NEXT GENERATION OF MS DATA ANALYTICS PROGRAMS - VISION

James Baldo Jr., Ph.D.
Director - MS Data Analytics Program

Harry Foxwell, Ph.D.
Associate Professor

Ioulia Rytikova, Ph.D.
Associate Professor and Associate Chair for Graduate Studies

3 June 2019
OVERVIEW OF THE CHALLENGES – DATA – PREPARING STUDENTS

- Conditioning
- Data Readiness
- Tagging/Security
- Storage/Archive/Sources/Real-time/...
- Search/Discovery
- Large Volume of Data: Petabytes -> Exabytes -> Zetabytes
DATA ANALYTICS – A KEYSTONE OF DATA SCIENCE

<table>
<thead>
<tr>
<th>Big Data</th>
<th>Data Analytics</th>
<th>Descriptive Analytics</th>
<th>Predictive Analytics</th>
<th>Prescriptive Analytics</th>
<th>Explicit Analytics</th>
<th>Implicit Analytics</th>
<th>Deep Analytics</th>
</tr>
</thead>
</table>

Data Science
TECHNICAL CHALLENGES - DATA ANALYTICS CONCEPTUAL ARCHITECTURE

Data Centric/Driven
- Data Lake – variety & veracity
- Data Catalog
- Data Conditioning & Pollution
- Data Security

Services
- Set of common micro services
- Service development
- Service versioning

Application Hosting
- Need Elastic Compute
- Need Shareable Data
- Need Access to Common set of Services
GMU MS DATA SCIENCE PROGRAM PERSPECTIVE - PEOPLE

Students

Degree/Certificate
Working
Full-time/Part-time

Industry Partners & Sponsors

Mason Data Analytics Program

Faculty

External Communities

• Application Domains:
  • Engineering
  • Business
  • Finance
  • Banking
  • Health
  • Medical
  • Insurance
  • Manufacturing
  • Education
  • Utilities
  • Public Sector
  • National Security

College of Health and Human Services
School of Business
Volgenau School of Engineering

Background: computer science/statistics/engineering/finance/etc.

Professional Societies
Standards Bodies
Government Organizations
Other Universities/Colleges
DAEN Advisory Board
COURSE TOPIC AREAS – IN SUPPORT OF NEXT GENERATION DATA SCIENTIST PROFESSIONALS

• Concentrations –Groupings
  • Statistical Analytics
  • Predictive Analytics
  • Applied Analytics
  • Business Analytics
  • Cyber Analytics
  • Health Data Analytics
  • Data Mining
  • Internet of Things
  • Financial Engineering

• Core Courses – Content Groupings
  • Analytics Big Data to Information (AIT 580)
  • Deterministic Models (OR 541)
  • Data Management & Mining (CS 504)
  • Applied Statistics/Visualization (STAT 515)
  • Capstone Project
    • Sponsored Project
    • Extend previous Capstone Project
    • New Project
    • Faculty mentor/Project
LEARNING ENVIRONMENTS AT MASON

Active Learning
Inquiry-based
Experiential
Project-based

Student-centric
Collaborative
Interactive
Engaging
SUPPLEMENTAL SLIDES
CURRENT LANDSCAPE FOR DATA ANALYTICS

• **Technical**
  - Data
  - Services
  - Computing

• **Process**
  - Problem definition
  - Data Readiness => identification, collection, and conditioning
  - Modeling
    - Development
    - Evaluation
  - Deployment

• **Roles**
  - Data Scientists
  - Data Engineer
  - Data Architect
DATA SCIENCE ROLE DEFINITIONS

• Data Architect – The data architect creates the framework that make data driven intelligence possible.
  • Details:
    • Create systems (e.g., procedures, governance, and architectures) to store, manage, process, and preserve or dispose of data.
    • Enable an organization to manage its data as an asset and increase the value it gets from its data by identifying opportunities for data usage, cost reduction, and risk mitigation.

• Data Engineer – The data engineer conditions data to fit within the data architecture and transforms it to be exploitable.
  • Details:
    • Transform data into usable and computationally accessible forms.
    • They condition data through extraction/cleansing/transformation/loading (ECTL, aka: data munging), they implement data systems which separate data from application and scale as required.

• Data Scientist – The data scientist creates repeatable means to draw key insight and signals from data.
  • Details:
    • Invent, perfect, or apply algorithms to extract insights from data.
    • They are specialists in a range of mathematical, computational, and visualization techniques that allow an organization to draw the greatest benefit from data holdings in terms of insight and decision advantage.
Some Organizations and Frameworks attempting to define data science roles:

- National Institute for Science and Technology (NIST)
- EDISON Project – 2-year project started in September 2015 funded by the European Union’s Horizon 202 research and innovation program
- SAIC – system integrator extended CRISP-DM model
- Springboard – data science education company
- Gartner

<table>
<thead>
<tr>
<th>Full Time/Title Only</th>
<th>Search for Number of Job Postings (dice.com 10Jun18)</th>
<th>EDISON</th>
<th>NIST</th>
<th>SAIC</th>
<th>Gartner</th>
<th>Springboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Science Research</td>
<td>1</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Scientist</td>
<td>526</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data Architect</td>
<td>457</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Data Analyst</td>
<td>731</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Data Science Programmer</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Engineer</td>
<td>577</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Evolutionary Impact: data science usage growth and team size growth

Individual Skills

Data Scientists

Collaboration

Data Engineers

Collaboration

Data Architects

Collaboration

Continuous Training – methods, technology, techniques, etc.

Team Skills
FOUNDATION AREAS FOR NEXT GENERATION DATA SCIENCE PROFESSIONALS

Mathematical & Statistical Foundations

Behavior & Event Processing
Data Storage & Management Systems
Data Quality Enhancement
Data Modeling & Representation
Deep Analytics, Learning & Discovery
Simulation & Experiment Design
High-performance Processing & Analytics
Analytics & Computing Architecture & Infrastructure
Networking, Communication, Interoperation

Social Issues: Privacy, Security & Trust
COURSES - INSTRUCTORS

• Objectives
  • Course uniformity across sections
    • Still provides instructor with ability to augment course content (e.g., lectures, quizzes, assignments, etc.)
  • Tools
    • Virtual Online Labs – Data Science Workbench Concept
      • Tools Loaded
      • Assignments Loaded – Instructors have support for designing and implementing assignments for environment
      • Metrics to show assignment progress, completion, problems, self-instruction/assistance, dashboard for visualization
  • Online Courses
    • Online courses should be evaluated for being converted to online
    • Online Course Structure:
      • Uniformed with other department courses
      • Still provides instructor with ability to augment course content (e.g., lectures, quizzes, assignments, etc.)
      • Online Courses are not static - course updates and modifications based on student feedback and faculty insights/observations
    • Online Course Monitoring
      • Missed assignments, quizzes, low marks, etc., should generate alerts to instructor/teaching assistant
    • Online Assistance
      • Blackboard
      • Email
      • WebEx
      • Etc.
DATA SCIENCE – PROGRAM VIEW POINT

Data Science

- Statistical Methodology
- Computer Science
- Areas of Applications (Domains)