

加拿大麦吉尔大学

大数据入门项目

McGill University

Introduction to Big Data Program

一、项目综述

随着云时代的来临，大数据正在吸引越来越多人的关注。大数据技术的战略意义不在于掌握庞大的数据信息，而在于如何对这些数据进行专业化处理与解读，从而真正发挥数据的重要作用。大数据入门项目是加拿大顶级学府麦吉尔大学开设的一门专业课程，旨在带领学生理解如何利用定量分析方面的优势，从大型数据集中推断出有意义的见解，以支持业务决策。同时，课程还将向学生介绍大规模处理数据的实用工具。

二、项目优势特色

- **【前沿的课程主题】**在“数据为王”的年代，由**计算机专业世界排名前 50**的麦吉尔大学设计的课程带你解读数据的奥妙；
- **【饱满实用的课程设置】**60 小时课程足质足量，有效强化学生对大数据的理解以及实用工具的运用；
- **【可接受四/六级成绩】**无需托福雅思成绩，四级 500/六级 470 即可申请；
- 【官方品质保障】**学生可获得麦吉尔正式的课程成绩单与参课证明，享受学校图书馆等资源，并可深度体验蒙特利尔地区的社会与文化。

三、项目详情

【项目日期】

2024 年 1 月 22 日 - 2 月 16 日（4 周，预估日期，待确认）

【报名要求】

- 1) 托福 79，或雅思 6.0，或大学英语四级 500、或大学六级 470；或 Duolingo 105
- 2) 大数据入门方向：学生需具备理工科或商科中的管理信息系统专业背景，且具备较强的数据与统计学知识，以及 Linux、Java 与 SQL 方面的技能；如不具备相关知识，学

生需在课程开始前参加以下在线课程学习：

Java Tutorial for Complete Beginners: <https://www.udemy.com/course/java-tutorial/>

Linux Linux/Unix Tutorial for Beginners: <https://www.guru99.com/unix-linux-tutorial.html>

SQL Intro to SQL for Data Science: <https://www.datacamp.com/courses/introduction-to-sql>

【课程内容】

项目为期四周，包含两门专业课，每门课 30 小时授课时间，共 60 小时。项目每周安排 15 小时授课，授课时间预计为当地时间每周一至周五上午 9-12 点或下午 1-4 点（以校方实际安排的课表为准）。

本项目共包含两门课程：

课程一：商业决策数据科学（30 小时）

课程将概述数据科学如何帮助推动业务决策和创建新的业务模型。课程重点关注如何将数据转化为业务见解，并且探讨数据科学流程以及数据驱动型企业所面临的各种挑战，包括伦理道德、数据治理和个人隐私等。课程还将通过银行、医疗保健、市场营销、农业等不同行业的实例分析，来探讨数据技术和存储的发展，以及数据科学工具和技术在不同业务领域的应用，如客户和 Web 分析、运营分析、人力资源相关分析等。

课程二：大规模数据（30 小时）

课程将帮助学生熟悉大型数据集的不同方面，以及如何在实地和云端管理它们。课程为参与者提供从数据摄取到大型数据集分析的亲身体验，包括静态数据或动态数据（流数据），亦包括定义大数据及其 5 V 特质：体积、速度、多样性、准确性和价值。同时，课程还将介绍分布式数据库和存储的体系结构、Hadoop 和 Spark 等生态系统，以及 Scala、Spark Shell 和 Pyspack 等工具。

【师资介绍】（往期师资，仅供参考，以实际安排为准）

Nabil Beitinjaneh

纳比尔·贝廷詹毕业于麦吉尔大学电子与计算机工程系（计算机方向），同时拥有 MBA 学位。目前他担任麦吉尔大学继教学院人工智能、大数据、以及机器学习项目的学术统筹，同时也在相关学院担任讲师。他拥有超过 25 年的商业和战略开发、项目管理、系统和流程工程经验；在商业情报、大数据和数据分析领域经验丰富；同时，他长期为非营利组织和非政府组织提供领导力、战略和组织发展方面的培训与服务。

【参考日程】

第一周

日期	日程安排
2024/1/21 (周末)	抵达蒙特利尔，入住校外公寓
2024/1/22 (星期一)	上午：项目启动；课程介绍；大数据简介 下午：参观校园，熟悉环境
2024/1/23 (星期二)	上午：数据科学和分析介绍 下午：参加学校的体育活动
2024/1/24 (星期三)	上午：大规模存储和处理数据 下午：游览蒙特利尔老城区
2024/1/25 (星期四)	上午：使用 Excel 进行数据探索（一） 下午：开展独立研究，完成作业任务
2024/1/26 (星期五)	上午：大规模分析结构化数据 下午：参加学校的兴趣社团活动
2024/1/27-1/28 (周末)	参加学校组织的活动，如渥太华一日游（需另付费）

第二周

日期	日程安排
2024/1/29 (星期一)	上午：使用 Excel 进行数据探索（二） 下午：参观蒙城唐人街
2024/1/30 (星期二)	上午：大规模吸收结构化和非结构化数据-实时摄取数据 下午：参加学校的体育活动
2024/1/31 (星期三)	上午：商业模型画布 BMC 下午：体验蒙特利尔地下城
2024/2/1 (星期四)	上午：学习定义将数据接收到 HDFS 的源/通道/接收器 下午：开展独立研究，完成作业任务
2024/2/2 (星期五)	上午：分析技术和机器学习概述-简介 Excel 中的练习会话 下午：参加学校的兴趣社团活动
2024/2/3-2/4 (周末)	自由安排

第三周

日期	日程安排
2024/2/5 (星期一)	上午：学习使用通道复制和通道多路复用 下午：参加学校的体育活动
2024/2/6 (星期二)	上午：Excel 中的机器学习和数据建模（一） 下午：游览皇家山公园

2024/2/7 (星期三)	上午：星火生态系统简介 下午：体验蒙城当地的特色市场
2024/2/8 (星期四)	上午：Excel 中的机器学习和数据建模 (二) 下午：开展独立研究，完成作业任务
2024/2/9 (星期五)	上午：Spark RDDs 和数据帧简介 下午：参加学校的兴趣社团活动
2024/2/10-2/11 (周末)	参加学校组织的活动，如魁北克一日游 (需另付费)

第四周

日期	日程安排
2024/2/12 (星期一)	上午：智能和商业决策 下午：参加学校的体育活动
2024/2/13 (星期二)	上午：概述 PySpark 和 Scala 语言 下午：参观蒙城当地博物馆
2024/2/14 (星期三)	上午：可视化技术介绍和讲述故事 Excel 中的可视化实践 下午：参加学校的兴趣社团活动
2024/2/15 (星期四)	上午：研讨会:Twitter 数据分析 下午：开展独立研究，完成作业任务
2024/2/16 (星期五)	上午：批判性思维和数据/统计分析问题 下午：结业评估
2024/2/17 (星期六)	项目结束，启程回国

(注：以上行程安排仅为参考，实际行程安排以最终校方出具的行程为准)

【文化活动】

麦吉尔大学会定期向暑期项目学生发送每周各类活动信息，学生可适时关注，并根据兴趣选择参加校方每周组织的丰富多彩各类课外活动，探索体验当地的社会与文化，比如游览麦吉尔大学校园、蒙特利尔老城区、唐人街、蒙特利尔地下城、皇家山公园、当地博物馆、公共市场等特色景点。此外，学生也有机会参加渥太华、魁北克等其它加拿大著名城市的一日游活动 (需另付费)。