BAN 7990: Big Data Warehousing

Myles D. Garvey, Ph.D

Summer, 2020

E-mail: garveym2@wpunj.edu
Office Hours: By Appointment
Office: Valley Road Room 3070
Phone: 973-720-3928

Virtual Hours: Friday: 7:00pm - 8:00pm
In-Class Hours: Virtual
Class Room: Zoom

Contents

1 Required and Recommended Materials 2

2 Course Description 2

3 Prerequisites 2

4 Course Format 2

5 General Course Policies 2

6 Evaluation 4

7 Blackboard 4

8 R Problem Sets 4

9 Virtual Sessions 5

10 YouTube 5

10.1 Tentative Course Schedule 5

1 Required and Recommended Materials

Required Textbooks:

- Books posted to blackboard.

2 Course Description

This course will cover the fundamental knowledge needed to handle very large sets of data. This is an extension of our Data Mining and Data Warehousing course. You will learn about relational database systems, the SQL language, Hadoop, Spark, and Parallel Computing, among other concepts.

3 Prerequisites

You should have a solid understanding of Algebra. Calculus is required. I have posted notes in the event you have never taken Calculus, or you have forgotten it. You must also have taken Data Mining and Data Warehousing.

4 Course Format

This course will run online. We will have an asynchronous and a synchronous format. Every Monday night before 11:59pm, your lecture videos will be posted, as well as your assigned readings from the text book. In addition, you will have live sessions for up to one hour per week. The live sessions will comprise of QA sessions and review. MAKE SURE TO WRITE QUESTIONS DOWN AS YOU WATCH/READ!!!

5 General Course Policies

1. Please adhere to professional behavior in class. Refrain from chatting, reading the newspaper, answering phones, wearing headsets etc. Such behavior is disruptive and discourteous and WILL result in you being asked to leave for the remaining time of the class. I cannot be more clear on this. If this is a continued pattern, this will result in you receiving an F for the course. Please note that if you are carrying private conversations with other students, this also classifies as a disruption. Do not be surprised if I call you out on it. It is nothing personal, however, private conversations will result in a warning and a second violation will result in you being asked to kindly leave the classroom.

2. Important announcements will be made in class and on Blackboard. So please make sure you are attending class and checking Blackboard! I ask everyone to check their email/Blackboard a MINIMUM of 30 minutes before class in the event of a last minute cancellation.

3. Final course grades are final. Let me repeat this. **Final course grades are final!** Changes will only be made if there is a mistake in the calculation of the final grade, but legitimate evidence suggesting the contrary must be presented to the professor. "Legitimate" constitutes the use of the professor's calculation in grade mismatching with the grade received. See below for more detail. It does NOT include a mistake made on a particular assignment or exam or project. Please keep in mind that grades are NOT rounded. So if you receive a 89.99, this constitutes a B+, not an A. Do NOT request me to change a grade due to the closeness of a letter grade. I'm informing you right now, this will not happen! Same for other grading boundaries.

4. Accommodating students with special learning needs: In accordance with the university policy, students with documented sensory and/or other learning disabilities should inform the professor, so that their special needs may be accommodated. Please let me know IMMEDIATELY following the first lecture.

5. As you may know, it is against university policy to cheat. It is a very serious violation of academic integrity. Please note that if cheating of any kind is observed in/out of the class, you will be reported to a higher authority in accordance with university policy on academic dishonesty.

6. I do not give extra credit just because you are falling behind. Please do not request me to do so.

7. All course material is posted on Blackboard. Our videos are posted on YouTube.

8. It is YOUR responsibility, not mine, to keep track of your grades. With that said, ensure that you use the formula indicated below to get an idea of your standing in my course. The "Total Score" grades on Blackboard **do not** properly reflect your grades. In order to determine your grade in the course, you must use the equation indicated below in this syllabus. Failure to keep track of your own grade is not an excuse for additional points, extra credit or additional revision on assignments outside the grace period for review for said assignments. It also is not a valid reason to contest a final course grade. If you receive a grade of F at the end of the semester, please keep in mind that this is not reason for me to change a final grade due to your lack of supervision of your own grade. If you need me to clarify or project what your final grade will be, or give you a comment on your progress in the course, I will be happy to do so, but YOU MUST first initiate that type of discussion with me, and do so **well before the semester ends**.

9. If you are having difficulty due to a death in the family, financial problems, or other personal issues, I MUST receive an email from you **in advance**. I do not accept such requests after due dates.

10. ALL submissions of anything in this course are digital. Furthermore, ALL documents MUST be of PDF file format upon submission. **DOC, TXT, DOCX, etc will NOT be accepted as valid submissions**. I have a hard absolute policy with this. So please, remember, don't let your hard work result in a 0 all due to you not submitting a PDF. We live in the 21^{st} century, and ALL modern operating systems have the easy ability to convert any of the aforementioned types into a PDF format. HENCE, please make sure that submissions are

indeed in PDF format. Last, and most importantly, hard-submission (paper-format) WILL **NOT** be accepted as a valid form of submission.

6 Evaluation

R-Problem Sets (4 Total)	25% each

Your grade is calculated according to the following formula:

$$Final\ Grade = (0.15) \cdot (0.25) \cdot (RProblemSet_1 + RProblemSet_2 + RProblemSet_3 + RProblemSet_4)$$

Once your number grade is calculated, you can use the table below to determine your final letter grade in the course. I use a mathematical interval notation. So if your grade x is in the interval $[a, b)$ this is the same as saying $a \leq x < b$

Numerical Grade	Letter Grade
$[94, 100]$	A
$[90, 94)$	A-
$[87, 90)$	B+
$[83, 87)$	B
$[80, 83)$	B-
$[77, 80)$	C+
$[70, 77)$	C
$[68, 70)$	C-
$[65, 68)$	D
$[0, 65)$	F

7 Blackboard

All submissions for everything are conducted through blackboard. Make SURE you have access to this! Everything will equally be posted on blackboard. Another side-note, please DO NOT EMAIL ME via blackboard. If you need to email me, please do so DIRECTLY from your WPUNJ email account.

8 R Problem Sets

I will be posting R-Problem Sets. These are essentially mini-projects. **You are required to work on them individually**. You also have three-weeks to complete these, but they will be posted on virtual weeks. Each problem set comprises of a few R-based questions that are statistically driven. You will submit to be a .R file on Blackboard. **Please comment your code!**. Failure to comment will result in a loss of points.

9 Virtual Sessions

Every week, we will "meet" live via Zoom. Our "online" portion will entail you asking questions live, as well as working through problems. It is intended to supplement the recorded sessions.

10 YouTube

Every class, make up class, extra course material and in some instances assignment solutions are recorded live using YouTube Live. While I offer this convenience to you, please do not be opportunistic of it. **You still are expected to attend class!**

10.1 Tentative Course Schedule

- Introduction to Data and Databases
- The SQL Language
- Interacting with a Database via R
- Amazon Web Services and the Cloud
- Introduction to Parallel Computing
- Working on a Hadoop Cluster
- Integrating Spark into the mix