Strengthening the Big Data & Analytics Ecosystem @ WSU

Ratna Babu Chinnam, Ph.D. & Patrick Gossman, Ph.D.
Ratna.Chinnam@wayne.edu | Patrick.Gossman@wayne.edu

Champion: Dr. Keith Whitfield, Provost

September 19, 2018
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00PM</td>
<td>Welcome Remarks: Provost K. Whitfield</td>
</tr>
<tr>
<td>3:20PM</td>
<td>TED Style Talks from WSU Faculty:</td>
</tr>
<tr>
<td></td>
<td>Prof. <strong>R.B. Chinnam</strong> Director, Big Data &amp; Business Analytics Group, Industrial &amp; Systems Engineering Department</td>
</tr>
<tr>
<td></td>
<td>Prof. <strong>M. Dong</strong> Co-Director, Big Data &amp; Business Analytics Group, Computer Science Department</td>
</tr>
<tr>
<td></td>
<td>Associate Prof. <strong>K. Ryzewski</strong> Archeology &amp; Digital Humanities, Anthropology Department</td>
</tr>
<tr>
<td></td>
<td>Dr. <strong>D. Martin</strong> Program Director, Urban Safety, Center for Urban Studies</td>
</tr>
<tr>
<td></td>
<td>Prof. <strong>P. Levy</strong> Assistant VP &amp; Associate Chair for Research, Department of Emergency Medicine</td>
</tr>
<tr>
<td></td>
<td>Prof. <strong>D. Ruden</strong> Director of Epigenomics, Institute of Environmental Health Sciences</td>
</tr>
<tr>
<td>4:15PM</td>
<td>Open Discussion: What are your needs?</td>
</tr>
<tr>
<td>4:40PM</td>
<td>Reception &amp; Posters</td>
</tr>
</tbody>
</table>
• McKinsey Global Institute predicted that an analytics mindset, coupled with the capability to harvest, mine, and exploit insights from ever-growing data streams, will facilitate new waves of innovation, productivity growth and consumer surplus across all sectors.

• This digital revolution is just as relevant for academia, whether it be for pursuing research, training the next-generation workforce through cutting-edge academic programs, or even improving university administration.

• Given this backdrop, how do we ensure that Wayne State University stays ahead of the curve?

MGI Reports on Big Data & Analytics (June 2011 | Dec 2016)
What is Big Data?

Big Data & Analytics Ecosystem @ WSU
**What is Big Data to Us?**

**Large Data Sets and/or Advanced Analytics**

<table>
<thead>
<tr>
<th></th>
<th>“Small” Data Sets</th>
<th>“Large” Data Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Standard” Analyses</td>
<td></td>
<td>Big Data</td>
</tr>
<tr>
<td>“Advanced” Analytics</td>
<td>Big Data</td>
<td>Big Data</td>
</tr>
</tbody>
</table>
Vision for WSU’s Big Data & Analytics Ecosystem

VISION:
“To be a premier center of innovation in big data science and analytics”

OPERATING PHILOSOPHY:
- Be a leading source of data science and analytics expertise
- Leverage and reinforce strengths of diverse groups across the university for synergistic collaborations that enable pursuit of large research grants and foster breakthroughs.
- Offer leading edge educational and training programs
- Actively collaborate with industry utilizing advanced research in computing tools and platforms
- Contribute to development of the broader SE-Michigan community through collaborative activities
- Enhance university image and reputation
Leadership & Steering Committees

Steering Committee
- K. Whitfield, Provost
- S. Lanier, OVPR
- F. Fotouhi, Engineering
- L. Hazlett, SoM
- W. Raskind, CLAS
- R. Forsythe, Business
- D. Hubbard, C&IT
- H.A. Coates, Development

Ecosystem Leadership
- R. Chinnam, Engineering
- D. Cinabro, CLAS
- S. Draghici, Engineering & SoM
- P. Gossman, C&IT
- P. Levy, SoM / OVPR
- T. Somers, Business
- L. Thompson, CUS
- D. Walz, SoM
• Pockets of strong expertise
• Market niche in select areas but lack critical mass
• Insufficient resources (people, skills, infrastructure) to serve ourselves and our community
  • Research Grants and Scholarship
  • Education
  • Administration & Governance (e.g., Student Success)
  • Community Engagement & Consulting
  • Service Centers of Excellence
Proposed Big Data Ecosystem Organizational Structure @ WSU

Big Data & Analytics Ecosystem @ WSU
Core Computing Infrastructure

- Add Hardware for Hadoop/Spark and Data Science Computing Platform/Software for Unstructured and Structured Data, and Visualization Platforms
- Add Data Storage
- Provide Secure Environments
- Provide Internal Cloud Services

Big Data & Analytics Ecosystem @ WSU
Initial Staffing Model

- Appoint Leadership Team
- Lead Systems Admin & Two (2) Data Engineers
- Program Manager / Administrative Assistant
- Develop Support Services
Establish Core Services Facility:
Data Science Faculty & Staff
Education & Training Programs
Research & Consulting Services
Accomplished

- Big Data Group
  - Annual Big Data Symposia since 2014
    - ~500 Registrations from ~150 companies for 2017/2018
  - M.S. in Data Science & Business Analytics
  - Build Infrastructure
    - Hadoop/Spark Cluster, Container Environments
- School of Library & Information Science
  - Certificate in Information Management
- NSF’s [Midwest Big Data Hub](#)
  - WSU Leads “Business Analytics” Spoke
- Sustained (Small) Research Grants/Funding
- Outreach to Other Groups
Action Plan

- Increase Synergies & Engagement
  - Interdisciplinary Faculty Lines
  - Regular Focus Group Meetings

- Enhance Infrastructure
  - Software for Unstructured Data / Visualization
  - Data Storage
  - Centralize computing where reasonable and fund using startup packages

- Establish Big Data Science Core Facility
  - Fund Critical Staff
    - Lead Systems Admin, Two Data Engineers, Faculty Time, GSAs, Program Manager / Administrative Assistant
  - Core Facility Setup (MEB)
• Education & Training Programs
  • Planning a menu of 1-credit courses on topics that will be of broad interest across the university
  • Planning Training Workshops (1 or 2 Days):
    • Data Science 101 Machine/Deep Learning
    • Visual Analytics Advanced Analytics
    • Distributed Computing Big Data Management

• Planning Interdisciplinary Research Stimulus Grants
  • Funding from Provost and VP for Research
  • Faculty have to work across colleges/disciplines

• Start Building Partnerships & Grow the Ecosystem
  • Private Sector / Government / Non-Profits / Others …
    • Membership, Investments …
  • Examples:
    • SOSCIP (Southern Ontario Smart Computing Innovation Platform)
    • IHBI - Central Michigan University (Primary Sponsor: Dow)
Support for Other Units

Big Data & Analytics Core
(Big Data, Machine Learning, Analytics, Training, Consulting)

Computing Infrastructure
Hardware, Software and System Staff
(C&IT and Colleges)

Research Design & Analysis (RDA)

Biostatistics Core (BERD)

Statistical Consulting (MATH)
Big Data & Analytics Lab

- Meeting Space
- Tutoring Space
- Student Research Workstations

- Volunteer Faculty & GSAs
- Conference Area for 6-8 People
- (7) Student Workstations
- Large TV, Glass Dry Erase Boards, Equipment for Online Meetings

Manufacturing Engineering Building
2nd Floor – Room 2011
Data Science & Business Analytics MS Program

- **Colleges Housing Program:**
  - College of Engineering
  - Mike Ilitch School of Business

- **Three Synergistic Tracks:**
  - **Analytics** – Industrial & Systems Engineering Department
  - **Engineering/Platforms** – Computer Science Department
  - **Business** – Information Systems Mgmt.

- **Inception Date:** Fall 2017
- **Applications To Date:** > 325
- **Enrollment:** ~80 Students
- **Coursework:**
  - Available to non-DSBA majors
- **Sponsor Practicum Projects**
- **Website:** bigdata.wayne.edu
Planning Short & Modular Courses: 1-Credit

Examples

- Big Data Essentials: HDFS, MapReduce and Spark (see Coursera Example – Course 1)
- Big Data Integration & Processing (see UC San Diego Coursera Example – Course 3)
- Getting and Cleaning Data (see Johns Hopkins Coursera Example – Course 3)
- Introduction to Data Science in Python (see UoM Coursera Example – Course 1)
- Data Visualization in Python / R / Tableau (see UoM Coursera Example – Course 1; UC Davis Coursera Example – Several Courses)
- Business Intelligence Concepts, Tools, Applications (see U. of Colorado Coursera Example – Course 4)
- Applied Statistics with R (see UoM Coursera Example – Course 5)
- Applied Machine Learning in Python (see UoM Coursera Example – Course 3)
- Applied Machine Learning with Big Data (see UC San Diego Coursera Example – Course 4)
- Applied Deep Learning (see Coursera Examples by Andrew Ng)
- Applied Text Mining in Python (see UoM Coursera Example – Course 4)
- Applied Social Network Analysis in Python (see UoM Coursera Example – Course 5)
- Applied Graph Analytics for Big Data (see UC San Diego Coursera Example – Course 5)

Survey to be sent out soon!
Strengthening the Big Data & Analytics Ecosystem @ WSU

How do we ensure that Wayne State University stays ahead of the curve?

Please return cards to:
Prof. Ratna Babu Chinnam
Director - Big Data & Business Analytics Group
ratna.chinnam@wayne.edu
MEB Building - Room 2161

For more information:
bigdata@wayne.edu or bigdata.wayne.edu

What is your most urgent need:
- [ ] Hardware
- [ ] Software
- [ ] Other
- [ ] Staff Support
- [ ] Training Workshops

What services would you like to see from the Big Data & Analytics Group?

Interested in meetings around Big Data & Analytics:
- [ ] Monthly
- [ ] Bi-Monthly
- [ ] Semester
- [ ] No Regular Meetings

Any other suggestions?

Would you like a follow-up?  [ ] Yes  [ ] No

Name: __________________________ E-Mail: __________________________

Department: __________________________
Thank you!

Ratna Babu Chinnam, Ph.D.
Co-Director, Big Data & Business Analytics Group
Ratna.Chinnam@wayne.edu