

Scheme R2012

Semester VIII

Course Code CPC801

Course Name Data Warehousing and Mining

Question No.	Answer-Key
1	C
2	C
3	A
4	A
5	A
6	C
7	A
8	D
9	A
10	A
11	A
12	A
13	C
14	D
15	C
16	B
17	D
18	D
19	B
20	B
21	B
22	A
23	C
24	B
25	A

Scheme	R2012
Semester	Data Warehousing and Mining
Course Code	CPC801
Course Name	Data Warehousing and Mining

Question No.	Question	a	b	c	d
1	Which of the following are not the desired characteristic of Data warehouse	Subject-oriented	integrated	volatile	time-variant
2	A data warehouse is _____.	updated by end users	Contains numerous naming conventions and formats	Organized around important subject areas.	Contains only current data.
3	Which of the following is a correct feature of Snowflake schema ?	Normalized data	Ease of Query processing	Intuitive Scephma	Easy browsing through the contents
4	_____ operation performs a selection on two or more dimension of the given cube, resulting in a subcube.	Dice	Slice	Roll-up	Pivot
5	Disadvantage of top-down approach is	Needs high level of cross-functional skills	Each data mart has its own narrow view of data	Permeates redundant data in every data mart	Perpetuates inconsistent and irreconcilable data
6	For a given Star Scephma if Time dimension = 5 years, Store dimension = 300 stores, and Product Dimension = 40,000. Then calculate the Maximum number of base fact table records.	1.2billion	1.5 billion	2 billion	2.5 billion
7	Deferred data extraction uses which of the given techniques ?	Extraction based on timestamp	Extraction using database triggers	Extraction using replication	Extraction based on index
8	following is task performed at Data Staging	Calculate and derive attribute values.	Provide backup and recovery for staging area repositories	Resolve inconsistencies for common data elements from multiple sources.	Create files as input to make changes to dimension tables.

9	CEO of company wants to compare product price in June with sales in July, then compare those results with the sales from another location, which might be stored in a different database. Which of the following will fetch him the desired results ?	OLAP	OLTP	Data Mining	Machine Learning
10	Which operation is the process of viewing the cross-tab (Single dimensional) with a fixed value of one attribute	Slicing	Dicing	Pivoting	Both Slicing and Dicing
11	Following is characteristic of ROLAP	Use of complex SQL to fetch data from warehouse	Creation of pre-fabricated data cubes	Moderate data volumes.	Large library of functions for complex calculations.
12	Suppose you have a variable, economic status with three categories (low, medium and high). This variable falls under which of the following types ?	Categorical	Ordinal	Binary	Continuous
13	Data Transformation by Normalization does not include	Min-max normalization	z-score normalization	chi-square test	Decimal scaling
14	A unique rule says	that there can be no missing values between the lowest and highest values for the attribute, and that all values must also be unique	that there can be no missing values	specifies the use of blanks, question marks, special characters, or other strings	that each value of the given attribute must be different from all other values for that attribute.
15	Following is not an Attribute Selection Measure	Information gain	gini index	covariance	gain ratio
16	Choose the curve analysis that is conducted on each predictor for classification?	NOC	ROC	COC	COC and NOC
17	Negative=5, True Positive=100 . Values of Sensitivity and Specifity are:	56.3 and 78.9	89 and 67	70 and 57	95.2 and 83.3
18	K-nearest neighbor is which of the following	OLAP tool	learning technique	strategy	Clustering technique
19	Problem of finding hidden structure in data that is not labelled is known as	Supervised learning	Unsupervised learning	Reinforcement learning	Smart learning
20	k-means is an _____	Re-inforcement learning	unsupervised algorithm	Supervised	unexpected learning
21	Which of the following is interestingness measures for association mining?	recall	support	accuracy	compactness
22	Select the correct statement based on Association Mining algorithms.	only twice when	carried out in FP tree	stored in a compact	generated by Apriori
23	Max_Confidence, Cosine Similarity and All_Confidence are what kind of measures?	pattern mining algorithms	efficiency of apriori	measure	Clusters
24	Each entry in the tree contains information about one cluster _____ in algorithm	DBSCAN	BIRCH	PAM	CURE
25	account amount based on the customer's attributes such as checking account	Association Rules	Chi-Squared	Dependency Analysis	Linear Regression

Semester
Course code
Course
Scheme

Q. No.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

VIII**CPC 802****Human Machine Interaction****R 2012 CBSGS**

Question	Answer-Key
_____are unintentional while_____occur through conscious deliberation.	Slips, mistakes
centered design-	interface
Getting response for a particular action can be considered as-	Feedback
something new about the domain of interest is :	Reasoning
The design of the HMI systems is generally influenced by :	the user
What is the Purpose of the Interview Process?	Users
The command line interface is used because:	flexible
The Goal Directed Design Proces is a ----- step process.	6
The mental model deals with :	the mind of the
and software applications on a single place. Which of the following will help in	Web
groupware product would help them to be more engaged would probably be	Field Study
the system.	Consistency
Which of the following is not an advantage of Graphical Systems ?	familiar
short-term memory is called :	Learning
Which is not a type of memory :	Elastic memory
short-term memory is called :	Learning
"Human eye tends to complete the missing part of an image" this defines :	Closure
What is advantage of trackball	obscure vision
Sentences and messages must be_____	Simple and Brief
Which of the following device can not be useful for a visually impaired person?	typical monitor
_____ window is used for displaying a message about a particular situation or condition	Dialog Box
Which of the following is not one of the primary colors?	Green
A _____is usually a collection of icons those are reminiscent of the purpose of the various modes.	Palette
If Password Is incorrect and System Provide facility to recollect or change Password is	Forgiveness
QR Code is_____	Access Simplifier

|

Semester **VIII**
 Course code **CPC 802**
 Course **Human Machine Interaction**
 Scheme **R 2012 CBSGS**

Q. No.	Question	A	B	C	D
1	through conscious deliberation.	Slips, mistakes	Errors, slips	errors	slips
2	considered about users in human-centered design-	interface	Experience	Knowledge	e of users
3	considered as-	Mapping	Feedback	Signifiers	e
4	conclusions or infer something new about the	Perceiving	Reasoning	solving	Learning
5	influenced by :	the user	of the system	system	taken by
6	What is the Purpose of the Interview Process?	Train The User	Understand Use	Analyse the Use	Financial
7	The command line interface is used because:	understand	DOS	UNIX system	and
8	process.	2	5	6	3
9	The mental model deals with :	the mind of the	the user	the user	style of the
10	print, audio video media and software applications	MS Word	Web	Newspaper	Icons
11	if an innovative new groupware product would help	testing	Field Study	evaluation	Study
12	response , is known as ---- of the system.	Transperancy	Consistency	Compactibility	Flexibility
13	Graphical Systems ?	familiar	Faster learning	Fewer errors	less space
14	information that is contained in short-term memory	Writing	Reading	Learning	Analysing
15	Which is not a type of memory :	memory	memory	memory	memory
16	information that is contained in short-term memory	Writing	Reading	Learning	Analysing
17	an image" this defines :	Proximity	Continuity	Similarity	Closure
18	What is advantage of trackball	Difficult to cont	Very tiring to us	requires differen	obscure
19	Sentences and messages must be _____	Simple and Brie	Unique	Passive Voice	Very Technical
20	Which of the following device can not be useful for a visually impaired person?	typical keyboard	typical monitor	typical speaker	typical procesor
21	_____ window is used for displaying a message about a particular situation or condition	Dialog Box	prompt box	message box	toolbox
22	Which of the following is not one of the primary colors?	Red	Yellow	Blue	Green
23	A _____ is usually a collection of icons those are reminiscent of the purpose of the various modes.	Button	Pointer	Palette	Toolbar
24	If Password Is incorrect and System Provide facility to recollect or change Password is	Resoonse Time	Forgiveness	User Communid	Feedback Secured
25	QR Code is _____	Web Based	Screen Based	Access Simplifi	Access

Semester
Course code
Course
Scheme

Q. No.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

VIII

CPC 802

Human Machine Interaction

R 2012 CBSGS

Question	Answer-Key
_____are unintentional while_____occur through conscious deliberation.	Slips, mistakes
centered design-	interface
Getting response for a particular action can be considered as-	Feedback
something new about the domain of interest is :	Reasoning
The design of the HMI systems is generally influenced by :	the user
What is the Purpose of the Interview Process?	Users
The command line interface is used because:	flexible
The Goal Directed Design Proces is a ----- step process.	6
The mental model deals with :	the mind of the
and software applications on a single place. Which of the following will help in	Web
groupware product would help them to be more engaged would probably be	Field Study
the system.	Consistency
Which of the following is not an advantage of Graphical Systems ?	familiar
short-term memory is called :	Learning
Which is not a type of memory :	Elastic memory
Multiple Option among available Options is which Type of Menu	Single Menu
"Human eye tends to complete the missing part of an image" this defines :	Closure
What is advantage of trackball	obscure vision
Sentences and messages must be_____	Brief
Which of the following device can not be useful for a visually impaired person?	typical monitor
	Dialog Box
Which of the following is not one of the primary colors?	Green
of the various modes.	Palette
Password is	Forgiveness
QR Code is_____	Simplifier

|

Semester **VIII**
 Course code **CPC 802**
 Course **Human Machine Interaction**
 Scheme **R 2012 CBSGS**

Q. No.	Question	A	B	C	D	Answer-Key
1	through conscious deliberation.	Slips, mistakes	Errors, slips	errors	slips	Slips, mistakes
2	considered about users in human-centered design-	interface	Experience	Knowledge	e of users	interface
3	considered as-	Mapping	Feedback	Signifiers	e	Feedback
4	conclusions or infer something new about the	Perceiving	Reasoning	solving	Learning	Reasoning
5	influenced by :	the user	of the system	system	taken by	the user
6	What is the Purpose of the Interview Process?	Train The User	Understand Use	Analyse the Use	Financial	Users
7	The command line interface is used because:	understand	DOS	UNIX system	and	flexible
8	process.	2	5	6	3	6
9	The mental model deals with :	the mind of the	the user	the user	style of the	the mind of the
10	print, audio video media and software applications	MS Word	Web	Newspaper	Icons	Web
11	if an innovative new groupware product would help	testing	Field Study	evaluation	Study	Field Study
12	response , is known as ---- of the system.	Transperancy	Consistency	Compactibility	Flexibility	Consistency
13	Graphical Systems ?	familiar	Faster learning	Fewer errors	less space	familiar
14	information that is contained in short-term memory	Writing	Reading	Learning	Analysing	Learning
15	Which is not a type of memory :	memory	memory	memory	memory	Elastic memory
16	Multiple Option among available Options is which Type of Menu	Hierarchical Menu	Cascaded Menu	Connected Menu	Single Menu	Single Menu
17	an image" this defines :	Proximity	Continuity	Similarity	Closure	Closure
18	What is advantage of trackball	Difficult to cont	Very tiring to us	requires differen	obscure	obscure vision
19	Sentences and messages must be _____	Simple and Brie	Unique	Passive Voice	Technical	Brief
20	a visually impaired person?	keyboard	typical monitor	typical speaker	procesor	typical monitor
21		Dialog Box	prompt box	message box	toolbox	Dialog Box
22	colors?	Red	Yellow	Blue	Green	Green
23	are reminiscent of the purpose of the various modes.	Button	Pointer	Palette	Toolbar	Palette
24	to recollect or change Password is	Resoonse Time	Forgiveness	User Communid	Feedback	Forgiveness
25	QR Code is _____	Web Based	Screen Based	Access Simplifi	Access	Simplifier

Scheme	R2012
Semester	VIII
Course Code	CPC803
Course Name	Parallel and Distributed Systems

Question No.	Answer-Key
1	systolic array
2	tightly coupled system
3	instruction-level parallelism
4	RAR
5	pipelining
6	structural hazard
7	multiple instructions are applied on single data
8	SIMD
9	the remaining sites can continue operating
10	scalable
11	middleware
12	inter-process
13	both transmitter and receiver has to be active at the same time
14	execute an entire process or parts of it at different sites
15	apriori knowledge of process is required.
16	dynamic load balancing
17	clock skew
18	active time server centralized algorithm
19	4
20	does not use coordinator but uses logical clock
21	ring algorithm
22	physical clocks measure the time of day and logical clocks are used to mark relationships
23	file name need not to be changed
24	hybrid model
25	read your writes

Scheme	R2012
Semester	VIII
Course Code	CPC803
Course Name	Parallel and Distributed Systems

Question No	Questions	a	b	c	d
1	Which of the following is not a type of general purpose architecture?	pipeline architecture	data flow computer	asynchronous multiprocessors	systolic array
2	Parallel computing is a _____ system	tightly coupled system	loosely coupled system	message passing	data parallelism
3	Reordering and grouping of instructions which are later on executed concurrently without influencing the result of the program. This is called _____.	instruction-level parallelism	bit-level parallelism	task parallelism	process parallelism
4	Which is not a type of data hazard?	WAW	WAR	RAR	RAW
5	When execution of multiple instructions are overlapped on time is called _____.	multitasking	multiprogramming	hardwired control	pipelining
6	If some combination of instructions cannot be accommodated because of resource conflicts, the processor is said to have a _____.	data hazard	structural hazard	pipeline hazard	stall
7	What is not true about SIMD	support some vector and possibly matrix operations in hardware	usually limit or provide less support for non-vector type operations involving data in the "vector components"	more support for traditional type operations (eg other than for vector / matrix data types)	multiple instructions are applied on single data
8	ILLIAC IV computer was an example of _____ machines	MISD	SISD	SIMD	MIMD
9	In distributed system if one site fails then what happens	the remaining sites can continue operating	all the sites will stop	directly connected sites will stop working	all process will be killed
10	The distributed systems should be _____ in a way that if the number of users or computers increases, the performance of the system should not be affected.	capable	available	manageable	scalable
11	_____ is important abstraction to construct distributed application	function	middleware	program	process
12	The _____ communication is the form of communication that makes interaction possible among multiple processes running on different machines over the network.	intra-process	inter-process	task	process
13	What is not true about Persistent Communication	message submitted for transmission is stored by communication system for as long as it takes to deliver it to receiver	not necessary for receiver to be executing at the time message submission	not necessary for sender to continue execution after submitting a message	both transmitter and receiver has to be active at the same time
14	What are the characteristics of process migration?	transfer data by entire file or immediate portion required	transfer the computation rather than the data	execute an entire process or parts of it at different sites	transfer only address space
15	What is not a desirable features of a good global scheduling algorithm	apriori knowledge of process is required.	ability to make dynamic scheduling decisions	flexible and scalable	no apriori knowledge of process is required.
16	"In which technique, processes are reassigned at the runtime depending upon the situation that is the load will be transferred from heavily loaded nodes to the lightly loaded nodes."	task assignment	static load balancing	dynamic load balancing	hybrid technique.
17	Two clocks are said to be synchronized at a particular instance of time if the difference in time values of the two clocks is less than some specified constant. The difference in time values of two clocks is called _____.	clock frequency	clock drift	clock skew	clock ticks
18	Berkeley Algorithm is _____.	passive time server centralized algorithm	active time server	distributed clock synchronization	logical clock synchronisation
19	In raymonds tree algorithm, under heavy load, how many messages are exchanged per critical section execution	n, where n is the sites	4	3	0
20	Distributed mutual exclusion algorithm _____ for event ordering.	does not use coordinator but uses logical clock	does not use physical clock but uses logical clock	does not use logical clock but uses coordinator	does not use logical clock but uses physical clock
21	Which one is coordinator selection algorithm	ring algorithm	lambport's algorithm	cristian's algorithm	berkeley algorithm

22	The difference between logical and physical clocks is	logical clocks measure the time of day and physical clocks are used to mark relationships among events in a distributed system.	both are the same	physical clocks measure the time of day and logical clocks are used to mark relationships among events in a distributed system.	physical clock is machines clock and logical clock is global clock
23	What is correct with respect to distributed file system, when a file's physical storage location changes	file name need to be changed	file name need not to	file's host name need to be changed	file's local name need to be changed
24	Which among these is a not a type of consistency models	sequential consistency	processor consistency	release consistency model.	hybrid model
25	"The effect of a write operation by a process P on a data item x will always be seen by a successive read operation on x by the same process."	monotonic reads	monotonic writes	writes follow reads	read your writes

Scheme	R2012
Semester	VIII
Course code	CPE8031
Course Name	Machine Learning

Question No.	Answer-Key
1	Filtering of spam messages
2	Because of unbalanced dataset the classifier will predict only one class of samples
3	Reinforcement Learning
4	Classification
5	0.7
6	Test specification
7	By pruning the longer rules
8	The entropy of a node typically decreases as we go down a decision tree.
9	Incrementing the number of categories
10	To transform the data from lower non linearly separable plane to higher linear separable dimension.
11	Gradient Descent with Lagrange Multiplier constraints
12	A supervised learning technique. It can be used for classification and regression
13	IF...Then rule
14	Learning problem
15	Attributes are statistically dependent of one another given the class value.
16	$P(Y X)$ decreases
17	Variance
18	Eigen values are not used in PCA
19	Eigenvalue
20	K-Means clustering
21	Agglomerative approach
22	2.23
23	Learning happens through interacting with the environment
24	Reinforcement learning.
25	Hyperplane

Scheme	R2012
Semester	VIII
Course code	CPE8031
Course Name	Machine Learning

Question No.	Question	a	b	c	d
1	Which of the following are classification problems?	Predicting the temperature (in Celsius) of a room from other environmental features (such as atmospheric pressure, humidity etc)	calculating the shorter route between two existing routes between two points.	Predicting the performance of cricket player given his playing records	Filtering of spam messages
2	A machine learning model gives 95% accuracy on an unbalanced dataset what can be correct conclusion about the classifier.	Since accuracy is 95% the classifier will perform well in real life scenario	Classifier will give good accuracy on validation dataset.	Unbalanced Dataset will not affect the performance of classifier	Because of unbalanced dataset the classifier will predict only one class of samples accurately
3	What does the diagram depicts (Link of the diagram: https://drive.google.com/file/d/1BXEb1KIZbhVEP1Osj-aykLF0a6QTYP5X/view?pli=1)	Supervised Learning	Semi Supervised Learning	Unsupervised Learning	Reinforcement Learning
4	Suppose you want to predict whether it will be raining at 5pm tomorrow or not. Which machine learning technique is suitable for this problem.	Regression	Classification	Optimization	Clustering
5	A classification model is built and confusion matrix is obtained with following values as True Positive=3, False Negative=1, False Positive =2 and True Negative=4. What will be the classification accuracy.	0.5	0.7	0.6	0.4
6	What does decision node illustrate in a decision tree?	Data process description	Test specification	Class instances	Data value description
7	How will you counter over-fitting problem in decision tree	By pruning the longer rules	By creating new rules	Cost Complexity Pruning	Time Complexity Pruning
8	Which of the following is true for a decision tree?	Decision tree is an example of linear classifier	The entropy of a node typically decreases as we go down a decision tree.	Entropy is a measure of purity	An attribute with lower mutual information should be preferred than other attributes
9	Handling missing or corrupted data in a dataset does not deal with_____.	Drop missing rows or columns	Replace missing values with mean/median/mode	Assign a unique category to missing values	Incrementing the number of categories

10	What is kernel trick in Support Vector Machine?	To transform the data from lower non linearly seperable plane to higher linear seperable dimension.	To transform the data from linearly seperable plane to higher non linear seperable dimension.	To convert regression problem to classification problem	To convert classification problem to regression problem
11	An appropriate learning algorithm for the Support Vector Machine is	Quadratic programming of soft margins	Quadratic programming via Gradient Descent	Gradient Descent with Lagrange Multiplier constraints	Quadratic programming via Sequential Minimal Optimization
12	Support Vector Machine can be classified as_____.	A unsupervised learning algorithm. It can be used for classification and regression.	A unsupervised learning technique . It can be used for classification but not for regression.	A supervised learning technique. It can be used for classification and regression.	learning technique. It can be used for classification but not for regression.
13	The rule based classifier generates rules for classification.	IF rule	IF..Then rule.	Do while	while
14	What is a fundamental problem in Hidden Markov Model.	Clustering problem.	Learning problem	Regression problem	dimension reduction problem
15	Which of the following statements about Naive Bayes is incorrect?	Attributes are equally important.	Attributes are statistically dependent of one another given the class value.	Attributes are statistically independent of one another given the class value.	Attributes can be nominal or numeric
16	X and Y are two events. If P(X, Y) decreases while P(X) increases, which of the following is true with respect to naive bayes classifier?	P(X Y) decreases	P(Y X) decreases	P(Y) decreases	P(Y X) increases
17	Which of the following is a measure of the spread of data?	Mean	Variance	Covariance	Summation
18	What is not true about Principal component analysis(PCA)?	First principal component has the largest possible variance	It orthogonally transforms the original 'n' coordinates of a data set	Applying PCA to your data set will result in loss of interpretability	Eigen values are not used in PCA
19	In Principle Component Analysis the _____ represents the total variance explained by each factor.	Eigenvalue	Correlation matrix	Bartlett's test of sphericity	Residuals
20	Which clustering algorithm terminates when current iteration mean values are identical to that of previous iteration?	Agglomerative clustering	Conceptual clustering	K-Means clustering	Expectation maximization
21	Which of the following is not the spatial clustering algorithms?	Partitioning based clustering	K-means clustering	Grid based clustering	Agglomerative approach
22	Which of the following will be Euclidean Distance between the two data point A(2,3) and B(4,4)?	2.23	2	4.5	8.2
23	How reinforcement learning is done?	Learning happens in the persence of supervision.	Learning happens through interacting with the environment	Learning happens by giving sample data to model/agent	Agent learns by itself without any feedback
24	In which of the following machine learning technique , Q learning algorithm is used?	Classification	Regression	Clustering	Reinforcement learning.
25	Which of the following is not related to Reinforcement Learning?	Envionment	Hyperplane	State	Action

Scheme	8
Semester	CPE8034
Course Code	Digital Forensics
Course Name	R2012
Question No.	Answer-Key
1	Spam
2	scan device, identify and eliminate
3	Report vulnerability in any system
4	Cyber crime category - Against person
5	Preparation, identification, containment, eradication, Recovery, lesson learned
6	Data in memory
7	The identification, preservation, recovery, restoration and presentation of digital evidence from systems and devices
8	Manipulation of computer data
9	Qualified Forensic duplicate
10	Chain of custody
11	ROM
12	Best evidence rule
13	System and network information
14	Confirming system date and time
15	snort
16	Click on unknown links to explore
17	Email Header
18	Zone transfer
19	Covert channels
20	HKEY_LOCAL_MACHINE
21	Deleted inodes are not accessible to the file system
22	dtSearch
23	The expense involved in creating the original material
24	Computer Fraud and Abuse Act (CFAA)
25	Computer Fraud and Abuse Act

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

|

Scheme	8
Semester	CPE8034
Course Code	Digital Forensics
Course Name	R2012

Question No.	Question	Option1	Option2	Option3	Option4
1	All of the following are examples of real security and privacy risks except	Hackers	Spam	Identity Theft	virus
2	Choose the best option : What is the purpose of antivirus program ?	identify and eliminate	scan device and identify	scan device, identify and eliminate	scan device and eliminate
3	Which of the following is not done by cyber criminals?	Unauthorized account access	Mass attack using Trojans as botnets	Email spoofing and spamming	Report vulnerability in any system
4	What are the examples of Harasment emails and cyber stalking ?	Cyber crime category - Against person	Cyber crime category - Against property	Cyber crime category - Against Government	Cyber crime category - Against computer
5	Choose one option for the incident handling cycle steps?	Preparation, identification, containment, eradication, Recovery, lesson learned	identification, Preparation, containment, eradication, Recovery, lesson learned	lesson learned, identification, Preparation, containment, eradication, Recovery.	identification, containment, Preparation,eradication, recovery, lesson learned
6	Which one of the following electronic evidence types has the highest degree of volatility?	Data in memory	Data stored in the cloud	Data stored on hard disks	Data stored on a flash drive
7	Digital Forensics entails to what ?	Accessing the system's directories viewing mode and navigating through the various systems files and folders	Undeleting and recovering lost files	Identifying and solving computer crimes	The identification, preservation, recovery, restoration and presentation of digital evidence from systems and devices
8	Choose one activity not involved in the Computer Forensics.	Preservation of computer data	Interpretation of computer data	Manipulation of computer data	Extraction of computer data
9	A file that contains every bit of information from the source but stored in altered format is	Qualified Forensic duplicate	Replica	Duplicate file	cloned file
10	The process of documenting the seizure of digital evidence and, in particular, when that evidence changes hands, is known as:	Chain of custody	Field notes	Interim report	Report book
11	The non volatile memory of the computer is known as	RAM	Monitor	ROM	CPU
12	The fact that with modern technology, a photocopy of a document has become acceptable in place of the original is known as:	Best evidence rule	Due diligence	Acceptance rule	Replication
13	What type of data that should be immediately preserved ?	USB drives	Digital picture frames	System and network information	USB bracelets
14	What is not included in the crime scene preservation process?	Protecting against unauthorized alterations	Acquiring digital evidence	Confirming system date and time	Controlling access to the crime scene
15	Which of the following deals with network intrusion detection and real-time traffic analysis?	John the Ripper	L0phtCrack	Snort	Nessus
16	Which of them is not a proper method for email security?	Click on unknown links to explore	Use email Encryption	Spam filters and malware scanners	Use Strong password
17	During email forensic Ip address of the machine from which email is sent is found out from	Email header	Email ID	Email Browser	Email Inbox
18	It is sometimes possible to obtain a list of all machines in the DNS belonging to a specific	Web crawl	Zone transfer	Reverse IP	IP transfer
19	What is an indirect way to transmit information with no explicit reading of confidential	Covert channels	Backdoor	Timing channels	Overt channels
20	Which key in Windows registry can reveal the software installed in the past	HKEY_CLASSES_ROOT	HKEY_CURRENT_USER	HKEY_LOCAL_MACHINE	HKEY_CURRENT_CONFIG
21	Deleting a file has the effect of preserving its inode until it is reused because	The inode is flagged as	The inode table entry is	Deleted inodes are not	The inode number is
22	Which of the following is the disk-search utility ,to perform search from a physical level	PsLogList	Dumpel.exe	dtSearch	netstat
23	Which of the following is not a consideration in determining "fair use" of copyrighted	The purpose and	The amount and	The expense involved in	Effect on the potential
24	A "protected computer" under this law is any computer that is connected to a government	Digital Millennium	Computer Fraud and	HIPAA	CAN-SPAM
25	Whoever intentionally accesses a computer without authorization or exceeds authorized	Computer Fraud and	Computer Fixing and	Customer Fraud and	Consumer Fraud and

Scheme	R2012
Semester	VIII
Course Code	CPE8035
Course Name	Big Data Analytics

Question No.	Question	a	b	c	d
1	Which one of the following is not a Hadoop limitation	Parallel Processing	High Availability	Multiple DataCenters	Security
2	Which of the following are among the duties of the Data Nodes in HDFS?	Manage the file system namespace.	Store and retrieve blocks	Maintain the file system tree and metadata for all files and directories.	Data processing layer of hadoop
3	Which of the following maps input key/value pairs to set of intermediate key/value pairs.	Reducer	Mapper	Both Mapper and Reducer	Combiner
4	The minimum (default) amount of block size, that HDFS can read or write in Hadoop 2.x and Hadoop 1.x respectively is	128 MB, 64 MB	64 MB, 128 MB	128 MB, 256 MB	256MB, 128 MB
5	Using Jacard similarity identify how similar these two sets are? A = {0,1,2,5,6} B = {0,2,3,4,5,7,9}	0.33	3	0	0.5
6	Which of the following is a Graph store ?	DynamoDB	cassandra	mongoDB	neo4j
7	Which of the following are the simplest NoSQL databases ?	Key-value	column family	document store	Graph store
8	Job Tracker is there in	Name Node	Laptop	Data Node	Computer
9	Which of the following for Stream data is true	They need not have the same data rates or data types, and the time between elements of one stream need not be uniform.	The time between elements of one stream must be uniform	They are all of same data types	They arrive at same data rate
10	Which of the following is similar to Euclidean distance?	Manhattan distance	Pythagoras metric	Chebyshev distance	Heuristic distance
11	Yahoo! Accepts billions of "clicks" per day on its various sites. This is an example of	Batch data	Stream data	Archival data	File data
12	Which node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker.	Mapreduce	Mapper	Task Tracker	Job Tracker
13	The source of HDFS architecture in Hadoop originated from which of the following?	Facebook distributed file system	Yahoo distributed file system	Google distributed file system	Amazon distributed file system
14	Which of the following does not accurately describe Hadoop:	Open source	Real-time	Java-based	Distributed computing approach
15	_____ can be used to describe nodes that contain the most amount of information about a network.	Social Networks	Betweenness Centrality	Degree Centrality	Broadcasters
16	_____ systems recommend items based on similarity measures between users and/or items.	Content-based filtering	general filtering	Collaborative Filtering	user-based filtering
17	Hadoop was first conceived to fix a scalability issue that existed in _____ crawler and search engine.	Bingbot	nutch	Yandex Bot.	Exabot
18	In CURE Algorithm, CURE stands for	Clustering Using Representatives	Clustering Using Repetation	Clustering Using Representations	Clustering Used for Representations

19	PageRank is a function	that assigns a number that is number of outgoing links on a page.	that assigns a number that is number of distinct words on a page.	that assigns a real number to each page in the web, based on its importance.	that assigns a number that is number of incoming links on a page.
20	Which of the following is not a high-level drivers associated with NoSQL movement	Agility	Velocity	Volume	Varsity
21	In implementation of Union of two relations using MapReduce the Map function	Turn each input tuple t into a key-value pair (t,t)	Turn each input tuple t into a key-value pair (t,1)	Turn each input tuple t into a key-value pair (t,'A') where 'A' is the name of Relation of t	Turn each input tuple t into a key-value pair (t,0)
22	Suggest appropriate type of distance measure for plagiarism detection?	Edit Distance	Cosine distance	Jaccard Distance	Hamming Distance
23	You run a movie empire, and want to build movie recommendation system based on Collaborative filtering. There are 3 popular review websites (X, Y and Z) which users to go to rate movies, and you have just acquired all three companies that run these websites. You like to merge the three companies' datasets together to build a single/unified system. On website X, users rank a movie as having 1 through 10 stars. On Website Y, users rank on a scale of 1 to 5. On Website Z, the ratings are from 1 to 50. You also have enough information to identify users/movie on one website with users/movies on a different website. Which of the following statement is true.	You can combine all three training sets into one without any modifications and expect high performance from a recommendation system.	You can merge three datasets into one but you should first normalize each dataset's ratings (say rescale each dataset's ratings to a 0-1 range)	You can combine all three training sets into one as long as you perform mean normalization and feature scaling after you merge the data.	It is not possible to combine these websites' data. You must build three separate recommendation systems.
24	For Counting Ones in a window which of the following algorithm is used	The Datar-Gionis-Indyk-Motwani	The Flajolet-Martin Algorithm	Bloom's Algorithm	The CURE Algorithm
25	SON Algorithm uses _____ passes	One	Two	Three	Four