Master Course Syllabus
School of Engineering and Computer Science
Washington State University Vancouver
CS 453
Cloud Data Management
3 Semester Hours
(3 lecture hours)

Catalog Description

Principles of cloud data management: data models, fragmentation, processing paradigms, consistency, storage, and commercial cloud data management platforms.

Prerequisite Courses

- CS 351 with a C or better

Prerequisite Topics

- Proficiency with the Java programming language
- Relational data model
- Relational algebra and SQL
- Basic discrete mathematics
- Basic logic
- Use of Unix environment for coding, compilation, debugging and testing

Measured Course Outcomes

Students taking this course will:

1. Analyze or solve problems using NoSQL database systems and suitable algorithms. (Contributes to performance criterion 1-b.)
2. Evaluate the merits of multiple solution designs in distributed data management systems with respect to performance, consistency and reliability. (Contributes to performance criterion 1-c.)
3. Produce group project reports, using appropriate formats, grammar, and content targeting technical and non-technical audiences. (Contributes to performance criterion 3-a.)
4. Deliver well-organized project presentation and answer questions professionally. (Contributes to performance criterion 3-b.)

Other covered course outcomes but not measured

1. Share project information and schedule with team members (contributes to performance criterion 5-b).
2. Develop NoSQL databases in a team (contributes to performance criterion 5-c).
3. Implement NoSQL DB queries corresponding to given data models (contributes to performance criterion 6-c).
4. Process a large volume of web data using the MapReduce programming model (contributes to performance criterion 6-d).

Required Textbooks
None Specified.

Reference Material
1. M. Tamer Ozsu and Patrick Valduriez, *Principles of Distributed Database Systems*, Springer.
2. "Hadoop: The Definitive Guide," Tom White (Oâ€™Reilly).

Major Topics Covered in the Course
1. Data models
2. Distributed query Processing & optimization
3. Distributed data storage
4. Data consistency
5. Reliable Data Management
6. Distributed key-value stores
7. Map-Reduce
8. Web programming

Projects

Programming Project Area	Weeks
Map data to a variety of data models, e.g., key/value model, document model, and column family model	3
Profile system characteristics of NoSQL databases	3
Design and implementation of data analytics using NoSQL databases, e.g., MongoDB	3

Design, Implementation, and Analysis
Students are guided through the several design phases in the development of a distributed data management system. Students analyze their solution to identify weaknesses in the design and implementation.
Students take real-world big data management problems and implement efficient solutions. The problem analysis has two related components. The first component is determining the data management needs encountered in a particular contemporary distributed computing scenario. Students learn the capabilities and limitations of current techniques, through application design and implementation. The second component is an analysis of the practical limitations of current distributed data management solutions.

CC2013
This course provides coverage of CS2013 knowledge areas. Values listed are minimum course hours dedicated to the topic, percentages indicate the fraction of CS2013 knowledge area topics covered (acceptable values are: <25%, 25-75%, >75%, or 100%).

Area	Tier 1	Tier 2	Elective
IM/Information Management Concepts		2 (25-75%)	
IM/Database Systems			3 (100%)
IM/Query Languages			4 (25-75%)
IM/Distributed Databases			10 (100%)

Course Coordinator:	Xuechen Zhang
Last Updated:	May 11, 2018
Syllabus Version Number:	2.1