CMPT 756 - G100 Systems For Big Data

Calendar Objective/Description:
From health care to social media the world generates a tremendous amount of data every day, often too much to be processed on a single computer or even sometimes a single data centre. In this graduate seminar we will learn about technologies and systems behind Big Data. In particular, we will discuss what challenges exist in processing and storing massive amounts of data. We will explore how these challenges are being solved in real-world systems as well as the limitations inherent in these designs. The evolution of these technologies will be explored by reading both current and historically significant research papers.

Instructor's Objectives:
Data engineering turns a proof-of-concept data model into an efficient, maintainable service well-matched to running in large-scale data centres. Such services are resilient against failure of either their own instances or a service dependency, deployed in a controlled way, suited for co-tenancy with services with different latency goals, instrumented to detect performance problems, and compatible with the data centre scheduler. Data engineering ensures that the services run all night while the operations staff sleep all night.

Prerequisites:
Operating Systems (CMPT 300) and Data Base Systems (CMPT 354), or equivalents. Students with credit for CMPT 886 when offered as a Special Topics course in Big Data may not take this course for further credit.

Topics:
- Design space analysis: Analyze the tradeoffs inherent in a cloud service design.
- Service level metrics, indicators, objectives, and agreements: Describe the differences.
- Resilience and recovery: Estimate, test, and improve the resilience of a design to failure.
- Latency versus consistency: Analyze tradeoffs between connected and partitioned performance.
- Data centre operating systems: Differentiate approaches to managing the resources of a data centre.
- Instrumentation: Classify the approaches to instrumenting a system and add instrumentation.
- Consensus: Summarize the need for consensus, particularly for service metadata and their algorithms.

Grading:
All assignments will be projects; there will be no tests or final exam. Projects will be a mix of individual and group efforts. The final project will be done in groups, for a total of 50% of the course grade.

Students must attain an overall passing grade on the weighted average of exams in the course in order to obtain a clear pass (C- or better).

Academic Honesty Statement:
Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct Policies (http://www.sfu.ca/policies/gazette/student.html).