**44517: Exam 2 Study Guide**

1. Understand the 5Vs of big data (Volume, Velocity, Variety, Veracity, and Value).
2. Identify which V is best suited for a given big data application.
3. Identify which V is best suited for a given tool or technology.
4. Comprehend the source code of Kafka consumers and producers.
5. Identify keys and values of a Consumer record given the source code of Kafka.
6. Identify the topic and content published to the Kafka broker given the source code of Kafka.
7. Write PySpark queries to solve simple problems (Use Marci's dashboard as a reference.)
8. Understand the different input file formats that can be used for performing batch operations using PySpark
9. Predict the output given source code in Java streams.
10. Write Java source code for basic stream operations similar to those discussed in class.
11. Understand how to use **IntStream.of(), Stream.of()**, and other methods discussed in class.
12. Understand how performance varies when executing streams in sequential and parallel mode.