

	MON	TUE	WED	THU	FRI	ENVIRONMENT
Week 1 Scala	<ul style="list-style-type: none"> • Introduction to Scala • Introduction to functional programming • Variables and values 	<ul style="list-style-type: none"> • Conditions • Loops • Classes and Objects • Structure of classes • expressions vs statements • Tuples 	<ul style="list-style-type: none"> • functions • pure and impure functions • higher order functions • Exception Handling • Error handling with Try • Pattern Matching 	<ul style="list-style-type: none"> • Scala collections • List/Set/Map • Filtering & transformation • FlatMap • Method Notations • Inheritance 	<ul style="list-style-type: none"> • Concurrency Overview • Futures • Handling Failures 	<ul style="list-style-type: none"> • VS Code
Week 2 Database-MongoDB	<div>Written evaluation</div> <div>Trainer interviews</div> <div>Quality Control Audit</div> <ul style="list-style-type: none"> • Intro to RDBMS • DML, DDL, DQL • Basic SQL queries • Aggregate functions • Multiplicity 	<ul style="list-style-type: none"> • NoSQL • MongoDB • Connection String • Schemas and Relations 	<ul style="list-style-type: none"> • Mongo Collections • Mongo Documents • Insert • Find 	<ul style="list-style-type: none"> • Query • Sort • Limit 	<ul style="list-style-type: none"> • Update • Delete • Indexing 	
Week 3 UNIX/Hadoop Fundamentals	<div>Project 1</div> <div>Written evaluation</div> <div>Trainer interviews</div> <div>Quality Control Audit</div> <ul style="list-style-type: none"> • Intro to Open Source Software • Linux commands brainstorm • Root [/] vs Home [~] • Commands: mkdir, rm, cp, mv, cd, ls, cat, grep, echo 	<div>Project 1</div> <ul style="list-style-type: none"> • Commands: df, fdisk, sfdisk, cfdisk, lsblk, blkid, mdadm • File Editors - vim, nano • Search and Replace commands • Intro to SSH (credentials/private key) • find, history, ps, kill, sed, nohup 	<div>Project 1</div> <ul style="list-style-type: none"> • Intro to data evolution • Intro to big data • Hadoop Ecosystem Introduction • Intro to HDFS • Evolution of Hadoop • HDFS Commands 	<div>Project 1</div> <ul style="list-style-type: none"> • Introduction to MapReduce • Mapper/Intermediate/Reducer phases • Partitioners • Combiners 	<div>Project 1</div> <ul style="list-style-type: none"> • YARN Overview • InputFormats 	

	MON	TUE	WED	THU	FRI	ENVIRONMENT
Week 4 Hive/Spark Fundamentals	Project 1	Project 1	Project 1	Project 1	Project 1	
	Written evaluation	<ul style="list-style-type: none"> Data Types 	<ul style="list-style-type: none"> Hive UDFs - Overview 	<ul style="list-style-type: none"> Introduction to RDDs 	<ul style="list-style-type: none"> Actions 	
	Trainer interviews	<ul style="list-style-type: none"> Managed vs External tables 	<ul style="list-style-type: none"> Introduction to Spark 	<ul style="list-style-type: none"> Basic RDD operations 	<ul style="list-style-type: none"> Shared variables 	
	Quality Control Audit	<ul style="list-style-type: none"> Partitions and Buckets 	<ul style="list-style-type: none"> Spark Ecosystem 	<ul style="list-style-type: none"> Local vs Cluster mode 	<ul style="list-style-type: none"> Accumulators 	
	<ul style="list-style-type: none"> Introduction to Hive Basic Hive Queries Hive Commands Table properties 	<ul style="list-style-type: none"> Indexes Joins Overview Views 	<ul style="list-style-type: none"> Hadoop vs Spark Spark Setup 	<ul style="list-style-type: none"> Working with Key/Value pairs Transformations 	<ul style="list-style-type: none"> Data Loading and Saving Through RDDs Key-Value Pair RDDs 	
Week 5 Spark Fundamentals	Project 1	Project 1	Project 1	<ul style="list-style-type: none"> Configure number of executors Spark cluster configuration Configure memory: Driver & executors Spark jobs troubleshooting and exceptional handling 	<ul style="list-style-type: none"> Spark caching Memory management In-memory processing 	<ul style="list-style-type: none"> AWS EMR
	Quality Control Audit	<ul style="list-style-type: none"> Creating Spark EMR cluster 	<ul style="list-style-type: none"> Spark Cluster Manager 			
	Trainer interviews	<ul style="list-style-type: none"> Spark Cluster mode 	<ul style="list-style-type: none"> Running Spark job on EMR 			
	Written Evaluation	<ul style="list-style-type: none"> Introduction to YARN 	<ul style="list-style-type: none"> Driver class configuration Executors Broadcast variables 			
	<ul style="list-style-type: none"> AWS Introduction AWS EC2 AWS S3 					
Week 6 Spark SQL/DataFrames	Project 2	Project 2	Project 2	Project 2	Project 2	
	Quality Control Audit	<ul style="list-style-type: none"> Introduction to DataFrames 	<ul style="list-style-type: none"> Creating DataSets 	<ul style="list-style-type: none"> Bucketing 	<ul style="list-style-type: none"> Working with JSON Datasets 	
	Trainer interviews	<ul style="list-style-type: none"> Entry point: SparkSession 	<ul style="list-style-type: none"> Working with RDDs 	<ul style="list-style-type: none"> Sorting and Partitioning 	<ul style="list-style-type: none"> Working with Parquet Files 	
	Written Evaluation	<ul style="list-style-type: none"> Creating DataFrames 	<ul style="list-style-type: none"> Using DataFrame aggregate functions Spark SQL Context vs Hive Context 	<ul style="list-style-type: none"> Unstructured social media data processing using Spark 		
	<ul style="list-style-type: none"> Introduction to Spark SQL Introduction to DataSets Agile/Scrum topics 					

	MON	TUE	WED	THU	FRI	ENVIRONMENT
Week 7 Streaming/Kafka	Project 2	Project 2	Project 2	Project 2	Project 2	<ul style="list-style-type: none"> Kafka
	Quality Control Audit	<ul style="list-style-type: none"> Apache Kafka architecture 	<ul style="list-style-type: none"> Create producer and consumer 	<ul style="list-style-type: none"> Introduction to Spark Streaming 	<ul style="list-style-type: none"> Processing data stream using Spark streaming 	
	Trainer interviews	<ul style="list-style-type: none"> Pub-Sub messaging 	<ul style="list-style-type: none"> Send messages from Producer 	<ul style="list-style-type: none"> Spark engine 		
	Written Evaluation	<ul style="list-style-type: none"> Creating a Kafka topic 	<ul style="list-style-type: none"> Producer API 	<ul style="list-style-type: none"> Sending data stream from Kafka to Spark 		
	<ul style="list-style-type: none"> Introduction to Streaming Introduction to Kafka Kafka Fundamentals Topics/Brokers/Consumer/Producer 	<ul style="list-style-type: none"> Retrieve list of topics 				
Week 8	Project 3	Project 3	Project 3	Project 3	Project 3	
	Written evaluation					
	Trainer interviews					
	Quality Control Audit					
Week 9	Project 3	Project 3	Project 3	Project 3	Project 3	
	QC Audit - Cumulative					
Week 10	Project 3	Project 3	Project 3	Project 3		
				Project showcase		

PROJECT

■ Project 3

■ Project 2

■ Project 1

TECHNOLOGIES

Spark, Spark SQL, Kafka, Spark Streaming

Spark RDD, Spark SQL

Hadoop, Hive



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.