



## Big Data- Python/Spark AIA only

	MON	TUE	WED	THU	FRI	ENVIRONMENT
<b>Week 1</b> <b>Database-RDBMS</b>	<ul style="list-style-type: none"> <li>Quick review of RDBMS concepts</li> <li>Quick review of DDL, DML, DQL, DCL</li> <li>Quick review of Data Types</li> <li>Executing CRUD operations</li> </ul>	<ul style="list-style-type: none"> <li>Defining Schema</li> <li>CREATE and ALTER</li> <li>Constraints</li> <li>Referential Integrity</li> <li>Table Relationships/Multiplicity</li> <li>What is a view, advantages of views</li> <li>Creating/Modifying views</li> </ul>	<ul style="list-style-type: none"> <li>Group by</li> <li>Having</li> <li>Aggregate Functions</li> <li>Normalization</li> <li>Atomicity</li> <li>Consistency</li> <li>Isolation</li> <li>Durability</li> </ul>	<ul style="list-style-type: none"> <li>Inner / Outer Joins</li> <li>Left / Right Joins</li> <li>Set Operations</li> <li>Subqueries</li> <li>DROP vs TRUNCATE</li> </ul>	<ul style="list-style-type: none"> <li>SQL Assignment</li> </ul>	
<b>Week 2</b> <b>Python</b>	Quality Control Audit  Trainer interviews  Written evaluation  <ul style="list-style-type: none"> <li>Interpreter vs Compiler</li> <li>REPL</li> <li>Identifiers &amp; Keywords</li> <li>Simple/Compound Statements</li> <li>Types &amp; Values</li> </ul>	<ul style="list-style-type: none"> <li>Namespaces</li> <li>String Literals</li> <li>Operators</li> <li>Lambdas</li> <li>Functions</li> <li>Exceptions</li> <li>OOP &amp; Python</li> </ul>	<ul style="list-style-type: none"> <li>Lists</li> <li>Sets</li> <li>Tuples</li> <li>Dictionaries</li> <li>Classes &amp; Objects</li> <li>Variable Scope/lifetime</li> </ul>	<ul style="list-style-type: none"> <li>File Handling: open()/close()</li> <li>Logging</li> <li>Collections</li> <li>Datetime</li> <li>PyPI &amp; pip</li> <li>Modules</li> <li>Unit Testing</li> </ul>	<ul style="list-style-type: none"> <li>Pylint</li> <li>Matplotlib</li> <li>Pandas</li> <li>Numpy</li> </ul>	
<b>Week 3</b> <b>Scala / AWS Fundamentals</b>	Quality Control Audit  Trainer interviews  Written Evaluation  <ul style="list-style-type: none"> <li>Installing JDK, Scala, SBT</li> <li>Introduction to Scala</li> <li>Introduction to Functional Programming with Scala</li> <li>Variables and Values</li> </ul>	<ul style="list-style-type: none"> <li>Conditions</li> <li>Loops</li> <li>Classes &amp; Objects</li> <li>Structure of Classes</li> <li>Expressions vs Statements</li> <li>Tuples</li> </ul>	<ul style="list-style-type: none"> <li>Functions</li> <li>Pure and Impure Functions</li> <li>Higher Order Functions</li> <li>Exception Handling</li> <li>Error Handling with Try</li> <li>Pattern Matching</li> </ul>	<ul style="list-style-type: none"> <li>Scala Collections</li> <li>List/Set/Map</li> <li>Cloud and AWS Fundamentals</li> <li>EC2 basics</li> <li>S3 basics</li> <li>RDS basics</li> </ul>	<ul style="list-style-type: none"> <li>Coding Assignment</li> </ul>	<ul style="list-style-type: none"> <li>VS Code</li> <li>IntelliJ</li> </ul>

	MON	TUE	WED	THU	FRI	ENVIRONMENT
<b>Week 4</b> <b>Spark &amp; SparkSQL</b>	Written evaluation Trainer interviews Quality Control Audit <ul style="list-style-type: none"> <li>Introduction to Spark</li> <li>Spark Ecosystem</li> <li>Hadoop vs. Spark</li> <li>Spark Setup</li> <li>Local vs Cluster Mode</li> </ul>	Project <ul style="list-style-type: none"> <li>Introduction to RDDs</li> <li>Working with Key/Value pairs</li> <li>Basic RDD Operations</li> <li>Data Loading - Conceptual</li> <li>Actions</li> <li>Transformations</li> <li>Shared Variables</li> </ul>	Project <ul style="list-style-type: none"> <li>Spark-submit Overview</li> <li>Launching a Spark Job using Scala</li> <li>Key-Value Pair RDDs</li> <li>Data Loading and Saving Through RDDs</li> </ul>	Project <ul style="list-style-type: none"> <li>Introduction to SparkSQL</li> <li>SparkSession vs SparkContext</li> <li>DataFrames</li> <li>DataSets</li> <li>SparkSQL Aggregate Functions</li> </ul>	Project <b>Portfolios Due to QC</b> <ul style="list-style-type: none"> <li>Working with JSON datasets</li> <li>Working with Parquet files</li> <li>Coding Assignment</li> </ul>	
<b>Week 5</b> <b>Data Warehousing Concepts / Apache Airflow / AWS EMR</b>	Project Quality Control Audit Written Evaluation Mock Interviews <b>OLAP vs. OLTP</b> <b>Data Warehousing and ETL</b> <b>Star Schema: Fact and Dimension Tables</b> <ul style="list-style-type: none"> <li>Dimensional Modeling Concepts</li> </ul>	Project <ul style="list-style-type: none"> <li>Spark Cluster Manager</li> <li>Driver Class Configuration</li> <li>Executors</li> <li>AWS EMR</li> <li>Running Spark Jobs with EMR</li> </ul>	Project <ul style="list-style-type: none"> <li>Configure Number of Executors</li> <li>Spark Cluster Configuration</li> <li>Configure Memory: Driver &amp; Executors</li> <li>Spark Caching</li> </ul>	Project <b>Apache Airflow</b>	Project <ul style="list-style-type: none"> <li>SQL Assignment</li> <li>Spark Assignment</li> </ul>	
<b>Week 6</b> <b>AWS/Project</b>	Project Written evaluation Trainer interviews Quality Control Audit	Project	Project	Project <ul style="list-style-type: none"> <li>Project Showcase</li> </ul>		<ul style="list-style-type: none"> <li>AWS</li> <li>AWS EMR</li> </ul>

**PROJECT**

Project

**TECHNOLOGIES**

Scala, Spark



Copyright © 2022 Revature, LLC. All Rights Reserved.

By viewing this document, you agree that under copyright law all content displayed is the sole intellectual property of Revature, LLC, a technology advancement and consulting company based in Reston, VA. All content generated by a representative of Revature which is used for the company's advancement, development, or have otherwise been developed at the company's request, are the sole property of the company. No intellectual property may be reproduced, distributed, altered, or shared without the explicit permission from a representative of Revature.