

---

## Research Resources

**Patricia Ledesma Liébana**  
(Director, Research Computing)

September 19, 2006

1

## Introductions

---

- New Faculty
- Research Computing
  - Alain Bonacossa, [Senior Research Associate](#)
  - Damba Lkhagvasuren, [Senior Research Associate](#)
- NU Library staff ([email @northwestern.edu](mailto:@northwestern.edu))
  - Leslie Bjorncrantz, [Bibliographer for Management and Psychology](#) ([email: l-bjorncrantz](mailto:l-bjorncrantz))
  - Jeannette Moss, [Reference Librarian / Instruction Librarian](#) ([email: j-moss](mailto:j-moss))
  - Jami Xu, [Reference Librarian & Liaison to the Kellogg School of Management](#) ([email: jamixu](mailto:jamixu))
  - Kathleen Murphy, [Social Science Data Services Librarian](#) ([email: kemurphy](mailto:kemurphy))

2

## Other NU library people

---

Bibliographers:

- Harriet Lightman, Bibliographer for Economics, Sociology
- Scott Garton, Bibliographer for Anthropology
- Lucy Lyons, Bibliographer for Political Science
- Other areas:  
[www.library.northwestern.edu/collections/selectors.html](http://www.library.northwestern.edu/collections/selectors.html)

Schaffner Library:

- Julie Borden, Head (j-borden, 3-0720)
- Carol Doyle, Reference & Instruction Librarian (c-doyle, 3-6618)
- Qiana Johnson, Reference & Instruction Librarian (q-johnson, 3-6617)

## Agenda

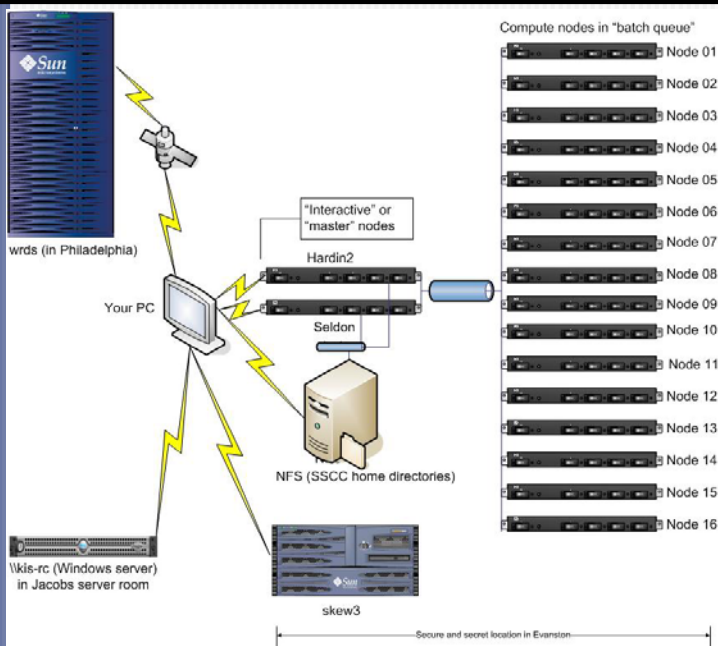
---

- What we do
- Research servers
- Software
- Data
- Experimental resources
- Some useful utilities
- Library resources

Feel free to ask questions as we go along.

## What do we do?

- Administer school wide resources
- Help faculty and doctoral students in search of data, taking a look at buggy programs...
- Provide training for faculty and doctoral students
- Assist faculty in the creation of training material for MBA students that uses research data or quantitative tools



## Research servers 1: skew3.kellogg.northwestern.edu

- Kellogg faculty and doctoral students only, no web services or email
- Access it using Secure Shell SSH or StarNet X-Win32
- Sun Fire V420, 4 CPUs, 16GB of RAM, running Solaris 8, 1.7TB of hard disk space
- Will be upgraded in early January (3 year cycle)
- Rich suite of software
- Datasets: Zacks History Files, Hemscoff Executive Compensation, R&C Futures, NAIC Property Casualty InfoPro, CIK to GVKEY match file, etc
- No multiple concurrent jobs

[www.kellogg.northwestern.edu/researchcomputing/skew3.htm](http://www.kellogg.northwestern.edu/researchcomputing/skew3.htm) 7

## Research servers 2: Social Science Computing Cluster (SSCC)

- Shared facility accessed by researchers in the social sciences (mostly Kellogg and economics)
- 18 dual CPU (AMD Opteron) nodes, two of them are “master” (interactive) nodes and 16 are compute nodes
- Linux based, master nodes have 16GB of RAM, while compute nodes have 8GB of RAM
- Batch management software: Altair PBS Pro
- Up to 8 concurrent jobs in compute nodes
- Rich suite of software
- Important datasets from ICPSR
- Parallel computing will be enabled soon (MPI)
- 6 nodes renewed every year

[sscc.northwestern.edu](http://sscc.northwestern.edu)

8

## Research servers 2: SSCC – how does this work?

---

- Why multiple jobs? (1) For a simulation, runs with different parameters; (2) For estimation, different specifications
- Need to modify program to be able to distinguish output when a job is done
- Login to a master node (seldon.it.northwestern.edu or hardin2.it.northwestern.edu)
- Write a shell script and use PBS Pro (“qsub” command) to send to the batch queue
- You will get email messages from the system when the program starts, ends or if there is an abnormal termination (error that causes program to end)

9

## Research servers 3: \\kis-rc

---

- MS Windows server (Dell)
- Holds datasets that have a Windows graphical interface: SDC Platinum, Economática, Datastream (finance department), Emerging Markets Database (EMDB)
- Holds installers for some research software: Matlab, and, coming soon, MediaLab and DirectRT
- Access by going to your desktop Start menu, select run and type \\kis-rc

10

## Research servers 4: Wharton Research Data Services (WRDS)

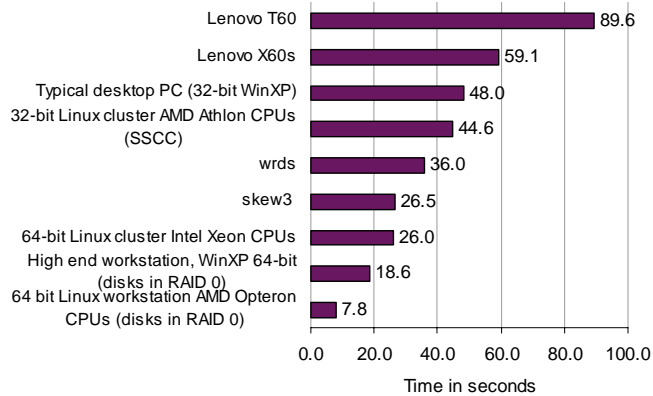
- Holds many of Kellogg’s data subscriptions
- Sun Fire E12K, 32 CPUs, 128MB of RAM
- Access with Secure Shell SSH or via the web
- Heavily used by subscribers, skew3 will be faster doing data processing ⇒ Use for data extraction
- Avoid the web interface:
  - Replicability
  - Rounding
  - Data “clean-up”: WRDS drops observations with missing values in selected variables
- Web interface useful for quick checks and for teaching

[wrds.wharton.upenn.edu](http://wrds.wharton.upenn.edu)

11

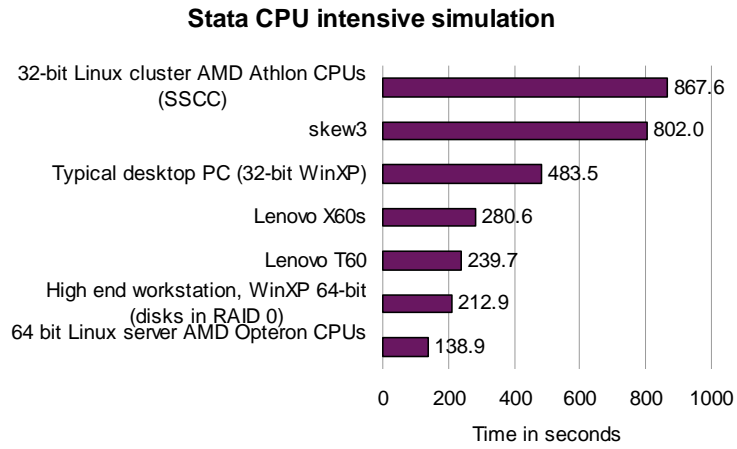
## Performance

Time used by SAS copying a 1GB data file



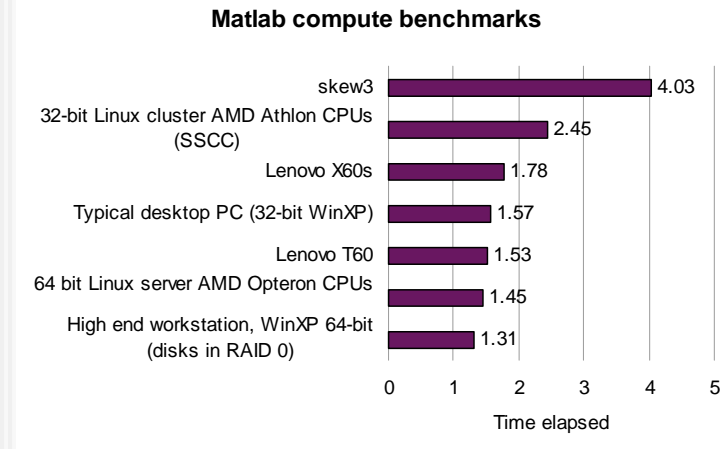
12

## Performance (continued)



13

## Performance (continued)



14

## Performance (continued)

---

- Keep in mind 32-bit operating systems have limitations: data in RAM cannot exceed 2GB (and that includes the software being used); 64-bit Windows is not ready for prime time yet.
- Since most work involves both computation and read/write operations, performance will be specific to the work at hand.
- There is not a best all-around system.
- Shared servers have capacity and speed that is prohibitively expensive with personal workstations.

15

## Main software packages in skew and SSCC

---

- SAS
- Stata
- SPSS (legacy version in skew3 since UNIX version no longer produced)
- Matlab
- Mathematica
- Gauss
- Limdep (legacy version in skew3 since UNIX version no longer produced)
- Lisrel, HLM (SSCC only)
- StatTransfer (skew3, and seldom in SSCC), DBMS/Copy

16

## Software for workstations & laptops

- NU licenses that require annual payment: SAS, SPSS, Mathematica, Maple, Matlab (handled through Kellogg Research Computing). Typical expiration date is August 31, 2006 (regardless of when you paid for it).
- NU site licenses that require no payment: Secure Shell SSH, X-Win32, EndNote
- Kellogg site licenses that require no payment: z-tree, MediaLab, DirectRT
- Special discount plans: Stata, StatTransfer (Kellogg Stata GradPlan)
- Other: Scientific WorkPlace/Word group order (to get volume discounts) when there is a new release

17

## Key corporate data holdings

- Accounting data: Compustat (Global, North America), Datastream, Global Access, Economatica
- Stock market: CRSP daily/monthly stocks, Datastream, TAQ, ISSM
- Bonds: FISD/Mergent, CRSP US Treasuries, Moody's DRS, TRACE
- Derivatives: OptionMetrics, R&C Futures, Datastream
- Mutual funds: CRSP Mutual Fund database, Thomson Financial Mutual Fund Holdings, MFlinks
- Analyst forecasts: Zacks, First Call, ValueLine
- Corporate governance: IRRRC (Governance, Directors, Dilution), Execucomp, Thomson Financial Insiders ("Lancer Analytics"), Hemscott Executive Compensation
- Financial transactions (mergers, acquisitions, initial public offerings, seasoned equity offerings, bankruptcies): SDC Platinum, most dates available in CRSP events files.
- Ownership: Thomson Financial 13F filings, Mutual Fund Holdings, Insiders, IRRRC Directors, ExecuComp, Hemscott, Blockholders

18

## Some macro data

---

- World Bank's World Development Indicators (Library)
- IMF's International Financial Statistics
- DRI/Basic
- Datastream
- Other data from multilateral organizations: check Library Gov Docs section

[www.library.northwestern.edu/govinfo/index.html](http://www.library.northwestern.edu/govinfo/index.html)

19

## NU subscriptions

---

- Inter-University Consortium for Political and Social Research (ICPSR): [www.icpsr.umich.edu](http://www.icpsr.umich.edu)
  - Includes many datasets from the Bureau of the Census: CPS, County Business Patterns...
  - Includes Panel Study of Income Dynamics (PSID), General Social Survey (GSS), World Values Survey, Eurobarometer surveys, etc
  - Can download directly with your NU email address
- Roper Institution: [www.ropercenter.uconn.edu](http://www.ropercenter.uconn.edu)
  - Poll data
  - Can download directly with "Roper Express" service (IP authentication)
- Subscriptions managed by the Library's Social Science Data Services (Kathleen Murphy)

20

## Getting new datasets

---

- Research budgets
- Department funds
- Dean's Office will match for school wide licenses
  - Talk to your chair, researchers with similar interests
  - Approach Research Computing
  - Why is this dataset interesting (especially if we have a similar resource)? Have a description at hand.
  - Have time horizons in mind: how long is the data going to be needed, are periodic updates needed

21

## Resource ("library") room – Jacobs 4219

---

- Documentation (user guides, related books) about the statistical / mathematical software and datasets available at the school
- One week borrowing – renewable unless someone is waiting for the title.
- Please remember to sign out and return titles – we rely on good citizenship.
- Suggestions for titles are welcome.

22

## Experimental research facilities

- Subject pool (currently about 3,500 subjects)
- 2 laboratories for computer based experiments (3<sup>rd</sup> and 4<sup>th</sup> floors)
- 6 break out rooms (3<sup>rd</sup> floor) with video cameras
- Audio/video equipment
- Web surveys: Cogix ViewsFlash (hosted by Kellogg), Qualtrics (hosted by vendor)  
[www.kellogg.northwestern.edu/kis/websurveys/](http://www.kellogg.northwestern.edu/kis/websurveys/)
- Software: MediaLab, DirectRT, z-tree, Multistage. Custom programs in languages such as C, java, or perl.

23

## Behavioral labs



24

## Institutional Review Board (IRB)

---

- [www.research.northwestern.edu/research/OPRS/irb/](http://www.research.northwestern.edu/research/OPRS/irb/)
- Get IRB certification by taking the CITI BASIC course: [www.citiprogram.org/](http://www.citiprogram.org/)
- When submitting a research proposal to the IRB...
  - Be certified!
  - Fill all the fields required
  - Check whether you qualify for an expedited review
  - Make sure your PhD students have a faculty as Principal Investigator

25

## Some useful utilities

---

- **SSH Secure Shell**
    - Encrypted communication
    - Very efficient (does not use a lot of resources)
    - Text-mode
  - **Starnet X-Win32**
    - Allows graphical interface emulation
    - Requires VPN from off-campus
    - Can be slow due to rendering of graphics
- [www.kellogg.northwestern.edu/researchcomputing/terminal-emulators.htm](http://www.kellogg.northwestern.edu/researchcomputing/terminal-emulators.htm)

26

## Some useful utilities (continued)

---

- **Endnote**
  - Site license provided by the Library
  - Bibliography management package
  - Can connect to various databases to retrieve citations
  - Can use it to format your bibliography in Word and also to generate a BibTeX file for use with LaTeX.
  - Can create keywords to categorize citations in the database
- **StatTransfer**
  - In skew3 or seldon (SSCC); also available for purchase
  - Translates datasets between different formats

27

## Full text data sources

---

- Many journals online
- Search NUcat for links
- Some resources omitted from library catalog
- Important databases:
  - Abstracts/citations: EconLit, PsychInfo, Sociological Abstracts, Social Science Citation Index (Web of Science)
  - Working papers: subscribe to SSRN (<http://www.kellogg.northwestern.edu/researchcomputing/ssrn.htm>)

28

## Full text data sources (continued)

---

- Firm “histories”, lists of subsidiaries and joint ventures
  - Mergent Online (FIS Online)
  - S&P NetAdvantage
  - Hoover’s Online
- Analyst reports
  - Investext Plus
  - Thomson ONE

29

## Research assistants

---

- Research Computing does not have a list; sometimes, based on work with doctoral students, we know what skills they have
- Paid from research budget
- Undergraduate RAs
  - Work study students  
<http://ug-finaid.northwestern.edu/work-study/jobs/index.html>
  - Kellogg Honors Research Assistants (KHRA)  
<https://vs.kellogg.northwestern.edu/khra/discussion/default.aspx>
    - We are helping to structure database based on criteria used by faculty
    - We hope to open a sign up system to graduate students, too

30

## Committees

- Research Computing committee:** Patricia Ledesma (chair), Kathleen Hagerty (Senior Associate Dean: Faculty & Research), David Keown (Assistant Dean, KIS), Alain Bonacossa, Damba Lkhagvasuren, Alex Chernev (Mktg), Tom Hubbard (M&S), Ithai Stern (MORS), Beverly Walther (AIM), Annette Vissing-Jørgensen (Fin), Bob Korajczyk (Fin), Bob McDonald (Fin), Vincent Nijs (Mktg),
- Behavioral Research committee:** Kathleen Hagerty (Senior Associate Dean: Faculty & Research), Patricia Ledesma, David Keown (Assistant Dean, KIS), Alain Bonacossa, Damba Lkhagvasuren, Angela Lee Mktg), Katherine Philips (MORS), Paola Sapienza (Fin), Brian Rogers (MEDS)

31

## Ready for lunch?

- Any other questions?
- Thank you
- Welcome to Kellogg!
- Where are we?

Who	Room	Phone	Email alias
Alain	4242	7-1854	a-bonacossa
Damba	4222	7-4070	d-lkhagva
Patricia	4218	7-7658	pledasma

<http://www.kellogg.northwestern.edu/researchcomputing/>

32