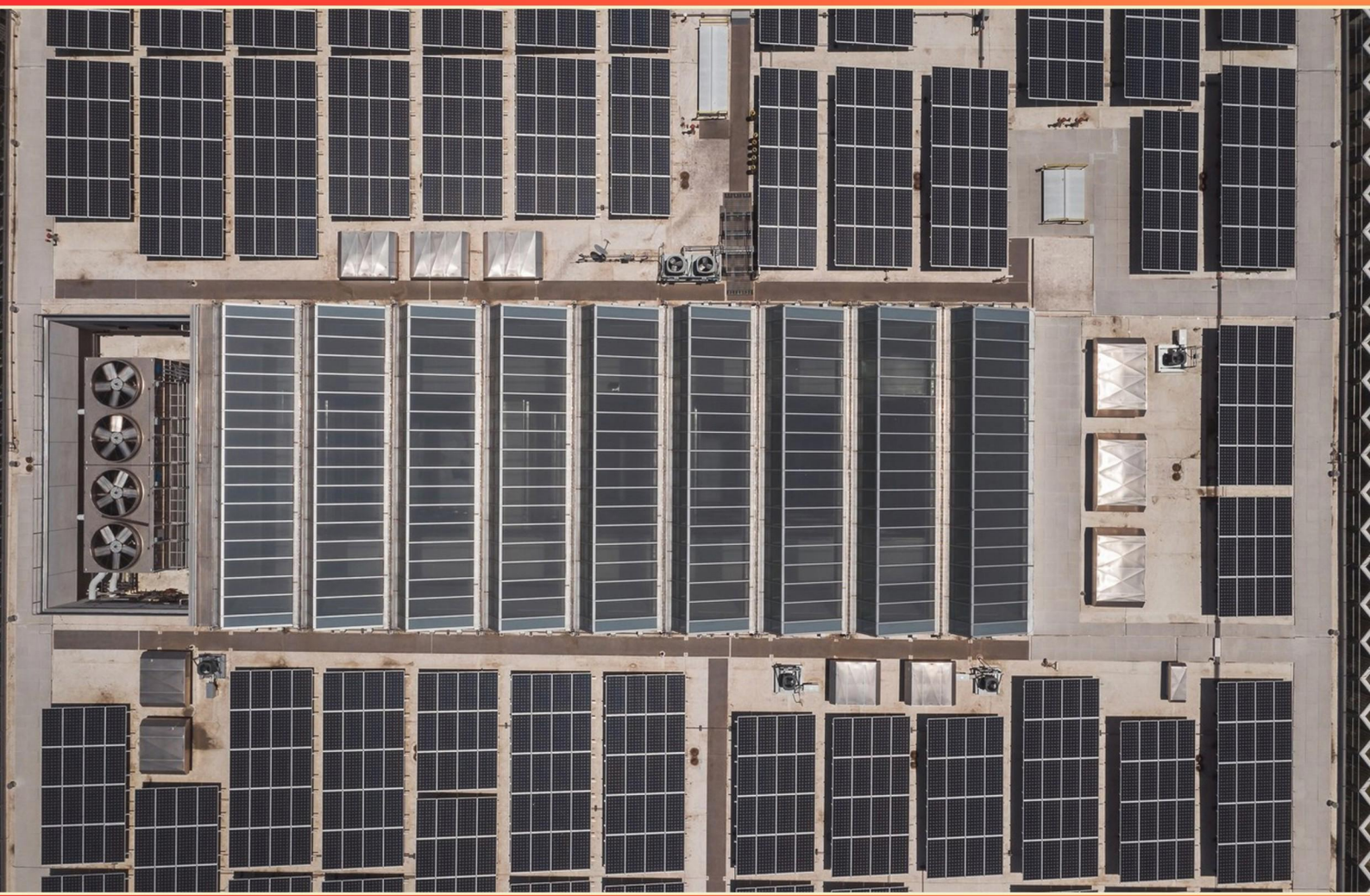
HPC BIG DATA CERTIFICATION PRACTICE TEST (SAMPLE)

STUDY GUIDE



SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

VISIT US ONLINE FOR THE FULL VERSION STUDY GUIDE

https://hpcbigdatacert.examzify.com

Copyright © 2026 by Examzify - A Kaluba Technologies Inc. product.

ALL RIGHTS RESERVED.

No part of this book may be reproduced or transferred in any form or by any means, graphic, electronic, or mechanical, including photocopying, recording, web distribution, taping, or by any information storage retrieval system, without the written permission of the author.

Notice: Examzify makes every reasonable effort to obtain accurate, complete, and timely information about this product from reliable sources.

SAMPLE

Table of Contents

Copyright 1

Table of Contents 2

Introduction 3

How to Use This Guide 4

Questions 5

Answers 8

Explanations 10

Next Steps 16

SAMPLE

Introduction

Preparing for a certification exam can feel overwhelming, but with the right tools, it becomes an opportunity to build confidence, sharpen your skills, and move one step closer to your goals. At Examzify, we believe that effective exam preparation isn't just about memorization, it's about understanding the material, identifying knowledge gaps, and building the test-taking strategies that lead to success.

This guide was designed to help you do exactly that.

Whether you're preparing for a licensing exam, professional certification, or entry-level qualification, this book offers structured practice to reinforce key concepts. You'll find a wide range of multiple-choice questions, each followed by clear explanations to help you understand not just the right answer, but why it's correct.

The content in this guide is based on real-world exam objectives and aligned with the types of questions and topics commonly found on official tests. It's ideal for learners who want to:

- Practice answering questions under realistic conditions,
- Improve accuracy and speed,
- Review explanations to strengthen weak areas, and
- Approach the exam with greater confidence.

We recommend using this book not as a stand-alone study tool, but alongside other resources like flashcards, textbooks, or hands-on training. For best results, we recommend working through each question, reflecting on the explanation provided, and revisiting the topics that challenge you most.

Remember: successful test preparation isn't about getting every question right the first time, it's about learning from your mistakes and improving over time. Stay focused, trust the process, and know that every page you turn brings you closer to success.

Let's begin.

SAMPLE

How to Use This Guide

This guide is designed to help you study more effectively and approach your exam with confidence. Whether you're reviewing for the first time or doing a final refresh, here's how to get the most out of your Examzify study guide:

1. Start with a Diagnostic Review

Skim through the questions to get a sense of what you know and what you need to focus on. Your goal is to identify knowledge gaps early.

2. Study in Short, Focused Sessions

Break your study time into manageable blocks (e.g. 30 – 45 minutes). Review a handful of questions, reflect on the explanations.

3. Learn from the Explanations

After answering a question, always read the explanation, even if you got it right. It reinforces key points, corrects misunderstandings, and teaches subtle distinctions between similar answers.

4. Track Your Progress

Use bookmarks or notes (if reading digitally) to mark difficult questions. Revisit these regularly and track improvements over time.

5. Simulate the Real Exam

Once you're comfortable, try taking a full set of questions without pausing. Set a timer and simulate test-day conditions to build confidence and time management skills.

6. Repeat and Review

Don't just study once, repetition builds retention. Re-attempt questions after a few days and revisit explanations to reinforce learning. Pair this guide with other Examzify tools like flashcards, and digital practice tests to strengthen your preparation across formats.

There's no single right way to study, but consistent, thoughtful effort always wins. Use this guide flexibly, adapt the tips above to fit your pace and learning style. You've got this!

SAMPLE

Questions

SAMPLE

1. What is meant by SSL handling in the context of a load balancer?

A. Transferring SSL certificates between servers

B. Encrypting data for secure transmission

C. Managing SSL traffic efficiently

D. All of the above

2. Which characteristic of Apache HBase allows for efficient scaling?

A. Database sharding

B. Data encryption

C. Compression techniques

D. Static memory allocation

3. Why are instance configurations beneficial in cloud deployment?

A. They allow for random instance deployment

B. They help standardize instance templates for easy deployment

C. They decrease the overall data storage requirement

D. They enhance network security

4. Which of the following has the highest price per core?

A. BM.Standard.2.52

B. BM.GPU2.2

C. BM.GPU3.8

D. BM.HPC2.36

5. In high throughput workloads, how important is latency?

A. Very important

B. Not important

C. Somewhat important

D. Depends on the node

6. What type of database is Apache HBase?

A. A relational database

B. A non-relational, distributed database

C. A time-series database

D. A file-based database

SAMPLE

7. What test measures throughput, IOPS, and Metadata performance for parallel file systems?

A. mdtest

B. IO500

C. Vdbench

D. IOzone

8. What is the storage capacity of the standard object storage in HPC?

A. Terabytes

B. Petabytes

C. Gigabytes

D. Megabytes

9. What is the network configuration for BM.HPC2.36?

A. TCP (2x25GBs)

B. TCP (25GBs); RDMA(100GBs)

C. TCP (100GBs)

D. TCP (10GBs); RDMA(50GBs)

10. What type of operations is TeraGen heavily dependent on?

A. Read intensive

B. Write intensive

C. Compute intensive

D. Network intensive

SAMPLE

Answers

SAMPLE

1. D
2. A
3. B
4. C
5. B
6. B
7. B
8. B
9. B
10. B

SAMPLE

Explanations

SAMPLE

1. What is meant by SSL handling in the context of a load balancer?

A. Transferring SSL certificates between servers

B. Encrypting data for secure transmission

C. Managing SSL traffic efficiently

D. All of the above

In the context of a load balancer, SSL handling involves several critical functions that ensure secure and efficient management of data traffic across servers. These functions include transferring SSL certificates between servers, encrypting data for secure transmission, and managing SSL traffic efficiently. When a load balancer handles SSL, it takes on the task of managing SSL certificates, which are essential for establishing secure connections between clients and servers. Load balancers can offload the SSL processing from individual servers, thereby simplifying the management of SSL certificates and improving performance. Additionally, the encryption of data for secure transmission is vital for protecting sensitive information as it travels over the network. By handling this encryption at the load balancer level, organizations can ensure that data remains secure while optimizing server resources. Finally, managing SSL traffic efficiently involves routing incoming SSL requests to the appropriate servers, balancing the load across multiple backend servers, and mitigating potential security threats. This efficiency not only maximizes server performance but also enhances the user experience by providing reliable access to services. Thus, SSL handling by a load balancer is a multi-faceted process that includes all the elements mentioned, making the comprehensive option the most accurate choice.

2. Which characteristic of Apache HBase allows for efficient scaling?

A. Database sharding

B. Data encryption

C. Compression techniques

D. Static memory allocation

Apache HBase's ability to efficiently scale is primarily due to its characteristic of database sharding. Sharding, also known as partitioning, involves splitting a large database into smaller, more manageable pieces called shards. Each shard can be hosted on different servers, which allows HBase to distribute the load across multiple nodes. This distribution is crucial for handling large volumes of data and for serving a high number of concurrent read and write requests. When data is sharded, it allows for horizontal scaling, meaning that as data grows, additional servers can be added to accommodate the increased load without significant reconfiguration. This capability helps maintain performance levels as the system scales, making it an essential feature for systems dealing with big data, where datasets can grow significantly. In contrast, the other options do not contribute to scaling in the same manner. Data encryption is important for security but does not directly impact the scaling capabilities. Compression techniques can reduce the size of data, which can help with storage efficiency but does not inherently allow for better scaling in terms of handling increased workloads. Static memory allocation is not conducive to dynamic scaling and can lead to inefficiencies in resource utilization as the demand fluctuates. Thus, sharding stands out as the key feature that facilitates effective scaling in HBase

SAMPLE

3. Why are instance configurations beneficial in cloud deployment?

A. They allow for random instance deployment

B. They help standardize instance templates for easy deployment

C. They decrease the overall data storage requirement

D. They enhance network security

Instance configurations are beneficial in cloud deployment primarily because they help standardize instance templates for easy deployment. By using predefined configurations, organizations can ensure that every instance is created with consistent settings, such as instance types, disk configurations, and security settings. This standardization simplifies the processes of scaling, updating, and managing instances, leading to more efficient use of resources and reduced potential for error. In addition to promoting efficiency, standardized instance configurations make it much easier to replicate environments across different stages of development, from testing to production. This uniformity can greatly streamline workflows and enhance collaboration among team members, as everyone can rely on the same setup. Variability in instance deployment—such as random configurations—could lead to complexities and inconsistencies that complicate scalability and management. While decreasing overall data storage requirements and enhancing network security are important aspects of cloud deployment, they are not the primary benefits of instance configurations.

4. Which of the following has the highest price per core?

A. BM.Standard.2.52

B. BM.GPU2.2

C. BM.GPU3.8

D. BM.HPC2.36

The choice of BM.GPU3.8 having the highest price per core is based on its configuration, which typically includes a powerful GPU architecture designed for high-performance computing tasks. This model is engineered for demanding applications that require not just raw processing power but also efficiency in handling complex computations. The GPU capabilities imply that it can perform parallel processing much more effectively than standard CPU-based configurations. In high-performance computing (HPC) environments, the inclusion of GPUs significantly influences pricing due to their advanced technology and capability to execute operations much faster than traditional CPU cores. As such, hardware with GPUs—especially more advanced versions like those found in the BM.GPU3.8—often costs more per core than standard processors or other types of node configurations. While other options might include powerful configurations as well, they don't combine the same level of advanced GPU integration that enhances performance and justifies a higher price point. This makes BM.GPU3.8 stand out in terms of price per core in the context of high-performance workloads.

SAMPLE

5. In high throughput workloads, how important is latency?

A. Very important

B. Not important

C. Somewhat important

D. Depends on the node

In the context of high throughput workloads, latency is generally considered less critical compared to other performance metrics such as throughput itself. High throughput workloads prioritize the volume of processed operations or data over the speed of response time for individual requests. This means that the system can tolerate longer wait times for individual tasks as long as it is capable of handling and processing large amounts of data efficiently. For example, in environments like batch processing or data analysis where tasks are accumulated and processed in chunks, the overall performance is measured by how many tasks can be completed in a given timeframe rather than how quickly each individual task is completed. Therefore, while there may be a certain level of latency that can affect system performance, in high throughput scenarios, minimizing this latency is typically not the main focus. In contrast, other scenarios such as interactive applications or real-time data processing place a much higher premium on low latency, which is why it is essential to consider the specific use case when evaluating the importance of latency in different workloads. Understanding this distinction clarifies why latency may be categorized as "not important" in high throughput settings.

6. What type of database is Apache HBase?

A. A relational database

B. A non-relational, distributed database

C. A time-series database

D. A file-based database

Apache HBase is recognized as a non-relational, distributed database specifically designed to handle large amounts of sparse data in a scalable manner. It is built on top of the Hadoop File System (HDFS) and provides real-time read and write access to big data. Unlike traditional relational databases that require a fixed schema and use SQL for queries, HBase employs a column-oriented storage format and supports flexible schema evolution, making it well-suited for big data applications where data structures can change over time. The non-relational aspect allows HBase to scale horizontally by distributing data across multiple nodes, ensuring high availability and fault tolerance. This design choice enables it to manage massive volumes of data efficiently, which is crucial in big data environments where datasets can grow dynamically. In contrast, the other options present various database types that do not match HBase's characteristics. Relational databases depend on fixed schemas and SQL, while time-series databases specialize in storing temporal data. Meanwhile, file-based databases typically manage data in a flat file format, lacking the robust scalability features and architecture of a distributed system like HBase.

SAMPLE

7. What test measures throughput, IOPS, and Metadata performance for parallel file systems?

A. mdtest

B. IO500

C. Vdbench

D. IOzone

The IO500 benchmark is specifically designed to evaluate the performance of parallel file systems, focusing on three key performance metrics: throughput, Input/Output Operations Per Second (IOPS), and metadata performance. This comprehensive testing framework is vital in assessing how well a parallel file system can handle both large data transfers (throughput) and a high volume of small requests (IOPS), in addition to managing metadata operations, which are crucial for file system navigation and organization. By using IO500, researchers and system administrators can gather a holistic view of a parallel file system's capabilities, enabling them to compare different systems effectively and identify areas for optimization. This makes IO500 an essential tool in high-performance computing environments where efficiency and speed in handling data are paramount. Other tools like mdtest, Vdbench, and IOzone focus on specific aspects or types of workloads but do not provide the comprehensive evaluation across all three metrics in the context of parallel file systems that IO500 does.

8. What is the storage capacity of the standard object storage in HPC?

A. Terabytes

B. Petabytes

C. Gigabytes

D. Megabytes

The storage capacity of standard object storage in High-Performance Computing (HPC) is indeed measured in petabytes. This is primarily because HPC environments often deal with vast amounts of data generated from simulations, analyses, and experiments, thus requiring storage solutions that can accommodate tremendous volumes. Object storage is designed to manage large amounts of unstructured data efficiently, which is a common scenario in HPC. It provides scalability, durability, and accessibility for big data applications, enabling organizations to store not just large files, but also a wide array of data types. Petabytes refer to 1,024 terabytes or about a million gigabytes, representing a substantial leap in storage capability compared to the other options listed. Terabytes, gigabytes, and megabytes are smaller units and are generally not sufficient for the data-intensive workloads typically encountered in HPC settings. Thus, choosing petabytes reflects the advanced capacity requirements of high-performance computing solutions for big data applications.

SAMPLE

9. What is the network configuration for BM.HPC2.36?

A. TCP (2x25GBs)

B. TCP (25GBs); RDMA(100GBs)

C. TCP (100GBs)

D. TCP (10GBs); RDMA(50GBs)

The network configuration for BM.HPC2.36 is indeed TCP (25GBs); RDMA(100GBs). This configuration is suitable for high-performance computing (HPC) environments where low latency and high throughput are essential. In this setup, TCP provides a reliable transport protocol for regular data transfer, while the RDMA (Remote Direct Memory Access) setup allows for direct memory access from one computer to another without involving the operating system, significantly lowering latency. The combination of these two protocols ensures that the system can handle a variety of tasks efficiently, optimizing performance for data-intensive applications. Having a TCP connection at 25GBs alongside a more robust RDMA capability at 100GBs means that while the system can manage standard TCP/IP traffic and provide reliable communication, it can also leverage RDMA for tasks that require higher throughput and lower latency, particularly beneficial in an HPC context where rapid data transfer can be a critical factor for performance. This dual-configuration allows for flexibility and adaptability in handling different workloads, making it ideal for HPC use cases.

10. What type of operations is TeraGen heavily dependent on?

A. Read intensive

B. Write intensive

C. Compute intensive

D. Network intensive

TeraGen is primarily focused on generating large datasets, which involves creating and writing substantial volumes of data. Therefore, it is heavily dependent on write-intensive operations, as the process of generating data entails continuously writing this newly generated information to storage systems. This characteristic distinguishes TeraGen from other processes that may emphasize reading from existing datasets or performing computations on existing data. Write-intensive operations are indicative of systems that must efficiently handle high rates of data insertion, making them critical when evaluating tools meant for data generation like TeraGen. Understanding the nature of write operations is crucial in optimizing performance for tasks relating to data handling, especially in high-performance computing and big data contexts, where the volume of data created can be substantial.

SAMPLE

Next Steps

Congratulations on reaching the final section of this guide. You've taken a meaningful step toward passing your certification exam and advancing your career.

As you continue preparing, remember that consistent practice, review, and self-reflection are key to success. Make time to revisit difficult topics, simulate exam conditions, and track your progress along the way.

If you need help, have suggestions, or want to share feedback, we'd love to hear from you. Reach out to our team at hello@examzify.com.

Or visit your dedicated course page for more study tools and resources:

https://hpcbigdatacert.examzify.com

We wish you the very best on your exam journey. You've got this!

SAMPLE