The first module talks about the basics of Big Data.

Here you will learn the most up to

date definition of big data,

along with its use cases impacting

everyday personal tasks and business transactions.

You'll also learn how big data uses parallel processing,

scaling, and data parallelism,

and explore commonly used big data tools.

The second module will introduce you to

the Apache Hadoop architecture and ecosystem.

The Hadoop ecosystem comprises of

Hadoop distributed file system or HDFS,

MapReduce, Hive, and HBase.

You'll also get the opportunity to

gain practical experience in Hive,

Hadoop cluster, and MapReduce.

In the third module,

you'll focus on the popular

Apache Spark platform where you'll

gain key insights about

functional programming and Lambda functions.

You will also explore resilient distributed datasets,

or RDDs,

parallel programming and resilience in Apache Spark.

In this module, you'll also work

with queries and learn about the functions,

parts, and benefits of Spark SQL and data frame queries.

The next module, Module 4,

talks about data frames and Spark SQL,

where you will learn about RDDs,

their uses in Apache Spark,

and RDD transformations and actions.

You'll compare the use of datasets with

Spark's latest data abstraction data frames.

You'll also explore Apache Spark

SQL optimization and learn

how Spark SQL and

memory optimization benefit from

using catalyst and Tungsten.

Module 5 will take you through

development and run time environment options.

Here you'll explore how Spark

processes the requests that your application submits.

You'll also be able to identify Apache Cluster Managers,

their components and benefits,

and know how to connect with each cluster manager.

The module will also explain

using Sparks unified interface,

Spark submit, and options and dependencies.

You'll also learn about setting

Apache Spark configurations and perform

hands on labs to use

Apache Spark on IBM Cloud and Kubernetes.

In the sixth module,

you'll learn about connecting to

the Apache Spark User Interface,

or UI web server and using

the same UI web server to manage application processes.

You'll also learn to debug

Apache Spark application issues

using the application UI and locate related log files.

The module will also help you discover and gain

real world knowledge about how Spark

manages memory and processor resources.

Module 7, the final module,

is where you will cement your learning.

In this module, you'll perform a practice lab to apply

various transformations and actions

on RDDs and dataframes.

Further, you'll apply your knowledge

to perform a final lab,

where you will create a data frame and finally,

you'll be assessed based

on your learning from the course.

Learn by doing.

To solidify your course knowledge,

you will perform activities

and gain practical experience in

the hands on labs

strategically placed across the learning modules.

Get the most from your course.

Watch all videos and attempt all quizzes.

Complete the hands on labs to practice

your new skills and create your sharable portfolio.

Use the discussion forums to connect and collaborate

with your peers or seek assistance from the course staff.

Congratulations on beginning the next step

on this exciting journe