DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE
(AUTONOMOUS)
(Approved by AICTE & Affiliated to Anna University, Chennai)
Re-Accredited by NAAC with 'A' Grade
Accredited by NBA for AERO, BME, CSE, ECE, EEE, IT & MECH.
PERAMBALUR-621212, TAMILNADU, INDIA.
Website: www.dsengg.ac.in

COURSE PLAN

Name of the Faculty				
Designation/Department	AP/CSE			
Course Code/Name	**P23CSE06/ HIGH PERFORMANCE COMPUTING FOR BIG DATA**			
Year/Section/Department	I/A/CSE			
Credits Details	L: 3	T: 0	P: 0	C:3
Total Contact Hours Required	45			

Syllabus:

UNIT I - INTRODUCTION	**No. of Periods: 9**
The Emerging IT Trends- IOT/IOE-Apache Hadoop for big data analytics-Big data into big insights and actions – Emergence of BDA discipline – strategic implications of big data – BDA Challenges – HPC paradigms – Cluster computing – Grid Computing – Cloud computing – Heterogeneous computing – Mainframes for HPC - Supercomputing for BDA – Appliances for BDA.	
UNIT II - NETWORK & SOFTWARE INFRASTRUCTURE FOR HIGH PERFORMANCE BDA	**No. of Periods: 9**
Technologies that support Real time analytics – MOA: Massive online analysis – GPFS: General parallel file system – Client case studies – Key distinctions – Machine data analytics – operational analytics – HPC Architecture models – In Database analytics – In memory analytics	
UNIT III - REAL TIME ANALYTICS USING HIGH PERFORMANCE COMPUTING	**No. of Periods: 9**
Procedures and work instructions – Supporting quality devices - Staff training and certification - Corrective and preventive actions – Configuration management – Software change control – Configuration management audit - Documentation control.	
UNIT IV - SECURITY AND TECHNOLOGIES	**No. of Periods: 9**
Security, Privacy and Trust for user – generated content: The challenges and solutions – Role of real time big data processing in the IoT – End to End Security Framework for big sensing data streams – Clustering in big data	
UNIT V - EMERGING BIG DATA APPLICATIONS	**No. of Periods: 9**
Deep learning Accelerators – Accelerators for clustering applications in machine learning - Accelerators for classification algorithms in machine learning – Accelerators for Big data Genome Sequencing	

TOTAL: 45 PERIODS

Objectives:

1. To learn the fundamental concepts of High Performance Computing.
2. To learn the network & software infrastructure for high performance computing.
3. To understand real time analytics using high performance computing
4. To learn the different ways of security perspectives and technologies used in HPC
5. To understand the emerging big data applications.

Text Book:

T1- **(AWS for Dummies)** → Basics of cloud computing & AWS services
T2 **(Windows Azure)** → Microsoft Azure programming & cloud platform
T3 - **(Virtualization Guide)** → Fundamentals of virtualization & infrastructure.

Reference Book:

R1.Pethuru Raj, Anupama Raman, Dhivya Nagaraj and Siddhartha Duggirala, "HighPerformance BigData Analytics: Computing Systems and Approaches", Springer, 1st Edition, 2015.
R2. Kuan-Ching Li , Hai Jiang, Albert Y. Zomaya, "Big Data Management and Processing", CRC Press,1st Edition,2017.
R3. Chao wang, "High Performance Computing for Big Data: Methodologies and Applications", CRC Press,1st Edition,2018.
R4.Khosrow Hassibi, "High-Performance Data Mining And Big Data Analytics", Create Space Independent Publishing Platform,1 stEdition,2014

Website:

W1: https://www.statlect.com/probability-distributions
W2: https://www.khanacademy.org/math/statistics-probability
W3: https://stattrek.org/probability/probability-rules.aspx
W4: https://www.geeksforgeeks.org/probability-in-maths//

Online Mode of Study:

W1: https://onlinecourses.nptel.ac.in/noc24_ma23/preview
W2: https://www.coursera.org/courses?query=probability%20and%20statistics
W3: https://www.khanacademy.org/math/statistics-probability
W4: https://stattrek.org/statistics/tutorials.aspx
W5: https://www.statlect.com/

Course Plan:

Topic Number	Topic	Reference Detail	Page Number	Mode of teaching	Number of Periods Required	Cumulative Period
UNIT I - INTRODUCTION						
1	Emerging IT Trends	T1	1-5	BB	1	1
2	Types of Virtual Machines	T1	6-10	BB	1	2

3	IoT / IoE	T1	11-15	BB	1	3
4	Apache Hadoop	T1	16-20	BB	1	4
5	Big Data → Insights & Actions	T1	21-25	BB	1	5
6	Emergence of BDA	T1	26-30	BB	1	6
7	Strategic Implications	T1	31-35	BB	1	7
8	HPC Paradigms	T1	36-40	BB	1	8
9	Computing Models	T1	41-45	BB	1	9
Outcome of Unit I: At the end of unit, Students should be able to • Understand the basics concepts of High Performance computing systems						
UNIT II - NETWORK & SOFTWARE INFRASTRUCTURE FOR HIGH PERFORMANCE BDA						
10	Procedures & Work Instructions	T1	46-50	BB	1	10
11	Supporting Quality Devices	T1	51-55	BB	1	11
12	Staff Training	T1	56-60	BB	1	12
13	Certification	T1	61-65	BB	1	13
14	Corrective Actions	T1	66-70	BB	1	14
15	Preventive Actions	R1	71-75	BB	1	15
16	Configuration Management	T1	76-80	PPT	1	16
17	Software Change Control	T1	81-85	BB	1	17
18	Audit & Documentation	T1	86-90	BB	1	18
• **Outcome of Unit II:** At the end of this unit, students should be able to: • Apply the concepts of network and software infrastructure for high performance computing.						
UNIT-III - REAL TIME ANALYTICS USING HIGH PERFORMANCE COMPUTING						
19	Procedures & Work Instructions	T1	91-95	BB	1	19
20	Supporting Quality Devices	T1	96-100	BB	1	20
21	Staff Training	T1	101-105	BB	1	21
22	Certification	T1	106-110	BB	1	22
23	Corrective Actions	T1	111-115	BB	1	23
24	Preventive Actions	T1	116-120	BB	1	24
25	Configuration Management	T1	121-125	BB	1	25
26	Software Change Control	T1	126-130	BB	1	26
27	Audit & Documentation Control	T1	131-135	BB	1	27

. Outcome of Unit III:

At the end of this unit, students should be able to:

- Use real time analytics using high performance computing.

UNIT-IV- SECURITY AND TECHNOLOGIES						
28	Security Concepts	T2	136-140	BB	1	28
29	User-Generated Content	T2	141-145	BB	1	29
30	Challenges	T2	146-150	BB	1	30
31	Solutions	T2	151-155	BB	1	31
32	Role of Real-Time Processing in IoT	T2	156-160	BB	1	32
33	Big Data in IoT	T2	161-165	BB	1	33
34	End-to-End Security Framework	T2	166-170	BB	1	34
35	Clustering in Big Data	T2	171-175	BB	1	35
36	Applications & Case Study)	T2	176-180	BB	1	36

Outcome of Unit IV:

At the end of this unit, students should be able to:

- Apply the security models and big data applications in high performance computing.

UNIT-V - PROGRAMMING MODEL						
37	Introduction to Emerging Applications	T2	181-185	BB	**1**	37
38	Deep Learning Accelerators	T2	186-190	BB	1	38
39	Need for Accelerators	T2	191-195	BB	1	39
40	Clustering Applications	T2	196-200	BB	1	40
41	ML Clustering Techniques Applications	T2	201-205	BB	1	41
42	Classification Algorithms	T2	206-210	BB	1	42
43	Accelerators for Classification	T2	211-215	BB	1	43
44	Genome Sequencing	T2	216-220	BB	1	44
45	Advanced Applications	T2	221-225	BB	1	45

Outcome of Unit V:

At the end of this unit, students should be able to:

- Understand the emerging big data applications.

Course Outcome:

At the end of the course the students would be able to

CO1: Understand the basics concepts of High Performance computing systems.
CO2: Apply the concepts of network and software infrastructure for high performance computing
CO3: Use real time analytics using high performance computing.
CO4: Apply the security models and big data applications in high performance computing
CO5: Understand the emerging big data applications.
CO6: Apply classification algorithms for accelerators

Course Outcome Vs Program Outcome Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO 1	2	1	-	-	-	-	-	-	-	1	-	-	2	1
CO 2	3	2	1	1	1	-	-	-	-	1	-	2	2	1
CO 3	2	1	-	-	-	-	-	-	-	1	-	-	2	1
CO 4	2	1	-	-	-	-	-	-	-	1	-	-	2	1
CO 5	2	1	-	-	-	-	-	-	-	1	1	-	2	1
CO 6	2	1	-	-	1	-	-	-	2	1	1	2	2	1
AVG	**2**	**1**	**1**	**1**	**1**	**-**	**-**	**-**	**2**	**1**	**1**	**2**	**2**	**1**

Internal Evaluation Components:

Web portal	Assignment	Components	Topic Number with Topic / Unit Details	Relevance to CO
Web portal 1	**--**	**Assessment – I (60)**	**Unit I and II**	**CO 1 & CO2**
	1	**Assignment – Handwritten (20)**	1 Explain the emerging IT trends and discuss the role of IoT/IoE in modern computing. 2 Explain the challenges in Big Data Analytics (Volume, Velocity, Variety, etc.)	CO1

	2	**Assignment – Poster Presentation / PPT (20)**	1 Explain the technologies supporting real-time analytics in detail... 2 Explain in-memory analytics with suitable examples.	CO2
Web portal 2	**--**	**Assessment – II (60)**	**Unit III and IV**	**CO3 & CO4**
	3	**Seminar (20)**	1. Explain the importance of procedures and work instructions in real-time analytics. 2.Explain the need for staff training and certification in analytics systems..	CO3
	4	**Case Study Report (20)**	**1**. Explain security, privacy, and trust issues in user-generated content... **2** .Explain clustering techniques in Big Data Analytics**..**	CO4
Web portal 3	**--**	**Model Exam (75)**	**Unit I to V**	**CO1 to CO6**
	5	**MCQ (15)**	**Unit I to V**	**CO1 to CO6**
	-	**Course Attendance (10)**	**--**	**--**

Submission Details:

Phase 1(Before AT 1)	**Phase 2 (Before AT 2)**	**Phase 3(Before Model Exam)**
Assignment 1	**Assignment 2**	**Assignment 3**

PLAN OF ASSESSMENT TEST –DISTRIBUTION OF MARKS:

TEST	**CO- MARK WISE DISTRIBUTION**						**BLOOM'S LEVEL MARK WISE DISTRIBUTION**					
AT-1	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**	**CO6**	**BTL1**	**BTL2**	**BTL3**	**BTL4**	**BTL5**	**BTL6**
	37	**23**	**-**	**-**	**-**	**-**						
AT-2	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**	**CO6**	**BTL1**	**BTL2**	**BTL3**	**BTL4**	**BTL5**	**BTL6**
	-	**-**	**37**	**23**	**-**	**-**						
MODEL	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**	**CO6**	**BTL1**	**BTL2**	**BTL3**	**BTL4**	**BTL5**	**BTL6**
	20	**20**	**20**	**20**	**10**	**10**						

Prepared By **Approved B**