Benchmarks



ABOUT BENCHMARK ONLINE

SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » Issues

Benchmarks - March, 2010

Campus Computing News



Social media, networks may use your personal information

By <u>Dr. Rich Anderson</u>, Information Security Coordinator, CITC, with <u>Jennifer Lafleur</u>, Assistant Director, Administration & Compliance, CITC, and the staff of <u>inHouse</u>

Guidelines for use of social networking and media sites such as Facebook, MySpace, LinkedIn, YouTube, Twitter, Flickr and others have been established by the Computing and Information Technology Center.



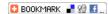
Spring Break Hours



By Claudia Lynch, Benchmarks Onlin Editor

Spring Break is here! If you're not leaving town for the entire week, the hours for various areas noted in this article could come in handy.

Read more



Staying Afloat in a Sea of Data



By DaMiri Young, HPC Systems Administrator

This article investigates different technologies for processing massive datasets that offer reasonable performance per unit



High Performance Computing Seminar Planned

By Dr. Philip Baczewski, Director of Academic Computing and User Services

Academic Computing and User services is planning a seminar entitled "An Introduction to High Performance Computing at

By the Numbers

UNT HPC Utilization for **February**

· Total number of completed jobs: 7,263

- · Average jobs per hour:
- Average job turnaround time (minutes): 5
- Average time consumed (CPU hours): 55
- · Total time consumed (CPU hours): 422,615 February was the first full month that the new Talon HPC cluster was in operation. For more information about UNT's central HPC resources, see: http://citc.unt.edu/hpc





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- <u>UNT Home</u>
- UNT System
- <u>UNT Dallas</u>
- <u>UNT Health Science Center</u>

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home

Campus Computing News



Social media, networks may use your personal information

By Dr. Rich Anderson, Information Security Coordinator, CITC, with Jennifer Lafleur, Assistant Director, Administration & Compliance, CITC, and the staff of inHouse

 ${f G}$ uidelines for use of social networking and media sites such as Facebook, MySpace, LinkedIn, YouTube, Twitter, Flickr and others have been established by the Computing and Information Technology Center.

These sites are tools to create, maintain, and rekindle social and professional relationships, but it may be difficult to maintain personal privacy or to restrict published information, the center cautions. Some sites may have the ability to harvest e-mail contacts or personal information to use without your consent.

The Safe Use of Social Networks and Social Media

The rise of social networks such as Facebook, MySpace, and LinkedIn, as well as social media sites such as YouTube, Twitter, and Flickr, has provided Internet users an unprecedented set of tools to create, maintain, and rekindle social and professional relationships. While these social networks and media sites have many useful features, users should be aware of some their downsides, particularly as it relates to use on UNT computers. These guidelines are intended to address individual use

As social networks and media sites are designed to make information available, it may be difficult to maintain personal privacy or to restrict information only to certain persons. Some sites may have the ability to harvest your email contacts or personal information and use this information in a manner in which you may not approve.

Individuals should use caution, particularly when accessing these sites from university computers.

Below are general principles for individuals to follow when using social network and media sites to avoid the potential

- 1. Read the privacy and acceptable use policies before joining or publishing information on a social network or social media site.
- 2. Avoid installing toolbars and plug-ins offered on social network or social media sites. These programs tend to slow down computers, cause technical problems and have a tendency to gather data about you that you may not wish to share.
- 3. Don't publish information that you wouldn't want your family, co-workers, supervisors, current and future employers, etc., to see.
- 4. Don't publish information about others that is personal, private or protected by law on a social network.
- 5. Remember that others can post information about you on these sites.
- 6. Social networks generate most of their revenue through advertising. Don't publish information about yourself that you don't want to be used for targeted marketing.
- 7. Choose your friends wisely. Avoid adding people as friends if you haven't met them in person or you don't have an established relationship with them through other channels. "Friends" on social networking sites usually have access to more of your information than the casual browser.

- 8. Social networks are as much about relationships as they are about sharing information. Think twice before becoming a "fan" of a controversial topic, group, or person if you wouldn't want your family, co-workers, supervisors, current and future employers to know about your support of that issue or person.
- 9. Avoid misrepresenting yourself or your background.
- 10. Don't forget that this is relatively new technology. New technology tends to be buggy, and bugs have a tendency to lead to security and privacy breaches.
- 11. Remember that not everything you see or read on the Internet is true. Always be sure to check sources of information before relying on what's published on social networking and media sites.

The above article originally appeared in *inHouse* February 23, 2010.

Administrative Announcement (3/2/10)-- Human Resources Featured Training: Social Networking: Friend or Foe?

Join us for this informative session designed for all employees. It explores how social networks affect the workplace and provides tips on how to stay safe from the potential personal and professional perils that can lie in wait. Participants will also learn what UNT policies relate to use of social networks in the workplace. **Date:** 3/25/2010 **Time:** 10:30am - 12pm **Location:** Marquis 118

Register Now

<u>Important Note:</u> When you click on the Online Registration Form below, you may be directed to a page that states 'There is a problem with this Website's Security Certificate'. If this occurs, please click 'Continue to webpage'. The page is safe.





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

<u>Disclaimer</u> | <u>AA/EOE/ADA</u> | <u>Privacy Statement</u> | <u>Web Accessibility Policy</u> | <u>State of Texas Online</u> | <u>Emergency Preparedness</u>

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » Spring Break Hours

Spring Break Hours

By Claudia Lynch, Benchmarks OnlinEditor

Spring Break is here! If you're not leaving town for the entire week, the hours for various areas noted in this article could come in handy.

Daylight Saving Time begins March 14

The Helpdesk will maintain normal hours through most of the break. The exception is that staff will be answering the phones and email from 8 a.m. - 5 p.m. on Monday, March 15 but the lobby will be closed to walk-in traffic. The University is officially closed Monday, March 15. The University will be open Tuesday through Friday, but NO CLASSES will be held. Some areas will also have abreviated hours during the week of Spring Break.

- Data Management Services will be closed over Spring Break.
- The ACS General Access/Adaptive Lab (ISB 104) will keep their normal hours all week.



Hours for Other Campus Facilities

Check out the UNT Shuttle Spring Break Schedule here: http://www.unt.edu/transit/routes_sched.html

General Access Labs

• WILLIS:

Saturday, March 13, Close at 11:50 p.m.

Sunday, March 14: Closed

Monday, March 15 - Friday, March 19 Open: 8 a.m.-7:50 p.m.

Saturday, March 20: Closed

Sunday, March 21, Open: 1 p.m.; resume 24hr schedule

• College of Information General Access Computer Lab (CI-GACLab) (B205):

Saturday, March 13 - Monday, March 15: Closed Tuesday, March 16 - Friday, March 19: 10:10 a.m. - 6 p.m. Sataurday, March 20: Closed Sunday, March 21, Open 10 a.m.; resume normal hours

• MUSIC:

Closed: Saturday, March 13 - Sunday, March 21

• PACS Computing Center (Chilton Hall):

Close: 5 p.m. Friday, March 12

Reopen: Saturday, March 20; resume normal hours

• CVAD (formerly SOVA):

Close: 6 p.m. Friday, March 12

Reopen: 7 a.m. Monday, March 22; resume normal hours

• <u>COE</u>:

Closed: Saturday, March 13 - Sunday, March 21

Reopen: 7 a.m. Monday, March 22; resume normal hours

• COBA:

Closed: Saturday, March 13 - Saturday, March 20 Reopen: Sunday, March 21; resume normal hours

• <u>CAS</u>:

GAB 330

Closed: Saturday, March 13 - Saturday, March 20

GAB 550

Closed: Saturday, March 13 - Sunday, March 21

Terrill 220

Closed: Saturday, March 13 - Sunday, March 21

Wooten 120

Closed: Saturday, March 13 - Sunday, March 21

• UNT Dallas Campus - 155A

Closed: Monday, March 15

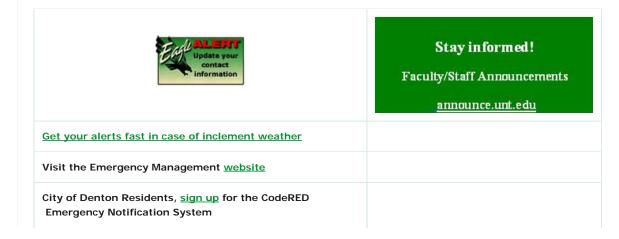
Open: Tuesday, March 16 - Friday, March 19: 7 a.m. - 6 p.m.

Open: Saturday, March 20: 7 a.m. - 5 p.m. **Closed:** Sunday, March 21; resuming normal hours

• Engineering General Access Lab (CENGAL, englab@unt.edu, Discovery Park, B129, 891-6733)

Closed: Saturday, March 13 - Sunday, March 21

Remember:







Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

<u>Disclaimer</u> | <u>AA/EOE/ADA</u> | <u>Privacy Statement</u> | <u>Web Accessibility Policy</u> | <u>State of Texas Online</u> | <u>Emergency Preparedness</u>

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » Staving Afloat in a Sea of Data

Staying Afloat in a Sea of Data

By DaMiri Young, HPC Systems Administrator

 ${f T}$ his article investigates different technologies for processing massive datasets that offer reasonable performance per

The turbulent sea

The perpetual quest continues to not only efficiently store but also efficiently search the sea of data we're immersed in. For instance, how useful would the World Wide Web be without web search engines? As Moore's law is in full effect, the ability and cost to store massive amounts of data is becoming more and more trivial. Also, raw textual data is still the de facto standard for storing and distributing large datasets. So this presents an interesting problem; we have lots of storage and lots of data. How do you process or search gigabytes or perhaps terabytes of raw data? Think of processing several thousand large text documents or DNA fragments in the textual "AATTGGCC" form, billions of characters long.

In much the same way that the web relies on web search engines, research in many fields dealing with text including Informatics, Computational Biology, and Natural Language Processing (NLP), rely on similar techniques to make problems solvable in a reasonable time. For some specific examples, see the article from last month detailing projects in Digital Humanities. Not to mention the massive amounts of data resulting from computations in countless other research fields that are virtually useless without efficient ways to store, search, and visualize it.

Paddling with one oar

A naive approach to the problem might be to attempt to use a tool like Awk, Sed, or Grep. Don't get me wrong, these tools have their rightful domain. In that domain they perform extremely well, just ask your local Linux system administrator. However, utilizing these tools for serious text processing would be futile at best. The main issue with these tools is dealing with the almost arbitrary data lengths at which they cease to function. Ever been greeted with the infamous "grep: Argument list too long." error? Supposedly, there are workarounds that allow searching through at least 100 thousand files. Be prepared to wait a while if you choose this route however. Lets say you do manage to get around the operating system and data length limits. There would still be the sizable and time consuming task of trying to apply these tools for data intensive research. Lets move on to the next potential solution, shall we?

Hoisting the sails

A more sophisticated approach would be to employ any of the high-level programming languages to craft a tool. These languages might include C, C++, Perl, Python, PHP, or Ruby for example. Most of these languages expose potent text processing facilities without the seemingly arbitrary data length limits found in the above basic tools. Indeed this facilitates intuitive manipulation of raw data among other tasks. Combine this with the free and opensource nature of these dialects and its a win-win situation right? Wrong.

Unfortunately, the learning curve for developing and employing a tool written in one of these languages is often steep. To acquire a working knowledge of PHP, Python, Perl, C++ or dare I say it-- C, it usually takes a savvy individual months at best weeks. Working knowledge in this context meaning enough grasp to actually write a useful piece of software. With any luck, the software will even facilitate reaching the research goals at hand. However, a useful piece of software does not equate to a maintainable or reusable piece of software. How useful is software that isn't maintainable or reusable? Without getting into the icky details of the software development lifecycle, I'll say oneshot-one-kill applications are hardly worth the bother.

Keep in mind that usually the people tasked with developing these tools are not computer scientists, programmers, or code hackers at all. They are Chemists, Biologists and Social Scientists who just happen have a need to process large data sets in order to get their work done. There are plenty of notable examples of success stories but there are many more examples that don't end so well.

A free powerboat

It's always a good thing when ingenuity produces breakthroughs in one area that are applicable in many other unrelated areas. Data intensive processing is one such example. Decreasing hardware costs, advances in distributed computing i.e., parallel processing, and algorithmic developments have decreased the cost and time it requires for data intensive tasks.

Among readily available technologies to aid in processing massive datasets is the proverbial MapReduce. MapReduce is a patented framework presented by Google to support parallel processing of huge datasets on clusters of computers. It implements the 2-step map and reduce procedure that is applicable to many real world problems. The real power of MapReduce is that it is relatively simple to implement on commodity hardware clusters and scales nicely. With a little elbow grease and some skill, one could be ready to process terabytes of data in a weekend. In fact here are the specs for a typical implementation:

- 100s-1000s of dual CPU x86 machines
- 2-4GB of memory
- Gigabit Ethernet
- Local IDE storage

Of course Google utilizes MapReduce extensively thus proliferating its use. However, as MapReduce is proprietary, it is the open-source Hadoop project by Apache that is providing MapReduce techniques to the masses. The open nature of Hadoop has lent the project to being supported and used by a large and diverse international community. Companies like Yahoo! have hopped on the bandwagon contributing to and even using Hadoop abundantly in its search functionality. Also, IBM and Google teamed up to announce a major initiative to use Hadoop to support university courses in distributed programming. Here is an extensive list of organizations and applications that utilize Hadoop.

Hadoop provides access to several popular "elastic" filesystems including Amazon S3, Cloudstore, FTP filesystem, HTTP filesystems and it's own Hadoop Distributed File System (HDFS). HDFS allows storage and processing of gigantic files across multiple machines in the cluster. It achieves significant redundancy by replicating the data across multiple hosts thereby avoiding the need for the hosts to have RAID storage. The default replication level is 3 meaning data is stored on three nodes, two in the same rack and one in a different rack. HDFS is not restricted to MapReduce jobs and in fact has been extending to the HBase database, Apache Mahout machine learning system, and of course matrix operations.

Boarding the powerboat

In theory, Hadoop can be used for any work that is batch-oriented, data-intensive, and able to be processed in parallel (hint, hint). Many strides have been made to extend Hadoop into several disciplines, making it essentially a swiss-army knife for research. Here are some projects that highlight the diverse ways in which Hadoop has been used:

- A group of biologists from the University of Maryland utilized open-sourced software <u>Crossbow</u> to perform over 2 billion reads, assembling a human genome sequence in roughly 6 and a half hours on 80 CPU cores at a cost of \$62 dollars. This was done via the Amazon EC2 cloud service.
- A group of NLP researchers at Stanford used MapReduce/Hadoop for around a 3x speed up in Machine Learning algorithms on a campus cluster.
- A <u>collaborative group of six teams</u> from the University of Maryland working on Statistical Machine Translation, used a campus cluster and Hadoop/MapReduce to estimate phrasal translation probabilities in several gigabytes of data in ~20 minutes.
- The <u>New York Times</u> used 100 Amazon EC2 instances and a Hadoop application to process 4TB of raw image TIFF data into 11 million PDFs in 24 hours at a cost of \$240 (Hello, Humanities scholars).

These are just examples of what has been accomplished with Hadoop. They are by no means the end of the story, only the beginning. There is a wealth of examples demonstrating the usefulness of Hadoop and MapReduce for research, especially collaborative research. However, like most things in life, there are trade-offs and Hadoop is no exception. To go back to the boating analogy, clearly a powerboat is not always the right vehicle for traveling the ocean. Also, it takes a bit of skill to pilot a powerboat. Heres what Hadoop is not:

- A substitute for a database, Hadoop's claim to fame is being able to successfully process raw unstructured text.
- MapReduce is not always the tool for the job, many operations are independent from others making it problematic for some tasks.

- Hadoop requires one to know a bit of Java, notice "know". It is not the correct place to learn Java programming.
- Hadoop clusters are not the place to learn Unix/Linux administration. It assumes you're familiarity with these concepts, to set it up at least.
- Hadoop Filesystem is no substitute for a High-Availability SAN filesystem. Systems like Lustre, GPFS, and GFS run circles around Hadoop's filesystem as they usually run over high-speed networks like infiniband.
- Hadoop throws away a lot of the performance gained by being a POSIX compliant filesystem like the said filesystems. This usually results in brute-force approaches to many task.

Conclusion

Tools like Hadoop and MapReduce are in many cases the only avenue for non-computer scientists to utilize distributed computing in their work. Also, the metaphor "not re-inventing the wheel" is applicable here. Why spend precious weeks writing your own basic tools when one can learn a bit of Java and spend the same hours writing advanced tools to quickly perform real research? On our own local campus, it seems facilities like Hadoop are crucial to support the collaborative research clusters that the Office of Research and Economic Development is advocating. These researchers need a quick and powerful way to express and implement their ideas.

This is exactly what makes MapReduce attractive to many researchers. It is a simple model by which one can quickly express complicated distributed programs. These fundamental concepts are what has brought Hadoop significant interest from the academic research community. Make no mistakes, we have all intentions of ensuring our local research community has the tools with which to collaborate and harness significant computing power quickly. That being said, take note of the increasing stories of computing that we once measured in months and thousands if not millions of dollars, that we'll now measure in just hours and hundreds of dollars.

External Links:

http://citc.unt.edu/hpc

http://research.unt.edu/

http://research.unt.edu/clusterpositions.htm

http://wiki.apache.org/hadoop/

http://bowtie-bio.sf.net/crossbow

http://labs.google.com/papers/mapreduce.html

http://research.yahoo.com/files/Lin-cloud.pdf

 $\underline{http://www.umiacs.umd.edu/\sim}jimmylin/cloud-computing/$

http://hpcinfo.com/2009/11/22/benchmarking-the-cloud-for-genomics/

 $\underline{http://www.cs.stanford.edu/people/ang//papers/nips06-mapreducemulticore.pdf}$

 $\underline{http://sourceforge.net/apps/mediawiki/cloudburst-bio/index.php?title=CloudBurst}$

http://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/

http://hpcinfo.com/2009/07/02/a-look-at-the-yahoo-systems-used-to-sort-a-petabyte-in-1625-hours-and-a-terabyte-in-62-seconds/





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » High Performance Computing Seminar Planned

High Performance Computing Seminar Planned

By Dr. Philip Baczewski, Director of Academic Computing and User Services

f Academic Computing and User services is planning a seminar entitled "An Introduction to High Performance Computing at UNT."

This 1.5 hour presentation will provide an overview of central HPC services, highlight the University's 224 node, 1792 processor Talon compute cluster system, and cover the basic information needed to apply for and access central HPC systems. The seminar is scheduled to be held on April 23, 2010 at 2 p.m. in Chilton Hall room 245. For more information, contact Dr. Philip Baczewski (baczewski@unt.edu), Director of Academic Computing and User Services.





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » network-connection

Network Connection

By Dr. Philip Baczewski, Director of Academic Computing and User Services

Here Comes the iPad

Mark your calendars. On April 3, you will be able to buy a "magical and revolutionary device" (Apple's words, not mine) and the Internet may never be the same. At least, that's what you'd think if you believe all of the prerelease chatter surrounding what's so far been a mostly theoretical device. The iPad could kill the Kindle, slay netbooks, and revolutionize education. On the other hand, the iPad is missing all kind of essential components, is bad for freedom, and is basically good for nothing. That's a pretty extensive record for a device that hasn't even been available yet for sale.

Once you get beyond the fact that the iPad won't have Flash or (GASP!) a floppy drive, it's interesting to consider how the iPad reflects the current state of the Internet and access to information and media. What the iPad seems to be is one of the first devices of its class that is media and information rich, but not designed as a computing platform. If you are reading this, you probably own or have access to a computer, but you may not actually do much computing. Sending e-mail, editing documents, and browsing the web all utilize computing resources, but the architecture on which you are doing that activity is more a matter of tradition than necessity.



Computer architecture

Computer architecture hasn't changed much over the past 50 years. Computers have a central processing unit (CPU), random access memory (RAM), a direct access storage device (DASD or hard disk), and an operating system to control the execution of instructions and transfer of information between the various system components. Computers were originally designed to be flexible calculators and this architecture is optimized for performing calculations. As I write this column, I'm using, at most, about 5% of my computer's calculating capacity. However, if I fire up a statistics application and crunch some serious data, or start a ray-trace application to render a nice 3-d image, I'd be using the computer architecture and capacity to a much greater extent and getting a much greater return on the investment that was made to buy the computer.

The iPad

Enter the iPad, a relatively inexpensive device that has the potential to let you use somebody else's computing capacity. The iPad still has components of the traditional computer architecture, but it has an operating system that is not as much about computation as it is about communication. The iPad will run local applications and store local files, but many of its functions are dependent upon remote sources of information. You won't run statistics or raytrace software on the iPad, but you might use it to connect to a remote service that can do those things.

What the iPad will do is concentrate your information activity on a convenient portable device. If you primarily access web pages, e-mail, online videos, music, photos, maps, and books, then perhaps a portable, well-connected device at about \$500 makes more sense than a computation device that costs \$1500. The iPad's not a computer. It's a mediacommunicationinformationater.

What Apple does well

Apple has again done what Apple does well. They have shifted a paradigm and made the available more accessible. First it was the computer for the "rest of us", and then the PDA (the long-forgotten Newton), and then the music player, and then the cell phone, and now a device to easily access the information cloud. The information cloud is not new. For centuries, though, it was locked in printed books and periodicals, images and photo albums, theaters, and concert halls. The Internet has enabled the information cloud to be more accessible. The iPad, whether it turns out to be a resounding success like the <u>iPod</u> or a hopeless failure like the <u>Newton</u>, just takes the next step.





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » link-of-month

Link of the Month



Acdording to an <u>article</u> on the UNT News Service website, students will soon be able to get help with their Income Tax preparation:

The Student Money Management Center will take the anxiety out of tax season by offering several tax-related events in the coming weeks. The first event will be called Preparing for the Tax Man and will be held from 5 to 6 p.m. on March 22 (Monday). This workshop will teach simple techniques to take stress out of the tax season.

The center, in partnership with the <u>United Way of Denton County</u>, also will host two free tax preparation sessions via the Volunteer Income Tax Assistance Program, which provides free income tax return preparation to taxpayers earning up to \$45,000 annually. The VITA sessions will be held on March 30 (Tuesday) from 5 to 8 p.m. and April 6 (Tuesday) from 5 to 8 p.m. The VITA sessions will be held in Chestnut Hall, room 324. Reservations are required for the sessions, call 940-369-7761 to reserve a space.

For more information about these events and other services, visit the Student Money Management Center website:

http://moneymanagement.unt.edu/





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- <u>UNT Home</u>
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

<u>Disclaimer</u> | <u>AA/EOE/ADA</u> | <u>Privacy Statement</u> | <u>Web Accessibility Policy</u> | <u>State of Texas Online</u> | <u>Emergency Preparedness</u>

Benchmarks





ABOUT BENCHMARK ONLINE

SEARCH ARCHIVE

SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » helpdesk-fyi

Helpdesk FYI

By Jonathan "Mac" Edwards, Assistant Manager of the CITC Helpdesk

Viewing Shared Outlook Calendars Via the Web

Soon you should be able to view published Outlook Calendars via Office Online, at this time though that functionality is not available, or at least not fully implemented.

Another method for viewing Calendars online is to view them directly using a web address.

Browser Requirements: To view Outlook Calendars in this method you are required to use IE 6+

Accessing the Calendar:

- To access a Calendar via the web you must first be given access rights by the Calendar creator, or another
 owner. To grant this access open the Outlook Client on your PC (not the web version), go to Calendar you
 wish to share, Right Click, Choose Properties, and Click on the Permissions Tab. Click Add..., add the user,
 choose Permission Level, click OK, and then Apply.
- You will need to know the address of the Calendar. Only the Calendar creator will have this information, which can be found by right clicking on the Calendar, Choose Properties, and select the General Tab.
 - Location: This is the location of the Calendar. It will be easiest to access a Calendar located in your inbox.
 - Next to the Hand I con: This is the Calendar's name. It will be easiest to access a Calendar
 whose name contains no spaces or special characters.

The Address

- Now that you have Access to the Calendar and know it's name and address you should be able to access it in your browser
- Address Format: https://webmail.unt.edu/OWA/first.last@unt.edu/?cmd=contents&f=calendar
 - first.last@unt.edu = Shared Calendar Owner's Email address
 - Calendar = name of the calendar
 - *note these are for calendars created in the inbox: ex. Under Properties the Location=\\Mailbox last, first. If the location is within a sub-folder please look at the Advanced Section of this article
- Example: Calendar Creator = john.doe@unt.edu
 - Calendar Location = \\Mailbox Doe, John

- Calendar Name = test
- $\bullet \ \ Final \ address = \underline{https://webmail.unt.edu/OWA/john.doe@unt.edu/?cmd=contents\&f=test}$

Advanced Section

While accessing Calendars in the inbox is fairly simple, accessing a shared Calendar that is located in a subfolder can be a little more difficult.

- Use the example below to access a Calendar in Subfolders
 - Example: Calendar Creator = <u>john.doe@unt.edu</u>
 - Calendar Location = \\Mailbox-Doe, John\subfolder
 - Calendar Name = test calendar
 - Monthly View
 - Final address = https://webmail.unt.edu/OWA/john.doe@unt.edu/?cmd=contents&f=subfolder%2...
- · Formatting to achieve final address:
 - subfolder is the name of the folder that contains the calendar
 - %2f is placed between each subfolder
 - %20 is used in place of a space.
 - So in our example you can see that the address is using subfolder%2ftest%20calendar
 - At the end we add a view option: **&view=weekly**. By default the calendar will display daily.
 - The three options for view are daily/weekly/monthly
- Longer Example
 - Example: Calendar Creator = john.doe@unt.edu
 - Calendar Location = \\Mailbox-Doe, John\subfolder\interiorfolder\supersubfolder
 - Calendar Name = shared test calendar
 - Monthly View
 - Final address = https://webmail.unt.edu/OWA/john.doe@unt.edu/? cmd=contents&f=subfolder%2finteriorfolder%2fsupersubfolder%2fshared%20test%20calendar&view=monthly

Additional Commands: You can find the entire list of commands at http://technet.microsoft.com/en-us/library/bb232199(EXCHG.80).aspx

Additional Information: Soon we will be adding an automatic address formatting application, allowing you to
more easily format the web address for shared Calendars. Check for updates to this tutorial at:
 http://remedy4.ars.unt.edu:8080/rkm/viewdoc.jsp?
 http://remedy4.ars.unt.edu:8080/rkm/viewdoc.jsp?
 doc=329&sid=10853&type=Published&terms=quick_searchTerms&user=Self%20Help





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- <u>UNT System</u>
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

<u>Home</u>

Benchmarks Issue content





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have guestions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- <u>UNT System</u>
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

<u>Disclaimer</u> | <u>AA/EOE/ADA</u> | <u>Privacy Statement</u> | <u>Web Accessibility Policy</u> | <u>State of Texas Online</u> | <u>Emergency Preparedness</u>

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » rss-matters

RSS Matters

Research and Statistical Support University of North Texas

Homogeneity of Variances

Link to the last RSS article here: Moving on up ... again: Introduction to new features of PASW Statistics 18 (i.e. SPSS _18) Take a look at the "new and improved" RSS website too! http://www.unt.edu/rss/ - Ed.

By Dr. Jon Starkweather, Research and Statistical Support Consultant

 ${f T}$ his month we touch on a fundamental issue in statistical evaluation that often gets overlooked. Testing assumptions for parametric analysis is a fundamental step and a necessary one. For example, let's consider some of the simplest experimental design analysis available; the independent t-test and analysis of variance F test—testing for mean differences among independent groups. These tests have three key assumptions; normality, independence of observations, and homogeneity of variances (HOV). Generally speaking, experimental design dictates random sampling from a well defined population and random assignment to groups (aka. conditions, levels, etc.), both of which should help take care of assumptions mentioned. But, let's focus our attention on the third assumption (HOV), which needs to be (and can be) tested to ensure accurate or valid interpretation of the mean differences. Luckily, most statistical software packages offer a way to test for HOV (including PASW/SPSS). Generally, the Levene's test is used to statistically test the amount of difference between variances (of groups selected for a t-test or F test).

A Royal Rumble

Means vs. variances, a Royal Rumble... Levene's test is testing for differences among our group's (2 or more) variances. A t-test is testing for differences among 2 group's means. An F test (one-way ANOVA) is testing for differences among more than 2 group's means. In these contexts, the independent variable is comprised of multiple groups; for the t-test, there are two groups; for the F test, there are more than two groups. Each group represents a treatment or lack of one in the case of a placebo. In essence, each group receives something different as stimulus, for example different drugs administered in each condition of an efficacy study.

Essentially, whether looking at 2 groups (t-test) or more than 2 groups (F test), we are concerned with the assumption of homogeneity of variances (among other assumptions). Recall that variance is a measure of dispersion, how much do the scores (of one group) VARY around the mean (whatever that mean happens to be). Mean is a measure of central tendency; arithmetic average. The HOV assumption states that our groups are similar in essence (similar variances), regardless of independent variable level (treatment or condition administered).

Providing some practical examples for the discussion.

Treatment administered is what each group experiences that is different, each level of the independent variable. Treatment in this sense represents your independent variable; that which is manipulated in the experiment and you are trying to establish that each level or group displays mean differences. The groups for our novel example will be Zoloft vs. Xanax vs. Lithium vs. Placebo (a lack of treatment). The dependent variable is that which is used to measure change in the independent variable, which for this example will be Statistics Anxiety scores

If we randomly sample introductory statistics class undergraduates at UNT, then randomly assign them to 4 treatment

groups; we are ASSUMING the students are similar (homogeneous). Because they are all undergraduates taking intro stats at the same university and we randomly sampled and randomly assigned; thus equalizing individual differences (hair & eye color, etc.). However, if we realize after randomly assigning them to our groups that most of one group was made up of engineering majors and one group was almost completely made up of music majors while the third was made up of primarily English majors and the fourth primarily physics majors; then we can see that the groups likely differ regardless of treatment administered.

Stated another way, we are likely to have significant group differences (and heterogeneous variances), because each group is different in essence and is likely to differ on our dependent variable (Statistics Anxiety scores). Already, you can imagine English and music majors having higher Statistics Anxiety scores than the physics and engineering majors, simply because they have different mathematics course requirements and likely different interests.

Why is this important? Well, if our groups are inherently different, then any differences we find in our dependent variable (the mean stats. anxiety scores) after administering treatments (the drugs) may have been due to the treatments OR the inherent differences of the groups (we wouldn't know which). In which case, whatever statistical test (*t*-test, or *F* test) results we find are of no practical validity. We cannot be confident that our Zoloft group displayed less Statistics Anxiety due to the Zoloft, because they may simply have been better or more relaxed with statistics due to a background heavy in mathematics.

A more precise example (stay with me now...)

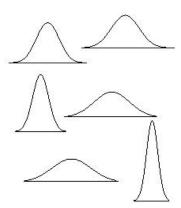
Imagine administering 150 mg. of Xanax (a commonly prescribed anti-anxiety drug) to one group and administering a placebo (inert pill) to the other group (2 groups only = *t*-test). Our dependent variable is Stats. Anxiety scores again. Well, does each individual respond to the same dosage (say 150 mg.) of any drug differently? Consider something as simple as body weight, more body weight = more volume of drug required to have the desired effect (makes sense doesn't it?). Obviously its more complex than that, physiology, liver functioning, tolerance, etc. each plays a part. BUT we all understand that six shots of vodka for me (approx. 180 lbs.) is going to have a different effect than for my fraternal twin (approx. 140 lbs.; usually on the floor drooling on himself after six shots!). SO; if each person in our Xanax group reacts differently to the 150 mg. Xanax, then that group is likely to have more variance than our placebo group. Which is likely to have very little variance because they were administered no active drug...therefore, each person's weight, physiology, liver functioning, tolerance, etc. will not matter in the placebo group—regardless of mean Statistics Anxiety score!

You can now see that when looking at the variances of statistics anxiety scores we might find differences based on body weight, not necessarily on who got the Xanax and who didn't.

REMEMBER, our *t*-tests or *F* tests are testing for differences among the group **means**. Levene's test is testing for differences among group **variances**.

Another way of Looking at it.

Consider a few distributions each with different variance:



Imagine each of these represents one of our groups; Zoloft, Xanax, Mountain Dew, coffee, alcohol and placebo... You can see it makes no difference what mean (stats. anxiety score) happens to be under the middle of each distribution; they are different from one another in their variance. Inherently different groups! Stated another way; each group responded to their respective treatment differently. Some group's participants were more similar (low variability or a narrow distribution), while others were more different (high variability or a wide distribution). How does this relate to the Levene's test of the HOV assumption?

Recall; the Homogeneity Of Variances assumption stipulates that our groups have similar variances; similar reactions to the treatment/condition/drug they received. If this assumption holds then we know that whatever test result (*t*-test or *F* test) we find is attributable to the different treatment (drug) each group received (treatment effects, not confounds). Furthermore, recall that Levene's test is testing whether or not the variances of our groups are

statistically different. We generally use the .05 probability level (or "Sig." value) to determine statistical significance; so, if Levene's test shows a "Sig." value of less than (<) .05; then we conclude that the variances are significantly different; meaning our statistical test (t-test or F test) is invalid and we can't make conclusive inferences from it. Likewise, if Levene's test shows a "Sig." value of greater than (>) .05; then we conclude the variances are NOT significantly different---which is what we **want** to see so that we can have confidence in the validity our t-test or F test result.

Additional discussion of Heterogeneity of Variance:

Bryk, A. & Raudenbush, S. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, *104*(3), 396 – 404. DOI: 10.1037/0033-2909.104.3.396

Until next time, you don't need a weatherman to know which way the wind blows...





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » short-courses

Short Courses

 ${f T}$ here are still some Spring Short Courses being offered through the first part of April, including the two new courses -- Applications in R: Latent Variable Modeling with Survey Data -- Part I and Applications in R: Latent Variable Modeling with Survey Data -- Part II. Surf over to the Short Courses page to see the the schedule.

Special classes can always be arranged with the RSS staff. See "Customized Short Courses" below for further information. Also, you can always contact the RSS staff for one-on-one consultation. Please read the FAQ before requesting an appointment though.

Especially for Faculty and Staff Members

In addition to the ACS Short Courses, which are available to students, faculty and staff, staff and faculty members can take courses offered through the Human Resources Department (they have a new comprehensive training curriculum), and the Center for Learning Enhancement, Assessment, and Redesign. Additionally, the Center for Achievement and Lifelong Learning offers a variety of courses, usually for a small fee.

EIS training is available. Questions or comments relating to EIS training should be sent to EISTCA@unt.edu.

High Performance Computing Seminar Planned -- April 29, 2010 at 2p.m.

Microsoft E-Learning

Microsoft E-Learning courses are now available for faculty and staff via our UNT-Microsoft Campus Agreement. Please contact Claudia Lynch at lynch@unt.edu for instructions on accessing this training.

Microsoft Outlook Training and more

The Messaging Systems Group has all sorts of useful information on their website, including training information.

Central Web Support

Consult Central Web Support for assistance in acquiring "Internet services and support." As described on their website:

CWS provides Internet services and support to UNT faculty, staff and students. Services include allocating and assisting departments, campus organizations and faculty with web space and associated applications. Additionally, CWS assists web developers with databases and associated web applications, troubleshooting problems, support and service.

CLEAR (was Center for Distributed Learning)

CLEAR offers courses especially for Faculty Members. A list of topics and further information can be found here.

The center also offers a "Brown Bag" series which meets for lunch the first Thursday of each month at Noon in Chilton 245. The purpose of this group is to bring faculty members together to share their experiences with distributed learning. One demonstration will be made at each meeting by a faculty member with experience in distributed learning. More information on these activities can be found at the CLEAR Website.

UNT Mini-Courses

There are a variety of courses offered, for a fee, to UNT faculty, staff and students as well as the general public. For additional information surf over to http://www.unt.edu/minicourses/

Information Security Awareness

The UNT Information Security team has been offering Information Security Awareness <u>courses</u> to all UNT faculty and staff. Topics to be covered will include workstation security, sensitive data handling, copyright infringement issues, identity theft, email security, and more.

For more information, or if you would like to request a customized course to be taught for your department, contact Gabe Marshall at x4062, or at security@unt.edu.

Also, Information Security Training is now available through Blackboard Vista (formerly known as WebCT).

Alternate Forms of Training

Many of the General Access Labs around campus have tutorials installed on their computers.

See http://www.gal.unt.edu/ for a list of labs and their locations. The Willis Library, for example, has a list of Iss of Tutorials and Software Support. The Library Instructional Unit also offers workshops and training, including "tech skills" training. Visit their website for more information: http://www.library.unt.edu/library-instruction

The <u>Training Website</u> has all sorts of information about alternate forms of training. Computer Based Training (CBT) and Web-based training are some of the alternatives offered, although due to the rising costs of training, shrinking budgets and changing technology, computer-based training at UNT is in a state of transition. For up-to-date information on CBT at UNT, see the CBT <u>website</u>.

Gartner Research Services

Way back in 2006 we announced <u>Gartner Core Research Services Now Available to the UNT Community</u>. Our subscription for Gartner services has always included **all** UNT faculty, students, and staff. All you need to do to access the subscription is to log into the UNT Gartner portal page at https://gartner.unt.edu/. Gartner is now offering "Webinar Wednesdays." To view all the offerings see: https://my.gartner.com/portal/server.pt?tbb=webinarcalendar You can also listen to Gartner podcasts here:

http://www.gartner.com/it/products/podcasting/asset_137461_2616.jsp.

State of Texas Department of Information Resources

Another possible source of training for staff and, perhaps, faculty members is the Texas Department of Information Resources. A look at their Education and Training <u>website</u> reveals some interesting possibilities.





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE

SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home » issues » 2010-03 » staff-activities

Staff Activities

Transitions

New Employees:

- Adam McMillon, CSS Tech, Classroom Support Services (part-time)
- Christie Jones, CSS Tech, Classroom Support Services (part-time).
- Beverly J. Forbes, IT Programmer Analyst, Admissions Management Systems Development Team (AIS).

No longer working in the Computing and Information Technology Center:

- Christina Marie Day, CITC Helpdesk Consultant (part-time).
- John J. McGraw, CSS Tech, Classroom Support Services (part-time).
- Kevin J. Wade, CSS Tech, Classroom Support Services (part-time).
- Gwynnethe Viban, CSS Tech, Classroom Support Services (part-time).
- Antonio Ventorini, Micro Maintenance Tech, Administrative Desktop Support (part-time).

Changes, Awards, Recognition, Publications, etc.

Service to UNT

The March 2 issue of InHouse recognized John Hooper, Execitve Director, Administrative Information Systems for his 20 years of service, Dorthy Flores, Director, Finance and Administrative Systems (AIS) for her 15 years of service and Tom McElwee, Senior Director, Enterprise Systems Technical Services, for his 5 years of service to the University. Congratulations!

Fun Fact Winners

Continuing the CITC tradition, we have some more "Fun Fact Winners." Congratulations to Jay Maxwell, Programmer Analyst, Contributor Relations, Finance & Administrative Systems (AIS) who was a winner in the March 3 InHouse prize giveaway. Bari Tinker, Programmer Analyst, Student Records, Campus Solutions Student Admin. Team (AIS) was a winner in the March 10 InHouse prize giveaway.





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338

http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- <u>UNT Home</u>
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016

Benchmarks



ABOUT BENCHMARK ONLINE SEARCH ARCHIVE SUBSCRIBE TO BENCHMARKS ONLINE

Columns

Network Connection

Link of the Month

Helpdesk FYI

ITC News

RSS Matters

Short Courses

Staff Activities

Home

Today's Cartoon



"One of my tweets is being turned into a screenplay and Broadway musical!"

From "Today's Cartoon by Randy Glasbergen", posted with special permission. For many more cartoons, please visit $\underline{www.glasbergen.com}.$





Contact Us:

University Information Technology

1155 Union Circle #310709 Denton, TX 76203 USA Voice: 940-565-4068 Fax: 940-565-4060

Visit Us:

Sage Hall, Room 338 http://it.unt.edu/benchmarks/



Email us:

Have questions on content or technical issues? Please contact us. unt.uit@unt.edu



UNT System:

- UNT Home
- UNT System
- UNT Dallas
- UNT Health Science Center

Site last updated on April 22, 2016