

# RX-M - Python + Data Science on AWS

---

<b>Code:</b>	RX-M-PDS-AWS
<b>Length:</b>	4 days
<b>URL:</b>	<a href="#">View Online</a>

---

## DAY 1: PYTHON BOOTSTRAP FOR FINANCE

This intensive one day hands on course is designed to provide technology and data science professionals with a rapid introduction to Python. This course covers the essential elements of the Python programming language needed to use Python in a range of Data Science settings. Basic coding skills are combined with a survey of some of the more popular analytics and stats resources for Python such as numpy, scipy, matplotlib and pandas. Attendees will leave with a clear understanding of Python basics and practical scripting skills.

## DAY 2-4: DATA-SCIENCE on AWS

Data science is an interdisciplinary field focused on developing insights from structured and unstructured data. Apache Hadoop and Apache Spark are some of the most important and most used tools in the data science field. The AWS EMR platform makes these tools accessible and scalable within the integrated AWS environment. This course will teach attendees how to use Apache Hadoop and Spark to solve sophisticated data science problems, producing valuable insights in a wide range of scenarios in the Amazon AWS environment. Day one focuses on data science basics, including data acquisition, scrubbing, manipulation and storage, as well as a general overview of data science applications and the analytics and machine learning processes typically used. A number of practical use cases are examined during class and lab sessions where students will gain exposure to S3, Sqoop and other tools. Day two focuses on AWS EMR and its ecosystem along with the types of data science applications typically handled by the Hadoop platform. The course outlines the statistical methods used to produce actionable business insights with Map Reduce, MetaStore/Presto, Mahout and other tools. Day three begins with an overview of the Apache Spark platform. Attendees will learn how to work with RDDs, DataFrames and SparkSQL, implementing recommendation engines and performing other common data science tasks using Spark batch, streaming, graph and machine learning capabilities. Upon course completion attendees will have a clear understanding of data science, its typical use cases and how data science is performed using a range of tools in the AWS EMR ecosystem.

## Skills Gained

- This course is designed to get technical staff up and running with Python in a financial environment.
- This course is designed to provide attendees with a comprehensive introduction to data science with Apache Spark and Hadoop on AWS.

## Prerequisites

Each attendee will require the ability to run a 64 bit virtual machine (provided with the course) with good internet access.

## Course Details

### Course Outline

Day 1

- Python Overview (console I/O, data types, conditionals and loops)

- Creating Programs (program structure, command line arguments)
- Working with Functions
- Using Python Packages (Numpy, SciPy, Matplotlib, Pandas)

#### Day 2

- Data Science Overview
- Structured and Unstructured Data
- Data Acquisition and Transformation
- Data Analysis and Machine Learning

#### Day 3

- Map Reduce Fundamentals
- Common Hadoop use cases
- Machine Learning with Amazon Machine Learning
- NLTK and Natural Language Processing

#### Day 4

- Apache Spark Overview
- SparkSQL
- Working with MLlib
- Spark Streaming

---

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy

Get Your Free Copy Now