Post mortem: Web Application Outage Incident

Issue Summary:

- Duration: March 10, 2024, 13:45 UTC to March 11, 2024, 02:30 UTC

- Impact: The outage affected approximately 30% of our users, resulting in degraded performance and intermittent service disruptions.

- Root Cause: A misconfiguration in the load balancer settings caused excessive traffic redirection to a single server, overwhelming its capacity.

Timeline:

- 13:45 UTC: Issue detected through monitoring alerts indicating increased latency and error rates.

- 13:50 UTC: Engineering team notified of the incident.

- 13:55 UTC: Initial investigation focused on backend services and database health.

- 14:20 UTC: Misleading assumption made about potential database overload due to recent schema changes.

- 14:45 UTC: Incident escalated to senior engineering management as the problem persisted.

- 15:30 UTC: Load balancer configurations reviewed, but no abnormalities initially found.

- 16:00 UTC: Further investigation revealed misconfigured settings causing traffic imbalance.

- 17:15 UTC: Immediate action taken to correct load balancer settings and redistribute traffic.

- 02:30 UTC: Service fully restored as traffic stabilized and performance metrics returned to normal.

Root Cause and Resolution:

- Root Cause: The misconfiguration in the load balancer caused it to favor one server over others, leading to an overload situation.

- Resolution: Load balancer settings were adjusted to evenly distribute traffic across available servers, resolving the imbalance and restoring normal operation.

Corrective and Preventative Measures:

- Improvements/Fixes:

1. Implement automated checks for load balancer configurations to detect abnormalities promptly.

2. Enhance monitoring systems to provide more granular insights into traffic distribution and server health.

3. Establish regular audits of critical infrastructure settings to prevent similar misconfigurations.

- Tasks to Address the Issue:

1. Develop and deploy scripts for automated load balancer configuration validation.

2. Enhance monitoring dashboard to display real-time traffic distribution across servers.

3. Schedule recurring reviews of load balancer settings to ensure alignment with traffic patterns and server capacities.

This incident exposed vulnerabilities in my infrastructure management processes, particularly in the area of load balancer configuration. By implementing the outlined corrective measures and addressing the identified tasks, I aim to bolster our system's resilience and minimize the risk of similar outages in the future.