DATA ENGINEERING and ANALYTICS with CO-OP

PROGRAM OUTLINE

Full-Time 90 Weeks Program
900 Hours of Classroom Instruction
900 Hours of Co-op
1800 Total Hours in Duration

PROGRAM DESCRIPTION

In this instructor led multidisciplinary program, students will receive knowledge and expertise to fulfill a position of a junior/intermediate level Information Systems Data Analyst in modern Software Development organization.

CAREER OCCUPATION PROGRAM

NOCs: 2173

This program prepares students for the following career occupations:

Data Administrator, Database Administrator, Database Analyst, Electronic Data Processing Systems Analyst, Database Architect, Data Miner, Data Mining Analyst, Data Processing Specialist, Database Designer, Database Management Supervisor, Information Resource Analyst, Technical Architect

ADMISSION REQUIREMENTS

Admission requirements may not be waived by either the student nor the Canadian College of Technology and Business (CCTB).

No funding may be disbursed to the student or received by CCTB until all admission requirements are met.
A payment plan can be offered to a student if student loan is not an available option.

Requirements:

1. Good command of English language IELTS 5.5 or equivalent outlined below
2. High school diploma or equivalent from an approved government institution of applicant's home country, or applicant is minimum 19 years of age.
3. Students are required to have and use their own personal computer in class.

Instruction at CCTB is conducted in English.
Students must be in possession of one of the English Language Equivalencies in the list below.
Language proficiency requirements are admission requirements and may not be waived by either the institution or the student.

The following are accepted:

- English Grade 12 (from a high school program in Canadian, US, UK, or other English-speaking country)
- Communications Grade 12 (from a high school program in Canadian, US, UK, or other English-speaking country)
- Successful completion of CCTB EAP (English for Academic Purposes) Level 4
- IB (International Baccalaureate) English A1/A2 (HL or SL)
- IELTS: (International English Language Testing System) Band 5.5 or higher
- TOEFL: (Test of English as a Foreign Language) IBT 46-59 or higher

- CAEL: (Canadian Academic English Language Assessment) Score of 50 or higher
- PTE: (Pearson Test of English) Score of 45.4 or higher
- CLB: (Canadian Language Benchmarks) Level 6 or higher
- TOEIC: (Test of English for International Communication) Score 605-690 or higher
- Successful completion of a Designated ESL school program with level equivalent to tests   outlined in section 1 above OR from an established pathway partner of CCTB.
- Applicants may choose to complete the TLGC Online Pre-Arrival Test and receive
- customized language study plan based on the applicant's test results if a passing grade is not achieved.
- Applicants that are not in possession of one of the credentials above, are advised to enroll in The Language Gallery Canada (TLGC) UPP (University Pathway Program). Successful completion of TLGC UPP level 4.5 is recognized by CCTB in lieu of the aforementioned assessments.

LEARNING OBJECTIVES

After completing this program, students will be able to:

- Identify all the organizational business and their respective roles and responsibilities
- Understand how each specific business unit contributes to the successful completion of the overall project as well as interoperates with the other units within the organization
- Understanding of different technologies and toolsets used by different roles within the organization
- Plan projects utilizing different SDLC models including Waterfall and Agile approach frameworks
- Analysis, interpretation, and creation of business requirements
- Translation of business requirements into strategic project objectives
- Be able to track issues and create defect reports in Application Lifecycle Management System
- Comprehend and apply the communication and defect lifecycle
- Describe frontend and backend development related concepts, tools and technologies including environment architecture, protocols, languages like Java and XML to the extent needed by Data Analyst
- Explain the underlying system architecture of Linux/Unix based systems
- Demonstrate proficiency in administering systems on a command line interface
- Utilize administrative commands on an advanced level
- Utilize package management, system services, text stream parsers, pseudo filesystem and system information commands
- Create, configure, and mount different filesystems and devices
- Demonstrate the ability perform networking configuration and administration on the command line interface

- Implement and configure Linux/Unix based graphical systems
- Implement, configure, and administer common Linux/Unix system services including SQL servers, printing systems and Web Servers
- Secure and audit systems
- Create and run bash shell scripts and system jobs for task management
- Understand the concept of data normalization and functional dependencies
- Understand database design methodologies and entity relationship modelling
- Use UML for diagramming notation
- Utilize SQL CRUD operations
- Demonstrate the knowledge and use of Data Definition Language (DDL) and Data Manipulation Language (DML)
- Understand database security, transaction management, triggers and stored procedures, replication, data warehousing and OLAP on a basic level
- Reliably demonstrate the ability to operate the fundamental data structures in computer science
- Demonstrate their knowledge and implementation of Object-Oriented Concepts
- Understanding intermediate programming concepts such as arrays, linked lists and recursion
- Understanding of advanced programming concepts such as trees, maps, hash tables, text processing and graphs
- Reliably demonstrate the ability to analyse business requirements how the new features and functions should be incorporated into the system
- Understand the fundamentals of Agile Methodologies and specific implementations such as the SCRUM framework
- Students will also learn the basics of the project management and Object-Oriented Analysis and Design using Unified Modelling Language, UML, as well as graphic representation tool, BPMN
- Reliably demonstrate the ability to operate different distributions of discrete and continuous random variables
- To compute a mean, variance, and standard deviation for various distributions
- Will be able to make inference and conclusions based on evidence – reasoning from collected data
- Students will be familiar with confidence intervals for the population mean and for the proportion based on a sample mean
- Understand how to make inferences from hypothesis testing
- Reliably demonstrate the ability to model and analyse the relationship between variables in a case of two quantitative variables: single response variable and a single exploratory variable
- Understand the basic principles of regression analysis
- Establish statistical inferences from sample population based on a simple lineal regression model
- Able to apply Microsoft Excel for regression analysis
- Understand the concepts of artificial intelligence, machine learning, deep learning and machine learning applications
- Differentiate between supervised, unsupervised and reinforcement learning

- Demonstrate an understanding of various models and algorithms for data manipulation
- Display practical application for supervised learning algorithms such as classification and regression trees, k-neighbours algorithm, naïve Bayes algorithm, support vector machines (SVM), random forest and classification and regression ensembles for supervised learning
- Demonstrate the ability to estimate important parameters of the multiple regression model and interpret inferences about these parameters.
- Predict the value of response (dependent) variable knowing the values of exploratory (independent) variables based on the statistical model for multiple linear regression
- Use to non-parametric statistical analysis (Wilcoxon test) as well as other predictive models.
- Use the R programming language that was created for statistical analysis. R is widely used for data mining, regression analysis, statistical tests, time-series analysis, and other data analysis techniques.
- Practically apply unsupervised learning algorithms such as k-mean clustering and hierarchical clustering algorithms, Gaussian mixture, and hidden Markov models
- Reliably demonstrate the ability to Extract, Transform and Load (ETL) principles and data warehousing life cycle
- Understand how to work with various dimensional models of data warehousing such as the star, the snowflake, and the fact constellation schemas as well as Online Analytical Processing (OLAP) servers
- Demonstrate the ability to use and apply a powerful software data visualization tool: Tableau
- Understand the fundamentals of data visualization
- Customize data view and create dashboards; extract, join and blend data; add, save, and delete Tableau worksheets and make calculations with Tableau
- Understand how to sort and filter the data and how to represent data graphically using bar charts, pie charts, scatter plots, box plots, tree maps, histograms, and others
- Demonstrate the ability to work with SQL Server services such as SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS), and SQL Server Reporting Services (SSRS)
- Understand how to create and maintain an Analysis Services databases, how to extract data from multiple data sources, and how to create multidimensional databases
- Create a project on data migration which implies transferring data from one system to another while the storage, database or application has changed
- Understand Data Mining fundamentals
- Perform data mining tasks and issues, and different algorithms and techniques
- Employ cloud computing technologies, architecture, and infrastructure as well as cloud computing models, services, management, data storage, and different applications

- Understand NoSQL databases and their implementation as well as their differences from traditional relational databases
- Understand the internet security protocol, secure sockets layer, email, web sites, and firewall security issues
- Recognize and prevent data breaches and protect personal data
- Security and privacy challenges concerns with Big Data regarding cloud storage

PROGRAM EVALUATION METHODS AND COMPLETION REQUIREMENTS

CCTB evaluates students using a variety of methods including projects, assignments, presentations, assessments, quizzes, and exams. Students will be given a performance evaluation before 30% of the hours of instruction of the program are completed. This evaluation will address any academic concerns that the college may have regarding student performance and/or learning outcomes. This evaluation will also ensure the student comprehensively understands the grading system, and what actions they can take moving forward to achieve or maintain a higher grade.

To complete the program, students will be required to achieve a minimum grade of 65% in each course, as well as complete the co-op/ practicum component of their program.

The co-op/practicum component of the program includes a performance-based evaluation conducted by the placement host and an analysis report created by the student relating to their work experience that must submitted to the faculty.

Additionally, to successfully complete the program, students must maintain a minimum attendance rate of 75%.

If a course is failed, the student must re-take the course within the next available cohort. The course re-take fee is $1100.

Please reference the CCTB Dispute Resolution and Grade Appeal Policy

HOMEWORK HOURS

A minimum of 2.5 - 3 hours of homework between lectures is to be expected.

DELIVERY METHODS

- In-Class
- Combined delivery (both in-class and distance)

REQUIRED PROGRAM MATERIALS

Resources in the form of custom learning materials will be provided by CCTB.

Software tools and user licenses will be provided by CCTB.

Instructors will provide students with additional educational resources that will be specific to the subject matter of each course. These resources will be used in conjunction with the class lectures.

These resources and learning materials will be made available online via the CCTB Canvas learning management system. Students are required to login to gain access to the e-materials.

Students must have and use their own personal computer in class.

Additional Recommended Learning Materials (not required):

- Nemeth, Snyder, Hein, Whaley (2010). Unix and Linux System Administration Handbook, 4th Edition. Prentice Hall.
- Goodrich M.T., Tamassia R. Data Structures & Algorithms in Java, 5th Ed. Wiley (2010)
- Albright S.Ch., Winston W.L.(2017). Business analytics: Data analysis and decision making. Sixth edition. Boston, MA. ISBN:978-1-305-94754-2.
- De Veaux, R., Velleman, P., and Bock, D. (2014). Intro Stats with CD. 4th ed. New York: Pearson/Addison-Wesley. ISBN: 978-0-321-82527-8 / 0-321-82527-6
- Moore, D.S., McCabe, G.P. (2006). Introduction to the Practice of Statistics, 5th ed., New York: W.H. Freeman and Company. ISBN-13:978-0-7167-6400-7, ISBN-10:0-7167-6400-8
- Judd, Ch.M., McClelland G.H. (2017). Data analysis: A model comparison approach to regression, ANOVA, and beyond. Third edition. New York: Routledge. ISBN-13: 978-1138819832.
- Moore, D.S., McCabe, G.P. (2006). Introduction to the Practice of Statistics, 5th ed., New York: W.H. Freeman and Company. ISBN-13: 978-0-7167-6400-7, ISBN-10: 0-7167-6400-8
- Everitt B., Hothorn T. (2011). An Introduction to applied multivariate analysis with R. Springer. ISBN-13: 978-1441996497.
- Bisgaard S., Kulahci M. (2011). Time Series Analysis and Forecasting by Example. ISBN: 9780470540640, Online ISBN: 9781118056943, DOI: 10.02/9781118056943, Wiley Series in Probability and Statistics.
- Murphy, K.P. (2012). Machine Learning: a probabilistic perspective. Massachusetts Institute of Technology. ISBN: 978-0-262-01802-9.
- Kimball, R., Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling. Third edition. John Wiley & Sons. Inc. ISBN: 978-1118530801.
- Ponniah, P. (2010). Data warehousing: Fundamentals for IT professionals. Second edition. John Wiley & Sons. Inc. ISBN:978-0470462072.
- Murray, D. Tableau your data!: Fast and easy visual analysis with Tableau software. John Wiley & Sons. ISBN: 9781118612040/ 1118612043.

- Aspin A. (2015). Business Intelligence with SQL Server Reporting Services. Apress. ISBN-13: 978-1484205334.
- LeBlanc P., Moss J.M., Sarka D., Ryan D. (2015). Applied Microsoft Business Intelligence. Indianapolis IN: John Wiley & Sons. ISBN: 978-1-118-96177-3.
- Williams S. (2016). Business Intelligence Strategy and Big Data Analytics. Cambridge, MA: Elsevier Inc. ISBN: 978-0-12-809198-2.
- White T. (2015). Hadoop: The definitive guide: storage and analysis at internet scale. Sebastopol, CA: O'Reilly Media, Inc. ISBN-13: 978-1491901632.
- Ledolter J. (2013). Data Mining and Business Analytics with R. Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN-13: 978-1118447147.
- Ruparelia N.B. (2016). Cloud Computing. Massachusetts Institute of Technology. ISBN: 9780262529099.
- Harrison G. (2015). Next generation Databases: NoSQL and Big Data. Apress. ISBN-13: 978-1484213308.
- Hayden, L. (2010). IT Security Metrics: A Practical Framework for Measuring Security & Protecting Data. McGraw-Hill Companies. ISBN-13: 978-0071713405.

PROGRAM ORGANIZATION

1.	Information Systems Management	60 HRS
2.	Introduction to Linux Systems Administration	60 HRS
3.	Relational Database Systems and SQL	60 HRS
4.	Introduction to Data Structures, Design and Analysis	60 HRS
5.	Business Analysis and System Design	60 HRS
6.	Probability and Statistics Fundamentals	60 HRS
7.	Statistics for Data Analysis	60 HRS
8.	Applied Multivariate Statistical Analysis	60 HRS
9.	Machine Learning	60 HRS
10.	Data Warehouse (EDW) Concepts	60 HRS
11.	Visual Analysis for Business Intelligence (BI) with Tableau	60 HRS
12.	Business Intelligence (BI) with SSIS, SSAS, and SSRS	60 HRS
13.	Advanced Topics in Data Analysis, Big Data, Data Mining, Cloud Computing	60 HRS
14.	Security and Privacy Issues in Data Analytics	60 HRS
15.	Employment Preparation Training	60 HRS
	CO-OP Placement	900 HRS
	Total Duration	**1800 HRS**