NEXT GENERATION OF MS DATA ANALYTICS PROGRAMS - VISION

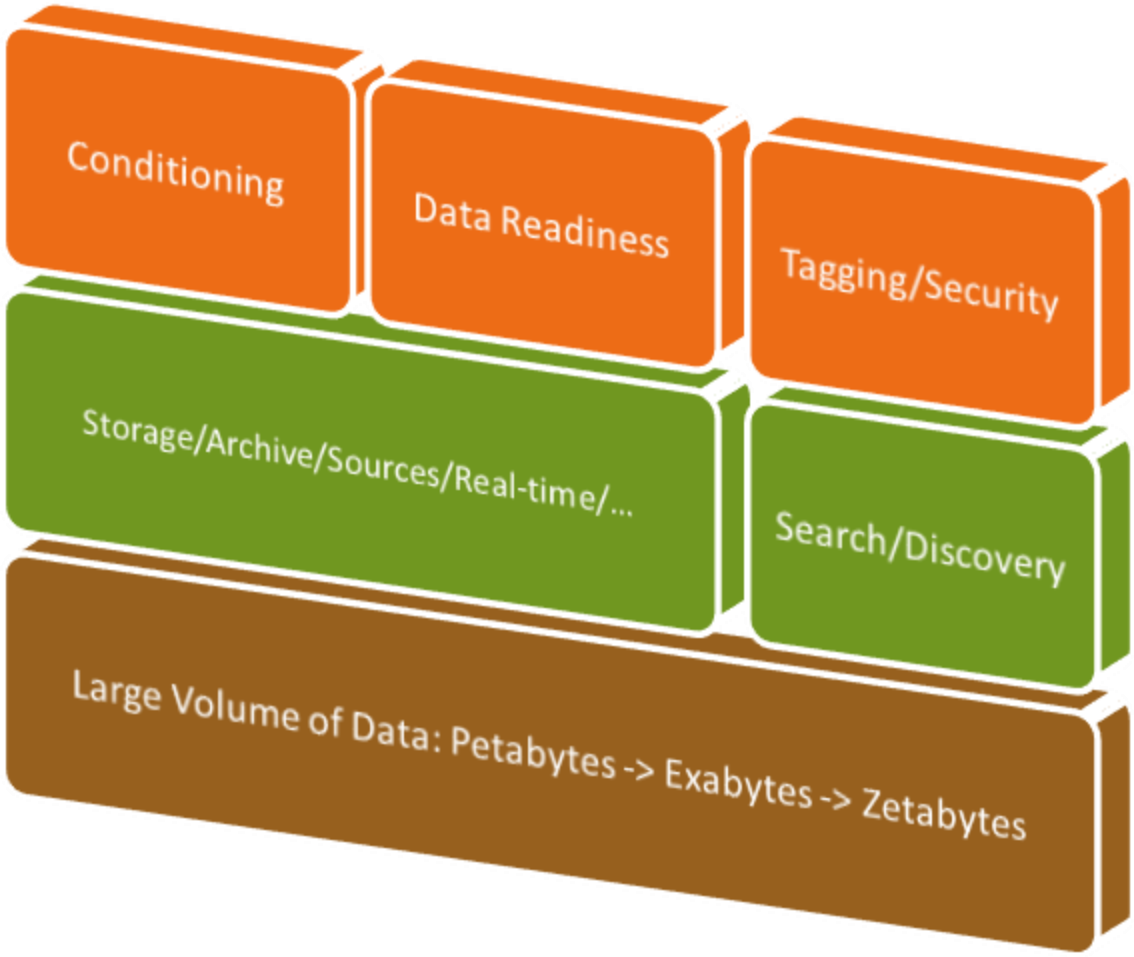
James Baldo Jr., Ph.D.
Director - MS Data Analytics Program
Harry Foxwell, Ph.D.
Associate Professor
Ioulia Rytikova, Ph.D.
Associate Professor and Associate Chair for Graduate Studies

3 June 2019

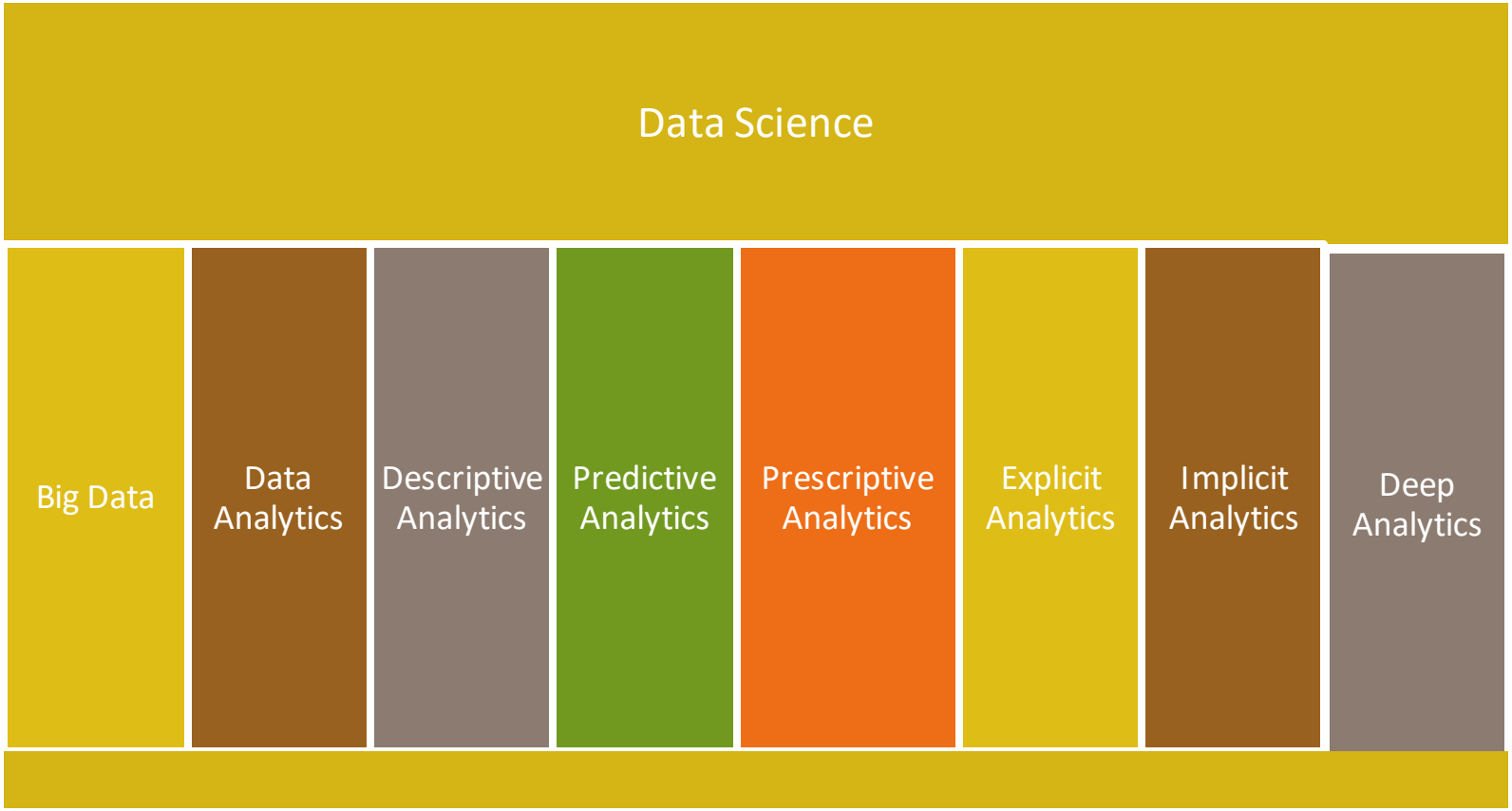
VOLGENAU SCHOOL MASON ENGINEERING

GEORGE MASON UNIVERSITY

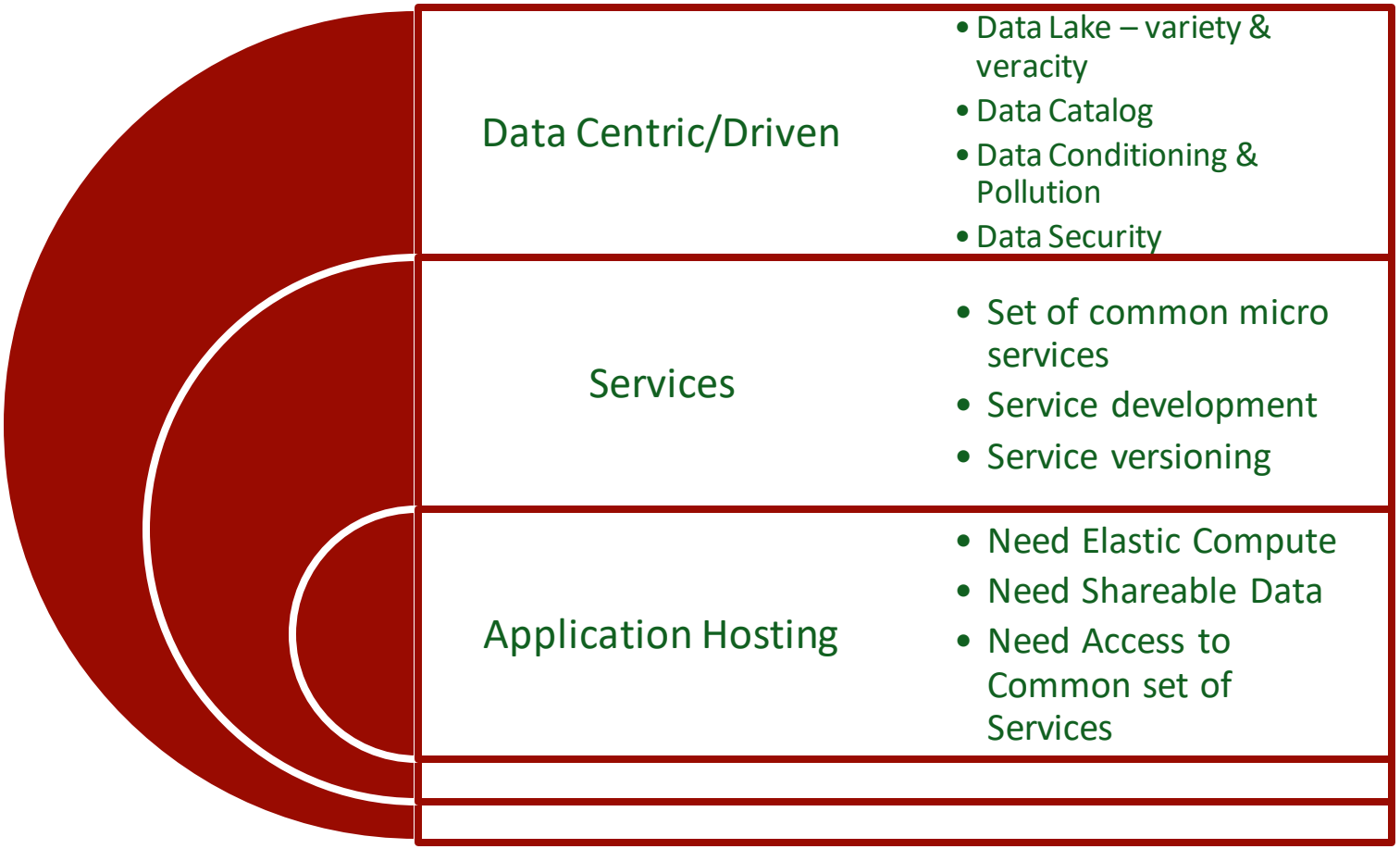
OVERVIEW OF THE CHALLENGES – DATA – PREPARING STUDENTS



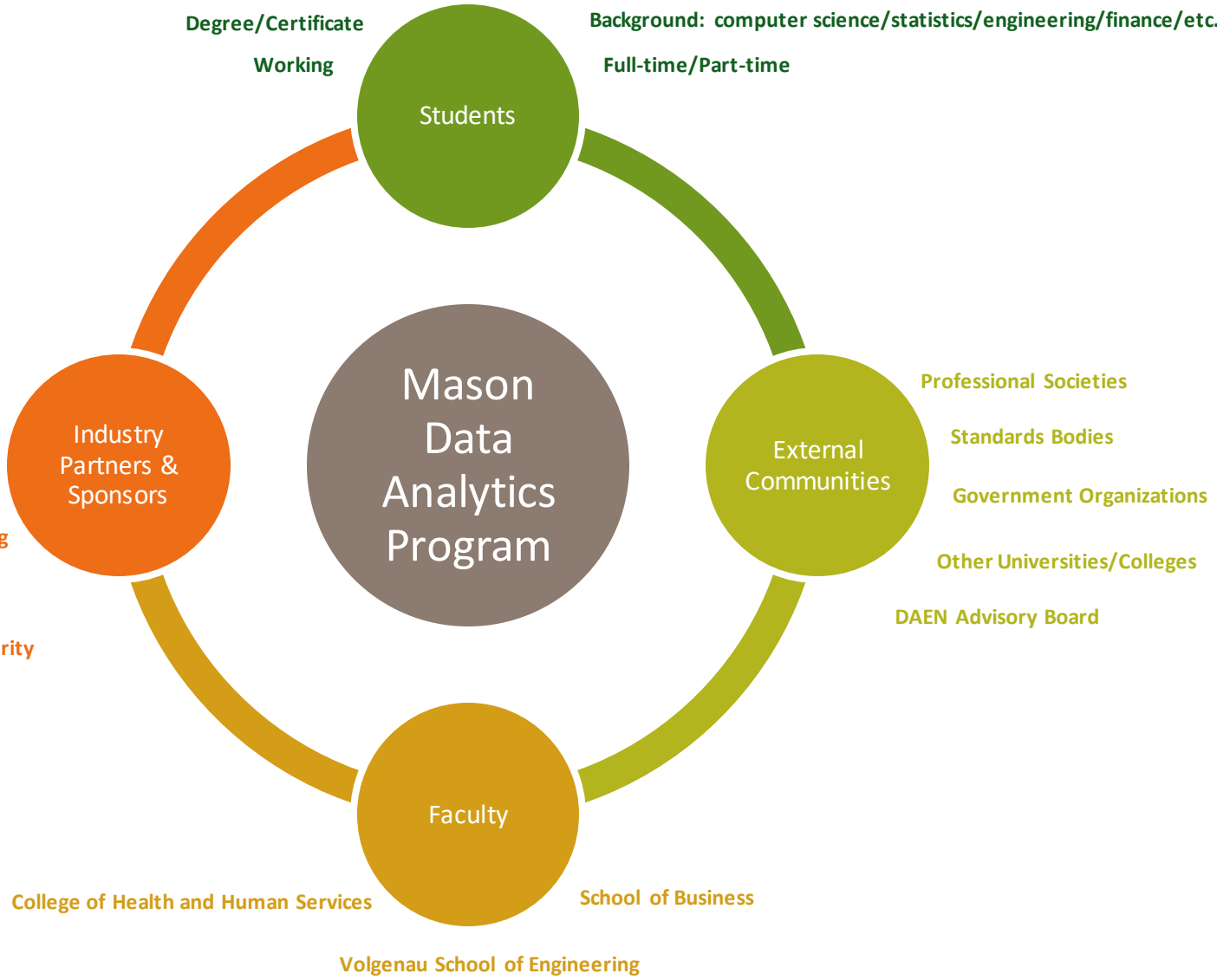
DATA ANALYTICS – A KEYSTONE OF DATA SCIENCE



TECHNICAL CHALLENGES - DATA ANALYTICS CONCEPTUAL ARCHITECTURE



GMU MS DATA SCIENCE PROGRAM PERSPECTIVE - PEOPLE

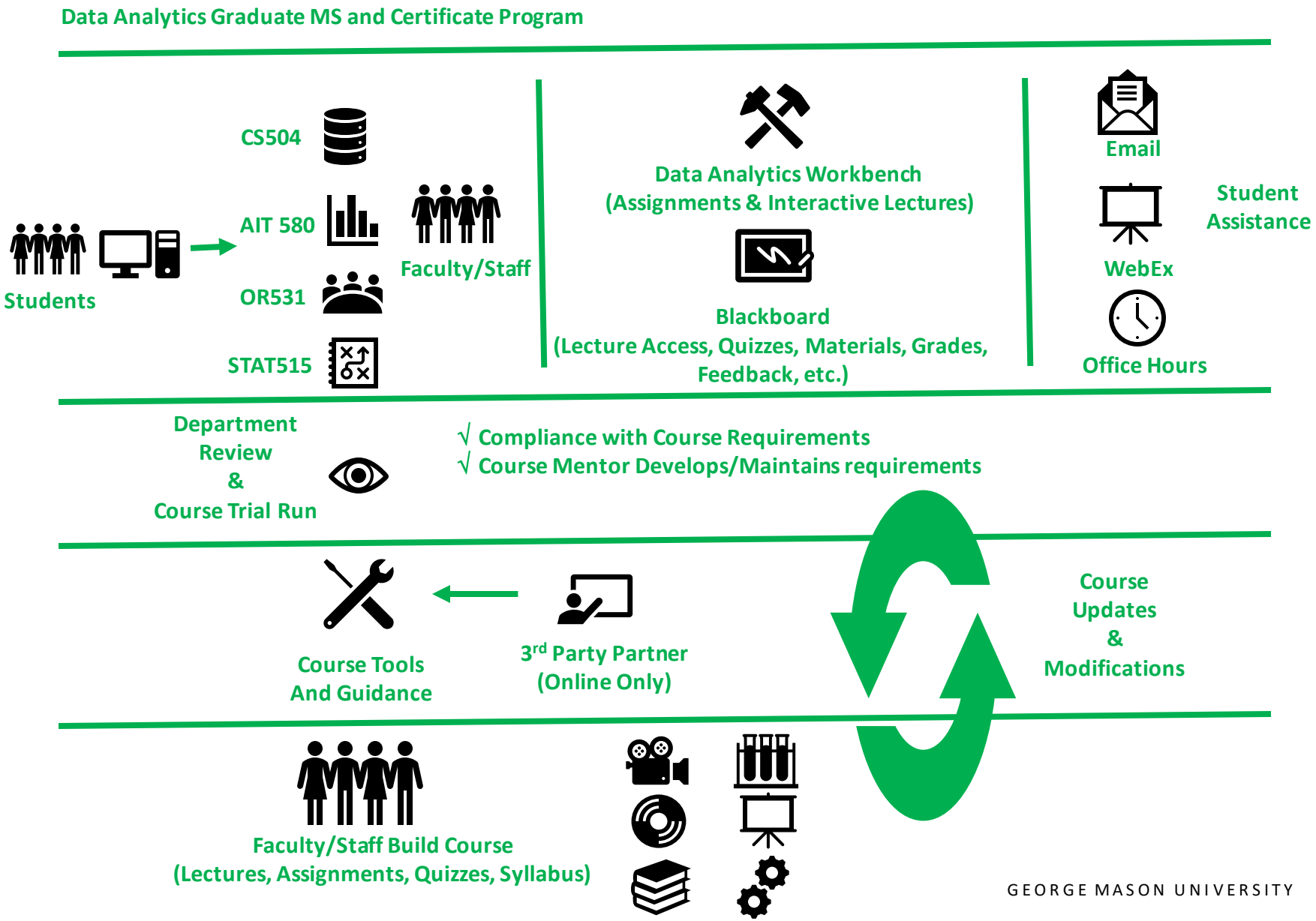


COURSE TOPIC AREAS – IN SUPPORT OF NEXT GENERATION DATA SCIENTIST PROFESSIONALS

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

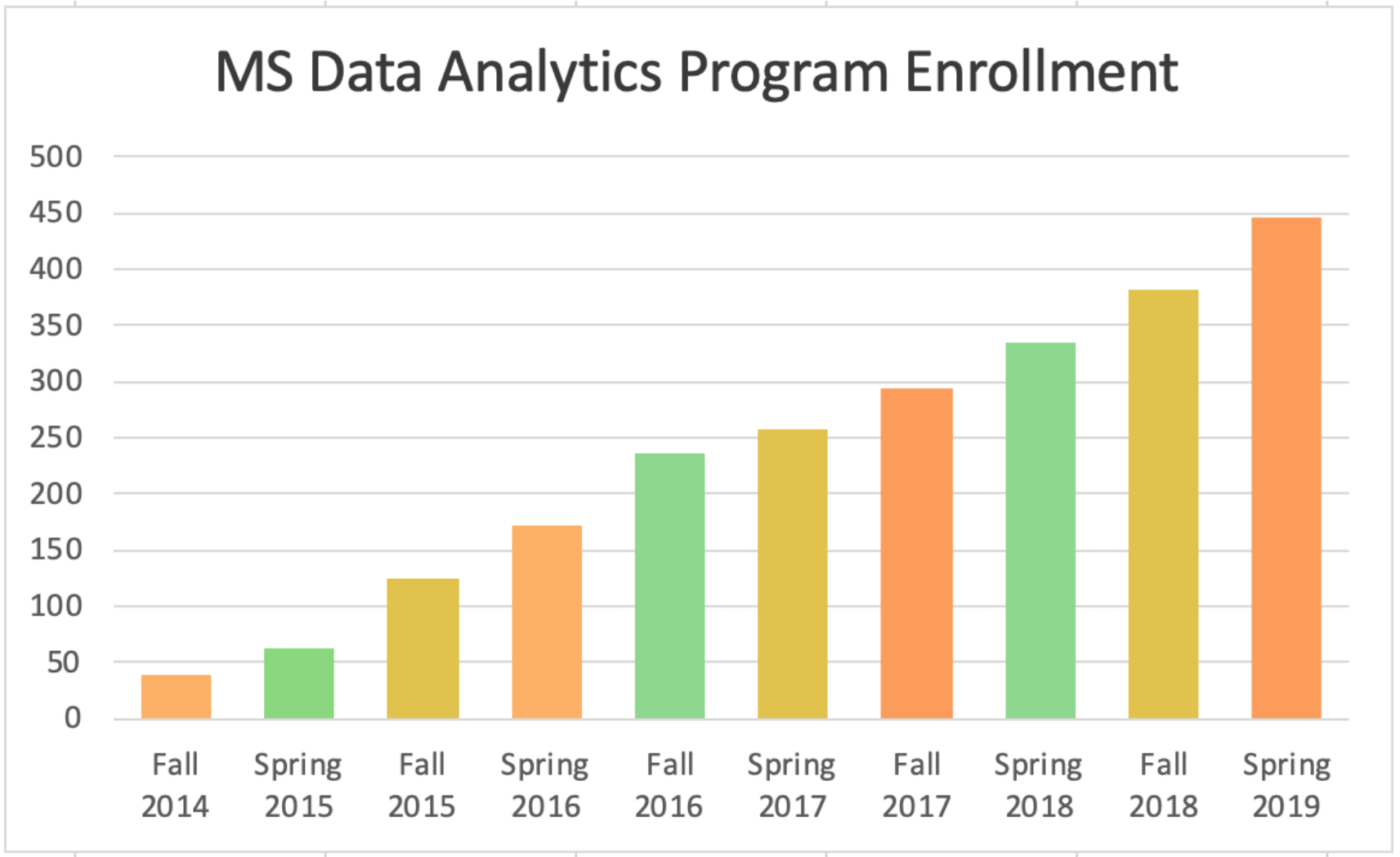
- **Core Courses – Content Groupings**
  - Analytics Big Data to Information (AIT 580)
  - Deterministic Models (OR 541)
  - Data Management & Mining (CS 504)
  - Applied Statistics/Visualization (STAT 515)
  - Capstone Project
    - Sponsored Project
    - Extend previous Capstone Project
    - New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

COURSE DEVELOPMENT/MAINTENANCE APPROACH



LEARNING ENVIRONMENTS AT MASON

MS DATA ANALYTICS PROGRAM ENROLLMENT



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.

DATA SCIENCE ROLES CHALLENGES

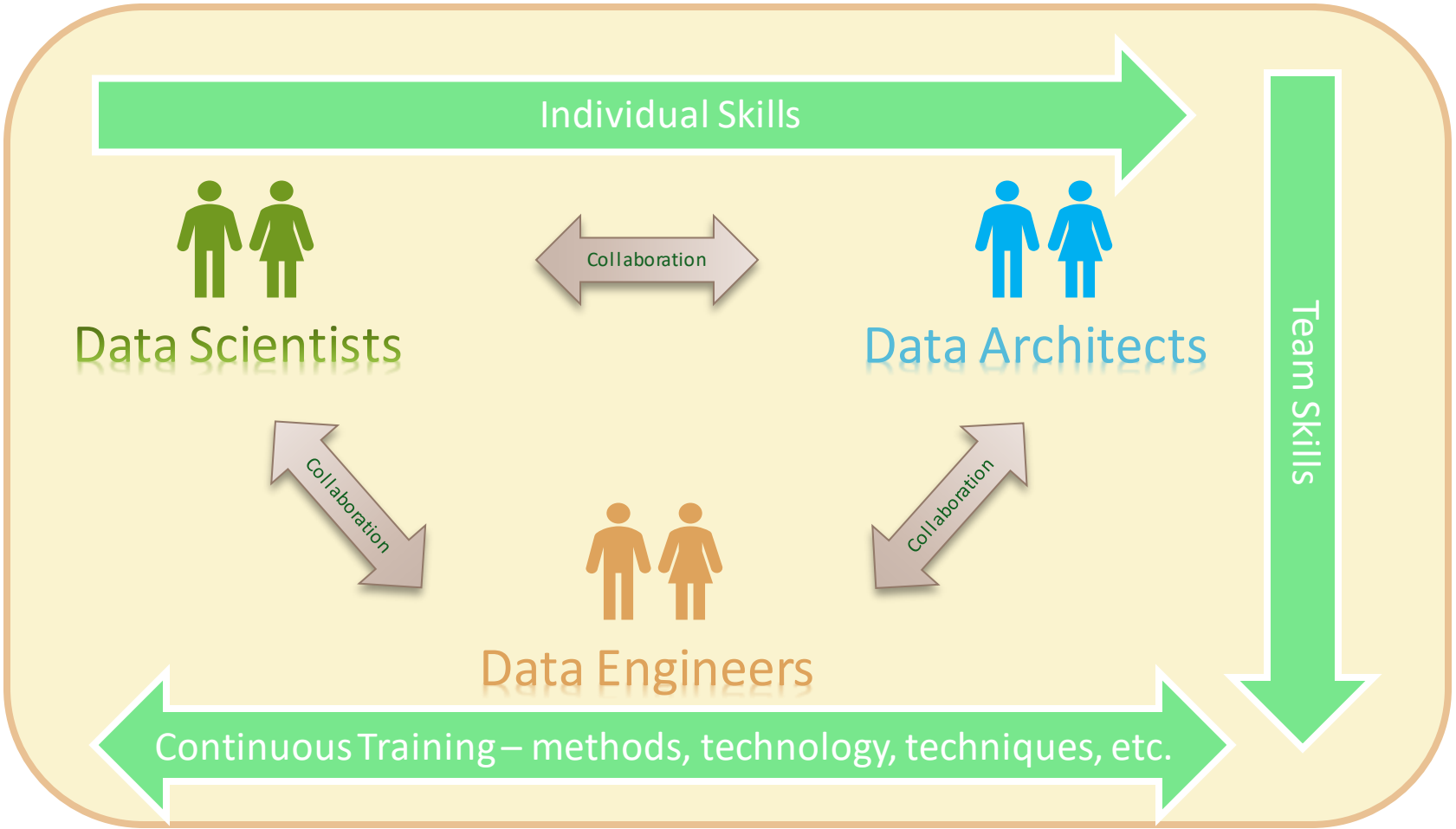
Some Organizations and Frameworks attempting to define data science roles:

- 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

Full Time/Title Only	Search for Number of Job Postings (dice.com 10Jun18)
Data Science Research	1
Data Scientist	526
Data Architect	457
Data Analyst	731
Data Science Programmer	0
Data Engineer	577

	EDISON	NIST	SAIC	Gartner	Springboard
Data Science Research	√				
Data Scientist	√		√	√	√
Data Architect	√				√
Data Analyst	√				√
Data Science Programmer	√				
Data Engineer			√	√	√

Evolutionary Impact: data science usage growth and team size growth



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