

Processing Big Data - CS 131

Fall 2026 - Section 03

Contact Information

Instructor: Robert Nicholson

Email: robert.nicholson@sjsu.edu

Class Days/Time: T/Th 12:00-1:15

Classroom: Science 311

Office Hours: Wednesday, 2:00pm-5:00 PM

via Zoom: <https://sjsu.zoom.us/j/8791841945>

(other times by appointment)

Course Description and Requisites

In-depth study of essential tools and techniques for processing big data over the UNIX operating system and/or other operating systems. On UNIX, it includes using grep, sed, awk, join, and programming advanced shell scripts for manipulating big data.

Prerequisite(s): CS 46B with a grade of “C-” or better; or CS 22B with a grade of “C-” or better AND graduate standing. Allowed Declared Majors: Computer Science BS, Data Science BS, MS Bioinformatics (MSBI).

Grading: Letter Graded

Classroom Protocols

Students will be dropped from the class by the instructor (and will not be given ADD codes) for either one of the following reasons:

- absence for 1st day of class without informing you before 2nd day of class
- lack of prerequisites.

Cheating will not be tolerable; a ZERO will be given to any cheated assignment/exams, and it will be reported to the Department and the University.

Do NOT share/post online any course materials, PPT slides, or homework solutions.

Laptops that are capable of running the Lockdown Browser are required for all exams and quizzes and should be brought to each class. No other electronic devices are allowed during exams and quizzes.

You are required to check Canvas for reading/assignments.

The information on this syllabus is subject to change; changes, if any, will be clearly explained in class, and it is your responsibility to become aware of them.

Attendance

University policy F69-24 at <http://www.sjsu.edu/senate/docs/F69-24.pdf> states that students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class.

I will not take attendance or grade on attendance. However, I will give periodic “pop quizzes” during class time. There will be no makeup for these quizzes unless you have a verifiable medical emergency.

Students should bring a laptop capable of running the Lockdown Browser to every class.

Consent for Recording of Class and Public Sharing of Instructor Material:

University Policy S12-7, <http://www.sjsu.edu/senate/docs/S12-7.pdf>, requires students to obtain instructor's permission to record the course: Common courtesy and professional behavior dictate that you notify someone when you are recording him/her. You must obtain the instructor's permission to make audio or video recordings in this class. Such permission allows the recordings to be used for your private, study purposes only. The recordings are the intellectual property of the instructor; you have not been given any rights to reproduce or distribute the material. Course material cannot be shared publicly without his/her approval. You may not publicly share or upload instructor generated material for this course such as exam questions, lecture notes, or homework solutions without instructor's consent.

AI Use Policy

AI tools (e.g., ChatGPT, Copilot, Gemini) can be helpful for learning, brainstorming, and debugging. In this course, AI use is permitted in limited ways, but your submitted work must reflect your own understanding and original effort.

You may use AI tools to get ideas, debug, or get unstuck. Any AI use must be disclosed in your submission (tool + how you used it). You may not copy/paste AI output into your reports or submit AI-generated solutions (text or code) as your own. All homework and project submissions must be your original work, and you must be able to explain your answers and code.

If you cannot sufficiently explain your submitted work, the penalty may include zero points on the assignment and a report to the Office of Student Conduct and Ethical Development.

Example

AI Use Disclosure: I used ChatGPT to help debug a Spark join issue and to brainstorm an approach for handling missing values. I did not copy code or text from the tool; I used the suggestions to guide my own implementation and wrote the final solution myself.

Program Information

Diversity Statement: At SJSU, it is important to create a safe learning environment where we can explore, learn, and grow together. We strive to build a diverse, equitable, inclusive culture that values, encourages, and supports students from all backgrounds and experiences.

Course Learning Outcomes (CLOs)

Upon successful completion of this course, students will be able to:

CLO 1 - Analyze, manipulate and process large-scale data with the UNIX/Linux command line and other operating systems.

CLO 2 - Develop shell scripts for use in data-intensive applications.

CLO 3 - Build data analysis pipelines, automate tasks, make analyses reproducible and shareable.

CLO 4 - Compare data analysis on the command line with use of graphical user interface and web-based tools.

CLO 5 - Solve big data challenges with the UNIX/Linux shell and command-line tools.

CLO 6 - Apply data science solutions to datasets from example domains, such as biology, business, and finance.

CLO 7 - Perform big data analysis efficiently, document and reproduce analysis, use cloud computing for data intensive problems.

You will follow industry-standard best practices and use software development tools that are common in today's software industry.

You will develop the *critical job skill* of working in a small project team to successfully develop a software application that uses shared interfaces and data formats.

Course Materials

There is no required text for this course. The following books are recommended:

UNIX Command Line: A Complete Introduction

Author: William Shotts Jr.

Publisher: No Starch Press

[Download it from the author's webpage (<https://linuxcommand.org/tlcl.php>)]

Data Science at the Command Line, 2nd Edition

Author: Jeroen Janssens

Publisher: O'Reilly Media, Inc

ISBN: 978149208791

[The book is available for free through the SJSU library

<https://learning.oreilly.com/library/view/data-science-at/9781492087908/>

(<https://learning.oreilly.com/library/view/data-science-at/9781492087908/>) .]

Course Requirements and Assignments

The course is delivered in person.

All students are required to have access to a wireless laptop (running OSX, Windows, or some version of LINUX), with a camera and microphone, upon which you can install the required software.

The technology used will include Canvas, programming in Java, and an IDE (Integrated Development Environment)

Course Requirements and Assignments

This course is designed to be hands-on and cumulative, preparing students to process, manipulate, and scale data using real-world tools. Success in this course requires regular engagement, thoughtful experimentation, and a consistent time investment both inside and outside the classroom

During the second part of the semester, there will be a team project that combines techniques covered during the course to solve a significant real-world problem.

Each team will submit its assignment and project into Canvas, which will display the scoring rubrics for grading.

No late submission of assignments will be accepted except for the verified emergency such as doctor's notes or family death certificates.

In general, all team members will receive the same score on each assignment. However, a small number of points may be awarded or deducted at the end of the semester based on feedback from your fellow team members.

If there are issues with team member conflicts or lack of contributions during the semester, contact the instructor as early as possible.

Midterm Exam

The Midterm will only be given during class time.

Makeup midterm exams will only be given in cases of verifiable emergency or illness.

The midterm exam date in this syllabus is approximate and subject to change.

Final Exam

The final exam will be cumulative (it will cover the entire course).

Makeup exams are only given if there is a verifiable emergency or illness.

Quizzes

There will be quizzes throughout the semester. The quizzes are designed to help students stay on top of the material and illustrate areas of confusion for both students and the instructor

Technology

Students are required to have laptop computer, and to bring the computer to each class. The laptop must be capable of running the **Lockdown Browser**, which is used for quizzes and exams.

If you do not have access to a suitable laptop, SJSU has a free equipment loan program available. You will need a reliable WIFI connection to attend class.

If you run into issues with technology or WIFI, please reach out to the instructor.

Grading Information

Final grades will not be adjusted in any way. No incomplete grades will be given. No late submission of assignments will be accepted except for the verified emergency such as doctor's notes or family death certificates.

Breakdown

Assignments (25%)

Semester Project (30%)

Pop Quizzes (10%)

Midterm (15%)

Final (20%)

Grade Criteria

Course grades will be based on a curve. Per CS Department policy, the median total score will earn a B–. Approximately one third of the class will earn higher grades, and another one third will earn lower grades.

University Policies

Per University Policy S16-9 (PDF) (<http://www.sjsu.edu/senate/docs/S16-9.pdf>), relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on the Syllabus Information (<https://www.sjsu.edu/curriculum/courses/syllabus-info.php>) web page.

Make sure to visit this page to review and be aware of these university policies and resources.

Key Dates

Midterm	Tuesday	10/13/26
Final Project Due	Thursday	12/3/26
Final Exam	Tuesday	12/15/26, 10:45 AM-12:45 PM