Program Name : **Diploma in Artificial Intelligence and Machine Learning**

Program Code : **AN**

Semester : **Fifth**

Course Title : **Cloud Computing for Data Science (Elective)**

Course Code : **22594**

1. RATIONALE

Cloud computing is one of the fastest growing domains. Cloud computing makes while allowing developers to build ML algorithms faster. Using cloud solutions for machine learning allow developers to build ML algorithms more accessible, flexible, and cost-effective. This course will cover the basic architecture of cloud environment, uses of various available cloud services to ease the machine learning process.

2. COMPETENCY

The aim of this course is to help the student to attain following industry identified competency through various teaching learning experience:

- **Use various cloud services for data science in machine learning.**

3. COURSE OUTCOMES (COs)

a. Identify the architecture and infrastructure of cloud computing
b. Classify various cloud service models
c. Explain cloud data warehouse functions
d. Describe data pipeline design in cloud environment
e. Demonstrate container for data validation
f. Demonstrate ML studio for model training

4. TEACHING AND EXAMINATION SCHEME

Teaching Scheme			Credit (L+T+P)	Examination Scheme												
				Theory							Practical					
L	T	P		Paper Hrs.	ESE		PA		Total		ESE		PA		Total	
					Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
3	-	2	5	3	70	28	30*	00	100	40	25#	10	25	10	50	20

(**) *marks should be awarded on the basis of internal end semester theoryexam of 50 marks based on the specification table given in S. No. 9.* ***(~[2]): For the practical only courses,*** *the PA has two components under practical marks i.e. the assessment of practicals (seen in section 6) has a weightage of 60% (i.e.30 marks) and micro-project assessment (seen in section 12) has a weightage of 40% (i.e.20 marks). This is designed to facilitate attainment of COs holistically, as there is no theory ESE.*

Legends: L-Lecture; **T** – Tutorial/Teacher Guided Theory Practice; **P** -Practical; **C** – Credit, **ESE** -End Semester Examination; **PA** - Progressive Assessment, **'#':** No Theory Examination

5. COURSE MAP (with sample COs, Learning Outcomes i.e. LOs and topics)

This course map illustrates an overview of the flow and linkages of the topics at various levels of outcomes (details in subsequent sections) to be attained by the student by the end of the

course, in all domains of learning in terms of the industry/employer identified competency depicted at the centre of this map.

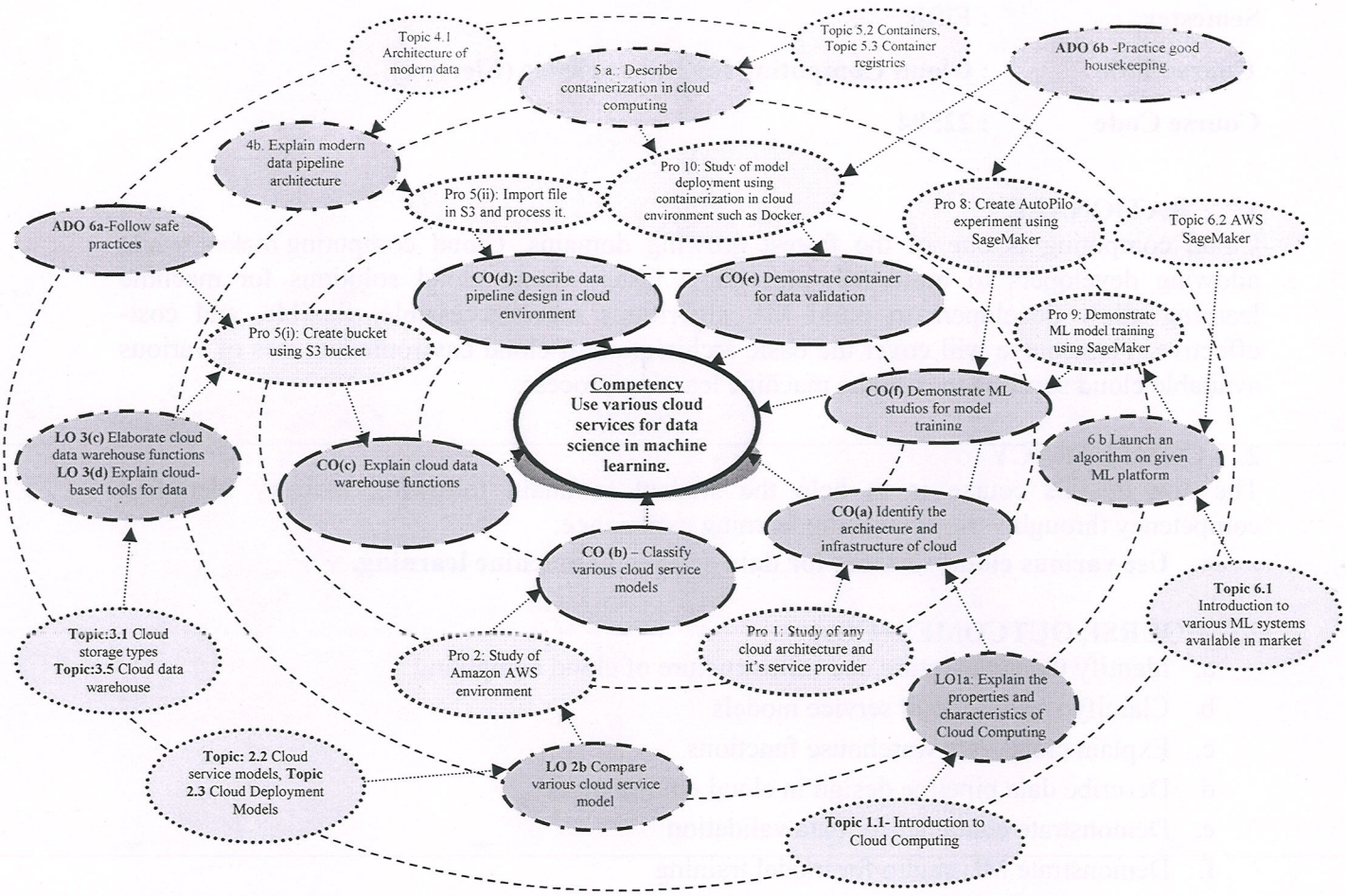


Figure 1 - Course Map

6. SUGGESTED PRACTICALS/ EXERCISES

The practicals/exercises/tutorials in this section are psychomotor domain LOs (i.e.sub-components of the COs) are to be developed andassessed in the student to lead to the attainment of the competency.

S. No.	Practical Exercises (Learning Outcomes to be achieved through practicals)	Unit No.	Approx. Hrs. Required
1	Study of any cloud architecture and it's service provider (with respect to front-end, back-end) such as Amazon AWS, Microsoft Azure, Google Cloud Platform (GCP) etc.	I	02
2	Study of Amazon AWS environment with respect to compute, storage, database, management tools, application services	I, II	02

S. No.	Practical Exercises (Learning Outcomes to be achieved through practicals)	Unit No.	Approx. Hrs. Required
3	(i) Create an AWS account with administrative role. (ii) Study of resource explorer in AWS	II	02
4	Create EC2 resources and launch the EC2 instance	III	02
5	(i) Create bucket using S3 bucket. (ii) Import file. Apply permissions to uploaded file. Perform processing such as authentication and authorization for selected file	IV	04
6	Prepare dataset using Data Wrangler. Process the created dataset using scikit-learn	III	04
7	(i) Create feature group and add features to the feature store (ii) Create dataset from features group	IV	04
8	Create Autopilot experiment using Amazon SageMaker Studio UI	VI	04
9	Demonstrate various training models on selected data set using SageMaker Autopilot	VI	04
10	Study of model deployment using containerization in cloud environment such as Docker.	V	04
	Total		**32**

'': compulsory practicals to be performed.*

Note

i. *Given in above tables is suggestive list of practical exercises. Teachers can design other similar exercises.*

ii. *Assessment of the 'Process' and 'Product' related skills in the laboratory/workshop/field work should be done as per suggested sample below:*

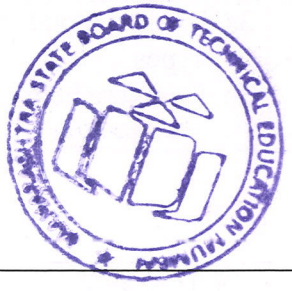
S.No.	Performance Indicators	Weightage in %
1	Preparation of experimental setup	30
2	Setting and Operation	20
3	Observation and Recording	20
4	Interpretation of result and conclusion	10
5	Answers to sample question	10
6	Submit Report in time	10
	Total	**100**

Additionally, the following affective domain LOs (social skills/attitudes), are also important constituents of the competency which can be best developed through the above-mentioned laboratory/field-based experiences:

a. Handle command prompt environment.
b. Plan, develop, assemble, link, debug and test the programs.
c. Demonstrate working as a leader/a team member.

The development of the attitude related LOs of Krathwohl's 'Affective Domain Taxonomy', the achievement level may reach:

- 'Valuing Level' in 1st year
- 'Organizing Level' in 2nd year and
- 'Characterizing Level' in 3rd year.



7. MAJOR EQUIPMENT/ INSTRUMENTS REQUIRED

The major equipment with broad specification mentioned here will usher in uniformity in conduct of experiments, as well as aid to procure equipment by authorities concerned.

S. No.	Equipment Name with Broad Specifications	Expt. S.No.
1	Computer System-Hardware: Personal computer, (i3 preferable) with min 8GB RAM, 512 GB HDD, Gigabit Ethernet network equipment, Software Requirement: Apache Tomcat, Java/Python/ equivalent programming language setup, Virtualization software. Academic version of any cloud service (AWS)	All

8. UNDERPINNING THEORY COMPONENTS

The following topics/subtopicsshouldbetaught and assessed in order to develop LOs in cognitive domain for achieving the COs to attain the identified competency.

Unit	Major Learning Outcomes (in cognitive domain)	Topics and Sub-topics
Unit - I Cloud Computing Fundamentals	1a. Explain the properties and characteristics of Cloud Computing 1b. Explain various layers and types of Cloud 1c. Identify the challenges and risks related to various aspects such as Security and Privacy 1d. Explain the evolution of cloud computing and virtualization	1.1. Introduction to Cloud Computing – Definition, Evolution of Cloud computing (from Mainframes to Clouds), Service – Oriented Architecture, Web Services, Grid Computing, Utility Computing, Hardware Virtualization. 1.2. Properties and Characteristics of a Cloud computing 1.3. Challenges and Risks: Security, Privacy, and Trust, Data Lock-In and Standardization, Availability, Fault-Tolerance, and Disaster Recovery, Resource Management and Energy-Efficiency. 1.4. Advantages of Cloud computing in Machine Learning
Unit - II Cloud Architecture and Cloud Service Management	2a. Explain the given component of cloud computing Architecture 2b. Compare various cloud service model 2c. Illustrate the services offered by various cloud computing models 2d. Compare various Cloud deployment models	2.1. Cloud computing architecture: basic components – front-end platform, back-end, platform, networking, cloud-based delivery 2.2. Cloud Service Models: - Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Continuous delivery using PaaS 2.3. Cloud Deployment Models: Public, Private, Community, Hybrid 2.4. Cloud service management – SLA, SLO, Policies and mechanisms for service management – admission control, load balancing, capacity allocation, energy optimization, QoS
Unit-III Cloud Data storage	3a. Explain cloud storage types 3b. Compare between storage types	3.1. Cloud storage types 3.2. Cloud data governance 3.3. Key-Value databases 3.4. Batch data and streaming data in

Unit	Major Learning Outcomes (in cognitive domain)	Topics and Sub-topics
	3c. Elaborate cloud data warehouse functions 3d. Explain cloud-based tools for data science	Machine learning 3.5. Cloud data warehouse – AWS Redshift 3.6. Various cloud-based tools used for data science in ML – GCP BigQuery,
Unit-IV Data Management using Cloud Computing	4a. Enlist data pipeline characteristics 4b. Explain Modern data pipeline architecture 4c. Describe ELT 4d. Describe data delivery in cloud computing	4.1 Architecture of Modern Data Pipelines 4.2 Data pipeline characteristics 4.3 Collecting and Ingesting Data 4.4 Transforming Data 4.5 Designing pipelines 4.6 Evolving from ETL to ELT 4.7 Delivering and sharing data
Unit – V Virtualization & Containerization & Elasticity in Cloud Computing	5a. Describe containerization in cloud computing 5b. Elaborate container registries 5c. Demonstrate model training pipeline in Kubernetes	5.1 Elastic Resources 5.2 Containers: Docker, Introduction to DevOps 5.3 Container Registries 5.4 Kubernetes in the Cloud – scaling, pipeline, microservices 5.5 Hybrid and Multi-cloud Kubernetes 5.6 Running Kubernetes locally with Docker Desktop and sklearn flask
Unit VI Managed Machine Learning Systems	6.a. Compare commercial and open-source ML systems 6.b. Launch an algorithm on given ML platform 6.c. Create log-in account and setup environment to train ML algorithm	6.1. Introduction to various ML systems available in market, Benefits of using managed ML platforms 6.2. Jupyter Notebook – Introduction, The workflow, 6.3. Azure ML Studio 6.4. Google AutoML Computer Vision 6.6. AWS SageMaker

Note: To attain the COs and competency, above listed Learning Outcomes (LOs) need to be undertaken to achieve the 'Application Level' of Bloom's 'Cognitive Domain Taxonomy'.

9. SUGGESTED SPECIFICATION TABLE FOR QUESTION PAPER DESIGN

Unit No.	Unit Title	Teaching Hours	Distribution of Theory Marks			
			R Level	U Level	A Level	Total Marks
I	Cloud Computing Fundamentals	8	6	6	-	12
II	Cloud Architecture and Cloud Service Management	8	2	6	4	12
III	Cloud Data storage	6	4	4	2	10
IV	Data Management using Cloud Computing	10	4	4	4	12
V	Virtualization & Containerization & Elasticity in Cloud Computing	10	4	4	6	14
VI	Managed Machine Learning Systems	6	2	4	4	10
	Total	**48**	**22**	**28**	**20**	**70**

Legends: R=Remember, U=Understand, A=Apply and above (Bloom's Revised taxonomy)
Note: This specification table provides general guidelines to assist student for their learning and to teachers to teach and assess students with respect to attainment of LOs. The actual distribution of marks at different taxonomy levels (of R, U and A) in the question paper may vary from above table.
This specification table also provides a general guideline for teachers to frame internal end semester practical theory exam paper which students have to undertake.

10. SUGGESTED STUDENT ACTIVITIES

Other than the classroom and laboratory learning, following are the suggested student-related ***co-curricular*** activities which can be undertaken to accelerate the attainment of the various outcomes in this course:

a. Prepare journals based on practical performed inlaboratory.
b. Library/E-Booksurveyregardingassembly language programmingusedinComputer industries.
c. Prepare power point presentation for showing different types of Assembly language ProgrammingApplications.

11. SUGGESTED SPECIAL INSTRUCTIONAL STRATEGIES (if any)

These are sample strategies, which the teacher can use to accelerate the attainment of the various outcomes in this course:

a. Massive open online courses (***MOOCs***) may be used to teach various topics/sub topics.
b. ***'L' in item No. 4*** does not mean only the traditional lecture method, but different types of teaching methods and media that are to be employed to develop the outcomes.
c. About ***15-20% of the topics/sub-topics*** which is relatively simpler or descriptive in nature is to be given to the students for ***self-directed learning*** and assess the development of the LOs/COs through classroom presentations (see implementation guideline for details).
d. With respect to item No.10, teachers need to ensure to create opportunities and provisions for ***co-curricular activities***.
e. Guide student(s) in undertaking micro-projects.
f. No. of practical's selection to be performed should cover all units.

12. SUGGESTED MICRO-PROJECTS

Only one micro-project is planned to be undertaken by a student assigned to him/her in the beginning of the semester. S/he ought to submit it by the end of the semester to develop the industry-oriented COs. Each micro-project should encompass two or more COs which are in fact, an integration of practicals, cognitive domain and affective domain LOs. The micro-project could be industry application based, internet-based, workshop-based, laboratory-based or field-based. Each student will have to maintain dated work diary consisting of individual contribution in the project work and give a seminar presentation of it before submission. The total duration of the micro-project should not be less than ***16 (sixteen) studentengagement hours*** during the course.

In the first four semesters, the micro-project could be group-based. However, in higher semesters, it should be individually undertaken to build up the skill and confidence in every student to become problem solver so that s/he contributes to the projects of the industry. A suggestive list is given here. Similar micro-projects could be added by the concerned faculty:

a. Create an EC2 instance with specified resources and manage resources.
b. Create cloud storage on any given platform to upload dataset.
c. Create dataset for given problem statement. Also maintain dataset on the platform

d. Create / Use ML model through available platforms such as AWS SageMaker / Microsoft ML studio
e. Train ML model to perform given classification problems
f. Deploy ML algorithm to perform given task.

13. SUGGESTED LEARNING RESOURCES

S. No.	Title of Book	Author	Publication
1	Cloud Computing: Principles and paradigm	Rajkumar Buyya	Wiley Publication
2	Machine Learning in the AWS Cloud	Abhishek Mishra	Sybex
3	Cloud Computing for Machine Learning and Cognitive Applications	Kai Hwang	The MIT Press (16 June 2017)
4	Cloud Data Engineering for dummies	David Baum	Jon Wiley & Sons, Inc.

14. SOFTWARE/LEARNING WEBSITES

Sr	Topic	Software / Website reference
1	Cloud computing architecture	www.nptel.ac.in
2	AWS service: EC2, S3	https://docs.aws.amazon.com/ec2/index.html?nc2=h_ql_doc_ec2
3	AWS S3 bucket user	https://docs.aws.amazon.com/pdfs/AmazonS3/latest/userguide/s3-userguide.pdf
4	AWS services	https://aws.amazon.com/getting-started/?nc2=h_ql_le
5	Importance of cloud computing for data science	• https://www.quora.com/How-important-is-cloud-computing-for-a-data-engineer • https://towardsdatascience.com/aws-essentials-for-data-science-why-cloud-computing-141cc6cee284
6	Course for data engineering using cloud computing	https://www.coursera.org/learn/cloud-data-engineering-duke
7	ML model creation and evaluation	https://scikit-learn.org/stable/modules/model_evaluation.html
8	Cloud computing for machine learning	https://www.javatpoint.com/machine-learning-and-cloud-computing
9	Introduction to AWS SageMaker	https://youtu.be/AVNqdT5ilOg
10	AWS Feature store	https://aws.amazon.com/blogs/machine-learning/automate-feature-engineering-pipelines-with-amazon-sagemaker/