



NORTHWEST
MISSOURI STATE UNIVERSITY

MARYVILLE | KANSAS CITY

School of Computer Science and Information Systems

44-517-02/03: Big Data (3 hrs.)

Fall 2024

Instructor: Dr. Ajay Bandi

Email: ajay@nwmissouri.edu

Office hrs: MW – 9:30 am to 10:00 am
MWF – 2pm to 4:00 pm

Office: CH 2250

Prerequisites:

Undergraduate: 44-242 with a grade of C or better.

Graduate: 44-542 with a grade of C or better, or consent of instructor

Textbook and supplementary materials: *Resources will be provided by the instructor as necessary through the semester.*

Course description: An introduction to the design of data-intensive, reliable, scalable, and maintainable systems. Includes an introduction to current and relevant tools, technologies, design principles, and frameworks. This may include concepts such as parallel programming, distributed computing, distributed file systems, MapReduce, regular expressions, and the ingesting and processing of data at rest and data in motion. Tools used may include Hadoop, HDFS, Pig, Hive, Spark, Storm, Kafka, Mahout, MLlib, etc.,

Student learning outcomes: *This course satisfies the following outcomes for the School of CSIS*

Outcome	Assessment Methods
DSI students will access, generate, and reorganize information using contemporary technologies.	Selected assignment(s)
DSI Students will work as a team to design, implement, and deliver solutions to problems using best practices with contemporary technologies.	Selected assignment(s)

In addition, after completing this course, a student should be able to:

- Describe the characteristics of big data systems.
- Describe the history of distributed data processing including Hadoop, HDFS, and MapReduce
- Describe big data integration platforms (e.g., Apache Beam).
- Compare choices for data processing engines.

Refer to Syllabus Addendum for additional information.

- Describe key processing concepts.
- Describe common options for programming languages.
- Describe common options for I/O connectors.
- Create and execute big data jobs using contemporary technologies.
- Function effectively on teams to accomplish a common goal.

Instructional methods: Class time will consist primarily of lecture and student discussion about course topics, with time allotted for individual or group work on projects, homework, and hands on activities.

Graded course requirements: *Tentatively, the grade will be weighted as follows:*

Category	Weight
Exam 1	10%
Exam 2	10%
Discussions	25%
Assignments and In Class Activities	25%
Final Project/Exam	30%

Final Exam: The class will meet during the final exam time as specified by the registrar's office at <https://www.nwmissouri.edu/registrar/finals.htm>

Grading scale: The grading scale is the standard 10-point grading scale but may be revised in the students' favor at the instructor's discretion.

Graduate Credit (44-517)	
Percent Range	Grade
> 90%	A
80%-89.9%	B
70%-79.9%	C
60%-69.9%	D
<60%	F

Undergraduate students will receive a 2% boost to their grade at the end of the semester.

Course outline/major topics studied:

- Introduction to Big Data
- Structured, Semi-structured, and Unstructured Data
- Hadoop
- Distributed Database Architectures
- Data Engineering
- Python for Data Science

Refer to Syllabus Addendum for additional information.

- Map Reduce
- Selected Java topics
- PySpark
- Apache Beam
- Apache Kafka

Note: Course schedule is subject to change with instructor notification and students will be responsible for abiding by these changes.

Artificial Intelligence Engines: Generative AI engines are fast becoming important tools to help improve various personal, professional, and educational tasks. Specifically for this course, the submission of academic work created by a generative AI engine is not allowed. All submitted coursework must be your own work. The goal is for you to learn and understand the course content. AI engines are a tool which can help facilitate the learning process. The understanding of course content is assessed by completing the required coursework. Your understanding cannot be gauged if AI-generated material is used.

Late Submission Policy for assignments: Each assigned, graded activity will have a due date posted and is expected to be completed by the due date. All the submissions must be submitted through Northwest Online. Students with incorrect or no submissions will receive zero credit.

Attendance: Students are expected to attend all classes as specified in the course syllabi for each course and are responsible for all material and work assigned during class lectures and hands on learning time. It is the responsibility of the student to promptly notify their instructor when unable to attend class. Please refer to the university policy on attendance at <https://www.nwmissouri.edu/policies/academics/Attendance.pdf>