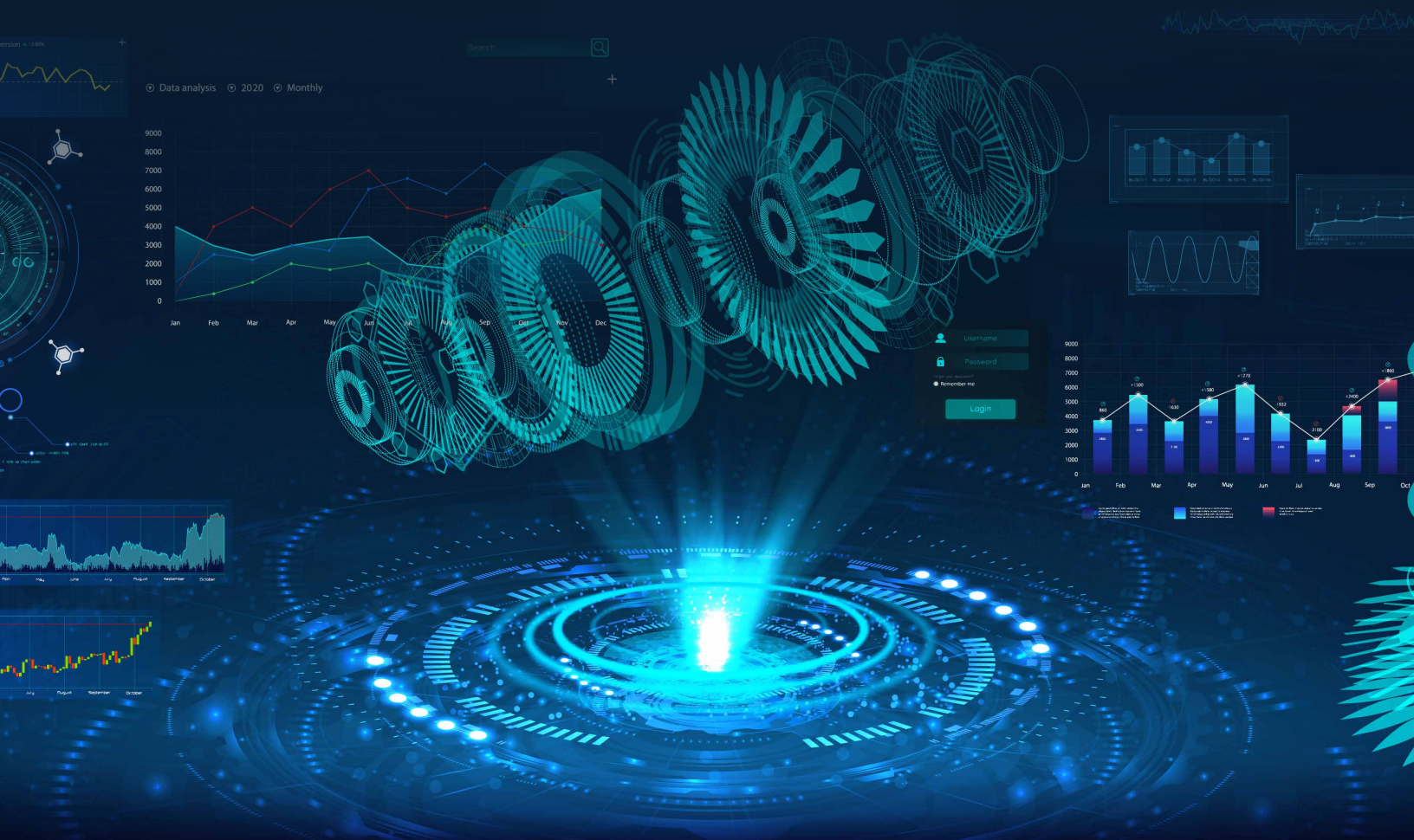
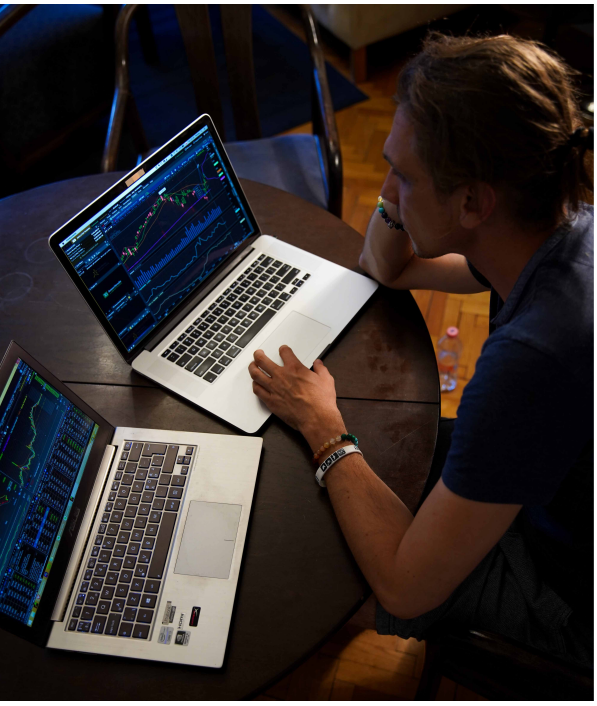
WeCloudData

Data Engineering Certificate Program

COURSE INFO & SYLLABUS



CAREER FACTS



WHAT DO DATA ENGINEERS DO?

Data Engineers play an essential role in organizations that collect and manage big data.

In a typical data organization, data engineers gather and collect the data, store it, do batch processing or real-time processing on it, and serve it via the database to a data scientist who can easily query it. In essence, data engineers help manage how data flows in an enterprise's data architecture.

JOB MARKET

The market has seen a surge in demand for data scientists in the past several years and we see almost all universities and colleges offer some kind of data science courses and programs.

However, data engineers are usually harder to train and source because the program needs to be very practical/hands-on and there is not much theory to teach.

JOB TITTLE

ETL/ELT Engineer

ETL Developer
Data Integration Engineer
SQL Developer
Big Data Engineer

Data-Savvy Software Engineer

Computer Science/Engineering
Mechanical Engineering
Electrical Engineering
Data Platform Engineer

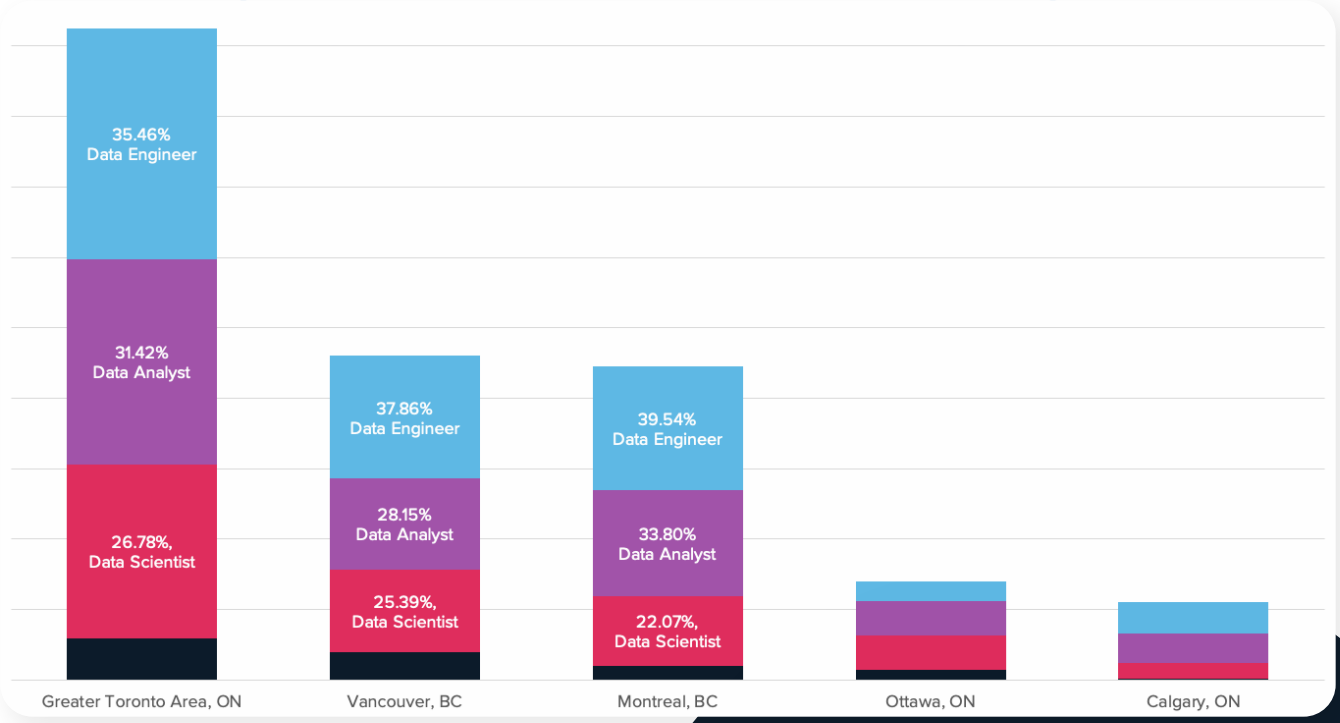
BI/Data Warehouse

Data Architect
BI Engineer
Big Data Developer
Data Warehouse Engineer
BI Specialist

ML Data Engineer

Data Engineer
Big Data Engineer
AI/ML Engineer
Model Automation Engineer

DEMAND FOR DATA ENGINEERING



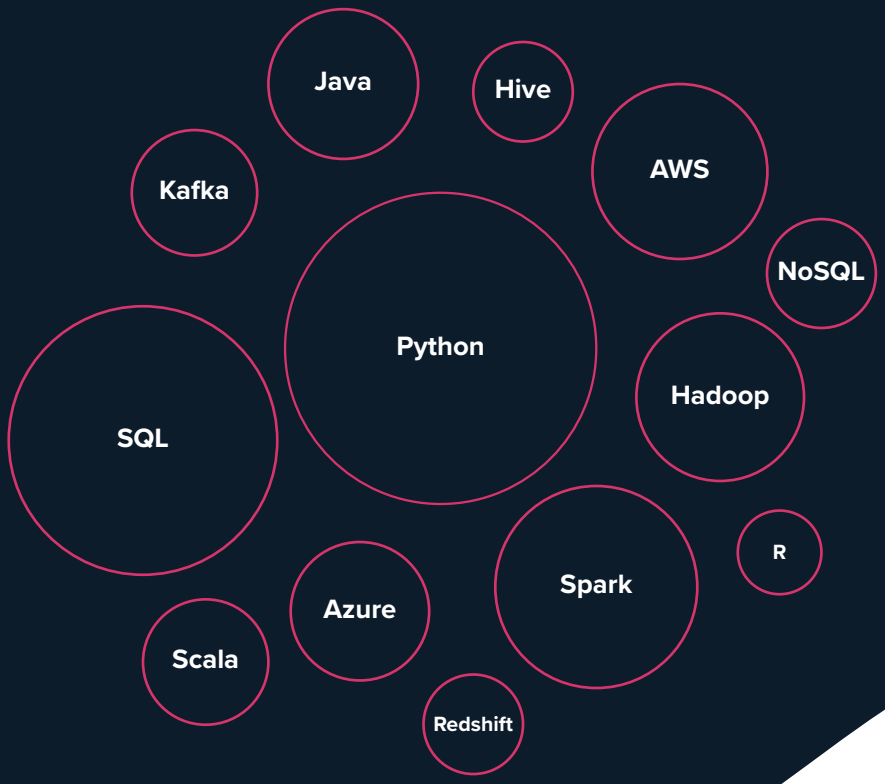
DATA ENGINEER SALARIES IN CANADA

Average in Canada

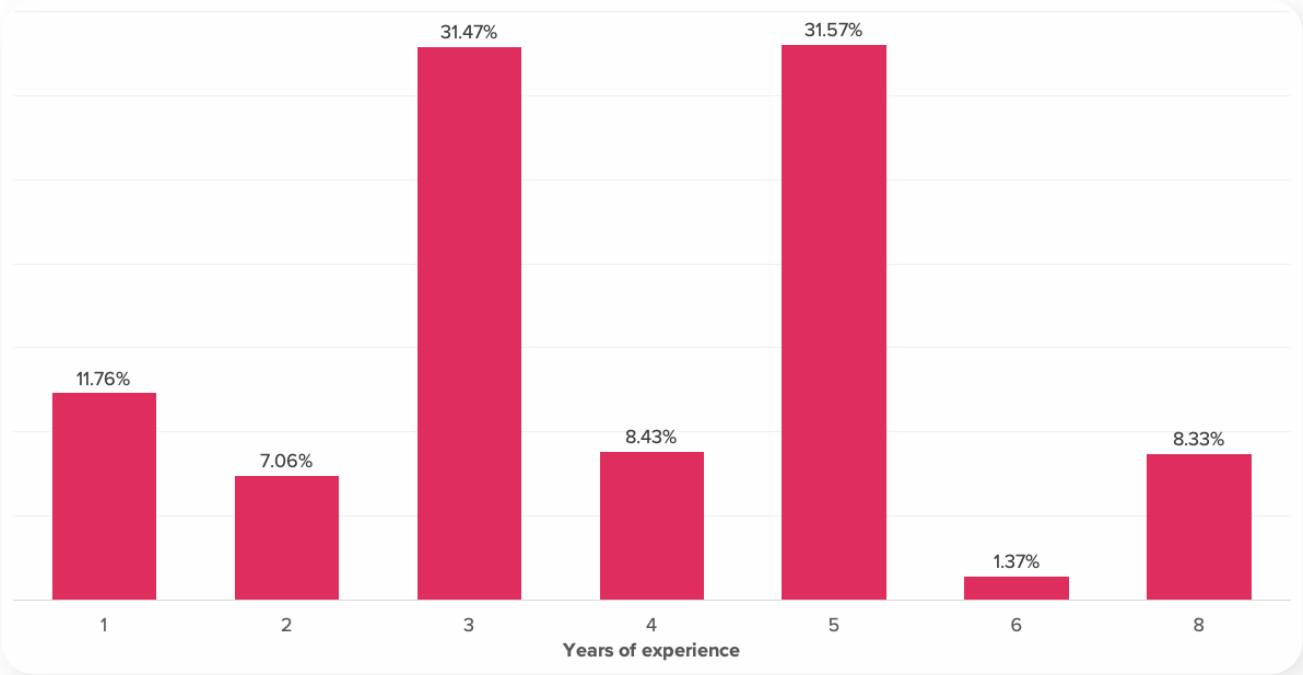
$83,662 per year

Salary Distribution

TECHNICAL SKILLS



REQUIRED YEARS OF EXPERIENCE

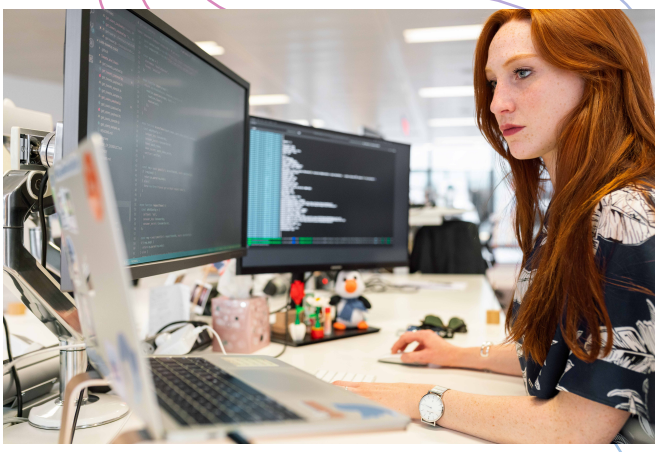


ABOUT THE PROGRAM

OVERVIEW

WeCloudData's Data Engineering certificate program focuses on helping students to acquire the essential data engineering skills, gain professional experience and prepare for data engineer careers.

WeCloudData is a Toronto-based data skill training academy. We've partnered with many Canadian corporations on upskilling their employees. No. 1 best data science bootcamp in 2020-2021 by SwitchUp.



WHO IS IT FOR?

This weekend program is suitable for students who want to become a Machine Learning Engineer or AI Engineer

Students accepted into the program should have acquired fundamental knowledge and skills of Python programming and machine learning already

Recent Grads

from CS and Engineering who want to gain advantage in the job market by gaining practical ML engineering skills

Data Scientists

who want to build more advanced skills in MLOps and model deployment

Entrepreneurs

who want to build ML systems and platforms for her/his next startup

Career Switchers

who have a full-time job and would like to switch career

PROGRAM DETAILS

24 WEEKS
Wednesday evening &
Saturday full day

TA OFFICE HOUR
Mon-Thu 6PM-8PM

10+ GUEST SPEAKERS
to make connection with

4 : 1 RATIO
students vs instructors

END-TO-END PROJECTS
to build your portfolio

1 BUSINESS PROJECT
from a real client
(full-time students only)

LEARNING OUTCOMES

- Solid understanding of major big data and cloud platforms such as Hadoop, Spark, Databricks, AWS, and GCP
- Able to build and deploy data pipelines on cloud platforms using workflow orchestration tools such as Apache Airflow and DBT
- Comfortable with building real-time and batch data ingestion
- Deep understanding of the pros and cons of different database systems
- Hands-on experience with data warehouse modeling and ETL
- Hands-on experience with DataOps tools such as Git, Docker, Kubernetes
- Familiar with all things data engineering for data scientists: maintaining pipelines, building feature stores, scheduling model training, monitoring, and deployment
- Solid understanding of data governance and managing data engineering projects

PREREQUISITES

- Know the basics of Python and SQL (We will also provide online Python and SQL courses for you)
- An operating system with at least 8G of RAM (16G is preferred), 4 cores of CPU

GUEST SPEAKERS

TOP-NOTCH CURRICULUM

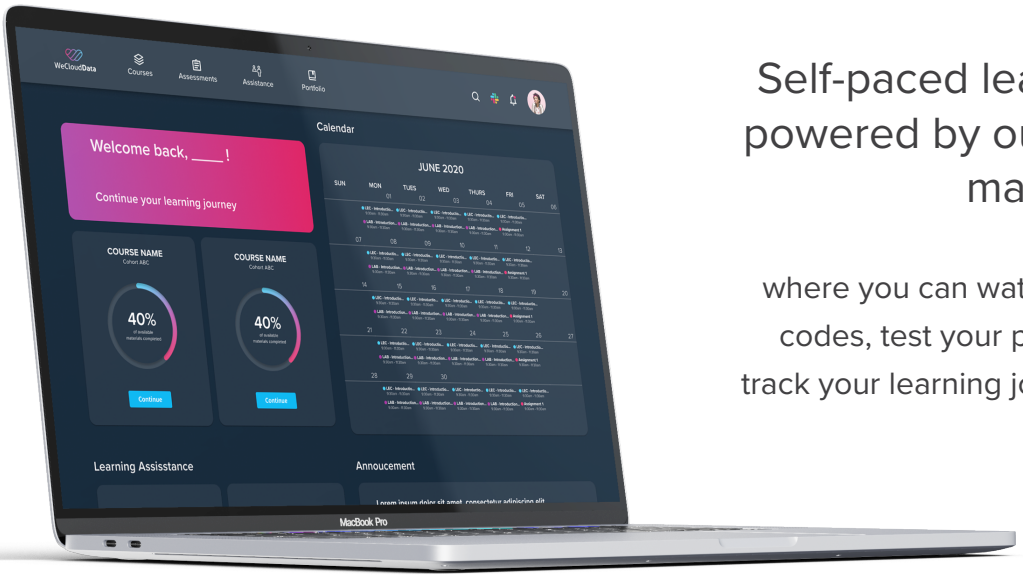
The only academy that provides the most comprehensive data engineering curriculum

Topics	Content	Topics	Content
Programming	Linux SQL Python/Scala	**Data Pipelines**	Apache Airflow CI/CD Basics
Cloud Computing	Amazon Web Service Google Cloud Platform	**Docker Kubernetes**	Containerization Container Orchestration
Big Data	Hadoop, MapReduce, Hive Apache Spark NoSQL Database Presto/Druid	**Data Governance**	Data Quality Data Cataloguing Metadata Management Data Privacy
Data Warehouse	Snowflake, Redshift, BigQuery Data Modelling Database Sharding, Replication	**AI for Data Engineers**	Machine Learning Processes Building Feature Stores Model Deployment
ETL/ETL	Data Ingestion Data Integration Streaming Data Processing	**Data Project Management**	Project Management Time Management Communications

FLEXIBLE SCHEDULE

PART-TIME	FULL-TIME
$9,100	**$13,000**
✓ Learning Schedule: • Wednesday Evening • Saturday Full Day	✓ Learning Schedule: • Wednesday Evening • Saturday Full Day
✓ 6-month Career Support after graduation	✓ 6-month Career Support after graduation
✓ Hands-on projects	✓ Hands-on projects
Full-time TA (additional projects)	✓ Full-time TA (additional projects)
Real client project	★ **Real client project**

CUSTOMIZED LEARNING PLATFORM



Self-paced learning guided and powered by our unique learning management system

where you can watch lectures, work on your codes, test your progress with live quizzes, track your learning journey, and communicate with instructors and TA's.

CURRICULUM

Semester 1 | 12 weeks

WEEK 1
Programming Basics

- Linux Commands
  - Bash
  - Text editing
  - CRON jobs
- Python Programming (Review)
  - Modules and Packages
  - OOP/Class
  - Key Data Structures (Queues, Stack, Trees, Heaps, etc.)

WEEK 2
Apache Spark I

- Scala Programming
- Introduction to Apache Spark (Scala)
  - Distributed Systems 101
  - RDD Programming
  - Spark DataFrame/SQL

WEEK 3
AWS I

- Cloud Computing 101
- Introduction to AWS
  - EC2/S3/Lambda
- Introduction to RDS
  - MySQL, Postgre

WEEK 4
AWS II

- Introduction to Streaming with Kinesis
- Introduction to Serverless with Lambda
- Data Ingestion using Kinesis Firehose, Lambda, and S3

WEEK 5
Docker

- Virtualization
- Docker Containers
- Docker Compose
- Deploying Flask applications using Docker and AWS ECR/EKS

WEEK 6
Kubernetes

- Kubernetes Introduction with MiniKF
- Working with Kubernetes on Google Cloud (GCP)
- Deploying Kubeflow Services on GCP
- Deploying Elasticsearch on Kubernetes

WEEK 7
RDBMS

- Relational Database Core Concepts: ACID, Normalization
- Replications, Sharding, Monitoring
- Google Big Query
- Amazon Redshift

WEEK 8
Data Warehouse

- Snowflake Database
- Data Modeling Deep Dive
- Query Performance Tuning
- Data Warehouse Design Patterns

WEEK 9
Data Lake: Hadoop

- Introduction to Big Data
- Distributed File Systems and Cloud Storage
  - AWS S3, Google Cloud Storage
- Data Warehouse with Apache Hive

WEEK 10
Data Lake: Presto/Spark

- SQL on Data Lake with Presto and Dremio
- Delta Lake with Spark on Databricks/AWS

WEEK 11
Data Lake: OLAP

- Data Warehouse/Lake Best Practices
- BI/OLAP on Lake with Apache Kylin and Druid

WEEK 12
Term 1 Project

- Modern BI on Lake/Warehouse Project
  - Build OLAP Cubes and Dashboard
  - Data Warehouse to Cloud Migration

CURRICULUM

Semester 2 | 12 weeks

WEEK 13

NoSQL Database I

- Introduction to NoSQL Databases (CAP)
- Column-family NoSQL
  - Cassandra
  - HBase
- Google BigTable

WEEK 14

NoSQL Database II

- Data Modeling for NoSQL
  - DynamoDB/Cassandra
- Full-Text Search Database
  - Elasticsearch

WEEK 15

Data Ingestion I

- Introduction to Real-time Data Ingestion
- Message Brokers (pub/sub)
  - Goolge Pub/Sub
  - Apache Kafka

WEEK 16

Data Ingestion I

- The Lambda and Kappa Architecture
- Stream Processing with Spark Streaming
- Stream Processing with Flink/Beam
- Data Lineage with Kafka, NiFi, and Atla

WEEK 17

Data Integration

- Introduction to Data Integration
- Data Integration: ETL vs ELT
- Data Integration with Spark and Talend
- Delta Lake

WEEK 18

Data Pipelines

- Building Data Pipelines with Apache Airflow
- Deploying Airflow in Production with AWS ECS

WEEK 19

Data Governance

- Data Quality and Metadata Management
- Data Quality: Data Profiling
- Data Quality: Validation, Standardization, Matching, and Enrichment
- Metadata Management: Cataloging and Discovery

WEEK 20

Machine Learning for Data Engineers

- Machine Learning Lifecycle Overview
- Building Feature Database/Stores for ML
- Deployment ML Pipelines using Apache Airflow, PySpark, and GCP

WEEK 21

Project Management

- Project Management 101
- Time Management
- Data Governance: The People, Policies, and Processes
- Managing Data Migration Projects
- Managing Data Lake Projects

WEEK 22/23/24

Capstone Project

- Capstone Project

ADMISSION PROCESS

Book a Counseling Meeting

Our program advisor will have a 1-on-1 meeting with you to see if the program is a good fit for you. There is no application fee. It takes about 30 minutes

Pass the Challenge

There will be a technical test and an interview. Applicants spend up to 2 hours on the challenge.

Reserve the Spot & Sign the Contract

Our admission officer will work with you directly to help you fill out a contract, pay $500 deposit and assist you to apply any kind of grant and finance options.

Start your Pre-bootcamp Learning

Starting ahead will gain you more experience and competence. Research shows that preview and preparation account for 73.7% success in academic achievement of university students.

TUITION, GRANTS AND FINANCE OPTIONS

$9,100
PART-TIME

$13,000
FULL-TIME

As an Ontario registered private career college, you can apply for student line of credit from BMO with lower interest rate.

Please check our website for early bird discounts

DO YOU KNOW?

With Ontario Second Career grant you may be eligible for up to $28,000 for costs including: tuition, books, manuals, transportation, basic living allowance, child care.

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