DALHOUSIE UNIVERSITY
Inspiring Minds

Winter 2017

CSCI 5408 Data Management, Warehousing, and Analytics

Course Syllabus

Instructor & TA Information: (Class time: TR, 2:35-3:55PM in FASS 2016)

- Instructors:
  - **Part-I, Data Management:** DB, Distributed System and Cloud Computing (Weeks: 1-6)
    - Suhaib Qaiser (suhaibqaiser@dal.ca); Office Hours: MW, 4:30-6:00PM, CS 427
  - **Part-II, Data Analytics:** Data Warehousing and Data Mining (Weeks: 8-13)
    - Dr. Qigang Gao (q.gao@dal.ca); Office Hours: MR, 12:30-2:00PM, CS 219
- TA: Abhinav Kalra (abhinav.kalra@dal.ca); Ass-help: WF, 1:00-2:30PM, CS 134/233

Important Dates:

- University Closed: Feb 3 (Munro Day) & Feb 20 (NS Heritage Day)
- Study Break: Feb 20-24
- Midterm Exam (2 hours): Feb 27, 7:00-9:00PM, Room: CS 127
- Final Exam (2 hours): TBA in the period of Apr 12-26
- Assignment Submission Deadlines:
  - Jan 24 (Ass1), Feb 7 (Ass2), Feb 21 (Ass3), Mar 14 (Ass4), Mar 28 (Ass5), Apr 11 (Ass6)
- Tutorial Dates: (1:00-2:30PM, Room: CS 134)
  - Ass1-tutorial: Jan 11
  - Ass2-tutorial: Jan 25
  - Ass3-tutorial: Feb 08
  - Ass4-tutorial: Mar 01
  - Ass5-tutorial: Mar 15
  - Ass6-tutorial: Mar 29

Course Abstract:

- In this course, we will focus on three pillars for managing and analyzing data in distributed and cloud environments: Management of data in distributed systems, Data Warehousing, and Data Analytics.
- **Prerequisites:** CSCI 2110, CSCI 2141

Evaluation Scheme:

- Assignments: 50% (Six, allowed to work in a team of 2 students)
- Midterm Exam: 25% (2 hours, Part-I, Closed book)
- Final Exam: 25% (2 hours, Part-II, Closed book)

Assignments:

- Assignments are for learning through developing small application tasks. All assignments are allowed to work as a team of 2 students.
- To access assignments and submit your completed work, login the Dal's Online Learning Management System "Brightspace".

- **No late submission will be accepted.** Accordingly, you should submit whatever you have at the due, even it has not been completed yet, to receive at least a partial mark. You are strongly recommended to work on each assignment as early as you can.
- There will be a tutorial for each assignment, which provides more detail instructions and examples in helping you to proceed for achieving better work result.
- Assignment Help-Hours are to provide you some handy help individually for any questions related to assignments. You may also access TA via email if you cannot attend the scheduled Ass-help hours.

Textbook & Lecture Slides:

- **Textbook for Part-I:** ***"****Cloud Computing: Concepts, Technology & Architecture"*, The Prentice Hall Service Technology Series from Thomas Erl,.
    - http://ptgmedia.pearsoncmg.com/images/9780133387520/samplepages/0133387526.pdf
- **Textbook for Part-II:** "*Data Mining: Concepts and Techniques"*, J. Han, M. Kamber, and J. Pei, Morgan Kaufmann, 2011 (3rd Edition). *The2nd Edition is also okay which is available from the Brightspace.
    - As a reference, another recent DM textbook example: "Data Mining: The Textbook" by C.C. Aggarwal, Springer Pub, 2015 is also available from the Brightspace.
- **Lecture Notes:** Available after each class from the Brightspace system.
  A set of questions will be posted at the end of each slides, it is strongly recommended to get your answers in writing for those questions after each class for examining your understanding and for preparing for exams.

Web Resource Examples:

- **Part-I,** DB, Distributed System and Cloud Computing:
    - Velte, A. T., Velte, T. J., Elsenpeter, R. C., & Elsenpeter, R. C. (2010). *Cloud computing: a practical approach* (p. 44). New York: McGraw-Hill. (available online)
    - AWS Certified Solutions Architect Official Study Guide: Associate Exam by Joe Baron (Author), Hisham Baz (Author), Tim Bixler (Author), Biff Gaut (Author), Kevin E. Kelly (Author), Sean Senior (Author), John Stamper
    - Learning Spark by O'Reilly (https://www.pdf-archive.com/2016/04/21/learning-spark-o-reilly-2015/preview/page/1/)
    - JAMES, BD. "Security and privacy challenges in cloud computing environments." (2010).
- **Part-II,** Data Mining and Data Warehousing:
    - General Data Mining Site: http://www.kdnuggets.com
        - A collection of DM publications, software/tools, data repositories, companies, etc.
    - Data Mining Software - Weka: http://www.cs.waikato.ac.nz/ml/weka/
        - A collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform.
    - Database Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
    - Wikipedia: http://en.wikipedia.org/wiki/Data_mining

Other Course Management Note:

- The primary source of communication will be in class. Attendance in class is expected; if you must miss a class, talk to a fellow student to see anything else you should know

besides the class note. Additional communications will be posted to the course email list. It is the student's responsibility to check their class emails on a regular basis.

Course Outline:

WEEK	TOPIC	EXERCISE
1	Lecture 1: Review relational DB topics: SQL, Primary key, Foreign key, Indexes, Triggers, SQL, Conventional DB, MySQL, Oracle, Performance limitations, memory limitations, fault tolerance, NoSQL, Big Data, Query Search, Elastic Search Lecture 2: Intro to Cloud Computing, Cloud Characteristics (on-demand usage, ubiquitous access, multi-tenancy, elasticity, measured usage, resiliency), Delivery models (IaaS, PaaS, SaaS), Deployment models (Public, Private, Hybrid, Community)	Ass1: DB "Implementing Search Query using RDBMS with and without Clouds"
2	Lecture 1: Cloud Architecture: a) Basics – Grid Computing, Distributed Computing, Parallel Computing, Cloud Computing; b) Virtualization – Hardware Independence, Server Consolidation, Resource Pooling, Resource Replication, Zero downtime architecture; c) Web platform – App Servers, HTML5, JavaScript, Deploying services, Web & REST services Lecture 2: Scalability in Clouds, Hypervisor, Load Balancers, Managed vs Unmanaged Services, SLAs, Billing, Resource Management, DaaS, replication, failover, build automation, disaster recovery, role management	
3	Lecture 1: Cloud Security : Symmetric and Assymetric encryption, Hashing, Digital Signature, PKI, Identity Management, Sign Sign-On, Security Groups, Firewalls, Security Threats, Performance Tuning, Security Measures Lecture 2: Amazon Cloud: EC2, S3, EMR, Data Pipeline, Elastic Cache (Redis), DynamoDB, Load Balancer, Elastic Search	Ass2: DC "Load Balancing in Cloud Environments with and without Firewall"
4	Lecture 1: Example of various Cloud providers such as Heroku, Salesforce, Azure, IBM, etc. Introduction to Big Data databases like Casandra and MongoDB. Lecture 2: Big Data in Clouds (part 1): Apache Hadoop and NoSQL databases. Virtualization in Hadoop. HDFS system. Need for Hadoop in current industry.	
5	Lecture 1: Big Data in Clouds (part 2): Data pipelines and Apache Sparks. Concept of RDD and Sparks SQL. Aggregations, Streaming and MLib in Sparks. Lecture 2: Protection of Privacy and Security in Cloud environments. Latest issues relating to privacy and security. Performance and other concerns and their resolution while keeping industry standard and practices. Review of a research paper on Security protection in Cloud environments.	Ass3: CC "Implementing Bid Data in Clouds. Using either Spark or Hadoop

6	Lecture 1: Review of a use case from Industry: Salesforce massive loss of data and their shift to new Cloud Service Provider. Discussion on reasons of failure and remedies. Lecture 2: Review of effective remedies to protect data from being lost, remedies to secure and save data from security breaches, effective monitoring. Examples of Amazon, Microsoft, Salesforce and other breaches of data from service providers. "What we have learned from our mistakes". Concluding Cloud topics and discussion.	
7	**Study Break**	**Midterm**
8	Overview on data analytics: DW for OLAP, DM for knowledge discovery Data preprocessing for data warehousing and data mining (by Q. Gao)	Ass4: DW
9	Multi-dimensional data model, DW schemas, OLAP operations	
10	Classification pattern discovery for class prediction, Decision tree based methods, Bayesian classification, etc.	Ass5: Classification
11	Text classification, social media data analysis, semantic classification, etc.	
12	Association rules discovery for variable relationship analysis, Recommender system	Ass6: Association
13	Clustering analysis for concept discovery	
		Final Exam

Academic Integrity[1] :

At Dalhousie University, we respect the values of academic integrity: honesty, trust, fairness, responsibility and respect. As a student, adherence to the values of academic integrity and related policies is a requirement of being part of the academic community at Dalhousie University.

What does academic integrity mean?

Academic integrity means being honest in the fulfillment of your academic responsibilities thus establishing mutual trust. Fairness is essential to the interactions of the academic community and is achieved through respect for the opinions and ideas of others. "Violations of intellectual honesty are offensive to the entire academic community, not just to the individual faculty member and students in whose class an offence occurs." (see Intellectual Honesty section of University Calendar)

How can you achieve academic integrity?

- Make sure you understand Dalhousie's policies on academic integrity.
- Give appropriate credit to the sources used in your assignment such as written or oral work, computer codes/programs, artistic or architectural works, scientific projects, performances, web page designs, graphical representations, diagrams, videos, and images.
- Use RefWorks to keep track of your research and edit and format bibliographies in the citation style required by the instructor http://www.library.dal.ca/How/RefWorks
- Do not download the work of another from the Internet and submit it as your own.
- Do not submit work that has been completed through collaboration or previously submitted for another assignment without permission from your instructor.
- Do not write an examination or test for someone else.
- Do not falsify data or lab results.

[1] Based on the sample statement provided at http://academicintegrity.dal.ca.

These examples should be considered only as a guide and not an exhaustive list.

What will happen if an allegation of an academic offence is made against you?
I am required to report a suspected offence. The full process is outlined in the Discipline flow chart, which can be found at: http://academicintegrity.dal.ca/Files/AcademicDisciplineProcess.pdf and includes the following:

1. Each Faculty has an Academic Integrity Officer (AIO) who receives allegations from instructors.
2. The AIO decides whether to proceed with the allegation and you will be notified of the process.
3. If the case proceeds, you will receive an INC (incomplete) grade until the matter is resolved.
4. If you are found guilty of an academic offence, a penalty will be assigned ranging from a warning to a suspension or expulsion from the University and can include a notation on your transcript, failure of the assignment or failure of the course. All penalties are academic in nature.

Where can you turn for help?

- The Academic Integrity website http://academicintegrity.dal.ca has links to policies, definitions, online tutorials, tips on citing and paraphrasing.
- The Writing Center provides assistance with proofreading, writing styles, citations.
- Dalhousie Libraries have workshops, tutorials, citation guides, Assignment Calculator, RefWorks, etc.
- The Dalhousie Student Advocacy Service assists students with academic appeals and student discipline procedures.
- The Senate Office provides links to a list of Academic Integrity Officers, discipline flow chart, and Senate Discipline Committee.

Student Accommodation

Students may request accommodation as a result of barriers related to disability, religious obligation, or any characteristic under the Nova Scotia Human Rights Act. Students who require academic accommodation for either classroom participation or the writing of tests and exams should make their request to the Advising and Access Services Center (AASC) prior to or at the outset of the regular academic year. Please visit www.dal.ca/access for more information and to obtain the Request for Accommodation – Form A.
A note taker may be required as part of a student's accommodation. There is an honorarium of $75/course/term (with some exceptions). If you are interested, please contact AASC at 494-2836 for more information.
Please note that your classroom may contain specialized accessible furniture and equipment. It is important that these items remain in the classroom, untouched, so that students who require their usage will be able to participate in the class.

Culture of Respect[2]

Every person has a right to be respected and safe. We believe inclusiveness is fundamental to education and learning. Misogyny and disrespectful behavior in our classrooms, on our campus, on social media, and in our community is unacceptable. We stand for equality. We hold ourselves to a higher standard. (http://www.dal.ca/faculty/computerscience/about/respect.html)

---

[2] Source: Speak Up! © 2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full "Speak Up" document found at: http://www.dal.ca/dept/dalrespect.html. Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University, 902.494.4140, lyndsay.anderson@dal.ca www.dal.ca/think.