SRI VENKATESWARA UNIVERSITY : : TIRUPATI

No.C-II(A)/2 year M.S.Data Science/2023-24   Date:12-03-2024

From   To
THE REGISTRAR   The Principal,
S.V.UNIVERSITY   S.V.U. College of Commerce
                                               Management and Information Sciences
                                               Tirupati

Sir,

Sub:- S.V.University – Courses of Study – Start New 2 year Master of Science in Data Science
       under self-supporting collaborative programme -   Scheme, Regulations and model question
       papers of the Course effective from the batch of students admitted into I-Semester during the
       academic year 2023-24 - Approval – Communicated – Regarding.

Ref: 1. Lr.No.APSCHE/ACIII/MS Data Science Programme/SVU/2023, dated 16-08-2023 of
        the Secretary, APSCHE, Atmakur, Mangalagiri, Guntur (dt)
     2. Lr. no.C-III(1)/M.S.Data Science/2023, dt:28-08-2023 from Registrar, SVU, TPT.
     3. MOU of EXAFLUENCE INC (Data Driven Influence),USA and S.V.University,
        Tirupati.
     4. Letter No;DCS/BOS/M.S.Data Science/2022, dated: 23-11-2022 received on 25-2-2024
                                    ******

        With reference   to the subject cited,   I am by direction, to inform you, that on the
recommendations of the Chairman, Board of Studies in   Computer Science(PG), the Vice-Chancellor
has approved Scheme, Regulations   and model question papers   for New 2 year M.S.Data Science Self
Supporting Course, effective from the batch of students admitted during the academic year 2023-24.

        Therefore, I request you to circulate the same   to the concerned for their information and
guidance.

        The receipt of this letter together with its enclosures may please be acknowledged.

                                               Yours faithfully,
                                               Sd/-O.Mohammad Hussain
Encl:As above   REGISTRAR

Copy to Chairman, Board of Studies in Computer Science (PG), Dept. of Computer Science,
        S.V.U. College of Commerce, Management and Computer Sciences, Tirupati, for information.
Copy to the Head, Dept. of Computer Science, S.V.U. College of Commerce, Management and
        Computer Sciences, Tirupati, for information.
Coy to Co-ordinator, M.S.Data Science, Department of Commerce, Management and Computer
        Sciences, Tirupati, for information
Copy to CEO, Exafluence Education, Data Science Education Centre, Next to Distance Education
        Centre Building, SV University, Tirupati-517 502 for information and necessary action.
Copy to the In-charge Director, S.V.U. Computer Centre, S.V. University, Tirupati for upload New
        Course syllabus.
Copy to the Controller of Examinations /A.A.O. (Examinations for P.G. Courses), S.V. University,
         for information and necessary action.
Copy to the Superintendents of B-VII, B-VIII and Computer Section in Examination Branch,
         S.V. University, for information and necessary action.
Copy to file.

M.S. Data Science

2 Years (4 Semesters)

Dept. Of Computer Science

SVU College of CM & CS, SRI VENKATESWARA UNIVERSITY: TIRUPATI

An Industry Collaborative Programme

WITH EXAFLUENCE INC., USA

SYLLABUS, ACADEMIC SCHEDULE & REGULATIONS

2023-25

Department of Computer Science

SVU COLLEGE OF COMMERCE MANAGEMENT & COMPUTER SCIENCE

SRI VENKATESWARA UNIVERSITY, TIRUPATI

M.S. DATA SCIENCE

An Industry Collaborative Programme in Hybrid Mode (online and offline)

CHOICE-BASED CREDIT SYSTEM (CBCS) Regulations 2023-2025 as per NEP-2020 for the batch admitted in the academic year 2023-25

EXAFLUENCE INC having its register office at 3759, US Route 1South, Suite 203, New Jersey, South Brunswick, United States of America- 08852 with overseas development centers in Bangalore, Chennai, Dubai, Canada and UK and ODC at #201, Ecco Greens- DTP, Upadhyaya Nagar, Tirupati 517507 Bangalore, Chennai in India and in Dubai entered a memorandum of understanding on 12th day of August 2022, with **SRI VENKATESWARA UNIVERSITY, TIRUPATI**, Andhra Pradesh, one of the oldest and renewed state-owned University in Andhra Pradesh.

Sri Venkateswara University was established in 1954, in Tirupati to cater the educational needs and aspirations of people of the Rayalaseema Region of Andhra Pradesh. After completing 68 years of excellence in teaching, research, extension and outreach activities, the University is committed to cater the needs of higher education offering a full range of post-graduate programs in Arts, Sciences, Law, Management, Education, Physical Education, Engineering and Pharmacy disciplines. From a humble beginning of one College with six departments, the University has now grown into the second largest University in Andhra Pradesh having five constituent Colleges Viz. College of Arts, College of Sciences, College of Commerce, Management & Computer Science, College of Pharmacy, and College of Engineering accommodating 52 departments offering 88 programs.

Exafluence head quartered in New Jersey USA with overseas development centers in Bangalore, Chennai, Tirupati in India, Dubai, Canada and UK, Exafluence has presence across global the intent is to enroll students in worldwide and bring in international students to SV University. Exafluence is promoted by alumni of SV University Mr. RaviKiran Dharmavaram and COO TN Sreenivas, and it's associated with MongoDB, AWS, Microsoft, Google etc., and with rich experience in data science, the students can get benefited with latest technologies and work in Real Time environment and connect with technical leaders spread across global.

1. Preamble:

M.S. Data Science Professional Programme is of two academic years with each academic year being divided into two consecutive (one odd + one even) semesters. Choice-Based Credit System (CBCS) is a flexible system of learning and provides choice for students to select from the prescribed elective courses. A course defines learning objectives and learning outcomes and comprises of lectures / tutorials / laboratory work / field work / project work /viva / seminars / assignments / presentations / self-study etc. or a combination of some of these. Under the CBCS, the requirement for awarding a degree is prescribed in terms of number of credits to be completed by the students.

The CBCS permits students to:

i.	Choose electives from a wide range of courses offered by the Departments of the College/University.
ii.	Opt for additional courses of interest
iii.	Adopt an inter-disciplinary approach in learning
iv.	Make the best use of expertise of the available faculty
v.	Skill Oriented Course and Multidisciplinary Course/Project are introduced as per Action Plan of National Education Policy-NEP-2020

- **Duration of the Program:**
  - The professional postgraduate programme leading to the degree of Master of Science in Data Science will extend 2 academic years/over a period of 4 semesters.
  - Each academic year being divided into two consecutive (one odd + one even) semesters
  - Each semester will normally have duration of 90 working days.
  - Minimum duration of the full-time M.S. Data Science Professional Programme is two c o n s e c u t i v e academic years i.e. four semesters and maximum period is four academic years, i.e. eight semesters.

3.Minimum Qualifications for Admission:

Candidates for admission into the 2first semester of the M.S. DATA SCIENCE

Professional programme must have a

- o B.Sc. : Computer Sciences| IT | Maths| Statistics| Physics| Chemistry| Electronics| Bio-technology| Bio-informatics
- o M.Sc.: Computer Sciences| IT | Maths| Statistics| Physics| Chemistry| Electronics| Bio-technology| Bio-informatics
- o B. Tech/ BE of all Streams , BBA/ MBA

of academic input of any university recognized by the Universities Grants Commission.

4. Admission Procedure

4.1. The first academic year admission into M.S. Data Science is through an online entrance examination which consisted of 50 multiple choice   Objective   type questions for 50 Marks. There is no negative marking for the wrong answers.

4.2   The selection process for the qualified students in the entrance examination, Personal Interview and UG/PG academic merit.

4.2   The short-listed candidates are called for counselling by Directorate of Admissions (DOA), SV University, for certificate verification, and first year course fee payment (through online portal https://pay.svupayonline.edu.in/.)

4.2   Admission into M.S. DATA SCIENCE Professional Programme for the academic year 2024-2025 will be through APPGCET Examination for Indian Students conducted by Andhra Pradesh State Council of Higher Education(APSCHE).

4.3   Admission of Foreign / NRI Students for M.S. DATA SCIENCE Professional Programme:

- o Students should have earned their qualifying degree from a University / Institute recognized by the Association of Indian Universities / similar Indian body.
- o Students nominated by Foreign Governments may be accepted to M.S. DATA SCIENCE Professional programme, without any further test / interview, if their request for admission is routed through MHRD / any agency of the Government of India.

4.4   The candidate at the time of admission should produce a Transfer certificate

(T.C.), Degree certificate, Marks statement or equivalent credit statements, Migration certificate and any other certificate as stipulated by the University.

4.5 Candidates should abide by the conditions, rules and regulations stipulated by the University / Exafluence Inc. and Laws of the University applicable from time to time.

4.6 For further information the candidate may contact the ACADEMIC INCHARGE - Exafluence Education Campus Centre,Next to Distance Education Centre Bldg. SV University,Tirupati – 517502.email ID: admissions@exafluence.com

Note: This MS programme is a self-financed course and is **not** applicable for benefits under schemes such as **Jagananna Vidya Deevena (JVD), Jagananna Vasathi Devana (JVaD) and other government scholarships.**

5. Structure of the Programme:

The M.S. DATA SCIENCE Professional programme has a curriculum, with syllabus consisting of:

- **Core courses**, which give a broad base in the main field of study in the academic programme concerned.

- **Elective courses** chosen by the student in consultation with the faculty adviser Elective course is a course which can be chosen from a pool of papers. It may be:
  o Supportive to the discipline of study
  o Providing an expanded scope
  o Enabling an exposure to some other discipline/domain
  o Nurturing student's proficiency/skill.

An elective may be "Generic Elective" focusing on those courses which add generic proficiency to the students. These electives shall be "Discipline centric".

- Second type elective may be open elective and shall be offered for other disciplines/ departments.

- **The Foundation Courses** may be of two kinds: Compulsory Foundation and Elective foundation. "Compulsory Foundation" courses are the courses based upon the content that leads to Knowledge enhancement. They are mandatory for all disciplines. Elective Foundation courses are value-based and are aimed at man-making education.

- **Skill Oriented Course** and Multidisciplinary Course/Project are introduced

as per Action Plan of National Education Policy-NEP-2020.

- o Laboratory, project work and software design courses where special emphasis is laid on the application of knowledge to real-life problems.
- o The program will also include involve lectures /tutorials/ lab work, Technical Seminars, Group discussions and practical Industry exposure training, as prescribed by the Board of Studies in the curriculum & syllabus and approved by the Academic Council.

- **Audit Course (100 Marks Internal – Zero Credits)**
  - o The main aim of Audit Course is too aware the Contemporary Societal issues. This course is under self-study.
  - o In addition, the students may have to choose audit courses, in consultation with the faculty adviser.

- **MOOCS and e-Learning:**

Discipline centric elective course through MOOCs (Massive Open Online Course) platform. Students can register for the course offered by authorized Institutions/Agencies through online with the approval of the course coordinator. The certificate issued by the Institutions/Agencies after successful completion of the course will be considered for the award of the Grade to that course in open electives only.

- **Open elective course offered by M.S. Data Science**

  a. Fundamentals of Programming

  b. Web Designing

6. The Credit System:

- Each course has a certain number of credits assigned to it, depending upon whether it is a lecture or tutorial, or practical work and the number of periods assigned per week. The practical courses shall include laboratory work and project work.

- **The credits are assigned according to the following pattern:**
  - 1 credit for each lecture period/week,
  - 1 credit for each tutorial period/week,
  - 1 credit for each practical session of 2 periods/week &
  - 2 credits for each project session of 3 periods / week.

7. Minimum Instruction Days:

- The normal duration of M.S. DATA SCIENCE Course is 4 semesters.
- Semesters, I, II, III, IV shall consist of a minimum of 90 instruction days (based on six instruction days per week) excluding the days allotted for internal assessment, semester end examinations and preparation holidays.
- 4th semester shall consist of a minimum of 15 weeks for undertaking major project work either in the College or any of the recognized Public Service Sectors/ Government Sectors, National Laboratories and Industrial Houses or any other organizations approved by the Department course coordinator & Exafluence Inc., USA.

8. Medium of Instruction and Examination:

The medium of instruction will be English, and the examination will be in English only.

9. Credits Required for Award of M.S. DATA SCIENCE Degree:

A student earns credits by passing courses every semester. A student, who has registered the M.S. DATA SCIENCE Professional programme, has to acquire 96 credits to become eligible for the award of the degree.

- It is mandatory for a student to complete successfully all the core courses pertaining to his/her specialization of study.
- A student may choose Generic Electives from the list of elective courses offered from his/her specialization of study.
- Further, a student may select from a list of Elective courses from other Departments as Open Electives to "suit the required" number of credits, such that the total credits is at least 96.
- There should be a register maintained by the course coordinator & Head of the Department indicating for each student, the course (s) registered by the student within the department, so that "Generic Electives" opted by the student are indicated.

- In the case of Open Elective, the course coordinator & Head of the Department should prepare a statement/register indicating the courses chosen/ opted by the students of the department in other departments.
- The Course coordinator (or) Head of the Department should send the list of registered papers (opted by the student) to the principal with a copy to the controller of examinations immediately within a week commencement of each semester.
- A copy of the courses registered by the students in each semester approved by Dept. Course coordinator shall be sent to the Academic Branch as well as Examination Branch.
- The list of students registered for MOOCs shall be furnished by dept. course coordinator giving details of the programme with a copy to the Principal and Controller of Examinations.
- A model of Registers to be maintained by the course coordinator & Head of the Department**.** It is mandatory on the part of the course coordinator & Head of the Department to maintain Register for each Course separately.

10. Academic Performance:

The performance of the students in each semester shall be evaluated paper wise.

The distribution of marks between sessional work (internal assessment) and University Examination is as follows:

I - SEMESTER (6 Core papers)

Code	Title of the Paper	Category of Courses	L	T	P	No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
MSP 101	Mathematics Foundation for Data Science	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 102	Statistics for Data Science	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 103	Python for Data Science	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 104	Fundamentals of Programming	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 105	Fundamentals of Database Systems: SQL & MongoDB	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 106	Data Visualization & Story Telling	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 107	Hands-on: Python	-	0	0	4	2	30	70	3 Hrs.	100
MSP 108	Hands -on: Data Visualization	-	0	0	4	2	30	70	3 Hrs.	100
Total						**28**				**800**

II - SEMESTER   (3 Core papers/2 Elective papers)

Code	Title of the Paper	Category of Courses	Teaching Hours Per Week			No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
			L	T	P					
MSP 201	Big Data Processing	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 202	Fundamentals of Machine Learning	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 203	Developing Web Applications	DSC	3	1	0	4	30	70	3 Hrs.	100
MSP 204	Elective – I	DSE	3	1	0	4	30	70	3 Hrs	100
MSP 205	Elective – II	DSE	3	1	0	4	30	70	3 Hrs	100
MSP 206	Hands-on: Elective -I	-	0	0	4	2	30	70	3 Hrs.	100
MSP 207	Hands-on: Elective -II	-	0	0	4	2	30	70	3 Hrs.	100
MSP 208	MongoDB Associate Developer Exam – MongoDB Certification	ABCC	3	1	0	4	-	-	-	-
Total						**28**				**700**

9

III - SEMESTER   (3 Core Papers/2 Elective Papers)

Code	Title of the Paper	Category of Courses	Teaching Hours Per Week			No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
			L	T	P					
MSP 301	Artificial Intelligence and Deep Learning	DSC	3	1	0	4	30	70	3 Hrs	100
MSP 302	Cloud Computing - AWS / Azure / GCP	DSC	3	1	0	4	30	70	3 Hrs	100
MSP 303	Soft Skills & Personality Development	SEC	3	1	0	4	30	70	3 Hrs	100
MSP 304	Elective – III	DSE	3	1	0	4	30	70	3 Hrs	100
MSP 305	Elective – IV	DSE	3	1	0	4	30	70	3 Hrs	100
OEC 306	Open Elective	OEC	3	1	0	4	30	70	3 Hrs	100
MSP 307	Hands-on: Elective – III	-	0	0	4	2	30	70	3 Hrs	100
MSP 308	Hands-on: Elective - IV	-	0	0	4	2	30	70	3 Hrs	100
Total						**28**				**800**

IV - SEMESTER (Project)

Code	Title of the Paper	Category of Courses	No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
MSP 401	Capstone Project	DSC	12	100	200	3 Hrs.	300
	Total		**12**				**300**

Total Credits	Total Marks
96	**2600**

SEMESTER II

Elective -I

Stream	Code	Title of the Paper	Category of Courses	Teaching Hours Per Week			No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
				L	T	P					
Data Science	MSP 204A	MLOps	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 204B	NLP Applications (Image Recognition, Speech-to-text)	DSE	3	1	0	4	30	70	3 Hrs.	100
Data Engineering	MSP 204C	Data Engineering with Python	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 204D	Learn AWS Data Engineering	DSE	3	1	0	4	30	70	3 Hrs.	100

* Student should select any one elective from each stream

SEMESTER II

Elective -II

Stream	Code	Title of the Paper	Category of Courses	Teaching Hours Per Week			No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
				L	T	P					
Data Science	MSP205A	Machine Learning & Prediction Algorithms	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP205B	Applied Machine Learning	DSE	3	1	0	4	30	70	3 Hrs.	100
Data Engineering	MSP205C	Data Engineering using Kafka and Spark Structured Streaming	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP205D	Data Engineering on Google Cloud platform	DSE	3	1	0	4	30	70	3 Hrs.	100

* Student should select any one elective from each stream

SEMESTER III

Elective -III

Stream	Code	Title of the Paper	Category of Courses	Teaching Hours Per Week			No. of Credits	Sessional Marks CIE	Semester End marks SEE	Exam Duration	Max Marks
				L	T	P					
Data Science	MSP 304 A	Data Science Tools	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 304 B	Data Science with PySpark	DSE	3	1	0	4	30	70	3 Hrs.	100
Data Engineering	MSP 304 C	Data Engineering using AWS Data Analytics	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 304 D	Data Engineering Essentials using SQL, Python, and PySpark	DSE	3	1	0	4	30	70	3 Hrs.	100

* Student should select any one elective from each stream

SEMESTER III

Elective IV

Stream	Code	Title of the Paper	Category	Teaching	No. of	Sessional	Semester	Exam	Max

			of Courses	Hours Per Week			Credits	Marks CIE	End marks SEE	Duration	Marks
				L	T	P					
Data Science	MSP 305 A	AI and Deep Learning through TensorFlow	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 305 B	Machine Learning Techniques	DSE	3	1	0	4	30	70	3 Hrs.	100
Data Engineering	MSP 305 C	Data Engineering for Beginner using Google Cloud & Python	DSE	3	1	0	4	30	70	3 Hrs.	100
	MSP 305 D	Data Engineering with Google Big Query & Google Cloud	DSE	3	1	0	4	30	70	3 Hrs.	100

*Student should select any one elective from each stream

Acronyms Expanded

- ➢ AECC: Ability Enhancement Compulsory Course
- ➢ DSC ©: Discipline Specific Core (Course)
- ➢ SEC-SB/VB: Skill Enhancement Course-Skill Based/Value Based
- ➢ OEC: Open Elective Course
- ➢ DSE: Discipline Specific Elective
- ➢ SEE: Semester End Examination
- ➢ CIE: Continuous Internal Evaluation
- ➢ L+T+P: Lecture + Tutorial + Practical(s)
- ➢ MSP: Master of Science Program

11. Evaluation of theory papers

11.1 Evaluation shall be done on a continuous basis i.e. through Continuous Internal Evaluation (CIE) in the Semester and Semester End Examination (SEE).

11.2 For both Theory and Laboratory papers two Internal Assessments will be conducted and best of two Assessments considered for award of Sessional marks. The duration of these tests will be 2 hours. These sessional marks are to be communicated to the Controller of Examinations Office on or before the commencement of the end Semester examination.

- o The first internal test shall be held immediately after the completion of 50% of the instruction days covering 50% of the syllabus. The second internal test shall be held immediately after the completion of 90 instruction days covering the remaining the 50% of the syllabus.

11.3 It is mandatory for a student to attend both the internal tests in each theory course.

11.4 The valuation of answer scripts of Sessional Tests shall be completed after the conduct of the internal Tests. The answer scripts shall be maintained in the department until the semester end results are announced.

11.5 The valuation of End-Semester Examination answer scripts shall be arranged by the Controller of Examinations as per the University procedures in vogue.

12. Evaluation of Practical

For each practical course, the sessional marks for a maximum of 100 shall be awarded by the teacher based on continuous assessment of practical work. The Semester End University Practical Examinations carrying 100 marks shall be conducted by i) Internal examiners and ii) external examiners permitted by the Board of Studies of the Department with a panel submitted to the Controller of examiners.

13. Project Work:

13.1 The work shall be carried out in the concerned department of the student or in any recognized Educational Institutions of Higher learning/Universities/Industry/ Exafluence Inc, Tirupati.

13.2 The student shall submit the outcome of the project work in the form of a report.

The Project work shall be evaluated at the end of the IV semester with 300 marks in which 100 marks are Internal and 200 marks are External evaluation of the project work.

Paper Category	Sessional Marks	University Examination Marks
Theory	30	70
Practical	30 Designing of the Program (10) + Program Execution (10) + Viva-Voce(10)	70 Record (20) + Designing of the Program (20) + Program Execution(20) + Viva Voce(10)
☐ Seminar / Technical Seminars / Group discussion	50 Content Preparation (12) + Communication Skills (12) + Presentation Skills (12) + Attempting Queries (14)	-
Minor Project Work	30 Presentation(15) + Viva Voce (15)	70 Documentation (20) + Presentation(20) + Execution(20) + Viva Voce (10)
Major Project Work	100 Project Seminar (Internal): 50 Marks Project Report (Internal): 50 Marks 18	200 1. National / International Conference Publication Proceedings (External): 50 Marks (Paper based on project should be submitted to conference and published in the form of proceedings) 2. National / International Journal Publication (External): 50 Marks (Paper based on project should be submitted to the journal and should be published) 3. Viva Voce (External): 50 Marks 4. Project Execution (External): 50 Marks

14. University Examinations:

14.1   For each theory subject, there shall be a comprehensive University Examination of three hours' duration. Setting of Question papers shall be done by external examiners from the panels recommended by the Board of Studies.

14.2   The semester end university examination in each course and paper is of three hours duration with a maximum of 70 marks. The Question paper consists of two parts, A & B. Part A consists of 10 short answer questions and 5 to be answered (5 X 4 = 20 Marks). Part B consists of 10 questions with choice from each unit(5 X 10 = 50 Marks).

14.3   (a) For each practical subject, the University Examination shall be conducted by one Internal examiner and one external examiner/two examiners.

(b) Viva -Voce examination in practical subject shall be conducted by Examiners (One Internal and One External Examiner/Two Examiners)

14.4   Viva-Voce Examination in Minor Project Work shall be conducted by One Internal Examiner and One External Examiner / Two Examiners

14.5   Viva-voce examination in Major Project Work shall be conducted by a committee consisting of Two External Examiners and One Internal Examiner / Three Examiners

14.6   (a) The Examiners for Practical examinations and Viva-voce examinations (Stated in 14.3) shall be appointed from among the panels submitted by the Chairperson/Chairman of the Board of Studies.

(b) The Examiners for Project report evaluation for Minor/Major Project Works and Viva Voce examination for Minor/Major Project Works (Stated in 14.4, 14.5) shall be appointed from among the panels submitted by the Chairperson/Chairman of the Board of Studies.

15. Attendance Requirements:

15.1   Regular course of study, in a semester, means a minimum average attendance of 75% in all the papers computed by totaling the number of periods of lectures, tutorials, and practical's and project as the case may be, held in every paper as the denominator and the total number of periods attended by the student in all papers, as the numerator.

15.2 However, a student has to put in a minimum attendance of 50% in each subject, in addition to the condition laid down in clause 15.1.

15.3 No consideration whatsoever in attendance will be shown to any student for late admission due to any reason.

15.4 Condonation of shortage of attendance may be recommended on medical grounds only provided a student puts in at least 62.5% attendance, in all the papers put together as calculated in clause 15.1 above, along with a minimum of 50% attendance in each subject as stated in clause 15.2.

15.5 A student, who could not satisfy these requirements of attendance as given in clauses above, in any semester, shall have to repeat that semester in the subsequent academic years with the written permission of the university.

15.6 A student shall not be permitted to study any semester for more than 2 times during the course of study.

15.7 A certificate of satisfactory attendance must be submitted by the student from the organization where he/she undertakes Project Work. For the above purpose, the candidate is not expected to avail more than 12 working days of leave of absence.

15.8 Further, a student is required to complete the course of study satisfying the attendance requirements in all the 4 semesters, within a period of first 8 semesters from the time of admission, failing which he/she shall forfeit his /her seat.

15.9 A student, who has not satisfied the minimum attendance requirements in any semester, may repeat that semester after obtaining written permission from the university, canceling the previous record of attendance and sessional marks of that semester. However, this facility may be availed by any student not more than twice during the entire course of study and the entire course of study shall be within the first eight semesters as stipulated in clause 15.8.

16   Academic Requirements:

16.1   A candidate shall be declared to have passed whole examination of a semester if he/she secures a minimum aggregate of 50% along with minimum marks of 40% in the University Examination in each theory and practical paper, including project work. Aggregate for this purpose shall mean the total marks obtained in the University Examination and sessional put together in all the papers of that semester.

16.2 A candidate who secures a minimum aggregate of 50% in any semester as specified in clause 16.1, but fails to secure the paper minimum of 40% in the University Examination in any paper, shall be declared to have failed in that paper. The candidate may appear for the University Examination in such papers as and when conducted and pass by securing the subject minimum of 40% in each of such papers.

16.3 A candidate, who could not secure the minimum aggregate of 50% in any semester, as given in clause 16.1, shall be declared to have exempted in such theory papers in which he/she secured a minimum of 50% of the marks in the University Examination.

16.4 A candidate who could not secure the minimum aggregate of 50% in any semester, as given in clause 16.1, shall be declared to have exempted in such practical papers in which he/she secured a minimum of 40% in the University Examination and a minimum of 50% of the marks in the University Examination and sessional put together in each of such papers.

16.5 A candidate has to appear for all the other papers in which he/she has not satisfied the stipulations of clauses 16.3 and 16.4, and pass compartmentally by satisfying the stipulations stated in clauses 16.3 and 16.4. Rechecking of the aggregate shall not arise when a candidate opts for compartmental passing.

16.6 Candidates who would not like to avail themselves of the facility of compartmental pass given in clauses 16.3, 16.4 and 16.5 may reappear for the whole examination of that semester, after canceling the previous result. In such a case, clauses 16.1 to 16.5 shall apply in that order.

16.7 Candidates shall register for improvement in one or more theory papers of a semester indicating the same in the application form. However, improvement facility is not extended to practical and project work papers.

NOTE: The question of checking the aggregate shall arise only when the candidate registers for the whole examinations for the first time or when he/she registers for the whole examinations at a subsequent attempt after having canceled the previous result.

17. Condition of Promotion:

A candidate shall be promoted to the next semester, if he/she satisfies the minimum attendance requirements of that semester of M.S. DATA SCIENCE as stipulated in clause 15.

18. Award of Degree:

The degree of M.S. DATA SCIENCE shall be conferred on a candidate who has satisfied the following conditions:

18.1 The candidate, after admission to the M.S. DATA SCIENCE Professional programme of the University, must have studied each of the 4 semesters at least once

18.2 The candidate must have satisfied minimum attendance requirements as stipulated in clause 15, and minimum academic requirements as prescribed in clause 16, in each of the 4 semesters of the M.S. DATA SCIENCE Professional programme.

18.3 The Programme of Study for the M.S. DATA SCIENCE shall cover normally a period of 2 academic years, comprising of 4 semesters and No student shall be permitted to complete the course of study of M.S. DATA SCIENCE earlier than 4 semesters or to take not more than 8 consecutive semesters, failing which he / she shall forfeits his / her seat (or) He/she will not be getting degree .

18.4 A student who has earned a minimum of 96 credits by passing in all the core courses and the minimum number of electives prescribed shall be declared to have passed the course work and shall be eligible for the award of degree.

18.5 A student who has earned extra credit shall be issued a separate certificate to that effect, mentioning the subject and grade.

19.0 Award Grades and Grade Points:

19.1 After a candidate has satisfied all requirements for the award of the degree as specified in clause 18, he/she shall be placed in one of the following three classifications, irrespective of whether the candidate passed compartmentally or otherwise, even after the regular period of study of four semesters.

Grade Point: It is a numerical weight allotted to each letter grade on a 10-point scale.

Letter Grade: It is an index of the performance of students in a said course. Grades are denoted by letters O, A+, A, B+, B, C, P and F.

Semester Grade Point Average (SGPA): It is a measure of performance of work done in a semester. It is ratio of total credit points secured by a student in various courses registered in a semester and the total course credits taken during that semester. It shall be expressed up to two decimal places

SGPA (Si) = $\sum(C_i \times G_i) / \sum C_i$

where C_i is the number of credits of the ith course and G_i is the grade point scored by the student in the ith course.

The **CGPA** is also calculated in the same manner taking into account all the courses undergone by a student over all the semesters of a programme, i.e.

Cumulative Grade Point Average (CGPA): It is a measure of overall cumulative performance of a student over all semesters. The CGPA is the ratio of total credit points secured by a student in various courses in all semesters and the sum of the total credits of all courses in all the semesters. It is expressed up to two decimal places

CGPA = $\sum(C_i \times S_i) / \sum C_i$

Where S_i is the SGPA of the ith semester and C_i is the total number of credits in that semester. The SGPA and CGPA shall be rounded off to 2 decimal points and reported in the transcripts

Letter Grades and Grade Points

A 10-point grading system with the following letter grades is to be followed:

S.no	Grade	Grade Points	Absolute Marks
1.	O (Outstanding)	10	90 and above
2.	A$^+$(Excellent)	9	80 to 89
3.	A (Very Good)	8	70 to 79
4.	B$^+$ (Good)	7	60 to 69
5.	B (Above Average)	6	50 to 59
6.	C (Average)	5	45 to 49
7.	P (Pass)	4	40 to 44
8.	F (Fail)	0	Less than 40
9.	Ab. (Absent)	0	-

A student obtaining Grade F shall be considered failed and will be required to reappear in the examination.

For non-credit courses 'Satisfactory' or "Unsatisfactory' shall be indicated instead of the letter grade and this will not be counted for the computation of SGPA/CGPA

For this purpose, aggregate shall mean the aggregate of the marks in the University Examinations and sessional put together at all 4 semester of M.S. DATA SCIENCE.

19.2 A Candidate, before and even after becoming eligible for the award of the Degree, may reappear for the University Examination, as and when conducted, in any of the theory subjects, which he/she has already passed, for the purpose of improving the aggregate. However, this facility cannot be availed by a candidate who has taken the Provisional Pass Certificate. Candidates shall not be permitted to appear for University examination in Practical papers including project work for the purpose of improvement.

20. Award of Rank:

The rank shall be awarded based on the following:

20.1 Ranks shall be awarded for the top ten students appearing for the regular M.S. DATA SCIENCE examination.

20.2 Only such candidates, who satisfy the minimum academic requirements as stipulated in clause 16, by the end of their 4th semester, and become eligible for the award of the degree, are eligible for the award of rank. Candidates, who lose one or more semesters for any reason whatever, are not eligible for the award of rank.

20.3 For the purpose of awarding rank, aggregate of marks in University Examination and Sessional put together at all the semesters of M.S. DATA SCIENCE, secured at the first attempt only, shall be taken into account.

20.4 Ranks shall be awarded in each branch of study on the basis of Cumulative Grade Point Average (CGPA) for top ten percent of the students or top three students whichever is higher.

20.5 The students who have become eligible for the award of M.S. Data Science by passing regularly   without break, shall only be considered for the award of ranks.

20.6 Award of Prizes, scholarships and other honors shall be according to the rank secured by a candidate, consistent with the desire of the Donor.

21. Transitory Regulations: 23

21.1 Candidates who studied the M.S. DATA SCIENCE Course under the   Regulations (Semester Pattern) and syllabus from 2023-2024 but who could not satisfy the minimum attendance requirements in any semester may join the appropriate semester in the Regulations (For the full time, regular course) – 2025-27 for the remaining part of the course and be governed by the Regulations of that batch from then on.

21.2 University Examinations shall be conducted 2 more times, after the last regular examination, in all those papers prescribed under the Regulations & syllabus from 2024-2025

21.3 Candidates, who satisfied the minimum attendance requirements in any semester under the Regulations & syllabus from 2024-2025 but who are yet to pass some papers of that semester even after Two chances shall appear for the equivalent papers under Regulations from time to time, specified by the Chairman / Chairperson of the Board of Studies.

22. Institutional Framework and Methodology:

22.1 Exafluence Inc. designed new and innovative, industry / job-oriented, international standard course curriculum, Academic Schedule, Regulations and the methodology of training along with the practical and procedural knowledge sharing practices to be conducted in hybrid mode as per the MOU with the S V University in the department of Computer Science SVU College of CM & CS Tirupati with appropriate approvals from APSCHE, Sri Venkateswara University & Department of Computer Science.

22.2 Exafluence Inc. will identify Industry partners, mentors, specialist trainers, certified trainers, experienced faculty, Guest faculties from the industry (Data Sciences | Machine Learning | Artificial Intelligence | Cloud Technologies Etc.) to execute the programme.

22.3 Exafluence Inc. has developed educational content, curetted course and curriculum to deliver, execution methodology, practical learning, Lab's setup, through its efforts for executing and conducting the programmes / courses in the SVU campus

in Hybrid mode internationally in association with partner international organizations.

22.4 Exafluence Inc. will coordinate with SVU in enrolment of students, collection of fees, submission of student rolls, supplying the study material to the students enrolled into the programme (offline and online), assisting in industry training, models and practices of evaluation and examinations process support to SVU and all other incidental activities thereto.

22.5 Exafluence Inc. will coordinate with SVU and the partner national and international organizations for international student's enrollments for the Hybrid course delivery.

22.6 Exafluence Inc. will provide all software's and courseware to execute the programme. Exafluence Inc. along with its international and national technology partners support the labs with relevant software and technology components and lab assistants. However, most of the Data Science technology work with the open-source components.

22.7 Exafluence Inc. will advertise and market the programme extensively though all mediums to propagate and get the right admission through the entrance examination and admission management system.

22.8 Exafluence Inc. will provide the necessary industry internship, project support and job assistance to at least 50% of the students on pro rata basis.

22.9 University will provide the necessary infrastructure facilities for the intended course to be conducted.

22.10 Students will be admitted into the Industry Collobarative programme to pursue M.S. Data Science, or any other programmes offered in per view of this MOU.

22.11 The Programme will be closely monitored by the department "course Coordinator" who is a University Professor from the respective faculty/ department and by an Academic Board consisting of both academicians and industry professionals.

22.12 SVU shall conduct period-end examinations and evaluate the enrolled students in such programmes/courses under this MOU. SVU will award DEGREES to all successful students.

23. FEE & REMUNERATION

23.1 The fee payable in each year is collected in one installment as detailed below:

Exafluence Inc. Programme	Tuition Fee Rs.	GST @ 18%	Total
II Years	3,00,000/-	54000/-	3,54,000/-
I year	150000/-	27000/-	1,77,000/-
II year	150000/-	27000/-	1,77,000/-

Note: The fee is subject to changes to be made by the University from time to time.

24. Remuneration / Honorarium and Allowances

a. Honorarium for Principal 3000 * $\frac{25}{10}$ rnonths = Rs 30000/-

b. Honorarium to Coordinator 3000 * 10 months = Rs 30000/-

c. Honorarium for AAO 1000 * 10 months = Rs 10000/-

d. Honorarium for Junior Assistant, Department of CS 850 * 10 months = Rs 8500/-

25. Amendments of Regulation:

The University may, from time to time, revise, amend, or change the Regulations, Scheme of Examinations and Syllabi, whenever necessary.

26. Careers

Data Scientists focus on helping their organization to use Data. They are experts in handling large quantum of data and are responsible for deriving business value. Data science has huge potential as career and with tremendous opportunities for advancement in the future. Already, demand is high, salaries are competitive, and the perks are numerous – which is why Data Scientist has been called "the most promising career".

Opportunities for Data Scientists:

Data Architect & Administrators, Data Engineer, Data Analyst, Data Scientist, Data Specialists, Statisticians and Mathematicians, Critical Data Managers, Data Mining Engineer, Business Intelligence Analyst.

Opportunities for Data Engineers:

Database Developer, Data Architect, Data Manager, ETL Developer, Big Data Developer, Business Intelligence Developer, Analytics Consultant, Data Engineer.

27. Internship & Industrial Training:

27. 2   The Industry will take students as "Interns" or "Trainees" for a

maximum period of 4 months and a minimum period of 2 Months.

27. 3   The Industry will select students who are ready for Industry Internship and then shortlist students for the purpose. The Programme Coordinator will make alternative arrangements for such students who are successful reaching the final semester but fail to meet the expectations of the industry.

27.4   The Industry may or may not pay any stipend to the students during the Internship period. If the industry is comfortable, it may plan for some incentives and other benefits when and where available.

27.5   The selected students will be under the direct control of the industry during the period of internship. A consultant/officer/executive will take care of the candidate and statutory requirements by a Flexi contract.

27.6   The Industry will render knowledge transfer on key practical areas to the students during the period of internship.

27.7   Each student will have a Project Guide in the Partner industry institution and or in the organization, where the student carries out his/her project work / internship.

27.8   The Industry will certify the project/work reports of the students which have to be submitted to the University as part of their course completion.

27.10 The "Interns" shall abide by all the terms and regulations of the   Internship Programme as stipulated by the industry

28. Important Instructions:

28.1 No refund of the fee will be made once remitted. There are no fee concessions for any category of students. The fee once paid by the candidate will neither be refunded nor adjusted. Please check the eligibility before obtaining the admission.

28.2 All the fee paid receipts must be preserved till the candidate completes the course and obtains the degree certificate

28.3 All matters relating to examinations / results / marks / certificates etc will be dealt

with only by the Controller of Examinations, Sri Venkateswara University, Tirupati.

28.4 For all legal matters the Jurisdiction is restricted to Tirupati city only.

28.5 Candidates are required to abide by the rules and regulations that are in force and those that will come into effect from time to time as formulated for M.S. Data Science by the Industry Collaborative Programme by Exafluence Inc and /or by the University.

28.6 Admission at any time may be cancelled by the coordinator/ Head/ Principal/ DOA - Industry Collaborative Programme if the candidates is found ineligible for admission into the said course or he/she is found to be guilty of indiscipline.

Website information:

www.svuniversity.edu.in & https://exafluence.education/ will give all the information about the M.S. Data Science Professional programme schedules,Syllabus,Regulations, academic calendar & Model Question Paper.

Visiting Faculty:

Faculty visits from industry as per their specializations will be organized strictly. The Programme schedule of the visiting faculty from industry with personal details, Years of experience, teaching experience, Telephone, E-mail along with their photos should be submitted to the department course Coordinator, S.V. University at regular intervals

Industry connectivity:

Exafluence Inc. should appoint a "placement and internship officer/coordinator" who should work along with SV University for achieving 100% internships and placements only for M.S Data Science students.

Communication:

Any communication to the University by the Exafluence Inc. should be through the department course coordinator only.

MS in Data Science (MSDS)

Vision:

"Our vision is to provide a holistic data science education, fostering a learning community that empowers students to pursue their professional goals with confidence and vision. We are committed to quality, innovation, and excellence, thereby enhancing our students' employability and enabling them to become trailblazers in the tech industry."

Mission:

Our mission is to equip the next cohort of Data Science professionals with ML & AI concepts and the skills, knowledge, and practical experience essential to excel in the dynamic and rapidly-evolving realm of data science domain. We aim to provide our students with a comprehensive education that accentuates critical thinking, problem-solving, and innovation, through real-time projects. Our unwavering commitment to offering students unparalleled opportunities to gain hands-on work experience, combined with our emphasis on practical experience and exposure to real-world challenges, ensures that our students are well-equipped to meet the ever-evolving demands of the dynamic field of data science.

Programme Educational Objectives:

The student will gain

PEO1:

Knowledge and Skills: Provide students with a strong foundation in the theories, principles, and technical skills of data science, enabling them to analyze complex data sets using various tools and techniques.

PEO2:

Problem Solving: Develop students' critical thinking and creative problem-solving skills to enable them to make informed decisions and solve complex data-driven problems across different domains and industries.

PEO3:

Leadership and Ethics: Prepare students for leadership roles in the field of data science and instil a commitment to ethical and responsible practices in data collection, management, and use.

PEO4:

Lifelong Learning: Prepare students to adapt to emerging technologies and tools in data science and continue learning throughout their careers.

Programme Outcomes:

PO1. Data Science Foundations: Provide students with a strong foundation in data science concepts and tools, including statistical methods, machine learning, and data visualization.

PO2. Data Engineering Fundamentals: Provide students with a strong foundation in data engineering concepts and tools, including data management, data integration, and data

processing.

PO3. Advanced Data Science Techniques: Enable students to analyze complex data sets using advanced data science techniques, such as deep learning and natural language processing.

PO4. Scalable Data Engineering Techniques: Enable students to design and implement scalable data architectures and pipelines using advanced data engineering techniques, such as big data technologies.

PO5. Domain-Specific Applications: Enable students to apply data science and data engineering techniques to solve real-world problems in different domains, such as healthcare, finance, and marketing.

PO6. Communication and Collaboration: Develop students' communication and collaboration skills, enabling them to work effectively in multidisciplinary teams and communicate results to technical and non-technical stakeholders.

PO7. Leadership and Ethics: Prepare students for leadership roles in the field of data science and instill a commitment to ethical and responsible practices in data collection, management, and use.

PO8. Data Science Project Management: Enable students to manage end-to-end data science and data engineering projects, including problem formulation, data collection, analysis, and interpretation.

PO9. Emerging Technologies: Prepare students to adapt to emerging technologies and tools in data science and data engineering, enabling them to stay up-to-date with the latest trends and innovations.

PO10. Lifelong Learning: Foster a commitment to lifelong learning and professional development, enabling students to continue learning and growing throughout their careers.

Programme Specific Outcomes:

Graduates of the program will be able

PSO 1: To design, develop, and implement data architectures and pipelines that can handle large and complex datasets, using various data engineering tools and techniques.

PSO 2: To apply data engineering and data science techniques to solve complex problems in different domains, such as healthcare, finance, and marketing.

PSO 3: To communicate effectively about data engineering and data science concepts and results to a range of audiences, including technical and non-technical stakeholders. They will also be able to collaborate effectively in multidisciplinary teams to achieve project goals.

PSO 4: To design and implement secure and efficient data management systems that ensure data integrity and privacy, using various data engineering tools and techniques. They will also

be able to apply ethical and responsible practices in data collection, management, and use, while complying with regulatory and legal requirements.

Academic Schedule

	Department of Computer Science **S.V.U College of Commerce, Management & Computer Science SRI VENKATESWARA UNIVERSITY : TIRUPATHI**	

M.S. DATA SCIENCE

(An Industry Collaborative Programme with Exafluence Inc, USA in Hybrid Mode (Online and Offline))

Tentative Academic Schedule for The Batch 2024-2026

Academic Year	AY 2024 – 2025		AY 2025 - 2026	
	I Semester	**II Semester**	**III Semester**	**IV Semester**
Commencement of Class Work	17.01.2024	18.07.2024	12.12.2024	16.06.2025
Last Day of Instruction	22.06.2024	20.11.2024	11.04.2025	06.10.2025
I Sessional Examination	14.03.2024 – 20.03.2024	17.09.2024 - 23.08.2024	11.02.2025 - 17.02.2025	-
II Sessional Examination	24.06.2024 – 29.06.2024	21.11.2024 – 26.11.2024	15.04.2025- 21.02.2025	-
University Theory Examination	Proposed Dates 03.07.2024 - 09.07.2024	Proposed Dates 02.12.2024- 06.12.2024	Proposed Dates 02.06.2025 - 06.06.2025	-
University Practical Examination	Proposed Dates 10.07.2024 – 16.07.2024	Proposed Dates 07.12.2024 – 11.12.2024	Proposed Dates 09.06.2025 - 14.06.2025	07.10.2025 - 16.10.2025 Major Project work Viva -Voce

Vacations

	2024-2025	**2025-2026**
Pongal	13th Jan – 16th Jan	11th Jan – 15th Jan
Summer	1ST May 2024 – 31st May 2024	1st May 2025 – 31st May 2025
Dussehra	11th Oct – 13th Oct	1ST Oct - 2nd October

University Examinations Schedule

Academic Year	**2024 – 2025**	**2025 - 2026**

	03.07.2024	02.12.2024	02.06.2025	-
	04.07.2024	03.12.2024	03.06.2025	-
Theory Examinations	05.07.2024	04.12.2024	04.06.2025	-
	06.07.2024	05.12.2024	05.06.2025	-
	08.07.2024	06.12.2024	06.06.2025	-
	09.07.2024			

Working Days (Month Wise)

Academic Year	2024 – 2025		2025 - 2026	
	I Semester	**II Semester**	**III Semester**	**IV Semester**
Working Days	Jan 24 (12)	July 24 (12)	Dec 24 (16)	June 25 (13)
	Feb 24 (24)	Aug24 (23)	Jan 25 (23)	July 25 (26)
	March 24 (16)	Sep24 (16)	Feb 25 (18)	Aug 25 (23)
	April 24 (21)	Oct 24 (23)	March 25 (24)	Sep 25 (25)
	May 24 (0)	Nov24 (16)	April 25 (09)	Oct 25 (03)
	June 24 (17)			
Total days	**90 days**	**90 days**	**90 days**	**90 days**

UNIT WISE DETAILS
MSP 101-Mathematics Foundation for Data Science
(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Mathematics Foundation for Data Science	**Course Code**	MSP 101
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Set Theory - Number system, Sets and their operations, Relations and functions - Relations and their types, Functions and their types. Straight Lines - Slope of a line, Parallel and perpendicular lines, Representations of a Line, General equations of a line, Straight-line fit Quadratic Functions - Quadratic functions, Minima, maxima, vertex, and slope, Quadratic Equations. Algebra of Polynomials - Addition, subtraction, multiplication, and division.

Unit II:

Functions - Horizontal and vertical line tests, Exponential functions, Composite functions, Inverse functions. Logarithmic Functions - Properties, Graphs, Exponential equations, Logarithmic equations. Sequence and Limits - Function of One variable - Function of one variable Graphs and Tangents, Limits for sequences, Derivatives, Tangents and Critical points - Differentiability and the derivative, Critical points: local maxima and minima.

Unit III:

Graph Theory - Representation of graphs, Breadth-first search, Depth-first search, Complexity of BFS and DFS, Graph theory Algorithms - Single-source shortest paths, Dijkstra's algorithm, Bellman-Ford algorithm, All-pairs shortest paths, Minimum cost spanning trees, Prim's algorithm, Kruskal's algorithm.

Unit IV:

Vector and matrices - Vectors; Matrices; Systems of Linear Equations. Linear dependence; Basis and dimension rank/dimension for a vector space; Rank and dimension using Gaussian elimination. Rank and Nullity of a matrix; Introduction to Linear transformation - The null space of a matrix. linear mapping, linear transformation.

Unit V:

Orthogonality, Orthonormality; Gram-Schmidt method - Orthogonality and linear independence; Projections using inner products; The Gram-Schmidt process; Orthogonal transformations and rotations, continuity and directional derivatives - Multivariable functions: visualization; Partial derivatives; Directional derivatives; Limits for scalar-valued multivariable functions; Continuity for multivariable functions.

Prescribed Books:

1. "Introductory Algebra: a real-world approach (4th Edition)" - by Ignacio Bello

2. "Mathematical Methods" by Dr.T.K.V. Ivengar, Dr.B. Krishna Gandhi, Dr.S. Ranganatham by S.Chand publications 6th revised edition 2011.

3. "Quantitative Techniques" by C. Satyadevi.

MSP 102- Statistics for Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Statistics for Data Science	**Course Code**	MSP 102
L-T-P	3-1-0	**Credits**	4

Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction and type of data - Types of data, Descriptive and Inferential statistics, Scales of measurement. Measures of central tendency - Mean, median and mode, Quartiles and percentiles, Measures of dispersion - Range, variance, standard deviation and IQR, five number summary. Association between two variables - Using relative frequencies in contingency tables, Association between two numerical variables - Scatterplot, covariance, Pearson correlation coefficient, Point bi-serial correlation coefficient.

Unit II:

Basic principles of counting and factorial concepts - Addition rule of counting, Multiplication rule of counting, Factorials. Permutations and combinations. Probability Basic definitions of probability, Events, Properties of probability. Conditional probability - Multiplication rule, Independence, Law of total probability, Bayes' theorem.

Unit III:

Random Variables - Random experiment, sample space and random variable, Discrete and continuous random variable, Probability mass function, Cumulative density function. Expectation and Variance - Variance and standard deviation of a discrete random variable. Binomial and Poisson random variables - Bernoulli trials, Independent and identically distributed random variable, Binomial random variable, Uniform distribution, Exponential distribution.

Unit IV:

Multiple random variables - Two random variables, Multiple random variables and distributions. Multiple random variables - Independence, Functions of random variables - Visualization, functions of multiple random variables. Expectations Casino math, expected value of a random variable, scatter plots and spread, Variance and standard deviation, Covariance and correlation, Inequalities.

Unit V:

Multiple continuous random variables - Height and weight data, two continuous random variables, Averages of random variables - Colab illustration, Limit theorems, Models with multiple random variables, dependency, Models for IPL powerplay, Models from data. Estimation and Inference - Bayesian estimation, Hypothesis testing.

References:

1. Introductory Statistics (10th Edition) - ISBN 9780321989178, by Neil A. Weiss published by Pearson.
2. Introductory Statistics (4th Edition) - by Sheldon M. Ross.

MSP 103-Python for Data Science

(Semester–1 of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Python for Data Science	**Course Code**	MSP 103
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs

SEE Marks	70	CIE Marks	30

Unit I:

Intro to Python Programming - Flowcharts, Data Types, Operators, Conditional Statements & Loops, Functions & Recursion, Strings, In-built Data Structures - List, Tuple, Dictionary, Set.

Unit II:

Advanced Python concepts - Lambda Functions, List Comprehension, Functional Programming, Decorator, Args, Kwargs, Object Oriented Programming, Exception Handling, Modules, Package, Library, Built-in Modules in Python, Basic DSA & Problem Solving, Time complexity, List, 2D List, Bit Manipulation, Strings, Searching, Sorting.

Unit III:

Introduction to Data Science - Different Sectors Using Data Science, Purpose and Components of Python, Data Analytics Overview: Data Analytics Process Knowledge Check, EDA-Quantitative Technique EDA - Graphical Technique, Data Analytics Conclusion or Predictions Data Analytics Communication. Data Types for Plotting Data Types and Plotting.

Unit IV:

Python for DS - NumPy, Pandas, Data Visualization using Matplotlib and Seaborn Regular Expressions/Pattern Matching, Machine Learning with Scikit–Learn, Natural Language Processing with Scikit Learn, Data Visualization in Python using matplotlib, Web Scraping with Beautiful Soup, Python integration with Hadoop MapReduce and Spark, Terminal/OS, Git and GitHub, File Handling.

Unit V:

Python with Data structures - Complexity, Notations, Sorting and Searching Algorithms, Arrays, Lists, Stacks, Queues, Hashing, Graph Algorithms, Priority Queue, Heap, Binary Search Tree, Balanced Search Tree, Greedy Algorithms, Divide and Conquer, Dynamic Programming, String or Pattern Matching Algorithms, Network Flows, Linear Programming, Class of Algorithms.

References:

1. "PYTHON DATA SCIENCE HANDBOOK: Essential Tools for Working with Data" by Jake VanderPlas.

2. "Python for Data Science for Dummies" by John Paul Mueller and Luca Massaron.

3. "Data Science from Scratch: First Principles with Python" by Joel Grus.

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Programming	**Course Code**	MSP 103
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I

Basics of PL: What is a program, Programming Language, types of Programming Languages, Areas of Programming Language, Language Translators: Compiler, Interpreter, Assembler, Introduction to IDE, Types of IDEs.

Unit II

Data: Variables, Data Types, Constants, Advanced data types, Operators usage, Arrays: Definition, declaration & initialization, Types of arrays, matrices, types of matrices, usage of matrices in PL.

Unit III

Structured Programming: Sequential PL vs Structured PL, Conditional Branching statements: Simple If, If-else, If-else, If ladder, Nested if else, Iterative statements: for loop, while, do-while, for each, Switch statement, break, continue, exit.

Unit IV

Modular Programming: Introduction to Modular Programming, Divide & Conquer Methods, Structure of function, Types of functions, Scope of variable, Global variables, Pointers, Call by value, Call by reference, Recursion.

Unit V

Object Oriented Programming: Structured Programming Vs Object-Oriented Programming, Objects, Class, Properties, Methods, Structure of a class, Usage of Methods of an Object, Object Oriented Programming principles: Inheritance, Polymorphism, Data Abstraction, Constructors: Method Vs Constructor, Types of Constructors.

References:

1. Fundamentals of Programming Languages by E. Horowitz
2. Programming Fundamentals: A Modular Structured Approach, 2nd Edition by Kenneth Leroy Busbee and Dave Braunschweig

MSP 105-Fundamentals of Database Systems: SQL & MongoDB

(**Semester-**I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Database Systems: SQL & MongoDB	Course Code	MSP 105
L–T–P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Types of DB - Relational, Non-relational, ER diagrams, SQL Commands - DDL, DML, DCL, Aggregate Functions, Joins, SubQueries, Normalization, Scaling patterns, ACID Properties, Relational Query Languages and Database Design.

Unit II:

Relational Query Languages and Database Design - Storage Management, Indexing and Hashing, Transactions, Backup and Recovery, Query Optimization, SQL Injection, DaskSQL, CloudSQL (Athena/Big Query).

Unit III:

Introduction to MongoDB – Databases, Collections, Architecture and Installation, Schema Design and Data Modelling, CRUD Operations – Insert, Query, Delete, Update Documents, Bulk Write Operations, Retriable Writes & Reads.

Unit IV:

Indexing - Need of Index, types of Indices, Aggregation Framework: Aggregation stages, Aggregation operations, Map Reduce, MongoDB Security – Authentication, Role Base d Access Control, Encryption, Auditing.

Unit V:

MongoDB on the Cloud – Intro to MongoDB Atlas, Atlas account set up, Usage of Mongo Cluster – Serverless, Dedicated, Shared, Deploy free tier Cluster, MongoDB Atlas Pricing, Mongo Realm.

References:

1." Database Management Systems" - Abraham Silber Schatz, Henry F. Korth, S. Sudarshan

2. "MongoDB: The Definitive Guide 1st Edition" by Kristina Chodorow and Michael Dirolf

3. "MongoDB in Action" by Kyle Banker

4. "NoSQL with MongoDB in 24 Hours, Sams Teach Yourself 1st Edition" by Brad Dayley

MSP 106- Data Visualization & Story Telling
(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Data Visualization & Story Telling	Course Code	MSP 106
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Visualization **-** Visualization and its key topics, Visualization Basics, Exploratory Vs Explanatory Analysis, Decluttering and Highlighting.   Multivariant Visualization Method - Geometric Representations, Icon Representation, Pixel-Oriented Representations, Types of Graphs.

Unit II:

Data visualization tools - Fundamentals of Data Visualization using Excel, fundamentals of Data Visualization using Tableau, fundamentals of Data Visualization using PowerBI. Plots - Area plots, Histograms, Bar Charts, Pie Charts, Box Plots, Scatter Plots.

Unit III:

Advanced Visualizations and Geospatial Data - Waffle Charts, Word Clouds, Seaborn and Regression Plots, Maps with Markers.

Unit IV:

Getting Started in Effective and Ineffective Visuals - The Human Brain and Data Visualization, Cognitive vs Perceptual Design Distinction, Tableau Desktop Vs Tableau Public. Dive into Visualization - Map Abstraction, Visual encoding of data, Visualization design, Clutter and Visual Perception, Visualization Aesthetics, Enhancing the Visualization.

Unit V:

Design for Understanding - Design for Purpose, Data, Relationships and Design, Static Vs Interactive Visualization, Multiple Connected Views, Language, Labelling and Scaling.Assignment - Tell your Great Story with Data.

References:

1. "Storytelling with Data: A Data Visualization Guide for Business Professionals" by

    Cole Nussbaumer Knaflic

2. "Fundamentals of Data Visualization: A Primer on Making Informative and

    Compelling Figures" by Claus O. Wilke

3. "Data Visualization: A Practical Introduction" by Kieran Healy and Kieran Joseph

    Healy

SEMESTER II
MSP 201- Big Data Processing
(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Big Data Processing	Course Code	MSP 201
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Big data Basics – Big Data definition, Sources of Big Data, 3 Vs of Big Data, Big Data Characteristics, Applications of Big Data, Big Data Benefits and Challenges, Big Data Engineering, Fundamentals of Hadoop - Introduction, Architecture, Ecosystem, Components.

Unit II:

Data preprocessing – Types of Data, Sources of Data, Stages of Data Processing, From Data Processing to Analytics, Big Data Processing – Types of Big Data, Big Data Sources, Big Data Storage, Big Data Architecture, Stages of Big Data Processing.

Unit III:

Big Data Processing and Distributed Systems - Big Data Processing Vs Distribution Systems, Batch processing, Batch processing with MapReduce, A simple MapReduce example, MapReduce vs classic divide-and-conquer approach.

Unit IV:

Real Time Processing - Real-time processing, Examples of Real time Processing, Real-time faces, Streaming data processing – Data Streaming, Examples of Data Streaming Platforms, SCV principle.

Unit V:

Future of Data Processing - The future of data processing, Challenges and new possibilities in big data preprocessing, from data processing to analytics, Big Data Analytics – Uses and Examples of Big Data Analytics.

References:

1. "Big Data: A Tutorial-Based Approach" by Nasir Raheem.

2. "Modern Big Data Processing with Hadoop" by V. Naresh Kumar, Prashant Shindgikar.

MSP 202-Fundamentals of Machine Learning
(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Machine Learning	**Course Code**	MSP 202
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Basics - Intro to Machine learning, Applications of Machine learning, Data mining process - Business understanding, Data understanding, Data preparation, Modelling, Evaluation, Deployment, Supervised (Classification and Regression) and Unsupervised learning techniques.

Unit II:

Exploratory Data Analysis - Deriving insights, Data summary, distributions of attributes, Univariate, Bivariate, Multivariate, Interactions, Visualization using Seaborn and Matplotlib, Statistical testing, Imbalanced data and challenges.

Unit III:

Feature Engineering – Creating new features, removing features, Data Pre-Processing – Cleaning, Type conversions, Null detection and treatment, Outlier detection and treatment Transformation, Scaling, Encoding (Label and One-hot).

Unit IV:

Dimension reduction - Principal Component Analysis, Factor Analysis, Unsupervised Learning: Clustering - Agglomerative, Hierarchical Cluster Profiling, Cluster visualization, Applications of clustering.

Unit V:

Supervised Learning – Intro to Supervised Learning, Classification problem, Decision tree (binary and multi-class classifiers), Decision tree implementation using SKLearn, Model Performance metrics – Classification, K-fold cross validation, Feature importance.

References:

1. "The Hundred-Page Machine Learning Book" by Andriy Burkov.

2. "Machine Learning for Absolute Beginners" by Oliver Theobald.

3. "Deep Learning" by Ian Goodfellow, Yoshua Bengio and Aaron Courville.

MSP 203-Developing Web Applications
(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Developing Web Analytics	**Course Code**	MSP 203
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs

SEE Marks	70	**CIE Marks**	30

Unit I:

Internet Basics - Protocols, Servers and their Functions, Internet Clients, Network Security, Internet Development, Design Functional Internet site & Business Concepts.HTML - Fundamentals/ Basic HTML, Text formatting on Web Pages, incorporate images, creating hyperlinks, complex image maps, tables and nested tables, inserting web page, Setting & modifying field properties, Validating HTML

Unit II:

Cascading Style Sheet (CSS) - Introduction, Designing with Style Sheets, Style Sheet Syntax, ID, Class Contextual Selectors, Cascading Order, Properties, Absolute and Relative Positioning, Layering Elements using Z-Index, Animating objects. Extensible Markup Language (XML) - Introduction to XML, Benefits, Holding Data, Separates Structure from Formatting, Data Sharing XM.

Unit III:

Document Type Definition (DTD) - Introduction, DTD Building Blocks, Elements & Attributes, Entities, Validation. PHP Hypertext Preprocessor (PHP) **-**Introduction, PHP Document, Language Fundamentals, Decision Making Statement, Loops, Statements, Operators, PHP functions, Arrays & Functions, String Functions, Server-Side Processing, Processing Forms via GET/POST, State and Persistence, Web Application Development, Introduction to PHP Frameworks & Basic OOP.

Unit IV:

Basics of JavaScript - Syntax, Comments, Variables, Operators, Data Types, Functions, Objects, Events, Strings, Arrays, Dates, Control Structures, Sets, Maps, Regexp. Advanced JavaScript - JS Objects, JS classes, JS HTML DOM, JS Browser BOM, JS web APIs, JS Ajax, JS JSON. Introduction to Web Frontend.

Unit V:

SQL & MySQL **-** Creating Databases & Tables, using keys, Types of Tables in MySQL, Data Types, deleting databases and tables, Inserting, Retrieving, Updating & Deleting data, User Accounts, Access Control & documentation. Integrating PHP and MySQL - PHP interfacing with MySQL, connecting to MySQL, Database Connecting, Executing SQL, Retrieving the data set & refining the fetch.

44

References:

1. Web Design with HTML, CSS, JavaScript and jQuery Set 1st Edition by Jon Duckett.

2. JavaScript: The Definitive Guide: Master the World's Most-Used Programming Language 7th Edition by David Flanagan.

3. Fullstack Vue: The Complete Guide to Vue.js 1st Edition by Hassan Djirdeh.

MSP 301 - Artificial Inteliigence and Deep Learning

(Semester-III of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Artificial Intelligence and Deep Learning	Course Code	MSP 301
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Machine Learning - Intro to ML, Types of Machine Learning algorithms- Supervised, Unsupervised, Reinforcement, ML Applications, ML Life Cycle, challenges in Machine Learning, ML vs AI, ML vs Deep Learning.

Unit II:

Intro to AI - Basics of Artificial Intelligence, Applications of AI, Types of AI, AI agents, Artificial Neural Networks, Fuzzy Logic, Introduction to Genetic Algorithm, Soft Computing.

Unit III:

Intro to Deep Learning - Fundamentals of Deep Learning, Architecture of Deep Learning, Types of Deep Learning Networks – FFNN, RNN, CNN, Restricted Boltzmann Machine, Autoencoders, Deep Learning Applications, Limitations, Advantages, Disadvantages.

Unit IV:

Introduction to AI Networks with Keras **-** Biological neurons, multilayer perceptron's, and back-propagation, multilayer perceptron using Keras and visualize the runs and graphs using Tensorboard, Training Deep Neural Networks - Reusing pre-trained layers, using faster optimizers and avoiding overfitting by regularization.

Unit V:

Processing Sequences using RNNs & CNNs - Recurrent Neural Networks - RNN predict the future, the problem they face like limited short-term memory and solutions to these problems - LSTM (Long Short-Term Memory) and GRU cells, NLP Concepts – NLP Techniques.

References:

1. S. Russel and P. Norvig, "Artificial Intelligence – A Modern Approach", Pearson Education.
2. B. Yegnanarayana, Artificial Neural Networks, Prentice Hall of India.
3. Neural Networks and Fuzzy Logic System by Bart Kosko, PHI Publications.

MSP 302 - Cloud Computing - AWS / Azure / GCP

(Semester-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Cloud Computing - AWS / Azure / GCP	**Course Code**	MSP 302
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:
Basics of Cloud Computing – Intro to Cloud Computing, Architecture, Cloud Technologies, Types of Cloud, Cloud Service Models, Cloud Service Providers, Cloud Server, Cloud Computing Applications, Security Risks of Cloud Computing.

Unit II:

Introduction to AWS – AWS Infrastructure, AWS Free tier, AWS IAM – User, Roles, Policies, EC2 – AWS EC2 & Pricing, EBS, EBS Volume, Security Group, AWS AMI, AWS Load Balancing - Create Load Balancer, Types of Load Balancers – Application Load Balancer, Network Load Balancer, Classic Load Balancer, Auto Scaling, Cloud Watch.

Unit III:
AWS Storage Service – S3 Buckets, Objects, Keys, Regions, Create S3 Bucket, AWS Versioning, Lifecycle Management, S3 Transfer, Application Services – AWS SQS, AWS SNS, API Gateway, AWS VPC – Custom VPC, NAT Gateways, AWS Route S3, AWS SFTP – Create Server, Users, File Transfer with SFTP, AWS EKS, AWS Lambda Function.

Unit IV:

Basics of GCP **-** Intro to GCP, Building Blocks – Compute, Storage, Network, Identity Management, Services – Databases, Data & Analytics, AI & ML, Developers Tools, Enterprise Services.

Unit V:

Basics of Azure – Intro to Azure, Azure Portal, Storage Services – Azure Storage account, Blob Storage, Storage Security, File Storage, Storage Monitoring, Compute Services – Azure VM Storage, VM Security, Monitoring, Network Services - Virtual Network, Network Security, Availability Zones & Sets, Azure Load Balancer, App Services – Azure Web App, Azure Mobile App, Azure API.

References:

1. "Cloud Computing for Dummies" by Judith S. Hurwitz, Fern Halper, Robin Bloor, Marcia Kaufman.

2. "AWS: The Complete Beginner's Guide" by Stephen Baron.

MSP 303-Soft Skills & Personality Development
(Semester-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Soft Skill & Personality Development	**Course Code**	MSP 303
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit-I

Introduction to Soft-Skills-Personal Skills: Knowing Oneself/Self-Discovery-Confidence Building- Defining Strengths- Developing Positive Attitude- Thinking Creatively-Improving Perceptions -Forming Values.

Unit-II

Interpersonal and Social Skills: Understanding Others-Developing Inter-personal relation-ship- Team Building-Group Dynamics-Networking-Problem-solving.

Unit-III

Communication Skills: Art of Listening-Art of Speaking-Art of Reading-Art of Writing-Art of Writing E-mails: Email etiquette

Unit-IV

Corporate Skills: Working with others- Developing a proper body language-behavioral etiquettes and mannerism- Time Management –Stress Management

Unit-V

Job-hunting skills: Writing Resume/CV- Interview Skills-Group Discussion-Mock interview-Mock GD-Goal Setting-Career Planning

References:

1. Meena K and V. Ayothi (2013) A Book on Development of Soft Skills (Soft Skills: A Road Map to Success), P. R. Publishers & Distributors, No. B-20 & 21, V. M. M Complex, Chatiram Bus Stand, Tiruchirappalli-620002. (Phone No: 0431-2702824Mobile No.: 9443370597, 9843074472)
2. Alex K. (2012) Soft Skills-Know Yourself & Know the World, S. Chand & Company LTD, Ram Nagar, New Delhi-110055. Mobile No.: 9442514814 (Dr.K.Alex)

SEMESTER II
48
ELECTIVE-I

MSP 204 A-MLOps

(Elective-I of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	MLOps	**Course Code**	MSP 204 A
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Basics – Intro to MLOps, Configuring Continuous Integration with GitHub Actions, DataOps and Data Engineering, Platform Automaton, Developing Machine Learning Model, Deployment patterns, Model CI/CD (Build, Test, Integration and Delivery of model).

Unit II:

MLOps Foundations – Maturity Levels in MLOps, MLOps Tools, Cloud Shell Development Environments, MLOps for Containers, MLOps for Edge Devices.

Unit III:

MLOps Life cycle - Introduction to MLOps Model life cycle management, Adaptation of DevOps for building and deploying machine learning systems, Model Deployment- Infrastructure Requirements.

Unit IV:

Models - Developing and Deploying Models, Model Assessment, Continuous Delivery for ML Models, Trends in Model deployment - ML on the Cloud / Edge / Browsers; VMs, Containers, Docker, Kubernetes (K8S), FaSS; ML-as-a-Service.

Unit V:

MLOps for AWS - AWS Lambda, MLOps for GCP, DataOps on GCP - Building MLOps Command Line Tools and Microservices, MLOps for Azure – Orchestrated ML Codes in Azure, CI/CD MLOps Pipeline in Azure.

References:

1. "How to Scale Machine Learning in the Enterprise" by Mark Treveil & the Dataiku Team.

2. "What is MLOps" by Mark Treveil, Lynn Heidmann.

3. "Practical MLOps" by Noah Gift & Alfredo Deza.

MSP 204 B- NLP Applications (Image Recognition, Speech-to-text)
(Elective-I of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	MLOps	**Course Code**	MSP 204 B
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Introduction - NLP tasks in syntax, semantics, and pragmatics. Applications such as information extraction, question answering, and machine translation, The problem of ambiguity, The role of machine learning. Brief history of the field, N- Gram Language Models, Simple N-gram models, estimating parameters and smoothing., Evaluating language models. Part Of Speech Tagging and Sequence Labelling.

Unit II:

Neural Networks - Basic introduction to perceptron, Multilayer Neural Network, Gradient Descent learning, Back propagation, Empirical Risk Minimization, regularization, Radial Basis Neural Network, Back propagation through time, Long Short-Term Memory, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs.

Unit III:

Syntactic Parsing - Grammar formalisms and treebanks, Efficient parsing for context-free grammars (CFGs), Statistical parsing and probabilistic CFGs (PCFGs), Lexicalized PCFGs. Neural shift-reduce dependency parsing.

Unit IV:

Semantic Analysis - Lexical semantics and word-sense disambiguation, Compositional semantics. Semantic Role Labelling and Semantic Parsing.

Unit V:

Information Extraction (IE) AND Machine Translation (MT) - Named entity recognition and relation extraction, IE using sequence labelling. Basic issues in MT, Statistical translation, word alignment, phrase-based translation, and synchronous grammars, Case Studies - Bidirectional Encoder Representations from Transformers (BERT), Transformer (machine learning model), Graph Neural Networks.

Reference:

1. "Speech and Language Processing" by Daniel Jurafsky and James H Martin. Pearson Education.

2. "Natural language Understanding" by James A.

3. "Natural language processing" by Bharati A., Sangal R., Chaitanya V.

MSP 204 C-Data Engineering with Python

(Elective-I of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering with Python	Course Code	MSP 204 C
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Data Engineering **-** Data Engineering Ecosystem and it's Lifecycle. Handling files with Python **-** Types of Files, Creating and Reading Text Data, File Methods to Read and Write Data, Reading and Writing Binary Files, Reading and Writing CSV Files, Python OS and OS.

Unit II:

Working with Text Data **-** Processing HTML Files, Processing Texts in Natural Languages, JSON and XML in Python, Regular Expression Operations **-** Using Special Characters, Regular Expression Methods, Named Groups in Python Regular Expressions.

Unit III:

Working with Data Series and Frames - Pandas Data Structures, Reshaping Data, Handling Missing Data, Combining Data, Ordering and Describing Data, Transforming Data, Taming Pandas File I/O.

Unit IV:

Working with Databases - Setting Up a MySQL Database, using a MySQL Database: Command Line, using a MySQL Database, Taming Document Stores: MongoDB Working with Tabular Numeric Data (Numpy with Python)

Unit V:

Plotting **-** Basic Plotting with PyPlot, Getting to Know Other Plot Types, Mastering Embellishments, Plotting with Pandas.

References:

1. Data Engineering with Python, Paul Crickard, 2022.

2. Introduction to Python Programming. Gowrishankar S., Veena A. CRC Press, Taylor & Francis Group, 2019

MSP 204 D-Learn AWS Data Engineering

Department	Computer Science	Course Type	DSCC
Course Title	Learn AWS Data Engineering	Course Code	MSP 204 D
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to AWS - Overview of AWS Glue Components, Data Ingestion using AWS Lambda Functions, Setup Spark History Server for AWS Glue Jobs, Deep Dive into AWS Glue Catalog.

Unit II:

Introduction to Cloud Computing & AWS - Introduction to Cloud Computing, Cloud Service & Deployment Models, Introduction to AWS S3, EC2.

Unit III:

Elastic Compute and Storage Volumes **-** Introduction of EC2, Regions & Availability Zone (AZs), Pre – EC2, EC2 Instance Types, comparing public IP and Elastic IP, demonstrating how to Launch an AWS EC2 instance. Virtual Private Cloud - What is Amazon VPC, VPS as a networking layer of EC2, Components of VPC – network Interfaces, route tables, internet gateway.

Unit IV:

Simple Storage Service **-** Introduction to AWS Storage, Pre-S3 – Online Cloud Storage, API, S3 Consistency models, Storage Hierarchy, Buckets in S3. Overview of Amazon AWS Athena, Amazon AWS Athena using AWS CLI, Amazon AWS Athena using Python boto3.

Unit V:

Data Storage **-** What is S3 Glacier, Accessing Amazon S3 Glacier, Glacier Vaults, Glacier Archives, Amazon DynamoDB and it's working, Accessing DynamoDB through portal and CLI. DynamoDB - Tables and items, Indexes, Streams and Replication, Backup and Restore.

Reference:

1. "Data Engineering with AWS: Learn how to design and build cloud-based data transformation pipelines using AWS" by Gareth Eagar.

2. "Fundamentals of Data Engineering" by Joe Reis & Matt Housley.

MSP 205 A-Machine Learning & Prediction Algorithms
(Elective-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Machine Learning & Prediction Algorithms	**Course Code**	MSP 205 A
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Basics of Analytics - Different Analytical Models, Data with numerical summaries, Statistical Inferences, Visualization, Probability Distributions, Sampling and Hypothesis Testing.

Unit II:

Prediction Algorithms in ML – Definition, 3 levels of predictive models – Defining the problem statement, building and testing the model, and uncertainty evaluation, Selection of Algorithms, Case Study.

Unit III:

Regression - Basics, Linear and Non-Linear Regression Models, Univariate, Bivariate and Multivariate Regression, Multicollinearity, Autocorrelation and Dummy Variables, Durbin-Watson Test.

Unit IV:

Classification **-** Fundamentals of Classification, Decision Tree, Random Forest Algorithms, Naïve Bayes Algorithm, Neural Networks – KNN Classification Algorithms, Support Vector Machines.

Unit V:

ML Future - Business Impact of Machine Learning Algorithms, Workshop with industry experts about machine learning and algorithms.

References:

1. Artificial Intelligence and Machine Learning (2014), by Vinod Chandra SS and Anand Hareendran S.

2. Fundamentals of Machine Learning for Predictive Analysis by John D. Kelleher, Brian Mac Namee and Aoife D'Arcy.

3. Applied Regression Analysis (Third Edition) by Norman R. Draper and Harry Smith.

MSP 205 B-Applied Machine Learning
(Elective-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Applied Machine Learning	**Course Code**	MSP 205 B
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Fundamentals of Machine Learning - Intro to SciKit Learn, Key Concepts in Machine Learning, Python Tools for Machine Learning, An Example Machine Learning Problem, Examining the Data.

Unit II:

Applications of Machine learning - Image Recognition, Speech Recognition, Traffic prediction, Product recommendations, Self-driving cars, Email Spam and Malware Filtering, Virtual Personal Assistant, Online Fraud Detection, Stock Market trading.

Unit III:

Supervised Machine Learning - Introduction to Supervised Machine Learning, Overfitting and Underfitting, Supervised Learning: Datasets, K-Nearest Neighbors: Classification and Regression, Linear Regression: Least-Squares, Linear Regression: Ridge, Lasso, and Polynomial Regression.

Unit IV:

Logistic Regression, Linear Classifiers - Support Vector Machines, Multi-Class Classification, Kernelized Support Vector Machines, Cross-Validation, Decision Trees, Naive Bayes Classifiers, Random Forests, Gradient Boosted Decision Trees, Neural Networks, Data Leakage, Dimensionality Reduction and Manifold Learning, Clustering.

Unit V:

Unsupervised Learning algorithms - K-means clustering, KNN (k-nearest neighbors), Hierarchal clustering, Anomaly detection, Neural Networks, Principal Component Analysis, Independent Component Analysis, Apriori algorithm, Singular value decomposition.

Reference:

1. "Applied Machine Learning" by David Forsyth.

2. "Applied Machine Learning and AI for Engineers" by Jeff Prosise.

3. "Machine Learning" by Tom M. Mitchell.

MSP 205 C-Data Engineering using Kafka and Spark Structured Streaming
(Elective-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering using Kafka and Spark Structured Streaming	Course Code	MSP 205 C
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Kafka **-** Producer, Consumer, Ecosystem, Kafka API's, Basic Kafka Operations, Zookeeper, Introduction to Spark -Spark and it's Features, Components of Spark, Spark on Hadoop, Spark RDD, Spark Shell.

Unit II:

Kafka - Managing Topics using Kafka CLI, Produce and Consume Messages using CLI, Kafka Connect to Produce, Validate Kafka Connect to Produce.

Unit III:

Kafka Connect to Produce Messages, Validate Kafka Connect to Produce Messages, Kafka Consumer Groups, HDFS 3 Sink Connector, Kafka Topic **-** Creation, Reading Data, Preview Data Using Console and Memory, Transform Data using Spark API's, Writing Data to HDFS using Spark and Validate.

Unit IV:

Spark **-** Spark SQL, Spark Streaming, Spark Lib, Spark ML, Spark Config and Spark session.

Unit V:

Using Spark in applications**:** Master and Slave Configurations, Data Sources, DAGS, Transformations, Actions, RDD, Datasets, Data Frames, Hadoop**:** HDFS, Hive, MapReduce, Logic to Upload GHArchive Files, Upload GHArchive Files to HDFS, Analyze GHArchive Data in Parquet files using Spark, Load Data Incrementally to Target Table, Validate Spark 2 using CLIs.

References:

1. Modern Data Engineering with Apache Spark, Scott Haines, 2022.
2. Spark: The Definitive Guide, Bill Chambers, Matei Zaharia, 2018.
3. Kafka Streams in Action, Bill Bejeck, 2018.
4. Learning Spark, Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee, 2020.

MSP 205 D- Data Engineering on Google Cloud platform

(Elective-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering on Google Cloud platform	**Course Code**	MSP 205 D
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Introduction to Data Engineering in GCP - what is Data Engineering and it's examples, what is Data Modelling, Features of Google Data Engineering, Google Data Engineering Tools.

Unit II:

Database - Relational Database, No SQL Database, PostgreSQL & with Python, Relational Database Model - Importance of Relational Database Model, OLTP Vs OLAP, Database Normalization, Snowflake Schema.

Unit III:

Storage systems - Cloud Storage, Big Table, Big Query, Data Lake, Elasticsearch - Basic Usage, Index & Document, Elasticsearch ETL, Query DSL, Elasticsearch ETL – Python.

Unit IV:

Hadoop & Spark - Hadoop Ecosystem, Introducing Spark, Spark Programming, Date Formats, Data Lake - Overview, Schema on Road, Lake, not Swamp, Google Data Catalogues.

Unit V:

Data Lake-Data Lake Vs Data Swamp, Spark read with or without schema, Cloud functions & high-level Architecture.

References:

1. Data Engineering with Google Cloud Platform, Adi Wijaya, 2022.

2. Google Cloud Platform in Action, JJ Geewax, 2018.

3. Google Cloud Platform for Architects, Janani Ravi, Judy Raj, Vitthal Srinivasan, 2018.

SEMESTER III ELECTIVES
MSP 304 A- Data Science Tools
(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Science Tools	**Course Code**	MSP 304 A
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Basics of DS - Fundamentals of Data Science, Methodologies, Benefits of Data Science platforms, Challenges of implementing Data Science tools and techniques, Popular Data Science tools.

Unit II:

Statistics for Data Science – Statistics for DS, Data collection, Data management, Data Visualization and Storytelling, Applications of Data Science.

Unit III:

Data Science Process - Different stages of Data Science process, Data Storage and Processing Tools – Apache Hadoop, Exploratory Data Analysis – Informatica PowerCenter.

Unit IV:

Data Visualization - Fundamentals of Data Visualization, Concept of reporting and dashboarding, Introduction to Tableau – Creating Insightful Dashboards, Basics of QlikView.

Unit V:

Basics of Python - Introduction to Python, Variables, Operators and Python Block, Read Functions, String, List and Dictionary, Object Oriented Programming (OOP), Libraries – Matplotlib and Scikit-learn.

References:

1. Data Science from Scratch by Joel Grus (2nd Edition).

2. Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic.

3. Python Data Science Handbook by Jake VanderPlas.

MSP 304 B- Data Science with PySpark

(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Science with PySpark	Course Code	MSP 304 B
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Intro to Python & Pyspark **-** Basics of Python, Programming with RDD: Foreach Loop, Using Reduce Function, Viewing Records from MySQL, Spark Joins, Pyspark Joins Examples, Word Count

Unit II:

Data Frames Essentials - Intro to Data Frames, Read, Write, validate &explore Clean, Manipulate, Join, Aggregate.

Unit III:

Classification of Spark MLlib - Intro to Spark MLlib, Model Selection and Tuning in MLlib, Data Formatting and Transformations, Train and Evaluate Models, Train & Test Models, Train & Test Models [Multilayer PC], Train & Test Models [Naive Bayes], Train & Test Models [Linear SVM], Train & Test Models [Decision Tree], Train & Test Models [Random Forest]

Unit IV:

Natural Language Processing - Intro to NLP, Feature Transformers, Feature Extractors, Data Prep, Vectorize, Train & Eval, Regression in MLlib, Regression in Pyspark MLlib, Data Prep, Linear Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosted Tree Regression

Unit V:

MLlib - Introduction to Clustering, K-Means & Bisecting K-Means in MLlib code, Gaussian Mixture Modeling in MLlib, Frequent Pattern Mining in MLlib, Intro to Spark Structured Streaming, Intro to Streaming Data Using Sockets.

References:

1. "Learning Pyspark" by Tomasz Drabas Denny

2. "Learning Spark: Lightning-Fast Data Analytics" by Jules S. Damji

MSP 304 C-Data Engineering using AWS Data Analytics
(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering using AWS Data Analytics	**Course Code**	MSP 304 C
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Data Engineering – Types of Data, Data engineering Data Process, Data Engineering Responsibilities, Types of Data Sources, Data Engineering Tools and Skills, Data Engineering Vs Data Science, Roles and Responsibilities of a Data Engineer.

Unit II:

AWS Data Engineering – Intro, The AWS Data Engineer's Toolkit, AWS Data Engineering Tools, Data Management Architectures for Analytics, Data Cataloguing, Security, and Governance.

Unit III:

AWS Technical Essentials - AWS Infrastructure, AWS Free tier, AWS IAM, EC2, EBS, EBS Volume, Security Group, AWS AMI, AWS Load Balancing Auto Scaling, Cloud Watch, S3 Buckets, Application, AWS SFTP – Create Server, Users, File Transfer with SFTP, AWS EKS, AWS Lambda Function.

Unit IV:

AWS Data Engineering - Identifying and Enabling Data Consumers, Loading Data into a Data Mart, Orchestrating the Data Pipeline. Architecting Data Engineering Pipelines, Ingesting Batch and Streaming Data, Transforming Data to Optimize for Analytics.

Unit V:

AWS Data Engineering - Ad Hoc Queries with Amazon Athena, Visualizing Data with Amazon Quick Sight, Enabling Artificial Intelligence and Machine Learning, Ingesting data using Rest API, Ingesting data using SFTP Server, Ingesting data into Database (AWS RDS - Postgre SQL), Incremental data loading.

References:

1. "Data Engineering with AWS: Learn how to design and build cloud-based data transformation pipelines using AWS" by Gareth Eagar

2. "Data Engineering with AWS" by Eagar Gareth)

304 D Data Engineering Essentials using SQL, Python, and PySpark
(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering Essentials using SQL, Python, and PySpark	**Course Code**	MSP 304 D
L-T-P	3-1-0	**Credits**	4
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit I:

Introduction to SQL – SQL using Postgres, CRUD operations Basic SQL Queries (Filtering, Joins, and Aggregations), Advanced - Predefined Functions using SQL (String, Date, and other functions), Creating Tables and Indexes using Postgres DDL Commands, Advanced SQL Queries using PostgreSQL.

Unit II:

Programming Essentials using Python, Using Pandas - Python Pandas Libraries, Data Frame by applying Standard Transformations such as filtering, joins, sorting, Database connectivity through python - Map Reduce Libraries in Python, Database Programming using Python - CRUD Operations, Batch Operations.

Unit III:

Introduction to Bigdata - Introduction to Bigdata and its applications using pyspark, Architecture of Spark and its components, SparkML, Spark Lib and Spark SQL, setting up Single Node Data Engineering Cluster for Practice Apache Spark integrated with HDFS and YARN, Deep dive into HDFS - Overview of HDFS Commands, Copy Files into HDFS using put or copyFromLocal command using appropriate HDFS Commands, HDFS commands such as du, df, HDFS block size, replication factor.

Unit IV:

Spark SQL and its components - Intro to Spark SQL, Basic Transformations using Spark SQL, Managing Tables -Basic DDL and DML in Spark SQL, Managing Tables - DML and Create Partitioned Tables using Spark SQL, Joins and transformations in Data frames using temp view, Spark SQL functions - Overview of DAGS, Overview of Spark SQL Functions to manipulate strings, dates, null values, Windowing Functions using Spark SQL for ranking, advanced aggregations.

Unit V:

Data frames and Dataset APIs - Data Processing Overview using Spark or Pyspark Data Frame APIs, Projecting or Selecting data from Spark Data Frames, renaming columns, providing aliases, dropping columns from Data Frames, processing Column Data using Spark or Pyspark Data Frame, Data frames and Transformations - Basic Transformations and Data Sets on Spark Data Frames using Pyspark Data Frame APIs, Windowing Functions on Spark Data Frames, Deployment as well as Execution Life Cycle of Spark Applications.

References:

1. "Data Engineering with Python: Work with massive datasets to design data models and

automate data pipelines using Python" by Paul Crickard.

2. "Advanced Analytics with Pyspark: Patterns for Learning from Data at Scale Using Python and Spark "by Akash Tandon.

305 A-AI and Deep Learning through TensorFlow

(Elective-IV of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	AI and Deep Learning through TensorFlow	Course Code	MSP 305 A
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs

SEE Marks	70	CIE Marks	30

Unit I:

Intro to AI - Basics of Artificial Intelligence, Applications of AI, Types of AI, AI agents, Artificial Neural Networks, Fuzzy Logic, Introduction to Genetic Algorithm, Soft Computing.

Unit II:

Intro to Deep Learning - Fundamentals of Deep Learning, Architecture of Deep Learning, Types of Deep Learning Networks – FFNN, RNN, CNN, Restricted Boltzmann Machine, Autoencoders, Deep Learning Applications, Limitations, Advantages, Disadvantages.

Unit III:

TensorFlow - What is TensorFlow, Installation Through pip, Installation Through conda, Architecture of TensorFlow, Advantage & Disadvantage of TensorFlow, Basics of TensorFlow – Representation of a Tensor, Types of Tensors, TensorFlow Perception – Single Layer, Hidden Layer, Multilayer Perceptions.

Unit IV:

ANN in TensorFlow – Intro, Implementation of NN, Classification of NN, Liner Regression, CNN in TensorFlow – Intro, Working of CNN, Training of CNN, MNIST Dataset in CNN, RNN in TensorFlow – Intro, Working of RNN, RNN Time Series, LSTM RNN in TensorFlow.

Unit V: Case Studies

Case 1: CNN for Digit Recognition

Case 2: CNN for Breast Cancer Detection

Case 3: CNN for Predicting the Bank Customer Satisfaction

Case 4: CNN for Credit Card Fraud Detection

Case 5: RNN - LSTM for IMDB Review Classification

Case 6: Google Stock Price Prediction with RNN and LSTM

References:

1. "Hands-on Machine learning with Scikit-Learn, Keras, and TensorFlow" by Aurelion Geron.

2. "Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API" by Amita Kapoor, Antonio Gulli, and Sujit Pal.

MSP 305 B-Machine Learning Techniques

(Elective-IV of MS Programme) 62

Department	Computer Science	Course Type	DSCC
Course Title	Machine Learning Techniques	Course Code	MSP 305 B
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Machine Learning Basics - Fundamentals of Machine Learning - Type of ML – Supervised, Unsupervised and Reinforcement Learning, Applications of Machine Learning.

Unit II:

Supervised Learning – Basics, Scope, Advantages and Disadvantages of Supervised Learning, Regression – Linear and Logistic Regression, Hypothesis Testing, Data Distribution and Transformation, Classification – Decision Tree and Random Forest Technique, Data Visualization**.**

Unit III:

Un-Supervised Learning - Basics, Scope, Advantages and Disadvantages of Un- Supervised Learning, Clustering, Concepts of Clustering, K-Means Clustering Technique, Pattern Search, Dimension Reduction.

Unit IV:

Reinforcement Learning – Basics, Scope, Advantages and Disadvantages of Reinforcement Learning, Markov Decision Process (MDP), Monte Carlo Prediction, Business aspect of Reinforcement Learning.

Unit V:

Deep Learning – Intro to Deep Learning, Scope, Advantages and Disadvantages of Deep Learning, Introduction to Neural Networks, Recurring Neural Networks (RNN), Convolutional Neural Networks (CNN), Natural Language Processing (NLP).

References:

1.Understanding Machine Learning from Theory to Algorithms by Shai Shalev- Shwartz, Shai Ben-David.
2. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Mueller and Sarah Guido.

MSP 305 C-Data Engineering for Beginner using Google Cloud & Python

(Elective-IV of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Machine Learning Techniques	Course Code	MSP 305 C
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Kafka **-** Producer, Consumer, Ecosystem, Kafka API's, Basic Kafka Operations, Zookeeper, Introduction to Spark - Spark and it's Features, Components of Spark, Spark on Hadoop, Spark RDD, Spark Shell.

Unit II:

Starting with Kafka - Managing Topics using Kafka CLI, Produce and Consume Messages using CLI, Kafka Connect to Produce, Validate Kafka Connect to Produce.

Unit III:

Kafka Connect to Produce Messages, Validate Kafka Connect to Produce Messages, Kafka Consumer Groups, HDFS 3 Sink Connector, Kafka Topic - Creation, Reading Data, Preview Data Using Console and Memory, Transform Data using Spark API's, Writing Data to HDFS using Spark and Validate.

Unit IV: Spark:

Spark Intro**,** Spark SQL, Spark Streaming, Spark Lib, Spark ML, Spark Config and Spark session.

Unit V:

Using Spark in applications - Master and Slave Configurations, Data Sources, DAGS, Transformations, Actions, RDD, Datasets, Data Frames, Hadoop - HDFS, Hive, MapReduce, Logic to Upload GHArchive Files, Upload GHArchive Files to HDFS, Analyze GHArchive Data in Parquet files using Spark, Load Data Incrementally to Target Table, Validate Spark 2 using CLIs.

References:

1. Building Google Cloud Platform Solutions, Ted Hunter, Steven Porter, Legorie Rajan PS · 2019

2. Data Science on the Google Cloud Platform, Valliappa Lakshmanan · 2017

MSP 305-D Data Engineering with Google BigQuery & Google Cloud
(Elective-IV of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering with Google Big Query & Google Cloud	Course Code	MSP 305 D
L-T-P	3-1-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I:

Introduction to Data Warehouse - Data Warehouse and it's life cycle, Characteristic, Architecture and Components, Introduction to Big Query - Big Query and it's Architecture, Data Storage and Execution, Big Query Vs MapReduce.

Unit II:

Big Query - Data Types, Functions, Query, Converting Data Types, Data Flow Basic - Data Quality, Clean and Transform, Store Data, upgrading from sandbox Account, Clean & Transform with Data prep, Schedule Query, Analyse Data, Data Visualization.

Unit III:

Load Data into Big Query **-** The basics, Handling Errors, Efficient Load, From your Data to Big Query, Microservice Architecture, Recurring Load.

.

Unit IV:

Data Visualization **-** Google sheets and Big Query, Google Data Studio, Query **- Join -** Theory, Hands-on, Union and Intersection, Basic Statistical Functions.

Unit V:

Query Essentials, DDL & DML, Integration with Google Cloud, Work flow of ETL, EL and ELT in Big Query.

References:

1. Google Big Query: The Definitive Guide, Valliappa Lakshmanan, Jordan Tigani, 2019.

2. Learning Google BigQuery, Eric Brown, Thirukkumaran Haridass, Jason Morris, Mikhail Berlyant, Ruben Oliva Ramos, 2017.

3. Google BigQuery Analytics, Jordan Tigani, Siddartha Naidu, 2014.

OPEN ELECTIVE

Fundamentals of Programming

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Programming	**Course Code**	
L-T-P	3-0-1	**Credits**	3
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

Unit-I:

Introduction to Python ,Features and Applications of Python; Python Versions; Installation of Python; Python Command Line mode and Python IDEs; Simple Python Program. Python Basics: Identifiers; Keywords; Statements and Expressions; Variables; Operators; Precedence and Association; Data Types; Indentation; Comments; Built-in Functions- Console Input and Console Output, Type Conversions; Python Libraries; Importing Libraries with Examples.

Unit-II:

Python Control Flow: Types of Control Flow; Control Flow Statements- if, else, elif, while loop, break, continue statements, for loop Statement; range () and exit () functions.

Unit-III:

Exception Handling: Types of Errors; Exceptions; Exception Handling using try, except and finally. Python Functions: Types of Functions; Function Definition- Syntax, Function Calling, Passing Parameters/arguments, the return statement; Default Parameters; Command line Arguments; Key Word Arguments; Recursive Functions; Scope and Lifetime of Variables in Functions.

Unit-IV:

Strings: Creating and Storing Strings; Accessing Sting Characters; the str() function; Operations on Strings- Concatenation, Comparison, Slicing and Joining, Traversing; Format Specifiers; Escape Sequences; Raw and Unicode Strings; Python String Methods.

Unit-V:

Lists: Creating Lists; Operations on Lists; Built-in Functions on Lists; Implementation of Stacks and Queues using Lists; Nested Lists. Dictionaries: Creating Dictionaries; Operations on Dictionaries; Built-in Functions on Dictionaries; Dictionary Methods; Populating and Traversing Dictionaries. Tuples and Sets: Creating Tuples; Operations on Tuples; Built-in Functions on Tuples; Tuple Methods; Creating Sets; Operations on Sets; Built-in Functions on Sets; Set Methods.

66

References:

1. Think Python How to Think Like a Computer Scientist, Allen Downey et al., 2nd Edition, Green Tea Press.

2. Introduction to Python Programming, Gowrishankar S et al., CRC Press, 2019.

3. Python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language, Fabio Nelli, Apress®, 2015

4. Advance Core Python Programming, MeenuKohli, BPB Publications, 2021.

5. Data Structures and Program Design Using Python, D Malhotra et al., Mercury Learning and Information LLC, 2021.

Department	Computer Science	Course Type	MSP
Course Title	Web Designing	**Course Code**	
L–T-P	3-0-1	**Credits**	3
Contact Hours	60 Hrs	**Duration of SEE**	3 Hrs
SEE Marks	70	**CIE Marks**	30

UNIT -I

INTERNET BASICS: Introduction to internet and its applications, working of the internet, Email, telnet, FTP, E-commerce, video conferencing. Internet service providers, domain name server, internet address, World Wide Web and its evolution, uniform resource locator (URL), browsers, search engine, web server, HTTP protocols.

UNIT -II

WEB DESIGNING TECHNOLOGIES :Introduction HTML – HTML editors, file creation, Basic HTML tags, text formatting, forms, images, lists, tables, linking documents, frame sets, lists and HTML layout elements.

UNIT -III

CASCADING STYLE SHEETS Introduction to CSS, importance of CSS, Types: inline, internal, and external with examples. Selector types: Element, id, class and overriding the styles, CSS box model.

UNIT IV

JAVA SCRIPT Introduction to Java script, writing java script into HTML, Java Script Syntax, Variables, Data types, arrays - properties and methods, operators, expressions, programming constructs – conditional and looping statements, dialogue boxes, web page events.

UNIT V

Introduction to Web Publishing or Hosting ,Creating the Web Site , Saving the site, Working on the web site, Creating web site structure, Creating Titles for web pages ,Themes-Publishing web sites.

References: 68

1. M. Srinivasan: Web Technology Theory and Practice, Pearson Education.
2. Jeffrey C. Jackson: Web Technologies- A Computer Science Perspective, Pearson Education, Eleventh Impression.
3. Chris Bates: Web Programming Building Internet Applications, 3rd Edition Wiley India.
4. Internet Technology and Web Design, Instructional Software Research and Development (ISRD) Group, Tata McGraw Hill.

QUESTION PAPER STRUCTURE FOR SEMESTER END EXAMINATIONS

The semester end university examination in each course and paper is of three hours duration with a maximum of 70 marks. The Question paper consists of two parts, A & B. Part A consists of 6 short answer questions and 5 to be answered. Part B consists of 10 questions with choice from each unit.

S V UNIVERSITY: TIRUPATI
MS in Data Science
Course code:

Time: 3hrs	Max. Marks: 70

PART A   (5 X4 =20 MARKS)
Answer any Four from the following Questions

1.
  a.
  b.
  c.
  d.
  e.
  f.

PART B   (5 X10 =50 MARKS)

Answer any one question from each unit.

Unit-I

2.

  Or

3.

Unit -II

4.

  Or

5.

Unit-III

```
6.
   Or
7.
                          Unit -IV
8.
   Or
9.
                          Unit-V
10.
   Or
11.
```

Internal Examination: There will be two written examinations for a student for each course in the Partner institution conducted by the faculty concerned. The examination will be for a minimum duration of 2hrs. The internal assessment final marks will be awarded for a maximum base of 30 marks. Apart from the two written tests, students may be given any kind of Assignment and exercise where the marks awarded can be part of the total 30 marks assessment. The faculty concerned will develop the structure of Assessment for 30 sessional marks and will inform the students at the beginning of the semester.

The evaluated papers and material of each semester shall be preserved by the registered industry institution till the completion of the succeeding semester of the students.