DATA SCIENCE AND BIG DATA ANALYTICS
An ‘open’ course to unleash the power of Big Data

“We live in a data-driven world. Increasingly, the efficient operation of organizations across sectors relies on the effective use of vast amounts of data. Making sense of big data is a combination of organizations having the tools, skills and more importantly, the mindset to see data as the new “oil” fueling a company. Unfortunately, the technology has evolved faster than the workforce skills to make sense of it and organizations across sectors must adapt to this new reality or perish.”

- Andreas Weigend, Ph.D Stanford, Head of the Social Data Lab at Stanford, former Chief Scientist, Amazon.com

DATA SCIENCE AND BIG DATA ANALYTICS COURSE
An ‘open’ course and certification focused on concepts and principles applicable to any technology environment and industry.

This course is intended for:
• Business and data analysts looking to add big data analytics skills
• Managers of business intelligence, analytics, or big data groups
• Database professionals looking to enrich their analytic skills
• College graduates considering data science as a career field

The course provides a hands-on* practitioner’s approach to the techniques and tools required for analyzing Big Data.

The course is designed to enable students to:
• Become an immediate contributor on a data science team
• Assist reframing a business challenge as an analytics challenge
• Deploy a structured lifecycle approach to data analytics problems
• Apply appropriate analytic techniques and tools to analyze big data
• Tell a compelling story with the data to drive business action
• Use open source tools such as R, Hadoop, and Postgres
• Prepare for EMC Proven™ Professional Data Scientist certification

Visit http://education.EMC.com/DataScience to
• Register for the Data Science and Big Data Analytics StarterKit
• Access additional information on the course and the certification

* The Video ILT includes recorded walk-through lab demonstrations.
# DATA SCIENCE AND BIG DATA ANALYTICS COURSE OUTLINE

Applying a hands-on practitioner’s approach to the techniques and tools required for Big Data Analytics.

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Analytics Lifecycle</th>
<th>Basic Methods</th>
<th>Adv. Methods</th>
<th>Tools</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Data Overview</td>
<td>Key roles for a successful analytics project</td>
<td>Introduction to R</td>
<td>K-Means Clustering</td>
<td>Using MapReduce/Hadoop for analyzing unstructured data</td>
<td>How to operationalize an analytics project</td>
</tr>
<tr>
<td>State of the practice in analytics</td>
<td>Main phases of the lifecycle</td>
<td>Analyzing and exploring data with R</td>
<td>Association Rules</td>
<td>Hadoop ecosystem of tools</td>
<td>Creating the Final Deliverables</td>
</tr>
<tr>
<td>The role of the Data Scientist</td>
<td>Developing core deliverables for stakeholders</td>
<td>Statistics for model building and evaluation</td>
<td>Linear and Logistic Regression</td>
<td>Naïve Bayesian Classifier</td>
<td>Data Visualization Techniques</td>
</tr>
<tr>
<td>Big Data Analytics in industry verticals</td>
<td>End-to-end data analytics lifecycle</td>
<td>Using R to execute basic analytics methods</td>
<td>Decision Trees</td>
<td>Time Series Analysis</td>
<td>Hands-on Application of Analytics Lifecycle to a Big Data Analytics Problem</td>
</tr>
</tbody>
</table>

## Preparing for certification?

- Use course materials or the Data Science and Big Data Analytics book to prepare for your certification exam. Make sure to register for the E20-007 Associate-level exam.

![QR Code](image-url)