



Data Science @ IBM

Education for our Data Scientists

Emily Plachy, Distinguished Engineer, IBM Global Chief Data Office
May 1, 2017

What is a Data Scientist?



▪ Data Scientists are Pioneers

- Work with business leaders to solve problems by understanding, preparing, and analyzing data to predict emerging trends
- Provide recommendations to optimize business results

▪ Who Typically Have Quantitative Academic Training Such As

- Statistics
- Operations Research
- Machine Learning
- Econometrics

▪ Use a Variety of Data and Analytics Tools and Languages

▪ Have Business Acumen

- to understand the business problem
- to influence strategic choices through data
- in deploying the solution

▪ Effectively Communicate Findings to Business Leaders Using Strong Skills

- communication
- visualization
- storytelling

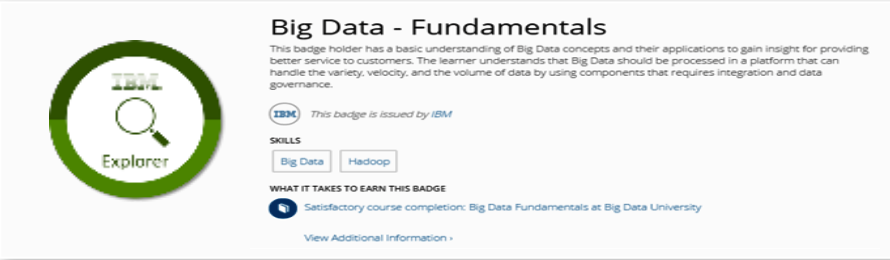
Data Science Profession with Open Badges for Skills and Certification

What are open badges?

- Contains metadata with skill tags and accomplishments
- Use to signal and verify skills and achievements
- Badging process provides feedback

Benefits:


- **Earners:** Instant recognition while building a personal brand. Captures skill profiles, while engaging employees to drive their own development
- **Clients:** Showcases verified credentials and the skills Clients value most
- **IBM:** Easy to assess employee talent and skill gaps; encourages retention



The screenshot shows an IBM Open Badge for 'Big Data - Fundamentals'. The badge is circular with a green border and the word 'Explorer' at the bottom. To the right, the text reads: 'Big Data - Fundamentals', 'This badge holder has a basic understanding of Big Data concepts and their applications to gain insight for providing better service to customers. The learner understands that Big Data should be processed in a platform that can handle the variety, velocity, and the volume of data by using components that requires integration and data governance.', 'This badge is issued by IBM', 'SKILLS: Big Data, Hadoop', and 'WHAT IT TAKES TO EARN THIS BADGE: Satisfactory course completion: Big Data Fundamentals at Big Data University'. A link 'View Additional Information' is also visible.

Easily-shared digital credentials:

- Open badges generate thousands of brand instances daily on LinkedIn, Twitter, Facebook and blogs
- Improves social connections with peers, employers and clients



Open badge users:



Establish Certification Roadmap and Framework

Data scientists expect employers to help them grow their skills and eminence

Data scientists in...

Technical skills profile

Business skills

Then

2000

SPSS

SQL

UNIX

Domain knowledge

Now

2017

Machine Learning

SPSS

R

KNIME

Spark

PySpark

Tableau

UNIX / Linux

NLP

Hadoop

Pig / Hive

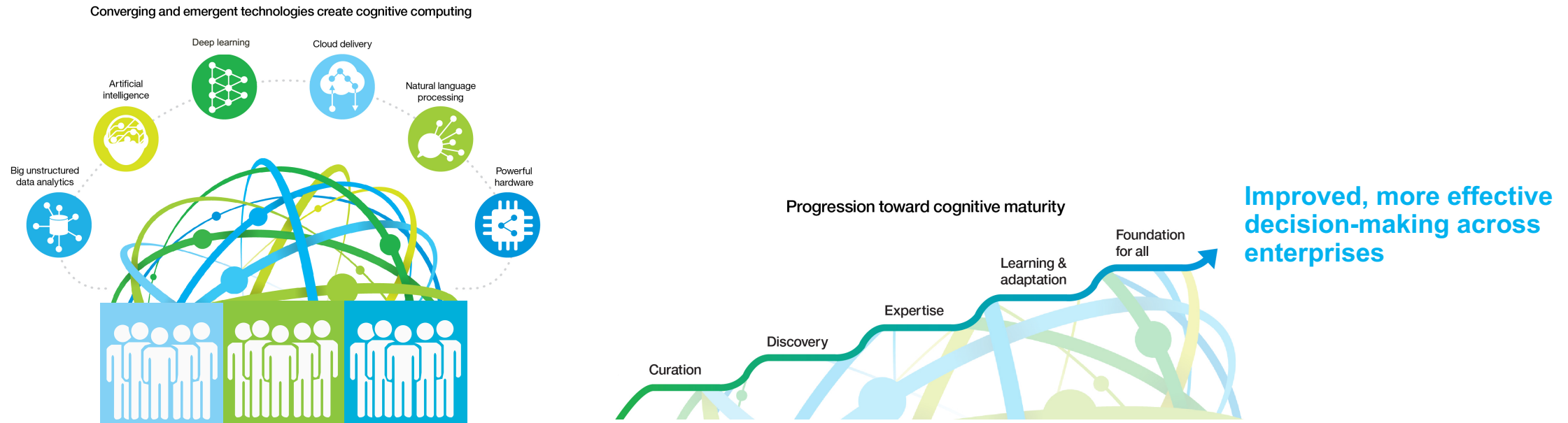
SMA

Consulting

Industry specialty

Business Transformation

Data Science Profession Evolves in the Cognitive Era



Data Science Profession evolves in Cognitive Era

- ✓ **Hybrid Analytics:** Unstructured data (text, voice, images) integrated alongside with traditional structured data sources and embedded into machine learning algorithms
- ✓ **Streaming Data:** Data Scientists use Big Data to develop analytical algorithms and provide real-time insights using Streaming Data
- ✓ **Artificial intelligence:** Solutions that augment human work by providing “what” (the insight and/or decision) and why (evidence and/or business rationale)
- ✓ **Watson Services:** Usage of API-based analytical “services”, such as Watson (send your data for processing) in conjunction with locally developed models (using client or served applications)
- ✓ **Cloud-based solutions:** Such as Watson Analytics that allows data scientists to prepare analytical solutions for consumption by business users (business users can experiment with various what-if scenarios and business options)

Data Science Education for IBM Data Scientists

- Boot camps
- Big Data and Analytics University
- Analytics product courses
- Chief Analytics Office development courses
- Analytics Education Series
- *Analytics Across the Enterprise* Book
- Cognitive Academy
- Data Science Experience
- Safari online courses and ebooks
- Kahn Academy, Coursera, and Universities



Data Science Bootcamp

TARGET AUDIENCE

- This boot camp is targeted for practitioners who intend to perform the role of a data scientist within IBM. They can be categorized:
 - Practitioners who are at the start of their data science journey and is looking to take their skills and knowledge to the next level.
 - Analytics practitioners with strong descriptive analytics background or performed roles in big data projects (like data mining, ETL).

PRE-REQUISITES

- Degree in a quantitative field.
- Working knowledge of mathematics up to algebra (either calculus or linear algebra are required).
- Background in statistics.
- Familiar with a standard statistical software package like IBM SPSS Modeler, SAS, R or Python.

OBJECTIVES

- Upon completion of the Data Science course, you should be able to:
 1. Gain insight into the role of a data scientist
 2. Understand data engineering and associated methodology.
 3. Learn tools and techniques for data transformation
 4. Analyze data using statistical modelling techniques.
 5. Implement various Machine Learning Algorithms using data analysis tools like SPSS Modeler, R, Python, etc.
 6. Ability to break down business problems using data science.
 7. Communicate insights on modelling results.

Big Data and Analytics University

Learn how Big Data & Analytics is important to our clients' success

Lobby

Learning Center

Industry Hall

Solution Hall

Auditorium

Library

Welcome to the Big Data & Analytics University!



IBM Watson and IBM Cognitive services are pioneering the cognitive domain across the globe and delivering value to clients in every industry ranging from health to games. The **Big Data & Analytics University** will be updated often to help IBMers stay tuned to of many of these advancement. Bookmark this Lobby page and check back often for new content!

Core Education: The basics or foundational information about big data and analytics.

Deeper Dive: More detailed information specific to your role.

Explore: Explore to get familiar with topics such as Watson, Big Data, Analytics Tools.



Analytics Product Courses

- IBM has a number of courses, both face to face and on-line, self-paced covering a multitude of analytics products.



IBM. **WebSphere.** software



CAO Development activities for Data Scientists: 1st half 2016

January	February	March	April	May	June
<p>Social</p> <ul style="list-style-type: none"> • Dev Session: Social listening • Bootcamp: Strategy-Ruslan 	<p>Watson</p> <ul style="list-style-type: none"> • Dev Session: Watson Analytics • Bootcamp: Watson Analytics use cases • Bootcamp: Strategy 	<p>Cognitive</p> <ul style="list-style-type: none"> • Dev Session: BlueMix • Bootcamp: Cognitive build updates • Bootcamp: Strategy 	<p>Agile</p> <ul style="list-style-type: none"> • Dev Session: Whitewater/Agile method for consultants • Mandatory Agile training for managers • Bootcamp- NLP/Social media analytics 	<p>Cloud</p> <ul style="list-style-type: none"> • Dev Session: Cloud strategy • Bootcamp: Watson API's in SPSS 	<p>Prescriptive Analytics</p> <ul style="list-style-type: none"> • Dev. Session: Causality • Bootcamp: Advanced methods • Bootcamp: MPW • New Hire Training



CAO Development activities for Data Scientists: 2nd half 2016

July	August	September	October	November	December
<p>Strategy</p> <ul style="list-style-type: none"> • Dev Session: Storytelling • Bootcamp: Problem Solving 	<p>IBM Offerings</p> <ul style="list-style-type: none"> • All Hands: • Bootcamps (x3): D3 	<p>IBM Innovation</p> <ul style="list-style-type: none"> • Bootcamp (x1): D3- • Bootcamp: Media Analytics 	<p>Machine Learning</p> <ul style="list-style-type: none"> • Dev Session: machine learning • Bootcamp (x5): Agile sessions 	<p>Recruiting</p> <ul style="list-style-type: none"> • Dev. Session: Intro to Interviewing • Bootcamp: Case Interview Prep 	<p>IBM Leadership</p> <ul style="list-style-type: none"> • Speaker Series • Speaker Series • Speaker Series • Speaker Series



Analytics Education Series

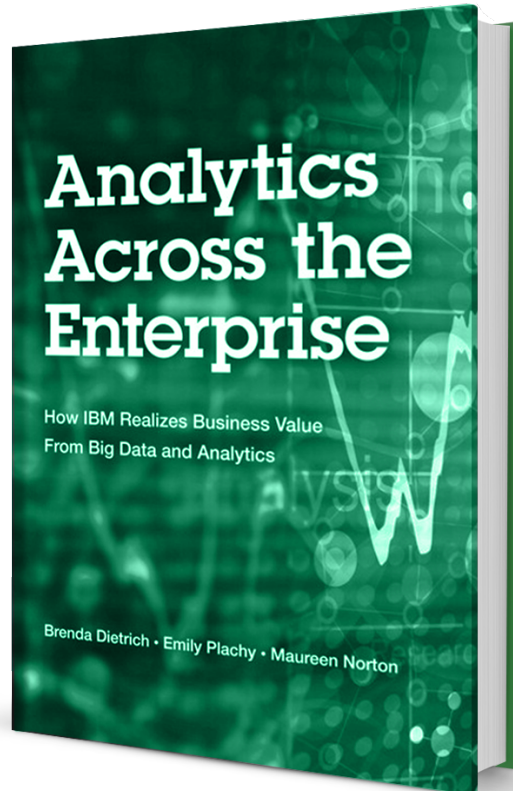
Audience	Content	Characteristics
Employees who want to become Analytic Practitioners or Data Scientists.	Introduce analytics topics to those that need a foundation before using Business Analytics tools.	30 plus 1-hour video lectures by experts across IBM. Select the topics of interest to you.

Example Topics:

- Data Visualization
- Data Modeling
- Machine Learning
- Time Series Analysis
- Artificial Neural Networks and Genetic Algorithms
- Text and Natural Language Processing
- Entity Resolution
- Spatio-Temporal Analytics
- Graph Analytics



Book: *Analytics Across the Enterprise: How IBM Realizes Business Value from Big Data and Analytics*, by Brenda Dietrich, Emily Plachy and Maureen Norton



“... give IBM credit: It practices what it preaches.... This book is valuable for anyone who wants to know how to apply analytics profitably.”

~ Wayne Eckerson, principal consultant, Eckerson Group and author, *Secrets of Analytical Leaders: Insights from Information Insiders*.

32 case studies spanning Human Resources, Supply Chain, Finance, Services, and Sales.

Audience:

- Business leaders across industries
- Academia – MBA / Analytics Students
- IBM employees

Cognitive Academy

Cognitive Academy

Home

Courses

Share

Guilds

Team

News

Search

Pick one of our selected learning paths, and get started today!

Everyone

Business Analyst or
SME

Cognitive
Developer

Data
Specialist

**Data
Scientist**



Database Fundamentals

This series covers all aspects of relational databases, NoSQL and DBaaS offerings.

Your Learning

1

20hrs 20min

Database Mining

Learn how to Analyze Text, Discover Patterns and Visualize Data.

Your Learning

2

5hrs

Machine Learning - Fundamentals

An introduction to machine learning.

Your Learning

3

5hrs

Data Visualization

A collection of courses focussed on data visualization.

Your Learning

4

8hrs

Machine Learning - Advanced

Learn to analyze large and complex datasets, create systems that adapt and improve over time.

Your Learning

5

8hrs

Data Science Experience



Learn

Get started or get better with built-in learning.



Create

Use the best of open source tooling with IBM innovation.



Collaborate

Work smarter using community, work faster with your team.

A Growing Set of Data Science Tools

Jupyter Notebooks

Create and collaborate on Python, R, and Scala notebooks that contain code and visualizations.

RStudio

Jumpstart your R experience with a free, open-source RStudio tool.

Machine Learning (Coming Soon)

Create, train and deploy machine learning models.



Wrap up

- What's remaining?
 - Data Science is broad and a rapidly evolving field. We will never be done with education.
 - We need more cognitive education
 - Add a mentoring program and an apprentice program
- What's motivating IBM to create this education?
 - Top Data Science talent wants a work environment where they can develop and grow their skills
- What skill gaps does IBM see with newly hired data scientists?
 - Cognitive skills, e.g., ML and AI; leveraging unstructured data
 - Good communication skills; curiosity, patience
 - Strong quantitative foundation

