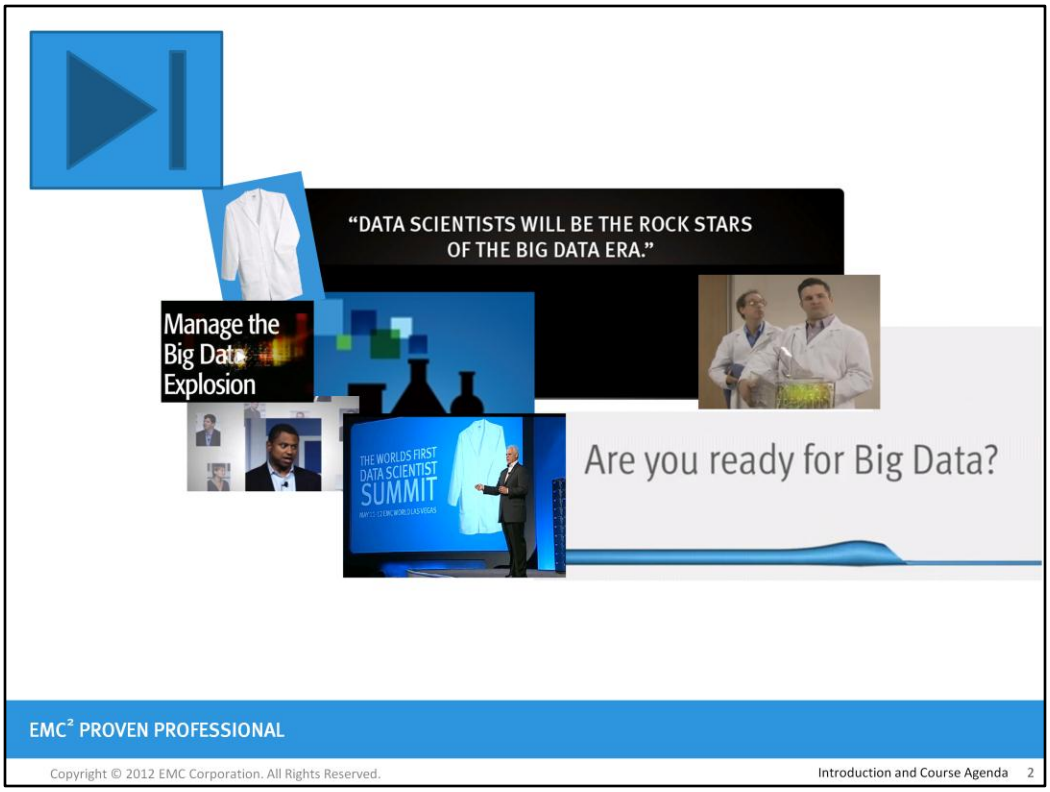
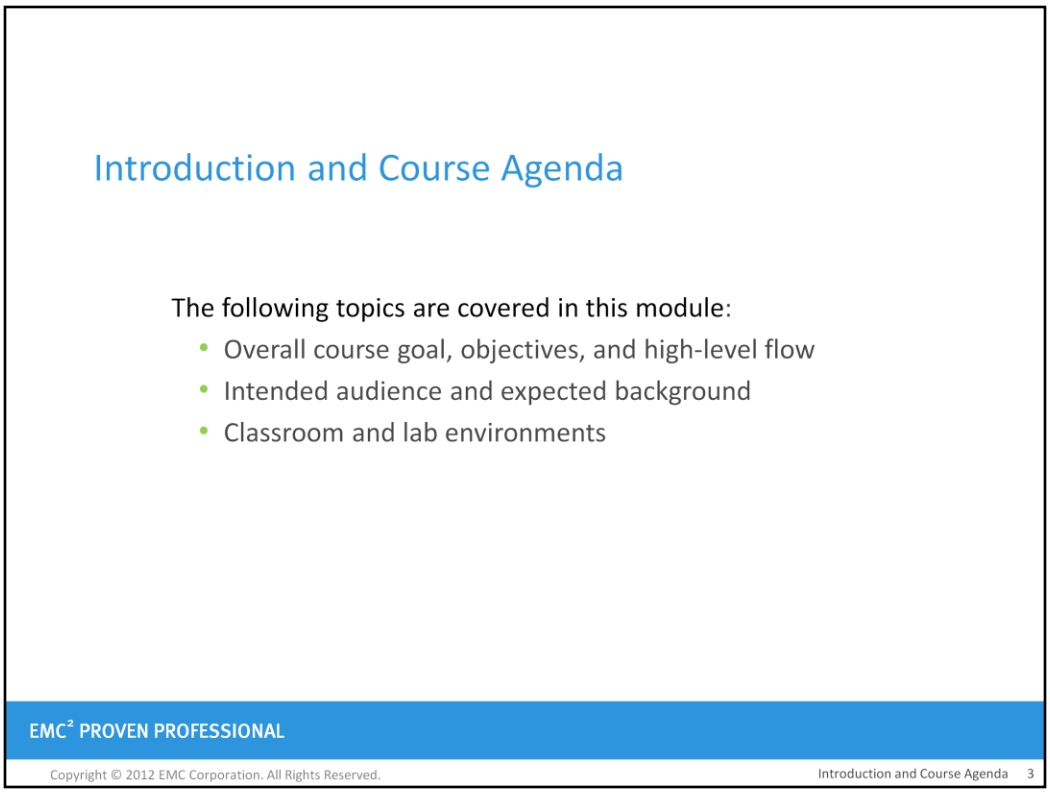
Welcome to Data Science and Big Data Analytics.

Revision Date: February 17, 2012
Revision Number: MR-1CP-DSBDA.1.2.1





This module focuses on introducing the course, its goal and objectives, its structure, and the classroom and lab environments.

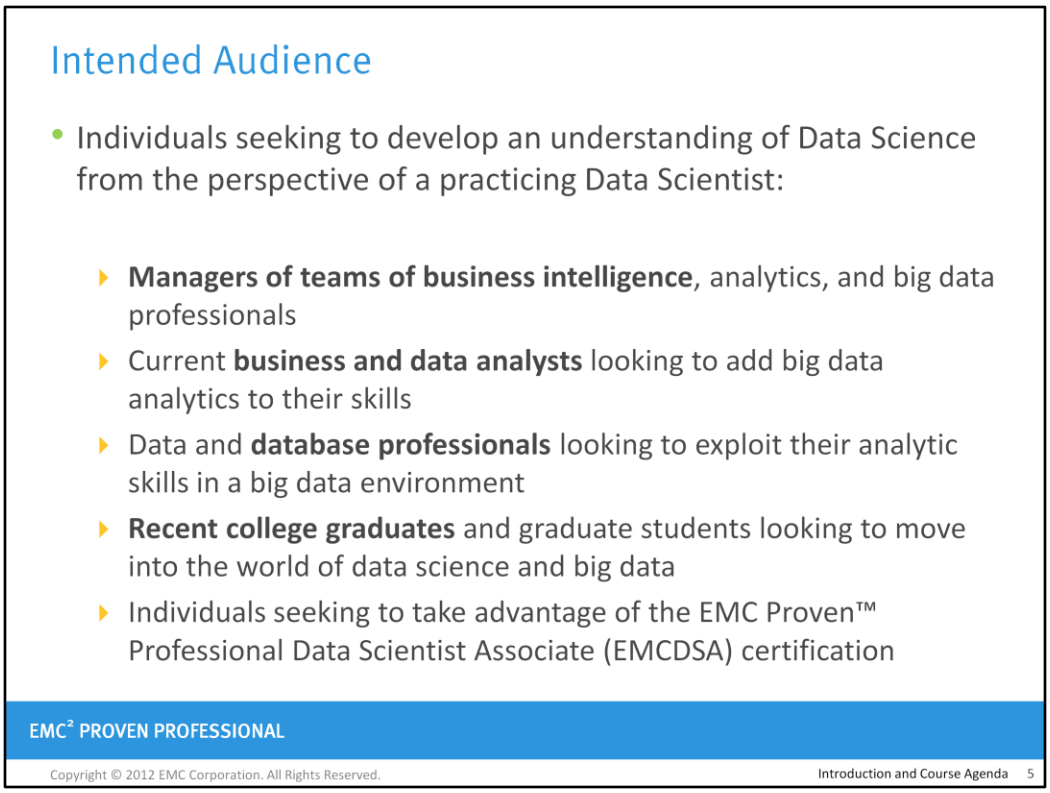
It also introduces students to other students and to the instructor(s).

Overall Course Goal

- The goal of the *Data Science And Big Data Analytics Course* is for you to be able *to* ***immediately participate as a Data Science team member*** *on big data and other analytics projects*
  - *Data Scientist p-o-v*
  - *Open*
  - *Practical*

EMC² PROVEN PROFESSIONAL

Here is the primary goal of the course. To achieve it, the course content is focused on the point of view (p-o-v) of a Data Scientist, it teaches concepts and principles in an open, vendor-neutral manner so they can be applied in any technology environment, and it provides many hands-on labs for practical experience with coaching from the instructor(s).



The audience for this course can come from a variety of backgrounds and be driven by different objectives.

Expected Background

- Strong mathematical, quantitative capability
- Experience with statistical methods and basic proficiency with a statistical software package, such as R or RStudio, Minitab, Matlab, SAS, or SPSS
- Experience with the conditioning and management of business data including databases
- Basic programming skills, preferably including SQL

EMC² PROVEN PROFESSIONAL

The lectures in this course assume students have a strong numerate background with some experience of statistical software packages and the conditioning of business data. The labs in this course assume some programming expertise (preferably with R or SQL).

Course Objectives

Upon completion of this course, you should be able to:

- Immediately participate and contribute as a data science team member on big data and other analytics projects by:
  - Deploy a structured lifecycle approach to data science and big data analytics projects
  - Reframe a business challenge as an analytics challenge
  - Apply analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results
  - Select optimal visualization techniques to clearly communicate analytic insights to business sponsors and others
  - Use tools such as R and RStudio, MapReduce/Hadoop, in-database analytics, and window and MADlib functions
- Explain how advanced analytics can be leveraged to create competitive advantage and how the data scientist role and skills differ from those of a traditional business intelligence analyst

EMC² PROVEN PROFESSIONAL

Please review these objectives of the course.

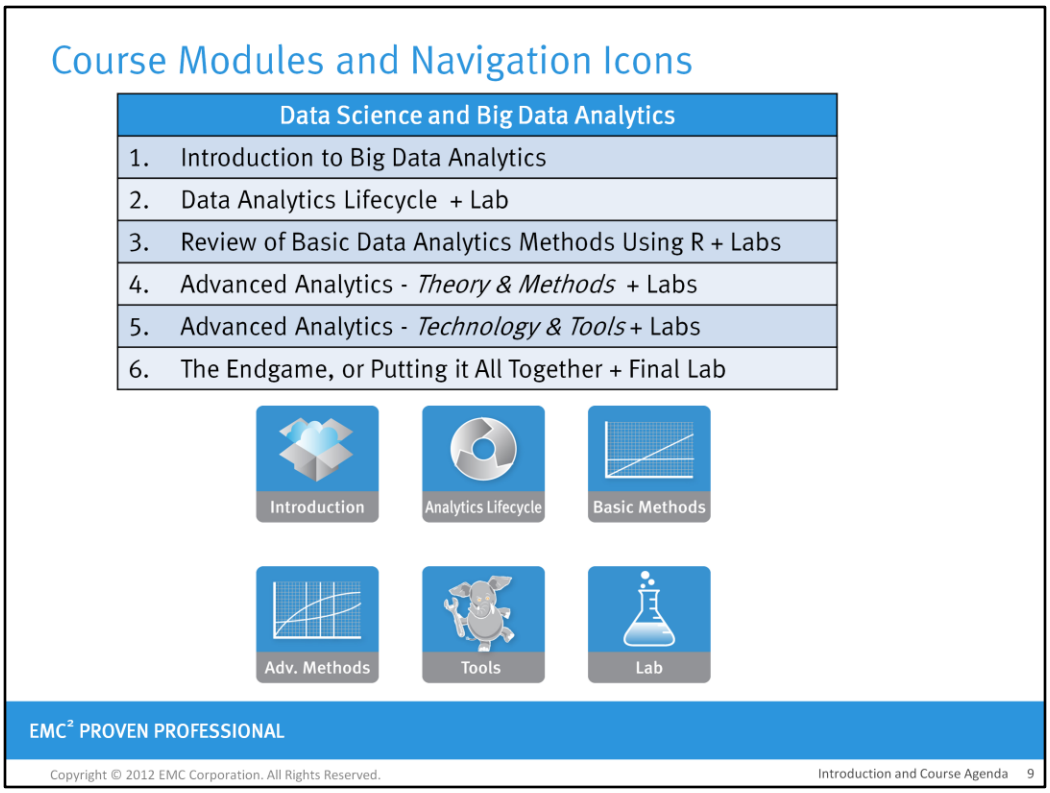
Please Briefly Introduce Yourself

- Name
- Company
- Work Location
- Role, and how analytics relates to it
- Analytics Expertise

- What you would like to achieve as a result of attending this course

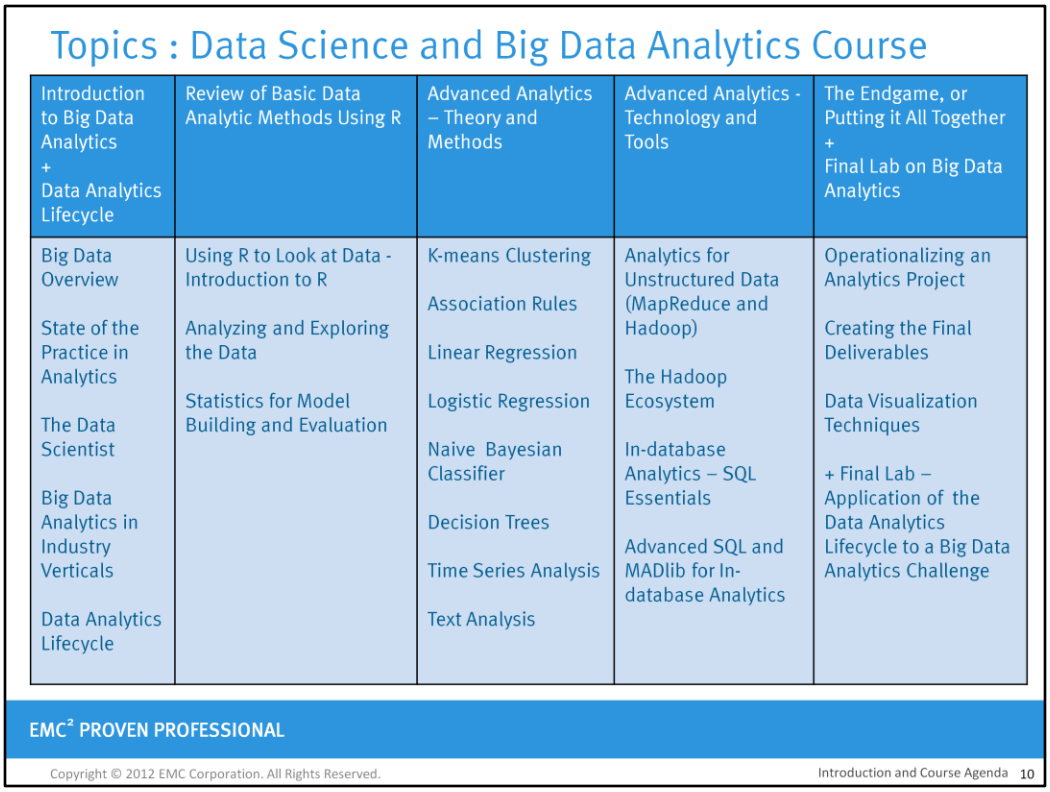
EMC² PROVEN PROFESSIONAL

Please briefly introduce yourself to the other students and the instructor(s) so we can all better understand your current role, background, and motivation for attending this class.



These is the progression through the training modules in this course. Most modules have one or more labs.

We will use the icons throughout the lectures and in the Student Resource Guide to assist in navigation.



This slide outlines the modules of the course (at the top) and the individual lesson topics that comprise those modules (below).

The flow of the course is from left to right on the diagram.

Most modules have one or more labs associated with them for students to gain practical, hands-on experience, individually or in pairs, depending on the needs and constraints of each specific class.

The Classroom Environment

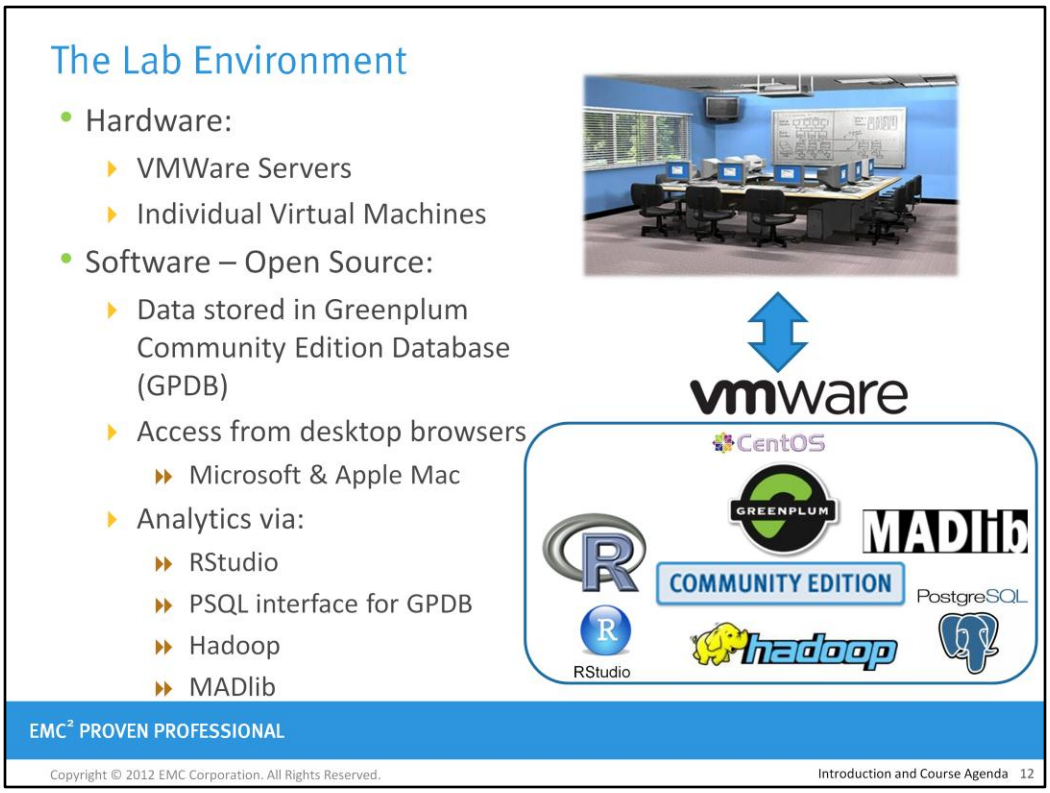
- Locations
  - Restrooms
  - Cafeteria – coffee
  - Network / Phone Access
  - Smoking Area
  - First Aid
  - Water cooler - coffee

- Hours of Class
  - 8:30am – 5pm each days
  - Lunch typically 12:00
  - Approximately 60% lecture, 40% lab

EMC² PROVEN PROFESSIONAL

The instructor will communicate the important facilities in the training center.

This course is intensive in lectures (50%) and labs (50%) and requires an 8:30 start each day.



This slide describes the technical environment that supports execution of the labs in the course.

Course Materials

- Student Reference Guide:
  - Lecture slides
  - Appendix:
    - References
    - Quick reference guides
      - LINUX
      - PSQL
      - R
- Student Lab Guide:
  - Lab instructions

EMC² PROVEN PROFESSIONAL

Course materials include a hardcopy Student Reference Guide with References in an Appendix.

There is a separate Student Lab Guide with the detailed instructions for each lab in the course.

Softcopy of “codes” are available at the Students home directories (Locations detailed in the Lab guide) to enable copy/paste (rather than keying) of commands required for the labs.

Classroom Etiquette

- **Limit usage of personal electronic devices**
  - Cell phones/PDAs (set to vibrate if possible)
  - Laptops (preferably closed during lecture)
  - If your phone rings, answer it as you step out of the classroom
- **Food and drink are allowed** in classroom
- **Inform the instructor** (and lab partner, if you have one) **of all absences** from classroom sessions
  - Excessive absences will be interpreted as non-attendance at the class
- Although we encourage collaboration during the class, ***please treat the data files, code and lab as intellectual property of EMC Education Services.***
  - Please do not redistribute without the consent of EMC Education Services

EMC² PROVEN PROFESSIONAL

Please observe these simple rules to ensure you and your co-students gain the maximum learning from this course.