The Center for Dynamic Data Analytics Quarterly Newsletter (Public Edition)

Edition: Second Quarter, 2014 | Published April 17, 2014
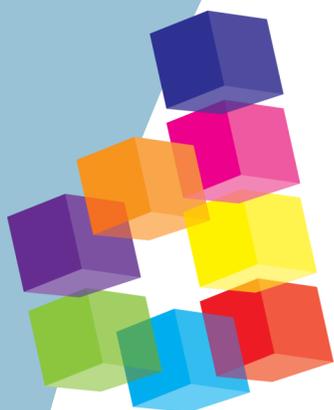
Contents

- Announcements
- Current Projects
- Projects Available for Collaboration
- Students Graduating
- Big Data News
- Upcoming Conferences
- Student Spotlight

Announcements

-The CDDA Spring 2014 Workshop and IAB Meeting is taking place on May 7-8th in Stony Brook. Please register here. The event is being co-promoted with the invitation-only XLDB Healthcare Workshop. To request an invite, please contact Rong Zhao.

-The University of Virginia, the Stevens Institute of New Jersey and George Mason University are set to speak at the upcoming Spring 2014 Workshop and IAB Meeting at Stony Brook. Peter Beling from UVA, Steve Yang from Stevens and Phil Auerswald from GMU will discuss the capabilities, resources and industrial connections they would like to bring to the CDDA.

Current Projects

◊   Developing Security Domain Ontologies (recently completed): PI, Paul Fodor

◊   Automatic Generation of Virtual Troubleshooting Systems Using Ontologies (recently completed): PI, Paul Fodor

◊   Tissue Quantification Project: PI, Dimitris Metaxas

◊   Anomaly Detection in Dynamic Networks: PI, Leman Akoglu

(Public details pending on further active projects)

Projects Available for Collaboration

⇒   Statistical Stylometry: PI, Yejin Choi, Stony Brook

⇒   Big Graph Mining: PI, Tina Eliassi-Rad, Rutgers

⇒   Privacy Preserving Data Mining: PI, Jaideep Vaidya, Rutgers

⇒   The Reality Deck - 1.5 Gigapixel Display: PI, Arie Kaufman, Stony Brook

⇒   4D Cardiac Fluid Flow Modeling: PI, Dimitris Metaxas, Rutgers

Screenshot of a real-time wind speed visualization. Viewable live here.

Students Graduating

- **Zhiyuan Zhang** - PhD, Computer Science, Stony Brook. Zhiyuan will be graduating in May, 2014.

  Email: zyzhang@cs.stonybrook.edu   For further info on Zhiyuan, please visit:

  http://www.cs.sunysb.edu/~zyzhang/

- **Timothy Shields** - PhD, Computer Science, Rutgers. Timothy will be graduating in May, 2014

  Email: timothy.shields@live.com   For further info on Timothy, please visit:

  http://www.linkedin.com/pub/timothy-shields/48/80/a23

"With too little data, you won't be able to make any conclusions that you trust. With loads of data you will find relationships that aren't real… Big data isn't about bits, it's about talent"

-Douglas Merrill

Big Data News

- White House — The FY 2015 Science and Technology R&D Budget
- White House — Tell Us What You Think About Big Data and Privacy
- The Performance Analytics and Decision Support (PADS) Framework
- Big Data Analytics Master's Degrees: 20 Top Programs
- IBM Watson Takes Aim at Cancer
- Hadoop vs. Redshift
- New Book: Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities

Upcoming Conferences

- 5/1-2/14: San Francisco, CA—The Innovation Enterprise: Sentiment Analysis Innovation
- 5/7-8/14: Stony Brook, NY - CDDA Spring 2014 Workshop and IAB Meeting
- 5/12-15/14: Toronto, ON—PAW: Predictive Analytics World
- 5/15-16/14: Philadelphia, PA—The Innovation Enterprise:

  Big Data and Analytics in Healthcare
- 6/10-11/14: Boston, MA—Data Driven Business: Useful Business Analytics Summit
- 6/11-12/14: Philadelphia, PA—The Innovation Enterprise:

  Big Data and Analytics in Pharma
- 6/16/14: NYC, NY - NYU, CUSP—Workshop on Privacy, Big Data, and the Public Good: Frameworks for Engagement
- 6/27—7/2/14: Anchorage, AK - IEEE—3rd International Congress on Big Data

OpenCB is open-source software for computational biology. Link here.

Student Spotlight

Zhiyuan Zhang is a PhD candidate in the Computer Science Department at Stony Brook University. Zhiyuan is graduating in May, 2014. He received his B.E. and M.E. degree in Computer Science from Shandong University, China. His research interests are visual analytics and information visualization, with a special focus on healthcare informatics, multivariate data visualization, and correlation analysis. During his graduate studies, he was awarded the Presidential Fellowship of Stony Brook University in 2008 and the IBM PhD Fellowship in 2013-

2014. He has been working as the chief architect and the developer in several research projects. His work on Visual Analytics in Health Care builds a system for clinical encounters using visualization and machine learning techniques. The system unifies all the patient's electronic medical record (EMR) information fragments into a single interactive visual framework. It seeks to improve the usability of information captured in the EMR, to help the physicians get a quick overview of the patient history, and to lower the time and effort needed to access the medical patient information required to arrive at a diagnostic conclusion. His work on Visual Association Mining of Multivariate Data implements a visual analytics framework that supports an overview

and detail-on-demand for the multivariate data. The interactive user interface provides an intuitive way for illustrative data representation, visual storytelling, and navigation through multi-dimensional data spaces. Geo-spatial data visualization is supported via a Google Earth plugin. This work has been applied to address real scientific problems in climate research and it already helped a team of climate researchers make a few important discoveries. The framework is currently being extended to help analysts gain valuable insights from large amounts of multivariate data using interactive visual correlation and causation analysis. Further information on Zhiyuan's background and research can be found here.
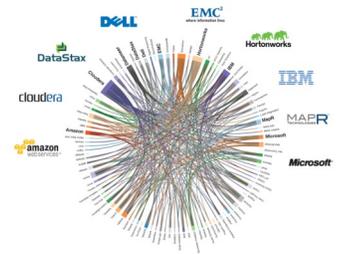
A visualization of the Hadoop ecosystem. Link here.

> "Listening to the data is important… but so is experience and intuition. After all, what is intuition at its best but large amounts of data of all kinds filtered through a human brain rather than a math model?"
>
> -Steve Lohr

The Four V's of Big Data infographic, showing Volume (Scale of Data), Velocity (Analysis of Streaming Data), Variety (Different Forms of Data), and Veracity (Uncertainty of Data).

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS
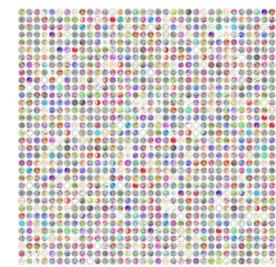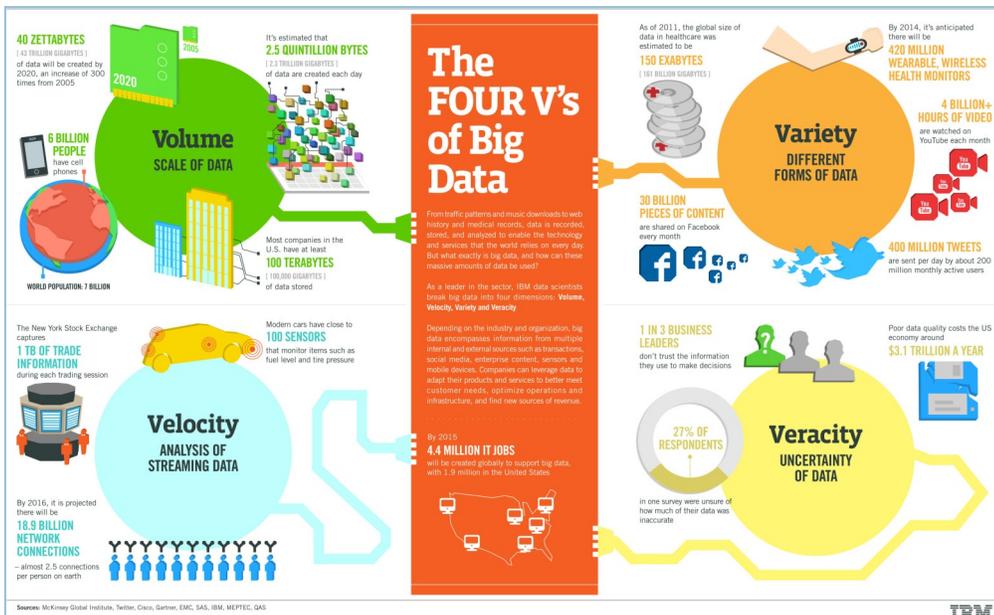
Tabular Data Visualization by Circos. Link here.

The Four V's of Big Data *Infographic courtesy of* IBM
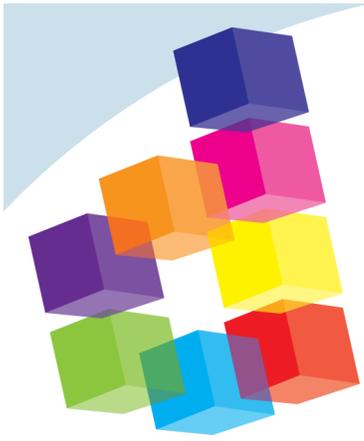
The Center for Dynamic Data Analytics

Rutgers Address:	Stony Brook Address:
617 Bowser Rd.	1500 Stony Brook Rd.
Piscataway, NJ	Stony Brook, NY
08854	11794

E-mail: james.mielke@rutgers.edu
Phone: 848-445-8824

E-mail: rzhao@cs.stonybrook.edu
Phone: 631-632-4627

"From Chaos to Knowledge"

cdda.rutgers.edu

About CDDA

The Center for Dynamic Data Analytics (CDDA) is a National Science Foundation (NSF) sponsored Industry and University Cooperative Research Program (I/UCRC) established between Rutgers University and the State University of New York (SUNY), Stony Brook.

The motivation for this center is the lack of scalable algorithms, methods and solutions for addressing the ever increasing amounts of industry-related data. The focus is on data sets that are massive, dynamic, complex and multidimensional, or what is commonly known as Big Data analytics. The goal of the center is to discover, develop and apply data analytics solutions to industry problems such that the chaotic data is transformed into knowledge and industry products.

NSF Factsheet—CDDA

CDDA Partners and Sponsors