

44-517



44-517 Course Syllabus

Area

School of Computer Science and Information Systems

Course Title

44-517 Big Data

Course Credit

3 hours

Placement in Curriculum

This course is typically offered in the later years of an undergraduate degree or after the first semester of a graduate program.

Prerequisites

Undergraduate prerequisites: MATH 17230 or MATH 17316 with a grade of C or better and CSIS 44242 with a grade of C or better. Graduate prerequisites: CSIS 44542 with a grade of B or better, or concurrent enrollment in CSIS 44542, or consent of instructor.

44-517

MWF 2-2:50pm CH 3400

[\(map\)](#)

Course Professor

Dr. Denise Case

Associate Professor

660.562.1136

dcase@nwmissouri.edu

Colden Hall 2280

Office Hours [\(link\)](#)

MW 3-3:50 PM

MW 4-4:50 PM

Tu 12:30-1:20 PM

Tu 1:30-2:20 PM

Tu 2:30-3:20 PM

Course Description

An introduction to the design of data-intensive, reliable, scalable, and maintainable systems. 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.

Course Rationale

This course involves an overview the design and implementation of big data solutions covering common approaches to processing big data at rest and data in motion.

Student Learning Outcomes

Competency	BS Data Science Program Outcome	Assessment
------------	---------------------------------	------------

44-517

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

Additional student learning outcomes include:

- Describe the characteristics of data-intensive systems.
- Describe key sources of big data.
- Describe the Vs of big data and identify what are (and are not) big data problems.
- Describe the Hadoop framework, HDFS file system, and MapReduce programming model.
- Design and execute a process using big data tools such as Hadoop & Spark.
- Design and implement MapReduce solutions.
- Describe distributed systems, their challenges and benefits, and apply concepts of good design to increasingly complex systems.
- Describe different ways to represent, store, manage, and process data.
- Describe different architectures and their strengths, weaknesses, and application areas.
- Describe current and relevant tools for data-intensive systems.
- Understand and employ basic regular expressions.
- Understand and employ basic shell commands.
- Implement big data solutions using standard algorithms.
- Implement data intensive systems using relational and nonrelational databases.

44-517

Materials

Recommended references

- [Mining of Massive Datasets](#) by Leskovec, Rajaraman, and Ullman, Third Edition, 2019.
- [Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems](#) by Martin Kleppmann, 2017 (preview).
- [Big Data](#) by Nathan Marz with James Warren, 2015.
- [Data-Intensive Text Processing with MapReduce](#) by Jimmy Lin and Chris Dyer, 2010.
- [MapReduce: Simplified Data Processing on Large Clusters](#) by Jeffrey Dean and Sanjay Ghemawat, 2004.
- [Cornell Java reference \(online book\)](#) by Professor David Gries.

Required

Students must have access to the following at every course meeting:

- A **bound notebook with pencil/pen** for taking notes and submitting written content (e.g., pop quizzes)
- Their campus-assigned laptop, in working order, with all required software
- Free [Git](#) distributed version control system
- Free [TortoiseGit](#) for integrating Git with Windows File Explorer
- Free [PuTTY](#) for creating SSH public/private key pairs
- Free [BitBucket](#) and [GitHub](#) educational accounts
- Free [GitHub Education Pack](#) (as needed)
- Free [Chocolatey](#) package manager for Windows
- Free [Notepad++](#) text editor
- Free [Visual Studio Code](#) integrated development environment

44-517

- Free, open-source Apache Kafka, Spark, Storm, Flume, Zookeeper, Hive, Pig and other tools and libraries as directed by the instructor.
- Access to our Proxmox Virtual Environment
- Access to free cloud accounts including [Amazon Web Services \(AWS\)](#), [Google Cloud Platform \(GCP\)](#), [IBM Cloud](#), [Microsoft Azure](#), and [Oracle Cloud](#) as required.
- Typing is a foundational skill for developing software systems. If additional practice would be helpful, try <https://www.typing.com/>.

Course Outline

The course is divided into seven content modules.

Module	Content
1	Course script intro, big data, shell processing
2	Hadoop HDFS and MapReduce, ecosystem, Pig, Hive, Impala, Hue
3	ZooKeeper, Kafka, Java apps
4	Python, map-reduce, architectures
5	Apache Spark
6	Demonstrations and peer review
7	Optional certifications or Splunk, final reflection

Course Schedule

The course spans 17 calendar weeks.

Wk	Start	Notes
----	-------	-------

44-517

1	Wed 8-19	Module 1 begins, assignments due first day
2	Mon 8-24	Module 1 continues
3	Mon 8-31	Module 1 continues
4	Mon 9-7	No class Monday, Module 2 begins
5	Mon 9-14	Module 2 continues
6	Mon 9-21	Module 3 begins
7	Mon 9-28	Module 3 continues
8	Mon 10-5	Module 4 begins, no class Friday (walkout)
9	Mon 10-12	Module 4 continues
10	Mon 10-19	Module 5 begins
11	Mon 10-26	Module 5 continues
12	Mon 11-2	Module 6 begins
13	Mon 11-9	Module 6 continues
14	Mon 11-16	Module 7 begins
15	Mon 11-23	Thanksgiving break (no class Wed-Thr-Fri)
16	Mon 1-30	Online prep week (no new content, no in-person classes)
17	Mon 12-7	Online finals week (no late submissions)

Schedule Subject to Change

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

44-517

SCHEDULE LINKS

- [Academic Calendar](#)
- [Final Exams Schedule](#)
- [University Events](#)

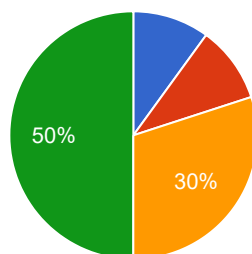
Instructional Methods and Techniques

Instructional methods include lectures, class discussions, individual work, small group work, learner presentations, discussions, guest speakers, and collaborative development.

Grading and Evaluation

Graded contributions

Category	Contribution
Engage	10%
Demo	10%
Quizzes	30%
Projects	50%
All	100%



- Engage
- Demo
- Quizzes
- Projects

Grading scale

Credit	
Percent Range	Grade
90-100%	A
$\geq 80\%$ and $< 90\%$	B
$\geq 70\%$ and $< 80\%$	C
$\geq 60\%$ and $< 70\%$	D
below 60%	F

Course Expectations

44-517

Course Email

Professional email practices are required.

1. Email must include course number and section (e.g., **44-517-01**) in the subject line.
2. Email should be professional, politely written, and use proper grammar, capitalization, and spelling.

Course Conduct Guidelines

1. Students will behave in a professional manner in all course interactions.
2. Devices including cell phones, ear phones, gadgets, and other items should be silent and put away during class.
3. Unauthorized use of devices during class may result in reduction in scores.
4. Be ready to work when the class is scheduled to begin.
5. Be polite and respectful to others in the class. Do not interrupt someone during a class discussion, and respect others' opinions/comments as you expect them to respect yours.

Course Citation Requirements

1. **Cite ALL Sources.** You must provide proper citation for any content including text, images, videos, and code that you did not personally develop.
2. Do not copy and paste content for any submission.
3. Failing to acknowledge assistance including internet photos, clipart, slide content, course files, shared solutions, etc. can be expected to result in a 0 on the assignment, a one-half letter grade reduction, a 0 in the course, or failing the course with an academic integrity violation.

Course Late Work Policy

Graded activities are expected to be completed by due date. If you must miss a due date, contact the instructor in advance or as early as possible. Due dates get you ready for coming material. Keeping up is critical. On rare occasions, special challenges or obligations may make it difficult to submit on time. You typically have a one-week grace period where the work may still be submitted and graded. Late penalties will be described in the rubric. No late submissions will be accepted for the final submission.

44-517

during class, in discussions, or in a recorded submission.

Course Grading

Grades will generally be posted within 2-7 days of the due date. If you wish to petition for additional points, please wait 24 hours. Petitions must be emailed, and must include the specific changes you are requesting, along with your reasoning on why the higher score is warranted. Challenges must be raised within seven days of the grade being posted; after seven days, grades are considered final.

Course Access Ends - Save Your Content

The course site will not be available after the end of the semester. Save the syllabus and any materials you wish to access after the course, you may need to submit this to a potential employer or another university in the future. We do not provide materials after completion of the course.

Grade of Incomplete (I)

Following Northwest Missouri State University policy, incomplete grades may only be given in extreme circumstances, such as illness, death in a student's immediate family, or similar circumstances beyond a student's control, and the student was unable to complete a significant portion of the course work.

University Attendance Recommendations

Students are expected to attend all classes as specified in the course syllabi for each course. It is the responsibility of the student to promptly notify their instructor when unable to attend class. Students receiving veterans' benefits should consult with the coordinator of Veterans' Affairs for the additional attendance requirements. Please review the [university policy on attendance](#).

Excused Absences

Excused absences include attendance at a university sponsored event (documented with an excuse signed by the university sponsor prior to the event) or by circumstances considered extenuating by the course instructor. A student may make up class work without penalty for excused absences.

Final Exams

44-517

time and the faculty member's considered judgment. Final exams will not be rescheduled, and a grade of "I" will not be given as a result of an institutional cancellation of a final examination.

Administrative Drop

An instructor may request the Office of the Registrar delete a student from a course roster if the student has not met the prerequisite for the course as stated in the catalog, or as a result of non-attendance in the course.

Academic Administrative Withdrawal

When it is in the best interest of Northwest Missouri State University for a student to withdraw, a student will be given a W, put on administrative hold, and given notice that they are about to be withdrawn. This action will result in removal of all credits associated with courses that have yet to be completed in the semester in question and the student will be administratively withdrawn from the University. An Administrative Withdrawal does not affect the student's grade point average. Please refer to Northwest Missouri State University's [Academic Administrative Withdrawal Policy](#).

Course Access

Course content is provided via Canvas and Northwest Online. Access to Northwest Online is at <https://nwmissouri.instructure.com/>.

University Communications

Students are expected to use their Northwest student email account for any electronic correspondence within the university. Students are also strongly advised to check their email and CatPAWS accounts on a regular basis.

Academic Integrity Policy

The students, faculty, and staff at Northwest endeavor to sustain an environment that values honesty in academic work, that acknowledges the authorized aid provided by and intellectual contributions of others, and that enables equitable student evaluation. Please refer to [Northwest Missouri State University's Academic Integrity Policy](#). In addition, students are expected to understand and abide by the [Computer Science and Information Systems Academic Integrity Policy](#)

44-517

Learning or Living Accommodations Request Process

Northwest Missouri State University complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990 (ADA) and the ADA Amendments Act of 2008 (ADAAA). If a student has a documented disability that qualifies under ADA/ADAAA and requests accommodations, they should review the Accessibility and Accommodations webpage at <https://www.nwmissouri.edu/titleixequity/accessibility/index.htm> for guidance, including the accommodations application and supporting documentation requirements. Contact ADA@nwmissouri.edu or 660.562.1873 for further assistance. For the university policy on disability accommodation refer to <http://www.nwmissouri.edu/policies/student/Disability-Accommodation.pdf>.

Non-Discrimination and Anti-Harrassment Policy

Northwest Missouri State University is committed to maintaining an environment for all faculty, staff, students, and third parties that is free of illegal discrimination and harassment. Please refer to the Non-Discrimination and Anti-Harassment Policy at <http://www.nwmissouri.edu/diversity/titlevi.htm>.

Family Educational Rights and Privacy Act (FERPA)

Family Educational Rights and Privacy Act of 1974, as amended (commonly known as the Buckley Amendment), is a federal law which provides that colleges and universities will maintain the confidentiality of student education records. Please refer to the [Family Educational Rights and Privacy Act \(FERPA\) Policy](#).

COVID-19 Classroom Mitigation

Northwest is committed to the health and safety of the University community and has therefore adopted COVID-19 mitigation policies. Every student must wear a face covering (such as a cloth facemask, bandana, scarf, neck gaiter, or medical mask) over their nose and mouth at all times in all academic building spaces, including classrooms (unless directed not to by the instructor), offices, hallways, and restrooms. Face shields may be worn in addition to, but not in place of, a face covering. Students without face coverings will not be allowed in the classroom until they comply with expectations. Students must also follow directions regarding entries, exits, and furniture, and maintain at least 6 feet of social distancing whenever possible. Northwest further asks all students

44-517

communicate with their instructors. Students who do not comply with these requirements will be subject to standard disciplinary procedures according to the Northwest Student Code of Conduct (i.e. verbal and written warnings followed by a hearing, if necessary). We thank you for doing your part to maintain our learning environment and to protect the health of fellow Bearcats.

Change in Course Delivery

It is our goal as a University to continue all courses as planned on campus. However, due to unforeseeable impacts of COVID-19, faculty and students must be prepared to move all courses to a remote/online learning format anytime during the semester, either permanently or for a short term. The type of format (synchronous or asynchronous) will be at the discretion of each faculty member. The University and faculty will communicate with students in the event such action is deemed necessary to preserve the health and safety of students and employees.

Course Evaluation

At the end of this course, you will be encouraged to complete a course evaluation.

Syllabus Subject to Change

The syllabus is subject to change with instructor notification and students will be responsible for abiding by these changes.

[Denise Case](#) ◆ [Site](#) ◆

★ [SOURCE CODE ON GITHUB](#) 