

Optimizing System Resources for Scalability

Part 1: Office Roleplay Dialogue

Scenario: A DevOps Engineer, Hiroshi, is discussing system optimization and scalability strategies with his colleague, Aisha.

Aisha: Hiroshi, we need to improve the system's **scalability** before the next product launch. Have you looked into our current capacity?

Hiroshi: Yes, I checked the servers. Right now, we're using **vertical scaling**, but it might not be enough if traffic spikes.

Aisha: I see. Should we consider **horizontal scaling** instead? Adding more instances might help distribute the load.

Hiroshi: That's a good idea. We can also implement **auto-scaling** to adjust resources automatically based on demand.

Aisha: Sounds like a plan. How about **load balancing**? Is it configured properly for traffic distribution?

Hiroshi: Not yet. I'll optimize the **load balancing** to ensure no single server is overwhelmed.

Aisha: Great. That should prevent downtime and keep performance stable. How do you want to test it?

Hiroshi: We can simulate high traffic and monitor the response times. If the system struggles, we'll tweak the scaling rules.

Aisha: Perfect. Let's schedule a test and adjust the **auto-scaling** thresholds accordingly.

Hiroshi: Agreed. Once we fine-tune it, the system should be able to handle sudden spikes smoothly.

Part 2: Comprehension Questions

1. What issue is Aisha concerned about?

- (A) Server security
- (B) System scalability
- (C) Software bugs
- (D) Network speed

2. What type of scaling is the company currently using?

- (A) Horizontal scaling
- (B) No scaling at all
- (C) Vertical scaling
- (D) Manual scaling

3. Why does Hiroshi suggest auto-scaling?

- (A) To reduce the cost of physical hardware
- (B) To automatically adjust resources based on demand
- (C) To replace the current system entirely
- (D) To improve security protocols

4. What does Hiroshi plan to do to test the system?

- (A) Shut down all servers
 - (B) Reduce the number of users
 - (C) Simulate high traffic and monitor response times
 - (D) Ignore the problem and wait for real traffic
-

Part 3: Key Vocabulary Definitions in Japanese

1. **Scalability (スケーラビリティ)** – システムの負荷増加に対応できる能力。
 2. **Load balancing (ロードバランシング)** – サーバー間でトラフィックを分散する技術。
 3. **Horizontal scaling (水平スケーリング)** – 複数のサーバーを追加して処理能力を向上させる方法。
 4. **Vertical scaling (垂直スケーリング)** – 既存のサーバーのリソースを増強する方法。
 5. **Auto-scaling (オートスケーリング)** – システムの負荷に応じてリソースを自動的に調整する機能。
-

Part 4: Questions & Correct Answers

1. **What issue is Aisha concerned about?**
☒ (B) System scalability
2. **What type of scaling is the company currently using?**
☒ (C) Vertical scaling
3. **Why does Hiroshi suggest auto-scaling?**
☒ (B) To automatically adjust resources based on demand

4. What does Hiroshi plan to do to test the system?

☒ (C) Simulate high traffic and monitor response times