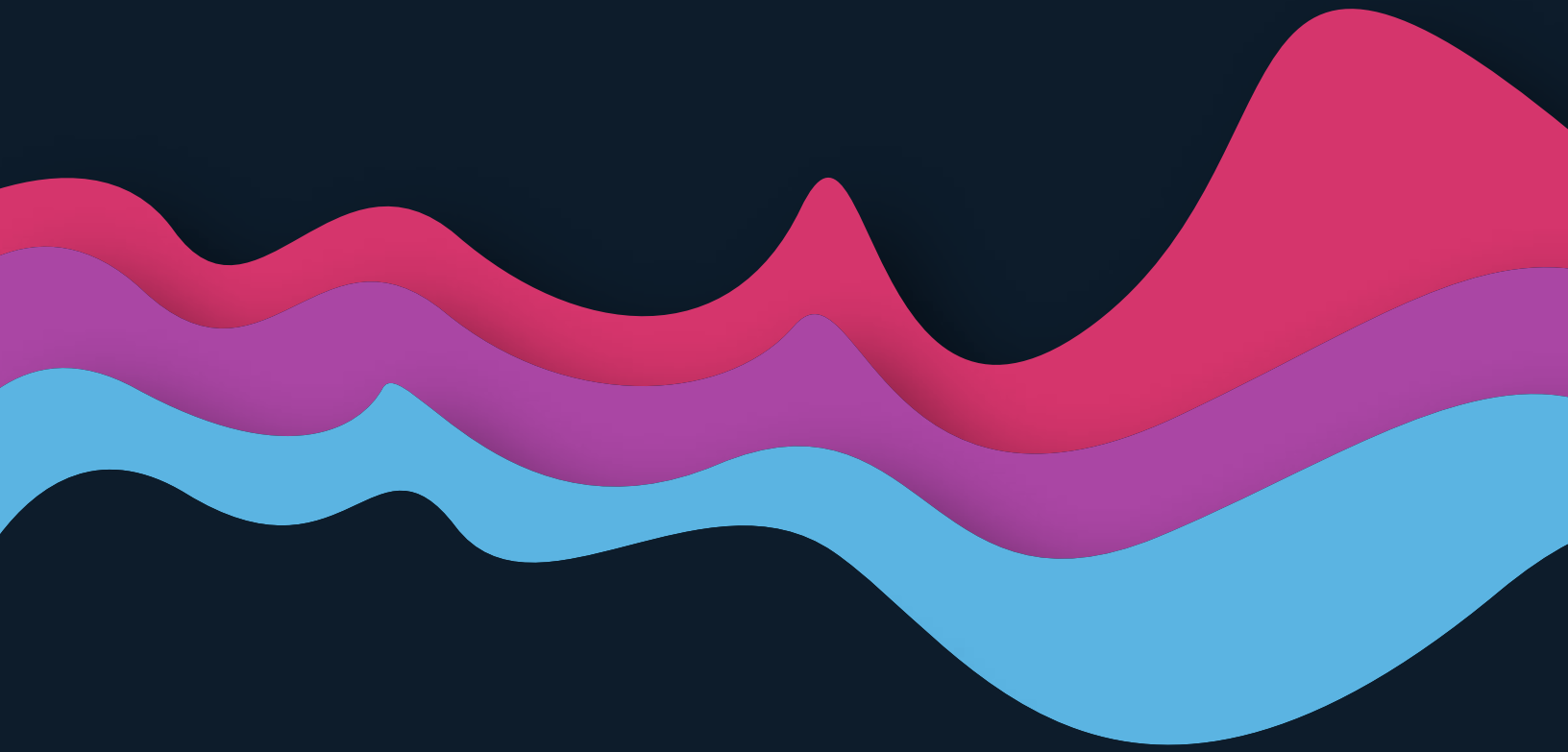
WeCloudData



BIG DATA FOR DATA SCIENTISTS

COURSE & SYLLABUS INFO

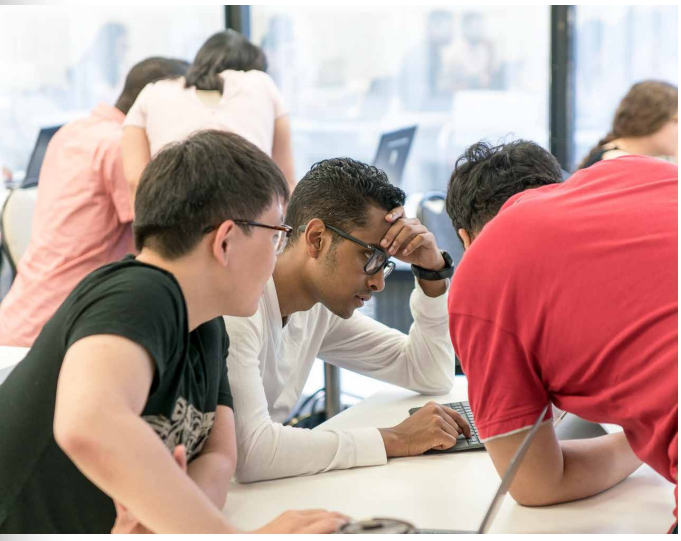
PROGRAM DETAILS

OVERVIEW

This course focuses on the Big Data & Spark Ecosystem tools that modern data scientists and analysts need to master.

Students learn throughout the course to work with components such as **Apache Hadoop/Hive**, **NosQL**, **AWS**, **Spark RDD**, **DataFrame**, **Spark ML**, **Spark Structured Streaming**, and **Kafka** to solve real-world advanced analytics problems.

Apache Spark is by far one of the most popular and successful big data processing frameworks that has been widely adopted by the industry and has become a must-have skill of modern data scientists. This course is designed to be hands on and teaches data scientists to scale their machine learning and advanced analytics to terabyte scale data (billions of records) in a cloud environment.



Students will learn how Spark can be used as a framework to train machine learning and deep learning models at scale and also learn how to deploy the models and build prediction services using Docker and Amazon Sagemaker .

Many big data use cases will be covered to help consolidate the learnings and most importantly students gain real-life experience and confidence to apply the knowledge learned back to their data science projects at work.

PREREQUISITES

You do not need prior experience with Scala programming since this course is taught mainly using PySpark.

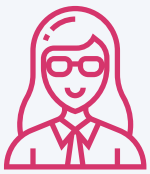
Students are required to be familiar with Linux Commands, SQL and relational database concepts. Some familiarity with distributed computing framework such as MapReduce is assumed.

WHO IS IT FOR?

New Grads & Job Seekers

who are looking for the first job

This course teaches you the spark skills required for modern data scientist jobs and then complementary spark interview questions will get your prepared for interview challenges.



Data Scientist

who are used to build ML models using R or Scikit-learn

This course will teach them how to scale their machine learning to a Spark Cluster.

Masters/ PhDs

who are well versed with the Hadoop ecosystem

This course will help you unleash the power of Spark and gain hands-on experience with machine learning pipelines and model deployment.

LEARNING OUTCOMES

After this course, the students will be able to:

- Gain competence to take on real big data challenges at workplace and demonstrate experience and advantage in the job market with the spark skills added to the resume
- Build machine learning models using Spark ML
- Handle real-time model scoring at scale using Spark Streaming
- Deploy Spark ML models in Amazon SageMaker and Kubernetes clusters
- Build and automate data pipelines with Apache Airflow and build a project demo via visualization dashboard with Superset
- Gain real world experience through a hands-on project and convince your manager/peers that you're up for big data related projects at work

Shaohua Zhang
Co-founder and CEO of WeCloudData

INSTRUCTORS

- Lead instructor for the Big Data course and the corporate training program
- Helped build and lead the data science team at BlackBerry (2010 – 2015)
- Helping Communitech incubator and Open Data Exchange mentor startups on data strategies
- Specialize in machine learning, big data, and cloud computing

PROGRAM HIGHLIGHTS

5 +3 WEEKS
to learn and build projects

50 HOURS
10 hours per week

WEDNESDAY
6:30PM - 9:30PM EST

SATURDAY
9:00AM - 5:00PM EST

INDUSTRY EXPERTS
to learn from the best

TA SUPPORT
through live calls and Slack

PROJECT-BASED
build an end-to-end project

CAREER SERVICES
to help you succeed

WeCloudData is intense...in a great way. Shaohua is experienced and leads an impressive team of instructors. The course material and the exercises are demanding, so you should be serious about a career in data science before you enrol. Unlike a lot of other data science bootcamps, WeCloudData also has advanced courses for people who are already working in the industry. Five stars.

- Mark -

Coming from Java and C# programming background, Big Data was a new concept for me. The syllabus is well put together, comprising of labs and theory. I got to work on a project and presented my findings to the class. I received good support during the entire coursework and felt encouraged to ask questions. All in all, I would recommend this course

- Hema Patil -

CURRICULUM

5 + 3 weeks

WEEK 1

Big Data

- Cloud: AWS EC2
- Cloud: AWS S3/Redshift

WEEK 2

Big Data

- Big Data: Introduction to Data Lake
- Big Data: Hadoop/Hive/EMR

WEEK 3

Big Data

- Big Data: Apache Spark Intro
- Big Data: Spark DataFrame

WEEK 4

Big Data

- Big Data: Distributed Machine Learning Algorithms
- Big Data: Spark ML Library

WEEK 5

Big Data

- ML Pipelines with Amazon SageMaker
- Real-time Prediction with Kinesis and Spark Streaming

STUDY BREAK

Capstone Project #1

- Agile Project Management (Jira/Confluence/Scrum)
- **Capstone #1 - Building End-to-end Big Data Pipeline**

STUDY BREAK

Capstone Project #1

- **Capstone #1 - Building End-to-end Big Data Pipeline**

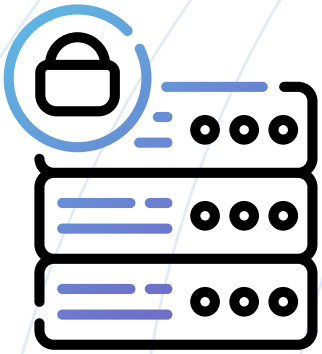
PRESENTATION

Capstone Project #1

- **Capstone #1 Building End-to-end Big Data Pipeline**

PROJECT DETAILS

This course is instructor-led and project-based. Students will be able to apply the big data knowledge acquired during the lectures build an end-to-end big data project.



Project: Large-scale data analysis with Spark ML

- Analyzing terabyte scale data using Apache Spark
- Machine Learning modeling at scale using Spark ML
- Real-time model prediction with Spark Streaming
- Real-time dashboard with Apache Superset
- Workflow automation with Apache Airflow Kinesis

Data Size: 500GB - 1TG
Records: 1 billion +

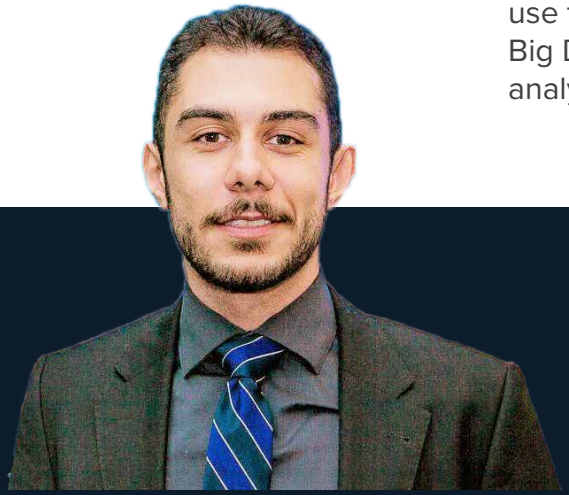
TUITION, GRANTS AND FINANCE OPTIONS

$2,500
CAD

DO YOU KNOW THAT MOST EMPLOYERS WILL REIMBURSE THE TRAINING COSTS?

- We have a detailed course syllabus and email template that you can use to convince your manager that this is the right course for you and a good investment for your company

- You will have a completed project and presentation that you can use to demo to your manager and showcase your newly minted Big Data and Spark skills and get ready for more interesting data analytics projects.



QUESTIONS?

Contact our program advisor, Amir, for more information **amir.asadian@weclouddata.com**