Q1) The line offset as the key. The content of the line is value as Text.

Q2) JAVA and HADOOP_CLASSPATH

Q3) The reason this is happening is that the user-defined Mapper and Reducer classes have to be defined as static classes; because they are not, we are facing this error.

Q4) FALSE

Q5) 1) It lacks a typical per-directory data structure to list each file in the directory and 2) it does not support aliases like sym links or hard links, unlike Unix

Q6) YES

Q7) Google Display Network

Q8) An ad exchange is a technology platform that facilitates the buying and selling of media advertising inventory from multiple ad networks

Whereas

An ad network is an aggregator that collects ad inventory from publishers and sells it to advertisers. It acts as an intermediary.

Q9) a) It's called the "Long Tail" effect. B) Infrequent events can cumulatively outweigh the initial portion of the graph, such that in aggregate they comprise the majority.

Q10) Likely less than $50.00

Q11) b. c. d. are the 3 correct answers (they would show up)

Q12) A) A rule of inference used to draw logical conclusions, which states that if p is true, and if p implies q ($p \longrightarrow q$), then q is true.
B) From fact#1, Rover barks and eats Alpo, so Rover is a dog (this is fact #5).
From #3, since Rover is a dog, Rover is brown (#6 is formed) which is what we want

Q13) We use it while computing the pagerank values for each website.

Q14) 1.0

Q15) <listener event = "newSearcher" class = "org.apache.solr.schema.ExternalFileFieldReloader"/>

<listener event = "firstSearcher" class = "org.apache.solr.schema.ExternalFileFieldReloader"/>

Q16) I chose the _text_ field (underscore text underscore) to make sure that Solr always queries from a particular field by default.

Q17) We use it to extract 1) attributes, 2) text and 3) HTML

Q18) query, index, delete, commit, and optimize

Q19) "resourcename", "article_tag", "article_author"

Q20) 1- Mapper, 2- StringTokenizer tokenizer = new StringTokenizer(line); 3-tokenizer.hasMoreTokens(), 4- Reducer, 5- reduce, 6- value.get()

Q21) edit distance algorithms or n-gram matching

Q22) It's a cognitive error

Q23) The name of the algorithm is "Stupid backoff algorithm".

Q24) a) Cosine Similarity: -1 to 1 (as per page 20 of clustering slides). B) Pearson Coefficient: -1 to 1, C) Jaccard Similarity: 0% to 100%

Q25) Who, When, Where, What

Q26) FALSE

Q27) TRUE

Q28) Featured snippets always show up above the #1 ranked result on the search results page. Typically, they appear above the "Fold". Google pulls snippet answers from pages that rank on Page 1 of the results for that query (spots #1 through #10) - but the page that wins the featured snippet isn't necessarily the #1 result.

Q29) FALSE

Q30) TRUE

Q31) Rich snippets allow websites to embed structured data in their webpages with the objective of displaying structured data to a user in a visually outstanding way. They give the users convenient summary information about their search results. Rich snippets are supported by Google's search engine whereas regular snippets are just generally extracted from webpages without any defining metadata for them.

Q32) It stands for "People also ask"

$$\lambda_{\text{LR}} = -2\ln\left[\frac{\sup_{\theta\in\Theta_0} \mathcal{L}(\theta)}{\sup_{\theta\in\Theta} \mathcal{L}(\theta)}\right]$$

Q33) Log Likelihood Ratio;

Q34) 1) Yes 2) Yes

Q35) 1) Inertia 2) Dunn Index, basically implying HIGH intra-class similarity and LOW inter-class similarity

Q36) Entity

Q37) Oslo and Paris, respectively, would be the leaf nodes for this tree.

Q38) Entity – Location – Region – Country – District

Q39) A) Estie and Dave have the greatest similarity, whereas B) Estie and Cindy have the least similarity.

Q40) Collaborative Filtering based recommendation system