



Estd:1980

SAGI RAMA KRISHNAM RAJU ENGINEERING COLLEGE (AUTONOMOUS)

(Affiliated to JNTUK, Kakinada), (Recognized by AICTE, New Delhi)

UG Programmes CE,CSE,ECE,EEE,IT & ME are Accredited by NBA

CHINNA AMIRAM (P.O):: BHIMAVARAM :: W.G.Dt., A.P., INDIA :: PIN: 534 204

Regulation: R20									
INFORMATION TECHNOLOGY (Honors)									
SCHEME OF INSTRUCTION & EXAMINATION (With effect from 2020-21 admitted Batch onwards)									
Course Code	Course Name	Year/ Sem	Cr	L	T	P	Int. Marks	Ext. Marks	Total Marks
B20ITH101	Advanced Data Structures	II-II	4	3	1	0	30	70	100
B20ITH201	Statistical Foundations for Data Science	III-I	4	3	1	0	30	70	100
B20ITH301	Mining Massive Data Sets	III-II	4	3	1	0	30	70	100
B20ITH401	Data Visualization	IV-I	4	3	1	0	30	70	100
B20ITH501	*MOOCS-I	II-II to IV-II	2	--	--	--	--	--	100
B20ITH601	*MOOCS-II	II-II to IV-II	2	--	--	--	--	--	100
TOTAL			20	12	4	0	120	280	600

*Two MOOCS courses of any INFORMATION TECHNOLOGY related Program Core Courses from NPTEL/SWAYAM with a minimum duration of 8 weeks (2 Credits) courses other than the courses offered need to be taken by prior information to the concern. These courses should be completed between II Year II Semester to IV Year II Semester

Code	Category	L	T	P	C	I.M	E.M	Exam
B20ITH101	Honors	3	1	-	4	30	70	3 Hrs

ADVANCED DATA STRUCTURES

(Honors Degree Course in IT)

Course Objectives: The objectives of the course are to impart:

1.	Describe and implement a variety of advanced data structures (hash tables, priority queues, balanced and digital search trees).
2.	Analyze the space and time complexity of the algorithms studied in the course.
3.	Identify different solutions for a given problem; analyze advantages and disadvantages to different solutions.
4.	Demonstrate an understanding of external sorting and string matching algorithms.

Course Outcomes: After completion of the course, students will be able to

S. No	Outcome	Knowledge Level
1	Be able to understand and apply asymptotic analysis on data structures, including search trees, heaps, and dictionaries.	K2
2	Understand the implementation and complexity analysis of external sorting and string matching algorithms.	K3
3	Have an idea of applications of data structures and algorithms in a variety of areas, including linear programming and duality, string matching, game -theory.	K2

SYLLABUS

UNIT-I (8 Hrs)	Algorithm Analysis and Hashing Introduction to Algorithm Analysis, Step Counts, Asymptotic Notations, Amortized Analysis (Textbook 2) Dictionaries, Hashing, Hash Functions, Open Hashing, Closed Hashing, Extendible Hashing (Text Book 1 & Reference 1)
UNIT-II (8 Hrs)	Priority queues and tournament Trees: introduction to priority queues and ADT, implementation with lists, Binary Heaps, operations, Build Heap, performance analysis, Binomial Queues: operations, Amortized analysis, Lazy Binomial Queues (Textbook 2 & Reference 5) Tournament Trees: Winner Trees and Loser Trees (Textbook 2)
UNIT-III (8 Hrs)	Efficient Binary Search Trees (Textbook 1) AVL Trees, Red-Black Trees and Splay Trees: Introduction, Operations, Maximum Height, Performance Analysis
UNIT-IV (8 Hrs)	Multiway and Digital Search Trees and External Sorting (Text Book 1) Multiway Search Trees: B-Trees and B+-Trees Digital Search Trees: Digital Search Trees, Binary Tries, PARTRICA and Multiway Tries introduction, k-way merging, buffer handling, run generation, optimal merging of runs

UNIT-V (8 Hrs)	String Matching (Text Book 3) String Operations, The Knuth-Morris-Pratt Algorithm, The Boyer-Moore Algorithm, Standard Tries, Compressed Tries, Suffix Tries, The Huffman Coding Algorithm, The Longest Common Subsequence Problem (LCS)
Text Books:	
1.	Fundamentals of Data Structures in C: Second Edition, Horowitz, Sahani, Anderson Freed, Universities Press.
2.	Data Structures, Algorithms and Applications in C++, Second Edition, Sartaj Sahani, Universities Press.
Reference Books:	
1.	Data structures and Algorithm Analysis in C, 2nd edition, Mark Allen Weiss, Pearson
2.	Introduction to Algorithms”, T. Cormen, R.Rivest, C. Stein, C. Leiserson, PHI publication, Second Edition, 004, ISBN 81-203-2141-3.
3.	Data Structures, a Pseudo code Approach, Richard F Gilberg, Behrouz A Forouzan, Cengage.



II B.Tech. II Semester MODEL QUESTION PAPER

ADVANCED DATA STRUCTURES

(Honors Degree course in IT)

Time: 3 Hrs

Max. Marks:70

Answer ONE Question from EACH UNIT

All questions carry equal marks

Assume suitable data if necessary

			CO	KL	M
UNIT-I					
1.	a).	Identify the basic requirements to achieve good Hashing mechanism	1	2	7
	b).	Apply Following elements 76, 40, 48, 05, 55 to inserted into an empty hash table with hash function $f(x) = x \% 7$ for quadratic probing.	1	3	7
OR					
2.		Explain the following overflow handling techniques with suitable examples? (i) Open Addressing. (ii) Chaining.	1	2	14
UNIT-II					
3.	a).	Construct the min and max priority queue with the following elements 20,10,5,18,6,12,14,4 and 22	1	3	7
	b).	Construct a binary heap with the following data 150, 110, 90, 80, 70, 100, 180	1	3	7
OR					
4.	a).	Show the result of constructing a binomial heap using the following elements 9, 11, 1, 13, 5, 4, 7, 14, 2, 8, 6, 3, 10, 12, and 15 one at a time, into an initially empty binomial heap.	1	3	7
	b).	Show the resultant Binomial heap after perform delete minimum element and reconstruct the binomial heap	1	3	7
UNIT-III					
5.	a).	Create an AVL Tree using the following data entered as a sequence set. Show the balance factors in the resulting tree: 13, 22, 6, 9, 32, 55, 79, 65, 70	1	3	7
	b).	Insert 42, 43, 46 and 49 in the above constructed AVL tree and show a balanced AVL Tree.	1	3	7
OR					
6.	a).	Create a Red-Black tree by inserting the following sequence of numbers 8, 18, 5, 15, 17, 25, 40 and 80 .	1	3	7
	b).	Explain the operations of Splay tree with an examples	1	2	7

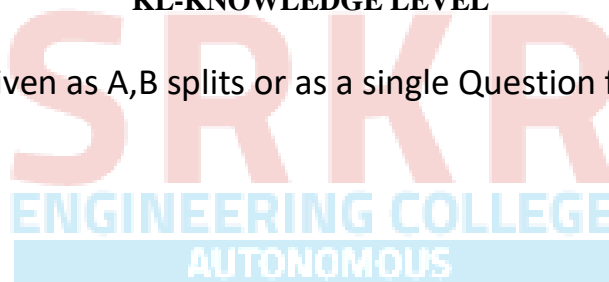
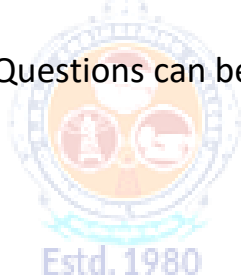
UNIT-IV					
7.	a).	Explain the insertion operation in B+ tree with suitable example?	2	2	7
	b).	Explain K-Way merging with suitable example?	2	2	7
OR					
8.	a).	Explain the insertion, deletion and search operations on Digital Search Trees with an example?	2	2	7
	b).	Briefly explain the construction of Multi-way tries with an example?	2	2	7
UNIT-V					
9.	a).	Explain working principal of Knuth Morris Pratt algorithm with example.	3	2	7
	b).	Suppose we have the weights $q_1=22$, $q_2=5$, $q_3=11$, $q_4=19$, $q_5=2$, $q_6=11$, $q_7=25$ and $q_8=5$. Write the step by step process to construct the Huffman tree?	3	3	7
OR					
10.	a).	Illustrate Boyers Moore algorithm with an example	3	3	7
	b).	Discuss about standard tries, Compressed tries and suffix tries with an examples	3	3	7

CO-COURSE OUTCOME

KL-KNOWLEDGE LEVEL

M-MARKS

NOTE: Questions can be given as A,B splits or as a single Question for 14 marks



Code	Category	L	T	P	C	I.M	E.M	Exam
B20ITH201	Honors	3	1	--	4	30	70	3 Hrs.

STATISTICAL FOUNDATIONS FOR DATA SCIENCE

(Honors Degree course in IT)

Course Objectives:

The course will introduce the fundamental concepts of probability and statistics required for a program in data science

Course Outcomes: After completion of the course, the student will be able to

S.No	Outcome	Knowledge Level
1	Use the probability and statistical concepts in the field of data science.	K3
2	Employ the techniques and methods related to the area of data science in variety of applications.	K3
3	Apply logical thinking to understand and solve the problem in context.	K3
4	Explore statistical learning methods and their application to modern problems in science, industry and society.	K3
5	Build analytics pipelines for regression problems and classification problems	K3

SYLLABUS

UNIT-I (12 Hrs)	Basics of Data Science: Introduction, Typology of problems, Importance of linear algebra, statistics and optimization from a data science perspective, Structured thinking for solving data science problems.
UNIT-II (10Hrs)	Probability, Statistics and Random Processes, Probability theory and axioms, Random variables, Probability distributions and density functions, Expectations, moments, characteristic function, Covariance and correlation. Statistics and sampling distributions, Hypothesis testing of means, proportions, variances, Chi-Square test and Confidence intervals.
UNIT-III (14Hrs)	Correlation functions- Autocorrelation and Cross correlation, Probabilistic formulations of prediction problems, Plug-in estimators, empirical risk minimization, Linear threshold functions, Risk bounds, Concentration inequalities, Rademacher averages.
UNIT-IV (12Hrs)	Linear Regression, Regularization and linear model selection, Feature selection methods, Cross-Validation, Game-theoretic formulations of prediction problems, High dimensional methods, Lasso, Ridge Regression, Dimensionality reduction, Minimax strategies for log loss, linear loss and Quadratic loss.
UNIT-V (12Hrs)	Neural networks: Stochastic gradient descent, Combinatorial dimensions and Rademacher averages, Recurrent neural networks-Vanila and Multilayer RNNs.

Text Books:	
1.	Bendat, J. S. and A. G. Piersol. Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA, 2010.
2.	Montgomery, D. C. and G. C. Runger. Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA, 2011.
3.	James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013.
Reference Books:	
1.	Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer, 2009.
2.	S. C. Gupta and V. K. Kapoor: Fundamentals of Mathematical Statistics, First edition 1970, S Chand & Son.
3.	Jianqing Fan, Ruze Li, Cun-Hui Zhang and hui Zou: Statistical Foundations of Data Science. First edition published 2020.



III B.Tech. I Semester MODEL QUESTION PAPER

STATISTICAL FOUNDATIONS FOR DATA SCIENCE

(Honors Degree course in IT)

Time: 3 Hrs

Max. Marks:70

Answer ONE Question from EACH UNIT

All questions carry equal marks

Assume suitable data if necessary

			CO	KL	M
UNIT-I					
1.	a).	Define Data Science and explain typology of the problems	1	2	7
	b).	Write the Importance of linear algebra in data science	1	3	7
OR					
2.		Explain about Structural thinking for solving data science problems.	1	3	14
UNIT-II					
3.	a).	Define the following i) Probability ii) Moment generating function iii) Covariance and iv) Expectation	2	2	7
	b).	A sample of size 400 was drawn and the sample mean was found to be 99. Test whether this sample could have come from a normal population with mean 100 and standard deviation 8 at 5% level of significance.	2	3	7
OR					
4.		Fit a Poisson distribution to the following data and write the goodness of fit X: 0 1 2 3 4 F: 109 65 22 3 1	2	3	14
UNIT-III					
5.	a).	Explain about auto and cross correlation functions	3	3	7
	b).	Define Rademacher average and discuss its structural properties	3	2	7
OR					
6.		Discuss the following concentration inequalities i) Markov's , i) Chebyshev	3	3	14
UNIT-IV					
7.		What is Cross-Validation and discuss the following i) K-fold Validation ii) Leave-One-Out Cross- Validation iii) Repeated K-fold Cross-Validation	4	2	14
OR					
8.	a).	Prove that Ridge regression estimator is equal to $\widehat{\beta}_\lambda = X^T [XX^T + \lambda I]^{-1} Y$ and the fitted value of Y at X is $\widehat{y} = x^T \widehat{\beta}_\lambda = x^T X^T [XX^T + \lambda I]^{-1} Y$	4	3	7

	b).	Explain about linear regression and regularization technique of weight delay.	4	3	7
		UNIT-V			
9.		Discuss about Stochastic gradient descent methods	5	3	14
		OR			
10.		Discuss the following concepts i) Rademacher averages, ii) Vanila Recurrent neural networks.	5	3	14

CO-COURSE OUTCOME

KL-KNOWLEDGE LEVEL

M-MARKS

NOTE: Questions can be given as A,B splits or as a single Question for 14 marks



Code	Category	L	T	P	C	LM	E.M	Exam
B20ITH301	Honors	3	1	--	4	30	70	3 Hrs.

MINING MASSIVE DATA SETS

(Honors Degree Course in IT)

Course Objectives:

1.	The course will discuss data mining and machine learning algorithms for analyzing very large amounts of data.
----	---

Course Outcomes: by the end of this course students can able to

S.No	Outcome	Knowledge Level
1.	Understand how data mining techniques are implemented in MapReduce environment.	K2
2.	Explore different approaches for finding similar items.	K4
3.	Outline different mining techniques to mine data streams.	K3
4.	Summarize various approaches for finding frequent item sets.	K3
5.	Illustrate various dimensionality reduction techniques on massive data.	K3

SYLLABUS

UNIT-I (10Hrs)	Data Mining: Data Mining, Statistical Limits on Data Mining, MapReduce: Distributed File Systems, MapReduce, Algorithms Using MapReduce, Extensions to MapReduce. .
UNIT-II (10 Hrs)	Finding Similar Items: Applications of Near-Neighbor Search, Shingling of Documents, Distance Measures, Theory of Locality-Sensitive Functions, Applications of LSH Hashing.
UNIT-III (10 Hrs)	Mining Data Streams: Stream Data Model, Sampling Data in Streams, Filtering Streams, Link Analysis: PageRank, Efficient Computational of PageRank, Link Spam, Hubs and Authorities.
UNIT-IV (10 Hrs)	Frequent Item sets: Market-Based Model, Market Based and A-Priori Algorithm, Limited- Pass Algorithms, Clustering: Introduction, Hierarchical Clustering and K-means Algorithm, CURE Algorithm
UNIT-V (10 Hrs)	Dimensionality Reduction: Eigenvalues and Eigenvectors, Principal-Component Analysis, CUR Decomposition, Large-Scale Machine Learning: Machine Learning Model, Perceptrons, SVM's, Nearest Neighbors.

Textbooks:

1.	Jure Leskovec, Anand Rajaraman, Jeffery D. ULLman, Mining of Massive Datasets, Cambridge University Press, 2014.
2.	Pattern Recognition and Machine Learning. Christopher Bishop. Springer-Verlag New York.

	2006.
Reference Books:	
1.	Machine Learning: A Probabilistic Perspective. Kevin Murphy. MIT Press. 2012
2.	The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer. 2013
e-Resources	
1.	https://www.edx.org/course/mining-massive-datasets
2.	https://www.cambridge.org/core/books/mining-of-massivedatasets/C1B37BA2CBB8361B94FDD1C6F4E47922



III B.Tech. II Semester MODEL QUESTION PAPER

MINING MASSIVE DATA SETS

(Honors Degree course in IT)

Time: 3 Hrs

Max. Marks:70

Answer ONE Question from EACH UNIT

All questions carry equal marks

Assume suitable data if necessary

			CO	KL	M
UNIT-I					
1.	a).	Define Data mining? Explain statistical limits on data mining?	1	2	7
	b).	Explain Distributed file system of MapReduce	1	2	7
OR					
2.	a).	Explain the details of execution of a MapReduce program.	1	2	7
	b).	How to use MapReduce program for relational algebra operations	1	2	7
UNIT-II					
3.	a).	Describe similarity of documents	2	2	6
	b).	Explain the following shingling of documents 1. K-shingles 2.Hashing shingles	2	2	8
OR					
4.	a).	Explain Euclidean distance, Cosine distance	2	2	7
	b).	Describe applications of Locality-Sensitive-Hashing	2	3	7
UNIT-III					
5.	a).	Explain data-stream-management system with neat diagram	3	3	6
	b).	Describe the Bloom Filter? Explain Analysis of Bloom Filtering	3	3	8
OR					
6.	a).	Explain Link analysis in Data mining	3	3	7
	b).	Explain how to compute Page Rank efficiently	3	3	7
UNIT-IV					
7.	a).	Explain Apriori algorithm with an example	4	3	7
	b).	Explain limited pass simple, randomized algorithm	4	3	7
OR					
8.	a).	Explain Hierarchical clustering	4	2	7
	b).	Explain CURE Algorithm	4	2	7
UNIT-V					
9.	a).	Describe Eigen values and Eigenvectors of Symmetric Matrices	5	2	7
	b).	Explain Principal-Component Analysis	5	2	7

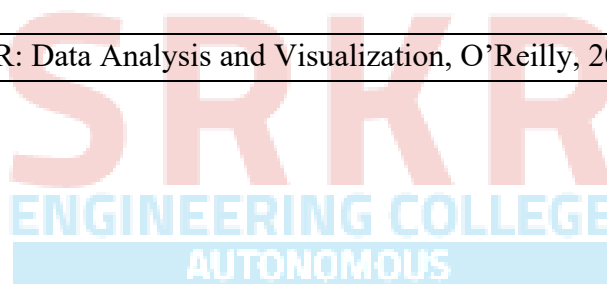
		OR			
10.	a).	Explain CUR Decomposition	5	2	7
	b).	Explain un-supervised machine learning models	5	2	7
		CO-COURSE OUTCOME	KL-KNOWLEDGE LEVEL		M-MARKS

NOTE: Questions can be given as A,B splits or as a single Question for 14 marks



Code	Category	L	T	P	C	I.M	E.M	Exam
B20ITH401	Honors	3	1	--	4	30	70	3 Hrs.
DATA VISUALIZATION								
(Honors Degree Course in IT)								
Course Objectives:								
1.	The main objective of this course is to make it easier to identify patterns, trends and outliers in large data sets							
Course Outcomes								
S. No	Outcome							Knowledge Level
1.	Understand basics of Data Visualization							K2
2.	Implement visualization of distributions							K3
3.	Write programs on visualization of time series, proportions & association							K3
4.	Apply visualization on Trends and uncertainty							K3
5.	Explain principles of proportions							K3
SYLLABUS								
UNIT-I (10Hrs)	INTRODUCTION TO VISUALIZATION: Visualizing Data-Mapping Data onto Aesthetics, Aesthetics and Types of Data, Scales Map Data Values onto Aesthetics, Coordinate Systems and Axes- Cartesian Coordinates, Nonlinear Axes, Coordinate Systems with Curved Axes, Color Scales-Color as a Tool to Distinguish, Color to Represent Data Values, Color as a Tool to Highlight, Directory of Visualization Amounts, Distributions, Proportions, x–y relationships, Geospatial Data							
UNIT-II (10 Hrs)	VISUALIZING DISTRIBUTIONS: Visualizing Amounts-Bar Plots, Grouped and Stacked Bars, Dot Plots and Heat-maps, Visualizing Distributions: Histograms and Density Plots- Visualizing a Single Distribution, Visualizing Multiple Distributions at the Same Time, Visualizing Distributions: Empirical Cumulative Distribution Functions and Q-Q Plots-Empirical Cumulative Distribution Functions, Highly Skewed Distributions, Quantile Plots, Visualizing Many Distributions at Once-Visualizing Distributions Along the Vertical Axis, Visualizing Distributions Along the Horizontal Axis							
UNIT-III (10 Hrs)	VISUALIZING ASSOCIATIONS: Visualizing Proportions- A Case for Pie Charts, A Case for Side-by-Side Bars, A Case for Stacked Bars and Stacked Densities, Visualizing Proportions Separately as Parts of the Total , Visualizing Nested Proportions- Nested Proportions Gone Wrong, Mosaic Plots and Tree-maps, Nested Pies ,Parallel Sets. Visualizing Associations Among Two or More Quantitative Variables-Scatter plots, Correlograms, Dimension Reduction, Paired Data.							

UNIT-IV (10 Hrs)	VISUALIZING TIME SERIES & UNCERTIANITY: Visualizing Time Series and Other Functions of an Independent Variable-Individual Time Series , Multiple Time Series and Dose– Response Curves, Time Series of Two or More Response Variables ,Visualizing Uncertainty-Framing Probabilities as Frequencies, Visualizing the Uncertainty of Point Estimates, Visualizing the Uncertainty of Curve Fits, Hypothetical Outcome Plots
UNIT-V (10 Hrs)	PRINCIPLE OF PROPORTIONAL INK: The Principle of Proportional Ink- Visualizations Along Linear Axes, Visualizations Along Logarithmic Axes, Direct Area Visualizations, Handling Overlapping Points Partial Transparency and Jittering, 2DHistograms, Contour Lines.
Textbooks:	
1.	Claus Wilke, “Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures”, 1st edition, O’Reilly Media Inc, 2019.
2.	OssamaEmbarak, Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems, Apress, 2018
Reference Books:	
1.	Tony Fischetti, Brett Lantz, R: Data Analysis and Visualization, O’Reilly, 2016



Course Code: B20ITH401

SAGI RAMA KRISHNAM RAJU ENGINEERING COLLEGE (A)

R20

IV B.Tech. I Semester MODEL QUESTION PAPER

DATA VISUALIZATION

(Honors Degree course in IT)

Time: 3 Hrs.

Max. Marks: 70 M

Answer **ONE Question** from **EACH UNIT**

All questions carry equal marks

Assume suitable data if necessary

			CO	KL	M
UNIT-I					
1.	a).	How Data-Mapping Data onto Aesthetics	1	2	7
	b).	How Cartesian coordinates, and nonlinear axes used for Data Visualization	1	2	7
OR					
2.	a).	Explain how color can be used to represent data values and tool to highlight	1	3	7
	b).	Explain Visualizing amounts	1	2	7
UNIT-II					
3.	a).	How to visualize multiple distributions at the same time	2	3	7
	b).	Explain empirical and highly skewed distributions	2	2	7
OR					
4.	a).	Explain how visualizing distributions in vertical and horizontal axis	2	2	7
	b).	Explain visualization of distributions by Histograms and Density-plots	2	2	7
UNIT-III					
5.	a).	Explain the case where side-by-side Bars are suitable than pie charts	3	3	7
	b).	Explain Visualization of Nested Proportions- Mosaic Plots and Tree maps	3	2	7
OR					
6.	a).	Explain Nested Pies and parallel sets	3	2	7
	b).	Explain Visualization of Associations Among Two or More Quantitative Variables	3	2	7
UNIT-IV					
7.	a).	Explain how to framing probabilities as frequencies	4	3	7
	b).	Explain Visualizing the Uncertainty of Curve Fits	4	2	7
OR					

8.	a).	Define Detrending? Explain Time Series Decomposition	4	3	7
	b).	Explain how to visualize response curves	4	2	7
UNIT-V					
9.	a).	What is the principle of proportional ink? Explain Visualization along Linear Axes?	5	2	7
	b).	Explain visualization along Logarithmic Axes	5	2	7
OR					
10.	a).	Explain Partial Transparency and Jittering	5	2	7
	b).	How to Use 2D Histograms, Contour Lines in visualization	5	2	7

CO-COURSE OUTCOME

KL-KNOWLEDGE LEVEL

M-MARKS

NOTE: Questions can be given as A,B splits or as a single Question for 14 marks

