

## Preparing Data for Analytics (IAF 603) Course Syllabus

Fall 2019

- **Credit Hours:** 3 credits
- **Class Hours and Location:** W 6 – 8:50 pm, Bryan 112
- **Instructor:** Dr. Regis Kopper (kopper@uncg.edu)
- **Office hours:** F 1-4:00 pm (Petty 160)
- **Description**

This course explores the topics of data volume, variety, velocity, and veracity as related to complex datasets. Students in the course will understand the key concepts in Big Data technologies as related to data access/collection, transformation, development of databases, issues with data, cleaning/normalization, efficient query of large-scale datasets, and models for data stewardship as pertaining to governance, sharing, and reuse.
- **Outline**

The course topics include data sources and collection, data cleaning, data transformation, preliminary data analysis, knowing the data and building knowledgebases, and heterogenous data. In addition, students will receive an introduction to Big Data analytics, get familiar with operations on Big Data, database query optimization and indexing, Hadoop and Apache Spark, dealing with structured and unstructured data, and automation of data preparation and data flows. Finally, students will gain knowledge about data stewardship, sharing and governance.
- **Materials**

There is no required textbook for the course. The instructor will provide lecture notes and suggested and required readings on Canvas. The following books touch on several topics of the course and are suggested for deeper understanding:

  - 1) Cielen, Davy, Arno Meysman, and Mohamed Ali. *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications Co., 2016.
  - 2) Kleppmann, Martin. *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. O'Reilly Media, Inc., 2017.
- **Grading**

- Class Participation:	5%
- Homework Assignments (3):	15%
- Class Project:	30%
- Midterm:	25%
- Final Exam:	25%
- **Grading scale**

- A :	94 - 100%
- A-:	90 - 93%
- B+:	87 - 89%
- B :	84 - 86%
- B-:	80 - 83%
- C+:	77 - 79%
- C :	74 - 76%
- F :	< 74%

- **Policies**

- **Attendance:** Students are expected to participate in class discussions. Many class activities will be interactive, and attendance is expected.
- **Due Dates:** Late assignments will be penalized 20 percent per day (per 24 hours).
- **Missed Exams:** Make-up exams are not generally arranged unless emergency approved by the instructor. Only if the student's absence is approved, it will be scheduled by the instructor.
- **Academic Integrity:** All work including assignments and exams is subject to the UNCG Academic Integrity Policy (<https://osrr.uncg.edu/academic-integrity/>). By submitting their assignments and exams, students are implicitly agreeing to this policy. Academic dishonesty is not acceptable and is subject to official sanctions. That is, the incident(s) will be reported to the department and it may result in zero point to the work and even failure in the course.
- **Academic Accommodations:** If you have disability-related requirements, please contact the Office of Accessibility Resources and Services (OARS) at <https://ods.uncg.edu/>.
- **Health and Wellness:** Student Health Services and The Counseling Center (<https://shs.uncg.edu/>) can help with health and wellness issues you may be experiencing (e.g., physical ailments, illnesses, strained relationships, anxiety, high levels of stress, alcohol/drug problems, feeling down, or loss of motivation).
- **Emergency University Closure:** Our course schedule follows the university's instructions in case of emergency such as bad weather.
- **Disruptive Behavior:** Upon any non-course-related or disruptive activities, the instructor may ask the student to leave the class room and count it as absence. If such behavior is continued, it may result in dropping the student from the course based on the UNCG Disruptive Behavior Policy (<https://osrr.uncg.edu/faculty/disruptive-behavior/>)
- **Copyright:** The course materials including lectures are provided only for this course. Redistribution of them is not recommended. Any commercial gain by such behavior is a violation of the University's Copyright Policy and of the Student Code of Conduct.

- **Course Schedule:** The detailed course schedule will be posted and updated on **CANVAS**. Students should check up-to-date schedule on **CANVAS**.

Week	Topic
1	Introduction and Data Sources
2	Data Collection and Introduction to Data Cleaning
3	Data Cleaning
4	Data Transformation: Motivation and Normalization
5	Data Transformation: Data Reduction, Handling Imbalanced Data
6	Preliminary Data Analysis
7	Knowing the Data and Building Knowledgebases
8	Heterogenous Data
9	Big Data Analytics: Database Operations
10	Hadoop and Apache Spark
11	Dealing with Structured and Unstructured Data
12	Automated Data Preparation and Data Flows
13	Thanksgiving Break
14	Data Stewardship, Sharing and Governance