

MILWAUKEE SCHOOL OF ENGINEERING
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE DEPARTMENT

CS4230 – Distributed and Cloud Computing 2023 - **SOLUTION**

Midterm

8.5 x 11 Note Sheet both sides

Name: _____

For FULL credit, please show all your work.

- | | |
|--|-------|
| 1) Cloud Computing Concepts | _____ |
| 2) Virtualization | _____ |
| 3) Containers, Orchestration, and Deployment | _____ |
| 4) Distributed Storage Technology | _____ |
| Total (100 Points) | _____ |

1) Cloud computing concepts (each question 5 points)

1a) Which resources are typically provided by an *Infrastructure as a Service* cloud computing delivery model?

- A. Applications
- B. Virtual machines
- C. Virtual private networks
- D. Middleware software stacks

B

1b) What is the role of *virtualization* in cloud computing?

- A. It removes operating system inefficiencies.
- B. It improves the performance of web applications.
- C. It optimizes the utilization of computing resources.
- D. It adds extra load to the underlying physical infrastructure and has no role in cloud computing.

C

1c) In cloud computing, which *term* is used to describe the nature of IT infrastructure and allows dynamic scaling depending on the demand for resources?

- A. Elasticity
- B. Clustering
- C. Consolidation
- D. Virtualization

A

1d) What *technology* is used to logically assign and separate physical resources such as memory and CPU in a cloud-computing model?

- A. Bios
- B. Hypervisor
- C. Load balancer
- D. Auto scaling

B

1e) Why can cloud providers experience lower variability in demand over time than individual users?

- A. Economies of scale
- B. Cost associativity
- C. Statistical multiplexing
- D. Fine-grained billing

C

2) Virtualization

2a) Hypervisor is often used to describe the interface provided by what cloud computing functional layer. *Select all that are true.*

- A. Software as a Service (SaaS)
- B. Platform as a Service (PaaS)
- C. Application as a Service (AaaS)
- D. Infrastructure as a service(IaaS)

D.

2b) Hypervisor. *Select all statements that are true.*

- A. Multiple instances of a variety of OS's may share the same server's virtualized hardware resources.
- B. Presents to the guest operating systems a virtual operating platform, and monitors the execution of the guest operating systems.
- C. Multiple instances of a variety of OS's may share the same server's virtualized hardware resources.
- D. A Type 2 hypervisor runs directly on the host's hardware to control the hardware and to monitor guest operating systems.

A, B, C

2c) Containerization. *Select all that are true.*

- A. Allows applications to be packaged with their dependencies and run in an isolated environment.
- B. Emulate the hardware of a physical computer and allow multiple operating systems to run on a single physical host.
- C. Allows applications to be packaged with their dependencies and run in an isolated environment.
- D. Share the same kernel as the host operating system, but have their own file system, networking, and process space.

A, C., D

2d) Clearly and succinctly explain the concept of paravirtualization and how it is different than using a traditional hypervisor.

- Presents a software interface to virtual machines that is similar to, but not identical to that of the underlying hardware – means guest OS's must be modified!
- The difference is that the guest OS needs to “know” that it is virtualized to take advantage of the functions.
- Operating systems require extensions to make API calls to the hypervisor.

2e) How could you use virtualization to increase world wide compute resources?

Virtualize all hardware.

3) Containers, Orchestration, and Deployment

3a) What are microservices?

Architectural pattern and organizational approach to software development where software is composed of small independent services that communicate over well-defined APIs.

3b) What are 3 key advantages, and 3 key disadvantages of microservice architectures?

Advantages: Scalability, code modularity and reuse, ability to leverage existing solutions.

Disadvantages: More complex development and deployment environment, lack of appropriate skills, response latency.

3c) Clearly and succinctly differentiate between load balancing and autoscaling.

- Load balancing distributes compute load across multiple nodes.
- Autoscaling increases or decreases available compute nodes.

3d) Clearly and succinctly differentiate between Docker Compose and Kubernetes. In your explanation provide a suitable use case for each technology.

Docker Compose deploys multi-container Docker apps to a single server, while Kubernetes is a production-grade container orchestrator that can run multiple container runtimes, including Docker's, across multiple virtual or physical machines.

3e) Identify 3 disadvantages of using docker.p

- Docker is not good for application that requires rich GUI.
- It is difficult to manage large amount of containers.
- Docker does not provide cross-platform compatibility means if an application is designed to run in a Docker container on windows, then it cannot run on Linux Docker container.

4. Large Distributed Key Value Storage Systems Part 1

4a) NoSQL – Key Value Data stores can be effective solutions for: (*select all that apply*).

A. Document indexing

B. Scalability

C. ACID

D. Flexible data model

A,B, D

4b) Consistent hashing is a strategy for dividing up keys/data between multiple machines. How does consistent hashing differ from traditional hashing techniques?

Answer: Traditional hashing techniques use a hash function to map a key to a specific node in a cluster. However, this can lead to imbalanced distribution of data when nodes are added or removed from the cluster. Consistent hashing, on the other hand, maps keys to a virtual ring, allowing for a more even distribution of data across nodes. When a node is added or removed, only a small portion of the data needs to be remapped, resulting in less disruption to the system.

4c) Vector Clocks. Which event is concurrent with vector clock (2, 8, 4)

- A. (3, 9, 5)
- B. (3, 8, 4)
- C. (1, 7, 3)
- D. (4, 8, 2)

(4, 8, 2)

4d) Cap Theorem. Is it possible to have a distributed system with high availability while still providing strong consistency guarantees? Justify your answer

No, it is not possible. The CAP theorem states that it is impossible for a distributed system to provide all three of the following guarantees:

- Consistency: All nodes in the system see the same data at the same time.
- Availability: All nodes in the system can be reached and data can be read or written at any time.
- Partition tolerance: The system can continue to operate even if some nodes are unavailable.

4e) Consistent hashing is a strategy for dividing up keys/data between multiple machines. How does consistent hashing differ from traditional hashing techniques?

Answer: Traditional hashing techniques use a hash function to map a key to a specific node in a cluster. However, this can lead to imbalanced distribution of data when nodes are added or removed from the cluster. Consistent hashing, on the other hand, maps keys to a virtual ring, allowing for a more even distribution of data across nodes. When a node is added or removed, only a small portion of the data needs to be remapped, resulting in less disruption to the system.