COURSE DETAILS & CURRICULUM

- **Module 1 - Introduction to Spark - Getting started**
    1. What is Spark and what is its purpose?
    2. Components of the Spark unified stack
    3. Resilient Distributed Dataset (RDD)
    4. Downloading and installing Spark standalone
    5. Scala and Python overview
    6. Launching and using Spark's Scala and Python shell ©

- **Module 2 - Resilient Distributed Dataset and DataFrames**
    1. Understand how to create parallelized collections and external datasets
    2. Work with Resilient Distributed Dataset (RDD) operations
    3. Utilize shared variables and key-value pairs

- **Module 3 - Spark application programming**
    1. Understand the purpose and usage of the SparkContext
    2. Initialize Spark with the various programming languages
    3. Describe and run some Spark examples
    4. Pass functions to Spark

5. Create and run a Spark standalone application
6. Submit applications to the cluster

- **Module 4 - Introduction to Spark libraries**
  1. Understand and use the various Spark libraries

- **Module 5 - Spark configuration, monitoring and tuning**
  1. Understand components of the Spark cluster
  2. Configure Spark to modify the Spark properties, environmental variables, or logging properties
  3. Monitor Spark using the web UIs, metrics, and external instrumentation
  4. Understand performance tuning considerations