

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
Q	01) Big Data uses which approach		L	1
A	Centralized		0	1
A	Distributed		1	2
A	Peer to Peer		0	3
A	Client Server		0	4
Q	02) Use cases for Marketing and Sales Applications are			1
A	No detection of fraud and compliances		0	1
A	Operational analytics for company operations optimization		1	2
A	Only helps to make organization customer-oriented		0	3
A	Informs customer about variety of products and services		0	4
Q	03) Grid computing refers to:			1
A	Distributed computing, in which a group of computers from several locations are connected with each other to achieve a common task.		1	1
A	Remotely connected computers using Internet		0	2
A	Computing using cloud		0	3
A	Computations and no under-performance even on failure of any of the participating nodes.		0	4
Q	04) Data architecture design considers:		M	1
A	Four design layers with the lowest being identification of internal and external data sources		0	1
A	Four design layers with the lowest being acquisition and next identification of internal and external data sources		0	2
A	Five layers with data consumption for analytics, business processes, business intelligence, data mining, pattern recognition and knowledge discovery.		1	3
A	Five layers with data storage processing and analytics being the highest layer.		0	4
Q	05) Automotive Maintenance Service Center Application in a company can			1
A	Use analytics followed by predictive analytics		1	1
A	Visualization with finer and coarse granulated and multi-dimensional data cubes for maintenance needs		0	2
A	Predict declining of rising sales		0	3
A	Plan strategies for boosting sales		0	4
Q	06) Cloud computing environment			1
A	Has restricted network access		0	1
A	Has no accountability		0	2
A	Has no abstraction and productivity		0	3
A	Performs parallel and distributed computing for processing and analyzing large datasets on computing nodes.		1	4
Q	07) Big Data also refer to		L	1
A	Large Dataset		0	1
A	Massive Dataset		1	2
A	Huge Dataset		0	3
A	Enormous Dataset		0	4
Q	08) _____ NameNode is used when the Primary NameNode goes down		L	1
A	Primary		0	1
A	Secondary		1	2

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	Master		0	3
A	Slave		0	4
Q	09)A _____ serves as the master and there is only one NameNode per cluster		L	1
A	DataNode		0	1
A	Namenode		1	2
A	Data Block		0	3
A	Rack		0	4
Q	10)_____ can change the maximum number of cells of a column family		L	1
A	set		0	1
A	reset		0	2
A	alter		1	3
A	select		0	4
Q	11)NoSQL databases is used mainly for handling large volumes of		L	1
A	Unstructured data		1	1
A	Structured Data		0	2
A	Semi-structured Data		0	3
A	Processed data		0	4
Q	12)Which of the following is a column-oriented database that runs on top of HDFS		M	1
A	Hive		0	1
A	Sqoop		0	2
A	Hbase		1	3
A	Flume		0	4
Q	13)Which of the following is correct statement?		M	1
A	Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data		1	1
A	Hive is a relational database with SQL support		0	2
A	Pig is a relational database with SQL support		0	3
A	Hive is a relational database having no SQL support		0	4
Q	14)Which of the following is a single point of failure for an HDFS cluster?		L	1
A	Data Node		0	1
A	NameNode		1	2
A	ActionNode		0	3
A	BlockNode		0	4
Q	15)For 132 MB file how many Blocks will be created ?		M	1
A	1		0	1
A	2		1	2
A	3		0	3
A	4		0	4
Q	16)What is the major advantages of storing data in block size 128MB		H	1
A	It saves disk seek time		1	1
A	It saves disk processing time		0	2
A	It saves disk processing time		0	3
A	It saves disk latency time		0	4
Q	17)Who provides File system and OS level abstraction in Hadoop ?		H	1
A	HDFS		0	1

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	MapReduce		0	2
A	Hadoop Common		1	3
A	YARN		0	4
Q	18)ZooKeeper provides which service?		M	1
A	Communication		0	1
A	Transmission		0	2
A	Coordination		1	3
A	Collaboration		0	4
Q	19)Which of the following is the major disadvantage of Key Value store type of NOSQL database?		M	1
A	There is no way to search the content based on value		1	1
A	There is no way of partitioning		0	2
A	Complex API		0	3
A	Keys are used to access and search the data		0	4
Q	20)Which of the following is graph based store?		L	1
A	Redis		0	1
A	Neo4j		1	2
A	CouchDB		0	3
A	Cassandra		0	4
Q	21)Which of the following statement is true?		H	1
A	Strong consistency makes avenues for higher performance and availability		0	1
A	Schema less architecture makes avenues for strong consistency		0	2
A	Relaxed consistency makes avenues for higher performance and availability		1	3
A	Agility makes avenues for strong consistency		0	4
Q	22)Which of the following is column based store?		H	1
A	Redis		0	1
A	Neo4j		0	2
A	CouchDB		0	3
A	Cassandra		1	4
Q	23)Which of the following metaphor map to Key Value Store?		H	1
A	Dustbin		0	1
A	Table		0	2
A	Locker		1	3
A	Cupboard		0	4
Q	24)Which scenario demands highest bandwidth for data transfer between nodes		H	1
A	Different nodes on the same rack		0	1
A	Nodes on the different racks on the same data center		0	2
A	Nodes in the different data centers		1	3
A	Data on the same node		0	4
Q	25) Which one of the following parts of the MapReduce is responsible for processing one or more chunks of data and producing the output results?		M	1
A	Maptask		1	1
A	Task execution		0	2
A	Mapper		0	3
A	Reducer		0	4

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
Q	26) Which of the following is the default InputFormat, which treats each value of input a new value and the associated key is byte offset?		M	1
A	FileInputFormat		0	1
A	KeyValueTextInputFormat		0	2
A	TextInputFormat		1	3
A	FileOutputFormat		0	4
Q	27) Which is the correct statement among the following statements		M	1
A	Mapper and Reducer interfaces form the core of the job		0	1
A	Mapper maps input key/value pairs to a set of intermediate key/value pairs		0	2
A	Applications typically implement the Mapper and Reducer interfaces to provide the map and reduce methods		0	3
A	MapReduce tries to place the data and the compute as close as possible		1	4
Q	28) Reason why MapReduce is not used for Amazon		M	1
A	Not Recording sales		0	1
A	Processes that involve relatively little calculation and that change the database		1	2
A	Does not responding to searches for products and services		0	3
A	Not able to perform analytic queries on large amounts of data		0	4
Q	29) Choose the incorrect statements among the following		M	1
A	The NameNode coordinates and monitors the data storage function (HDFS), while the JobTracker coordinates the parallel processing of data using MapReduce.		0	1
A	The DataNode is a slave to the NameNode.		0	2
A	SlaveNode does the actual work of storing the data and running the computations.		0	3
A	Each Master runs both a DataNode and a TaskTracker daemon that communicate with their respective Slave Nodes.		1	4
Q	30) Which of the following is responsible to cope up with the failure node in MapReduce?		M	1
A	NameNode		0	1
A	Mapper		0	2
A	TaskTracker and JobTracker		1	3
A	Reducer		0	4
Q	31) For which relational algebra operation using MapReduce, the relation R (A, B) with S (B, C), it is required to find tuples that agree on their B components, that is, the second component from tuples of R and the first component of tuples of S.		M	1
A	Aggregation		0	1
A	Natural join		1	2
A	Grouping		0	3
A	Intersection		0	4

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
Q	32) In MapReduce model to handle the upcoming data in a parallel and distributed form, the data has to flow from following various phases		M	1
A	Input Reader→ Map Function→ Shuffling and Sorting→ Reduce Function→ Output Writer		1	1
A	Map Function→ Input Reader→ Shuffling and Sorting→ Reduce Function→ Output Writer		0	2
A	Map Function→ Input Reader→ Reduce Function→ Shuffling and Sorting→ Output Writer		0	3
A	Input Reader→ Map Function→ Reduce Function→ Shuffling and Sorting→ Output Writer		0	4
Q	33) Which configuration file is used to control the HDFS replication factor?		M	1
A	mapred-site.xml		0	1
A	core-site.xml		0	2
A	yarn-site.xml		0	3
A	hdfs-site.xml		1	4
Q	34) Point out the correct statement about Key-Value pair.		M	1
A	Value class must extend WritableComparable		0	1
A	Value class must implement WritableComparable		0	2
A	Key class must implement WritableComparable		1	3
A	Key class must implement Writable		0	4
Q	35) In Map Reduce 3.0, the actual work of storing the data and running the computation is done by		H	1
A	Data Node		1	1
A	Task Tracker		0	2
A	Resource Node		0	3
A	Application Node		0	4
Q	36) In Map Reduce 2.0, the actual work of storing the data and running the computation is done by		H	1
A	Name Node		0	1
A	Task Tracker		1	2
A	Resource Node		0	3
A	Application Node		0	4
Q	37) In Map Reduce 3.0, the work of coordination and monitor is done by		H	1
A	Name Node		1	1
A	Task Tracker		0	2
A	Resource Node		0	3
A	Application Node		0	4
Q	38) In Map Reduce 2.0, the work of coordination and monitor is done by		H	1
A	Name Node		0	1
A	Job Tracker		1	2
A	Resource Node		0	3
A	Application Node		0	4
Q	39) The mapper and reducer classes extend classes from package		H	
A	org.apache.hadoop.core.mapreduce		0	
A	config.apache.hadoop.mapreduce		0	
A	org.apache.hadoop.mapreduce		1	
A	org.apache.hadoop.common.mapreduce		0	

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
Q	40) In MapReduce, OutputFormat can be from?		H	
A	mapper		0	
A	Reducer		0	
A	mapper or reducer		1	
A	combiner		0	
Q	41) Map process runs computation on their		H	
A	glocal block of data		0	
A	local block of data		1	
A	external block of data		0	
A	extended block of data		0	
Q	42) In Map Reduce traffic condition is referred as		H	
A	Fan-in		0	
A	Fan-out		0	
A	Out-Cast		0	
A	Fan-in or Incast		1	
Q	43) In MapReduce, initializing and controlling of Map Reduce happens at		H	
A	Mapper		0	
A	Reducer		0	
A	Driver		1	
A	Partitioner		0	
Q	44) If reduce node fails , the Job tracker set it to		M	
A	terminate state		0	
A	idle state		1	
A	waiting state		0	
A	blocked state		0	
Q	45)Which one of the following is not for F-M Algorithm		M	1
A	Classification Algorithm		1	1
A	Stream Processing Algorithm		0	2
A	Counting Distinct element		0	3
A	Finds Unique Element		0	4
Q	46)How Stream Data can be Processed in real world		M	1
A	Using SQL Data		0	1
A	Using Technology Like Apache Spark		1	2
A	Using Ansi SQL		0	3
A	Using RDBMS		0	4
Q	47)Let us consider stream as 5 ,9, 1, 6, 3. Hash functions as $h(x) = 3x + 7 \text{ mod } 32$.What is the value of 2R using Flajolet-Martin algorithm.		M	1
A	4		0	1
A	8		1	2
A	16		0	3
A	32		0	4
Q	48)Which of the following statements about data streaming is true?		M	1
A	Stream data is always unstructured data		0	1
A	Stream data often has a high velocity		1	2
A	Stream elements cannot be stored on disk		0	3
A	Stream data is always structured data		0	4
Q	49)If you want to count number of unique data element from very large data set which algorithm will be used		M	1
A	Sampling Algorithm		0	1

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	Page rank		0	2
A	Clustering		0	3
A	Flajolet-Martin (F-M) Algorithm		1	4
Q	50)The Stream data refer as which V from Big Data		M	1
A	Volume		0	1
A	Velocity		1	2
A	Variety		0	3
A	Varacity		0	4
Q	51)Consider the scenario of 3 star Hotel in the metro city, Management wants to store all the feedback of the visited users, which of the approach is feasible in this case		M	1
A	Big Data Processing Using Hadoop.		0	1
A	Data Stream Processing.		0	2
A	Traditional RDBMS Processing		1	3
A	NoSql Data Processing		0	4
Q	52)How Stream queries are differ than traditional data queries		M	1
A	The traditional data are dynamic in nature so query fetch is not correct, whereas the stream query processing is straight forward		0	1
A	The traditional data are mostly static in nature so query fetch is straight forward, whereas the stream query processing need to process standing query		1	2
A	The Traditional data need to find exact location of Query, whereas in Streaming it's not necessary		0	3
A	The Traditional query processing mechanism is not capable to handle small data processing, whereas Stream processing model handle query using SQL Format.		0	4
Q	53)Typically in DGIM Algorithms count numbers of 1's from		M	1
A	Left to right in window		0	1
A	Right to left in window		1	2
A	From Bottom to Top of Window		0	3
A	From Top to Bottom of Window		0	4
Q	54)To Implement the F-M Algorithm which of the sequence should follow		M	1
A	Input Stream , Apply Hash Function, Remainder From Hash, Binary Conversion of Remainder, Calculating Trailing zeros from binary, apply 2^k formula where k is trailing zeros.		1	1
A	Input Stream , Apply Hash Function, Remainder From Hash, Octal Conversion of Remainder, Calculating Non Trailing zeros from binary, apply 2^k formula where k is Non trailing zeros.		0	2
A	Input Stream , Apply Hash Function, Remainder From Hash, Binary Conversion of Remainder, Calculating Trailing ones binary, apply $2k$ formula where k is trailing ones.		0	3

Q=QUESTION	question_description	question_expl anation	question_type	question_difficulty
A=ANSWER	answer_description	answer_expla nation	answer_isright	answer_position
A	Input Stream , Apply Hash Function, Remainder From Hash, Octal Conversion of Remainder, Calculating Non Trailing ones from binary, apply 2^k formula where k is Non trailing ones.		0	4
Q	55)In PLANET algorithm, mapper does not keeps track of		M	1
A	Set of Split S		0	1
A	Subset of dataset D*		0	2
A	each leaf node and its subtree		0	3
A	Number of iterations		1	4
Q	56)PLANET algorithm runs the following mapper n number of times		M	1
A	MR Initialization		0	1
A	MR inMemoryBuild		0	2
A	MR FindBestSplit		1	3
A	MR CleanUp		0	4
Q	57)In presence of noisy data in K Nearest Neighbor, you would		M	1
A	Increase the value of K		1	1
A	Decrease the value of K		0	2
A	Noisy data is not affected by K		0	3
A	Opt for 1 Nearest Neighbor classifier		0	4
Q	58)Which algorithm would give faster results for larger datasets		M	1
A	SVM		0	1
A	Fast-SVM		0	2
A	SG-SVM		1	3
A	1NN		0	4
Q	59)Which of the following is true for SON algorithm		M	1
A	SON algorithm gives no false negative		1	1
A	In SON algorithm the dataset is not divided into chunks		0	2
A	SON algorithm gives no false positive		0	3
A	SON algorithm does not guarantee that all frequent itemsets are frequent in at least one chunk		0	4
Q	60)Confidence of an Association rule A->B is _____		M	1
A	$Conf(A \rightarrow B) = Support(A \cup B) / Support(B)$		0	1
A	$Conf(A \rightarrow B) = Support(A \cap B) / Support(B)$		0	2
A	$Conf(A \rightarrow B) = Support(A \cap B) / Support(A)$		0	3
A	$Conf(A \rightarrow B) = Support(A \cup B) / Support(A)$		1	4
Q	61)In CURE algorithm, representing clusters are by.....		M	1
A	Representative points		1	1
A	mean		0	2
A	Median		0	3
A	Centroid		0	4
Q	62)One of the following is NOT a goal of an SVM is		M	1
A	maximizes the Slack penalty C		1	1
A	To select a hyperplane that maximizes the distance between the hyperplane and any point in training set.		0	2
A	maximize margin (γ) that is multiple of unit vector ($w/ w $) between the separating hyperplane and the upper and lower hyperplane.		0	3

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	Minimize $\ w\ $ subject to constraint that for all $i = 1, 2, \dots, n$, $y_i(w \cdot x_i + b) \geq 1$		0	4
Q	63)For Large data sets, _____ does provides an advantageous solution for SVM		M	1
A	Linear Solver		0	1
A	Quadratic Solver		0	2
A	Batch Gradient Descent		0	3
A	Stochastic Gradient Descent		1	4
Q	64)Slack variable		M	1
A	Is the variable that tells us about the class information		0	1
A	Uses Margin in order to penalize error/misclassification		1	2
A	Is the extra time required for making classification		0	3
A	Is the extra time required for training the model		0	4
Q	65)To implement Nearest Neighbor we would NOT compulsorily require the following point		M	1
A	distance metric		0	1
A	weighting function		1	2
A	number of neighbors to look at		0	3
A	fitting with local points		0	4
Q	66)What is the relation between candidate and frequent itemsets?		M	1
A	A candidate itemset is always a frequent itemset		0	1
A	A frequent itemset must be a candidate itemset		1	2
A	No relation between the two		0	3
A	Both are same		0	4
Q	67)Methods to store counts of frequent pairs in main memory are		M	1
A	Triangular matrix		0	1
A	Frequent triples		0	2
A	Triangular matrix and frequent triples		1	3
A	Neither Triangular matrix and frequent triples		0	4
Q	68)Which method of storing frequent pair counts occupies more memory		M	1
A	Triangular matrix		0	1
A	Frequent triples		0	2
A	Both take same amount of memory		0	3
A	Depends on the number of frequent triples		1	4
Q	69)A triangular matrix is stored as a		M	1
A	2 dimensional array		0	1
A	1 dimensional array		1	2
A	A list of pairs		0	3
A	A heap		0	4
Q	70)The hash bucket in pass 1 of the PCY algorithm		M	1
A	stores the frequent pairs and their counts of those that hash to it		0	1
A	only the frequent pairs that hash to it		0	2
A	only counts of the frequent pairs that hash to it		1	3
A	Does not store either the frequent pairs or their counts		0	4
Q	71)The pairs (i, j) that are generated in pass 2 of the PCY are those		M	1
A	where one of i or j hash to a frequent bucket		0	1
A	where both i and j hash to a frequent bucket		1	2

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	where neither i or j hash to a frequent bucket		0	3
A	does not depend on the frequent bucket		0	4
Q	72)In the second pass of the PCY algorithm the hash table of frequent pairs is stored in the form of		M	1
A	Hash table		0	1
A	a list of frequent pairs		0	2
A	Bit Map		1	3
A	Triangular array		0	4
Q	73)The Multistage algorithm uses		M	1
A	multiple hash tables in one pass		0	1
A	multiple hash tables one in each pass		1	2
A	only one hash table		0	3
A	No hash table		0	4
Q	74)In the triangular matrix representation the count for pair {i, j} is found in position :		M	1
A	$(i + 1)(n - i / 2) + j - i$		0	1
A	$(i - 1)(n - i / 2) + j - i$		1	2
A	$(i - 1)(n + i / 2) + j - i$		0	3
A	$(i - 1)(n - i / 2) + j + i$		0	4
Q	75)The SON algorithm for finding frequent itemsets uses the concept of		M	1
A	Hashing		0	1
A	Pruning		0	2
A	Partitioning		1	3
A	Sorting		0	4
Q	76)Curse of dimensionality in Clustering occurs due to		M	1
A	very few data points		0	1
A	very large number of data points		0	2
A	Very large number of features per data point		1	3
A	Very few features per data point		0	4
Q	77)Which of the following is not an issue with Clustering data with high dimensionality		M	1
A	The dataset becomes exponentially very large		0	1
A	Normal clustering algorithms like K Means will not be sufficient		0	2
A	In high dimensions distances between points become uniform so difficult to cluster		0	3
A	Data collection is difficult		1	4
Q	78)Which of the following is not a Clustering technique		M	1
A	Partition based		0	1
A	Hierarchical based		0	2
A	Density based		0	3
A	temporal based		1	4
Q	79)In the CURE initialization phase the representative points of clusters are picked such that		M	1
A	Random points are picked		0	1
A	the centroid of the cluster		0	2
A	they are very close to each other		0	3
A	they are far away from each other		1	4
Q	80)The representative points picked in the CURE initialization phase		M	1
A	are directly stored for the next phase		0	1

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	are discarded in this phase		0	2
A	are combined together to form one point		0	3
A	they are shrunk towards the centre of the cluster		1	4
Q	81)In the CURE completion phase		M	1
A	All clusters formed are output without further merging		0	1
A	If representative points from 2 clusters are close to each other they are merged		1	2
A	All outlier points are discarded		0	3
A	All other points in the secondary memory are not considered		0	4
Q	82)Canopy Clustering is usually used as		M	1
A	Pre Clustering techniques for very large datasets		1	1
A	Clustering techniques for high dimensional datasets		0	2
A	Clustering techniques for very small datasets		0	3
A	post clustering techniques for Large datasets		0	4
Q	83)What is NOT applicable about online learning		M	1
A	It is the a way in which we our model learns from continuous stream of incoming data and slowly adapts to the changes in data		0	1
A	The trained model receives data that has few aspects changing over time and it learns the slow shift in data		0	2
A	We first train the model with existing data and as new data come we rebuild the model from scratch periodically		1	3
A	We first train the data and for every incoming data that we misclassify we use it to update the model using a very small learning rate		0	4
Q	84)A _____ selects one particular hyperplane that not only separates the points in the two classes, but does so in a way that maximizes the margin i.e. – the distance between the hyperplane and the closest points of the training set.		M	1
A	Support Vector Machine		1	1
A	Bayes Classifier		0	2
A	Decision Tree		0	3
A	Regressor		0	4
Q	Margin in SVM is defined as		M	1
A	the distance of closest example from the decision boundary/hyperplane		1	1
A	the distance between other side of the margin and data point		0	2
A	distance between two data points on the same side of hyperplane		0	3
A	distance between two data points on the opposite side of hyperplane		0	4
Q	85)Slack Variable in SVM is defined as		M	1
A	the distance of closest example from the decision boundary/hyperplane		0	1
A	the distance between other side of the margin and data point		1	2
A	distance between two data points on the same side of hyperplane		0	3

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	distance between two data points on the opposite side of hyperplane		0	4
Q	86)In closeness centrality , a node is considered important if it is relatively ----- to all other nodes		M	1
A	Far		0	1
A	Close		1	2
A	Between		0	3
A	Similar		0	4
Q	87)Google PageRank is a variant of ----- centrality		M	1
A	Betweenness		0	1
A	Closeness		0	2
A	Eigenvector		1	3
A	Vector		0	4
Q	88)A network is ----- connected if each pair of vertices is connected by a semipath.		M	1
A	Weakly		1	1
A	Strongly		0	2
A	both (a) and (b)		0	3
A	Tightly		0	4
Q	89)A user views the first 13 seconds of a 5 minute video on YouTube, then browses away. What kind of recommendation input is this?		M	1
A	A Rating		0	1
A	A Recommendation		0	2
A	Implicit input		1	3
A	Indication that the user likes the video		0	4
Q	90)All of these statements about information filtering and information retrieval are true EXCEPT ONE. Which one is NOT TRUE?		M	1
A	Information filtering evaluates new content items for match against user profiles.		0	1
A	Information filtering focuses on building profiles of long-term user interest while information retrieval focuses on building indexes of content.		0	2
A	Information filtering involves matching a user-entered query to document terms or item attributes.		1	3
A	Information retrieval often uses the TFIDF approach where terms are more relevant if they occur in few documents, but frequently in the matched document.		0	4
Q	91)In our taxonomy of recommender systems, what do we mean by "ephemeral personalization?"		M	1
A	Ephemeral personalization is based on the products preferred by people like you -- your neighborhood of users.		0	1
A	Ephemeral personalization is based on demographics or similar characteristics rather than on actions or purchases.		0	2
A	Ephemeral personalization means that you may get different recommendations the next time you log in.		0	3
A	Ephemeral personalization is based on your current navigation or market basket, but not a long-term profile of your preferences.		1	4

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
Q	92)Which of the following is a problem with using Pearson correlation (as opposed to other similarity metrics) for computing user similarities in user-user collaborative filtering?		M	1
A	Users may use different portions of the rating scale.		0	1
A	Users may not have rated any of the the same items.		0	2
A	The user may not know any other users in the system.		0	3
A	If users have only rated a small number of the same items, their correlation may be too high.		1	4
Q	93)Either vector cosine or Pearson correlation are often used to compute a weight in user-user collaborative filtering. What are these metrics trying to measure?		M	1
A	These are measures of the number of ratings users have in common.		0	1
A	These are measures of the similarity of ratings history between users.		1	2
A	These are measures of how well the recommendations match the user's preferences.		0	3
A	These are measures of how much the target user likes popular items.		0	4
Q	94)Random Surfer Model is used to find		M	1
A	Pagerank		1	1
A	Deadends		0	2
A	Spidertraps		0	3
A	Inlinks		0	4
Q	95)Suppose the Web consists of a clique (set of nodes with all possible arcs from one another) of n nodes and a single additional node that is the successor of each of the n nodes in the clique. For n = 4, determine the PageRank of the single additional node after the first iteration.		M	1
A	4 over 20		1	1
A	3 over 20		0	2
A	0		0	3
A	1 over 4		0	4
Q	96)If $\beta = 0.9$, then by how much percentage, the Spam Farm structure has amplified the external PageRank contribution?		M	1
A	46%		0	1
A	360%		0	2
A	526%		1	3
A	100%		0	4
Q	97)What is Geodesic distance?		M	1
A	is the number of position in the shortest walk from one actor to another		0	1
A	is the number of relation in the shortest walk from one actor to another		1	2
A	both		0	3
A	is the number of edges in the shortest walk from one actor to another		0	4
Q	98)The ratio of the number of relations which are reciprocated over the total number of relations in the network		M	1

Q=QUESTION	question_description	question_explanation	question_type	question_difficulty
A=ANSWER	answer_description	answer_explanation	answer_isright	answer_position
A	Clustering		0	1
A	Reciprocity		1	2
A	Transitivity		0	3
A	Classification		0	4
Q	Measures of Centrality		M	1
A	Degree & Betweenness		0	1
A	Closeness & Eigenvector		0	2
A	both (a) and (b)		1	3
A	closeness and vector		0	4
Q	99)If data set is noisy then what would be prediction of 1-Nearest neighbor classifier		M	1
A	the predicted value jumps around quite a lot		1	1
A	the predicted value is quite stable		0	2
A	there are no errors in the output		0	3
A	nothing can be said based on just noise		0	4
Q	100)Margin in SVM is expected to be		M	1
A	Large		1	1
A	Small		0	2
A	Zero		0	3
A	Infinite		0	4