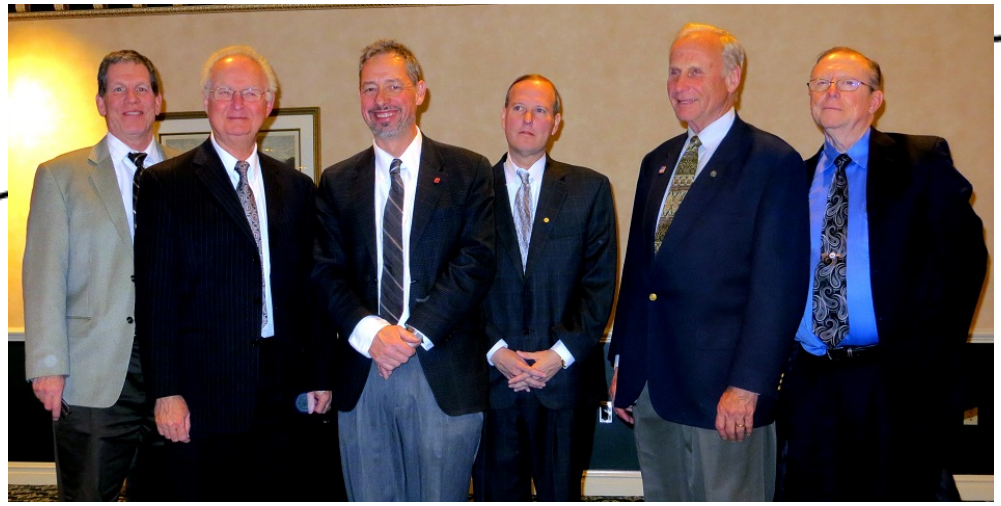
Big Data – Perspectives of the IEEE Reliability Society

Christian K. Hansen, PhD
President, IEEE Reliability Society
EWU Associate Dean for Computing and Engineering Science
and Professor of Statistics

IEEE Big Data Workshop, Oct 1-2, 2014

IEEE ReliabilitySociety EASTERN WASHINGTON UNIVERSITY start something big

IEEE Reliability Society



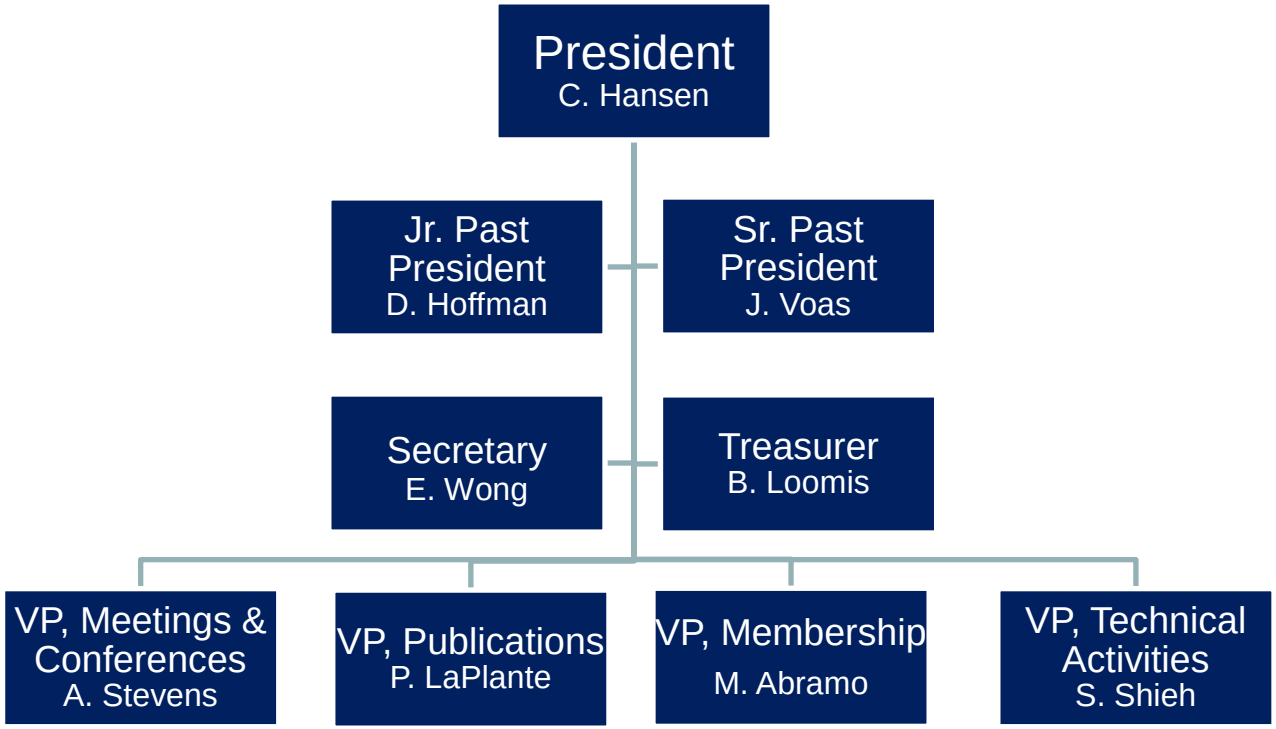
ReliabilitySociety

Visit us at **rs.ieee.org**

- IEEE Society # 7 founded 1949.
- Field of Interest (FoI): Reliability, quality, maintainability, safety, security and privacy, PHM and more (specialty engineering disciplines = TRUST).
- 27 active chapters and members in 60 countries.
- Two new chapters in China: Beijing and Nanjing (Changsha, Guangzhou and Xi'an chapters are currently pending)
- Small number of members (2000+) but big impact through publications and conferences.

Reliability Society Organization

start something big



IEEE Society/Council Divisions

Division I
IEEE Circuits and Systems; IEEE Council on Electronic Design Automation; IEEE Electron Devices Society IEEE Nanotechnology Council; IEEE Solid-State Circuits Society

Division II
IEEE Components, Packaging, and Manufacturing Technology Society; IEEE Dielectrics and Electrical Insulation Society; IEEE Industry Applications Society; IEEE Instrumentation and Measurement Society; IEEE Power Electronics Society; IEEE Ultrasonics, Ferroelectrics and Frequency Control

Division III
IEEE Communications Society

Division IV
IEEE Antennas and Propagation Society; IEEE Broadcast Technology Society; IEEE Consumer Electronics Society; IEEE Electromagnetic Compatibility Society; IEEE Magnetics Society; IEEE Microwave Theory and Techniques Society; IEEE Nuclear and Plasma Sciences Society; IEEE Council on Superconductivity

Division V
IEEE Computer Society

Division VI
IEEE Education Society; IEEE Industrial Ele … sional Communication Society; **IEEE Reliability Society**; IEEE Society o…

Reliability Society FoI overlaps with every division and virtually every IEEE Society

BIG Data overlaps with many Society FoIs

Division VII
IEEE Power & Energy Society

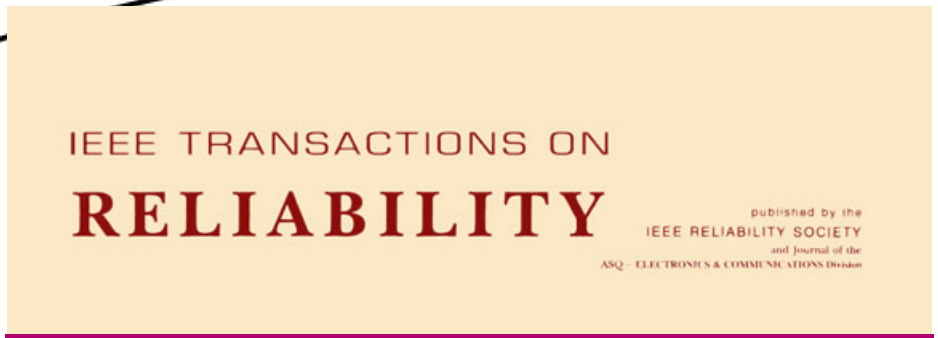
Division VIII
IEEE Computer Society

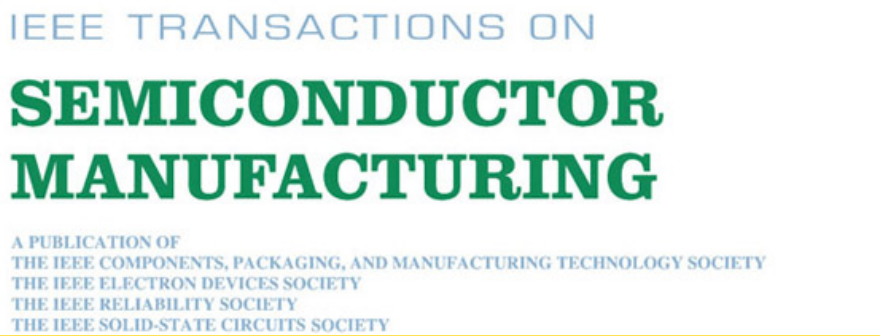
Division IX
IEEE Aerospace and Electronic Systems Society; IEEE Geoscience and Remote Sensing Society; IE… …y; IEEE Intelligent Transportation Systems Society; IEEE Oceanic Engineering Society; IEEE Signal Processing Society; IEEE Vehicular Technology Society

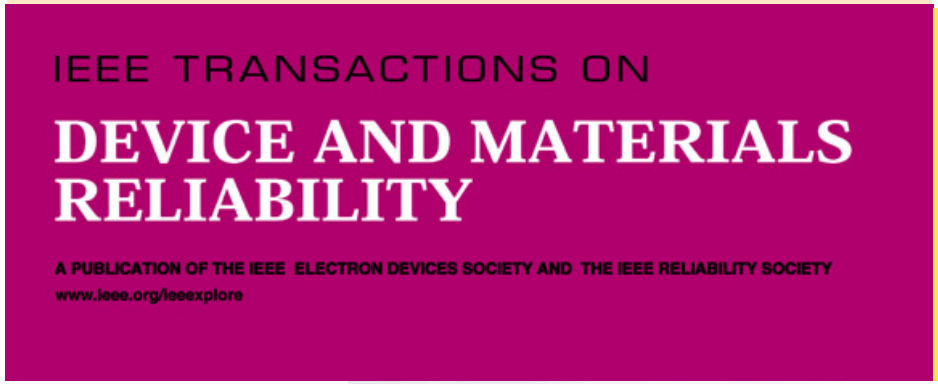
Division X
IEEE Biometrics Council; IEEE Computational Intelligence Society; IEEE Control Systems Society; IEEE Engineering in Medicine and Biology Society; IEEE Photonics Society;IEEE Robotics and Automation Society; IEEE Sensors Council; IEEE Systems, Man, and Cybernetics Society; IEEE Systems Council[6]

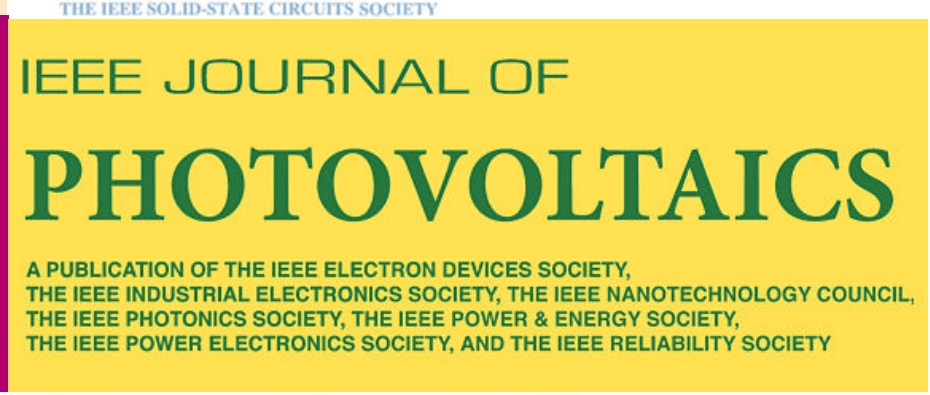
IEEE Reliability Society – Publications

start something big









IEEE Reliability Society – Conferences

Visit us at rs.ieee.org

Select Conferences (sponsored and co-sponsored)

- Annual Reliability and Maintainability Symposium (RAMS)
- International Reliability Physics Symposium (IRPS)
- Conference(s) on Prognostics and Health Management (PHM)
- Conference on Software Security and Reliability (SERE)
- International Conference on Quality Software (QSIC)
- World Forum of Internet of Things (WF-IoT)
- Symposium on Software Reliability Engineering (ISSRE)
- Conference on Trustworthy Systems and Applications (TSA)
- System of Systems Engineering Conference (SoSE)
- Photovoltaic Specialists Conference (PVSC)
- Reliability Science for Advanced Materials and Devices (RSMAD)
- Physical and Failure and Analysis (IPFA)

IEEE Reliability Society is active in BIG Data

start something big

Plenary Speech II: The Role of Trust in the Future World of Internet of Things

Abstract:

Over the past two decades we have seen the internet evolve from being strictly a network of con
towards today's and the future's Internet of Things (IoT), involving a much broader range of conn
generation of internet, the IoT is expe
namely time, money, physical, hum
ivity, issues surrounding trust, includi
ecome sources of ever increasing con
storical trends and challenges related
and discuss areas of most likely researc
on this subject over the next decade. In particular, ongoing efforts and future directions for rese
the IEEE and the IEEE Reliability Society will be presented.

Speaker:

Dr. Christian K. Hansen

CCIOT 2014

Dec 13-14, 2014 Changchun, CHINA

2014 International conference on cloud computing and Internet of things

IEEE Computer Magazine, April 2014

Big Data, Networked Worlds

George F. Hurlburt, *STEMCorp*
Jeffrey Voas, *IEEE Fellow*

BIG Data in the BIG Picture: Challenges and Projections for the Future

ISCC2013 参会指南

Drilling for Useful Information in the Landscape of Big Data

by 李亦昊 • 4 months ago • 144 views

This colloquium is given by Professor **Christian Hansen** President of IEEE Reliability Society Professor and Associate Dean of ...

HD

IEEE Reliability Society is active in Continuing Education

start something big

RS Tutorial Certificate Program on Software Reliability

- A series of tutorials offered at IEEE-RS sponsored software reliability related conferences including:
  - International Conference on Quality Software (QSIC 2014)
  - International Symposium on Software Reliability Engineering (ISSRE 2015)
- Tutorial lectures are available live to conference attendees as well as through recorded video hosted on UT-Dallas server.
- Must attend a minimum of three classes over a two year period
- Sample of topics covered:
  - Introduction to software reliability concepts and analysis techniques
  - Pros and cons of different models and tools
  - Use of techniques such as failure modes, effects and criticality
- For more information visit http://paris.utdallas.edu/RS-Certificate/ or contact Professor W. Eric Wong (ewong@utdallas.edu)

BIG Data Challenge – The STEM Talent Gap

start something big

- Will we have the expertise to manage and analyze the data in the future?
- The US alone needs 140,000 to 190,000 more workers with “deep analytical” expertise and 1.5 million data-literate managers. Lohr (2012)
- Microsoft currently has 6,000 STEM related job openings.
- In Washington State, only 1 out of 100 high school students are projected to earn a STEM related degree. Lazowski (2012)

File Edit View History Bookmarks Tools Help

Big Data Education | What'... | Graduate Programs in Big ... | 500 - Internal server error.

whatsthebigdata.com/2012/08/09/graduate-programs-in-big-data-and-data-science/

University programs in Bi

CanvasUser Dashboard | Google Calendar | Visit Spokane Media | Google Scholar | SharePoint - Home | Welcome - myEWU | RefWorks Login | text2

Search Web

Central Michigan University

Graduate Certificate in Data Mining

Columbia University

Master of Science in Data Science

Master of Science in Information and Knowledge Strategy

Certification of Professional Achievement in Data Sciences

Cornell University

Masters of Engineering in Operations Research and Information Engineering – Data Analytics Concentration

DePaul University

M.S. in Predictive Analytics

Drexel University

Most university programs in BIG Data
Are at the graduate level

How can we prepare undergraduate
Students for careers in BIG Data?

How BIG Data and IoT has changed our lives

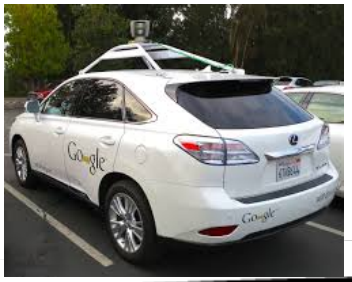
start something big

Old World

- How are you?
- Where is my iPhone?
- Where is my pizza?
- How do I get to your house?

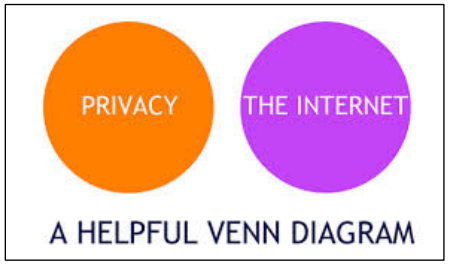
New World

- Where are you?
- Where is my husband?
- What is my heartbeat?
- Drive me to your house!
- **Who do you TRUST?**



Does BIG Data make our lives better?

start something big



The Good News

- Email/SMS has made it much faster an easier to communicate.
- BIG Data enables better decision making for businesses and consumers.
- BIG Data gives consumers more choices

decide. + ebay

Decide is now part of the eBay family. Thank you to all our supporters who made it possible.

The Bad News

- Loss of privacy
- Loss of effective communication
- BIG Data expands opportunities for cybercrime.
- Will the choice of being online/offline exist in future?

BIG Data – Does Bigger Data Mean More Useful Information?

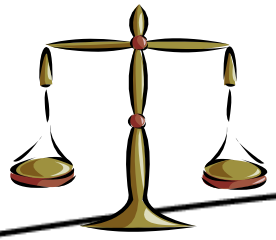
start something big

- If the digital world doubles in size every two years do get twice as much useful information?
- There are about 30 million songs available in Apple's iTunes store. If that number doubled would there be twice as many of your favorite songs to download?

- We want the "needle in the haystack", but we really don't care much about the hay.

- What if Google and YouTube were operated using a TV remote control?

Scalability



start something big

- Computation time (in seconds) can be reduced by a factor *n* when computing is distributed among *n* identical computers.
- Computation time (on each machine) often measured in terms of “Flops” = # Floating Point Operations
- **Easily scalable:** Repeated simulation runs
- **Not scalable:** Iterative or recursive computations where each computation requires the result of the previous iteration.
- Yahoo uses more than 4000 **commodity** computers, connected with high speed network, Tian (2013).

BIG Data – Predictive Analytics Using Linear Models

More observations drive more complex calculations

General Linear Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $\varepsilon_i \sim N(0, \sigma^2)$ (i.i.d)

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1k} \\ x_{21} & & x_{2k} \\ \dots & \dots & \dots \\ & & \\ x_{n1} & \dots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \dots \\ \beta_{k-1} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \dots \\ \varepsilon_n \end{bmatrix}$$

Least Squares Estimation: $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}$

Matrix Operations – Scalable?

Matrix Multiplication $\mathbf{A} \cdot \mathbf{B}$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ a_{21} & & a_{2k} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nk} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & \dots & b_{1p} \\ b_{21} & & b_{2p} \\ \dots & \dots & \dots \\ b_{k1} & \dots & b_{kp} \end{bmatrix}$$

Requires $k\,p\,(2n\text{-}1)$ FLOPs (Floating Point Operations)

Can be distributed to up to k x p x n computers

Matrix Inversion – Scalable?

Matrix Inversion: Solve $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}$ $\mathbf{B} = \mathbf{A}^{-1}$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \dots & & \dots \\ a_{k1} & \dots & a_{kk} \end{bmatrix}$$

Naive Gauss Elimination requires $\propto k^4$ FLOPs (for large *k*)

LU Decomposition Method requires $\propto k^3$ FLOPs (for large *k*)

Q-R Method helps reduce rounding errors

Only partially scalable – all methods require recursive algorithms

Drilling for Big Answers Using BIG Data

start something big

A single airplane dumbs approximately 20 TB of data during one hour of flight time*

facebook

500+ TB of data per day*

If ML Flight 370 did indeed crash, it stopped transmitting data well before it happened.

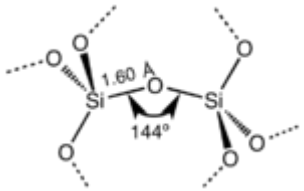
Satellite image data much more difficult to mine!!

*SOURCE: Big Data and Hadoop Tutorial – Edureka.in

BIG Data – Moore's Law

start something big

- Will the volume of the digital world continue to double every two years?
- Will the cost of storage continue to be reduced by a factor 2 or more every two years?
- Will I ever own a cellphone with a Terabyte internal memory? A Petabyte?
- Will Intel develop a processor with 1 THz clock speed?



0.3 lbs = 0.13kg

1 TB = 1 Trillion bytes = 8 Trillion flip-flops
Physical Dimensions of HD: 82 x 15 x 111 mm = 1.365 10^{-4} m^3
Current CMOS technology is 25 nm = 2.5 10^{-8} m diameter per memory cell
Diameter of a silicon molecule is in the order of 160 pm = 1.6 10^{-10} m.

If my house was a hard disk …

start something big

1984:

College student

Dorm Room

100 ft^2

If my house was a hard disk …

1994: Young Professional

5 bedroom house

2 car garage

3200 ft^2

0.25 acre lot

If my house was a hard disk …

2004: Full Professor

160 bedroom estate

64 car garage

100,000 ft^2

8 acre lot

(about the size of a “Walmart-type” store)

If my house was a hard disk …

start something big

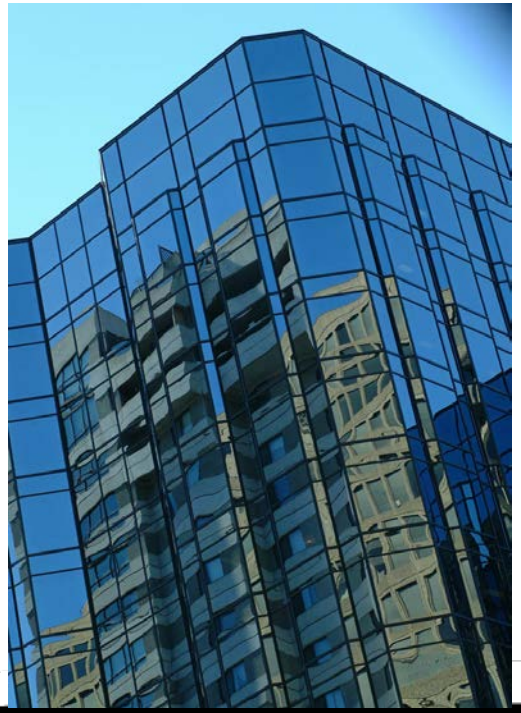
2014: Associate Dean

5000 bedrooms

2000 car parking garage

3.2 million ft^2

250 acre lot



(probably bigger than a New York City sky scraper)

Summary – Challenges and Projections

- The deployment of the Internet of Things (IoT) will impact BIG Data and how it will grow.
- Ratio of valuable to useless data/content is likely to decline.
- Reduction in size of memory cells will eventually be limited by physical constraints.
- TV, Internet and Communication will eventually converge into one technology (long overdue!)
- Many new applications of BIG data will be developed over the next decade, e.g. higher education and medicine.

Acknowledgements

Dr. Eric Wong
UT-Dallas

Dr. Jeff Voas, NIST

Dr. Yves Nievergelt, Professor of Mathematics
Eastern Washington University

Dr. Daniel Canada, Professor of Mathematics
Eastern Washington University

Dr. Tony Tian, Assistant Professor of Computer Science
Eastern Washington University

References

start something big

- Boyd, D. and Crawford, K. (2011), Six Provocations for Big Data. *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011.
- Hurlburt, G.F. and Voas, J.M. (2014) “Big Data, Networked Worlds” *IEEE Computer Magazine*, April 2014.
- Lohr, Steve. (2012). The age of big data. *New York Times* November, 2012.
- Nievergelt, Yves. "Total least squares: State-of-the-art regression in numerical analysis." *SIAM review* 36.2 (1994): 258-264.
- Sagiroglu, S.; Sinanc, D. (2013). Big data: A review, *Collaboration Technologies and Systems (CTS), 2013 International Conference on* , vol., no., pp.42,47, 20-24 May 2013
- Tian, T. (2013), Lecture Notes for course CSCD 467 Concurrent Systems
- Trahan, J., Kaw, A.,Martin, K.; Computational Time for Finding the Inverse of a Matrix: LU Decomposition vs. Naive Gaussian Elimination http://nm.mathforcollege.com/simulations/nbm/04sle/nbm_sle_sim_inversecomptime.pdf