



Center for Dynamic Data Analytics (CDDA)

*An NSF Supported
Industry / University Cooperative Research
Center (I/UCRC)*



University Consortium

Stony Brook University¹ and Rutgers University²
Co-Directors:

¹Arie Kaufman

**Distinguished Professor & Chair
Computer Science Department
Chief Scientist, Center of Excellence
in Wireless & IT (CEWIT)
Stony Brook University**

²Dimitris Metaxas

**Distinguished Professor
Department of Computer &
Information Science
Director, CBIM Center
Rutgers University**

CDDA Mission

- Conduct integrative, multi-disciplinary research to **manage, analyze, and visualize** massive, complex, multidimensional and multi-scale dynamic data in order to **turn chaos into knowledge** and **unleash the transformative potential of big data** in a wide range of application domains, such as IT, healthcare, pharmaceutical, biotechnology, commerce, retail, finance, insurance, media, entertainment, transportation, logistics, manufacturing, defense, security, education, and public administration.

What Is Big Data?

*“**Big data** refers to datasets whose **size is beyond the ability of typical database software tools to capture, store, manage, and analyze.**”*

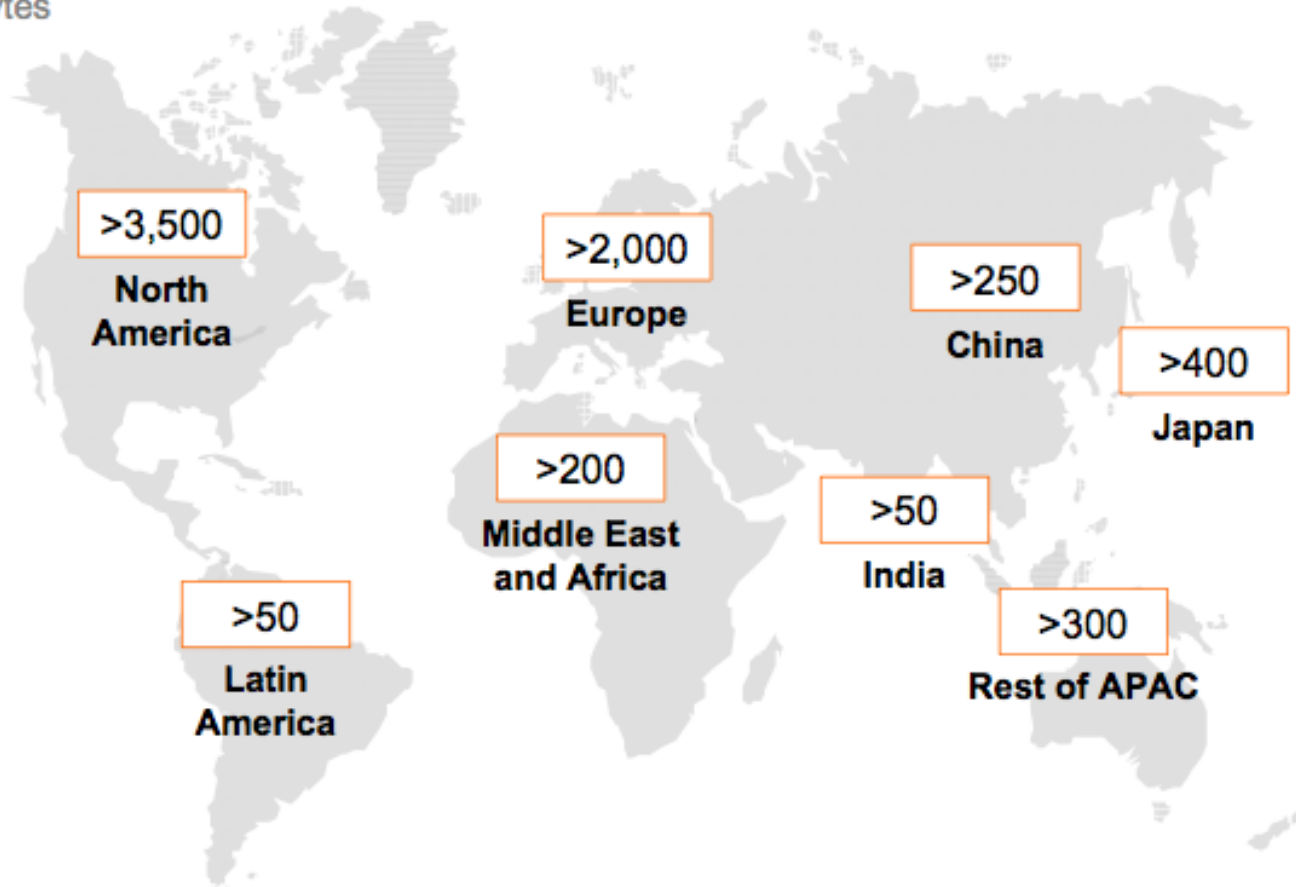
*“**Big data** can be used to create value across sectors of the global economy ... we are on the cusp of a tremendous **wave of innovation, productivity, and growth, as well as new modes of competition and value capture** — all driven by **big data** as consumers, companies, and economic sectors exploit its potential ... that the scale and scope of changes that big data are bringing about are at an **inflection point, set to expand greatly, as a series of technology trends accelerate and converge.**”*

“Big Data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011

Big Data Is Growing Fast

Amount of new data stored varies across geography

New data stored¹ by geography, 2010
Petabytes



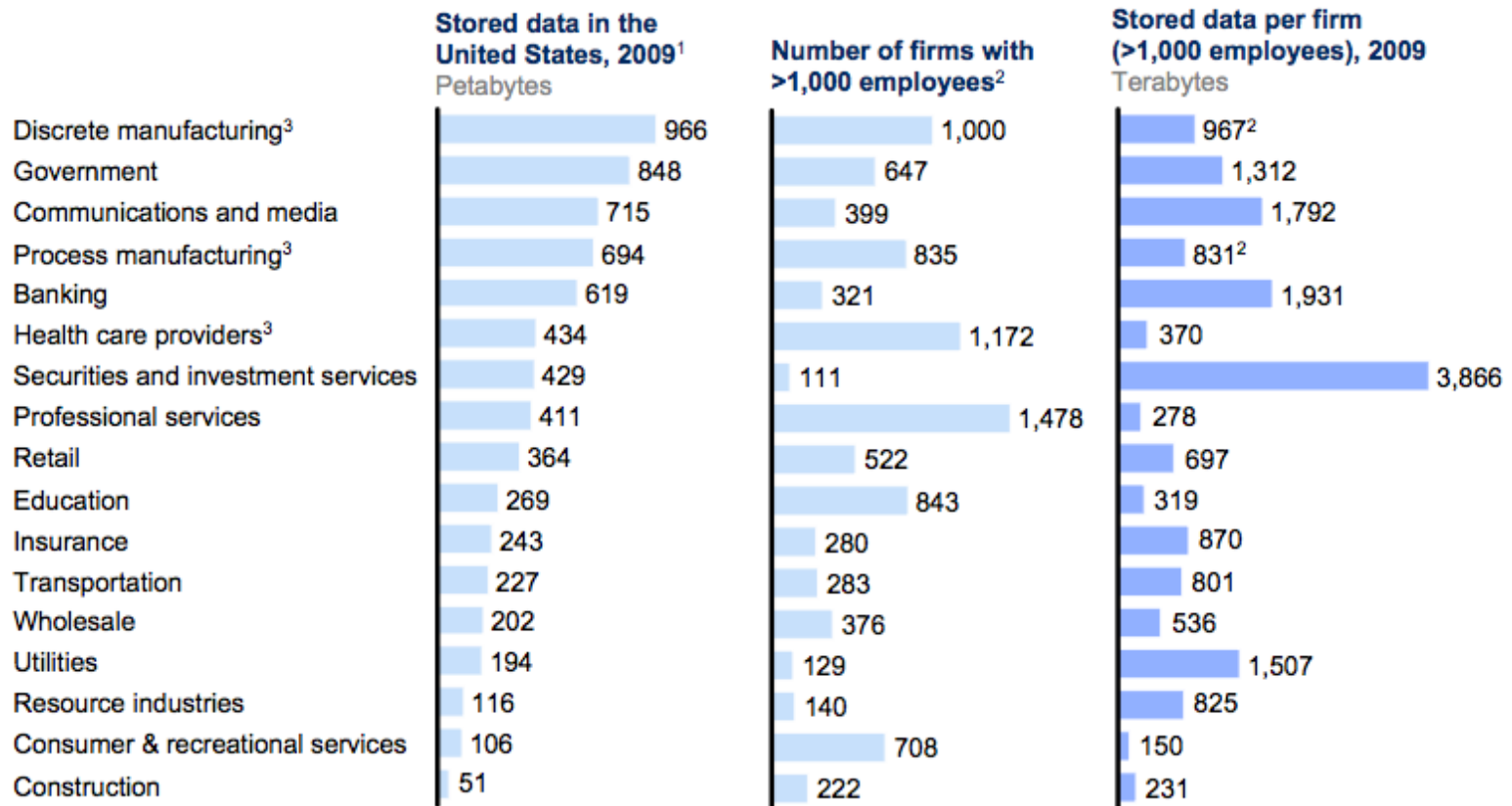
¹ New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.

SOURCE: IDC storage reports; McKinsey Global Institute analysis

“Big Data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011

Big Data Is in Many Industries

Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.

2 Firm data split into sectors, when needed, using employment

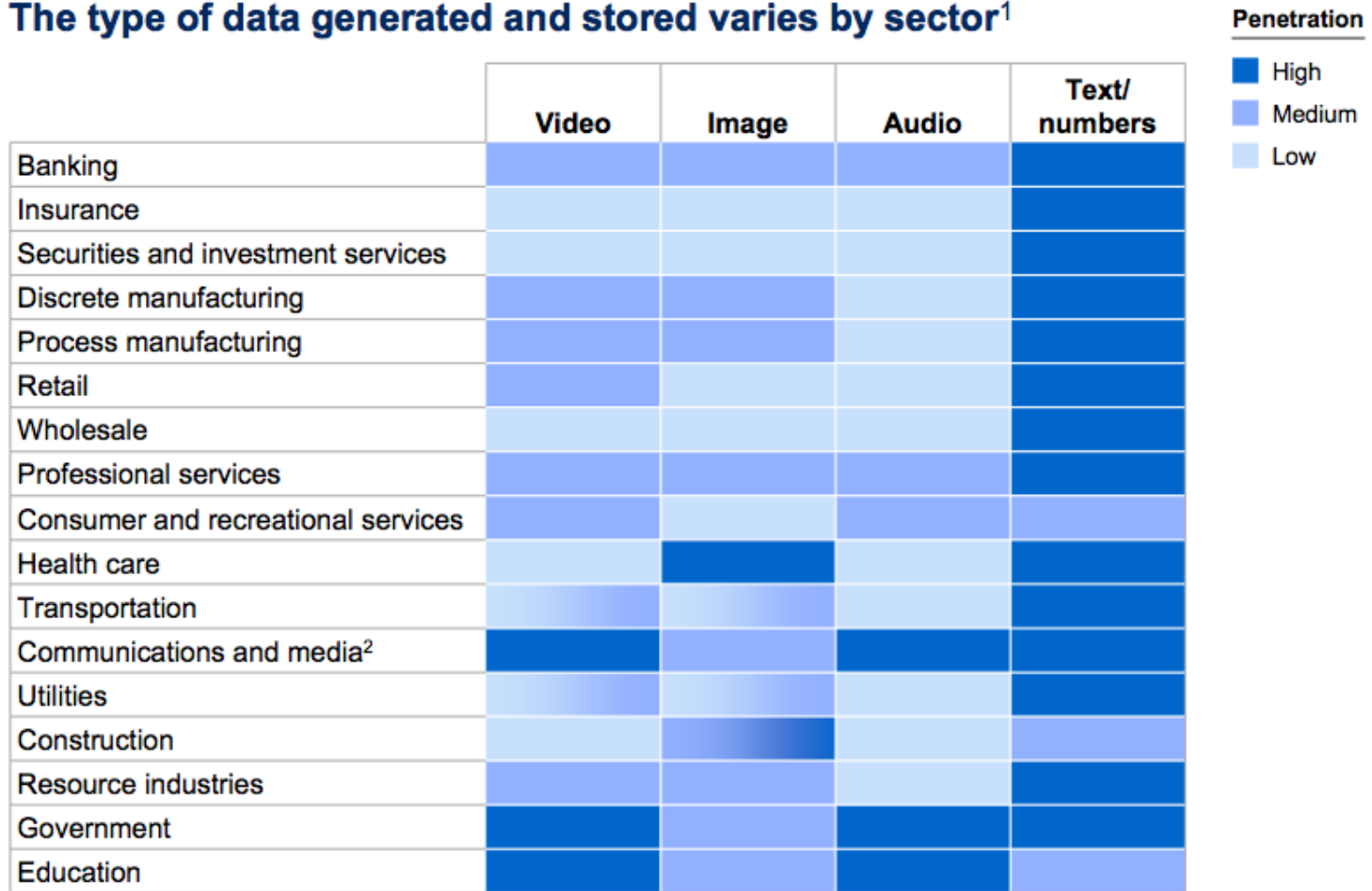
3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

“Big Data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011

Big Data Is in Many Forms

The type of data generated and stored varies by sector¹



¹ We compiled this heat map using units of data (in files or minutes of video) rather than bytes.

² Video and audio are high in some subsectors.

SOURCE: McKinsey Global Institute analysis

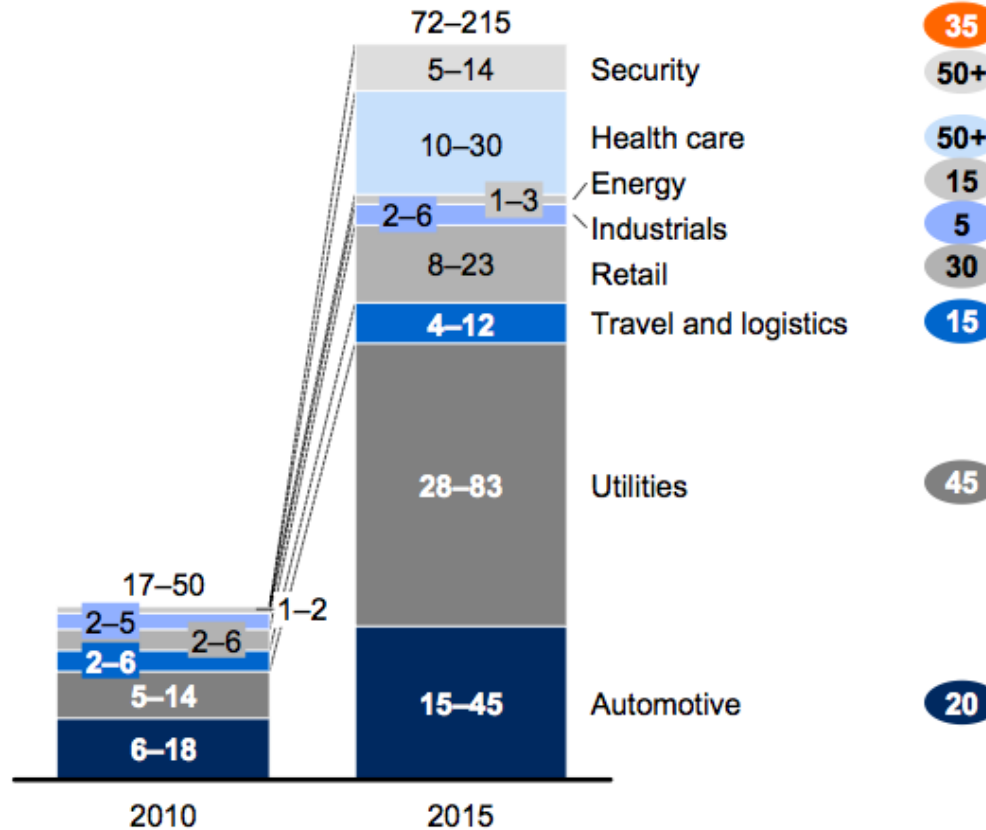
Big Data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011

Big Data and the Internet of Things

Data generated from the Internet of Things will grow exponentially as the number of connected nodes increases

Estimated number of connected nodes
Million

Compound annual
growth rate 2010–15, %



NOTE: Numbers may not sum due to rounding.

SOURCE: Analyst interviews; McKinsey Global Institute analysis

Big Data Has Big Potential

Big data can generate significant financial value across sectors



US health care

- \$300 billion value per year
- ~0.7 percent annual productivity growth



Europe public sector administration

- €250 billion value per year
- ~0.5 percent annual productivity growth



Global personal location data

- \$100 billion+ revenue for service providers
- Up to \$700 billion value to end users



US retail

- 60+% increase in net margin possible
- 0.5–1.0 percent annual productivity growth



Manufacturing

- Up to 50 percent decrease in product development, assembly costs
- Up to 7 percent reduction in working capital

SOURCE: McKinsey Global Institute analysis

Big Data: The next frontier for innovation, competition and productivity,” McKinsey Global Institute, May 2011

What Does the CDDA Do?

- Create a partnership between **academia, industry and government** to advance dynamic data analytics and to address big data challenges in the application domains
 - Understand the technology needs and evaluate real-world problems through constant, direct contact with industry partners
 - Identify pre-competitive research themes and conduct collaborative, interdisciplinary research
 - Produce highly trained students with advanced knowledge and hands-on skills in pertinent technologies
 - Educate policy makers about the emerging importance of dynamic data analytics and the transformative potential of big data to ensure and sustain our nation's competitive advantage

Stony Brook Resources & Experience

- **Center of Excellence in Wireless & Information Technology (CEWIT)**
- **Advanced Energy Research & Technology Center (AERTC)**
- **Center for Advanced Technologies (CAT) in Sensors**
- **Center for Advanced Technologies (CAT) in Biotechnology**
- **Center for Visual Computing (CVC)**
- **Center for Cloud Computing (C³)**
- **Center for Cyber Security (CCS)**
- **Strategic Partnership for Industrial Resurgence (SPIR) Program**
- **Brookhaven National Laboratory (BNL)**
- **New York Blue Supercomputer (>100Tflops)**
- **New York Center for Computational Science (NYCCS)**
- **Consortium for Digital Arts, Culture, and Technology (cDACT)**
- **Stony Brook University High Technology Incubators**
- **I/UCRC in Bioenergy Research and Development**
- **Brookhaven National Laboratory (BNL)**
- **Cold Spring Harbor Laboratory (CSHL)**

Stony Brook Niche

- **Managing Big Data**

- Cloud Computing
- Distributed and scalable storage systems
- Cybersecurity
- Wireless sensor networks
- RFID

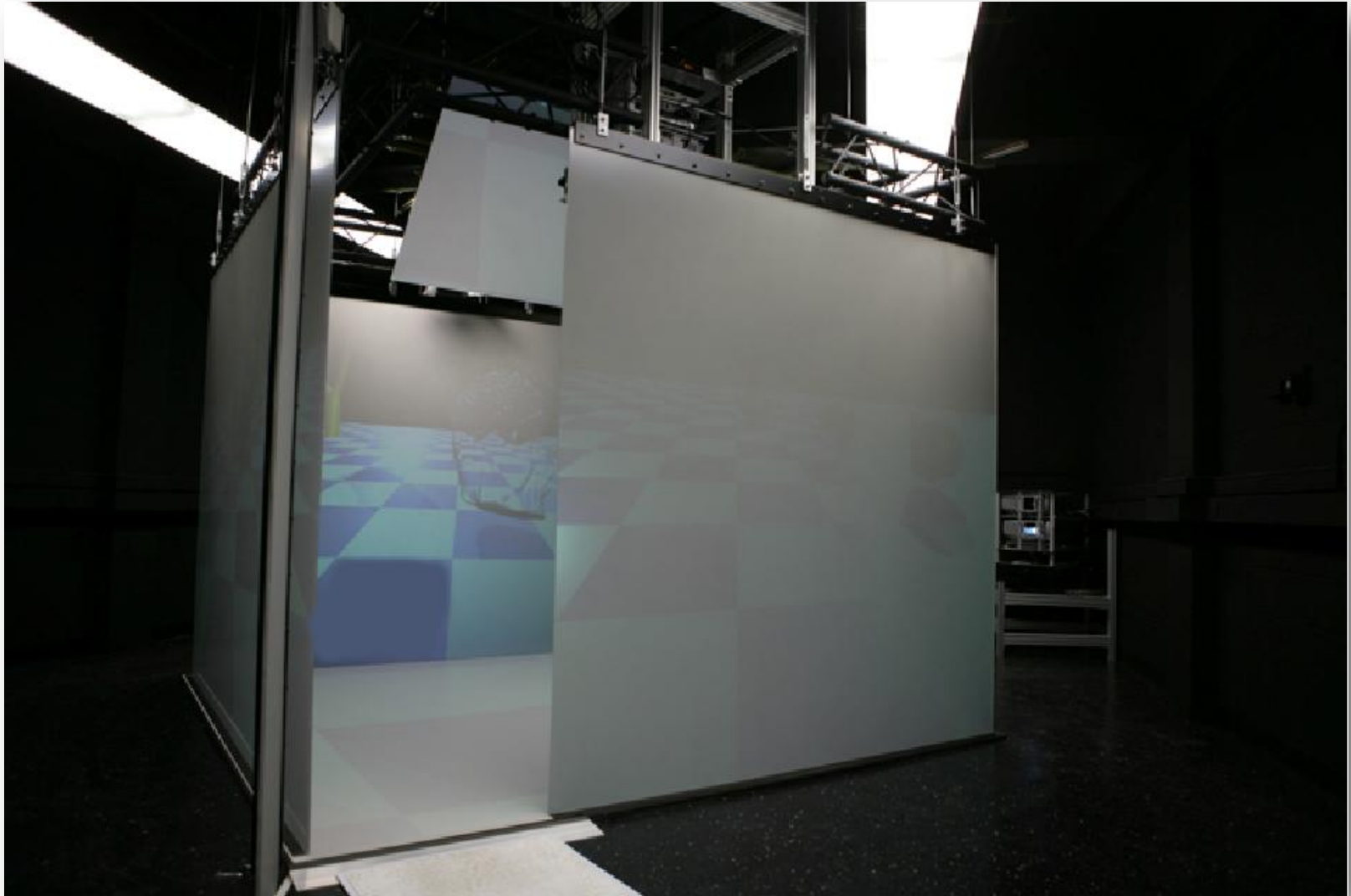
- **Analyzing Big Data**

- Machine learning and logic programming
- Natural language processing and text mining
- Semantic Web
- Sentiment analysis
- Image and video processing
- Biomedical imaging
- Medical informatics

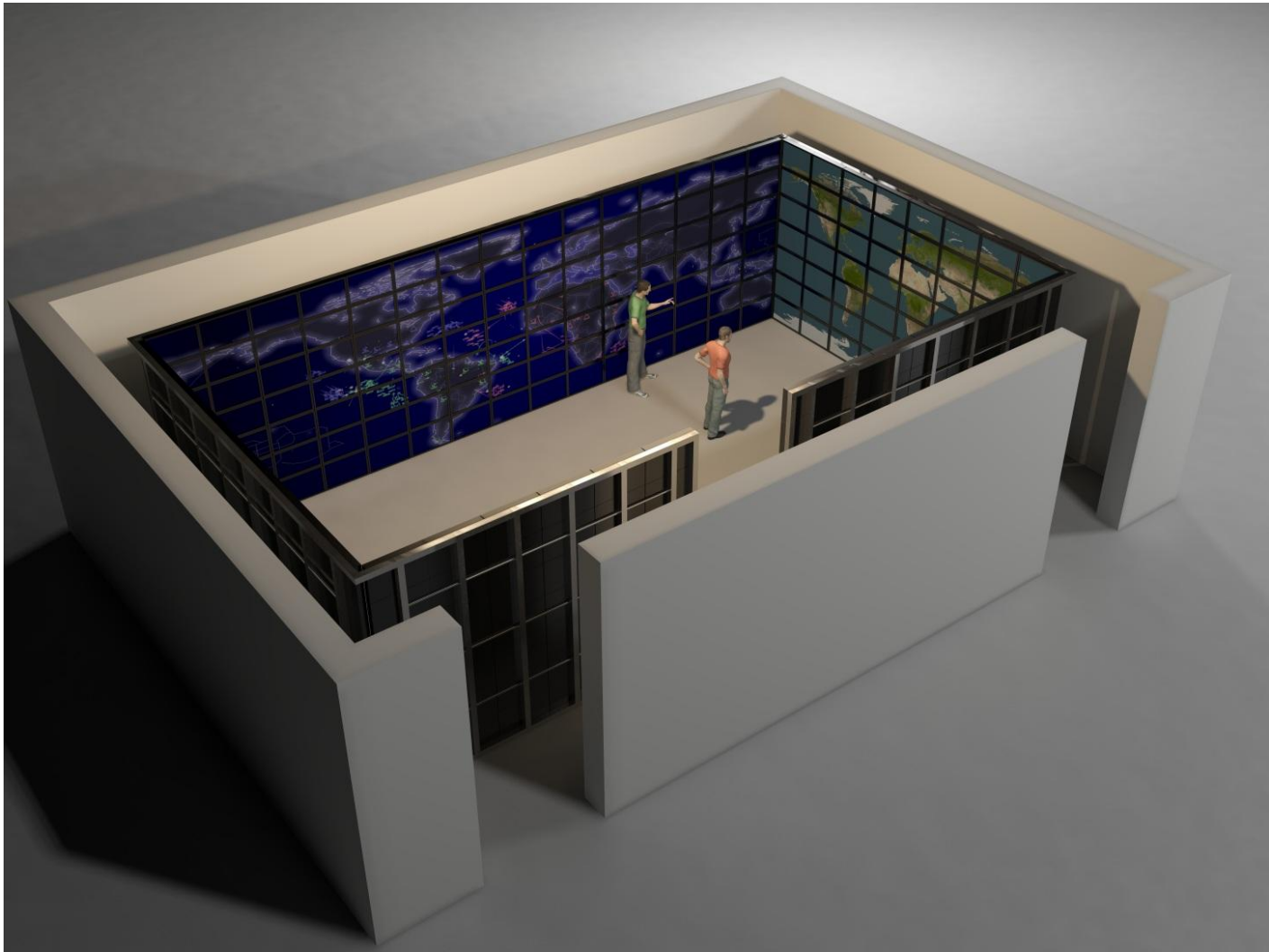
- **Visualizing Big Data**

- Visual analytics
- Human-computer interaction
- Immersive Cabin and RealityDeck

Immersive Cabin



Reality Deck



1.5 Billion pixels

Reality Deck with First Immersive Image



Rutgers Resources & Experience

- **Center for Computational Biomedicine, Imaging and Modeling (CBIM)**
- **Biological, Mathematical, and Physical Sciences Interfaces Institute for Quantitative Biology (BIOMAPS)**
- **Institute of Biostatistics**
- **Rutgers Center for Cognitive Science (RUCCS)**
- **Center for Discrete Mathematics and Theoretical Computer Science (DIMACS)**
- **Institute of Marine and Coastal Sciences (IMCS)**
- **Wireless Information Network Laboratory (WINLAB)**
- **Center for Alcohol Studies**
- **The Rutgers Business, Engineering, Science and Technology Institute (BEST)**
- **Other I/UCRC's at Rutgers**
 - **Center for Autonomic Computing (CAC)**
 - **Ceramic and Composite Materials (CCMC)**
 - **Integration of Composites into Infrastructure (CICI)**

Rutgers Niche

- **Bioinformatics and Biomedical**
 - Patient treatment
 - Biological and Clinical data analysis and mining
- **Business**
 - Scalable mining methods
- **Scalable Visual Search, Monitoring/Surveillance Systems**
 - Behavior analysis, crowd monitoring, multisensor fusion, learning, scalable learning and mining methods
- **Distributed Robotics**
 - Distributed learning, cooperative tasks, environmental applications
- **Programming Languages and Architectures**
 - Distributed systems languages and architectures, low power systems
- **Entertainment**
 - Computational modeling, special effects modeling

How Does the Partnership Work?

- **Company joins the CDDA as a member of IAB**
 - Full membership: \$35,000/year
 - Affiliate membership: \$10,000/year
- **Membership on Industrial Advisory Board (IAB)**
 - Review and rank projects of interest to member
 - Vote and fund projects
 - Provide guidance to research activities
- **Receive annual resume book of CDDA students**
- **Receive pre-publication reports of all projects**
- **Receive nonexclusive access to all intellectual property**

Benefits of Membership

- **Royalty-free licenses to all intellectual property generated through the Center**
- **The right to vote on research projects carried out by the center**
- **Direct collaboration with center faculty, post-doctoral researchers, and graduate students**
- **Center promotes industry standards; members have direct role**
- **Influence policy makers through the Center on important science and technology issues**
- **Recruit top graduate students with hands-on technical skills in big data**

NSF I/UCRC Program

- Consortium of universities AND industrial partners
- Conduct pre-competitive research
- Industry members select the projects
- “Tried and true” IP agreement and organization
- NSF covers some of the university overhead, leaving more industry \$\$ for research
- Over the history of the program, I/UCRC member companies benefited from a 37:1 leverage on their investments
- Member companies estimated the commercial value of projects directly resulting from I/UCRC research at \$1.6B over 30-year history of the I/UCRC program

CDDA History

- NSF awarded Stony Brook and Rutgers an I/UCRC CDDA Planning Grant in 2009
- CDDA held a Planning Workshop at Stony Brook Manhattan in January 2010
- CDDA applied for an NSF Center Designation in October 2010
- CDDA was designated by NSF as an I/UCRC center in March 2011
- CDDA held its Inauguration Workshop at Rutgers in November 2011
- **CDDA 2nd Workshop at Stony Brook, May 2012**