**CLOUD COMPUTING FOR DATA SCIENCE** **Course Code : 315327**

**Programme Name/s** : **Artificial Intelligence/ Artificial Intelligence and Machine Learning/ Data Sciences**
**Programme Code** : **AI/ AN/ DS**
**Semester** : **Fifth**
**Course Title** : **CLOUD COMPUTING FOR DATA SCIENCE**
**Course Code** : **315327**

## I. RATIONALE

Cloud Computing play an important role for Data Science and Machine Learning Algorithm. It provides a flexible, scalable and cost-effective environment for data science which is important for computational needs, enhancing collaboration and ensuring robust data management with security. This Course will cover basic architecture of Cloud environment, use of various available cloud Services, build and develop machine learning algorithm to train and test the machine learning Model.

## II. INDUSTRY / EMPLOYER EXPECTED OUTCOME

Use Cloud computing platform to solve real-world data science problems.

## III. COURSE LEVEL LEARNING OUTCOMES (COS)

Students will be able to achieve & demonstrate the following COs on completion of course based learning

- CO1 - Use Cloud-Based software services by comprehending the cloud Computing Architecture.
- CO2 - Configure Virtual Machines using Virtualization techniques.
- CO3 - Implement Virtualized storage system in Cloud.
- CO4 - Use Machine Learning algorithms in Cloud Environment.
- CO5 - Deploy Machine Learning Models on Cloud.

## IV. TEACHING-LEARNING & ASSESSMENT SCHEME

| Course Code | Course Title | Abbr | Course Category/s | Learning Scheme | | | | | Credits | Assessment Scheme | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Actual Contact Hrs./Week | | | SLH | NLH | | Paper Duration | Theory | | | | Based on LL & TL | | | | Based on SL | | Total Marks |
| | | | | | | | | | | | | | | | Practical | | | | | | |
| | | | | CL | TL | LL | | | | | FA-TH | SA-TH | Total | | FA-PR | | SA-PR | | SLA | | |
| | | | | | | | | | | | Max | Max | Max | Min | Max | Min | Max | Min | Max | Min | |
| 315327 | CLOUD COMPUTING FOR DATA SCIENCE | CCD | DSE | 4 | - | 2 | - | 6 | 2 | 3 | 30 | 70 | 100 | 40 | 25 | 10 | 25# | 10 | - | - | 150 |

**Total IKS Hrs for Sem. :** 0 Hrs
Abbreviations: CL- ClassRoom Learning , TL- Tutorial Learning, LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, IKS - Indian Knowledge System, SLA - Self Learning Assessment
Legends: @ Internal Assessment, # External Assessment, *# On Line Examination , @$ Internal Online Examination
Note :

1. FA-TH represents average of two class tests of 30 marks each conducted during the semester.
2. If candidate is not securing minimum passing marks in FA-PR of any course then the candidate shall be declared as "Detained" in that semester.
3. If candidate is not securing minimum passing marks in SLA of any course then the candidate shall be declared as fail and will have to repeat and resubmit SLA work.
4. Notional Learning hours for the semester are (CL+LL+TL+SL)hrs.* 10 Weeks
5. 1 credit is equivalent to 30 Notional hrs.
6. * Self learning hours shall not be reflected in the Time Table.
7. * Self learning includes micro project / assignment / other activities.

## V. THEORY LEARNING OUTCOMES AND ALIGNED COURSE CONTENT

| Sr.No | Theory Learning Outcomes (TLO's)aligned to CO's. | Learning content mapped with Theory Learning Outcomes (TLO's) and CO's. | Suggested Learning Pedagogies. |
|---|---|---|---|
| 1 | TLO 1.1 Explain the Evolution of Cloud Computing.<br>TLO 1.2 Explain characteristics of Cloud Computing.<br>TLO 1.3 Describe the Architecture of Cloud Computing. | **Unit - I Introduction to Cloud Computing**<br>1.1 Introduction to Cloud Computing: Definition, Evolution of Cloud computing (from Mainframes to Clouds), Service – Oriented Architecture, Web Services, Grid Computing, Utility Computing<br>1.2 Characteristics of a Cloud computing<br>1.3 Cloud computing architecture: Basic components: front-end platform, back-end, platform, Networking, cloud-based delivery<br>1.4 Cloud Service Models: Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Continuous delivery using PaaS | Lecture Using Chalk-Board<br>Flipped Classroom<br>Presentations |
| 2 | TLO 2.1 Explain Concept of Virtualization.<br>TLO 2.2 Explain the given type of Virtualization.<br>TLO 2.3 Compare various cloud service models.<br>TLO 2.4 Describe various Cloud Deployment Models.<br>TLO 2.5 Explain role of Cloud computing in Data Science. | **Unit - II Virtualization and Cloud Computing**<br>2.1 Introduction to Virtualization, types of Virtualization, Application Virtualization, Network Virtualization, Desktop Virtualization, Storage Virtualization, Server Virtualization, Data virtualization<br>2.2 Cloud Deployment Models: Public, Private, Community, Hybrid<br>2.3 Role of Cloud computing in data Science<br>2.4 Advantages of Cloud Computing in Machine Learning | Lecture Using Chalk-Board<br>Presentations<br>Flipped Classroom |
| 3 | TLO 3.1 Explain Cloud Storage.<br>TLO 3.2 Describe the given Cloud storage system.<br>TLO 3.3 Explain features and limitations of Key-Value databases.<br>TLO 3.4 Compare Batch data and Streaming data in Machine learning.<br>TLO 3.5 Explain GCP BigQuery cloud platform. | **Unit - III Cloud Storage**<br>3.1 Cloud Storage: Introduction, Benefits of using Cloud Storage, Use cases of Cloud Storage (Backup, Archiving, Disaster recovery, Data processing, Content delivery)<br>3.2 Cloud storage system: Block-Based, File-Based, Object-Based Storages<br>3.3 Key-Value Databases: Introduction, features, limitations<br>3.4 Batch data and Streaming data in Machine learning<br>3.5 Cloud Data Warehouse– AWS Redshift<br>3.6 Various Cloud-based tools used for data science in ML– GCP BigQuery | Lecture Using Chalk-Board<br>Presentations<br>Demonstration |
| 4 | TLO 4.1 Explain benefits of using Machine Learning in the Cloud.<br>TLO 4.2 Explain the given type of Cloud-Based Machine Learning Service.<br>TLO 4.3 Describe various Machine Learning systems.<br>TLO 4.4 Describe features of the given Cloud Machine Learning Platforms. | **Unit - IV Cloud Computing for Data Science**<br>4.1 Machine Learning in the Cloud: Benefits and Limitations<br>4.2 Types of Cloud-Based Machine Learning Services: Artificial Intelligence as a Service (AIaaS), GPU as a Service (GPUaaS)<br>4.3 Introduction to various ML systems and benefits of using Managed ML platforms<br>4.4 Cloud Machine Learning Platforms: AWS SageMaker, Azure Machine Learning studio, Google Cloud AutoML | Lecture Using Chalk-Board<br>Presentations<br>Demonstration |

| Sr.No | Theory Learning Outcomes (TLO's)aligned to CO's. | Learning content mapped with Theory Learning Outcomes (TLO's) and CO's. | Suggested Learning Pedagogies. |
|---|---|---|---|
| 5 | TLO 5.1 Explain factors for Selection of Cloud Machine Learning Platform.<br>TLO 5.2 Describe ETL and ELT Pipelines.<br>TLO 5.3 Explain Pre-Tuned AI Services.<br>TLO 5.4 Write step wise procedure to train a Machine Learning Project. | **Unit - V Training and Deployment of ML on Cloud**<br>5.1 Factors for Selection of Cloud Machine Learning Platform, Support for ETL or ELT Pipelines, Support for Scale-Up and Scale-Out Training, Support for Machine Learning Frameworks, Pre-Tuned AI Services, Monitor Prediction Performance<br>5.2 Training Machine Learning Projects in the Cloud: Steps to train Machine Learning project in the Cloud, Identify and Understand Data Sources, Engineer the Features, Train and Validate Model, Deploy and Monitor Model | Lecture Using Chalk-Board Presentations Demonstration |

## VI. LABORATORY LEARNING OUTCOME AND ALIGNED PRACTICAL / TUTORIAL EXPERIENCES.

| Practical / Tutorial / Laboratory Learning Outcome (LLO) | Sr No | Laboratory Experiment / Practical Titles / Tutorial Titles | Number of hrs. | Relevant COs |
|---|---|---|---|---|
| LLO 1.1 Use Google Cloud Software Services. | 1 | * Use Google Services to create Doc/Sheet/Keep/Forms | 2 | CO1 |
| LLO 2.1 Create Virtual Machine. | 2 | *Create Virtual Machine using VMware workstation for Windows/Linux | 2 | CO2 |
| LLO 3.1 Create Web Server on Virtual Machine. | 3 | Create Web server (WAMP/XAMP/APACHE) on Virtual Machine | 2 | CO2 |
| LLO 4.1 Create account on any cloud platform. | 4 | *Create an account on AWS or Azure, or Google Cloud Platform | 2 | CO2 |
| LLO 5.1 Use BigQuery to create tables and views. | 5 | Create an account on Google Cloud Platform. a) Create a project and access BigQuery. b) Query data directly from Google Sheets. c) Create tables and views using BigQuery | 2 | CO3 |
| LLO 6.1 Create instance and configure Block-Based storage. | 6 | Launch an EC2 instance with a specified configuration and configure Block-Based storage | 2 | CO3 |
| LLO 7.1 Configure File-Based storage. | 7 | *Create and Configure File-Based storage on EC2 instance | 2 | CO3 |
| LLO 8.1 Configure Object-Based storage. | 8 | Create and Configure Object-Based storage on EC2 instance | 2 | CO3 |
| LLO 9.1 Upload dataset to Aws S3 bucket using Python script. | 9 | Write a script in Python to upload a dataset to an S3 bucket, list the files in the bucket, and download a file | 2 | CO3 |
| LLO 10.1 Use Amazon Sagemaker notebook. | 10 | *Create instance of Amazon Sagemaker notebook | 2 | CO4 |
| LLO 11.1 Build and Deploy a model using Amazon Sagemaker. | 11 | * Build, test, tune, train, deploy and validate a model using Amazon Sagemaker | 2 | CO4 CO5 |

**Note : Out of above suggestive LLOs -**

- '*' Marked Practicals (LLOs) Are mandatory.
- Minimum 80% of above list of lab experiment are to be performed.
- Judicial mix of LLOs are to be performed to achieve desired outcomes.

## VII. SUGGESTED MICRO PROJECT / ASSIGNMENT/ ACTIVITIES FOR SPECIFIC LEARNING / SKILLS DEVELOPMENT (SELF LEARNING)

**Micro project**

- Use AWS Glue, Google Dataflow, or Azure Data Factory to create a data pipeline that ingests, transforms, and loads data from one storage service to another. Document the pipeline configuration and execution results.
- Create a Machine on AWS and make basic services available on machine like word, power point etc.
- Set up a relational database using AWS RDS, Azure SQL Database, or Google Cloud SQL. Create a database schema, insert sample data, and perform queries using SQL. Document the steps and results.
- Launch an EC2 instance with a specified configuration (e.g., type, region). Install software (e.g., Python, Jupyter Notebook) on the instance. Connect to the instance and run a basic Python script.
- Create a local cloud on Virtual Machine on VmWare workstation. Deploy a web application and make it accessible by URL.
- Write and deploy a simple AWS Lambda function or Azure Function that processes data (e.g., transforms a JSON file). Test the function with sample data and set up triggers for automatic execution.

**Note :**

- Above is just a suggestive list of microprojects and assignments; faculty must prepare their own bank of microprojects, assignments, and activities in a similar way.
- The faculty must allocate judicial mix of tasks, considering the weaknesses and / strengths of the student in acquiring the desired skills.
- If a microproject is assigned, it is expected to be completed as a group activity.
- SLA marks shall be awarded as per the continuous assessment record.
- For courses with no SLA component the list of suggestive microprojects / assignments/ activities are optional, faculty may encourage students to perform these tasks for enhanced learning experiences.
- If the course does not have associated SLA component, above suggestive listings is applicable to Tutorials and maybe considered for FA-PR evaluations.

## VIII. LABORATORY EQUIPMENT / INSTRUMENTS / TOOLS / SOFTWARE REQUIRED

| Sr.No | Equipment Name with Broad Specifications | Relevant LLO Number |
|---|---|---|
| 1 | VmWare Workstation Software | 2,3 |
| 2 | Aws, EC2 and Sagemaker instance with services | 5,6,7,8,9,10,11 |
| 3 | Minimum Hardware requirement: Personal computer, (i3-i7 preferable), RAM minimum 8 GB. | All |
| 4 | Minimum system requirement: 64-bit operating system such as Windows 10, macOS 10.13, or a recent version of Linux. | All |

## IX. SUGGESTED WEIGHTAGE TO LEARNING EFFORTS & ASSESSMENT PURPOSE (Specification Table)

| Sr.No | Unit | Unit Title | Aligned COs | Learning Hours | R-Level | U-Level | A-Level | Total Marks |
|---|---|---|---|---|---|---|---|---|
| 1 | I | Introduction to Cloud Computing | CO1 | 6 | 4 | 8 | 0 | 12 |
| 2 | II | Virtualization and Cloud Computing | CO2 | 8 | 4 | 6 | 4 | 14 |
| 3 | III | Cloud Storage | CO3 | 8 | 4 | 6 | 4 | 14 |
| 4 | IV | Cloud Computing for Data Science | CO4 | 10 | 4 | 8 | 4 | 16 |
| 5 | V | Training and Deployment of ML on Cloud | CO5 | 8 | 4 | 6 | 4 | 14 |
| **Grand Total** | | | | **40** | **20** | **34** | **16** | **70** |

## X. ASSESSMENT METHODOLOGIES/TOOLS

**Formative assessment (Assessment for Learning)**

- Laboratory Performance, Unit Tests , Midterm Exam,Term Work, Seminar/Presentations.

- Continuous assessment based on process and product related performance indicators
- Each practical will be assessed considering 60% weightage to process and 40% weightage to product

**Summative Assessment (Assessment of Learning)**

- End Semester Exam, Practical exam, viva voce.

## XI. SUGGESTED COS - POS MATRIX FORM

| Course Outcomes (COs) | Programme Outcomes (POs) | | | | | | | Programme Specific Outcomes* (PSOs) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | PO-1 Basic and Discipline Specific Knowledge | PO-2 Problem Analysis | PO-3 Design/ Development of Solutions | PO-4 Engineering Tools | PO-5 Engineering Practices for Society, Sustainability and Environment | PO-6 Project Management | PO-7 Life Long Learning | PSO-1 | PSO-2 | PSO-3 |
| CO1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | | | |
| CO2 | 2 | 3 | 2 | 3 | 3 | 1 | 1 | | | |
| CO3 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | | | |
| CO4 | 3 | 2 | 3 | 3 | 2 | 1 | 1 | | | |
| CO5 | 3 | 2 | 3 | 3 | 2 | 2 | 2 | | | |

Legends :- High:03, Medium:02,Low:01, No Mapping: -
*PSOs are to be formulated at institute level

## XII. SUGGESTED LEARNING MATERIALS / BOOKS

| Sr.No | Author | Title | Publisher with ISBN Number |
|---|---|---|---|
| 1 | Dr. Anand Nayyar | Handbook of Cloud Computing | BPB Publication First Edition (1 January 2019) ISBN-10:9388176669 ISBN-13:978-9388176668 |
| 2 | Toby Velte, Anthony Velte, Robert C | Cloud Computing: A Practical Approach By Toby Velte, Anthony Velte, Robert C | McGraw Hill Professional ISBN-978-0-07-162965-8 |
| 3 | Noah Gift, Alfredo Deza | Cloud Computing for Data Analysis | Pragmatic AI Labs (No ISBN) |
| 4 | Valliappa Lakshmanan | Data Science on the Google Cloud Platform: Implementing End-to-End Real-Time Data Pipelines | O'Reilly Media, Inc. ISBN: 9781098118952 |
| 5 | Abhishek Mishra | Machine Learning in the AWS Cloud: Add Intelligence to Applications with Amazon SageMaker | Wiley Publication ISBN: 978-1-119-55671-8 |

## XIII . LEARNING WEBSITES & PORTALS

| Sr.No | Link / Portal | Description |
|---|---|---|
| 1 | https://www.geeksforgeeks.org/virtualization-cloud-computin g-types/ | Introduction to virtualization and cloud Computing and its types |
| 2 | https://www.javatpoint.com/virtualization-in-cloud-computing | Introduction to virtualization and cloud Computing |

**CLOUD COMPUTING FOR DATA SCIENCE** **Course Code : 315327**

| Sr.No | Link / Portal | Description |
|---|---|---|
| 3 | https://www.hostitsmart.com/blog/types-of-virtualization-in-cloud-computing-complete-overview/ | Overview of cloud computing, types of Virtualization advantages and Application |
| 4 | https://aws.amazon.com/what-is/virtualization/ | How can AWS help with virtualization and cloud computing? |
| 5 | https://www.run.ai/guides/machine-learning-in-the-cloud | Machine Learning in the Cloud,AWS Sagemaker Service for Machine Learning |

**Note :**

- Teachers are requested to check the creative common license status/financial implications of the suggested online educational resources before use by the students