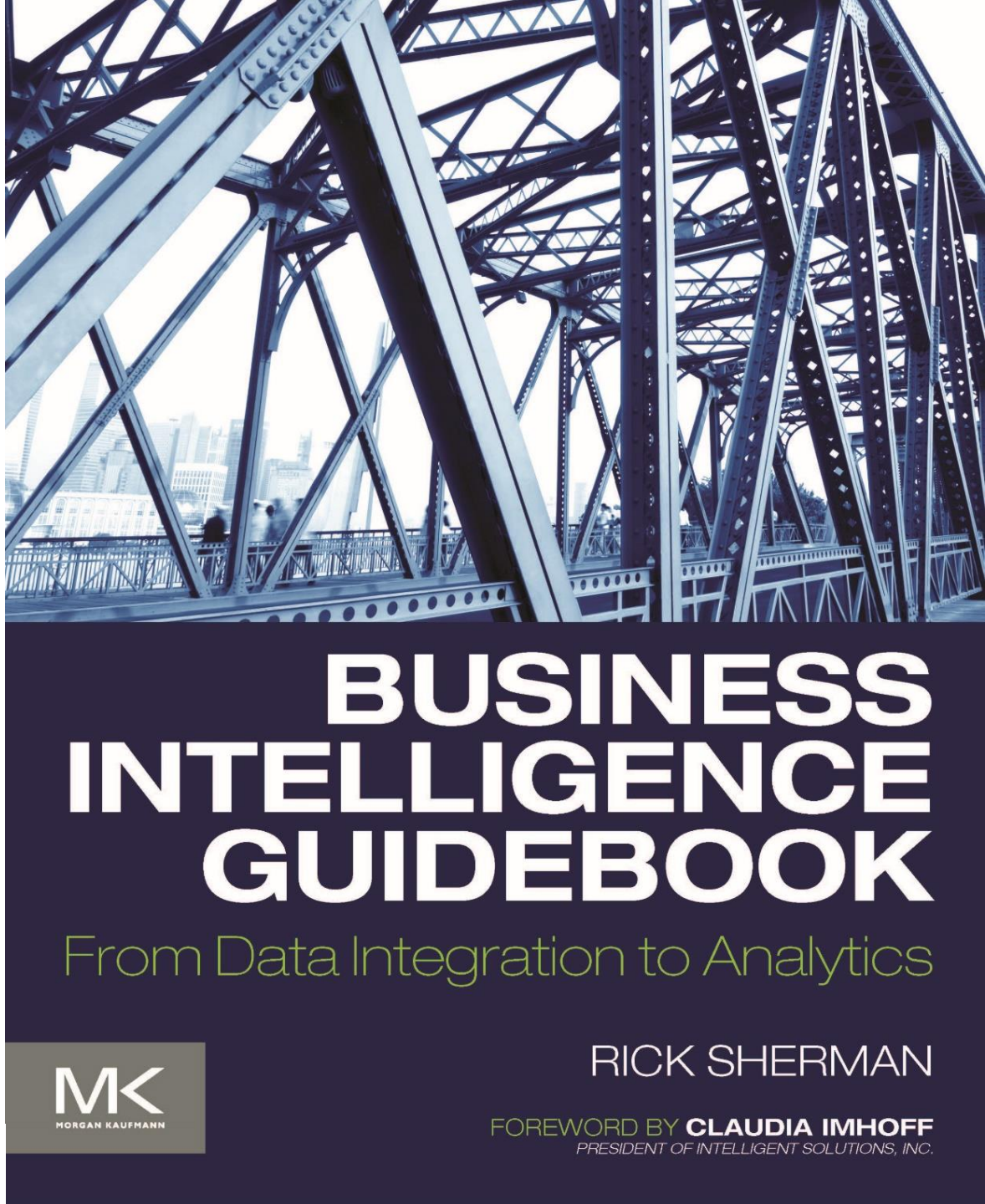
DAMG7370: Designing Advanced Data Architectures For Business Intelligence

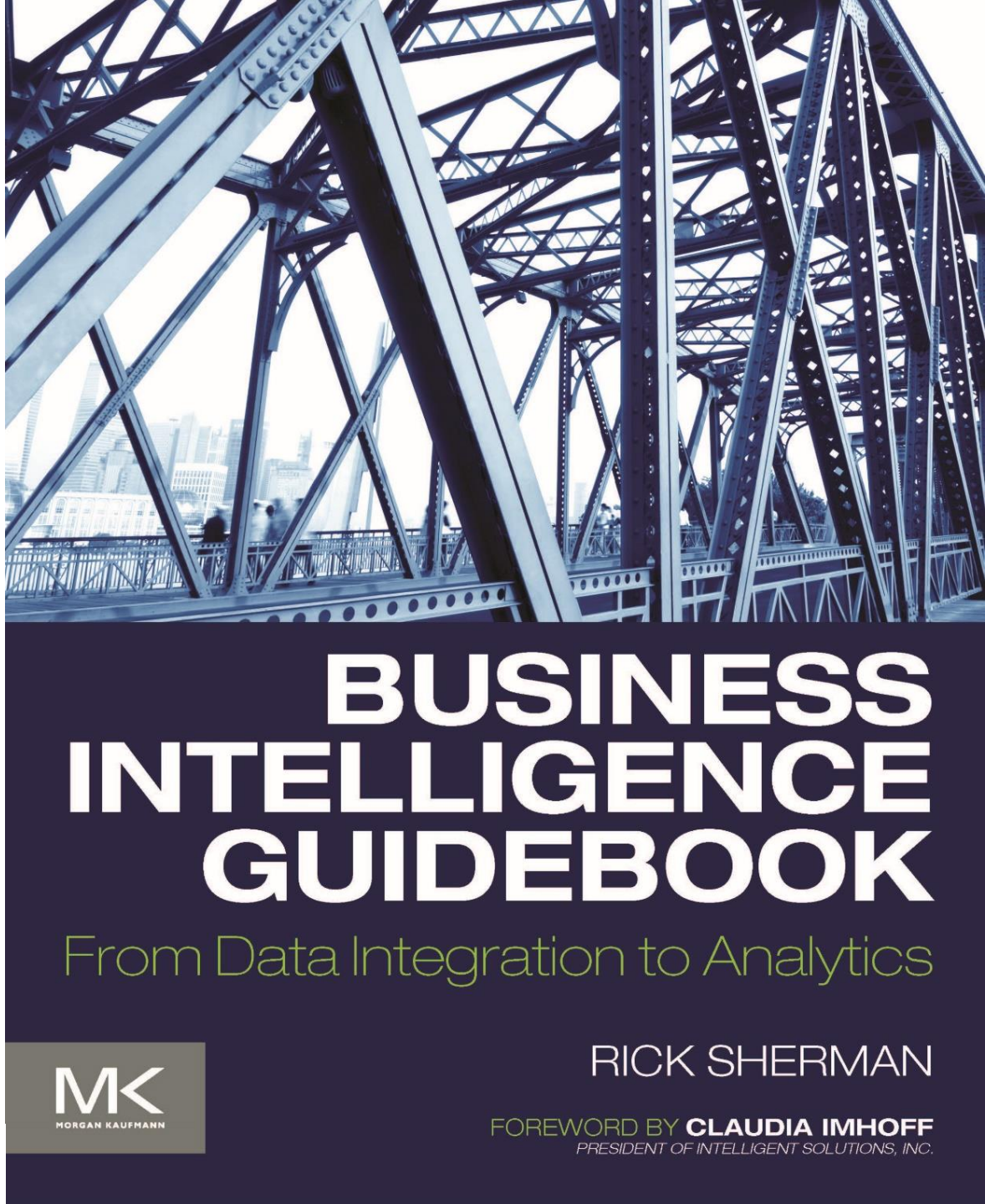
Topics:

- Architecture
- Data Modeling
- Data Integration & Data Preparation
- Business Intelligence & Analytics



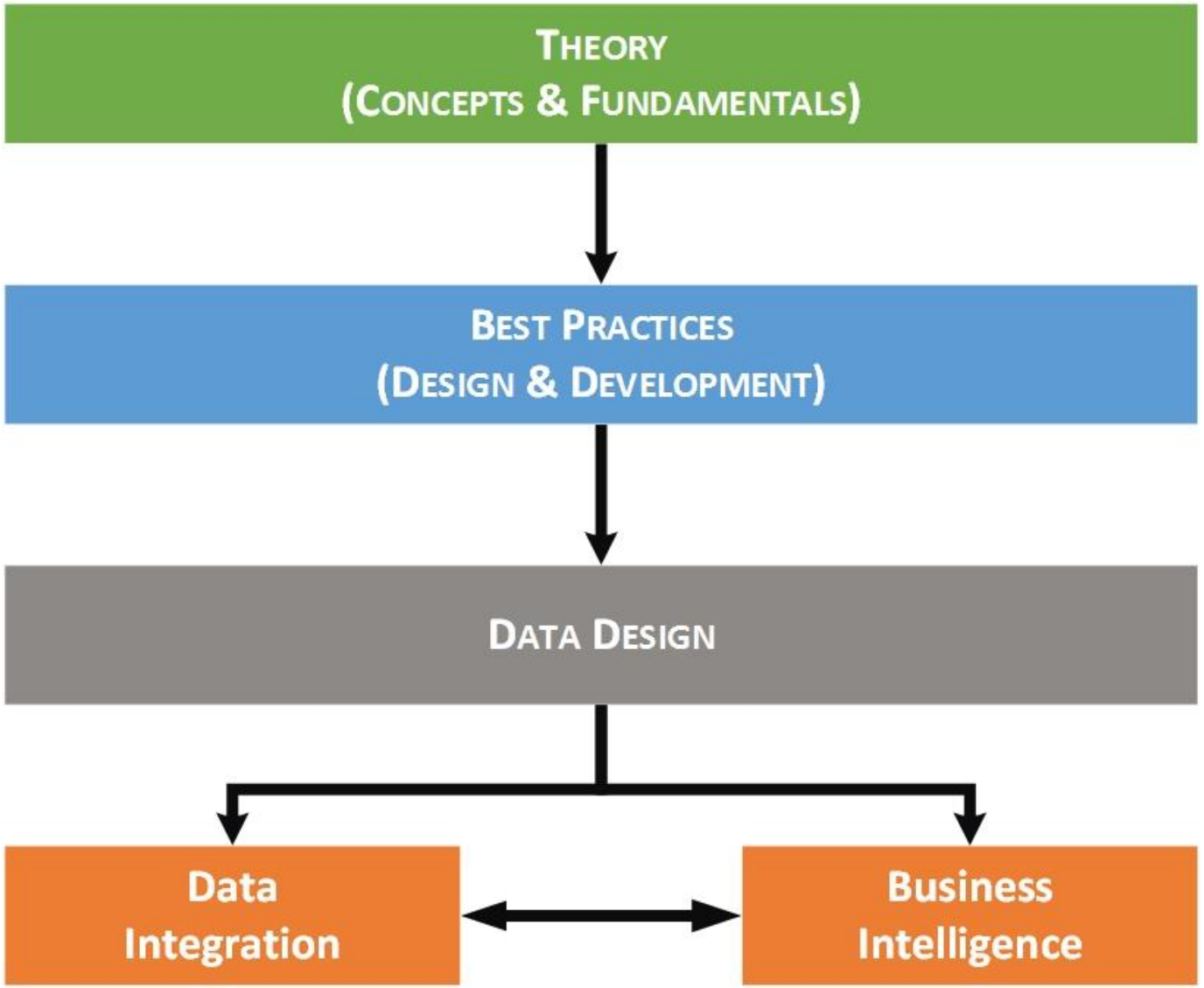
DAMG7370: Designing Advanced Data Architectures For Business Intelligence

- Architecture
  - Information, Data, Technology, Product
- Dimensional Modeling
  - BI-Oriented (Dimensional & Hybrid)
  - Traditional (ER Model) - as background
- Data Integration & Data Engineering
  - Concepts & Use Cases
  - Data integration & data engineering processes
  - Design, Development & Best Practices
- Business Intelligence
  - Concepts & Use Cases
  - Design, Development & Best Practices
  - Advanced & Big Data Analytics
  - Data Shadow Systems



ATHENA IT SOLUTIONS

Theory to Best Practices to Implementing with Tools



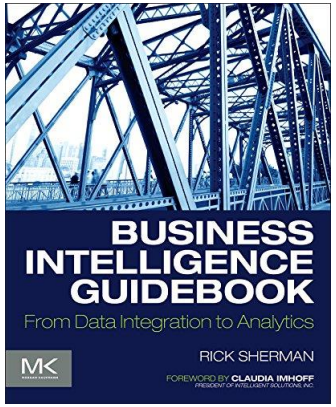
- Fundamentals, Architecture and Best Practices Continue to Expand and Evolve
- Tools & Coding Languages become hot & then obsolete by next “wave”
- People, Politics & Poor Practices are a Constant!
  …and create the primary problems with data (and tools do not fix by themselves)

Take-away:

- Tools help get you a job
- Being an expert in concepts and best practices builds your career

Course Textbook

- Required to buy paperback copy of textbook:
  - Business Intelligence Guidebook: From Data Integration to Analytics
  - Author: Rick Sherman (your instructor)
  - Publisher: Morgan Kaufmann; 1st edition (November 21, 2014)
  - ISBN-10: 012411461X
  - ISBN-13: 978-0124114616
- Industry Sources
  - My published articles, columns & podcasts
    - The Data Doghouse
    - Athena IT Solutions Library
  - Industry research – Gartner, Forrester, etc.



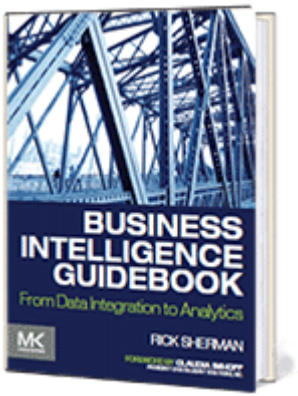
THE Data Doghouse℠
Unleashed observations on the business and technology of performa management, business intelligence, and data warehousing

Gartner IDC Analyze the Future FORRESTER®

ATHENA IT SOLUTIONS

Topics by Textbook Chapter

	Chapter
• Overview	
▪ State of Data in Enterprises	1
▪ Business Demand for BI	2
• Starting a BI Project	
▪ Defining Requirements	3
• Architecture	4
• Architecture Framework Data Design	
▪ E-R Modeling	8
▪ Dimensional Modeling	9
▪ Advanced Dimensional	10

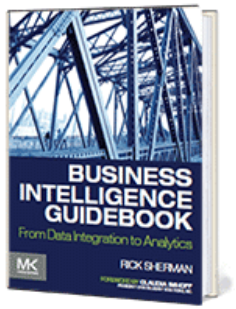


	Chapter
• Architectures	
▪ Information Architecture	5
▪ Data Architecture	6
▪ Tech & Product Architecture	7
• Data Integration	
▪ Design & Develop	11
▪ Processes	12
• Business Intelligence	
▪ BI Applications	13
▪ Design & Development	14
▪ Advanced Analytics	15
▪ Data Shadow Systems	16

Teaching Approach

Theory, Concepts & Real-World Experience

- Textbook
- Lectures
- Workshops
- Tutorials
  - Concepts
  - Tool Training
- Discussions
- Quizzes
- Exams



Hands-on Learning: Applying Concepts with Tools & Best Practices

Assignments

- Workshops (live or recorded with assignment)
- Homework
- Tutorials

Development-based projects

- Specific deliverables defined
- Objective: Apply concepts learned using hands-on approach with tools
- Each project reviewed like a code review

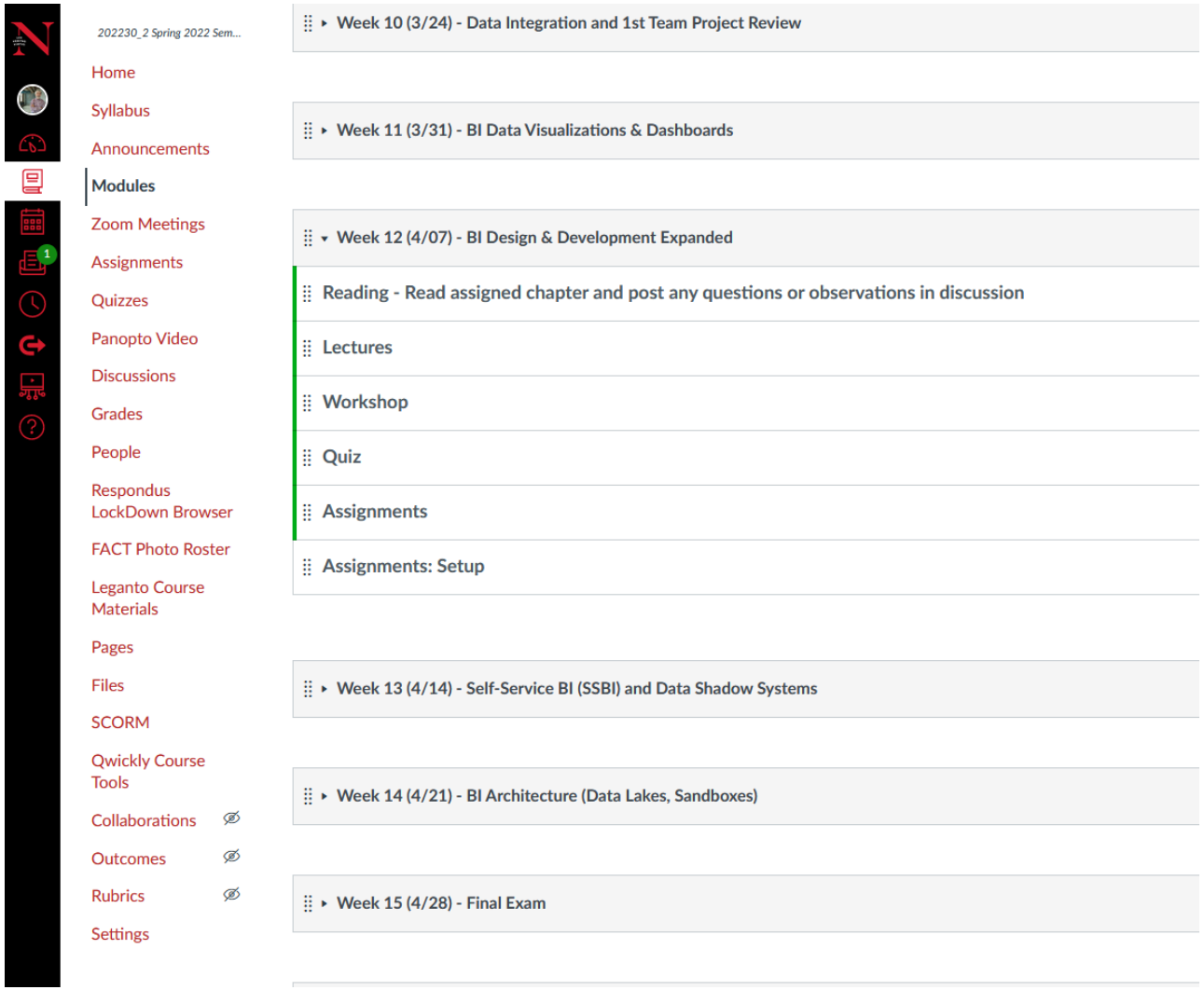
Data Samples (Projects, workshops & assignments later in semester)

My own <u>case studies</u>

Although some samples from vendors are used, I have synthesized data to better reflect real-world examples & use cases :

- Multiple data sources that overlap & are inconsistent
- Data anomalies & bad data quality
- Data gaps both missing & different granularity

Course Structure



- Canvas Modules – Weekly
  - Readings
  - Lecture
  - Workshop
  - Quiz
  - Assignments
  - Setup Assignments
  - References
  - Projects
- Zoom Meetings
- Assignments
  - Posted in Canvas Modules
  - Supplemental materials on class
- Quizzes (organized in Modules)
- Discussions (assigned in Modules)
- Recorded lecture & workshops
- Course communication

Grading Scheme

Course	Grade %	Rules of the Road
Lectures, Readings, Discussions & Quizzes	25%	Discussions: • Provide feedback, comments & analysis on lectures, readings, workshops Quizzes: • Applying what your read, heard and worked on • Mix of Multiple Choice, True/False & Matching, Problem set, Essay
Workshops, Tutorials & Assignments	20%	You learn by applying concepts & doing the work yourself, you will not learn by just watching me or your friend do workshops, assignments or projects
Exams – Midterm, Final	30%	Grading Criteria: • Do you understand the concepts and can you apply them to problems & use cases • More problems sets & essays
Team Projects	25%	Team (4 students maximum) • All team members get same grade with exceptions if… • Projects need to work on all team members' notebooks & cloud portals Grading Criteria: • Quantitative & Qualitative • Compared to other teams, top teams get bonuses
Late assignments, projects & quizzes lose 10% per week unless have an approved excuse		

DAMG 7370

Development Environment

Software Development Environment

Business Intelligence & Analytics

Tableau Online (Class Online Site)
Tableau Desktop

Power BI Service (Cloud)
(was Power BI Pro)
on NU Workspaces
Power BI Desktop

Looker Studio

Data Integration & Data Engineering

Talend Real-time Big Data Platform
Talend Studio

alteryx

Tableau Prep
Alternative for Alteryx (Mac)

Data Modeling

ER/Studio Data Architect

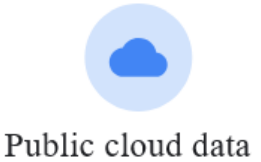
Navicat
Alternative for Idera (Mac)

Software Development Environment

Cloud Databases, Storage & Apps

Databases, Files, APIs

Cloud, Cloud VM, Notebook/Mac



- OpenData: NYC, Chicago, London…
- Google BigQuery Public Datasets
- Other sources

ATHENA IT SOLUTIONS

Why don’t we do all our class work in the cloud?

- Although *we will use the cloud*, using your notebook for much of your homework and projects allows me to assign more complex data problems that better reflect real-world situations (that recruiters will like) because…
- Database design, from a business data perspective, is independent of whether stored on the cloud or on-premise (or a database on a cloud VM)
- Because you have limited free student credits available, if you ran my assignments and projects after we get several weeks into the semester it would cost hundreds of dollars for each assignment
  - I’ll make sure you know what resources a free through your academic trials, such as BigQuery and Oracle ADW, and I’ll give you recommendations how to use cloud resources, so you’ll have more than enough for the course
- Cloud costs are not cheap. Cloud costs compare better than paying for servers, supporting infrastructure and IT to manage an environment but do not compare well to using your notebooks (free except for your time.)

You’d pay several hundred dollar$ per month to do your assignments & projects if we exclusively worked on the cloud rather than using both your notebook & the cloud

Development setup:
Windows-based development environment

- Setup development environment on your notebook
  - Windows-based OS 64-bit
  - Memory: 8Gb minimum but 16Gb is much better
    - You can upgrade memory on your notebook easily

  - May be useful to external disk storage such as SSD USB-c external drive

- Useful if at least one team member has a powerful notebook

- *Note: Many students have taken course with Macs (next slide)*

Development setup: Mac

- Macs with M1 or M2 chip: Docker Desktop for Apple silicon (M1/M2) (free)
- Macs with M1 or M2 chip: Parallels Desktop for Mac (note: there is student pricing)
  - M1/M2 chip architecture ARM not licensed to run Windows directly
  - Some Windows tools do NOT work in Parallels yet

- Create a VM on Azure with Windows
  (can use credits, students have used this at the end of the semester.)

- Students with Macs with Intel chip
  - Use Boot Camp & Install Windows 10 on 2nd OS.
    - Use Widows 10 OS for course development environment
  - Alternative: virtualization disk to host Windows above such as VMware, VirtualBox , Parallels
    - MacBook Pro with 16Gb memory is terrific but 8Gb will work
    - May be slow based on amount of memory you have on Mac because running Mac OS & VMware with Windows OS & rest of software

DAMG 7370

My Credentials

My Background

- Experience
  - 40 years relational & other databases
  - 30+ years of BI, DW & data integration
  - Consulting, IT and software engineering
- Consulting
  - Business groups, IT & software vendors
- Teaching
  - Northeastern University, Graduate COE
  - Conferences & Online courses
- Writing
  - Business Intelligence Guidebook: From Data Integration to Analytics
    - For practitioners and graduate programs
  - Over 200 published articles plus white papers, webinars, podcasts & seminars
- Connecting with thought leaders:
  - TDWI – Boston User Group Officer
  - Boulder BI Brain Trust

My customers

ATHENA IT SOLUTIONS