**Birla Institute of Technology & Science, Pilani**

**Work Integrated Learning Programmes Division**

**First Semester 2022-2023**

**Mid-Semester Test (EC-2 Regular)**

Course No. : SE ZG583

Course Title : Scalable Services

No of questions: 4

No of pages: 4

Nature of Exam : Open Book

Weightage : 30%

Duration : 2 Hours

Date of Exam : 24/09/2022 (FN)

Note to Students:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Question 1: Set A [4\*2=8]

1. Differentiate between horizontal scalability and vertical scalability. Hadoop belongs to which of this class?
2. What is the use of the blockreport received from the DataNode?
3. What is the single point of failure in Hadoop? How Hadoop does manages this problem?
4. Consider a file of size 1024 MB to be stored in a Hadoop cluster with block size 32MB. How many blocks will be occupied in total after the file is stored? Assume the replication factor is configured 3x in the version of Hadoop.

Question 1: Set B [4\*2=8]

1. Differentiate between horizontal scalability and vertical scalability. Hadoop belongs to which of this class?
2. How namenode knows if any slave node has crashed? What steps does it take to handle the failure of slave node.
3. What is the single point of failure in Hadoop? How does Hadoop manages this problem?
4. Consider a file of size 1024 MB to be stored in a Hadoop cluster with block size 32MB. How many blocks will be occupied in total after the file is stored? Assume the replication factor is configured 4x in the version of Hadoop used.

Question 2: Set A [2+2=4]

a) Among the three dimensions of the Scale cube, which of the dimension/s is/are right fit to limit the effects of failure and isolate them. Give Reason.

b) Which dimension of scaling is required for a multi-tentant application that requires each client or group of clients is assigned to a specific cluster of nodes

Question 2: Set B [2+2=4]

a) Split by X-axis implementations do not scale well with increases in data or application size. State True or False with justification.

b) Which dimension of scaling is required for a multi-tentant application that requires each client or group of clients is assigned to a specific cluster of nodes

Question 3: SET A

For each of the scenario given below analyse which of the listed tools/technology would you use and the approach you will follow to have an efficient scalable system. Elucidate the reason for your choice. List out the pros and cons (if any) for your choice as well. [Note: You can propose an approach with other tools/technology, not in the listing as well, but with proper justification.] [6]

a) A cloud kitchen facility wants to extract any feedback it receives on its instagram handle and move it to a Hadoop cluster for analysis. The feedback also needs to be persisted for future use. What would you use [Apache Kafka/Spark/RabbitMQ]?

b) If you have a Critical API that must respond fast to its clients, but has downstream service that could take a while to respond. Would you use Kakfa or RabbitMQ?

Question 3: SET B

For each of the scenario given below analyse which of the listed tools/technology would you use and the approach you will follow to have an efficient scalable system. Elucidate the reason for your choice. List out the pros and cons (if any) for your choice as well. [Note: You can propose an approach with other tools/technology, not in the listing as well, but with proper justification.] [6]

a) Suppose a large retail chain firm wants to get a real-time dashboard to track how many products are being purchased, shipped and delivered to customers. The order management system keeps pushing the order status to the Kafka queue. What would you use (Spark/Mapreduce/Hive)?

b) When the ads are displayed to the user, you need to track how many advertisements the user saw, in which position, and under which search criteria ads were chosen. You’d need to track and store the context information and certain user activities for future analysis. What would you use (Spark/Mapreduce/Hive)?

**Question 4 Set A**

A popular newspaper plans to launch a separate digital edition of a newspaper. The newspaper offers both free and payed content. To view the payed content, users need to subscribe. The newspaper staff using the site should be able to publish content in the application. The newspaper is delivered via website, tablet, and mobile app versions to the end users.

Functionalities:

* The content consists of articles, each of which has its own separate page.
* The system needs to have a user management that includes both readers and authors, with appropriate rights for different levels of access.
* The website, tablet, and app versions should all use a common REST API, provided by the server side of the system.
* The readers can provide feedback rating to the content
* The site is under continuous development, new features need to be added quickly. They include to review analytics dashboard for authors pertaining to the content they’ve written so that they can get feedback on their effectiveness, interactive content.

Design a scalable architecture for this application so that they can plan to expand it features later.

a) Identify the system operations (commands and queries) for this application based on the core functional use case of the application [3]

b) Identify the business capabilities and map them to services [3]

c) Assign the system operations to the services and identify the collaboration between the services [2]

d) How will the different services of your system communicate with each other and with client? Justify your answer. [2]

e) Suggest caching ideas to achieve faster response time for end users and reduce the load on the servers. [2]

**Question 4 SET B:**

A popular newspaper plans to launch a separate digital edition of a newspaper. The newspaper offers both free and payed content. To view the payed content, users need to subscribe. The newspaper staff using the site should be able to publish content in the application. The newspaper is delivered via website, tablet, and mobile app versions to the end users.

Functionalities:

* The content consists of articles, each of which has its own separate page.
* The system needs to have a user management that includes both readers and authors, with appropriate rights for different levels of access.
* The website, tablet, and app versions should all use a common REST API, provided by the server side of the system.
* The readers can provide feedback rating to the content,
* The site is under continuous development, new features need to be added quickly. They include to review analytics dashboard for authors pertaining to the content they’ve written so that they can get feedback on their effectiveness, interactive content.

Design a scalable architecture for this application so that they can plan to expand it features later.

a) Identify the system operations (commands and queries) for this application based on the core functional use case of the application [3]

b) Identify the business capabilities and map them to services [3]

c) Assign the system operations to the services and identify the collaboration between the services [2]

d) How will the different services of your system communicate with each other and with client? Justify your answer. [2]

e) What kind of data consistency model will you provide when multiple readers are reading the same content from different parts of the world and would like to add feedback to the content? [2]

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*