Big Data Bootcamp Curriculum

Adastra Academy and The Institute of Privacy and Big Data at Ryerson University bring students an eight-week intensive course on Big Data development. Through the span of five units, our experienced and passionate instructors teach students a cross-section of critical topics including Big Data architecture and processing frameworks, Hadoop tools and distributions, data mining and analysis, and data security. As a team, students develop an end-to-end application as their Capstone Project and feedback will be provided by our instructors. To conclude the Bootcamp, students take part in a careers workshop to prepare for the Big Data job market.

18 days of in-class instruction

50% lectures and 50% hands-on exercises

Student to instructor ratio is 7:1

20 students per cohort

Team Capstone Project with feedback from instructors

Careers workshop provided industry professionals

Learning Objectives

- Understand core Big Data concepts

- Understand key components of the Hadoop ecosystem

- Install and configure Hadoop environment

- Perform basic Hadoop administration tasks, such as starting and stopping various hadoop services, understand resources allocation and memory consumption, track and kill running Hadoop jobs

- Create, run, and test MapReduce streaming

- Install and use development environment to build and deploy Spark applications

- Acquire and store Data

- Process structured and semi-structured Data

- Understand core Data Management concepts: Data Quality, Reference Data Management, and Master Data Management

- Create a data profiling application

- Run and interpret a Data profile

- Learn the basic concepts of Data Science and Data Analytics lifecycle

- Learn the fundamental concepts of Machine Learning and Data Mining

- Learn about Data visualization and work with Data visualzataion tools

- Create Machine Learning applications

- Analyze and visualize Machine Learning application results

Unit 1 | Introduction to Big Data

Day 1: On the first day of the Bootcamp, students are introduced to Big Data and learn about Hadoop. Equipped with this knowledge, students install and configure Hadoop on their machines.

Day 2: Furthering their knowledge, students explore Hadoop administration, distributions and ecosystem, and administer their own cluster. After learning about the concept of MapReduce, students implement a MapReduce streaming exercise.

Day 3: Through an overview of Apache Spark, students learn the key components including resilient distributed datasets (RDDs), directed acyclic graph (DAG), Spark Core, they will the set up Apache Spark and NetBeans development environment.

Day 4: Students learn about Scala which is the Spark's programming languages.

Unit 2 | Data Acquisition & Storage

Day 5 & 6 & 7: Using an array of tools, students acquire data in different forms from different sources. Tools include Apache Sqoop, Apache Flume, Spark SQL, and Spark Streaming.

Day 8 & 9: With the acquired data, students explore methods of storing data in both SQL and NoSQL databases. Afterwards, students process the data using HiveQL and Spark SQL.

Unit 3 | Data Management Competencies

Day 10: Students are introduced to key data management concepts, including Data Quality, Reference Data Management, and Master Data Management.

Unit 4 | Data Analysis & Visualization

Day 11: Rounding out the Bootcamp, students are introduced to Data Science concepts, including Statistical Analysis, Visualization, and Machine Learning.

Day 12: Students learn about Data visualization further and perform exercises with Apache Zippelin.

Day 13 & 14: After setting up Spark MLlib and Apache Zeppelin, students create applications of supervised predictive models such as regression and classification.

Day 15: Continuing with Machine Learning, students deploy unsupervised learning methods, such as K-means, to cluster data. Then students analyze and visualize their results.

Unit 5 | Capstone Project

Day 16: Prior to the Capstone Project, students can participate in a Careers Workshop in preparation for the Big Data job market. Students receive insight to what employers are looking for, résumés tips, and interview questions.

Day 17 & 18: As a team, students collaborate on an end-to-end application leveraging their newly acquired knowledge and skills. By creating the application, students apply their learnings to current, real-world situations.

Our instructors review the project outputs and provide students with individual feedback on their contributions to the Capstone Project.

About Adastra Academy www.adastra.academy

Adastra Academy is a leading source of training and development for Information Management professionals and individuals interested in Data Management and Analytics technology. Our dedication to identifying and mastering emerging technologies guarantees our students are the first to gain access to critical skills. Our programs consist of hands-on labs and real-world examples allowing students to easily apply their new knowledge.

About Ryerson University www.ryerson.ca

Ryerson is Canada's leader in innovative, career-focused education and a university clearly on the move. It is a distinctly urban university with a focus on innovation and entrepreneurship. Ryerson has a mission to serve societal need and a long-standing commitment to engaging its community.

About The Privacy and Big Data Institute www.ryerson.ca/pbdi

The Privacy and Big Data Institute at Ryerson University is a multi-disciplinary hub for experts and students engaged in data-driven research, innovation and education. The Institute's mission is to become Canada's pre-eminent institution on privacy and data analytics issues, and an international leader in socially responsible policy development and technology innovation, through applied research, education, and community outreach.