

SRE Implementation for Global E-Commerce Platform

Utilizing a hybrid SRE model and data-driven automation to handle massive traffic spikes while achieving significant annual cost savings.

Overview

QBurst implemented a hybrid Site Reliability Engineering (SRE) model focusing on proactive observability, data-driven scaling, and automation to stabilize a global e-commerce platform. It established shared ownership of reliability and performance.

- Over \$1.5 million saved annually across 10 regions through automated, predictive scaling.
- Achieved 99.999% uptime goals while significantly improving site responsiveness and checkout success rates.



Client Profile

Based in Asia, the client is one of the world's largest apparel retailers, operating a massive manufacturing and sales network across 2,500+ stores. Their global e-commerce presence requires extreme reliability to support millions of customers across diverse overseas markets.

Challenges: High-Stakes Traffic and Reliability Gaps

Seasonal surges like Black Friday created immense pressure on the infrastructure, leading to unsustainable costs and performance bottlenecks.

- Cost-prohibitive 24/7 manual scaling was used to prevent downtime during unpredictable traffic surges.
- Latency issues caused immediate cart abandonment, as even minor half-second delays eroded customer trust and revenue.
- Traditional reactive IT silos resulted in a lack of proactive ownership regarding system reliability and performance benchmarks.

- Achieving a "five-nines" (99.999%) uptime goal while maintaining cost efficiency was a significant technical and operational hurdle.

QBurst Solution: Hybrid SRE and Observability Framework

We selected a hybrid SRE model, embedding developer representatives within a central SRE team to share ownership of features and reliability. The solution utilized a robust observability framework with Grafana and New Relic to track KPIs such as RDS utilization, error rates, and container performance.

- **Observability & Log Analysis:** Built a foundation for reliability by tracking traffic patterns and conducting frequent log scans with Splunk and Datadog to proactively address anomalies.
- **Predictive, Data-Driven Scaling:** Analyzed traffic on daily and weekly scales to implement automated scaling via Jenkins and Terraform, scaling down during low activity and up before peak surges.
- **Automation & ASG Optimization:** Reduced manual effort and human error by automating infrastructure tasks, maintaining high maximum container limits for rapid emergency scaling.
- **Resilient Incident Response:** Established a structured framework for swift failure mitigation, supported by detailed manuals, escalation protocols, and regular mock drills to improve team coordination.
- **No-Blame Root Cause Analysis (RCA):** Adopted a "Five Whys" methodology and structured documentation to ensure continuous learning and prevent recurrence without fostering a culture of blame.
- **Continuous Performance Optimization:** Utilized Gatling for proactive load testing and optimized slow database queries to refine indexing and application logic efficiency.

Technical Highlights

- **Hybrid Team Integration:** Merged developers and SREs into a unified workflow for collaborative reliability management.
- **Predictive Scaling Logic:** Leveraged historical traffic data to automate capacity planning and cost control.
- **Automation Suite:** Integrated Jenkins, Maven, and Terraform to enable error-free, rapid infrastructure deployments.
- **Observability Stack:** Deployed a comprehensive suite including AppDynamics, Dynatrace, and ELK for 360-degree system visibility.

Impact: Performance Excellence and Cost Leadership

- **Substantial Annual Savings:** Automated scaling reduced costs by \$10K–\$13K per region, totaling over \$1.5 million in annual savings across 10 regions.
- **Peak Event Success:** Saved \$45K–\$50K during Black Friday alone compared to previous manual scaling years.
- **Improved Performance:** Achieved up to 60% faster loading times and maintained a 4.3+ rating on global app stores.
- **Enhanced Resilience:** Drastically reduced downtime through a proactive incident management flow and structured RCA processes.