. If you add another hard disk drive, that hard disk drive's letters will appear after the CD-ROM's.

ScanDisk

Microsoft's ScanDisk is a lightweight utility that is best known for automatically running after an improper exit from Windows. Most common causes are pressing the hardware reset button due to a frozen system, a "blue screen of death" crash, or worse.

A sector is the basic storage unit of 512 bytes. A cluster is a series of logical sectors whose total size is predetermined by the operating system's file system. For example, FAT32 (ideally) uses 4-kilobyte clusters on partitions that are less than 8-gigabytes, 8-kilobyte clusters on partitions that are less than 16-gigabytes, 16-kilobyte clusters on partitions that are less than 32-gigabytes, and 32-kilobyte clusters on partitions that are greater than 32-gigabytes, The hard disk drive may not physically store data in this manner, but may perform translations as needed.

Windows uses a FAT (file allocation table) as a filename directory. Every file has an entry in the FAT that contains a pointer to the first cluster of the file itself. If the current cluster is not large enough to contain the whole file, a pointer within the cluster will point to the next cluster in the chain. It is possible for a file to be composed of millions of clusters.

ScanDisk's primary functionality is to test the chain of clusters that comprise each file. A file that contains a pointer to a cluster owned by another file is a cross-linked file. An allocated cluster that is not contained in a valid file is a lost cluster. A FAT entry may be invalid if it does not point to a file.

Run ScanDisk after occasions when Windows was not able to complete a "Shutdown" instruction by itself.

Defragmentation

Files are created and deleted in the normal course of operation. Some files were persistent and some were temporary (for use during the current session or for some short operation).

Imagine a wall with many random groups of bricks (deleted file clusters) removed. As a new large batch of bricks (the clusters in a single new file) comes in, they fill in the random openings as possible. These new bricks now appear in groups on various different rows.

In a hard disk drive, bricks are clusters and the many rows of bricks are tracks. The read/write heads take a long time (for computers) to step from track to track. A file read/write operation is fastest if the file is

in consecutive clusters and in adjacent tracks. If the clusters in a single file are scattered randomly about the disk, reading/writing will take a large amount of time since stepping is so slow.

By periodically running the defragmentation utility, the gaps in the fragmented files are removed. Disk performance can be noticeably improved. Do not reset or turn off the power to the computer until you have terminated the defragmentation. Nasty things can happen if you disrupt a defragmentation.

Microsoft claims to enhance load times on some of their operating systems by moving boot files to the outer tracks of the disk (since they have the fastest data transfer rate). In addition, they claim that some files (such as .DLL) are intentionally split up so that important sections of the file are moved to the outer tracks for speed. All of this rearranging of files is "done in the background." This explains some of the mysterious and otherwise inexplicable hard disk drive activity that occurs on some computers.

Many people maintain a small logical drive as a place to hold file before burning them to a CD-R/RW. The files are subsequently deleted. Rumor has it that some file systems do not efficiently clean up after such system. It is claimed that the logical drive must be formatted (or defragmented) again to be sure that the drive is truly defragmented.

Two Heads Are Better Than One

A system that has two physical hard disk drives can be slightly optimized by changing a few system defaults. Under Windows, the boot drive (drive C:) is the default drive for temporary files (ex. - C:\temp) and system memory-page swap-files (ex. – pagefile.sys). Writing these files can be significantly faster if these defaults are changed to another physical hard disk drive.

The explanation is quite simple. As the first physical hard disk drive's read/write heads are locating specific files, the second physical hard disk drive's read/write heads can quickly and efficiently move to other files without the interference that would result if only a single drive was serving the same purpose.

Cable select

The low cost of hard disk drives allows even basic systems to have two devices. Since there is not a Microsoft operating system that fills everyone's needs, many people want to run two different operating systems on one computer.

The lowest hardware cost technique is to use two existing drives and select the boot operation system through a minor change in the BIOS. This circumvents the need to purchase additional software or drive bays with removable drive chassis.

Follow these procedures. Remove every hard disk drive, except for the target hard disk drive. Set the jumpers on the hard disk drive to master. Create the primary and extended partitions as desired. Install your operating system of choice on the primary partition and format the logical drives as needed. Test this configuration to your satisfaction.

Repeat this procedure for the second target drive. You can install any operating system you choose on the second drive.

Change the jumpers on both hard disk drives to the cable select setting. The jumper settings are printed on most newer hard disk drives. Install both of these hard disk drives on the same ribbon cable on the primary port (usually labeled "port 0") of the motherboard.

You can now change the hard disk drive to boot from by selecting the appropriate drive from the BIOS (CMOS setup). In the AwardBIOS Setup Utility, select the "Boot" menu. Under the "IDE Hard Disk" item, select the desired hard disk drive as the onscreen instructions indicate. For example, I have a Seagate drive indicated by "[ST380021A]" next to the "IDE Hard Disk" text. The secondary drive might be indicated by something like "[MX548075A]." To boot from this device, change the setting to "IDE Hard Disk [MX548075A]."

Using the cable select feature of the ATA hard disk drive to change boot devices may not be the fastest way to change between operating systems. However, if you want a very inexpensive hardware solution, it does not get any better than this.

I left out some minor installation details. If you have further questions and do not think that you can get this to work, then you probably cannot. Any questions about using this cable select technique will be answered with instructions to read this article. There are too many variables involved in this type of system configuration to attempt to provide useful answers. However, any further information about this little documented feature will be used to update this page.

Performance

A hard disk drive transfers data faster from the outer tracks than the inner tracks. The outer tracks are longer and contain more data. As a result, for each revolution, outer tracks transfer more data. The outer tracks are filled first.

Sooner or later, you'll fill up your hard disk drive. It seems to have gotten slower and slower. It has! The inner tracks on a hard disk drive may transfer data about half as fast as the outer tracks!

One way to minimize this effect is to create several logical drives (on a physical disk) and use the second one (ex. – drive D:). Now you can put all of the other junk files on the inner tracks of drive E:, F:, G:, etc. Drive D: remains towards the outer tracks and your important data is faster.

When you buy a hard disk drive, look for the sustained transfer rate (data transfer rate). Of any single specification, this one will give you the best indication of a hard disk drives performance. Average access time is also important for random access applications (such as databases).

Safety First

Do not mix the data that you create and cherish with installed program files! Create several logical drives and store the valuable things on a separate drive. Now it is easy to identify all of your important data. It is anything on drive D:!

Going one-step further, I create a subdirectory (folder) called C:\Herbert Wong, Jr. (you might call your folder something else) that contains other folders of the supremely important data (.\telephone numbers, .\finances, etc.). Another folder (D:\ASUSP3V4X-SystemInstallationFiles) might contain every driver and document file needed to reinstall that computer.

Now I know where to find everything easily. I can easily determine if it will fit on a CD-R. And, on those many occasions over the years, when Windows self destructs, I can confidently format and reinstall on drive C: without fear of losing anything of much importance.

Relocate your MY DOCUMENTS and Favorites folders to another drive (ex. – E:\MY DOCUMENTS). Then you will not even lose those files during an emergency.

The best way to back up a hard disk drive is to use another hard disk drive as the storage medium. The price of tape backup drives and blank tapes is much too great to be effective in a home environment. CD-R/RW and DVD-R/RW drives and media are also too expensive.

Conclusion

Hard disk drives are larger, less expensive, and more reliable than ever before. With a little planning and a lot of maintenance, a new hard disk drive will well for years.

This article first appeared in the North Orange County Computer Club's (http://www.noccc.org/) Orange Bytes newsmagazine for May 2002. You can find the latest version of this article at http://www.singularitytechnology.com/articles/HDDOptimization.html. You can contact me, Herbert Wong, Jr. at NOCCCArticles@SingularityTechnology.com.

A USB story

by Jim Sanders, NOCCC (jSanders@ligasmicro.com)

There was a local sale this last month on the Epson 1640SU flatbed scanner. This was a factory refurbished unit, bought for $50, and at that price a real steal. The specifications for this model are really quite good and one of the items that caught my eye, was the fact that it has both a USB and A SCSI interface. Umax has basically stopped supporting my trusty old 1200S model. So I thought I would upgrade to a newer, fancier machine. The scanner worked just fine when I hooked it up to my SV-24 computer. A friend of mine was impressed with what the scanner could do and was interested in upgrading her scanner if the Epson would work on her K6-3/350 computer. No problem I thought, both of them have a USB port. In fact, her machine had an older, and slow, flatbed USB scanner installed and working. That was not to be the case however, when hooked up to her Tyan S1590s motherboard's USB port, based on the VIA MVP3 chipset, it did not work correctly. The Tyan S1590S motherboard is a couple years old, and because of that I thought perhaps the USB controller design might have a problem with the newer scanner. I thought a simple solution to that problem would be to install one of the PCI bus to USB upgrade cards. I happened to have of one of those in stock. So I went into the bios and disabled the on board USB controller, and installed the new card. When Windows ME re-booted, it found the new hardware and installed the the CMD USB card driver. The scanner worked better but still had problems, especially if you were trying to do a high resolution scan.

The next day I took the scanner into my class at Santa Ana College to try on my class demonstration machine that has the same motherboard. It failed the same way on the second motherboard of the same type. After doing a bit of research with Google on the Internet, I found a number of interesting comments about USB failures. Several comments claimed that the VIA Corp. was known to have produced motherboard chipsets that had a poorly designed USB controller. Additionally, I discovered that there was considerable discussion about different brands of upgrade cards and the different chips that they used. Naturally, the upgrade card that I had in stock was one that used the CMD brand of chip which was alleged to have known problems. One website had a list of card brands and what chip that they were built with. That list indicated that the chip of choice was made by Lucent, with Opti as the second place choice, and a recommendation to avoid those that used the CMD chip.

The next class day had two things occur that were interesting. First, when the computer and scanner were initially turned on in the cool of the morning, it worked flawlessly for about five minutes. Then it started failing. By experimenting with the bus speed settings and the CPU multiplier settings, we discovered that the machine would continue to work OK if the bus speed setting was set to 66 Mhz instead of 100Mhz. Alternately, if the bus speed was at 100, but the CPU multiplier was set to 2 X, the scanner would almost work OK. I also ran a test where I waited till the computer warmed up enough that the scanner failed, and then applied freeze spray to the chipset. When the chipset was cool again, the scanner started working properly. I have seen symptoms like these that were caused by a design flaw that did not allow for adequate timing margins. After running these experiments, I could not help but wonder if perhaps there was some validity to the comments on the Internet about the quality of the early VIA USB Controller design. Second, a student brought in his PCI to USB upgrade card made by the ADS Corporation. This card had one of the Lucent brand of chips. It worked flawlessly.

I thought I would solve the USB problems by going to a local source and purchasing two of the ADS brand of the USB upgrade card. I was a bit chagrined (actually that is far too mild of a word) when I opened up the box to find that they had changed the brand of the USB controller chip. The second choice Opti controller was now on the board. On the plus side, it worked fine for about 20 minutes, then it stopped working at all. After thinking a few choice words, I

remembered that I did not clean the PCI bus connection, and tried reseating the board several times. After doing that, it worked fine again.

I think there are several morals to this story. One is that if you have an older VIA chipset and you buy a newer USB device which fails when you attach it to your computer, it may be your computer and not the new device. Second, just because a chip purports to meet an industry standard, and is sold with that implied functionality, it ain't necessarily so. Third, just because you purchase the same brand of widget, that doesn't mean that you will get exactly the same product. The chips can change, the BIOS can change, the board revision level can change, etc. Caveat emptor, ask questions, do research, and read the bloody fine print.

DEADLINE

All articles and reports for the BYTES must be sent by the Wednesday after the meeting to **EDITOR@NOCCC.ORG**.

Dave Keays, BYTES editor

NOCCC Officers

The area code is 714 unless noted otherwise.

President

Alan Pearlman . . . . . PearlmanA@aol.com

Vice President

James C. Smith . . . . . Jamescsmith@bigfoot.com

Secretary

John Johnson . . . . . Jrosjohnson@juno.com

Treasurer

Elise Edgell. . . . . eliseme@aol.com

Directors

Cathy Grammer-Margolin . . . . . inventor00@pobox.com

Dave Keays . . . . . dave@rdksoftware.com

George Margolin . . . . . @pobox.com

Gerry Resch . . . . . gresch@hotmail.com

Jim Sanders . . . . . jsanders@ligasmicro.com

Herb Wong . . . . . ocug@singularitytechnology.com

Past President

Terry Warren . . . . . twarren@alumni.caltech.edu

Webmaster

Jim Peugh . . . . . Webmaster@citivu.com

Business solicitations/ lecture series

George Margolin . . . . . inventor@pobox.com

Consignment Table

Cathy Shimozono . . . . . shimca@charter.net

Classified Advertising

D.J. Jennings . . . . . dj97@hal-pc.org

Helpline

Ted Williams . . . . . tedwilliams@alum.mit.edu

Membership Database

Alan Pearlman . . . . . PearlmanA@aol.com

Programs

George Margolin . . . . . inventor@pobox.com

SIGs

Dave Keays . . . . . dave@rdksoftware.com

University Liaison

Terry Warren . . . . . twarren@alumni.caltech.edu

The Data Warehouse Toolkit

Practical Techniques for Building Dimensional Data Warehouses

by Ralph Kimball
Reviewed by Ron Rose, NOCCC (bf799@lafn.org)

A data warehouse is a copy of transaction data specifically structured for query and analysis. It should reflect the business plan for answering the necessary questions.Generally a structured query language (SQL) is used in the data queries. Even though SQL can do rather complicated queries, that is not easy or the most stable way to get the answers.The structure described in this book is a dimensional model. It is also call the star join schema. The central tables are fact tables consisting of numerical data that can be added. Dimension tables that relate to the fact tables contain dimensions only, no data.

The relation between the data and dimensions is established by the use of index keys. Some examples found in a dimension table would be the dimensional index keys: timeKey, dayOfWeek, month, year, holidayFlag; productKey, description, brand, category; storeKey, storeName, address, floorPlanType. The sales fact table would then have: timeKey, productKey, storeKey, dollarsSold, unitsSold, dollarsCost. Note that the fact table contains only keys and additive quantities.

The "grain" in the example is the day. A much finer grain of time would be the hour. One wants to construct with the largest grain and still be able to find the needed answers.

A standard SQL query would be similar to the following excluding the [comments]: SELECT p.brand, SUM(f.dollars), SUM(f.units) [select-list] FROM salesfact f, product p, time t [f,p,t are used to represent full table names (aliases)] WHERE f.productKey = p.productKey [join constraint] AND f.timeKey = t.timeKey [join constraint] AND t.month = 'MAY 2000' [dimension value] GROUP BY p.brand [group by clause] ORDER BY p.brand [order by clause].

Throughout the text, key words are highlighted, and they are also found and defined in appendix C.

A number of data warehouses of different businesses are described each with their own structure and variation. Some of them are the grocery store chain, the distribution warehouse, and the shipments systems. Their dimensions are part of the value chain. The demand side and the supply side are explained and organized.

Throughout the book, he gives design principles for a dimensional data warehouse which are collected by chapter and listed in appendix A. At the conclusion of the design description of each type of business, such as insurance or voyage, he shows how to estimate the disk space required to store the data base.

In chapter 12, he gives the details in building a data warehouse. Starting with the nine decision points, on to interviewing the end users and database administrators, assembling the team, filling in the details of the tables, and how to choose hardware and software. He has a chapter using aggregates, which are the precalculated and prestored summaries that are in the data warehouse and are used to improve query performance. Included with the book is a CD-ROM containing "Star Tracker" a query tool made to perform the needed functions to analyze data in the star schema data bases.

Star Tracker explicitly supports Red Brick, Oracle, Sybase, and Informix using Microsoft's Open Database Connectivity (ODBC) server software. In the PC, MS Access is the example used in the book. The installation and how to use is given in appendix D. It tells how to assemble complex reports, and do things such as, drill-down and drill-across plus many more.

If you have a business that involves thousands of items/combinations and need to track which gives greater and lesser value, then this book describes how to make data bases that can give the answers. The book was written for owners, managers, and implementers of a data warehouse. Even though it may appear to be highly technical, it is a fairly easy read.

If you are into the web aspects of the above, check out: wiley.com, search: kimball for a recent book involving the web.

The book was copyrighted in 1996 by John Wiley & Sons, Inc. and published by the same.

ISBN 0-471-15337-0.
Pages 388.
CD-ROM included.
The cost is $55.

Please make sure we have your correct email address and phone number.

We will give it out only with your permission.

No one from NOCCC other than the President or Treasurer is authorized to call a member at home for club business.

You can update your membership at any regular meeting or by emailing membership@noccc.org

Object Oriented Data Warehouse Design:

Building a Star Schema

by William Giovinazzo
Reviewed by Ron Rose, NOCCC

"In order for an organization not to just survive but to thrive, a strategic plan must guide its actions. The business strategist defines the strategy in the context of the environment in which the organization competes."

Where is a good source for the organized information the strategist needs? Enter the data warehouse.

"The data warehouse is the heart of BI [business intelligence]. It is more than just an archive of operational data. W. H. Inmon defines the data warehouse as 'a subject oriented, integrated, nonvolatile, time variant collection of data in support of management's decisions.' The data warehouse derives data from the operational environment. As the data is brought into the warehouse, it is transformed into a single cohesive image. From this image the business strategist through a decision support tool is able to analyze behaviors, giving insight into current trends. This enables the strategist to project future conditions."

This book is about data warehouse design. The design is object oriented using the development spiral. This spiral starts with definition, followed by analysis, design, development, implementation, maintenance, evaluation, and back through definition adding and refining on each go around.

The data has data about itself, its context, relations to other data, in other words 'metadata'. Two types of metadata are static (structural) and dynamic (operational). The static consists of name, format (data type, picture), relation, domain, and business rule (rules on the data element). The dynamic consists of data quality, statistical, usage and status."The question facing the data warehouse architect is how to discover and document this metadata in a way that is easily understood. The answer is 'objects'."

"What is an object? It is simple enough to describe an object as a thing." Objects can be tangible or intangible like concepts. "... the analyst can begin to view the organization in terms of objects." Objects can be categorized / grouped into classes. A classification of classes would be superclasses, such as, cars and people. In the next section the author diagramed the superclass as a rectangle with rounded corners (people) branching to two rectangles (male, female) classes. The other superclass diagramed is cars branching to sports car, practical car, and economy car classes. The people and cars are connected. "People drive cars", and "cars transport people". An instance of the people class is the individual person, such as, "Lisa Miller, Jed Perkins". "The object model makes clear the attributes and behaviors of objects as well as the relationships between objects."

In chapter 3 is a case study that starts with the definition phase. The example business does brokerage, property management, appraisal, reality advice, and investments.

What is their mission statement?

Analysis phase consists of the interview process: the project sponsor, the management, the users, the information technology staff, and asking the right questions.

The analysis model shown consists of arrow lines showing the activity between the circles (the objects).

In the design phase, superclass, class, subclass, and aggregation (summed objects) are identified. These are connected by relations of the type, exactly-one, one-to-many, zero-to-one, and zero-to-many. A layout is made using the above. The objects are the tables and the attributes are the fields or columns which have metadata consisting of field name, format (data type), null option (is null allowed), comments (descriptions) and system of record (where from).

Multidimensions: time is generally a given. In an auto sales example: time=365, products=200 and dealerships=100 cc 3 dimensions, a cube of 7,300,000 cells already. Time, products, dealerships and customers cc 2 cubes. Time, products, dealerships, customers and payment cc 4 cubes, the number of cells is getting out of control.

Fortunately in the star schema data base, many fewer cells (fields) require data. The non-normalized dimension tables, which have mostly descriptions and primary keys, would be 'time' with 365 rows per year, 'products' with 200 rows, 'dealerships' with 100 rows and 'customers' being continually added. The rows in these dimension tables have few changes (updates).

The fact table 'payment' contains only the above dimensional foreign (secondary) keys and payment columns. The payment column rows are addible which is why it is called a fact table. Dimension tables have nothing that is addible. Additional subjects covered include shadow dimensional tables (non-analytical), snowflake dimensional tables (normalized), degenerate dimensions and facts.

Implementation considerations of parallel processing, symmetric multiprocessing, bitmapped indexing, summation tables, and web-enabled data warehousing showing SQL (Structured Query Language) statements conversion to use summary tables instead of using raw data, are some of the topics covered. The appendixes cover a spatially enabled data warehouse, extraction, transformation and loading of data, metadata standards, and conclude with conventional wisdom, tips, hints, and general advice.

Publisher: Prentice HallI
SBN 0-13-085081-0
QA76.9.D37 G56 2000
Price: $49

CONSIGNMENT TABLE

1. The consignment table is for members-only. Only current members can place items for sale, but non-members are welcome to purchase items from the table.

2. The consignment table operates on a 90%/10% basis - with the owner getting 90%.

3. Fill out tag on each item! It must contain Name, NOCCC#, item name, short description and asking price.

4. Also fill out this USER LIST with Name. Address, Phone#, and a complete list of items and their asking prices.

5. All items and/or money may be picked up by owner at any time — but MUST be picked up no later than 2 PM on day of sale.

6. Any items and/or money not picked up by above stated time will become property of NOCCC and will be subject to disposal at the club's discretion.

7. NOCCC assumes NO RESPONSIBILITY for ANY items bought or sold. Each item is placed and sold on an AS-IS BASIS.

8. We accept only cash or checks for payment.

Macintosh SIG

by John Willner,NOCCC (Willner@earthlink.net)

After setting up the projector and lap top computer, several questions about OS X were asked. One was about whether the apple icon at upper left of screen could be used to denote whether or not the software application was carbonized or written only for OS 9 and earlier. It can, but a deeper question is, "Can one tell which type of program one has before opening?" One of the reasons is that it takes time for Classic mode to come up, and one might be better off from a performance standpoint by booting in OS 9.2 anyway. If programs are stored correctly in the hierarchical folders of OS X, that may be a good clue. However, we really did not resolve the question. Many of us are still in the mode of discovering this operating system's capabilities and ways to get around.

An old copy of "Spring Cleaning" , version 3.5 was installed just to see how it performed under OS X. It triggered the Classic Mode startup session, but ran flawlessly. Claire Lemire gave a very good presentation on the capabilities of "Spring Cleaning". This is a great program that you can love to hate if not careful. It can get one into trouble by deleting the wrong duplicate files or eliminating aliases that really are needed by programs that share resources. However, it has many useful capabilities, including the ability to find files of the same type, fix aliases, search for duplicate files, find empty folders, locate FAT binary files wherein the 68K versions are no longer needed, check for damaged files, and remove unwanted fonts or help data. During her presentation cache files were emptied on the fly.

We also took a look at iClean, and discussed ways to eliminate unwanted cookies. This is another one of those useful free programs. While hard drive memory and RAM tend to be bigger than ever, it is amazing how fast file space is constantly being taken by unwanted entries. On the laptop computer especially, the hard drive capacity is not so large as to allow one to ignore this problem.

A new visitor who is considering a Macintosh for his wife was concerned about networking to his other computers. Among the many suggestions was the use of the Apple Store brain trust. We have two Apple Stores in Orange County (Fashion Island and South Coast Plaza), and both have Brain Counters at the right rear of each store. Both have been tested several times over the past several months with outstanding results. These people have enormous knowledge and T-1 lines to download the latest updates.

Next month we will continue exploring OS X capabilities. If Apple sends CDs with information of interest, those files will be displayed.

NOTE:

You need a password to get a copy of the **current** PDF version of ORANGE BYTES.

The passwords will be emailed 2 weeks before each meeting.

Office Suites

By John Heenan (jc_heenan@a-wares.com)

In a prior session of the SIG, we used the Wizard to create the skeleton of an Order Entry application. The application is to be customized to our own needs. Part of that customization is to use the Contacts in MS Outlook as a replacement for the Customer table. The impact of this is that references to the Customer table will have to be changed throughout the application. For example, the original Customers table uses a label "Company Name", where as the Contacts from Outlook uses "Company". All forms, queries and reports that use the Customer-Contacts table would require changes to references to the company name. This can be a tedious task. We will see the impact as we reconstruct Our Order Entry application.

To use the Contacts in Outlook, we have a choice of either an importing or a linking method to create the table. The first difference we see in the two methods is that the lists of field names are not the same. An exported MDB (Access) file from Outlook contains the Home address of the individual as well as his business address. So, each of the two lists need to be reviewed to see which of the two methods of input will provide the data necessary for our application.

The Link Method has an advantage over the import method by allowing one to update the Outlook Contacts in real-time. That is, any changes made to a Contact record as an item in a table are also made in Outlook at the same time, and vice-versa. The disadvantage to this method is the time for Access to open the table. It must go through a conversion from Outlook to Access file formats.

With either method of input, i.e. Import or Link, the data base has no primary key, requirement for major tables in Access. We might try using the Phone field, because most likely it will be unique to each record in the Contacts Table. On the other hand, if the Link method is used, we discover that one of the data fields is named "Account". This field can be used to assign each Customer-Contact a unique identification. The problem, we discover while returning to Outlook, is that the contact form shows no field named Account. However, if we open a contact and select the "All Fields' tab, and select from Miscellaneous Fields, there at the top of the list is "Account".

It would be more convenient to have this field under the General tab, so let's modify the outlook Contacts form. We will continue this task next month.

COMPUTER AIDED INVESTING

By Aurora Singer, NOCCC
(ForInterpreting@aol.com)

In order to research data on mutual funds you may use MetaStock and Morningstar. Morningstar, for example, provides a unique rating system that takes into account both profitability and risk. You may use the ratings and choose those funds that have the highest four or five star. Or obtain more details on a fund that looks attractive from a charting standpoint. That way you have both the fundamental and technical factors together.

The New York Stock Exchange (NYSE) Advance-Decline (AD) is used to help determine the direction of the stock market. So far, according to Bob Krishfield the AD line is flat, and there is no indication whether the trend will be up or down.

Chip from Chart Watchers, states that Tech stocks are supposed to lead the way out of recession, but their recent weakness, makes him believe that the market haven't reach a bear market bottom yet.

PC Q&A

by Jim Sanders, NOCCC (jSanders@ligasmicro.com)

We opened this months Q&A SIG with some interesting questions. One question dealt with the problem of upgrading a computer system and not having the Win modem work after upgrading the system. As most of you know I am not a big fan of Win modems. I prefer what I think of as a "real" modem, what is now referred to as the controller type of modem as opposed to the controller-less modem. In this instance, one of the last symptoms mentioned was that the operating system had changed from Win9X to Win2K. The Win9X software driver was what was installed for Win2K. No attempt was made to check the web site for a updated driver that was specifically for Win2K. It is a fairly well known fact that a lot of the older Win9x drivers do not work properly with Win2K. My guess is that if he can find a Win2k driver that his problem will be fixed.

The primary topic of discussion for this SIG however was the IBM ViaVoice for Windows release 9 voice recognition software. I have, over the last several years, tried several different versions of the VIAVoice Software. Each, at the time, seemed pretty impressive and it has gotten steadily better. Not only has the software improved in terms of its ability to run the algorithms that processed the sounds from the microphone, but the horsepower of the computer has increased greatly, which makes it much easier to run the rather complex software. The manual makes note of the fact that the computer is at a severe disadvantage when it comes to understanding speech, compared to human beings. At the very least, people can fill in the missing speech information they did not hear correctly based on the context of the subject that is under discussion. In addition to that, you can often get visual clues based on body language and facial expressions. The computer is limited to simply hearing the sound information and trying to figure out what it was that was said based on the proper volume of the speech, the correctness of the pronunciation of all the words, and to a minor extent the context and grammar. For example, the program knows that you go "to" the store, not "two " the store.

As with all speech recognition software, in order to use it you must undergo a training session first. The training session consisted of reading two or more short stories. While the software listens to you reading the story and tries to match the words that it knows you are reading from the screen with the sounds that it is hearing with the microphone. The directions suggest that you do all four of the of stories that are available, and that it will take about an hour to do this. Being in a bit of a rush to get ready for the SIG, I only finished two of them. The manual also warns you that a different environment can affect the recognition ability of the software. For instance, the reverberation that is present in the auditorium were we hold the SIG, can significantly reduce the accuracy of the recognition of the program. As I did not have time to go through a separate training session that I could select for the auditorium, I was reluctant to try the program in that setting. I did give it a try though, and for the most part I thought it did pretty good. The audience did get a few laughs when some interesting or funny lines showed up on the screen.

The program has an extensive repertoire of a voice commands, both for dictation purposes and for control purposes. Like any large complex program the learning curve to become an efficient user of the software is significant. On the other hand, there is no question that I can speak a great deal faster than I can type. When I finish the dictation to SpeakPad and transfer it to the word processor, I get a second check of the grammar. Some of the dictation errors that are difficult to catch are the small things like the missing S on the end of a word that you wanted to be plural. One of the other common errors, is the substitution of a word that sounds very similar to the word spoke, but in fact is totally different. At times, this can result in some very humorous sentences. There are, however, a lot of situations where that type of humor would not be

appreciated. It highly recommended that you proof read your dictation.

The box suggests that a 266 MHz Pentium with 64 Meg of RAM is the minimum size computer that this software should be run on. The minimum size computer may run the program OK, but as a general rule more horsepower never hurts. My demonstration machine is approximately four times the required minimum and works noticeably better than the slower machine. Next month I will talk a bit more about speech recognition and on some of the USB problems that I encountered this week.

Current SIGS

SIG	Time	Building	Leader	E-mail
Autocad	9:00	Wilkinson 130	Joe Mizer	Jmmizer@Juno.com
Hardware Essentials	9:30	Science 306	Herbert Wong	ocug@singularitytechnology.com
Hardware Essentials	2:30	Science 109	Herbert Wong	ocug@singularitytechnology.com
Internet	2:30	Irvine Hall	James C. Smith	Jamescsmith@bigfoot.com
Java	9:30	Science 203	Terry Warren	Twarren@Alumni.caltech.edu
Beginning Linux	9:00	Wilkinson 210	Bob Ray	bobcray@Pacbell.net
Intermediate Linux	9:30	Wilkinson 210	Jim Holder	caholder@surfbest.net
Macintosh	11:15	Wilkinson 210	John Willner	willner@earthlink.net
New Technologies	9:30	Irvine Hall	George Margolin	Inventor@Pobox.com
Office Suites	11:15	Science 111	John Heenan	Jc_Heenan@Csi.com
Os/2	2:30	Science 203	Terry Warren	Twarren@Alumni.caltech.edu
PC Q&A	11:15	Irvine Hall	Jim Sanders	jSanders@ligasmicro.com
QuickBooks	11:15	Science 203	Linda Russell	goodnewsent@compuserve.com
CAI	9:30	Wilkinson 221	Bob Krishfield	bobkrish@socal.rr.com
Visual Programming I	8:30	Science 109	Anson Chapman	aeccrcss@Hotmail.com
Visual Programming II	9:00	Science 109	Anson Chapman	aeccrcss@Hotmail.com
Visual Programming III	10:00	Science 109	Anson Chapman	aeccrcss@Hotmail.com
Visual Programming IV	11:15	Science 109	Anson Chapman	aeccrcss@Hotmail.com
Windows	9:00	Science 111	James C. Smith	Jamescsmith@bigfoot.com
Understanding O.S.s	11:15	Wilkinson 111	Charlie Moore	Mooreca@aol.com

SIG ALERT ! ! !

New information about special interest groups at NOCCC

A new SIG on geneology is being formed and is replacing the internet SIG at 2:30 in Irvine Hall.

Members have asked for a group on Gaming. If you can lead one of them, call or email Dave Keays.

Send the meeting reports to EDITOR@NOCCC.ORG by Sunday after the previous meeting.

Publication information

Orange Bytes is published monthly by:
North Orange County Computer Club,
PO Box 3616, Orange, CA 92857.
Phone 714-998-8080

Commercial Display Advertisements
Are Welcome
Contact - Editor@noccc.org

RATES

One page vertical	7”x9.5”	$200
Half-page horizontal	7”x5 1/8	$110
Half-page vertical	3.5”x9.5”	$110
Quarter-page	3.5”x5 1/8”	$60
Business card size	3.5”x2 1/8”	$20
Center spread		$400

Deadline: Tenth of the month

If you are interested in advertising in the publications of multiple User Groups Throughout the region or nationally, You can learn more by contacting the User Group National Ad Program Hal-PC Advertising- dj97@hal-pc.org Please tell our advertisers that you saw their ad in the *Orange Bytes*

Printed by:
Federal Printing & Mailing Services
16601 Gothard St. #F
Huntington Beach, CA 92647
714-842-7237-qpfedmail@aol.com

Guidelines for Submittal

You can now e-mail articles directly to the Editor through the North Orange County Computer Club’s Web Site: editor@noccc.org, or just give the file on disk and printout directly to Dave Keays at the next meeting.

To transfer your article from your Windows word processor, click at the beginning of the article with your mouse, shift down arrow to the end of the article, then Edit Copy (^C), open your e-mail program, and Edit Paste (^V). If your article is too long to include in an e-mail, please save as ASCII file with a .TXT extension. Then zip the article and attach it to your e-mail.

All documents should have flush left margins, and only one carriage return between paragraphs. The editors will bold your paragraph headings, etc., in order to obtain consistent formatting throughout the Bytes.

Don’t use CAPS for emphasis; that’s like shouting at someone! Use CAPS for computer commands or file names, like AUTOEXEC.BAT or KNOW.TXT. For titles of books and software, capitalize the first letter of each word.

Also please spellcheck your article, and try to follow the ordinary rules of grammar. Don’t use any kind of formatting (columns, tabs, indents, justification, hyphenation, etc.) If columns or tables are needed, send us a hard copy.

NOCCC Membership

If you haven't joined, you're missing a lot!

Meetings - Keep up with what's going on in the computer world. Hear outstanding industry representatives make presentations of the latest and greatest in computer products.

Special Interest Groups - Our 23 SIGs cover a broad spectrum of user interest; they invite you, whether you rank as beginner or seasoned computerist, to take in the lectures and demonstrations they sponsor, and to share computer knowledge.

Get Help with your current Computer Problems - In the Random Access portions of the meetings you ask your question of the entire assemblage, and more than likely someone will have the answer.

The NOCCC HelpLine - Some 100 NOCCC volunteers, experts in their fields, stand ready to assist with your problems, as close as your telephone.

The Orange Bytes Newsmagazine - Our award winning newsmagazine mailed monthly to your address reports on current activities and gives you articles and reviews geared toward your needs.

Raffles - We distribute thousands of dollars worth of hardware and software raffle prizes at our general and SIG meetings.

Product Review - Write a review for the newsmagazine and keep the software, hardware, book, or Cd-ROM. Members review more than 20 products a month. For a list of products available email (items@noccc.org).

Consignment Table - We have a thriving consignment table on our regular meeting day, in which we assist members to sell or buy all kinds of computer items. Use our handy on-line form to prepare your paperwork in advance.

Volunteer Work - You are given opportunities to help our activities with interesting assignments. An all-volunteer organization, you can join with other members in a variety of activities like writing articles for our newsmagazine to conducting a seminar, and more

You may attend a meeting without joining NOCCC. Dues are $30.00/year. Members are entitled to the many benefits listed above. OK, Sign me up!

QUICK MEMBERSHIP APPLICATION

PLEASE PRINT!

Date:_____ I am a new member_____

This is a renewal, my membership number is_____

Name _____

Address_____

City _____

State _____ Zip _____ Home Phone _____

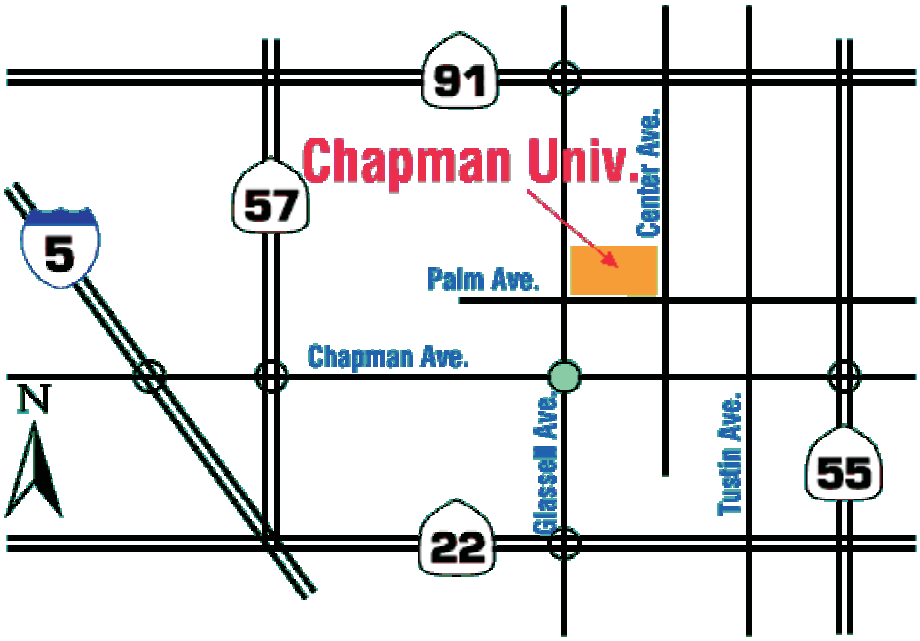
Fees: $30 per year plus $15 for each additional family member

Amount enclosed $_____

MC __ or VISA__

Credit Card Number _____Expires_____

Please make check payable to **North Orange County Computer Club**



parking information

Chapman University's main parking lot is on the north side of the campus (enter from Walnut) and is free of charge for NOCCC's meetings on Sunday. Please feel free to park in the parking lots.

Parking is also free on the campus-side of the surface streets. The city of Orange's parking laws prohibits parking in front of residential housing. Expensive parking tickets will be issued to violators.