	marks	question	A	B	C	D	ans
0	1	A group of 4 bits is also called?	Nibble	Byte	Kb	None	4 bits make one nibble.
1	1	There are how many types of Big Data:	3	2	1	None	Big Data is of 3 types.
2	1	Which of the following are the V's of Big Data:	All	Volume	Variety	Velocity.	This is an explaination.
3	1	Which of these is not a characterstic of Big data?	Storage	Volume	Variety	Velocity.	This is an explaination.
4	2	Which of the following is a drawback of Big Data:	Cost	Significant	Process	Fraud Detection	Big Data requires high cost to maintain huge amount of data
5	2	Fullform of GINA is:	Global Innovation Network and Analysis.	Global Invention in Networks and Analytics	Globally Investment in Neurons and Analytics	None	GINA stands for Global Innovations Networks and Analysis.
6	2	Which is the phase 3 in Data Analytics Life cycle.	Model Planning	Model Building	Data Preparation	Operationalize	Model Planning is the 3rd phase in life cycle.
7	2	GINA team thought to accomplish mainly____ goals:	3	2	1	5	GINA targeted to achieve three goals for the project.
8	2	The Data Preparation stage doesn't involve:	Analyzation	Collection	Cleansing	Processing.	This is an explaination.
9	2	Unstructured Data is further divided into how many types?	2	3	4	5	Unstructured data is divided into 2 types.
10	2	The GINA team mainly used which software tool to analyze the Data	Tableau	Hadoop	HIVE	SQL	The team used Tableau to visualize the Data.
11	2	Which of the follwing is the first step of Data Analytics Life Cycle:	Discovery	Data Preparation.	Model Planning	Data Aware	This is an explaination.
12	2	There are how many phases in data analytics life cycle:	6	5	4	7	there are 6 stages in data analytics life cycle.
13	2	SEMMA Methodology has how many stages:	5	4	6	7	SEMMA methodology has five stages.
14	2	Which phase of Life Cycle requires collaboration with stakeholders?	Phase 5	Phase 6	Phase 4	Phase 3	Phase 5 involves collaboration with stakeholders.
15	2	In Building a Model, how many phases are required:	2	3	4	5	This is an Explaination.
16	2	How much Data in the whole world is structured:	0.2	0.4	0.6	0.5	Only 20% of world's total data is structured.
17	2	10^7 bytes of memory is equal to:	1ZB	1TB	1YB	1XB	10^7 B is equal to 1 ZB.
18	2	Data Scientists in the GINA team used which technique on the textual Description of the Innovation Roadmap Idea.	Natural Language Processing(NLP)	Hadoop	HIVE	SQL	NLP technique was used on the description of Innovation Roadmap Idea.
19	2	How many types of data analytics methodologies are there?	2	4	3	6	Two types of data anlytical methodologies are there. EDA and CDA
20	3	Other name for Bell Curve is:	Normal Distribution.	Poisson Distribution	Bionomial Distribution	Bernoulli Distribution.	Bell Curve is also known as normal distribution.
21	3	One of the most important tasks in big data analytics is:	Statical Modeling	Testing of Data	Visualization	Operationalize	One of the most important tasks in big data analytics is statistical modeling
22	3	Some of the approaches considered for building the data analytics lifecycle framework best practices are:	All	CRISP-DM	SEMMA	MAD Skills	This is an explaination.
23	3	In Phase 4, the team develops datasets for:	All	Testing of Data	Training of Data	Production purposes	This is an explaination.
24	3	Fullform of CRISP-DM Methodology is:	Cross Industry Standard Process for Data Mining	Cross International Standard Process for Data Modeling	Common Industry Standard Program for Data Mining	Company's Initial Standards Progress for Data Methods	CRISP-DM stands for Cross Industry Standard Process for Data Mining.
25	3	SEMMA Methodology doesn't include which of the following stages:	Evaluate	Sample	Explore	Asses	This is an Explaination.
26	3	In Which stage, the data is monitored and analyzed to see if the generated model is creating the expected results.	Operationalize	Collection	Plan Model	Data Aware	In last phase i.e. Opeartionalize Data is monitored and analyzed to see if the generated model is creating the expected results.
27	3	Data is captured in how many ways:	3	4	5	6	Data is captured in 3 main ways.

	marks	question	A	B	C	D	ans
28	3	In phase 2 of the Data Anlaytics Life Cycle, the team performs how many analytics to get the data in the sandbox.	3	2	4	6	The team performs ETL and ELT and ETLT in 2nd phase of the cycle.
29	3	The total area under the bell curve is____unit.	1	2	3	4	Area under the bell curve is 1 unit.
30	1	Wilcoxon rank-sum test is also known as?	Mann-Whiteney U test	Mean Difference	Alternative Hypothesis	Null Hypothesis	Wilcoxon rank-sum test is also called Mann- Whiteney U Test.
31	1	Which test is also known as T-test?	Hypothesis Test	Mean Difference	K-means test	None	This is an explaination.
32	1	This equation is of which test?	Mean Difference	K-Means	Null Hypothesis	Alternative Hypothesis	This eqn is of Mean difference test.
33	1	A test of a statistical hypothesis, where the region of rejection is on a side of the sampling distribution, is called_____.	One tailed test	Two-tailed test	Tailed test	Null test	A test of a statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a one-tailed test
34	1	How many types of Statical Hypothesis is there?	2	3	4	6	There are two types of Statical Hypothesis.
35	1	Analysis of Variance is also refered as?	ANOVA	Mean Difference	Alternative Hypothesis	Null Hypothesis	ANOVA stands for Analysis of Variance.
36	1	How many steps are involved in a Hypothesis Testing?	4	2	3	5	There are 4 steps in Hypothesis testing.
37	2	The strength of evidence in support of a null hypothesis is measured by?	P-value	K-value	H-value	Null-value	The strength of evidence in support of a null hypothesis is measured by the P-value.
38	2	Difference in means is also called?	Two sample t-test	T- test	M-test	Two sample test	Difference in means is also known as two sample t test.
39	2	The k-medoids is also called_____ algorithm.	Partitioning Around Medoids (PAM)	Lloyd's Algorithm	Poisson's Algorithm	Regression	The k-medoids is also called partitioning around medoids (PAM) algorithm .
40	2	Clustering is an example of ____?	Unsupervised Learning	Supervised Learning	Classification	Regression	Clustering is an example of unsupervised learning.
41	2	Which of the following is not an advantage of K means Clustering?	Requires a Priori	Fast	Robust	easy to evaluate.	This is an explaination.
42	2	The probability of committing a Type 2 error is called	Beta	Alpha	Delta	Theta	The probability of committing a Type II error is called Beta
43	2	The_____ variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.	Less	More	Variable	Fixed	The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.
44	2	Which hypothesis is usually the hypothesis in which sample observations result is purely from chance?	Null-Hypothesis	Mean Difference	K-means test	Alternative Hypothesis	Null Hypothesis is usually the hypothesis that sample observations result purely from chance.
45	2	Classical" ANOVA for balanced data does how many things at once?	3	2	1	4	Classical" ANOVA for balanced data does three things at once.
46	2	K-mean clustering is used to solve which problems?	NP-hard problems	NP Problems	Hypothesis Problems	P problems	NP hard problems are solved using K means clustering.
47	2	The probability of committing a Type I error is called?	Alpha	Beta	Gama	Delta	The probability of committing a Type I error is called alpha
48	2	K means Clustering is also known as?	Lloyd's Algorithm	Gaussian Algorithm	Poisson's Algorithm	None	K means clustering is also called Lloyds algo.
49	3	Which algorithm requires the user to specify the number of clusters k to be generated.	K-means clustering	Gaussian Algorithm	Alternative Hypothesis	Null Hypothesis	k-means clustering requires the user to specify the number of clusters k to be generated.
50	3	K means clsutering uses which approach to solve the problems?	Expectation-maximization	Greedy Approach	Divide and Conquer	None	expectation-maximization technique is used by k means clustering.
51	3	How many factors affect the power of a hypothesis test?	3	2	1	4	The power of a hypothesis test is affected by three factors.
52	3	Law of Variance is called?	Eve's Law	Laplace Law	Poisson's Algorithm	Regression	Law of variance is also called Eve's law.
53	3	K-Medoids use which approach to solve problems?	Greedy Approach	Divide and Conquer	Recursive	None	K Medoids use greddy approach to solve problems
54	3	The time complexity of k means clustering is?	O(n^2)	O(nlogn)	O(n)	O(1)	Time complexity is O(n^2) of k means clustering.
55	3	the number (k) of clusters assumed in k-medoids is known as?	Priori	Null Hypothesis	ANNOVA	Effect size	The number k of clusters assumed known as priori.

	marks	question	A	B	C	D	ans
56	3	What is the difference between the true value and the value specified in the null hypothesis.	Effect -size	Null Hypothesis	Alternative Hypothesis	ANOVA	The effect size is the difference between the true value and the value specified in the null hypothesis.
57	3	Time complexity of k medoids is?	O(n^2)	O(nlogn)	O(n)	O(n^3)	This is an explaination.
58	3	Which algorithm aims at minimizing an objective function know as squared error function	K-means	Mean Difference	Alternative Hypothesis	ANOVA	K means algorithm aims at minimizing an objective function know as squared error function
59	1	Which algorithm was the earliest of the association rule algorithms?\n	Apriori Algorithm	Gaussian Algorithm	K means clustering	Bernoulli Distribution.	Apriori Algorithm was earliest in the association of algorithms.
60	1	The Apriori algorithm takes a_____ iterative approach to uncovering the frequent itemsets by first determining all the possible items	Bottom-Up	Top-Down	Recursive	None	The Apriori algorithm takes a bottom-up iterative approach to uncovering the frequent itemsets by first determining all the possible items
61	1	Apriori uses which structure to count candidate item sets efficiently?	BFS	DFS	Queue	Stack	Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently
62	1	"y=a+b*x^2". This equation shows which regression?	Polynomial Regression	Logistic Regreasion	Linear Regression	Lasso Regression	This is an explaination.
63	2	_____ is defined as the measure of certainty or trustworthiness associated with each discovered rule.	Confidence	Recursion	Item-set	None	Confidence is defined as the measure of certainty or trustworthiness associated with\neach discovered rule.
64	2	In which Regression, we predict the value by 1 or 0?	Logistic Regression	Linear Regression	Both	None	In Logistic Regression, we predict the value by 1 or 0.
65	2	The formula for linear regression is:	Y' = bX+A	Y' = bX - A.	Y' = bX /A.	Y' = bX * A.	The formula for linear regression is: Y' = bX + A.
66	2	Which regression is useful when there are a large number of independent variables.	Partial Least Squares(PLS) Regression	Cox Regression	Lasso Regression	Logistic Regression	PLS regression is also useful when there are a large number of independent variables.
67	2	Which regression is an approach for predicting a response using a single feature.	Linear-Regression	Logistic Regreasion	Elasticnet Regression	None	Simple linear regression is an approach for predicting a response using a single feature.
68	2	Association rule mining consists of_____ steps.	2	3	4	5	Association rule mining consists of 2 steps
69	2	Which type of regression is suitable when dependent variable is ordinal in nature?	Ordinal Regression	Linear Regression	Cox Regession	Logistic Regression	Ordinal regression is suitable when dependent variable is ordinal in nature
70	2	Which regression is used for support vector machines	ElasticNet Regression	Linear Regression	Logistic Regression	None	ElasticNet regression is used for support vector machines,
71	2	Which regression can solve both linear and non-linear models?	Support Vector Regression	Linear Regression	Logistic Regression	ElasticNet Regression	Support-Vector Regession can solve both linear and non linear models.
72	2	Which is the most common method used for fitting a regression line	Least Square Method	Mean Difference	Null Hypothesis	Classification	Least Square Method is the most common method used for fitting a regression line
73	2	_____problems are when the output variable is a real or continuous value.	Regression	Classification	Recursive	Hypothesis	A regression problem is when the output variable is a real or continuous value.
74	2	Linear Regression is a machine learning algorithm based on _____ learning regression model.	Supervised Learning	Unsupervised Learning	Recursive Learning	All	Linear Regression is a machine learning algorithm based on supervised regression algorithm.
75	2	When dependent variable's variability is not equal across values of an independent variable, it is called	Heteroscedasticity	Homooscedasticity	Multicolinearity	Outliers.	When dependent variable's variability is not equal across values of an independent variable, it is called heteroscedasticity
76	2	_____requires large sample sizes because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square	Logistic Regression	Linear Regression	Lasso Regression	ElasticNet Regression	Logistic Regression requires large sample sizes because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square
77	2	PCR Regression is divided into how many steps?	2	3	4	5	PCR regression is divided into 2 steps
78	3	L2 regularization is also called?	Tikhonov Regularization	Norm Regularization	Poisson's Regularization	None	This is an explaination.
79	3	When the variance of count data is greater than the mean count, it is a case of?	Overdispersion	Underdispersion	Dispersion	High dispersion	When the variance of count data is greater than the mean count, it is a case of overdispersion

	marks	question	A	B	C	D	ans
80	3	Which regression assumes the normal distribution of the dependent variable?	Linear-Regression	Logistic Regreasion	Elasticnet Regression	None	Linear regression assumes the normal or gaussian distribution of the dependent variable.
81	3	Nature of predicted data in regression is?	Ordered	Unordered	Both	None	Nature of predicted data in regression is ordered.
82	3	Which regression uses a binary dependent variable but ignores the timing of events.	Logistic Regression	Linear Regression	Cox Regession	Lasso Regression	Logistic regression uses a binary dependent variable but ignores the timing of events.
83	3	The Ridge Regression is also known as?	Shrinkage Regression	Percentile Regression	Elasticnet Regression	Lasso Regression	The ridge regression is also known as Shrinkage Regression.
84	3	In which regression, we calculate Root Mean Square Error(RMSE) to predict the next weight value.	Linear-Regression	ElasticNet Regression	Logistic Regression	All	In Linear Regession we calculate Root Mean Square Error(RMSE) to predict the next weight value.
85	3	The_____ is the standard deviation of the observed residuals.	Residual standard error	Mean Difference Error	Data Error	All	The residual standard error is the standard deviation of the\nobserved residuals.
86	3	Which Regression is used when dependent variable has count data.	Poisson Regression	Linear Regression	Cox Regession	Lasso Regression	Poisson regression is used when dependent variable has count data.
87	3	_____regression can handle both over-dispersion and under-dispersion.\n	Quasi-Poisson regression	Cox Regression	Elasticnet Regression	Linear Regression	Quasi-Poisson regression can handle both over-dispersion and under-dispersion.\n
88	3	___ is the regularization parameter in Lasso Regression?	λ	θ	Ω	β	λ is the regularization parameter in lasso regression.
89	1	Decision Tree is a hierarchical model that does the separation of the\ninput space into class regions using:	Recursion	Pointers	Greedy Approach	Divide and Conquer	Decision Tree is a hierarchical model that recursively does the separation of the\ninput space into class regions
90	1	Learning Algorithm of Decision Tree is:	Greedy Approach	Divide and Conquer	Both	None	Decision Tree uses greedy approach for learning algorithm.
91	1	Normal Distribution is also called?	Gausiann Distribution	Bernoulli Distribution	Naïve Bias	Binary Distribution	This is an explaination.
92	1	Classification has how many phases:	2	3	4	5	There are 2 phases of classification.
93	1	"Every pair of features being classified is independent of each other".This principle is used by:	Naïve Bais Classifier	Decision Tree	Bernoulli Distribution	Normal Distribution	Naïve Bias uses the principle that every pair of features being classified is independent of each other.
94	2	This equation is of which theorem?	Gausiann Distribution	Binary Distribution	Naïve Bias	Gross-Entrpoy	This is an explaination.
95	2	In Naïve Bias, The Datasets are divided into how many types?	2	3	4	5	data sets are divided into two types in naïve bias.
96	2	Decision trees can be used to predict non-categorical values is called?	Regression Trees	Categorial trees	Normal tree	None	Decision trees can be used to predict non-categorical values is called regression trees
97	2	An attribute with____Gini index should be preferred in a decision tree.	Lower	Higher	Recursive	Negative	an attribute with lower Gini index should be preferred.
98	2	In Naïve Bias, if any two events A and B are independent, then,	P(A,B)=P(A)P(B)	P(A,B)=P(A)/P(B)	P(A,B)=P(B)	P(A,B)=P(B)P/(A)	If any two events A and B are independent, then,P(A,B)=P(A)P(B)
99	2	What is the measure of uncertainty of a random variable in a decision tree.	Entropy.	Gain	Gini Index	None	Entropy is the measure of uncertainty of a random variable
100	2	Which of the following is not true for decision trees?	Stable	Easy to understand	Easy to explain	Easy to evaluate.	this is an explaination.
101	2	Decision tree algorithm falls under the category of which learning?	Supervised	Unsupervised	Regression	Classification	Decision tree algorithm falls under the category of supervised learning
102	2	False Positives and False Negatives is an application of which theorem?	Bayes' Theorem	Binary Distribution	Bernoulli Distribution	Normal Distribution	One of the use Bayes Theorem is false positives and false negatives.
103	2	Decision Tree used in mining the data are of how many types?	2	3	4	5	There are 2 types of decision trees used in data mining.
104	3	In Bayes' Theorem, P(A) and P(B) are the probabilities of observing A and B respectively; they are known as:	Marginal Probability	Normal Distribution	Bernoulli Distribution	Parallel Algorithm.	P(A) and P(B) are the probabilities of observing A and B respectively; they are known as the marginal probability.

	marks	question	A	B	C	D	ans
105	3	ID3 Algorithm in a decision tree stands for?	Iterative Dichotomiser 3 (ID3)	Interval Driven	Interconnected Decision	None	ID3 stands for Iterative Dichotomiser 3 (ID3)
106	3	Probably the best way of estimating performance for very small\ndata sets is:	Boot Strapped Method	Normal Distribution	Naïve Bias	Binary Distribution	Probably the best way of estimating performance for very small data sets is bootstrapped method
107	3	The Decision Tree works on which form?	Disjunctive Normal Form	Product of Sum	Bijective Form	Conjuctive Form	Decision Tree works on Disjunctive normal form.
108	3	The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a_____ distribution.	1-D	2-D	3-D	NONE	The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution.
109	3	Theoretical concept to evaluate Classfiers is:	COLT	PAC Model	Naïve Bias	Prediction.	This is an explaination.
110	3	_____is a metric to measure how often a randomly chosen element would be incorrectly identified	Gini Index	Entropy	Pointer	Gross-Entrpoy	Gini Index is a metric to measure how often a randomly chosen element would be incorrectly identified
111	3	The most notable types of decision tree algorithms are:	3	2	1	4	The most notable types of decision tree algorithms are 3
112	3	Which process is completed when the subset at a node all has the same value of the target variable?	Recursive Partitioning	Termination	Transformation	Prediction.	The recursive partition is completed when the subset at a node all has the same value of the target variable
113	3	The_____ method reserves a certain amount for testing and uses the remainder for training.	Holdout	Parallel Algorithm	Naïve Bias	Normal Distribution	The holdout method reserves a certain amount\nfor testing and uses the remainder for training
114	3	This equation is of which theorem?	Bayes' Theorem	Normal Distribution	Bernoulli Distribution	Gross-Entrpoy	This is an explaination.
115	3	"Independence among the features". This is an assumption in:	Naïve Bais Classifier	Bernoulli Distribution	Parallel Algorithm	Binary Distribution	Independence among the features is an assumption in Naïve bias.
116	3	Error rate obtained from training data is called:	Resubstitution Error	Grid	Gini Index	True error	error rate obtained from training data is called resubstitution error.
117	3	In Decision Tree entropy is _____ to content.	proportional	inverse	High	Less	This is an explaination.
118	3	In Decision Tree, No root-to-leaf path should contain the same discrete attribute _____.	Twice	Once	Thrice	Four Times.	No root-to-leaf path should contain the same discrete attribute twice
119	1	Using_____, designers can make information understandable for stakeholders.	Data Visualization	Classification	Regression	Supervised Learning.	Using data visualization methods, designers can make information understandable for stakeholders.
120	1	The additional visual methods include:	All	Tree Map	Parallel Coordinates	Semantic Networks.	This is an explaination.
121	1	Data Visualization tools Doesn't include:	Ms--Excel	Tableau	Power BI	Jupyter	This is an explaination.
122	1	Which of the following requires Javascript Knowledge to run the visualization tool?	All	Chart.js	Polymap	Sigmajs	This is an explaination.
123	1	Merits of Tableau doesn't include which factor:	Cost	Performance	Usage	Computation	Merits of tableau doesn't include the cost factor.
124	1	Which of these is not a type of Big Data Visualization.	Pictograph	Bar-Graph	Line-Chart	Pie-Chart	This is an explaination.
125	2	The drag-and-drop editor od which tool makes it easy to create professional-looking designs without a lot of visual design skill.	Infogram	Google Chart	Tableau	Grafana	The drag-and-drop editor of Infogram makes it easy to create professional-looking designs without a lot of visual design skill.
126	2	How many V's are defined for Data Visualization.	4	6	2	3	There are 4 V's of Data visualization.
127	2	Which of the following is not a free Data Visualization tool?	Tableau	Google Chart	Jupyter	Hub-Spot CRM	Tableau is a chargeable tool of data visualization.
128	2	Companies that work with both traditional and big data use which technique to look at customer segments or market shares?	Pie-Chart	Bar-Graph	Stream graph	Line-Chart	Companies that work with both traditional and big data may use pie chart to look at customer segments or market shares
129	2	Visualization of Data includes which of the following problems:	All	Information Loss	Visual Noise	Large Image Perception.	This is an explaination.
130	2	Mainly, Data Visualization has how many types of challenges?	5	6	4	2	There are 5 main challenges to data visualization.

	marks	question	A	B	C	D	ans
131	2	Which tool uses HTML5/SVG to visualize data	Google Charts	Jupyter	Grafana	Tableau	Google charts uses HTML5/SVG since its browser compatible.
132	2	According to Colin Ware's Information Visualization: Perception for Design, he defines_____ pre-attentive visual properties.	4	2	1	3	According to Colin Ware's Information Visualization: Perception for Design, he defines four pre-attentive visual properties
133	2	_____ is based on space-filling visualization of hierarchical data.	Tree-Map	Stream graph	Bar-graph	Line-Chart	Tree map method is based on space-filling visualization of hierarchical data
134	2	Which graph shows the dependency relationships between activities and current schedule status.	Gantt-Chart	Line-Chart	Pie-Chart	Bar-Graph	Gantt chart show the dependency relationships between activities and current schedule status.
135	2	Another name for distribution free data is:	Non parametric data	Parametric Data	static data	Dynamic data	Non parametric data is also called distribution free data.
136	2	Which chart is used for comparison of values, such as sales performance for several persons or businesses in a single time.	Bar-Graph	Gantt-Graph	Line-Chart	Pie-Chart	Bar Graph is used for Comparison of values, such as sales performance for several persons or businesses in a single time
137	2	_____are graphics in the field of statistics used to visualize quantitative data.	Graphical-Techniques	Line-Chart	Regression	Classification	Graphical Techniques are graphics in the field of statistics used to visualize quantitative data.
138	2	_____ can handle several factors for a large number of objects per single screen, so it satisfies the data variety criterion.	Parallel Coordinates	Stream graph	Google Chart	Jupyter	Parallel Coordinates can handle several factors for a large number of objects per single screen, so it satisfies the data variety criterion
139	3	Chart.js provides how many types of charts?	8	5	3	6	This is an explaination.
140	3	Which visualization tool supports mixed data sources, annotations, and customizable alert functions, and it can be extended via hundreds of available plugins.	Grafana	Tableau	Google Chart	Jupyter	Grafana supports mixed data sources, annotations, and customizable alert functions, and it can be extended via hundreds of available plugins.
141	3	Which tool was created specifically for adding charts and maps to news stories.	Data Wrapper	Tableau	Google Chart	Jupyter	Datawrapper was created specifically for adding charts and maps to news stories.
142	3	Conventional Visualization methods doesn't include:	Mekko Chart	Pie-Chart	Bar-graph	Histogram	Mekko chart is a new technique to visualize data.
143	3	_____ is a type of a stacked area graph, which is displaced around a central axis, resulting in flowing and organic shape.	Streamgraph	Bar-Graph	Pie-Chart	Line-Chart	Streamgraph is a type of a stacked area graph, which is displaced around a central axis, resulting in flowing and organic shape
144	3	Which visual tool includes over 150 chart types and 1,000 map types?	Fusion charts	Tableau	Google Chart	Jupyter	Fusion charts includes over 150 chart types and 1,000 map types
145	3	Which graph/chart is a graphical representation of logical relationship between different concepts. It generates directed graph, the combination of nodes or vertices, edges or arcs, and label over each edge.	Semantic Networks	Bar-Graph	Pie-Chart	Line-Chart	A semantic network is a graphical representation of logical relationship between different concepts. It generates directed graph, the combination of nodes or vertices, edges or arcs, and label over each edge
146	3	According to SAS we can process only_____ of information per second on a flat screen.	1 Kilobit	1 Byte	1 Bit	1 MB	According to SAS we can process only 1 kilobit of information per second on a flat screen
147	3	There are____ steps for interactive data visualization:	4	5	3	6	This is an explaination.
148	3	When working with big data, companies can use which visualization technique to track total application clicks by weeks, the average number of complaints to the call center by months, etc.\n\n	Line-Chart	Bar-Graph	Pie-Chart	Stream graph	When working with big data, companies can use the line chart visualization technique to track total application clicks by weeks, the average number of complaints to the call center by months, etc.\n\n
149	1	Which of the following Enterprises use HBase?	All	Facebook	Netflix	Adobe	This is an explaination.

	marks	question	A	B	C	D	ans
150	1	Which NLP is used in the present era?	Neural NLP	Symbolic NLP	Statical NLP	None	From 2010, Neural NLP is being used.
151	1	The Computer World magazine states that unstructured information might account for more than_____of all data in organizations.	70-80%	0.9	0.5	0.6	The Computer World magazine states that unstructured information might account for more than 70%–80% of all data in organizations.
152	1	Almost all of the information we use and share every day, such as articles, documents and e-mails, are completely_____.	Unstructured	Structured	Semantic	None	Almost all of the information we use and share every day, such as articles, documents and e-mails, are completely or partly unstructured
153	1	Which standard provided a common framework for processing information to extract meaning and create structured data about the information?	Unstructured Information Management Architecture (UIMA)	Management Architecture for Data	Data Architecure	None	The Unstructured Information Management Architecture (UIMA) standard provided a common framework for processing this information to extract meaning and create structured data about the information.
154	2	The base Apache Hadoop framework is composed of the how many modules?	4	2	3	6	The base Apache Hadoop framework is composed of the four modules.
155	2	No-SQL doesn't include which software?	MS-SQL	HBASE	DyanoDB	MongoDB	This is an explaination.
156	2	There are _____main types of OLAP systems.	3	2	5	6	There are 3 types of OLAP systems.
157	2	SQL alternative in Apache HIVE is called?	HIVEQL	BASEQL	SPARK-QL	H-QL	HIVE-QL is the alternative to SQL in Apche Hive family.
158	2	MapReduce program executes in how many stages?	3	2	5	4	MapReduce program executes in three stages.
159	2	How many types of NO-SQL database are there?	4	3	2	6	There are 4 types of databases in NO-SQL.
160	2	MapReduce is a processing technique and a program model for distributed computing based on which programming Language?	JAVA	Python	C++	R	MapReduce is a processing technique and a program model for distributed computing based on java
161	2	Hive supports how many properties of transactions?	4	3	2	1	Hive supports all four properties of transactions
162	2	HDFS consists of only one Name Node that is called as?	Master Node	Slave Node	Both	None	HDFS consists of only one Name Node that is called the Master Node.
163	2	Which Apache Software is needed to process massive amounts of data for the purposes of natural-language search?	Apache HBASE	Apache Spark	Apache-PIG	Apache-mahout	Hbase to process massive amounts of data for the purposes of natural-language search
164	2	Which database store data in a format other than relational tables	NO-SQL	HIVESQL	SPARK-QL	H-QL	No-sql databases that store data in a format other than relational tables.
165	2	Which is a project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms focused primarily on linear algebra?	Apache Mahout	Apache Spark	Apache-PIG	Apache HBASE	Mahout is a project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms focused primarily on linear algebra.
166	2	Which model is a specialization of the split-apply-combine strategy for data analysis?	MapReduce	Hadoop	HBASE	HIVE	MapReduce model is a specialization of the split-apply-combine strategy for data analysis.
167	2	All Hadoop commands are invoked by which command?	$HADOOP_HOME/bin/hadoop	$HADOOP/bin/hadoop	$HADOOP_HOME/hadoop	$HADOOP_HOME/bin	All Hadoop commands are invoked by the $HADOOP_HOME/bin/hadoop command
168	3	The table typically enforces the schema when the data is loaded into the table. This enables the database to make sure that the data entered follows the representation of the table as specified by the table definition. This design is called?	Schema on Write	Schema on Read	Schema for Read Write	None	The table typically enforces the schema when the data is loaded into the table. This enables the database to make sure that the data entered follows the representation of the table as specified by the table definition. This design is called schema on write.

	marks	question	A	B	C	D	ans
169	3	Which command formats the DFS filesystem?	Namenode -format	Node -format	Name -format	Format	Namenode -format command formats the DFS file system.
170	3	Which command applies the offline fsimage viewer to an fsimage?	oiv	fs	fc	ov	oiv applies the offline fsimage viewer to an fsimage.
171	3	Hadoop requires which Java Runtime Environment (JRE) or higher version?	1.6	1.2	1.5	1	Hadoop requires Java Runtime Environment (JRE) 1.6 or higher
172	3	Every Data node sends a Heartbeat message to the Name node every____ seconds and conveys that it is alive.	3	2	4	1	Every Data node sends a Heartbeat message to the Name node every 3 seconds and conveys that it is alive
173	3	HDFS can store files upto:	1 TB	1 GB	1ZB	1PB	HDFS can store upto1 TB of files.
174	3	Which of the following is a wide-column store?	HBase	SQL	DyanoDB	MongoDB	HBASE is a popular wide columnn store.
175	3	Which node acts as both a DataNode and TaskTracker in Hadooop.	Slave Node	Data Node	Admin Node	Name Node	A slave or worker node acts as both a DataNode and TaskTracker.
176	3	HDFS system uses which protocol for communication?	TCP/IP	TCP	UDP	IP	HDFS system uses TCP/IP sockets for communication
177	3	HDFS has how many services?	5	4	2	6	HDFS has five services.
178	3	_____is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis	Apache HIVE	Apache Spark	Apache-PIG	Apache HBASE	HIVE is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis