

COSC 448 (3) Direct Studies

Instructor: Gema Rodríguez-Pérez gema.rodriguezperez@ubc.ca

Duration: Winter 2026, 3 credits

Academic Calendar Entry

COSC 448 (3) Direct Studies: Development and Deployment of a Scalable Data Pipeline for Research

Description: Students will customize and extend GrimoireLab, an open-source software development analytics platform, to meet the specific data collection and analysis needs of the STASER lab, led by Dr. Gema Rodríguez-Pérez. This work involves building a robust and modular data pipeline to automate the collection, cleaning, enhancement, and storage of data from diverse software repositories and communication platforms. Through this project, students will deepen their understanding of scalable data engineering workflows while adhering to modern software engineering best practices.

Prerequisite: and permission of the department head.

Course Format, technology requirements, and communication

This directed study will follow a project-based and collaborative format, where students will work together to design, implement, and evaluate a data pipeline tailored to the STASER lab's needs.

The **course format** will include:

- Weekly team meetings (in person or virtual) to review progress, plan tasks, and discuss challenges
- Sprint-based development using agile methodologies (e.g., weekly planning and retrospectives)
- Independent and team-based work, including coding, documentation, testing, and system integration

Students are expected to dedicate approximately 10 hours per week on project-related tasks, including meetings, development, reading, experimentation, and documentation.

Technology Requirements: Students are required to have a personal computer and use it for the duration of the directed studies.

Communication: Attendance and participation in the weekly scheduled meetings is required. Students should come prepared with updates and discussion points. Students are expected to communicate regularly through the selected team channel (Slack, teams, or similar)

Course Overview, Content and Objectives

This directed study emphasizes the practical application of core software engineering practices, including requirements gathering, modular system design, testing, validation, and agile development. Students will gain hands-on experience with state-of-the-art tools and technologies for mining, storing, and visualizing data such as Docker, Elasticsearch, and Kibana and learn how to integrate components that leverage large language models (LLMs) where appropriate to address data enrichment or analysis challenges.

The course will cover:

- Core Software Engineering processes such us Requirements gathering, modular design, testing, validation, and agile methodologies

- Understanding and implementation of Open-source Analytics Systems
- Developing customized scripts for automating data collection from software repositories (e.g., GitHub, GitLab), structuring and enriching collected data using scripts or auxiliary tools
- Indexing and storing structured data using Elasticsearch
- Creating interactive dashboards and visual summaries with Kibana or similar platforms
- Applying language models to improve or analyze collected data where appropriate

Learning Outcomes

By the end of the course, the student will be able to

- 1) Design and implement a modular data pipeline for software analytics using GrimoireLab and related technologies.
- 2) Apply foundational software engineering practices, including planning, testing, documentation, and agile development.
- 3) Automate the collection and preprocessing of data from multiple heterogeneous sources.
- 4) Store and query enriched datasets using Elasticsearch and similar indexing tools.
- 5) Build and deploy visual dashboards to support interactive data exploration for researchers.
- 6) Integrate LLM-powered enhancements to process or interpret the collected data, based on project scope.
- 7) Collaborate effectively in a team setting, contributing to shared codebases and participating in regular sprint meetings.

Schedule

We cover the following topics:

Week	Topic
1	Introduction and project overview
2 - 3	Introduction to GrimoireLab & Data Sources
4 - 5	Requirements Gathering & System design & Project planning
6-9	Software development & Testing
10	Midpoint Review
11-13	Software development & Testing
14	Delivery and presentation

Evaluation Criteria and Grading

Task	Weight %
Weekly Progress and participation	42
Code quality	30
Documentation	20
Final presentation	8

The **Weekly Progress and Participation** will account for 42% of the grade. Each week students are expected to actively contribute to the project each week by making regular commits to the GitHub repository, managing issues, and participating in discussions. This includes creating and updating GitHub issues to track tasks and bugs, as well as engaging constructively in Pull Request (PR). Responsiveness to feedback and collaborative communication within the team are important aspects of this criterion. The weight per week will be 3.5% between weeks (2-11).

Code Quality represents 30% of the grade and focuses on the quality of the submitted work. Students should ensure their code is clean, readable, and adheres to best practices and coding standards. Proper use of Git workflows, such as creating feature branches and submitting well-documented PRs with clear descriptions, is expected. Additionally, implementing appropriate testing such as unit or automated tests to verify the correctness and reliability of the code is essential for a high score in this area. This will be evaluated after the students have submitted their code.

Documentation is worth 20% of the total grade. Students should provide clear, comprehensive, and up-to-date documentation that helps others understand and use the project. This includes a well-maintained README file that explains the project setup, usage instructions, and overall goals. Inline comments and explanations within the code are also important for clarity. Furthermore, all GitHub issues and PRs should be properly described to reflect the current state of the project. This will be evaluated after the students have submitted their code.

Finally, the **Final Presentation** accounts for 8% of the grade. The presentation should effectively communicate the project's objectives, implementation process, and outcomes. It should demonstrate the developed features and functionality clearly and address any challenges encountered along the way, including how they were resolved. The final presentation will be on December 3rd, 2025.

Academic Integrity

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work. Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in failing the course and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences.

A more detailed description of academic integrity, including the University's policies and procedures, may be found in the Academic Calendar at <http://okanagan.students.ubc.ca/calendar/index.cfm?tree=3,54,111,0>.

More details on collaboration and plagiarism is available at <https://my.cs.ubc.ca/docs/collaboration-plagiarism>

Student Service Resources

UBC Okanagan Disability Resource Centre

The Disability Resource Centre ensures educational equity for students with disabilities and chronic medical conditions. If you are disabled, have an injury or illness and require academic accommodations to meet the course objectives, please contact Earlene Roberts, the Diversity Advisor for the Disability Resource Centre located in the University Centre building (UNC 214).

UNC 214 250.807.9263
email: earlene.roberts@ubc.ca
Web: www.students.ok.ubc.ca/drc

UBC Okanagan Equity and Inclusion Office

Through leadership, vision, and collaborative action, the Equity & Inclusion Office (EIO) develops action strategies in support of efforts to embed equity and inclusion in the daily operations across the campus. The EIO provides education and training from cultivating respectful, inclusive spaces and communities to understanding unconscious/implicit bias and its operation within in campus environments. UBC Policy 3 prohibits discrimination and harassment on the basis of BC's Human Rights Code. If you require assistance related to an issue of equity, educational programs, discrimination or harassment please contact the EIO.

UNC 216 250.807.9291
email: equity.ubco@ubc.ca
Web: www.equity.ok.ubc.ca

Health & Wellness

At UBC Okanagan health services to students are provided by Health and Wellness. Nurses, physicians and counsellors provide health care and counselling related to physical health, emotional/mental health and sexual/reproductive health concerns. As well, health promotion, education and research activities are provided to the campus community. If you require assistance with your health, please contact Health and Wellness for more information or to book an appointment.

UNC 337 250.807.9270
email: healthwellness.okanagan@ubc.ca
Web: www.students.ok.ubc.ca/health-wellness



Student Learning Hub

The Student Learning Hub (LIB 237) is your go-to resource for free math, science, writing, and language learning support. The Hub welcomes undergraduate students from all disciplines and year levels to access a range of supports that include **tutoring in math, sciences, languages, and writing, as well as help with study skills and learning strategies**. For more information, please visit the Hub's website (<https://students.ok.ubc.ca/student-learning-hub/>) or call 250-807-9185.

Safewalk

Don't want to walk alone at night? Not too sure how to get somewhere on campus? Call Safewalk at 250-807-8076.

For more information, see: www.security.ok.ubc.ca