SAMPLE CLOUD COMPUTING QUESTIONS

1. Is it possible to have an application-level virtualization? Support your answer by explaining the reasons and examples.

Answer: *any answer that conveys the following information is acceptable.*

Application-level virtualization allows applications to be run in runtime environments that do not natively support all the features required by such applications. In this scenario, applications are not installed in the expected runtime environment but run as they were. These techniques are generally concerned with partial file systems, libraries, and operating system component emulation. Such emulation is performed by a thin layer—a program or an operating system component—that executes the application. Emulation can also be used to execute program binaries compiled for different hardware architectures.

One of the most popular solutions for implementing application virtualization is Wine, a software application that allows Unix-like operating systems to execute programs written for the Microsoft Windows platform.

2. Design a WordCount system using the MapReduce techniques. Draw a figure and explain how your design works based on the given two sentences.
Input:
```
This is an apple.
Apple is red in color.
```

Solution:
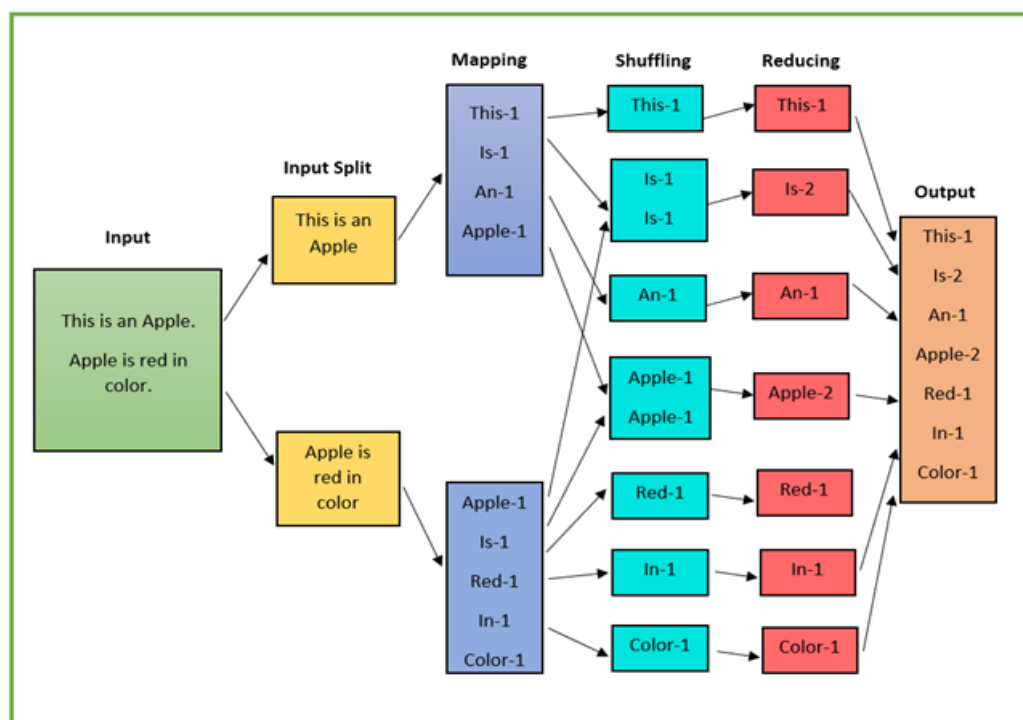
3. What is serverless computing? Is it the same as function as a service? Why is it called serverless?
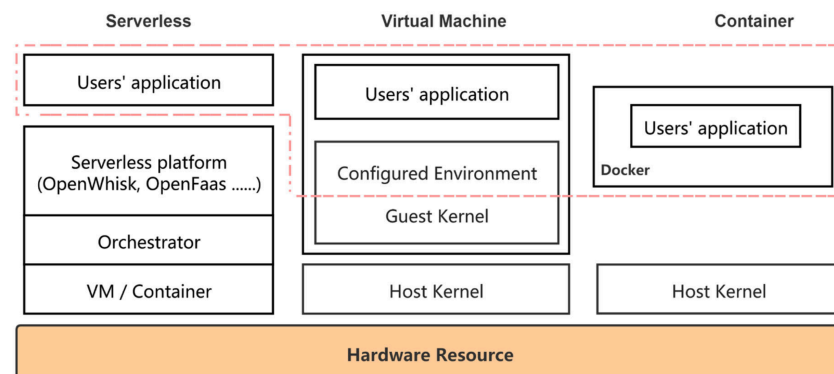
Answer:

Serverless Computing = FaaS(Function-as-a-Service) + BaaS(Backend-as-a-Service). One fallacy is that Serverless is interchangeable with FaaS. To be precise, they both are essential to serverless computing. The FaaS model enables function isolation and invocation, whereas Backend-as-a-Service (BaaS) provides overall backend support for online services.

In the FaaS model (aka the Lambda paradigm), an application is sliced into functions or function-level microservices. The function identifier, the language runtime, the memory limit of one instance, and the function code blob URI (Uniform Resource Identifier) together define the existence of a function.

The BaaS covers a wide range of services that any application relies on and can be categorized into it—for example, the cloud storage (Amazon S3 and DynamoDB), the message bus system for passing (Google Cloud pub/sub), the message notification service (Amazon SNS), and DevOps tools (Microsoft Azure DevOps).

A serverless platform serves as the interaction medium between the application and the underlying infrastructure, simplifying application development by sheltering resource management for developers.

Consequently, no server can be observed from the user perspective, even though serverless platforms still rely on servers to set up the basic runtime environment.

4. Considering the use of a control theory to optimize cloud resource management, what are the sources of instability when applying the control theory?

Answer:

- The delay in getting the system reaction after a control action.
- The granularity of the control, the fact that a small change enacted by the controllers leads to very large changes of the output.
- Oscillations, when the changes of the input are too large and the control is too weak, such that the changes of the input propagate directly to the output.

Sample Multiple Choice Question

1. Which component in GFS is primarily responsible for managing metadata, such as namespace and file-to-chunk mapping?

    A. The chunk server

    B. The client interface

    C. The replication manager

    D. The single master node

    E. The distributed lock service

    F. The metadata proxy

Correct Answer: D. The single master node

2. Which statement best describes the role of chunk servers in GFS?

    A. They maintain the file namespace and metadata

    B. They manage client authentication and access control

    C. They store the actual data chunks and handle read/write requests

    D. They provide encryption services for data in transit

    E. They balance the overall network load

    F. They directly communicate with the master for all operations

Correct Answer: C. They store the actual data chunks and handle read/write requests

3. Why does GFS use a single master node?

    A. To eliminate the need for distributed consensus

    B. To centralize metadata management and simplify coordination

    C. To increase data replication performance

    D. To allow clients to access data directly without authentication

    E. To balance network traffic between different data centers

    F. To store both metadata and actual file contents

Correct Answer: B. To centralize metadata management and simplify coordination

4. What is the primary goal of the Ray framework?

    A. To provide a specialized deep learning framework for neural networks

    B. To optimize traditional batch processing for big data analytics

    C. To support distributed execution of AI and machine learning applications

    D. To replace Kubernetes for container orchestration

    E. To provide a real-time streaming system for IoT data processing

    F. To offer a cloud-based service for GPU-based computations

Correct Answer: C. To support distributed execution of AI and machine learning applications

5. What component in Ray is responsible for managing distributed task execution?

    A. The Ray Client

    B. The Ray Master Node

    C. The Global Control Store (GCS)

    D. The Ray Trainer Module

    E. The Cluster State Manager

    F. The AI Execution Engine

Correct Answer: C. The Global Control Store (GCS)

6. Which of the following is NOT a key feature of Ray?

    A. Support for heterogeneous computing environments

    B. Built-in distributed reinforcement learning support

    C. Automatic scheduling of machine learning training jobs

    D. Direct execution of SQL queries on large datasets

    E. Fault-tolerant task execution

    F. Actor-based execution for dynamic workloads

Correct Answer: D. Direct execution of SQL queries on large datasets

7. What is a Borgcell?

    A. A collection of machines managed by a single Borg master

    B. A lightweight virtual machine used in Borg

    C. A distributed file system for storing application data

    D. A backup cluster used for fault tolerance

    E. A specialized scheduling algorithm within Borg

    F. A network partitioning mechanism

Correct Answer: A. A collection of machines managed by a single Borg master

8. What component of Borg is responsible for tracking job and task states?

    A. The Borgmaster

    B. The Node Agent

    C. The Borg UI

    D. The Borg Database

    E. The Priority Queue Manager

    F. The Job Monitor

Correct Answer: A. The Borgmaster