UNIVERSITY OF TORONTO
SCHOOL OF CONTINUING STUDIES

COURSE OUTLINE

COURSE CODE: 3252-064

COURSE TITLE: Big Data Management Systems and Tools

INSTRUCTOR: Stephen Giles

416.917.6823

stephen.giles@utoronto.ca

Please communicate with the instructor primarily through the online discussion boards. They will be checked once a day. For urgent requests, please email the instructor.

DAY/TIME:

January 16, 2025 **to** April 3, 2025

SCHEDULE: 6 PM

REQUIRED TEXT(S): (1) **[DDA]** Kleppmann, Martin. Designing Data-Intensive Applications. O'Reilly. 2017. (2) **[LS]** Damji et al. Learning Spark, 2nd Edition. O'Reilly. 2020.

PREREQUISITE(S)/ RECOMMENDATIONS: 3253 Machine Learning

Recommended:

You should have a laptop with at least 8 GB of RAM to bring to class. You will need to have access to a laptop or desktop outside class with an i5 or preferably i7 processor, that can run recent Windows, Mac or Linux operating systems. Any software you'll need is free and mostly open source. You will receive further instructions in class.

CERTIFICATE(S): This course is applicable to the following SCS Certificates

Data Science: Required Courses

COURSE DESCRIPTION:

Big Data involves massive data volumes and diverse data types. Modern organizations need people who can help implement the tools they need to deal with these huge data sets. In this course, you'll learn the technology of Big Data and build in-demand skills. You'll get hands-on experience using up-to-date database management systems and tools. Get in on the explosion in new NoSQL technologies and Big Data tools including Spark, Cassandra and Kafka.

LEARNING OUTCOMES:

By the end of this course, learners will be able to:

- Understand the architecture of reliable Big Data systems
- Describe how they differ from traditional systems
- Use several NoSQL database management systems
- Address the many challenges of working with data at scale
- Use tools such as Spark to process large datasets

ASSOCIATIONS: This course is not recognized by an Association.
https://learn.utoronto.ca/programs-and-courses/association-partnerships

COURSE FORMAT:

Course Format	Definition
[Classes and/or Webinars]	In-Class Thursdays
Assignments	Four assignments.
Term Project	Group or individual.

QUERCUS and ONLINE RESOURCES

Quercus is a learning management engine, and is the University of Toronto's main online teaching and learning environment. It is web-based software which gives you and your instructor a shared learning space online to receive and exchange course content as well as to communicate using a range of tools. More information about online learning in Quercus can be found here: https://help.learn.utoronto.ca/hc/en-us/sections/115000462414-Online-learning-in-Quercus

UNIVERSITY OF TORONTO
SCHOOL OF CONTINUING STUDIES

COURSE PLAN:

MODULES	DATE	TOPIC/LESSONS/WEBINARS	READINGS, ASSIGNMENTS, ACTIVITIES, ETC.
1 – Reliable and Scalable Data-Intensive Applications	Click or tap here to enter text.	• Course Logistics • Challenges of Big Data • Practices that make 24/7/365 Operations possible • Reactive Design	• Read DDA Chap. 1
2 – Relational Databases and Data Warehouses	Click or tap here to enter text.	• The Relational Model • SQL • Data Warehouses	• Assignment 1 assigned
3 – NoSQL	Click or tap here to enter text.	• Key-value stores • Column-oriented databases • Document-oriented databases • Object, graph and triple stores • MongoDB	• Read DDA Chap. 2 • Assignment 1 due • Assignment 2 assigned
4 – Distributed Datastores	Click or tap here to enter text.	• Distributed filesystems • The scalability problem for relational databases • The CAP theorem • Replication and partitioning	• Read DDA Chap. 8 • Assignment 2 due
5 – Introduction to Spark	Click or tap here to enter text.	• Setting up Spark • Spark datasets and dataframes	• Read LS Chaps. 1-3
6 – Analytics with Spark	Click or tap here to enter text.	• Spark actions • Spark transformations	• Read LS Chaps. 4-7 • Assignment 3 assigned
7 – Spark Streaming	Click or tap here to enter text.	• What's different about streaming data • Stream processing in general • Spark stream processing	• Read LS Chap. 8 • Assignment 3 due • Assignment 4 assigned
8 – Data in Motion	Click or tap here to enter text.	• Metadata • Patterns of data transfer • Formats for serializing and transmitting data	• Read DDA Chap. 4 • Assignment 4 due
9 – Big Data Architecture	Click or tap here to enter text.	• Lambda architecture • Lakehouses • Kafka and realtime analytics	• Read DDA Chaps. 9 and 11

10 – Cloud Analytics	Click or tap here to enter text.	• Amazon • Azure • Google • Snowflake	
11 – MLOps	Click or tap here to enter text.	• Intro to MLOps • Continuous Deployment • Tools for MLOps	
12 – Term Project Presentations 2	Click or tap here to enter text.	• Term project presentations	• Term project due

GRADING AND EVALUATION:

Project	30%
Assignments	60%
Participation	10%

A note about Participation Marks:

Participation marks are for attendance and, more importantly, constructive contributions to class discussion on a regular basis.

Learners can expect to receive feedback and marks, if applicable, before the course end date, for all their submitted assignment(s) and test(s) other than the final exam, project or course paper. However, it is the sole responsibility of learners to make sure that they do get these marks from their Instructor and have all related questions answered before the course ends.

MISSED TEST/ASSIGNMENT GUIDELINES

If you miss a test or assignment please work directly with your instructor to make alternate arrangements. There may be penalties for missed or late assignments and tests.

SCS GRADING SCALE:

A 80% to 100% Excellent

B 70% to 79% Good

C 60% to 69% Adequate

D 50% to 59% Marginal

FX Less than 50%

INC Incomplete

DNW Did not write

FINAL GRADE:

To view your final grade, please login to "My Access – Student Login" at: https://learn.utoronto.ca/login. Please note that your final grade will not be posted on Quercus.

More information regarding Academic Policies and Guidelines is located here: https://help.learn.utoronto.ca/hc/en-us/sections/207314307-Academic-Policies-and-Guidelines

CODE OF STUDENT CONDUCT AND CODE OF BEHAVIOUR ON ACADEMIC MATTERS:

All School of Continuing Studies learners are required to comply with the University of Toronto Academic Policies including, but not limited to the Code of Student Conduct and the Code of Behaviour on Academic Matters.

Information regarding University of Toronto Academic Policies can be reviewed here: https://help.learn.utoronto.ca/hc/en-us/articles/235279047-Academic-Policies-and-Student-Code-of-Conduct

ACADEMIC HONESTY:

Course work that is not appropriately cited may be in violation of the Code of Behaviour on Academic Matters (see above).

For guidelines about plagiarism and properly citing your sources, please visit:

https://help.learn.utoronto.ca/hc/en-us/articles/115006427548-Academic-Honesty

AUDIO/VIDEO RECORDINGS:

You are not permitted to record lectures without the written consent of your instructor(s).

ACADEMIC ACCOMMODATIONS:

If you require accommodation for a disability, please contact Enrolment and Learner Services at 416-978-2400, email scs.accessibility@utoronto.ca or fill out the form at the following link to arrange this service. https://learn.utoronto.ca/help/forms-and-applications/accommodation-request-form