Course name: Apache Spark – Unified Engine for large-scale data analytics

Week 1-2: Big Data and Distributed Systems Fundamentals (6 hours)

Week 1:

- Hour 1: Introduction to Big Data (definition, characteristics, challenges) - Read Chapter 1 of "Big Data: The Missing Manual" by Tim O'Reilly (book)
- Hour 2: Distributed Systems basics (scalability, fault tolerance, distributed computing) - Watch "Distributed Systems" video by Martin Kleppmann (YouTube)
- Hour 3: Overview of Hadoop and Spark ecosystems - Read Chapter 1 of "Hadoop: The Definitive Guide" by Tom White (book)

Week 2:

- Hour 4: Data processing in distributed systems (batch processing, stream processing) - Read Chapter 2 of "Streaming Systems" by Tyler Akidau, Slava Chernyak, and Reuven Lax (book)
- Hour 5: Data storage in distributed systems (HDFS, NoSQL databases) - Watch "HDFS and MapReduce" video by Cloudera (YouTube)
- Hour 6: Spark architecture and components (RDDs, DataFrames, Spark SQL) - Read Chapter 2 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 3-4: Python Programming in Spark (6 hours)

Week 3:

- Hour 7: Setting up Spark environment (installation, configuration) - Follow instructions on Apache Spark website (online resource)
- Hour 8: Introduction to PySpark (SparkContext, RDDs, DataFrames) - Read Chapter 3 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

- Hour 9: Data manipulation and analysis with PySpark (filtering, mapping, reducing) - Complete Lab 1 of "PySpark Tutorial" by DataCamp (online resource)

Week 4:

- Hour 10: Working with DataFrames and Spark SQL (data manipulation, querying) - Read Chapter 4 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 11: Data visualization with PySpark (Matplotlib, Seaborn) - Complete Lab 2 of "PySpark Tutorial" by DataCamp (online resource)
- Hour 12: Best practices for PySpark development (coding standards, debugging) - Read Chapter 5 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 5-8: Spark Machine Learning (12 hours)

Week 5:

- Hour 13: Introduction to Spark MLlib (machine learning library) - Read Chapter 6 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 14: Supervised learning with Spark MLlib (regression, classification) - Complete Lab 3 of "Spark MLlib Tutorial" by DataCamp (online resource)
- Hour 15: Model evaluation and selection with Spark MLlib (metrics, cross-validation) - Read Chapter 7 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 6:

- Hour 16: Unsupervised learning with Spark MLlib (clustering, dimensionality reduction) - Complete Lab 4 of "Spark MLlib Tutorial" by DataCamp (online resource)
- Hour 17: Advanced machine learning topics with Spark MLlib (ensemble methods, neural networks) - Read Chapter 8 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

- Hour 18: Using Spark MLlib for customer analysis (customer segmentation, churn prediction) - Complete Lab 5 of "Spark MLlib Tutorial" by DataCamp (online resource)

Week 7:

- Hour 19: Working with Spark MLlib and PySpark (integration, data preparation) - Read Chapter 9 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 20: Advanced topics in Spark MLlib (natural language processing, recommender systems) - Complete Lab 6 of "Spark MLlib Tutorial" by DataCamp (online resource)
- Hour 21: Best practices for Spark MLlib development (coding standards, debugging) - Read Chapter 10 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 8:

- Hour 22: Case studies of Spark MLlib applications (real-world examples) - Read Chapter 11 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 23: Future directions in Spark MLlib (new features, research areas) - Watch "Spark MLlib: Future Directions" video by Apache Spark (YouTube)
- Hour 24: Review and practice Spark MLlib concepts (practice problems, projects) - Complete Lab 7 of "Spark MLlib Tutorial" by DataCamp (online resource)

Week 9-15: Project Development and Advanced Topics (21 hours)

Week 9-12:

- Hours 25-36: Work on a project that applies Spark MLlib to customer analysis (data preparation, model development, evaluation)
- Use PySpark and Spark MLlib to develop a predictive model for customer churn or segmentation
- Use DataCamp labs and Apache Spark documentation as resources

Week 13:

- Hour 37: Advanced topics in Spark (GraphX, SparkR) - Read Chapter 12 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 38: Using Spark with other big data tools (Hadoop, NoSQL databases) - Watch "Spark and Hadoop" video by Cloudera (YouTube)
- Hour 39: Best practices for Spark development (coding standards, debugging) - Read Chapter 13 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 14:

- Hour 40: Spark performance optimization (caching, parallelism) - Read Chapter 14 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)
- Hour 41: Spark security and authentication (SSL, Kerberos) - Watch "Spark Security" video by Apache Spark (YouTube)
- Hour 42: Advanced Spark topics (Spark Streaming, Structured Streaming) - Read Chapter 15 of "Learning Spark" by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia (book)

Week 15:

- Hour 43: Review and practice Spark concepts (practice problems, projects) - Complete Lab 8 of "Spark Tutorial" by DataCamp (online resource)
- Hour 44: Prepare for Spark certification (study guide, practice exam) - Use Apache Spark documentation and DataCamp resources
- Hour 45: Final project presentation and review (present project, receive feedback) - Use Apache Spark documentation and DataCamp resources

Recommended Books and Resources

- "Learning Spark" by Holden Karau
- "PySpark Cookbook" by Tomasz Drabas
- "Big Data: The Missing Manual" by Tim O'Reilly
- "Hadoop: The Definitive Guide" by Tom White
- "Designing Data-Intensive Applications" by Martin Kleppmann
- edX's "Apache Spark Essentials" course
- edX's "PySpark Essentials" course
- edX's "Apache Spark Machine Learning" course
- edX's "PySpark Machine Learning" course