TEACHING AND EXAMINATION SCHEME Master of Technology (Computer Science) W.E.F. 2018 - 2019

Semester I

			.ec Tut	Exam	MAX MARKS	
	Paper Name (Theory)	Lec		Hours	Sess- ional	Sem Exam
mtc-101	Information & Network Security	5	1	3	20	80
mtc-102	High Performance Computing	5	1	3	20	80
mtc-103	Wireless Ad hoc Networks	5	1	3	20	80
mtc-104	Advance Data Mining	5	1	3	20	80

Total of Sessional & Semester Exam Marks 400

	Paper Name (Practical)	Pract Hours	Exam Hours	Max Marks
mtc-105	Lab-High Performance Computing & Data Mining	4	3	50
mtc-106	Lab-Ad hoc Networking	4	3	50
mtc-107	Lab-Information & Network Security	4	3	50
		Total of Practica	al Marks	150

Total of Theory & Practical Marks 550

TEACHING AND EXAMINATION SCHEME Master of Technology (Computer Science) Semester II

						Exam	MAX MARKS	
	Paper Name (Theory)	Lec	Tut	Hours	Sess- ional	Sem Exam		
mtc-201	Advance Storage Area Network	5	1	3	20	80		
mtc-202	Advance Java	5	1	3	20	80		
mtc-203	Advance Computer Graphics	5	1	3	20	80		
mtc-204	Machine Learning	5	1	3	20	80		

Total of Theory (Sessional + Semester Exam Marks)

400

	Paper Name (Practical)	Pract Hours	Exam Hours	Max Marks
mtc-205	Lab-Advance Java	4	3	50
mtc-206	Lab-Computer Graphics	4	3	50
mtc-207	Lab-Machine Learning	4	3	50
		Total of Pract	ical Marks	150

Total of Theory & Practical Marks 550

TEACHING AND EXAMINATION SCHEME Master of Technology (Computer Science) (2019-20)

Semester III

				Exam Hours	MAX MARKS	
	Paper Name (Theory)	Lec	Tut		Sess- ional	Sem Exam
mtc-301	Mobile Application Development	5	1	3	20	80
mtc-302	Advance Image Processing for Computer Vision	5	1	3	20	80
mtc-303	Big Data Analysis	5	1	3	20	80
mtc-304	Minor Project	5	1	3	20	80

Total of Theory (Sessional + Semester Exam Marks) 400

	Paper Name (Practical)		Pract Hours	Exam Hours	Max Marks
mtc-305	Lab-Mobile Application Devleopment		4	3	50
mtc-306	Lab-Image Processing		4	3	50
mtc-307	Lab-Big Data		4	3	50
		Total of	Practica	al Marks	150

Total of Theory & Practical Marks 550

TEACHING AND EXAMINATION SCHEME Master of Technology (Computer Science)

Semester IV

																													Evam	MAX	MARKS
Paper Name (Theory)	Lec	Lec	Lec	Tut	Hours	Sess-ional	Sem Exam																								
Dissertation	36		3	90	360																										
Seminar	6	1	3	20	80																										
	Dissertation	Dissertation 36	Dissertation 36	Dissertation363	Paper Name (Theory)LecTutExam HoursDissertation36390																										

Total of Theory (Sessional + Semester Exam Marks) 550

Scheme of Examination (For M. Tech. – Computer Science)

Theory:

Attempt 5 questions out of 10 questions set by the Examiner.

Sessional:

There will be sessional (internal assessment) of 20 marks conducted by the department.

Practical:

Practical exams shall be conducted by one internal and one external examiner of a batch of 20 students in a day.

Duration of Practical exam is 3 hours.

A Laboratory Exercise File should be prepared by each student for each practical paper and should be submitted during practical examinations.

Dissertation: 6 hours per student

Seminar: 3 hours per student

One External Examiner and one Internal Examiner will evaluate the Dissertation and Seminar. A group of three students will be assigned to one faculty for Dissertation and Seminar.

Practical of 50 marks distribution is as under:

- a. 30 marks for practical examination exercise for 3 questions
- b. 10 marks for Viva-voce
- c. 10 marks for Laboratory Exercise File

Eligibility:

M. Sc. (CS) / M. Sc. (IT), M. Sc. (Physics, M. Sc. (Math), MCA and B.E. /B. Tech. with minimum 60%

Scheme of Examination (For M. Tech. – Computer Science)

Reg. 17 (a)

The examination for the M. Tech. Computer Science will consist of 4 semesters. The examination shall consist of (a) Theory papers (b) Laboratory / Practical work and project work. Candidates will be required to pursue a regular, full time course of study at the University department for a period of two academic years in order to be eligible for appearing in the examination.

- 1. Eligibility for M. Tech. Computer Science: M.Sc (CS)/M.Sc. (IT), M. Sc. (Physics), M. Sc. (Maths), MCA and B.E. / B. Tech. with minimum 60%.
- 2. Examination:
 - i. There shall be 23 papers (4 theory, 3 practical in the I, II and III semester and in the IV Semester 1 dissertation and 1 seminar) of 2200 marks (first to fourth semester). Theory paper shall be of 100 marks out of which 20 marks shall be considered as internal assessment based on internal test and seminars and 80 marks will be of examination of 3hours duration at the end of each semester as determined by the University. The practical examination shall be of 50 marks assessed by external examiner. The minor project work will be of 80 marks, dissertation of 360 marks and seminar of 80 marks based on presentation and viva-voce assessed by external examiner.
 - ii. To pass a semester a candidate shall have to score 25% marks in each subject (theory, internal assessment, dissertation, seminar and practical) separately and also 36% marks in aggregate of all the papers prescribed for the examination.
 - iii. Due paper(s) will be applicable if a candidate obtains 36% marks in aggregate and fails in not more than two (2) papers (theory). Due paper(s) of first semester will be held along with the third semester and the due paper(s) of second semester will be held along with the fourth semester. The third and fourth semester due paper(s) will be held in the first and second semester respectively of the next year. The chance of due paper(s) will be given thrice in each semester.
 - iv. Wherever a candidate appears at for a due paper examination he/she will do so according to the syllabus in force.
 - v. A candidate not appearing at any examination/absent in any paper of term end examination shall be deemed as fail.
- 3. A candidate for a pass in the examination shall be required to obtain:
 - At least 36% marks in the aggregate of all the papers prescribed for the examination i. and
 - At least 36% marks in the practical(s), dissertation, seminar wherever prescribed at the ii. examination, provided that if a candidate fails to secure at least 25% marks in each individual paper at the examination notwithstanding his having obtained the minimum percentage of marks required in the aggregate for that examination.

No Division will be awarded in the first, second and third semester examinations. Division shall be awarded at the end of the fourth semester Examination on the combined marks obtained at the first, second third and fourth semester taken together as noted below:

Passed with First Division	60% of the aggregate marks taken together of all the
	foursemester examinations
Passed with second division	48%

Passed with second division

All the rest will be declared to have passed the examination if they secure more than 36% in aggregate.

Provided that if a candidate clears any paper after a continuous period of two years since he/she was admitted to the M. Tech. Computer Science then for the passing marks, i.e. 36% marks, shall be taken into account in the case of such course(s).

- 4. The grace marks shall be given up to 1% of the total aggregate marks of theory and practical of that semester in maximum one paper.
- 5. Candidates reappearing at an examination in a subsequent year shall be examined in accordance with the scheme and syllabi in force and shall be entitled to the award of the degree of year in which they clear the last failing/unclear paper.

Duration: 3 hours	Max Marks: 80
Duration: 3 hours	Max Marks: 80

mtc-101Information and Network Security

Classical Encryption Techniques Symmetric Cipher Model, Cryptography, Cryptanalysis and Brute-Force Attack, Substitution Techniques, Caesar Cipher, Monoalphabetic Cipher, Playfair Cipher, Hill Cipher, Polyalphabetic Cipher, One Time Pad. Block Ciphers and the data encryption standard: Traditional block Cipher structure, stream Ciphers and block Ciphers, Motivation for the feistel Cipher structure, the feistel Cipher, The data encryption standard, DES encryption, DES decryption, nature of the DES algorithm.

Public-Key Cryptography and RSA: Principles of public-key cryptosystems, applications, requirements, cryptanalysis. The RSA algorithm, Key Management and Distribution: Symmetric key, distribution scenario, Hierarchical key control, session key lifetime, a transparent key control scheme, Decentralized key control, controlling key usage, Symmetric key distribution using asymmetric encryption, simple secret key distribution, secret key distribution with confidentiality and authentication, A hybrid scheme, distribution of public keys, public announcement of public keys, publicly available directory, public key authority, public keys certificates, X-509 certificates. Certificates, X-509 version 3, public key infrastructure, User Authentication

Wireless network security: Wireless security, Wireless network threats, Wireless network measures, mobile device security, security threats, mobile device security strategy Web Security Considerations: Web Security Threats, Web Traffic Security Approaches. Secure Sockets Layer: SSL Architecture, SSL Protocols, Transport Layer Security, Electronic Mail Security, IP Security.

Duration: 3 hours	Max Marks: 80
-------------------	---------------

mtc-102High Performance Computing

Introduction to Grid Architecture Characteristic, standard bodies, Grid types, Topologies, Components and Layers, Comparison with other approaches System Infrastructure, traditional paradigms for distributed computing, web services,

Grid standards: OGSA & WSRF, introduction to Globus Toolkit 3 & Globus Toolkit 4 Semantic Grid & Autonomic Computing Metadata & Ontology in semantic web, Semantic Web Services, Layered Structure of Semantic Grid, Semantic Grid Activities, Autonomic Computing, Basic Services – Security, monitoring, review criteria overview of Grid Monitoring system – autopilot, computational grids, Data grids, architecture of Grid systems, Grid security infrastructure.

Grid Scheduling & Resource Management, scheduling paradigms, how scheduling works, review of Condor.

Introduction to Cloud Computing, Definition, Characteristics, Components, Cloud provider, SAAS, PAAS, IAAS / HAAS, and Others, Organizational scenarios of clouds, Administering & Monitoring cloud services, benefits and limitations

Virtualization&Cloud, virtualization characteristics, managing virtualization, virtualization in cloud, Virtualization desktop and managing desktops in the cloud and security issues Cloud Storage and Data Security Storage basics, Storage as a service providers, security, aspects of data security, data security mitigation, provider data and its security.

Duration: 3 hours Max Marks:

mtc-103Wireless Ad hoc Network

Ad hoc Wireless Networks: Introduction, Issues in Ad hoc Wireless Networks, Ad hoc Wireless Internet; MAC Protocols for Ad hoc Wireless Networks: Introduction, Issues in Designing a MAC Protocol, Design Goals of MAC Protocols, Classification of MAC protocols, Contention-Based Protocols, Contention-Based Protocols with Reservation Mechanisms, Contention-Based Protocols with Scheduling Mechanisms, MAC Protocols that Use Directional Antennas

Routing Protocols for Ad hoc Wireless Networks: Introduction, Issues in Designing a Routing Protocol for Ad hoc Wireless Networks; Classification of Routing Protocols; Table Driven Routing Protocols; On-Demand Routing Protocols, Hybrid Routing Protocols, Hierarchical Routing Protocols and Power-Aware Routing Protocols

Multicast Routing in Ad hoc Wireless Networks: Introduction, Issues in Designing a Multicast Routing Protocol, Operation of Multicast Routing Protocols, An Architecture Reference Model for Multicast Routing Protocols, Classifications of Multicast Routing Protocols, Tree-Based Multicast Routing Protocols and Mesh-Based Multicast Routing Protocols.

Transport Layer and Security Protocols for Ad hoc Networks: Introduction, Issues in Designing a Transport Layer Protocol; Design Goals of a Transport Layer Protocol; Classification of Transport Layer Solutions; TCP over Transport Layer Solutions; Other Transport Layer Protocols for Ad hoc Networks; Security in Ad hoc Wireless Networks, Issues and Challenges in Security Provisioning, Network Security Attacks, Key Management and Secure Touting Ad hoc Wireless Networks.

Quality of Service and Energy Management in Ad hoc Wireless Networks: Introduction, Issues and Challenges in Providing QoS in Ad hoc Wireless Networks, Classification of QoS Solutions, MAC Layer Solutions.

Duration: 3 hours Max Marks: 8

mtc-104Advance Data Mining

Data mining Overview and Advanced Pattern Mining: Data mining tasks – mining frequent patterns, associations and correlations, classification and regression for predictive analysis, cluster analysis, outlier analysis; advanced pattern mining in multi-level, multidimensional space – mining multilevel associations, mining multidimensionalassociations, mining quantitative association rules, mining rare patterns and negative patterns.

Advance Classification: Classification by back propagation, support vector machines, classification using frequent patterns, other classification methods – genetic algorithms, roughest approach, fuzzy set approach

Advance Clustering: Density – basedmethods –DBSCAN, OPTICS, DENCLUE; Grid-Based methods – STING, CLIQUE;Exception – maximization algorithm; clustering High- Dimensional Data; Clustering Graph andNetwork Data.

Web and Text Mining: Introduction, web mining, web content mining, web structure mining, we usage mining, Text mining –unstructured text, episode rule discovery for texts, hierarchy of categories, text clustering.

Temporal and Spatial Data Mining: Introduction; Temporal Data Mining – Temporal Association Rules, Sequence Mining, GSP algorithm, SPADE, SPIRIT Episode Discovery, Time Series Analysis, Spatial Mining – Spatial Mining Tasks, Spatial Clustering. Data Mining Applications

Duration: 3 hours	Max Marks: 80
Duration. 5 hours	

mtc-201Advance Storage Area Network

Introduction: Server Centric IT Architecture and its Limitations; Storage – Centric IT Architecture and its advantages. Case study: Replacing a server with Storage Networks, the Data Storage and Data Access problem; The Battle for size and access. Intelligent Disk Subsystems: Architecture of Intelligent Disk Subsystems; Hard disks and Internal I/O Channels; JBOD, Storage virtualization using RAID and different RAID levels; Caching: Acceleration of Hard Disk Access; Intelligent disk subsystems, Availability of disk subsystems.

I/O Techniques: The Physical I/O path from the CPU to the Storage System; SCSI; Fiber Channel Protocol Stack; Fiber Channel SAN; IP Storage. Network Attached Storage: The NAS Architecture, The NAS hardware Architecture, The NAS Software Architecture, Network connectivity, NAS as a storage system. File System and NAS: Local File Systems; Network file Systems and file servers; Shared Disk file systems; Comparison of fiber Channel and NAS.

Storage Virtualization: Definition of Storage virtualization; Implementation Considerations; Storage virtualization on Block or file level; Storage virtualization on various levels of the storage Network; Symmetric and Asymmetric storage virtualization in the Network.

SAN Architecture and Hardware devices: Overview, Creating a Network for storage; SAN Hardware devices; The fiber channel switch; Host Bus Adaptors; Putting the storage in SAN; Fabric operation from a Hardware perspective. Software Components of SAN: The switch's Operating system; Device Drivers; Supporting the switch's components; Configuration options for SANs.

Management of Storage Network: System Management, Requirement of management System, Support by Management System, Management Interface, Standardized Mechanisms, Property Mechanisms, Inband Management, Use of SNMP, CIM and WBEM, Storage Management Initiative Specification (SMI-S), CMIP and DMI, Optional Aspects of the Management of Storage Networks, Summary

mtc-202Advance Java

Servlet Structure, Servlet packaging, HTML building utilities, Lifecycle, Single Thread model interface, Handling Client Request: Form Data, Handling Client Request: HTTP Request Headers. Generating server Response: HTTP Status codes, Generating server Response: HTTP Response Headers, Handling Cookies, Session Tracking.

Overview of JSP Technology, Need of JSP, Benefits of JSP, Advantages of JSP, Basic syntax, invoking Java code with JSP scripting elements, creating Template Text, Invoking java code from JSP, Limiting java code in JSP, using JSP expressions, comparing servlets and JSP, writing scriptlets. Using Scriptlets to make parts of JSP conditional, using declarations, declaration example. Controlling the Structure of generated servlets: the JSP page directive, import attribute, session attribute, isElignore attribute, buffer and auto flush attributes, info attribute ,errorPage and is errorPage attributes, is Thread safe Attribute, extends attribute, language attribute, Including files and applets in JSP Pages, using java beans components in JSP documents

Java Beans & Annotations: Creating Packages, Interfaces, JAR files and Annotations. The core java API package, New Java Lang Sub package, Built-in Annotations. Working with Java Beans. Introspection, Customizers, creating java bean, manifest file, Bean Jar file, new bean, adding controls, Bean properties, Simple properties, Design Pattern events, creating bound properties, Bean Methods, Bean an Icon, Bean info class, Persistence ,Java Beans API.

JDBC: Talking to Database, Immediate Solutions, Essential JDBC program, using prepared Statement Object, Interactive SQL tool. JDBC in Action Result sets, Batch updates, Mapping, Basic JDBC data types, Advanced JDBC data types, immediate solutions.

Introduction to EJB, the Problem domain, Breakup responsibilities, CodeSmart, the Enterprise Java bean specification, components types, server side component types, session beans, message driven beans, entity beans, the Java Persistence Model, container services. Dependency injection, concurrency, instance pooling and caching, transactions, security, timers, naming and object stores, interoperability, Life Cycle callbacks, interceptors, platform integration. Definitions, naming conventions, coding the EJB, the contract, the bean Implementation class, out of Container Testing, Integration Testing.

Duration: 3 hours	Max Marks: 80
-------------------	---------------

mtc-203Advance Computer Graphics

Three-Dimensional Object Representations:Polyhedra, OpenGL Polyhedron Functions, Curved Surfaces, Quadric Surfaces, Super quadrics, OpenGL Quadric-Surface and Cubic-Surface Functions, Blobby Objects, Spline Representations, Cubic-Spline Interpolation Methods, Bezier Spline Curves, Bazier Surfaces B-Spline Curves, B-Spline Surfaces, Beta- Splines, Retional Splines, Conversion Between Spline Representations, Displaying Spline Curves and rfaces, OpenGL Approximation-Spline Functions, Sweep Representations, Constructive Solid –Geometry Method, Octrees, BSP Trees, Fractal-Geometry Methods, Shape Grammars and Others Procedural Methods, Particle Systems, Physically Based Modeling, Visualization Of Data Sets.

Visible-Surface Detection Methods:Classification Of Visible –Surface Detection Algorithms, Back-Face Method, Depth-Buffer Method, A-Buffer Method, Scan-Line Method, BSP-Tree Method, Area-Subdivision Method, Octree Methods, Ray-Casting Method, Comparison of Visibility –Detection Methods, Curved Surfaces, Wire-Frame Visibility –Detection Functions.

Illumination Models and Surface- Rendering Methods:Light Sources, Surface Lighting Effects, Basic Illumination Models, Transparent Surfaces, Atmospheric Effects, Shadows, Camera parameters, Displaying light intensities, Halftone patterns and dithering techniques, polygon rendering methods, ray-tracing methods, Radiosity lighting model, Environment mapping, Photon mapping, Adding surface details, Modeling surface details with polygons, Texture mapping, Bump mapping, OpenGL Illumination and surface-rendering functions, openGL texture functions. General computer-animation functions, Computer-animation languages, Key-frame systems, Motion specification, Articulated figure animation, Periodic motions, OpenGL animation procedures.

Color models, color applications and Computer animation:Properties of light, Color models, Standard primaries and the chromaticity diagram, The RGB color model, The YIQ and related color models, The CMY and CMYK color models, The HSV color model, The HLS color model, Color Selection and applications. Raster methods for computer animation, Design of animations sequences, Traditional animation techniques,

Hierarchical modeling and Graphics file formats:Basic modeling concepts, Modeling packages, General hierarchical modeling methods, Hierarchical modeling using openGL display list, Image-File configurations, Color-reduction methods, File-compression techniques, Composition of the major file formats.

Duration: 3 hours	Max Marks: 80
-------------------	---------------

mtc-204Machine Learning

Introduction

Definition of learning systems, Goals and applications of machine learning, designing a learning system: training data, concept representation, function approximation, well posed learning problems, perspective & issues in machine learning

Concept Learning

The concept learning task, Concept learning as search through a hypothesis space General-tospecific ordering of hypothesis. FIND-S, candidate elimination algorithm

Decision Tree Learning

Introduction, Decision tree representation, appropriate problems for decision tree learning, basic decision tree algorithm, hyperspace search in decision tree learning, issues in decision tree learning

Bayesian Learning

Probability theory and Bayes rule. Naive Bayes learning algorithm, Parameter smoothing, Generative vs. discriminative training, Logistic regression, Bayes nets and Markov nets for representing dependencies

Instance Based & Unsupervised Leaning

Introduction, K-nearest neighbour learning, case based learning, radial basis functions learning from unclassified data. Clustering, HierarchicalAgglomerative Clustering, K-means partitional clustering, Expectation maximization (EM) for soft clustering, Semi-supervised learning with EM using labeled and unlabeled data

Duration: 3 hours	Max Marks: 80
Duration: 5 hours	

mtc-301Mobile Application Development

Introduction to mobile communication and computing:, Introduction to mobile computing, Novel applications, limitations and GSM architecture, Mobile services, System architecture, Radio interface, protocols, Handover and security. Smart phone operating systems and smart phones applications

Fundamentals of Android Development: Introduction to Android., The Android 4.1 Jelly Bean SDK, Understanding the Android Software Stack, Installing the Android SDK, Creating Android Virtual Devices, Creating the First Android Project, Using the Text View Control, Using the Android Emulator, The Android Debug Bridge (ADB), Basic Widgets Understanding the Role of Android Application Components, Event Handling, Displaying Messages Through Toast, Creating and Starting an Activity, Using the Edit ext Control.

The Android Debug Bridge (ADB), basic widgets understanding the role of Android Application Components, event handling, displaying messages through toast, creating and starting an activity, using the Edit ext Control Building Blocks for Android Application Design, Laying Out Controls in Containers, utilizing resources and media

Using Selection Widgets and Debugging Displaying and Fetching Information Using Dialogs and Fragments Advanced, Android Programming: Internet, Entertainment, and Services, Implementing drawing and animations, displaying web pages and maps, communicating with SMS and emails, creating and using content providers: creating and consuming services, publishing android applications.

Duration: 3 hours	Max Marks: 80
-------------------	---------------

mtc-302Advance Image Processing

Elements of Visual Perception, A Simple Image Formation Model, Basic Concepts in Sampling and Quantization, Representing Digital Images, Spatial and Gray-level Resolution, Zooming and Shrinking Digital Images, Some Basic Relationships Between Pixels, Linear and Nonlinear Operations.

Image Enhancement in the Spatial Domain: Some Basic Gray Level Transformations, Histogram Processing, Enhancement Using Arithmetic/Logic Operations, Basics of Spatial Filtering, Smoothing Spatial Filters, Sharpening Spatial Filters, Combining Spatial Enhancement Methods.

Image Enhancement in the Frequency Domain: Introduction to the Fourier Transform and the Frequency Domain, Smoothing Frequency-Domain Filters, Sharpening FrequencyDomain, Filters, Homomorphic Filtering.

Image Restoration: A Model of the Image degradation/Restoration process, Noise Models, Restoration in the Presence of Noise Only–Spatial Filtering, Periodic Noise Reduction by Frequency Domain Filtering, Inverse Filtering ,Minimum Mean Square Error (Wiener) Filtering,Constrained Least Square Filtering, Geometric Mean Filter.

Color Fundamentals: Color Models, Pseudocolor Image Processing, Basics of Full-Color Image Processing, Color Transformations, Smoothing and Sharpening, Color Segmentation, Noise in Color Images, Color Image Compression.

Wavelets and Multi-resolution Processing: Image Pyramids, Subband coding, The Haar Transform, Multi-resolution Expansions, Wavelet Transforms in one Dimension, Fast Wavelet Transform, Wavelet Transforms in Two Dimensions, Wavelet Packets.

Image Compression: Fundamentals, Image Compression Models, Error-free (Lossless) compression, Lossy Compression.

Duration: 3 hours	Max Marks: 80
-------------------	---------------

mtc-303Managing Big Data

What is big data, why big data, data, data storage and analysis, comparison with other systems, rational database management system, grid computing, volunteer computing, convergence of key trends, unstructured data, industry examples of big data, web analytics, big data and marketing, fraud and big data, risk and big data, credit risk management, big data and algorithmic trading, big data and healthcare, big data in medicine, advertising and big data, big data technologies

Introduction to HADOOP – open source technologies, cloud and big data, mobile business intelligence, crowd sourcing analytics, inter and trans-firewall analytics.

Introduction to NoSQL, aggregate data models, aggregates, key-value and document data models, relationships, graph databases, schema less databases, materialized views, distribution models, shading, version, map reduce, partitioning and combining, composing map-reduce calculations.

Basics of HADOOP, Data format, analyzing data with HADOOP, scaling out, HADOOP streaming, HADOOP pipes design of HADOOP distributed file system (HDFS), HDFS concepts, Java interface, data flow, HADOOP I/O, data integrity, compression, serialization, Avro, file-based data structures.

MAPREDUCE applications, MapReduce workflows, unit tests with MR Unit, test data and local tests, anatomy of MapReduce job run, classic Map-reduce, YARN, failures in classic Map-reduce and YARN, job scheduling, shuffle and sort, task execution, MapReduce types, input formats, output formats