



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited A++ Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

1.1.3 Courses having focus on employability/ entrepreneurship/ skill development

B.E Computer Science and Engineering With Specialization In Data Science			
Employability			
S.No	Course Code	Course Name	Page No
	SCSA1704	Data Management in Cloud	2
	SCS3007	Introduction to Visual Computing (Only for 2019 and 2020 Batches)	4
	SCSB1232	Exploratory Data Analysis	6
Skill Development			
	SCSA2702	Data Management in Cloud Lab	7

SCSA1704	DATA MANAGEMENT IN CLOUD	L	T	P	Credits	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- To understand manage data across cloud platforms, either with or instead of on-premises storage
- To understand data storage tier for disaster recovery, backup, and long-term archiving
- Data prepared by vendor, delivered ready to use
- Can be integrated into larger data set to increase analytic validity

UNIT 1 DISTRIBUTED DATABASE 9 Hrs.

Distributed Database Concepts, Data Fragmentation, Replication, Structure of Distributed Database, trade-offs in Distributing the Database, Advantages and Disadvantages of Data Distribution, Allocation Techniques for Distributed Database Design, Design of Distributed Database

UNIT 2 CLOUD DATA MANAGEMENT 9 Hrs.

Data as a service (DaaS)-Data Location- Data Control- Securing data for transport, Scalability and Cloud Services- Large Scale Data Processing- Databases and Data Stores- Data Archival- Data quality - Data Security.

UNIT 3 DATA MANAGEMENT-AS-A-SERVICE 9 Hrs.

Data Management Essentials, Metadata, Auto-tiered Storage, Single Instance Storage, Time-indexed Data System, Data management challenges. Case study: Fast and secure enterprise cloud backup, Cloud disaster recovery: Bounce back with cloud disaster recovery.

UNIT 4 OPEN SOURCE CLOUD SYSTEM 9 Hrs.

Server virtualization techniques- big data processing frameworks (e.g., Apache Hadoop, Spark and Spark Streaming)- MapReduce programming model and their applications. Cloud computing and big data- efficient cloud resource management- performance control- energy efficiency.

UNIT 5 INFORMATION STORAGE IN CLOUD COMPUTING 9 Hrs.

Storage as a Service, Storage Providers- Amazon Simple Storage Service- Nirvanix- Google Bigtable Datastore- MobileMe- Live Mesh, Storage Security, Merits and Demerits of Storage. . Cloud Computing for Storing IoT Data, AWS IoT-Data, Google Cloud IoT.Technology and Tools – MapReduce/Hadoop , In- database Analytics, MADlib and advanced SQL Tools

Max. 45 Hrs.

COURSE OUTCOMES

- CO1: Able to Design and handle Distributed database in cloud computing environment.
- CO2: Able to manage cloud database and apply distributed database with DaaS cloud
- CO3: Analyze DMaaS for Fast and secure enterprise cloud backup, Recovery
- CO4: Design Cloud server environment/ setup for implement Apache Hadoop, Spark and Spark Streaming to manage cloud big data
- CO5: Design and can be able to use cloud storage services, servers.

CO6: Evaluate cloud storage for storing IoT data at cloud computing environment .

TEXT/REFERENCE BOOKS

1. Data Management in the Cloud: Challenges and Opportunitieis, Divyakant Agarwal, Sudipto Das, and Amr EL Abbadi,
Morgan & Claypool Publishers, 2013
2. Cloud Data Management and Storage A Standards-Based Approach, Mark Carlson, McGraw-Hill Osborne Media; 1
edition (September 5, 2013)
3. <https://blog.panoply.io/data-management-concepts-and-techniques-in-a-cloud-based-world>
4. <https://vilmate.com/blog/data-management-in-cloud-computing-best-practices-challenges-trends/>
5. <https://www.networkworld.com/article/3290240/the-5-pillars-of-cloud-data-management.html>
6. <https://www.druva.com/blog/how-the-cloud-impacts-data-management/>
7. https://link.springer.com/chapter/10.1007/978-3-642-12038-1_1#citeas
8. <https://www.rocketsource.co/blog/data-as-a-service/>.

SCSA3007	INTRODUCTION TO VISUAL COMPUTING	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES

- To gain in depth knowledge of fundamentals of Visual Computing
- Modeling and solving image and video processing and analysis problems.
- Exploiting high level knowledge about human visual perception
- To contribute state-of-the-art solutions for multimedia applications
- To learn the field of visualisation, determining how best to turn data into information

UNIT 1 VISUAL COMPUTING INTRODUCTION

Physiological Foundations-Representation of Light and Color-Image and Noise Models- Basics of Fourier series, Sampling

Theorem-Vector Quantization- k-Means Clustering- Mixture Models- Gray Level and Color Quantization-Relationships between Pixels- Camera Geometry- 2D and 3D Transforms, Projections

UNIT 2 DIGITAL IMAGE PROCESSING

Digital Image Filtering-Image Transforms-Image Enhancement and Restoration, Wiener Filters, Nonlinear Image Processing

(Median filtering)-Nonlinear Diffusion, Gauss-Laplace Pyramid, Wavelets-Scale Space, (Image and Video Compression Image

Segmentation-Optical Flow-Stereo Vision-Template Matching, Point Matching

UNIT 3 DIGITAL IMAGE GENERATION

The Graphics Pipeline- Lighting and Reflection Models- Shading-Texture Analysis and Texture Mapping-Aliasing- Global

Illumination-Radiosity-Ray Tracing-Graphics Systems- APIs-3D Graphics Hardware

UNIT 4 REPRESENTATION OF GEOMETRY

Parametric Curves-Bézier Curves-B-Splines-NURBS- Tensor Product Surfaces, Triangle Meshes- Subdivision Methods-Shape

Models, linear (Gaussian) Diffusion

UNIT 5 LEARNING METHODS IN VISION

Classifier Learning-Support Vector Machines-Radial Basis Function Networks-Dimension Reduction: PCA, ICA- Linear

Discriminant Analysis-Graphical Models- Markov Random Fields-Maximum Entropy Inference and Bayesian Image Analysis

COURSE OUTCOMES

On completion of the course, student will be able to

CO1: Acquaint students with core knowledge in visual information processing and learning
CO2: Implement Digital Image Processing Mechanisms. .
CO3: Analyze and design Digital Image Generation Mechanisms
CO4: Representation of geometry and subdivision methods.
CO5: Describe the Learning Methods in Vision
CO6: Comprehend the concepts related three dimensional object representations

TEXT/ REFERENCE BOOKS

1. Donald D Hearn, M. Pauline Baker, Computer Graphics C version, Pearson Education.
2. Dave Shreiner, Mason Woo, Jackie Neider, Tom Davis, OpenGL Programming Guide: The Official Guide to Learning OpenGL, (2013).
3. James D. Foley, Andries van Dam, Steven K. Feiner, John F. Hughes, Computer Graphics: Principles & Practice in C, Addison Wesley Longman
4. Zhigang Xiang, Roy A Plastock, Computer Graphics, Schaums Outline, TMH.

END SEMESTER EXAM QUESTION PAPER PATTERN

Max. Marks : 100

Exam Duration : 3

Hrs.

PART A : 10 Questions of 2 marks each-No choice 20 Marks

PART B : 2 Questions from each unit with internal choice, each carrying 16 marks 80 Marks

SCSA2702	DATA MANAGEMENT IN CLOUD LAB	L	T	P	Credits	Total Marks
		0	0	4	2	100

COURSE OBJECTIVES

- To Learn to build and maintain reliable, scalable, Service Oriented Architectures. Web Services. Business considerations Business considerations
- To Able to apply Cloud application with existing tools.
- To set up single and multi-node Hadoop Clusters.
- To implement the concept of On demand Application Delivery like SaaS using google cloud
- To implement the Various Cloud services provided by Amazon Web

SUGGESTED LIST OF EXPERIMENTS

- 1 Introduction to cloud computing.
2. Creating a Warehouse Application in Salesforce.com.
3. Creating an Application in Salesforce.com using Apex programming Language.
4. Implementation of SOAP Web services in C#/JAVA Applications
5. Implementation of Para-Virtualization using VM Ware's Workstation/ Oracle's Virtual Box and Guest O.S.
6. Installation and Configuration of Hadoop.
7. Create an application (Ex: Word Count) using Hadoop Map/Reduce.
8. Case Study: PAAS (Facebook, Google App Engine)
9. Case Study: Amazon Web Services.
10. Implementation of IoT Data stream in Cloud Storage (Google IoT, Azure IoT)

COURSE OUTCOMES

Upon completion of the course, the student will be able to

- CO 1:** Build and maintain reliable, scalable, service oriented architectures, web services business applications.
- CO 2:** Setup different types of Hypervisors in virtualization for cloud computing.
- CO 3:** Design new algorithms that uses Map Reduce on structured and unstructured data.
- CO 4:** Set up single and multi-node Hadoop clusters for data management in cloud.
- CO5:** Represent NoSQL data in cloud environment
- CO 6:** Load big data into Hadoop, handle files in Hadoop and extract big data from Hadoop

SCSB1232	EXPLORATORY DATA ANALYSIS	L	T	P	EL	Credits	Total Marks
		3	0	0	0	3	100

COURSE OBJECTIVES

- To focus on the significance of data and software tools.
- To emphasize the relationship between variables and gain knowledge about univariate data analysis.
- To explore multivariate data analysis and visualize the data.

UNIT 1 INTRODUCTION 9 Hrs.

EDA fundamentals – Understanding data science – Significance of EDA – Making sense of data - Comparing EDA with classical and Bayesian analysis – Software tools for EDA – Visual Aids for EDA- Data transformation techniques-merging database, reshaping and pivoting, Transformation techniques.

UNIT 2 THE EXPLORATORY PERSPECTIVE 9 Hrs.

Introduction: Overview – Sources of Data – Process for Making Sense of Data - Describing Data: Observations and Variables – Types of Variables – Central Tendency – Distribution of the Data – Confidence Intervals – Hypothesis Tests - Distributions of Single Variables.

Displaying Data: The Stem and Leaf – Summarizing Data – Resistant Statistics and Number Summaries - The Box and Whisker -Understanding Data: Skewness – Outliers – Gaps and Multiple Peaks.

UNIT 3 RELATIONSHIPS BETWEEN VARIABLES 9 Hrs.

Preparing Data Tables: Cleaning - Displaying Relationships =The Scatter Plot –Summarizing Relationships – Fitting A Line – Smoothing the Data – Median and Hinge Traces –Examining Residuals – Understanding Relationships: Outliers – Nonlinear Monotonic Relationships.

UNIT 4 REEXPRESSION AND UNIVARIATE ANALYSIS 9 Hrs.

Choosing Re-expressions: Nonlinear Monotonic Functions – Nonmonotonic Functions. Introduction to Single variable: Distribution Variables – Numerical Summaries of Level and Spread -Scaling and Standardizing – Inequality - Univariate data: measures of center and spread, transformations.

UNIT 5 MULTIVARIATE ANALYSIS AND VISUALIZATION 9 Hrs.

Introducing a Third Variable – Causal Explanations – Three-Variable Contingency Tables and Beyond - Bivariate Analysis – Multivariate Analysis – Casual Analysis.

Basics of Matplotlib, Plotting with Pandas and Seaborn, Other Python Visualization Tools.

Max. 45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to:

CO1 - Understand about the significance of exploratory data analysis.

CO2 - Analyze the data statistically.

CO3 - Summarize data and understanding their relationship.

CO4 - Perform univariate data exploration and analysis.

CO5 - Apply multivariate data exploration and analysis.

CO6 - Perform data exploration and visualization techniques.

TEXT / REFERENCE BOOKS

1. Glenn J. Myatt, Wayne P. Johnson – Making Sense Of Data | A Practical Guide To Exploratory Data Analysis And Data Mining, Second Edition, 2014.
2. Frederick Hartwig, Brian E. Dearing , Exploratory Data Analysis, 1979.
3. Suresh Kumar Mukhiya, Usman Ahmed, Hands-On Exploratory Data Analysis with Python, Packt Publishing, 2020.
4. Jake Vander Plas, “Python Data Science Handbook: Essential Tools for Working with Data”, First Edition, O Reilly, 2017.
5. Catherine Marsh, Jane Elliott, Exploring Data: An Introduction to Data Analysis for Social Scientists, Wiley Publications, 2nd Edition, 2008.

END SEMESTER EXAMINATION QUESTION PAPER PATTERN

Max.Marks:100 Exam Duration: 3 Hrs.

PART A : 10 questions of 2 marks each –No choice **20 Marks**
each unit of internal choice; each carrying 16 marks **PART B :** 2 questions from
80 Marks