ITP 419 – Advanced Cloud Analytics
Units: 4
Spring 2025

Date/Time: Tue, Thu 5:30-7:20pm
Location: ZHS252

Instructor: Patrick Alexander
Contact Info: patrick.alexander@usc.edu
Office Hours: TBD

Learning Assistant: TBD

IT Help:
USC IT (ITS): https://itservices.usc.edu/contact/

Course Description
This course provides students with a deep understanding of the tasks, tools, and strategies used in Data Engineering and Machine Learning. Students will gain hands-on experience in collecting, storing, preparing, analyzing, and visualizing data for analytics and ML applications. Through exploration of real-world use cases and a guided, hands-on approach, students will learn to make informed decisions in building data pipelines, select and apply machine learning services to address business challenges, and label, build, train, and deploy custom machine learning models. This course equips students with the skills needed to harness the power of cloud analytics for their specific applications.

Who Should Take This Course
The course is designed for those who have a foundational understanding of analytics and machine learning and are looking to delve deeper into the intricacies of data engineering and machine learning in the cloud environment.
- **Aspiring Data Engineers and Data Scientists**: Individuals looking to enter the field of data engineering or data science and want a comprehensive understanding of both data pipelines and machine learning applications.
- **Current IT Professionals**: Those already in the tech industry, such as software developers, system architects, or database administrators, who want to expand their skill set into data engineering and machine learning.
- **Business Analysts and Decision Makers**: Professionals who want to understand the technical aspects of data analytics and machine learning to make informed decisions for their organizations.
- Students in Computer Science or Related Fields: University or college students studying computer science, information technology, or related fields who want to specialize in data analytics or machine learning.
- **Tech Enthusiasts**: Individuals with a keen interest in the latest trends in cloud analytics, data engineering, and machine learning, even if they don't come from a strictly technical background.
- **Organizations and Teams**: Companies or teams that are looking to upskill their members in the areas of data analytics and machine learning to harness the power of data more effectively.

Learning Objectives

The primary objective of the "Advanced Cloud Analytics" course is to equip students with the skills and knowledge to effectively harness data through engineering and machine learning in the cloud. The academic rationale lies in the increasing demand for professionals adept at transforming vast data sets into

actionable insights. As organizations gravitate towards data-driven decision-making, there's a growing need for expertise in managing, analyzing, and visualizing data efficiently and securely in the cloud. This course bridges the gap, preparing students to meet industry demands, fostering innovation, and driving organizational efficiency through informed, data-backed strategies.

- Understand the significance of data science in data-driven organizations and the foundational principles of machine learning.
- Design and implement scalable and secure data pipelines, considering elements like data sources, infrastructure, and cloud services.
- Address challenges in data collection, including scalability, cost, fault tolerance, and latency.
- Choose appropriate data storage solutions tailored to specific analytics and machine learning use cases.
- Process diverse data formats, leveraging concepts like MapReduce for optimized data workflows.
- Recognize and mitigate risks, ensuring data security and governance throughout the data and ML pipelines.
- Differentiate between general data analytics pipelines and machine learning-specific workflows.
- Utilize cloud tools and platforms for effective data analysis, model training, and prediction.
- Craft tailored data visualization strategies and understand the application of ML in areas like forecasting, vision, and language processing.
- Engage in hands-on exercises, applying both data engineering and machine learning concepts to real-world scenarios.

Prerequisite(s): ITP 349 – Enterprise Data Analytics (formerly ITP 487).

Course Notes
All course materials will be made available through Brightspace. These include:
- Lecture slides
- In-class exercises
- Homework Assignments
- Grades and feedback
- Office hours
- Online discussion forums will be used for out-of-class discussions

Announcements made in class and content posted on the course management tool will supersede the contents of this syllabus.

USC Technology Support Links
Zoom information for students
Software available to USC Campus

Technological Proficiency and Hardware/Software Required
The assignments for this class will include both reading assignments as well as hands-on computer assignments. Students must bring their laptop computers (phones/tablets are not sufficient) to lecture sessions to participate in hands-on activities. Students will be given tutorials to gain familiarity with software tools.

- Amazon Web Services Lab (you will build your own)
- Python/Pandas/Numpy/Matplotlib
- SQL
- Amazon Athena/Redshift/EMR/Sagemaker/Forecast/Rekognition
- Apache Hadoop/Spark
- Amazon IoT Analytics and QuickSight

- Cloud technologies will vary by semester depending on industry trends and innovations

Readings and Supplementary Materials

Required reading and supplementary materials listed in the weekly breakdown section of this syllabus, additional materials will be announced in class and published online.
- "Designing Data-Intensive Applications" by Martin Kleppmann (ISBN: 1449373321)
- "Machine Learning Design Patterns" by Valliappa Lakshmanan, Sara Robinson, and Michael Munn (ISBN: 1098115783)

Description and Assessment of Assignments

Homework: Most homework is computer based. Homework should be turned into the course management tool on time. Grading will be based on completeness, accuracy, and timeliness. Feedback will be provided through course management tool. These are individual effort assignments.

In-Class Exercises: are guided Q&A and hands-on exercises that are used to spark additional discussion and deeper understanding of the materials and concepts before the student leaves the class. Announcement of in-class exercises may or may not be given prior to the class. In-class exercises can be a team or individual exercises. The score used for grading is the percentage of in-class exercises completed and turned in in-class vs what was assigned in the semester.

Exams: will be online using the course management tool.

Grading Breakdown

Homework	30%
In-Class Exercises	10%
Midterm	30%
Final Exam	30%
TOTAL	100%

Grading Timeline

Grading will typically be completed 7 days after submission. Any variations will be announced in class or on Brightspace. ***Regrade requests must be submitted within a week of the grades being published unless otherwise communicated in class.***

Generative AI Policy

Use of Generative AI technologies, including ChatGPT, are encouraged and allowed unless explicitly stated otherwise. ***YOU MUST CITE THAT YOU USED THE TECHNOLOGY AND INCLUDE ALL PROMPTS THAT YOU HAVE USED.***

Policies

Students are expected to attend and participate in lecture discussions, in-class exercises, and team meetings.

Assignments turned in late will have 25% of the total points deducted from the graded score for each late day.

No make-up exams (except for documented medical or family emergencies) will be offered. If they will not be able to attend an exam due to an athletic game or other valid reason, then they must coordinate with the instructor before the exam is given. They may arrange to take the exam before they leave, with an approved university personnel during the time they are gone, or within the week the exam is given. If students do not take an exam, then they will receive a 0 for the exam. Accommodations religious observance must be arranged with the Professor at least two weeks before the exam.

If students need accommodations authorized by OSAS (Office of Student Accessibility Services), notify the instructor at least two weeks before the exam. This will allow time for arrangements to be made.

Sharing of course materials outside of the learning environment
SCampus Section 11.12(B)

Distribution or use of notes or recordings based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study is a violation of the USC Student Conduct Code. This includes, but is not limited to, providing materials for distribution by services publishing class notes. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the Internet or via any other media. (See Section C.1 Class Notes Policy).

Week	Lecture Topic(s)	Readings & Assignments
1	Introduction to Advanced Cloud Analytics • Course Overview • Overview of Cloud Analytics • Career Relevance • USC ADAM: Applied Data Analytics Methodology • Overview of Tools & Data Sets	HW: Set up a basic cloud environment in AWS and write a brief report on the process and key components involved. Due: End of following week.
2	Data Driven Organizations • Importance of data-centric approaches • Benefits and challenges • Case studies of successful data-driven organizations Elements of Data • Types: structured, semi-structured, unstructured • Data sources and collection methods • Data quality and cleaning *In Class Exercise: Building Your Own Lab in AWS* *In Class Exercise: Accessing and Analyzing Data using S3*	Book: "Designing Data-Intensive Applications" by Martin Kleppmann Chapters: Chapter 2: Data Models and Query Languages (covers types of data and data models) Chapter 3: Storage and Retrieval (covers data storage and accessing methods) HW: Use a provided dataset to perform basic data analysis using cloud-based tools, focusing on data-centric approaches. Due: End of following week.
3	Design Principles and Patterns for Data Pipelines • Basics of data pipeline design • Common patterns and best practices • Real-world examples of data pipeline architectures *In Class Exercise: Querying Data by Using Athena*	Book: "Designing Data-Intensive Applications" by Martin Kleppmann Chapters: Chapter 6: Partitioning (covers data partitioning and patterns) Chapter 11: Stream Processing (covers data processing patterns) HW: Design and implement a simple data pipeline using cloud services, documenting the architecture and tools used. Due: End of following week.

4	Securing and Scaling the Data Pipeline • Data security measures and best practices • Scaling strategies: vertical, horizontal, and functional • Monitoring and maintaining data pipelines	Book: "Designing Data-Intensive Applications" by Martin Kleppmann Chapters: Chapter 8: The Trouble with Distributed Systems (covers challenges and solutions in scaling) Chapter 9: Consistency and Consensus (covers data security and consistency) HW: Apply security best practices to your Week 3 pipeline and simulate scaling it for increased data load. Due: End of following week.
5	Ingesting and Preparing Data • Data ingestion methods and tools • Data transformation and cleaning techniques • ETL vs. ELT approaches • Batch processing: advantages and use cases • Stream (or system) processing: advantages and use cases • Choosing between batch and stream processing *In Class Exercise: ETL using Glue*	Book: "Designing Data-Intensive Applications" by Martin Kleppmann Chapters: Chapter 3: Storage and Retrieval (covers data storage options and retrieval methods) Chapter 4: Encoding and Evolution (covers data encoding and transformation) HW: Ingest a raw dataset into a cloud database and perform data cleaning and transformation operations. Due: End of following week.
6	Storing and Organizing Data • Data storage options: databases, data lakes, data warehouses • Data organization strategies: partitioning, indexing • Data lifecycle management *In Class Exercise: Storing and Analyzing Data using Redshift*	We will be referring to the readings that we did in Week 5. HW: Implement a data storage solution using a cloud data warehouse or data lake, and demonstrate data querying. Due: End of following week.
7	Processing Big Data • Introduction to big data challenges • Big data processing frameworks and tools • Real-world big data use cases *In Class Exercise: Processing Logs using EMR*	We will be referring to the readings that we did in Week 5. HW: Use a cloud-based big data processing tool to analyze a large dataset and summarize your findings. Due: End of following week.

8	Analyzing and Visualizing Data • Considering factors that influence tool selection • Comparing tools and services • Examples and use cases *In Class Exercise: Analyzing and Visualizing Streaming Data with Kinesis Data Firehose, OpenSearch Service, and OpenSearch Dashboards*	HW: Create an interactive data visualization dashboard using a cloud analytics service. Due: End of following week.
9	Automating the Pipeline • Automating infrastructure deployment • CI/CD • Automating with Step Functions *In Class Exercise: Building and Orchestrating ETL Pipelines using Athena*	HW: Automate a data pipeline component using cloud orchestration tools. Due: End of following week.
10	Machine Learning in the Cloud • Overview of cloud-based ML platforms • Benefits of cloud ML: scalability, flexibility, cost-effectiveness • ML lifecycle in the cloud	Book: "Pattern Recognition and Machine Learning" by Christopher M. Bishop Chapters: Chapter 1: Introduction (covers the basics of machine learning and its applications) Chapter 5: Neural Networks (covers the implementation of ML models) HW: Deploy a pre-built ML model in a cloud environment and test its performance with a dataset. Due: End of following week.
11	Processing Data for Machine Learning • Data preprocessing techniques for ML • Feature engineering and selection • Data splitting: training, validation, testing *In Class Exercise: Preparing Data and Training a Model using SageMaker & SageMaker Canvas*	Book: "Pattern Recognition and Machine Learning" by Christopher M. Bishop Chapters: Chapter 3: Linear Models for Regression (covers data preprocessing and feature engineering) Chapter 4: Linear Models for Classification (covers data splitting and model training) HW: Preprocess a dataset specifically for machine learning, including normalization and feature encoding. Due: End of following week.

12	Forecasting using Machine Learning • Basics of forecasting • Improvements using machine learning • Popular forecasting models and algorithms • Use cases: stock predictions, weather forecasting	Book: "Pattern Recognition and Machine Learning" by Christopher M. Bishop Chapters: Chapter 8: Graphical Models (covers machine learning models for forecasting) HW: Build and train a forecasting model using a time-series dataset in a cloud ML platform. Due: End of following week.
13	Computer Vision • Introduction to computer vision tasks • Tools and libraries for computer vision in the cloud • Use cases: image recognition, facial recognition	Book: "Pattern Recognition and Machine Learning" by Christopher M. Bishop Chapters: Chapter 14: Kernel Methods and Support Vector Machines (covers tools and models for computer vision tasks) HW: Implement a cloud-based computer vision model to classify or detect objects in images. Due: End of following week.
14	Natural Language Processing • Basics of NLP: tokenization, POS tagging, named entity recognition • Cloud tools for NLP tasks • Use cases: sentiment analysis, chatbots	Book: "Pattern Recognition and Machine Learning" by Christopher M. Bishop Chapters: Chapter 13: Sequential Data (covers models and techniques for NLP tasks) HW: Develop a cloud-based NLP model to perform tasks like text classification or sentiment analysis on a text dataset. Due: End of following week.
15	Course Wrap Up & Review • Recap of key concepts and learnings • Professional Certifications • Future trends in cloud analytics • Feedback and Q&A session	HW: Complete a capstone project that involves building a comprehensive cloud analytics solution, integrating multiple aspects of the course. Due: End of following week.
Finals Exam	Final exam will be administered according to USC's schedule of classes	

Statement on Academic Conduct and Support Systems

Academic Integrity:
The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Course Content Distribution and Synchronous Session Recordings Policies
USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion.

This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - *(213) 740-9355 – 24/7 on call*
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - *988 for both calls and text messages – 24/7 on call*
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - *(213) 740-9355(WELL) – 24/7 on call*
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - *(213) 740-5086*
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - *(213) 740-5086 or (213) 821-8298*
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - *(213) 740-0776*
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - *(213) 740-0411*
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - *(213) 740-2101*
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - *UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - *UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call*
Non-emergency assistance or information.

Office of the Ombuds - *(213) 821-9556 (UPC) / (323-442-0382 (HSC)*
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.