Postgraduate Diploma in Data Science (PGDDS)

I.Programme Structure

Sem	Code	Title of the Subject	L	TCP	P	IM	SM	TM	CD
I	CADD111	Introduction to Python	4			50	50	100	4
	CADD112	Introduction to Data Science	4			50	50	100	4
	CADD113	Introduction to Data warehouse and Data Mining	4			50	50	100	4
	CADD114	Applied Statistics using R	4	1		50	50	100	4
			16	**1**		**200**	**200**	**400**	**16**
II	CADD211	Data Analytics using Python	4			50	50	100	4
	CADD212	Machine Learning	3	1		50	50	100	4
	CADD213	Deep Learning	4			50	50	100	4
	CADD214	Technologies for Data Science	3	1		50	50	100	4
	CADD215	Big Data and Internet of Things	3	1		50	50	100	4
	CADD216J	Internship			10	50	50	100	9
			17	**3**	**10**	**300**	**300**	**600**	**29**
	Total Credits								**45**

a. Coding Scheme

PGD	X	X	X	X
Programme Code	Semester Number 1-2	Curriculum Revision Number 0-9	Course Serial Number 0-9	Course Type*

Course Type: T–Theory, I – Internal Papers, J – Internship..

II.SYLLABI IN DETAIL

I SEMESTER

INTRODUCTION TO PYTHON **4-0-0:100**

Introduction

Data Analytics using Python has become the most preferred and popular mode of data analysis since python provides a range of libraries making data analytics simple. Several packages like Numpy and Pandas makes data analytics applicable in python.

This course aims to deliver the necessary skills to perform data analytics in python.

Prerequisite

Python and Data Mining concepts.

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO1	Observe and Discuss the relevance of python in Data Analytics.	K1,K2
CO2	Recognize and Use the numpy package in python for data analytics.	K1,K3
CO3	Observe, Draft and Develop data manipulation applications using pandas.	K1,K4,K6
CO4	Enumerate and Practice hierarchical indexing in pandas.	K1,K3
CO5	Devise and Create data visualization with MATPLOTLIB.	K1,K6
CO6	Identify and Integrate various components of python to perform data analytics.	K1, K6

Mapping of CO with PO and PSO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	3	3	2	2	3	3	2	2	2	2.5
CO2	2	3	2	3	2	3	3	2	2	3	2.5
CO3	2	3	2	3	2	3	3	3	2	2	2.5
CO4	3	3	3	2	2	3	3	2	2	3	2.6
CO5	2	3	2	3	2	3	2	2	2	2	2.3
CO6	3	3	3	2	2	3	3	2	2	2	2.5
Mean Overall Score											2.48333333
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	20
Apply	10	10	20
Analyze	-	-	-
Evaluate	10	10	20
Create	10	10	20

Participatory Assessment

- Quiz
- Problem Solving

Course Content

Unit I

Features of Python - How to Run Python DATA TYPES AND OPERATIONS: Numbers-Strings-List-Tuple-Set-Dictionary. FUNCTIONS: Function Definition-Function Calling - Function Arguments - Anonymous Functions.

Unit II

Built-in Modules - Creating Modules - import Statement - Locating Modules - Namespaces and Scope - dir() function - reload() function - Packages in Python - Date and Time Modules.

Unit III

Class Definition - Creating Objects - Built-in Attribute Methods - Built-in Class Attributes - Destructors in Python-Encapsulation - Data Hiding-Inheritance - Method Overriding-Polymorphism.

Unit IV

match() function - search() function - Search and Replace - Regular Expression Modifiers: Option Flags - Regular Expression Patterns - find all() method - compile() method.

Unit V

Connecting to a Database – Creating Tables – Insert, Update, Delete and Read Operation – Disconnecting from a Database

TEXT

1. Jeeva Jose and P. Sojan Lal, “Introduction to Computing and Problem Solving with PYTHON”, Khanna Book Publishing Co. (P) Ltd., 2016.
2. Jake Vander Plas, “Python Data Science Handbook: Essential Tools for Working with Data”, 1st Edition, O'Reilly Media, 2016

REFERENCE

1. Wesley J. Chun, “Core Python Programming”, Second Edition, Prentice Hall Publication, 2006.
2. Timothy A Budd, “Exploring Python”, Tata McGraw Hill, New Delhi, 2011
3. Alberto Boschetti and Luca Massaron, “Python Data Science Essentials”, Packt publishing, 3rd Edition, 2018

Course Designer: Dr.A.George Louis Raja

INTRODUCTION TO DATA SCIENCE 4-0-0:100

Introduction

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models. The course on Introduction to Data Science provides an overview of Data Science, covering a broad selection of key challenges in and methodologies for working with big data.

This course is integrative across the core disciplines of Data Science, including databases, data warehousing, statistics, data mining, data visualization, high performance computing, Artificial Intelligence and Machine Learning.

Prerequisite

A basic background in computer programming and statistics

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Discover and Discuss the various technologies used in data science	K1.K2
CO 2	Recognize and Elicit the founding principles of Data Science.	K1,K2
CO 3	Identify, compare and correlate the Artificial intelligence concepts, searching and learning algorithms.	K1,K2,K4
CO 4	Identify and illustrate the methods and techniques commonly used in data science.	K1,K3
CO 5	Analyze and Evaluate how data analysis, inferential statistics, modeling, machine learning, and statistical computing can be utilized in an integrated capacity.	K4,K5
CO 6	Observe and Demonstrate the ability to clean and prepare data for analysis and assemble data from a variety of sources.	K1,K5

Mapping of CO with PO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	3	2	2	2	3	3	2	2	2	2.4
CO2	3	3	2	2	2	3	3	2	2	2	2.4
CO3	3	3	2	2	2	3	3	2	2	2	2.4
CO4	3	3	2	2	2	3	3	2	2	2	2.4
CO5	3	3	3	2	2	3	3	2	2	2	2.5
CO6	3	3	3	2	2	3	3	2	2	2	2.5
Mean Overall Score											2.4
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	30
Apply	10	10	10
Analyze	10	10	10
Evaluate	5	5	10
Create	5	5	20

Participatory Assessment

- Learn data collection techniques and data pre-processing from various domains.
- Problem Solving in BFS, DFS and Searching algorithms.
- Construct and formulate Natural Language Statements into syntax and semantics using Propositional logic.
- Practice and analyze real time problems using Machine Learning Algorithms.

Course Content

Unit I

Foundation of Data science, Area and Scope of Data Science, Steps of Data Science Process: Data collection, Preprocessing, training, and testing. Use cases in various domain such Image, Natural Language, Audio and Video.

Unit II

Introduction to Artificial Intelligence: Introduction Artificial Intelligence, The Foundations of AI, AI Technique, Production system characteristics, Production systems: 8-puzzle problem. Searching: Uniformed search strategies – Breadth first search, depth first search.

Unit III

Searching Algorithms and Learning : Local Search Algorithms: Generate and Test, Hill climbing, simulated annealing search, Constraint satisfaction problems, Greedy best first search, A* search, AO* search.

Unit IV

Learning Algorithms : Propositional logic - syntax & semantics Game Playing: Overview, Minimax algorithm, Alpha-Beta pruning, Additional Refinements.

Unit V

Metadata - Submission of Data - Access and Reuse of Data - Preservation of Data. Case Studies: GitHub, UCI Repository.

TEXT

1. Rachel Schutt, Cathy O'Neil, "Doing Data Science: Straight Talk from the Frontiline", Schroff/O'Reilly, 2013.

2. S. Russell and P. Norvig, “Artificial Intelligence A Modern Approach”, Second Edition. Pearson Education, 2007.

Note: Unit 5 notes will be compiled by course teacher.

Course Designer Prof. S. Anthony Philomen Raj

INTRODUCTION TO DATA WAREHOUSE AND DATA MINING 4-0-0:100

Introduction

Data mining is the analysis of data and the use of software techniques for finding patterns and regularities in sets of data.

This course is designed to expand students' knowledge and skills gained in database management courses and look in depth at data warehousing and data mining methods. The course examines the database architecture and technologies required for solving complex problems of data and information management, information retrieval, and knowledge discovery facing modern organizations.

Prerequisites

Database Architecture and Statistics.

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Remember, Understand and explain the fundamentals of Data Warehouse and Data Mining	K1, K2, K4
CO 2	Apply the concepts of association mining, clustering classification and Regression	K2, K3
CO 3	Analyze and choose a suitable data mining task for a specific problem and support the choice of approach adopted.	K4, K5
CO 4	Compare and Correlate the results various data mining techniques for a specific problem.	K2,K4
CO 5	Identify and Apply real-world problems in business and scientific information using data mining techniques	K1,K5
CO 6	Draft and Build statistical predictive models using various techniques such as neural networks, decision trees and logistic regression.	K4,K6

Mapping of CO with PO and PSO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	3	2	2	2	2	1	2	2	2	2.1
CO2	3	3	3	3	3	3	2	3	2	2	2.7
CO3	3	3	3	3	3	3	2	2	3	2	2.7
CO4	3	3	3	3	3	3	2	2	3	3	2.8
CO5	3	3	3	2	3	3	2	3	2	2	2.6
CO6	3	3	3	3	3	3	2	3	3	3	2.9
Mean Overall Score											2.6
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	20
Apply	10	10	20
Analyze	10	10	20
Evaluate	5	5	10
Create	5	5	10

Participatory Assessment

- Problem Solving in Association, classification and Clustering algorithms.
- Online Quiz
- Apply the KDD process for a specific problem.
- Case studies of various domains using these technologies to support business intelligence gathering and decision making are examined.

Course Content

Unit I

Data Warehouse – Definition – Multidimensional Data model – Data Cube – Dimensional Modelling – Lattice of Cuboids – Summary Measures – OLAP Operations – Slicing – Dicing – Drilling – Data Warehousing Architecture

Unit II

Data Mining – Definitions – KDD Vs Data Mining – Stages of KDD – Selection – Preprocessing – Transformation – Data Mining – Interpretation and Evaluation – Data Visualization Data Mining Techniques – Verification Model – Discovery Model – Discovery of Association Rules – Clustering – Discovery of Classification rules – Frequent Episodes – Deviation Detection – Issues and Challenges in Data Mining.

Unit III

Introduction – Association rules - Definitions – Support- Association rule – Methods to discover association rules – Problem decomposition – Frequent set – Maximal Frequent set – Border set – A Priori Algorithm – Candidate generation – Pruning – Example of APriori.

Unit IV

Introduction – Clustering Paradigms, Clustering Methods – Partitioning Algorithms – K-means, Hierarchical clustering - DBSCAN, Agglomerative clustering – AGNES, Divisive clustering – DIANA, Categorical Clustering Algorithms – STIRR.

Unit V

Classification – Basic Concepts – Decision Tree Induction – Attribute Selection Measures – Bayes' Classification methods – bayes' theorem – Rule Based Classification – Using IF_THEN rules for classification – Rule extraction from a decision tree.

TEXT

1. Data Mining Techniques, Arun K Pujari, University Press, 2001
2. Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2011.

Course Designer Dr. K. Saravanpriya

II SEMESTER

DATA ANALYTICS USING PYTHON **4-0-0:100**

Introduction

Data Analytics using Python has become the most preferred and popular mode of data analysis since python provides a range of libraries making data analytics simple. Several packages like Numpy and Pandas makes data analytics applicable in python.

This course aims to deliver the necessary skills to perform data analytics in python.

Prerequisite

Python, Data Mining concepts.

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Observe and Discuss the relevance of python in Data Analytics.	K1,K2
CO 2	Recognize and Use the numpy package in python for data analytics.	K1, K3
CO 3	Observe, Draft and Develop data manipulation applications using pandas.	K1,K4,K6
CO 4	Enumerate and Practice hierarchical indexing in pandas.	K1,K3
CO 5	Devise and Create data visualization with MATPLOTLIB.	K1,K6
CO 6	Identify and Integrate various components of python to perform data analytics.	K1, K6

Mapping of CO with PO and PSO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	3	3	2	2	3	3	2	2	2	2.5
CO2	2	3	2	3	2	3	3	2	2	3	2.5
CO3	2	3	2	3	2	3	3	3	2	2	2.5
CO4	3	3	3	2	2	3	3	2	2	3	2.6
CO5	2	3	2	3	2	3	2	2	2	2	2.3
CO6	3	3	3	2	2	3	3	2	2	2	2.5
									Mean Overall Score		2.5
									Result		High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	20
Apply	10	10	20
Analyze	-	-	-
Evaluate	10	10	20
Create	10	10	20

Participatory Assessment

- Quiz in basics of numpy and pandas.
- Problem Solving in data manipulation, hierarchical indexing and visualization.
- Designing programs performing data analytics.

Course Content

Unit I

Introduction to Numpy - Basics of NumPy Array – Computation on NumPy Array – Aggregations – Broadcasting – Comparisons, Masks and Boolean Logic – Sorting Arrays – NumPy Structured Array.

Unit II

Introducing Panda Objects – Data Indexing and Selection - Operating Data on Pandas – Handling Missing Data

Unit III

Combining Data Sets – Vectorized String Operations – Working with Time Series.

Unit IV

Simple Line Plots – Simple Scatter Plots – Density and Contour Plots – Histograms, Binnings and Density

Unit V

Customising Color bars – Multiple Subplots – Text and Annotation – Three Dimension Plotting in Matplotlib – Geographic Data with Base Map – Visualization with Seaborn

TEXT

1. Jeeva Jose and P. Sojan Lal, "Introduction to Computing and Problem Solving with PYTHON", Khanna Book Publishing Co. (P) Ltd., 2016.

2. Jake Vander Plas, "Python Data Science Handbook: Essential Tools for Working with Data", 1st Edition, O'Reilly Media, 2016

REFERENCE

1. Wesley J. Chun, "Core Python Programming", Second Edition, Prentice Hall Publication, 2006.

2. Timothy A Budd, "Exploring Python", Tata McGraw Hill, New Delhi, 2011

3. Alberto Boschetti and Luca Massaron, "Python Data Science Essentials", Packt publishing, 3rd Edition, 2018

Course Designer: Dr.A.George Louis Raja

MACHINE LEARNING 3-1-0:100

Introduction

The objective of this course is to introduce the fundamentals of Machine Learning and Algorithms in computing environment. It enables the learner to develop machine learning techniques associated with the computing for the classification and clustering. It also covers the supervised and unsupervised algorithms.

Prerequisite

Python, Data Mining concepts.

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Understand and Comprehend the Machine Learning Concepts	K1, K2
CO 2	Identify the use cases of the supervised and unsupervised learning algorithms	K3
CO 3	Analyse the logic behind the execution of various classifiers	K4
CO 4	Compute and compare the performance of different algorithms for mining data	K2, K6
CO 5	Demonstrate and analyse the clustering methods	K3, K4
CO 6	Propose solution for real word problems	K5

Mapping of CO with PO and PSO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	3	2	2	1	3	2	3	1	1	2.1
CO2	3	3	3	3	1	3	2	3	2	1	2.4
CO3	3	2	1	3	1	3	2	3	1	1	2
CO4	3	3	3	2	1	3	2	2	1	1	2.1
CO5	3	3	3	2	1	3	1	2	1	1	2
CO6	3	3	3	2	1	3	3	2	1	1	2.2
Mean Overall Score											2.133333333
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	05	05	10
Understand	05	05	10
Apply	20	20	30
Analyze	10	10	30
Evaluate	05	05	10
Create	05	05	10

Unit I

Introduction: Machine learning – Examples and Applications - Perspectives and Issues in Machine learning - Input: Concepts, Instances, and Attributes - Output: Knowledge Representation - Training and Testing – Predicting Performance

Unit II

Decision Tree Learning: Decision tree representation – Decision tree learning – Random forest - Bayesian Learning: Naïve Bayes classifier - k- nearest neighbour Learning - Case based reasoning

Unit III

Artificial Neural Network - Introduction – Neural Network Representation - Multilayer Networks and Back propagation Algorithm - Linear models for Regression- PCA

Unit IV

SVM : Introduction – Kernel methods - formulation and computation- SVM Linear classifier – SVM with two variables – Clustering Methods.- Introduction – K- Means - Hierarchical Clustering - Choosing the Number of Clusters

Unit V

Deep Learning – Convolutional neural network– Auto encoders – Recurrent Neural Network - Use Cases: Finding similar users in Twitter (Mahout), Email marketing system (Mahout)

TEXT

1. Shalev-Shwartz, Shai, and Shai Ben-David, "Understanding machine learning: From theory to algorithms", Cambridge university press, 2014.

2. Duda, Richard O., Peter E. Hart, and David G. Stork, "Pattern classification", John Wiley & Sons, 2012.

3. Witten, Ian H., et al, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, 2016.

Course Designer Prof. A. John Martin

DEEP LEARNING 4-0-0:100

Introduction

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain. Deep Learning uses mathematical functions to map the input to the output. These functions can extract non-redundant information or patterns from the data, which enables them to form a relationship between the input and the output.

This Course aims to deliver the basic concepts of deep learning, Deep Feed forward networks, learning and optimization algorithms, regularization patterns, convolutional networks and applications of deep learning.

Prerequisite

- Optimization Techniques
- Graph Theory
- Data Analytics

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Observe and Apply the concepts of Deep Learning and its applications	K1,K3
CO 2	Identify and Use the configuration of Deep Feed Forward Networks	K1,K3
CO 3	Comparing and Correlating the Learning and Optimization Algorithms	K2,K4
CO 4	Identify and Practice Regularization for Deep Learning	K1,K3
CO 5	Observe and Discuss the concepts of Convolutional networks	K1,K2
CO 6	Identify, Analyze and Evaluate the applications of Deep Learning	K1,K4,K5

Mapping of CO with PO and PSO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	1	2	3	3	3	1	2	3	3	2.4
CO2	3	1	2	3	2	3	1	2	3	2	2.2
CO3	3	1	1	1	1	3	1	1	1	1	1.4
CO4	3	1	1	1	1	3	1	1	1	1	1.4
CO5	3	1	1	1	1	3	1	1	1	1	1.4
CO6	3	1	1	1	1	3	1	1	1	1	1.4
Mean Overall Score											2
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100)
	I CA (50)	II CA (50)	
Remember	15	15	30
Understand	10	10	20
Apply	15	15	30
Analyze	5	5	10
Evaluate	5	5	10
Create	-	-	-

Participatory Assessment

- Paper work has to be carried out in Applied Mathematics .
- Problem solving in applied mathematics.
- Problem solving using deep learning methods or algorithms.
- MCQ

Course Content

Unit I

Deep Feed forward Networks: Learning XOR – Gradient Based Learning – Architecture Design – Back Propagation and other Differentiation Algorithms.

Unit II

Regularization for Deep Learning: Dataset Augmentation – Noise Robustness – Semi Supervised Learning – Multi Task Learning – Early Stopping.

Unit III

Optimization for Training Deep Models: How Learning Differs from Pure Optimization – Challenges in Neural Network Optimization – Basic Algorithms – Parameter Initialization Strategies – Algorithms with Adaptive Learning Rates.

Unit IV

Convolutional Networks: Convolution Operation – Motivation – Pooling – Convolution and Pooling as an Infinitely Strong Prior – Variants of the Basic Convolution Function.

Unit V

Applications: Large Scale Deep Learning – Computer Visions – Speech Recognition –Natural Language Processing – Other Applications.

TEXT

Ian Goodfellow, Yoshua Bengio, and Aaron Courville, “Deep Learning-Adaptive Computation and Machine Learning”, 2015

REFERENCE

1. Ethem Alpaydin, "Introduction to Machine Learning" Second Edition, The MIT Press, 2009.
2. Tom M. Mitchell, "Machine Learning", First Edition, Tata McGraw-Hill Education, 2013.
3. Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2007.
4. Mevin P. Murphy, "Machine Learning: A Probabilistic Perspective", The MIT Press, 2012.

Course Designer Prof. V. Thomas Immanuel

TECHNOLOGIES FOR DATA SCIENCE 3-1-0:100

Introduction

Data science uses techniques such as machine learning and artificial intelligence to extract meaningful information and to predict future patterns and behaviors. This course will cover the various technologies using in data science and installation of HIVE, SQOOP and PIG tools.

Prerequisite

Basics of Data Science

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Identify and Discover the various technologies used in Data Science	K1,K2
CO 2	Recognize and Discuss Big Data and its analytics in the real world	K1,K2
CO 3	Identify, Draft and Develop Big Data Solutions using Hadoop	K1,K4, K6
CO 4	List and Leverage Hadoop as a reliable, scalable MapReduce framework	K1,K3
CO 5	Demonstrate and Install and interact the HIVE, SQOOP and PIG tools	K2,K5
CO 6	Apply and Demonstrate the ability to clean and prepare data for analysis using HIVE, SQOOP and PIG tools	K3,K5

Mapping of CO with PO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	2	2	2	2	3	2	2	2	2	2.2
CO2	3	2	2	2	2	3	2	2	2	2	2.2
CO3	3	2	2	2	2	3	2	2	2	2	2.2
CO4	3	2	2	2	2	3	2	2	2	2	2.2
CO5	3	2	2	2	2	3	2	2	2	2	2.2
CO6	3	2	2	2	2	3	2	2	2	2	2.2
Mean Overall Score											2.2
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	30
Apply	10	10	10
Analyze	10	10	10
Evaluate	5	5	10
Create	5	5	20

Participatory Assessment

- HIVE installation and commands.
- SQOOP installation and commands.
- PIG installation and commands

Course Content

Unit I

Big Data and Hadoop : Hadoop architecture, Hadoop Versioning and configuration, Single node & Multi-node Hadoop, Hadoop commands, Models in Hadoop, Hadoop daemon, Task instance, illustrations.

Unit II

Map-Reduce : Framework, Developing Map-Reduce course, Life cycle method, Serialization, Running Map Reduce in local and pseudo-distributed mode, illustrations.

Unit III

HIVE : Installation, data types and commands, illustration.

Unit IV

SQOOP: Installation, importing data, Exporting data, Running, illustrations

Unit V

PIG : Installation, Schema, Commands, illustrations.

Text

1. Chuck Lam, "Hadoop in Action", 2010, ISBN: 9781935182191

2. Jimmy Lin and Chris Dyer, "Data- intensive Text Processing with Map Reduce", Morgan & Claypool Publishers, 2010.

Course Designer Prof. S. Anthoiny Philomen Raj

BIG DATA AND INTERNET OF THINGS 3-1-0:100

Introduction

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

The Internet of Things (IoT) is the ability to have devices communicate with one another via the internet or other networks, remotely tracking information to provide feedback to assist with decision making for commercial, industrial and residential purposes.

This course is for those new to data science and interested in understanding why the Big Data Era has come to be. It is for those who want to become conversant with the terminology and the core concepts behind big data problems, applications, and systems and to understand the architecture of IoT devices and deals with the basics of electronics and programming.

Prerequisite

Data Structures, Data Mining, Computer Networks.

Course Outcomes

At the end of this course, the students will be able to

CO. No.	Course Outcome Statement	Cognitive Level
CO 1	Observe and Discuss the role of Big Data and IoT.	K1,K2
CO 2	Recognize and Correlate the Big Data and analytics in the real world.	K1,K4
CO 3	Observe, Analyze and Evaluate the Big Data framework like Hadoop and NOSQL to efficiently store and process Big Data to generate analytics.	K1,K4,K5
CO 4	Understand and Apply the application areas of IOT	K1,K3
CO 5	Determine and Justify the revolution of IoT Devices.	K3,K5
CO 6	Identify and Demonstrate the building blocks of Internet of Things and characteristics.	K1,K2

Mapping of CO with PO

CO	Programme Outcomes (PO)					Programme Specific Outcomes (PSO)					Mean Scores of COs
	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4	PSO5	
CO1	3	2	2	2	1	3	2	3	1	2	2.1
CO2	3	3	2	3	1	3	3	1	1	2	2.2
CO3	3	2	2	2	1	3	2	1	1	1	1.8
CO4	3	3	2	3	1	3	2	1	2	2	2.2
CO5	3	3	3	2	1	3	2	1	1	2	2.1
CO6	3	2	2	3	1	3	2	1	2	1	2
Mean Overall Score											2.1
Result											High

Assessment Pattern

Bloom's Category	CA Tests (Marks Allotment)		Term End Exam (100) Marks Allotment
	I CA (50)	II CA (50)	
Remember	10	10	20
Understand	10	10	20
Apply	10	10	20
Analyze	10	10	20
Evaluate	10	10	20
Create	-	-	-

Participatory Assessment

- Paper work to be prepared in related to Big data and IoT.
- MCQ can be practiced

Course Content

Unit I

Big Data Science-Historical Review of Big Data-Historical Interpretation of Big Data-Defining Big Data From 3Vs to 32Vs-Big Data Analytics and Machine Learning-Big Data Analytics and Cloud Computing-Hadoop, HDFS, mapreduce, Spark, and Flink-Database Techniques for Big Data-nosql Movement-nosql Solutions for Big Data Management-nosql Data Models-Future Directions.

Unit II

Resource Management in Big Data Processing Systems-Types of Resource Management-Big Data Processing Systems and Platforms Single-Resource Management in the Cloud-Multiresource Management in the Cloud-Related Work on Resource Management-Open Problems-Local Resource Consumption Shaping: A Case for MapReduce-Motivation-Local Resource Shaper-Evaluation.

Unit III

IoT Architectures – IoT Functional Stack, Sensors, and Actuators Layer, Communications Network Layer, Applications and Analytics Layer – IoT Data Management and computer Sack, Fog Computing, Edge Computing, Cloud Computing - Smart Objects, Sensor Networks.

Unit IV

Design Methodology – Case study – Basic blocks of IoT device – Arduino – Raspberry Pi – Board, Interfaces, Setting up, Programming – Other IoT Devices.

Unit V

Introduction, Functional View, Information View, Deployment and Operational View, Other Relevant architectural views.

TEXT

1. Rajkumar Buyya,Rodrigo N. Calheiros,Amir Vahid Dastjerdi "Big Data Principles and Paradigms", Elsevier 2016.

2. Jan Holler, VlasiosTsiatsis, Catherine Mulligan, Stefan Avesand, Stamatis Karnouskos, David Boyle, "From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence", 1st Edition, Academic Press, 2014.

Course Designer Prof. R.Veeraragavan