Français

Data at Scale

Language of Delivery: English

Delivery Format(s): Facilitated Online Learning

✏️ ✉️ 🕐 ❓ 🖨️

A section of this course has been added to your shopping cart or you may have registered for this course previously. You may proceed to <u>checkout</u> or <u>continue shopping</u>.

Official Description

9.0 Continuing Education Units (CEUs)

Overview of various aspects of large data sets and how they are managed both on site and in the Cloud. Emphasis on hands-on experience from data ingestion to analysis of large data sets, both data-at-rest and data-in-motion (streaming data), including defining Big Data and its 5 V's: Volume, Velocity, Variety, Veracity, and Value.

Supplementary Information

24 hours of lectures and 66 hours of independent study; course includes synchronous and asynchronous activities.

Topics Covered

- Distributed Storage & Distributed Processing
- Analyzing Structured & Un-structured data at scale
- Row Oriented and Columnar Oriented Files Formats
- NOSQL Databases main categories
- Ingesting Data at Scale (High Velocity & Large Volume)
- Apache Hadoop and Apache Spark ecosystems (On-premise and in the Cloud)
- Scala language overview

Learning Outcomes

The course is designed to enable you to:

- Use Hadoop & Spark to store and process data at scale (using MapReduce and Spark)
- Use Hive, Impala and Spark SQL to analyze data at scale (using Hive Query Language, Scala, Python)
- Use Pig to analyze unstructured data at scale (using Pig Latin language)
- Improve querying data time (using Avro and Parquet files formats)
- Import and Export data at scale (using Sqoop)
- Install and configure ODBC/JDBC connectors to connect third party tools (MS Excel, Tableau Software, MicroStrategy, etc...)
- Build Nifi Data Flows to ingest, transform and route data at scale
- Implement Real-Time dashboard (using Nifi, HBase, Kafka, Banana Solr Dashboard)

Required Hardware Configuration:

To fully engage in the hands-on activities of this course, you will need a computer with the following specifications:

- **RAM**: At least 32 GB (a minimum of 24 GB is required to run the virtual machine)
- **Processor**: 6 cores or more (required to run the virtual machine)
- **Storage**: At least 100 GB of free disk space

Please note that the MacBook Pro M series is not supported (Oracle Virtualbox does not support yet).

Notes

This course is supported by [DataCamp](#), the most intuitive learning platform for data science. Learn R, Python and SQL the way you learn best through a combination of short expert videos and hands-on-the-keyboard exercises. Take over 100+ courses by expert instructors on topics such as importing data, data visualization or machine learning and learn faster through immediate and personalized feedback on every exercise.
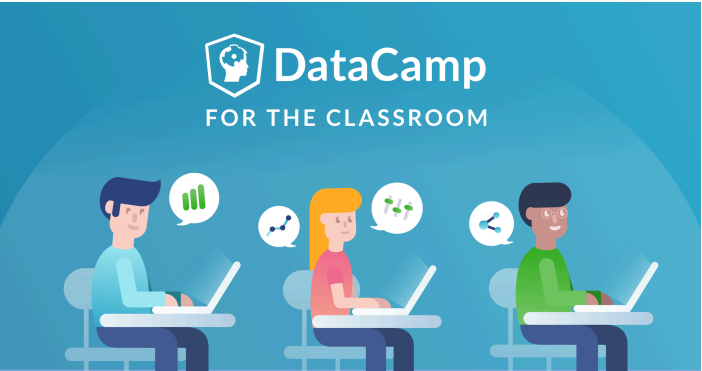
Prerequisite(s) and Corequisite(s)

- [Statistical Machine Learning (YCBS 255)](#)

Applies Towards the Following Programs

[Professional Development Certificate in Data Science and Machine Learning](#) : Required Courses

---

YCBS 257 - 23
Summer 2025
Expand to view schedule
Available
$1,482.64

Section Title:	Data at Scale
Language of Delivery:	English
Section Schedule:	May 08, 2025 to Jul 24, 2025
Type:	Online Course
Days:	Th
Time:	6:00PM to 9:00PM
Dates:	May 08, 2025 to Jul 17, 2025
Type:	Online Course
Days:	Th
Time:	6:00PM to 8:00PM
Dates:	Jul 24, 2025
Schedule and Location:	[View Details](#)
Contact Hours:	35.0
Delivery Format(s):	[Facilitated Online Learning](#)
Course Fee(s):	Tuition Fee non-credit $1,482.64
Potential Discount(s):	[McGill Employee Discount](#) [Senior Discount (65+)](#) [McGill Alumni Discount (Coupon Code)](#)

[McGill Alumni Discount](#)
[Dependents of McGill Employees](#)

Drop Request Deadline: Jan 28, 2025 to May 15, 2025

Transfer Request Deadline: Jan 28, 2025 to May 15, 2025

Withdrawal Request Deadline: May 15, 2025 to May 22, 2025

Section Notes:

A minimum number of registrations is required for this course section to be offered. The School reserves the right to cancel any course section when a minimum number of registrations has not been reached 7 days prior to the start date. In the event of a cancellation, the course fee will be refunded in full.

Course Format

This online course is a combination of weekly live online instructor-led sessions from 6:00 - 7:30 PM and self-directed learning activities and assignments.

Course Drop/Withdrawal Policy

- Any time prior to the 1st class: Course Drop Period with Full Refund.
- After the 1st and before the 2nd class: Course Withdrawal Period with Full Refund.
- After the 2nd class before the 3rd class: Course Withdrawal with No Refund.