The George Washington University
School of Engineering and Applied Science
Department of Computer Science
Department of Engineering and Systems Management

DBMS for Data Analytics

Professor: Joel Klein

Bulletin Course Description*:*

Study and design of database and data management systems for big data and data analytics; design of relational database systems and the SQL query language; NoSQL databases for unstructured data, including key-value, distributed table, graph databases, parallel processing databases. Prerequisite: EMSE 6574.

Prerequisites:

EMSE Students:
- EMSE 6574

CSCI 3907
- CSCI 2541/2441 and CSCI 3212

CSCI 6907
- Familiarity with python and CSCI 6212.

Or
- Instructor Approval

Required Text(s):

No Textbook is required, although online readings will be provided

Learning Outcomes:

- Learn the differences between SQL and NoSQL databases

- o SQL, Graph, Document, etc
- Understand what database best suits a problem and or available data
- Learn how to insert, update, and manipulate data in each type of database
- Experience the benefits of distributed vs singular database structures
- Gain insights into database design
  - o Limited to best practices, not DBA level
- Integration of databases with data analytic pipelines
  - o How to effectively utilize each type of database to fuel analytics
- Explore how cloud services are changing the way we think about databases

Attendance:

Regular class attendance is strongly encouraged. You will be held responsible for all the class discussions as well as the reading assignments. Here is the university policy:
https://registrar.gwu.edu/university-policies#attendance

Independent Learning:

In a **15**-week semester, including exam week, students are expected to spend a minimum of 100 minutes of out-of-class work for every **50** minutes of direct instruction, for a minimum total of **2.5** hours a week.

A **3**-credit course should include **2.5** hours of direct instruction and a minimum of **5** hours of independent learning or **7.5** hours per week. More information about GW's credit hour policy

can be found at: https://provost.gwu.edu/policies-procedures-and-guidelines and click on Assignments of Credit Hour Policy (PDF), or see the PDF pages (webpage);

https://provost.gwu.edu/files/downloads/Resources/Assignment-of-Credit-Hours_Final_Oct-2016.pdf

Class Schedule: [week-by-week]

Week	Topic(s) and readings

1/12	• Introduction to Course • Python Review • Introduction to Course Data • What is NoSQL
1/19	• Intro to Mongo (Document Database) ○ Discussion about Document Databases • Querying data in Mongo • Introduction to data manipulation in Mongo
1/26	• Mongo Cont.
2/2	• Intro to MySQL (SQL) ○ SQL vs NoSQL • Querying and Data Manipulation in MySQL • SQL vs Document Databases
2/9	• MySQL Cont.
2/16	• RDS and managed SQL • AWS Athena and S3
2/23	• Overflow
3/2	• Intro to Arango (Graph) • Querying and Data Manipulation in Graph Databases • Graph Visualizations • Graph vs SQL vs Document Databases
3/9	• Graph Cont.
3/16	• Spring Break (No Class)
3/23	• Intro to Cosmos (Document/SQL/Graph Database) • Advantage of Multi-Model DB's • Value of Managed DBs cont. • Sharding/Distributing Databases
3/30	• Cosmos Cont.
4/6	• Hadoop and Spark (Distributed Computation)
4/13	• Hadoop and Spark (Distributed Computation)
4/20	• Final Presentations

Assignments and Grades:

1. Assignments:

Assignment	Total Points
Mongo	10
MySQL	10
Graph	10
Final Project: Deliverable 01	10
Total Possible Points	**40**

2. Grading:

Final grades will be based on the standard grade scale.

Grade Percentages:
- Assignments – **40%**
    - Assignments submitted up to **1 week late** will earn a **maximum of 90/100**
    - Assignments submitted up to **2 weeks late** will earn a **maximum of 70/100**
- In Class Work – **10%**
- Final project – **50%**
    - Final Project Deliverables 02 & 03 – 20%
    - Final Project Presentation – 30%

Note: Any answers to assignments/exams should not be shared prior to receiving a grade for said assignment/exam.

Academic integrity:

Academic integrity is central to the learning and teaching process. Students are expected to conduct themselves in a manner that will contribute to the maintenance of academic integrity by making all reasonable efforts to prevent the occurrence of academic dishonesty. Academic dishonesty includes, but is not limited to, obtaining or giving aid on an examination, having unauthorized prior knowledge of an examination, doing work for another student, and plagiarism of all types. Ignorance is no excuse.

The number one problem that students run into with regards to academic integrity is plagiarism. It is not okay to copy, use, or otherwise exploit other people's ideas, words, or creations without giving them credit in the proper form. Sometimes this means you must use quotation marks;

while other times a simple source citation will do the trick. Changing a few words in a paraphrase is not enough to turn source material into "your own words" – in fact, that's a really bad idea to even try. Changing the phrasing order of sentences is not okay and using the thesaurus to find ways to change "happy" to "glad" is also a very bad idea. It is expected that students know how to correctly quote and cite material, and also how to write well. For those students who need assistance, the GWU Writing Center is available. Please see:

http://www.gwu.edu/~gwriter/ or
https://writingcenter.gwu.edu/

Academic Integrity Code:

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For the remainder of the code, see:

http://studentconduct.gwu.edu or

studentconduct.gwu.edu/code-academic-integrity