San José State University
Department of Computer Science
CS 267 – Topics in Database, Section 1, Fall 2017

Course and Contact Information

Instructor:	Ahmed Ezzat
Office Location:	**MH**, Room 218
Email:	Ahmed.Ezzat@sjsu.edu
Office Hours:	Tu, Th 8:00AM – 9:00AM (by advanced reservation)
Class Hours:	Tu,Th: 9:00AM – 10:15AM
Classroom:	**DH-450**
Prerequisites:	CS 157B (Database Management Systems II) with a grade of C- or better, or instructor's consent. The Department of Computer Science strictly enforces prerequisites. The instructor may drop any student who does not show up for the first two class meetings without providing a valid excuse ahead of time.
Grader: TBD	**Email:** TBD

Course Description

Operating Systems: The curse covers the following topics with Big data: NoSQL data modeling. Large-scale data processing platforms. HDFS, MapReduce and Hadoop. Scalable algorithms used to extract knowledge from Big data. Advanced scalable data analytics platforms. **Prerequisite:** CS 157B (with a grade of "C-" or better). Computer Science, Applied and Computational Math or Software Engineering Majors only; or Instructor Consent

Course Learning Outcomes (CLO)

Upon completion of this course, students would achieve the following:

- Explain key concepts, algorithms, techniques related to Big Data.
- Become familiar with Mining data streams.
- Become familiar with Apache Hadoop architecture, and Map-Reduce.
- Become familiar with the different data models used by NoSQL Big Data platforms.
- Become familiar with tradeoffs between SQL and NoSQL: Data model, Query language, guarantees provided.

Required Texts/Readings

- **Mining of Massive Datasets,** Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman, Cambridge University Press, ISBN: 978-1-107-01535-7. Download copy: http://i.stanford.edu/~ullman/mmds.html

- **Hadoop: The Definitive Guide,** Tom White, O'Reilly, 4rd Edition, 2015, ISBN: 978-149-190-1687, http://www.iteblog.com/downloads/OReilly.Hadoop.The.Definitive.Guide.4th.Edition.2015.3.pdf
- **Cassandra: The Definitive Guide,** Eben Hewitt, O'Reilly, http://www.gocit.vn/files/Cassandra.The.Definitive.Guide-www.gocit.vn.pdf

Online Reading Material:

- **Publications for relevant algorithms in Big Data**: http://pagesperso-systeme.lip6.fr/Marc.Shapiro/pubs.html#top10
- **Strong vs. Eventual Consistency Models**: http://ehcache.org/documentation/get-started/consistency-options
- **Balancing Strong and Eventual Consistency-** case study: https://cloud.google.com/developers/articles/balancing-strong-and-eventual-consistency-with-google-cloud-datastore
- **Apache Hadoop**: http://hadoop.apache.org/
- **Hadoop HDFS**: http://wiki.apache.org/hadoop/HDFS
- **MapReduce Overview**: http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
- **Analytics Overview**: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=5&ved=0CFgQFjAE&url=http%3A%2F%2Fwww.people.vcu.edu%2F~randrews%2FAnalytics%2FEvans2012slides%2FEvans_Analytics1e_ppt_01.pptx&ei=pzaMUvXDEoGbiAL0i4CABg&usg=AFQjCNFQBmmysw_WXGraHOZE7F69Ot4luw&sig2=0qk000GeJBcA-dDDTaqtjg
- **Mahout - Scalable Data Mining Algorithms Over Hadoop**: http://mahout.apache.org/
- **Apache Hive Home Page**: http://hive.apache.org/
- **Apache Pig Home Page:** http://pig.apache.org/
- **Hbase Home Page**: http://hbase.apache.org/
- **Cassandera Home Page**: http://cassandra.apache.org/
- **CouchDB Home Page**: http://couchdb.apache.org/
- **MongoDB Home Page**: http://couchdb.apache.org/

Course Requirements and Assignments

All the assignments and related documents must be handed in the classroom on due date. Students will lose 10% of the homework or project grade for each day delay, and after 5 days, homework or projects will not be accepted.

Homework and Project descriptions are available on Canvas

- Homework-1: Assignment is on Sept. 12, 2017, and is due back on Sept. 21, 2017.
- Homework-2: Assignment is on Oct. 10, 2017, and is due back on Oct. 19, 2017.

- Homework-3: Assignment is on Nov. 7, 2017, and is due back on Nov. 23, 2017.

- Group Paper: Assignment is on Aug. 29, 2017, and is due back on Nov. 28, 2017.
- Group Project: Assignment is on Aug. 29, 2017, and is due back on Nov. 30, 2017.

Exams or Evaluation

The midterm and final examinations will be closed book and no notes. There will be no laptops, or any personal digital devices allowed. There will be no make-up exams. If a student misses an exam without a legitimate excuse, a grade of zero will be recorded. If a student missed an exam with a legitimate excuse then the grade for that exam will be prorated. More details can be found on final examination in *University Policy S06-4 (http://www.sjsu.edu/senate/docs/S06-4.pdf) which states that "There shall be an appropriate final examination or evaluation at the scheduled time in every course, unless specifically exempted by the college dean who has curricular responsibility for the course."*

Grading Information

Your individual class grade will be weighted as follows:

• Three HW Assignments	15%	15 points	Individual
• Project	15%	15 points	Group score
• Research Paper	10%	10 points	Group score
• Class Quizzes	10%	10 points	individual score
• Midterm	25%	25 points	individual scores
• Final exam	25%	25 points	individual scores

Each assignment, project, and exam will be scored (given points) but not assigned a letter grade. The mean score for the Midterm will be announced after the exam.

Important NOTE:

In some occasions, the instructor may decide to administer evaluations where students are allowed to use their "paper notes"- (NO BOOK OR ELECTRONIC COPIES), so it is in your best interest to attend to class and take good notes; they may be handy in such situations. These individual in-class quizzes are not scheduled in-advance and they and class participation account to 10% of the overall class grade.

Determination of Grades

Final individual class letter grades will be assigned based on the class curve (i.e. relative grading). Your final class grade can be adjusted up or down depending on your level and quality of participation on your project team.

Classroom Protocol

It is expected that student attend classes, be active and participate in the class by asking/answering questions, arrive in time and leave only after the class is ended. No eating is allowed in the classroom, and it is expected to turn your cell off before entering the classroom.

University Policies

General Expectations, Rights and Responsibilities of the Student

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/

CS 267, Topics in Database, Section 1, Course Schedule

Any changes in the schedule will be sent to registered students through SJSU email 1 week earlier.

Tentative Course Calendar

Week	date	Item
1	August 24th	Lecture: Introduction to Big Data
2	August 29th	Lecture: Contd. + **Groups are formed & Project Review & Paper Review**
2	August 31st	Lecture: Hadoop Anatomy: HDFS + MapReduce Parallel Computing Model
3	Sept 5th	Lecture: Contd.
3	Sept 7th	Lecture: Contd.
4	Sept 12th	Lecture: Big Data Preprocessing: Data Cleaning Outliers, Integration, Reduction and Transformation + **HW1 is assigned**
4	Sept 14th	Lecture: Contd. + **Project Outline is due**
5	Sept 19th	Lecture: Online Analytical Processing (OLAP) + **Research Paper Outline is due**
5	Sept 21st	Lecture: Contd. + **HW1 is due**
6	Sept 26th	Lecture: Scalable Data Mining Algorithms: Frequent Itemsets and Mahout +
6	Sept 28th	Lecture: Contd.
7	Oct 3rd	Lecture: Contd.
7	Oct 5th	Lecture: Finding Similar Items: Locality Sensitive Hashing and Theory of Locality Sensitive Hashing + **Midterm Review**
8	Oct 10th	Lecture: Contd. + **HW2 is assigned**
8	Oct 12th	Lecture: **Midterm (Closed book)**
9	Oct 17th	Lecture: Mining Social Network Graphs
9	Oct 19th	Lecture: Contd. + **HW2 is due**

10	Oct 24th	Lecture: Dimensionality Reduction
10	Oct 26th	Lecture: Mining Data Streams
11	Oct 31st	Lecture: Contd. + Outliers
12	Nov 2nd	Lecture: SPARK Architecture, and YARN vs. Mesos
13	Nov 7th	Lecture: SPARK Architecture (Contd.) + Big Data K/V-based Model + **HW3 is assigned**
13	Nov 9th	Lecture: Big Data K/V-based Data Model: Hive, Pig, HBase (Contd.)
14	Nov 14th	Lecture: Big Data Document-based Data Model
14	Nov 21st	**Thanksgiving Holliday**
15	Nov 23rd	Lecture: Big Data Document-based Data Model (Contd.) + Scalability Model + **HW3 is due**
15	Nov 28th	Lecture: Scalability Models (Strong vs. Eventual Consistent Models) and Big Data Issues + **Research Paper is due**
16	Nov 30th	Lecture: Tradeoffs between SQL and NoSQL + **Final Project Report is due**
16	Dec 5th	Lecture: Course Review + **Final (Preview)**
17	Dec 7th	**Project Demo**
17	Dec 19th	**Final (Closed book) DH-450, Time: 7:15am – 9:30am**