

INFO7370: Designing Advanced Data Architectures For Business Intelligence

Topics:

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



BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

MK
MORGAN KAUFMANN

FOREWORD BY **CLAUDIA IMHOFF**
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.

INFO7370: 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



BUSINESS INTELLIGENCE GUIDEBOOK

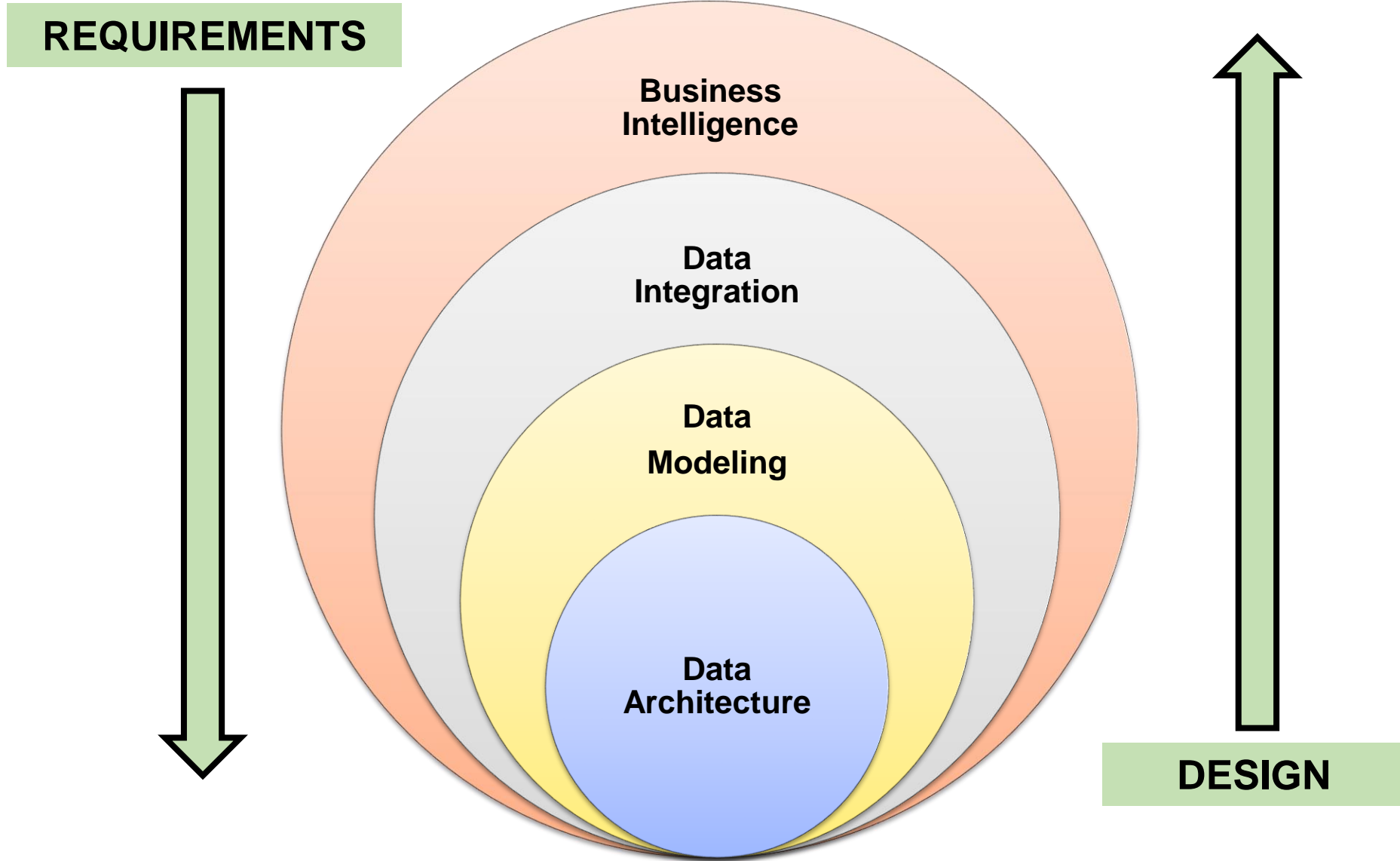
From Data Integration to Analytics

RICK SHERMAN

MK
MORGAN KAUFMANN

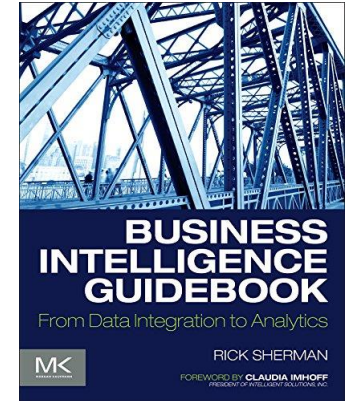
FOREWORD BY **CLAUDIA IMHOFF**
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.

Topics



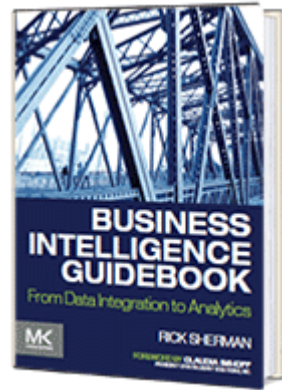
Course Textbook

- 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.

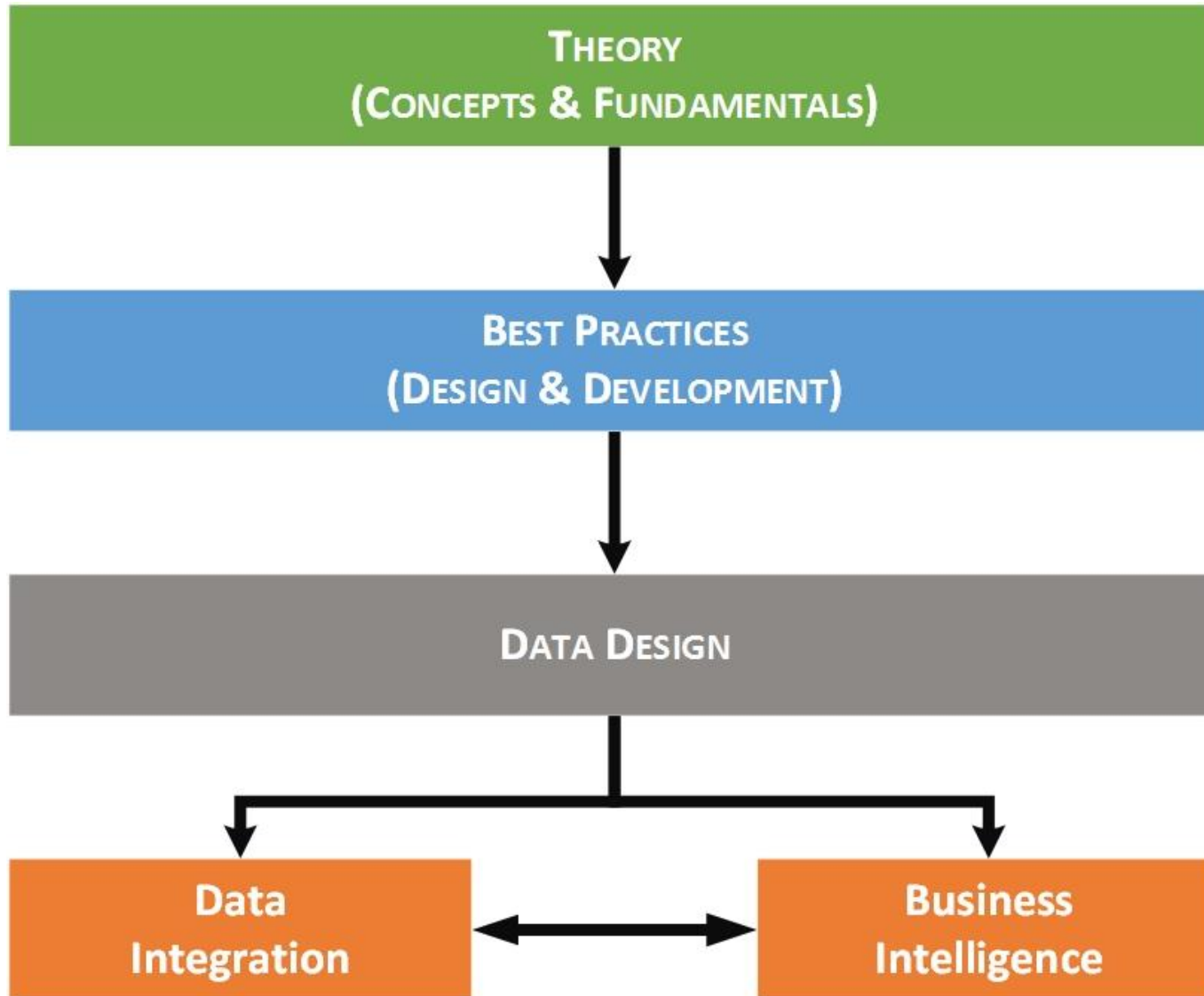


Topics by Textbook Chapter

	<u>Chapter</u>		<u>Chapter</u>
• Overview		• Architectures	
▪ State of Data in Enterprises	1	▪ Information Architecture	5
▪ Business Demand for BI	2	▪ Data Architecture	6
• Starting a BI Project		▪ Tech & Product Architectures	7
▪ Defining Requirements	3		
• Architecture	4	• Data Integration	
		▪ Design & Develop	11
		▪ Processes	12
• Architecture Framework Data Design		• Business Intelligence	
▪ E-R Modeling	8	▪ BI Applications	13
▪ Dimensional Modeling	9	▪ Design & Development	14
▪ Advanced Dimensional	10	▪ Advanced Analytics	15
		▪ Data Shadow Systems	16



Theory to Practice to Tools

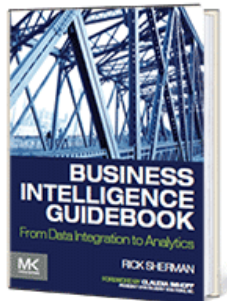


- Tools & Coding Languages become hot & then obsolete
- Fundamentals, Architecture & Best Practices & Evolve
- People, Politics & Poor Practices are a Constant!

Teaching Approach

Theory & Real-World Experience

- Textbook
- Online Lectures
- Online Software Tutorials
 - Basics
 - In-depth (as needed)
- Quizzes
- Exams



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

- Assignments
 - Workshops (in class)
 - Homework (outside of class)
 - Required tutorials (out of class)
- Development-based projects
 - Deliverables
 - Similar to Code Review



Data Samples (Projects & HW later in semester)

Although using samples from vendors, they have been significantly modified to reflect real-world examples & use cases :

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

Our own **case studies**

Software Development Environment

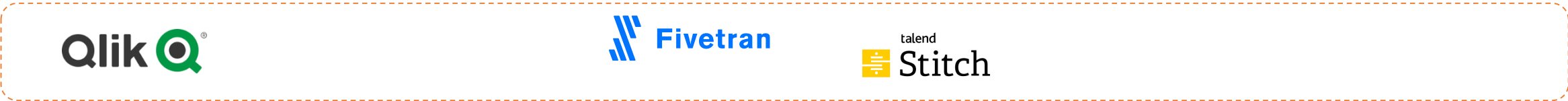
Business Intelligence & Analytics

Data Integration & Data Engineering

Databases
Used as data sources or targets

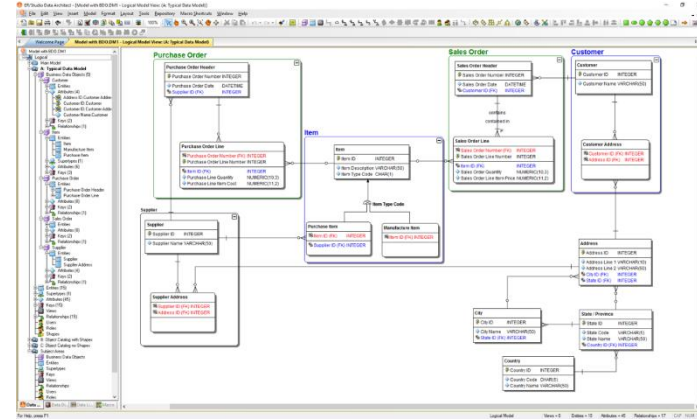


One or more (but not all) of applications listed below may be used but To Be Determined (TBD) based on student progress & interest



Software Development Environment

- Data Modeling:
 - Idera ER/Studio Data Architect
 - Database-specific data modeling



- Microsoft (on-premise):
 - SQL Server Data Tools (SSDT)
 - SQL Server Management Studio (SSMS)
 - Visual Studio
 - Azure Data Studio

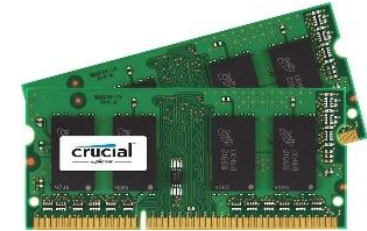


Developer

SQL Server 2017 Developer is a full-featured free edition, licensed for use as a development and test database in a non-production environment.

Development setup - Windows

- 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
 - Microsoft Office



Development setup - Mac

- Students with Macs – Two Options
 - Best Option: Use Boot Camp & [Install Windows 10 on 2nd OS](#).
 - Use Windows 10 OS for course development environment
 - Can use virtualization disk to host Windows environment 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



Parallels

Note: Many students have taken course with Macs

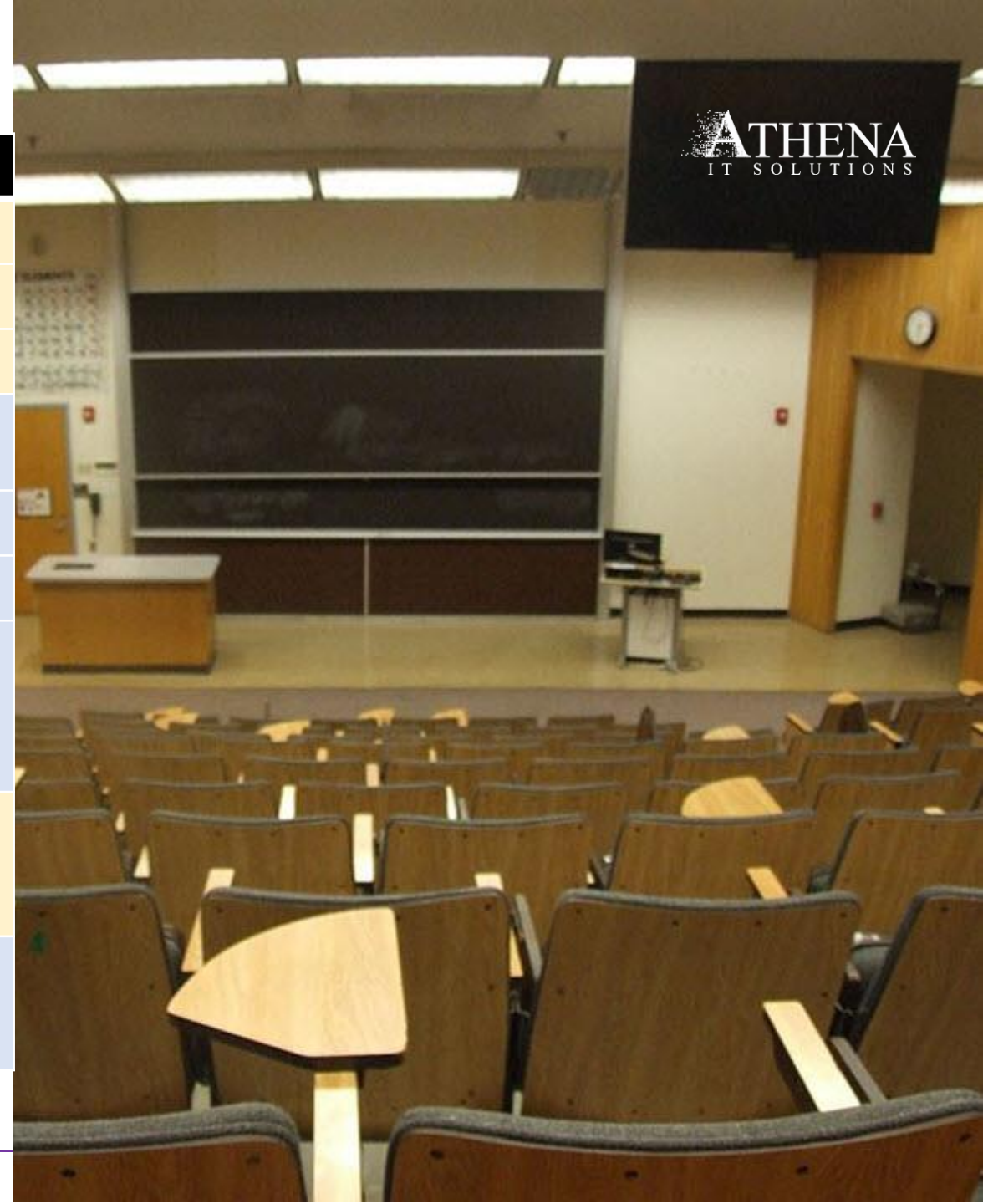
Software Development Environment

- Virtualization (if needed)
 - Mac Bootcamp – for Macs to run Windows Development environment
 - EMC VMware or Oracle VirtualBox
 - If use vendor create “Sandboxes”
 - Docker – not using yet for class



Course Organization

Delivery	Content
Online Recorded	Lectures (<u>Mandatory</u>)
	Workshops
	Tutorials (External)
Online Live Class Additional help & support as needed	Quizzes <ul style="list-style-type: none">Lectures & textbook
	Q&A on lecture
	Assignment Review (Students Present)
	Hands-on Workshop <ul style="list-style-type: none">Tools trainingApplication of ConceptsLearn best practices
Online Live Class 1-2 hours	Exams – Midterm & Final <ul style="list-style-type: none">Lectures & textbookAssignments (not tool-specific)
Online Live Class 20-30 min/team	Team Project Presentations & Reviews <ul style="list-style-type: none">1st Preliminary Review2nd Final Review



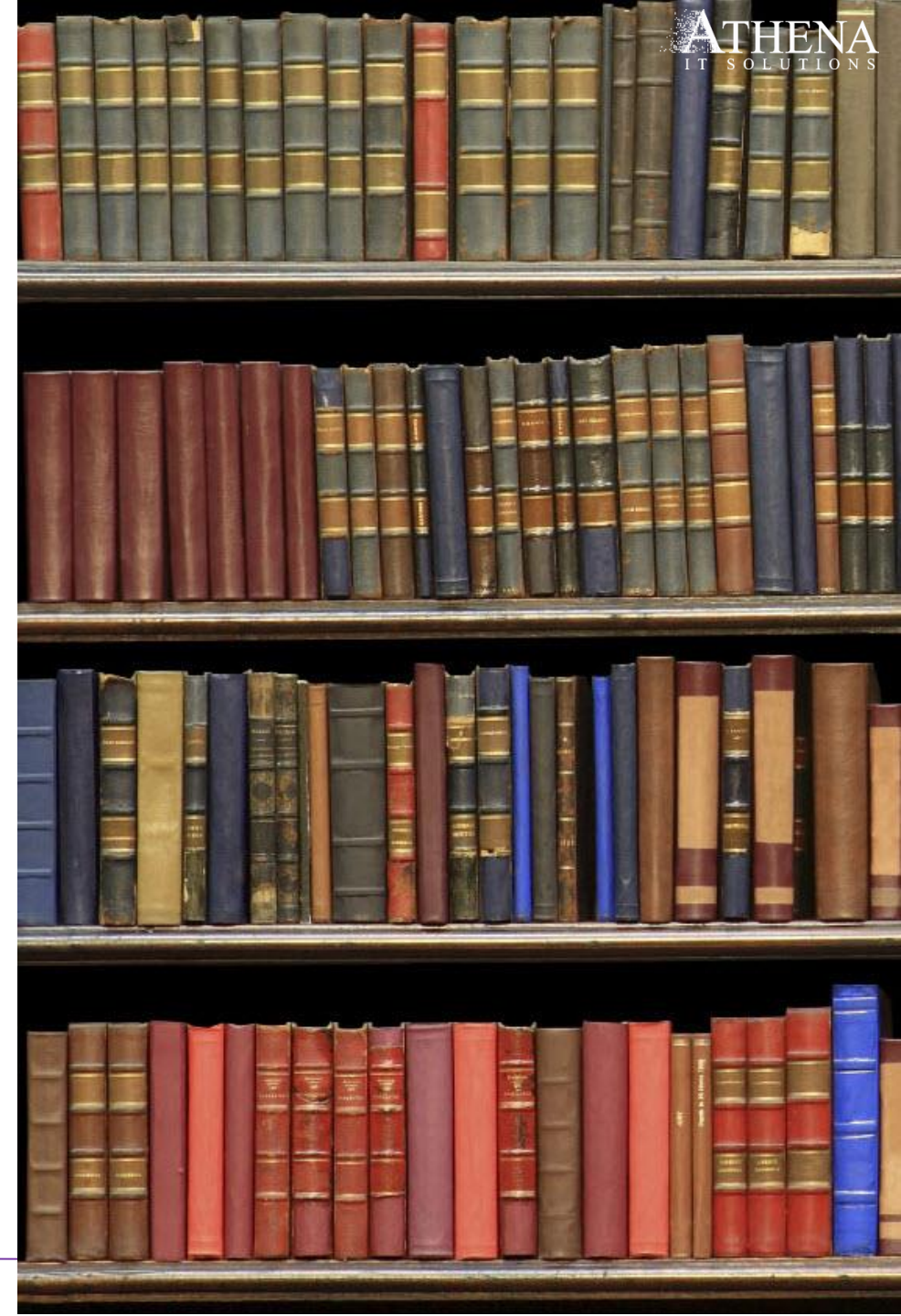
Course Communication

- Blackboard
 - Online Lectures & how-to videos
 - Submission Links - assignments, projects
 - Links – Tutorials & reference materials
- One Drive
 - Assignments & Project Materials
- Slack
 - Announcements
 - Q&A
- Web Conferencing (if/as needed)
 - Tutorials
 - Q&A
 - Project Reviews
 - Note: Zoom, Panopto or other application may be used



Course Support

- Weekly TA Sessions to review & help on
 - Course materials
 - Workshops
 - Projects
- TA will review & grade homework assignments
 - Confirm software installations
 - Confirm completion of software tutorials
 - Workshops & other assignments



Course Deliverables

Deliverables	When Due	Follow-on
Online Lectures	Prior to class	<ul style="list-style-type: none">• Q&A• Quiz
Online Training & Tutorials <ul style="list-style-type: none">• Required Assignment• Supplemental (Optional)	Prior to workshop	<ul style="list-style-type: none">• If required then submit on BB• Supplemental is provided for your interest
Workshop Assignment	End of workshop	Submit on Blackboard (BB)
Homework Assignment	Prior to next class	Submit on BB
Team Projects	Prior to presentation	Submit on BB
Software Installations & Development Setup	Prior to workshop need to verify setup.	If installing during workshop (unless it is your 1 st class) you will be asked to leave workshop.

Grading – will NOT be curved anymore!

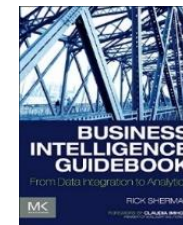
Grading Distribution	Rules of the Road
Assignments: 10% <ul style="list-style-type: none"> Workshops, Homework, Installations, Tutorials 	<ul style="list-style-type: none"> Initially individual but eventually team submissions Note: If you complete assignments on time then homework should be an A or A-
Team Projects: 25% <ul style="list-style-type: none"> Data Integration & Modeling BI, Data Integration & Data Engineering 	<ul style="list-style-type: none"> All team members: <ul style="list-style-type: none"> Get same grade with exceptions if... Projects need to work on all students' notebooks
Quizzes: 15% <ul style="list-style-type: none"> Online lectures & book 	<ul style="list-style-type: none"> Mix of Multiple Choice, True/False & Matching
Online Tutorials 10% <ul style="list-style-type: none"> Online lectures & book 	<p>You do not understand the why or how to be an architecture if you do not learn!</p>
Exams: 40% <ul style="list-style-type: none"> Midterm Final 	<ul style="list-style-type: none"> Two types of questions: <ul style="list-style-type: none"> Problem sets Essays
<p>Late assignments, projects & quizzes lose 10% per week unless there is an approved excuse</p>	

Tests are a majority of grade



My Background

- Experience
 - 30+ 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; onsite & 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
 - [DataDoghouse.com](#) blog on BI/DW industry
- Connecting with thought leaders:
 - TDWI – Boston User Group Officer
 - Boulder BI Brain Trust



My Clients:

