This is an open book exam but no Internet access allowed. If you are found using the Internet, you will be reported for cheating.

Section A: Based on Assignment 1 working knowledge [25 Points]

1. In assignment 1, you created an API Gateway with a POST method. You also created a Lambda function that can handle the POST request. What are some reasons for using API Gateway instead of directly invoking the Lambda function?

2. In assignment 1, what is a session in Lex? How can it be useful?

3. In assignment 1, you notice that some of the documents are not being indexed in your elasticsearch index, and there are no error messages. How would you troubleshoot this scenario and ensure that all documents are indexed correctly? List at least 2 points of failures.

4. In assignment 1, When developing a RESTful API with AWS API Gateway, you encounter an issue where certain HTTP requests, particularly those with non-standard headers, are failing unexpectedly. You suspect that preflight requests might be involved. Explain the concept of preflight requests, why they are important for web security, and how you would troubleshoot and resolve this issue in your API Gateway setup.

5. In assignment 1, you created a DynamoDB table for storing restaurant data. You used restaurant/business ID as the primary key. What could possibly go wrong if you use insertedTimestamp as the primary key? How about using insertedTimestamp as the secondary key?

Section B: System Design [30 Points]

You want to extend the Assignment 1 to add the following features:
1. Based on user preferences and past search, 5 restaurants are recommended to the user via email or SMS everyday.
2. When a user searches for restaurants, the daily specials from those restaurants, if any, are also shown to the user. Note that these deals are released by restaurants on a daily basis through an imaginary Yelp API, say dailySpecial that you are allowed to invoke.

You need to extend the architecture to accommodate the above two features. You need to state all the required data model details, APIs that need to be supported. You need to draw the end-to-end event interaction for the above two features to illustrate your idea details.

Section C: Papers [25 points]

1. GFS: Using Fig 1, describe how a chuck server failure is handled in GFS.
2. GFS: Why single GFS master is used? How does GFS recover from GFS master failure?
3. Dremel: Explain briefly but precisely how Dremel architecture is designed both based on concepts from Web Search and Parallel DBMS.
4. Dremel: What are the two key reasons Dremel is so fast and why?
5. BigQuery: What are key differences between BigQuery and MapReduce? Give examples of use cases where BigQuery is preferred over MapReduce and vice versa.

Section D: Lecture Notes [20 Points]

1. Why is hybrid cloud beneficial for an enterprise compared to either public or private cloud?
2. What is paravirtualization and why is paravirtualization more efficient than full virtualization?
3. Explain how iterative memory copy is leveraged for live migration?
4. What are the key differences between VM and containers?