USC Marshall
School of Business

DSO 599 – Hands-on Data Analytics and Machine Learning in AWS Cloud
Fall 2023
Saturday: 9:00 - 12:00 noon *(3.0 Units)*

Instructor:	*Sudi Bhattacharya*
Office:	*Virtual*
Office Hours:	*After Class by Appointment*
Phone:	*(818) 264 8006*
Email:	sudibhat@marshall.usc.edu

COURSE DESCRIPTION

This introductory course is intended for students interested in pursuing a career in data science, ML, and AI. The course enables aspiring data analysts, data engineers and data scientists to design, develop and build machine learning solutions on the AWS Cloud. Machine Learning has the potential to reshape, redefine, and transform business processes. Advances in Cloud Computing is making more processing power, unlimited storage, and elastic resources available to develop high-performance ML applications. Consequently, many companies are using public Cloud to experiment, develop and deploy high impact machine learning solutions. By the end of this course, students will be able to select and apply machine learning services in AWS Cloud to solve business problems. They will also be able to label, build, train, and deploy a custom machine learning model in AWS Cloud through a guided, hands-on approach.

COURSE OBJECTIVES

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

1. Understand how high-performing teams modernize today's enterprise using public Cloud
   1.1. Describe AWS Cloud core infrastructure components
   1.2. Create AWS Compute (EC2)
   1.3. Create AWS Storage (S3)
2. Understand how high-performing teams modernize today's enterprise using Data Analytics in the Cloud
   2.1. Describe AWS Cloud core data analytics components
   2.2. Describe big data analytical concepts
   2.3. Ingest, store, and secure data
   2.4. Query a data store with manual schema specification
   2.5. Query a data store with automated schema generation
   2.6. Load and query data in a data warehouse
   2.7. Analyze unstructured and IoT data
3. Understand how high-performing teams modernize today's enterprise using Machine Learning in the Cloud
   3.1. Describe machine learning (ML)
   3.2. Implement a machine learning pipeline using Amazon SageMaker
   3.3. Use managed Amazon ML services for forecasting
   3.4. Use managed Amazon ML services for computer vision
   3.5. Use managed Amazon ML services for natural language processing

STUDENT PREQUISITES

To ensure successful completion of this course, students should have:

- Experience scripting with Python
- A basic understanding of statistics

DELIVERY METHOD

This course will be delivered through a combination of the following material:

- Synchronous online lectures,
- Online class demonstrations of Cloud technology
- Asynchronous video lectures or learning material
- Asynchronous lab assignments that students can complete independently.
- Online class discussions on assigned articles and Whitepapers

We will use AWS Academy Leaner lab for hand-on work in this course. An AWS Academy Learner Lab offers a long-running lab environment suitable for student projects over a period designated by the educator. Each student in an AWS Academy Learner Lab will have a $100 AWS Platform Credit to use for the duration of the lab. Within this lab, students will have access to a restricted set of AWS services for project work, and a set of lab exercises specifically designed for this course by the instructor. Services deployed by students are available until the end date designated by the educator.

All the the material (videos, presentations, guided labs, knowledge check quizzes) used in this course will be developed specifically for this course.

This course will be delivered using USC Blackboard LMS, except for AWS Academy learner Lab which can only be accessed through AWS Academy LMS.

LEARNING RESOURCES

- Lecture materials
- Online multiple-choice knowledge checks
- Guided Lab exercises on AWS Cloud
- Digital -prerecorded lectures
- Software Demonstration
- Articles, Videos and Whitepapers

A list of key representative whitepapers that we will use in class is below. These papers are publicly available. We may add to or take away from this list as appropriate.

1. Overview of Amazon Web Services
2. AWS Well-Architected Framework
3. Big Data Analytics Options on AWS
4. Data Analytics Lens - AWS Well-Architected Framework
5. AWS Cloud Data Ingestion Patterns and Practices
6. Build Modern Data Streaming Architectures on AWS
7. The Machine Learning Lens – AWS Well Architected Framework
8. Accenture Enterprise AI – Scaling Machine Learning and Deep Learning Models
9. Hybrid Machine Learning

GRADING

An individual student's grades will be calculated based on the following components. It is important to note that exceptional work or effort by a student will be recognized and rewarded. Similarly, plagiarism in written assignments will be penalized according to USC rules and guidelines.

Categories	**Assignments**	**Points**	**% of Overall Grade**
Class Participation		100	10%
Guided hands-on labs	*There will be 18 hands-on guided labs in this course. At least 15 of these labs need to be completed.*	300	30%
Knowledge Check Quizzes	*There will be 10 knowledge Check quizzes (15 points each)*	150	15%
Mid-term Examination	*Multiple Choice Scenario based (1hr, 20 Questions, 7.5 points each)*	150	15%
Final Examination	*Multiple Choice Scenario based (2hrs, 40 Questions, 7.5 points each)*	300	30%
	Total	1000	100%

CLASS PARTICIPATION

Students are expected to participate in the lectures, an especially when discussing questions on reading assignments. Participation grades will be based on the quality of a student's contribution to the lectures. Participation will be graded by the professor in each class.

Students who are passionate, curious and willing to engage in class debate will receive higher participation grades. More importantly, this will create more exciting and engaging learning experience.

A full explanation of what constitutes *effective class participation* is described in Appendix II.

LIST OF GUIDED HANDS-ON LABS ON AWS CLOUD

Out of 18 labs, nine labs will need to be completed during class hours with the professor or the instructional assistant present, other nine can be completed any time asynchronously before the beginning of the next class.

Lab #	Activity
Guided Lab 1	Ingesting Data into Amazon S3
Guided Lab 2	Querying Amazon S3 Data Using Amazon Athena
Guided Lab 3	Transforming Data Using Amazon S3, AWS Glue, and Amazon Athena
Guided Lab 4	Loading the Amazon Redshift Cluster With Data and Querying
Guided Lab 6	Setting up and Executing a Data Pipeline Job to Load Data into Amazon S3
Guided Lab 7	Streaming Data with AWS Kinesis Firehose, Amazon Elasticsearch Service, and Kibana
Guided Lab 8	Using AWS IoT Analytics for Data Ingestion and Analysis
Guided Lab 9	Exploring Amazon SageMaker
Guided Lab 10	Visualizing Data
Guided Lab 11	Encoding Categorical Variables
Guided Lab 12	Splitting Data and Training a Model with XGBoost
Guided Lab 13	Hosting and Consuming a Model in AWS
Guided Lab 14	Evaluating Model Accuracy
Guided Lab 15	Tuning with AWS SageMaker
Guided Lab 16	Creating a Forecast with Amazon Forecast
Guided Lab 17	Facial Recognition in AWS
Guided Lab 18	Create a Chatbot to schedule appointments

KNOWLEDGE CHECK QUIZZES

Lab #	Topic to be Tested
Knowledge Check Quiz 1	AWS Cloud Fundamentals
Knowledge Check Quiz 2	AWS Compute, Storage and Databases
Knowledge Check Quiz 3	AWS Architecture
Knowledge Check Quiz 4	Data Warehouse and Data Lake Concepts
Knowledge Check Quiz 5	Data Pipeline Implementation
Knowledge Check Quiz 6	Machine Learning Concepts
Knowledge Check Quiz 7	Machine Learning Pipeline Implementation
Knowledge Check Quiz 8	Managed Services for Forecasting
Knowledge Check Quiz 9	Computer Vision
Knowledge Check Quiz 10	Natural Language Processing

MID-TERM AND FINAL EXAMINATION

The mid-term examination will be on the 7th week of class for a period of one hour. The exam will consist of 20 scenario based multiple choice questions.

The Final examination will be on the 16th week of class for a period of two hours. The exam will consist of 40 scenario based multiple choice questions.

THE IMPORTANCE OF COURSE EVALUATIONS

Students will be asked to submit an "unofficial" course evaluation to the instructor halfway through the class, as well as the official evaluation at the end of the semester. These evaluations are valuable because they allow the course to be continuously improved, based on feedback from students and instructor observations. Furthermore, the unofficial evaluation allows you to provide feedback directly to the instructor, if there is anything that you would like to see changed in the course. An example of what will be asked of students during the midpoint evaluation is shown in Appendix IV.

EMERGENCY PREPAREDNESS

In case of a declared emergency if travel to campus is not feasible, the USC Emergency Information web site (https://emergency.usc.edu/) will provide safety and other information, including electronic means by which instructors will conduct class using a combination of USC's Blackboard learning management system (blackboard.usc.edu), teleconferencing, and other technologies.

OPEN EXPRESSION AND RESPECT FOR ALL

An important goal of the educational experience at USC Marshall is to be exposed to and discuss diverse, thought-provoking, and sometimes controversial ideas that challenge one's beliefs. In this course we will support the values articulated in the USC Marshall "Open Expression Statement."

COURSE OUTLINE AND ASSIGNMENTS

Session	Topic of Focus	Labs for the Week	Whitepaper Reading Assignment	Knowledge Check Quiz
Week 1 Saturday 9:00 - 12:00 PT **Course Overview and Introduction**	**Course Introduction Cloud Concepts, AWS Services Overview**		Overview of Amazon Web Services	
Week 2 Saturday 9:00 - 12:00 PT AWS Cloud Services Fundamentals	**AWS Global Infrastructure, AWS Compute, Storage and Databases**	Guided lab 1 – Ingesting Data into S3 (to be completed during class hours) Guided lab 2 – Querying S3 data using Amazon Athena (to be completed asynchronously before the start of next class)	AWS Well-Architected Framework	Knowledge Check Quiz 1 – AWS Fundamentals
Week 3 Saturday 9:00 - 12:00 PT Architecting on AWS	**AWS Cloud Architecture, Security, Autoscaling and Monitoring**			Knowledge Check Quiz 2 – AWS Compute, Storage and Databases
Week 4 Saturday 9:00 - 12:00 PT Data Engineering Basics	**Data Warehousing, Data Lake and Analytics on AWS**	Guided lab 3 - Transforming Data Using Amazon S3, AWS Glue, and Amazon Athena (to be completed during class hours) Guided lab 5 - Delivering Insights using Amazon QuickSight (to be completed asynchronously before the start of next class)	Big Data Analytics Options on AWS	Knowledge Check Quiz 3 - AWS Architecture
Week 5 Saturday 9:00 - 12:00 PT Modern Data Storage	**Data Ingestion, Transformation and Query and Data Pipelines in AWS Cloud**	Guided lab 4 - Loading the Amazon Redshift Cluster With Data and Querying (to be completed during class hours) Guided lab 6 - Setting up and Executing a Data Pipeline Job to Load Data into Amazon S3	Data Analytics Lens - AWS Well-Architected Framework AWS Cloud Data Ingestion Patterns and Practices	Knowledge Check Quiz 4 - Data Warehousing and Data Lake Concepts

		(to be completed asynchronously before the start of next class)		
Week 6 Saturday 9:00 - 12:00 PT	**Modern Streaming Data Analytics**	Guided lab 7 - Streaming Data with AWS Kinesis Firehose, Amazon Elasticsearch Service, and Kibana (to be completed during class hours)	Build Modern Data Streaming Architectures on AWS	Knowledge Check Quiz 5 – Data Pipelines on AWS
Week 7 Saturday 9:00 - 12:00 PT	**IoT Data Analytics on AWS** **Mid-Term Exam (1 hr)**	Guided lab 8 - Using AWS IoT Analytics for Data Ingestion and Analysis (to be completed asynchronously before the start of next class)		
Week 8 Saturday 9:00 - 12:00 PT	**Introduction to Machine Learning on AWS**	Guided lab 9 – Exploring Amazon SageMaker (to be completed during class hours) Guided lab 10 - Visualizing Data (to be completed asynchronously before the start of next class) Guided Lab 11 – Encoding Categorical Variables (to be completed during class hours)	The Machine Learning Lens – AWS Well Architected Framework	Knowledge Check Quiz 6 – Machine Learning Concepts
Week 9 Saturday 9:00 - 12:00 PT	**Implementing a Machine Learning Pipeline with Amazon SageMaker**	Guided lab 12 - Splitting Data and Training a Model using XGBoost (to be completed asynchronously before the start of next class) Guided lab 13 – Hosting and Consuming a Model on AWS (to be completed during class hours) Guided lab 14 Evaluating Model	Accenture Enterprise AI – Scaling Machine Learning and Deep Learning Models Hybrid Machine Learning	Knowledge Check Quiz 7 – Machine Learning Pipeline Implementation

		Accuracy (to be completed asynchronously before the start of next class)		
Week 10 Saturday 9:00 - 12:00 PT	**Introduction to Forecasting in AWS**	Guided lab 15 – Tuning with Amazon SageMaker (to be completed during class hours) Guided lab 16 – Creating a forecast with Amazon Forecast (to be completed asynchronously before the start of next class)		Knowledge Check Quiz 8 – Managed Services for Forecasting
Week 11 Saturday 9:00 - 12:00 PT	**Introduction to Computer Vision in AWS**	Guided lab 17 – Computer Vision (to be during class hours)		Knowledge Check Quiz 9 – Computer Vision
Week 12 Saturday 9:00 - 12:00 PT	**Introduction to Natural Language Processing in AWS**	Guided lab 18 – Creating a bot to schedule appointments (to be completed asynchronously before the start of next class)		Knowledge Check Quiz 10 – Natural Language Processing
Week 13 Saturday 9:00 - 12:00 PT	**Course Wrap**			
Week 14 Saturday 9:00 - 12:00 PT	**Thanksgiving Break**			
Week 15 Saturday 9:00 - 12:00 PT	**Study Break**			
Week 16 Saturday TBD	**Final Exam**			

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.

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.

Appendix I. MARSHALL GRADUATE PROGRAMS LEARNING GOALS

How GSBA 5## Contributes to Marshall Graduate Program Learning Goals

Marshall Graduate Program Learning Goals	GSBA 5## Objectives that support this goal	Assessment Method*
Learning Goal #1: Develop Personal Strengths. **Our graduates will develop a global and entrepreneurial mindset, lead with integrity, purpose and ethical perspective, and draw value from diversity and inclusion.**	Objectives 1, 2 and 3 specifically contribute to 1.2 and 1.3 of learning goal # 1, specifically global and entrepreneurial mindset by making students aware of how modern ML is reshaping our world.	*We assess using hands-on guided labs, knowledge check quizzes, mid-term and final exams.*
1.1 Possess personal integrity and a commitment to an organization's purpose and core values.		
1.2 Expand awareness with a global and entrepreneurial mindset, drawing value from diversity and inclusion.		
1.3 Exhibit awareness of ethical dimensions and professional standards in decision making.		
Learning Goal #2: Gain Knowledge and Skills. **Our graduates will develop a deep understanding of the key functions of business enterprises and will be able to identify and take advantage of opportunities in a complex, uncertain and dynamic business environment using critical and analytical thinking skills.**	Objectives 1.1 – 1.3, 2.1-2.7 and 3.1 – 3.5 contribute to learning goal #2 by helping students gain deep knowledge and skills in modern Cloud, Data Analytics and ML technologies.	*We assess using hands-on guided labs, knowledge check quizzes, mid-term and final exams.*
2.1 Gain knowledge of the key functions of business enterprises.		
2.2 Acquire advanced skills to understand and analyze significant business opportunities, which can be complex, uncertain and dynamic.		
2.3 Use critical and analytical thinking to identify viable options that can create short-term and long-term value for organizations and their stakeholders.		

Learning Goal #3: Motivate and Build High Performing Teams. **Our graduates will achieve results by fostering collaboration, communication and adaptability on individual, team, and organization levels.**	Objectives 1, 2, and 3 contribute to learning goal #3, by helping students understand how high-performing teams with a variety of perspective, talents and skills come together to execute technology driven business transformation	*We assess using hands-on guided labs, knowledge check quizzes, mid-term and final exams.*
3.1 Motivate and work with colleagues, partners, and other stakeholders to achieve organizational purposes.		
3.2 Help build and sustain high-performing teams by infusing teams with a variety of perspectives, talents, and skills and aligning individual success with team success and with overall organizational success.		
3.3 Foster collaboration, communication and adaptability in helping organizations excel in a changing business landscape.		

Appendix II

CLASS PARTICIPATION STATEMENTS

Class participation is an extremely important part of the learning experience in this course as the richness of the learning experience will be largely dependent upon the degree of preparation by *all* students prior to each class session.

A course that incorporates the frequent use of case analyses to illustrate the practical application of concepts and practices requires the student to diligently and thoroughly prepare cases and actively offer the results of the analyses and conclusions derived as well as recommendations during each class session. My expectation and that of your classmates are that you are prepared for *all* classes and will actively participate in and meaningfully contribute to class discussions.

In-class participation is also a critical part of this course's learning experience. Cold calling may take place to encourage active participation and to gain multiple perspectives and points of view, thus lending itself to the richness of the learning experience. In-class participation grading will be based on students' demonstrated willingness to participate and the quality of the comments expressed, rather than quantity. While some students are far more comfortable than others with class participation, *all* students should make an effort to contribute meaningfully.

Students will offer their opinions in group settings many times in their careers; thus, class participation serves to prepare students for this business experience.
The evaluating of in-class participation is based on the following:

- *Relevance* – Does the comment or question meaningfully bear on the subject at hand? Irrelevant or inappropriate comments can detract from the learning experience.
- *Responsiveness* – Does the comment or question connect to what someone else has said?
- *Analysis* – Is the reasoning employed consistent and logical? Has data from course materials, personal experience, or general knowledge been employed to support the assertions/findings?
- *Value* – Does the contribution further the understanding of the issues at hand?
- *Clarity* – Is the comment concise and understandable?

During class sessions, I frequently assume the role of a facilitator to encourage a discussion that includes perspectives from a variety of viewpoints and, secondly, to help pull together prevailing analyses and recommendations. The direction and quality of a discussion is the *collective responsibility of the class*.

To underscore the importance of participation, 10 percent of the course grade or 100 of 1000 points are allocated to class participation.

For each in-class session ten (10) points will be awarded to a student for relevant and meaningful participation, seven (7) point for modest contributions to the class, five (5) point for minor contributions to the class and zero (0) points for no participation or absence. We will take the 10 best scores to calculate the participation score (maximum score of 100)

Class Participation—Behavioral Anchor Rating Scale:

Excellent Performance

- Initiates information relative to topics discussed
- Accurately exhibits knowledge of assignment content
- Clarifies points that others may not understand
- Shares personal experiences or opinions related to topic
- Offers relevant / succinct input to class
- Actively participates in class exercises
- Demonstrates ability to apply, analyze, evaluate & synthesize course material.
- Demonstrates willingness to attempt to answer unpopular questions
- Builds on other students' contributions

Average Performance

- Participates in group discussions when asked
- Demonstrates knowledge of course material
- Offers clear, concise, "good" information on class assignments
- Offers input, but tends to reiterate the intuitive
- Attends class regularly

Unacceptable Performance

- Fails to participate even when directly asked
- Gives no input to discussions
- Does not demonstrate knowledge of the readings
- Shows up to class: does nothing
- Distracts group / class
- Irrelevant discussion