NC State University Initiatives in “Big Data”

Randy K. Avent
23 May 2014
National Interests in “Big Data”

Managing, processing and exploiting massive data sets for better decision making will continue to grow in importance.
Data Analytics System Model

Applied Problem

“Synthesis” Postulate Options

Options Supplied

“Analysis” Assessment

Control

Inadequate Options

Data Management

Collection

Insufficient Information

Storage

Option Scoring

Supplied Collection Storage
Outline

• Introduction

• Departments

• Institutes

• Summary
College of Engineering

- Computer Science
  - Advanced Learning
  - Algorithms
  - Analytics
  - AI & Intelligent Agents
  - Cloud Computing
  - Cyber-security
  - Electronic Commerce
  - Information/Knowledge Management
  - Parallel & Distributed Systems
  - Scientific & HPC Computing

- Electrical Engineering
  - Bioelectronics Engineering
  - Communications & Signal Processing
  - Computer Architectures
  - Electronic Circuits & Systems
  - Nano-electronics & Photonics
  - Networking
  - Power Electronics
College of Sciences

- Statistics
  - Data Mining/Machine Learning
  - Biological Applications
  - Mathematical Statistics
  - Measurement Error/Mixed Models
  - Model Selection
  - Non-parametric Statistics
  - Time Series

- Mathematics
  - Algebra
  - Analysis, Geometry & Topology
  - Numerical Analysis
  - Optimization & Control
  - ODE & Dynamical Systems
  - PDE
  - Financial Mathematics
  - Symbolic Computation
Analytic Demographics

- Engineering: 35%
- Social Sciences: 13%
- Sciences: 8%
- Education: 5%
- Mgmt: 2%
- Design: 2%
- Natural Resources: 1%
## Faculty Gap Analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>CompSci</th>
<th>Math</th>
<th>Statistics</th>
<th>Physics</th>
<th>Business</th>
<th>CHASS</th>
<th>ECE</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical models</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Social models</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Symbolic models</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Numerical solvers</td>
<td>Red</td>
<td>Green</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>HPC, HPD, OS</td>
<td>Green</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Software architectures</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Privacy</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>Yellow</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Discrete mathematics</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>AI/Knowledge Mgmt</td>
<td>Red</td>
<td>Yellow</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Database</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>Data integration</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
<td>Red</td>
</tr>
</tbody>
</table>

### Coverage:
- **Red**: Little
- **Yellow**: Weak
- **Green**: Strong

---

RKA20140523-8

NC STATE UNIVERSITY
The Chancellor’s Faculty Excellence Program is bringing some of the best and brightest minds to join NC State University’s interdisciplinary efforts to solve some of the globe’s most significant problems. Guided by a strong strategic plan and an aggressive vision, the cluster hiring program is adding 36 new faculty members in 12 select fields to add more breadth and depth to NC State’s already-strong efforts.

**Bioinformatics**
This field develops and deploys tools to help make sense of the vast, complex, and diverse data-sets coming from studies in biological and medical science.

**Data-driven Science**
Nearly every research field and industry sector is struggling with managing and extracting useful information from massive data sets. Modeling, data management, and analytics are key elements in understanding and using these data sets, and their importance in academia and industry is predicted to grow exponentially.

**Strategic Planning**
In a changing world, NC State faces new challenges and new demands. Learn how we plan to stay ahead.

**Recent News**
- Doing ‘Cluster’ Hiring Right (9/30/13, Inside Higher Ed)
- Clusters Create New Campus Culture (9/19/13, NC State Bulletin)

**New Hires**
- Bioinformatics: David Reif
- Bioinformatics: Fred Wright
- Data-driven Science: Ayon Wilson
- Digital Transformation of Education: Roger Azeneto
- Digital Transformation of Education: Min Chiu
- Environmental Health Science: Michael Bereman
- Environmental Health Science: Jane Hopper
Analytic-Related Cluster Hires

- Bioinformatics
- Data-Driven Science
- Digital Transformation of Education
- Geospatial Analytics
Department Structures & Goals

- Departments primarily consists of Tenured/Tenure Track (T/TT) Faculty
- Faculty primarily conduct fundamental research
- Primary funding through federal agencies like NSF, DOE Science, OxRs, ....
- Faculty measured largely on publications and knowledge creation
- Faculty tend to be application-independent
Translational Research Model

• Medical Domain

- Academics
  - Fundamental Research
  - Systems Biology & Computational Biology
  - Clinicians

- Analysts & Industry
  - Institutes & Research Scientists

• Analytics Domain
Outline

• Introduction
• Departments
• Institutes
• Summary
## Data Science Institute Matrix

<table>
<thead>
<tr>
<th>National Security</th>
<th>Information Content</th>
<th>Quantitative Sense-making</th>
<th>Prediction Markets</th>
<th>Knowledge Representation</th>
<th>...</th>
<th>Narrative Processing</th>
<th>T/TT Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Research Faculty

Software Engineers

Industry Engagement Managers
Laboratory for Analytic Sciences (LAS)

• An NSA-funded laboratory exploring the challenges and opportunities created by “Big Data”
  – Lease agreement provides facility to collocate NSA researchers and analysts with faculty
  – Research contract funds fundamental research through development

• Primary research thrusts
  – Data discovery, curation, processing and storage, hypothesis generation, pattern recognition, classification, fusion, visualization, and dynamically reconfigurable workflows

• Team Members
  – Academic: UNC and Duke
  – Industry: SAS, IBM, Cisco, Signalscape, Tigerswan

• NSA goal to build an advanced data innovation hub on Centennial Campus with LAS as the anchor tenant
Why NC State?

- Deep ingrained land-grant culture drives faculty willingness to engage and work on “use-inspired” research
- “Triple helix” vision for Centennial Campus
- Research Triangle Park cluster economy on Big Data
- Long history of research in Big Data technologies and analytics
- Faculty willingness to engage across all colleges
LAS Current Research Efforts

• **Tradecraft – The Science of Analysis**
  – Instrumentation
  – Analyst Workflow
  – Cognitive Thinking

• **Technology – The Science of Analytics**
  – Data Readiness Levels
  – Future States

• **User Experience**
  – Narrative Processing
Summary

• Managing and extracting information from complex data sets continues to grow in importance in most all sectors of the US economy

• NC State recently awarded significant program in data science that will be leveraged for future growth

• Importance of data science recognized through the Chancellor’s Faculty Excellence Cluster Hires

• Building a research institute to capitalize on multi-disciplinary opportunities, build significant program for LAS and bring faculty together to maintain our national position in analytics and data science
Departmental Demographics

- Advanced Learning Technologies
- Algorithms and Theory of Computation
- Analytics
- Artificial Intelligence & Intelligent Agents
- Cloud Computing
- Cybersecurity
- Electronic Commerce
- Information & Knowledge Management
- Parallel & Distributed Systems

- Data Mining/Machine Learning
- Biological Applications
- Mathematical Statistics
- Measurement Error/Mixed Models
- Model Selection
- Nonparametrics/Semiparametric Statistics

- Algebra
- Analysis, Geometry, and Topology
- Numerical Analysis
- Optimization and Control
- ODE & Dynamical Systems
- PDE
- Probability, Stochastic Processes, and Financial Mathematics
- Symbolic Computation

- Bioelectronics Engineering
- Communications & Signal Processing
- Computer Architecture & Systems
- Electronic Circuits & Systems
- Nanoelectronics & Photonics
- Networking
- Power Electronics & Systems
Translational Research Model

Data Sciences Institute

Department
TT/T Faculty
Basic/Applied

Solutions, ideas, …

Problems, funding, context, validity, …

Institute
Research Faculty
Applied/TechDev

Infrastructure & Services

NSF
DOE
OxRs

National Security
Business
Health Care
NC State Faculty