8 west
CONSULTING

APRIL / 2025

WELL ARCHITECTED PERFORMANCE

CONFIDENTIAL

All Rights Reserved. This material contains the valuable properties and trade secrets of 8 West Consulting. No part of this material may be reproduced or transmitted in any form or by any means, electronic, mechanical or otherwise, including photocopying and recording, or in connection with any information storage or retrieval system without the permission of a Director of 8 West Consulting.

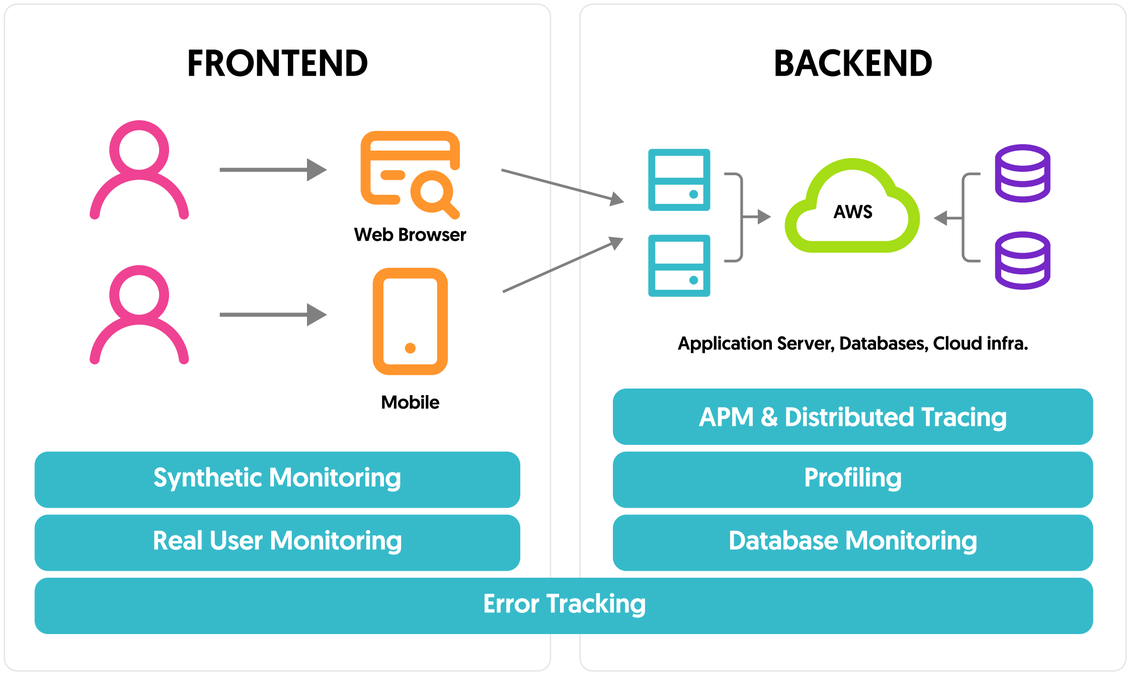
INFORMATION SECURITY
ISO/IEC 27001:2022
NSAI Certified

Performance Management: Efficiency and Beyond

Performance management covers all aspects of system performance, from client-side responsiveness and user experience optimization to load testing and capacity planning.

While Performance Efficiency, a pillar of the AWS Well-Architected Framework, focuses on efficient cloud resource use, it is just one part of a broader approach that ensures scalability, responsiveness, and system optimization as demands and technologies evolve.

Monitoring the application, end-to-end



The Aim of Performance Management

DevOps Goals

- Faster, better product delivery
- Effective collaboration between teams
- Faster feedback
- Quick resolution of issues
- Innovative mindsets

Migrating to the cloud provides flexibility and benefits like auto-scaling to handle capacity, but it doesn't ensure optimized performance or eliminate potential bottlenecks.

Various components within the pipeline can experience both catastrophic failures and subtle performance issues. For example, during Azure migration load testing, one client found a suboptimal Redis caching setup, causing high response times and limited user capacity. Fixing this before production improved response times by 80% and user capacity by 400% without extra resources!

Typical Performance Challenges

1. Architecture selection
2. Compute and hardware decisions
3. Data management
4. Networking and content delivery
5. User experience
6. Process and culture



1. Architecture Selection

To enhance architectural decisions and boost performance efficiency, continuously explore available cloud services and configurations.

Guidance

Make use of guidance from your cloud provider or an appropriate partner to learn about architecture patterns and best practices.

Cost

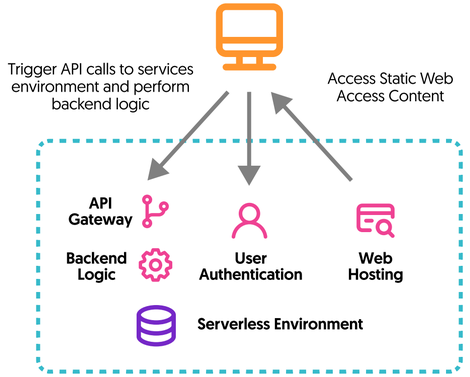
Factor cost into architectural decisions and evaluate how trade-offs impact customers and architecture efficiency.

Benchmarking

Use a data-driven approach to drive architectural decisions.

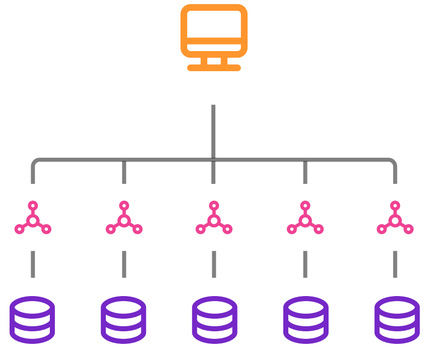
For example...

SERVERLESS ARCHITECTURE



Ideal for applications with variable workloads and a focus on reducing operational complexity.

MICROSERVICES ARCHITECTURE



Great for complex, large-scale applications needing high flexibility and resilience

2. Compute and Hardware Decisions

Choose the most suitable compute options for your workload to enhance performance, cut costs, and reduce operational efforts.

Knowledge Base

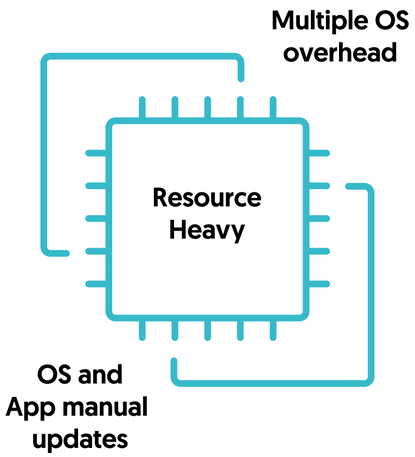
Understand the available compute features, collect compute-related metrics.

Optimal Configuration

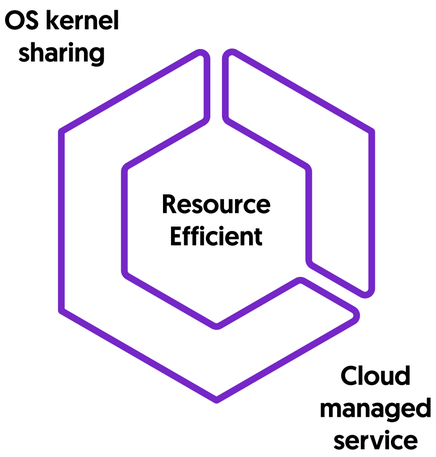
- Configure and right-size compute resources.
- Consider scaling compute resources dynamically.
- Use optimized hardware-based compute accelerators.

For example...

VIRTUAL MACHINES



CONTAINERS



3. Data Management

To effectively support your data access and storage needs, choose purpose-built data stores by understanding key characteristics like shareability, size, cache size, access patterns, latency, throughput, and persistence.

Measure

Track performance metrics to optimize your data store, ensure it meets workload requirements, and gain a comprehensive view of its performance.

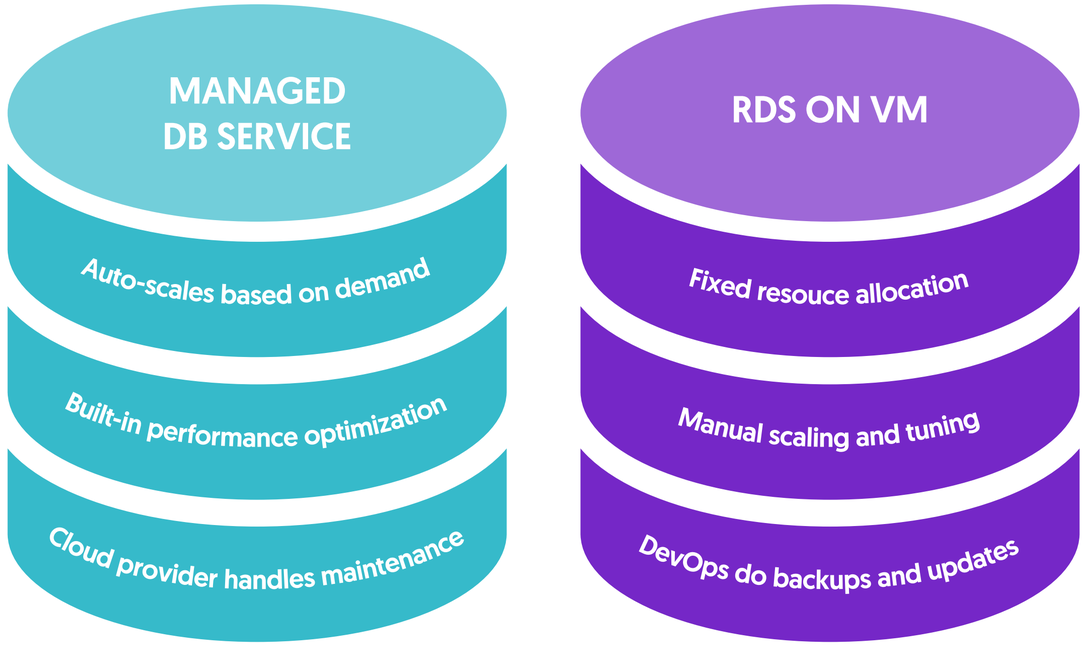
Evaluate

Understand the configuration options available for your data stores to maximize storage space and performance.

Implement

Strategies to enhance query performance and utilize caching for quick retrieval of frequently accessed data.

For example...



4. Networking and Content Delivery

Analyse how network decisions impact your workload and evaluate cloud networking features using collected data to optimize your network configuration for the best results.

Provisioning

Allocate adequate bandwidth and estimate requirements to meet performance needs.

Distribution

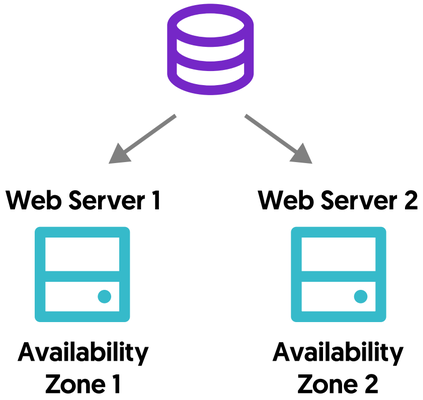
Using load balancing to leverage cloud elasticity and manage traffic effectively.

Efficiency

Choose network protocols and resource locations to reduce latency and improve throughput.

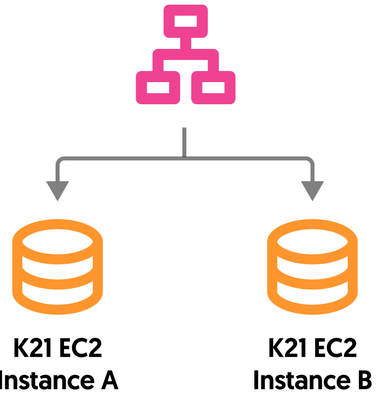
For example...

NETWORK ELASTIC LOAD BALANCER



Ideal for applications requiring high performance and low latency at the transport layer.

APPLICATION LOAD BALANCER



Suitable for HTTP/HTTPS traffic and applications requiring advanced routing features.

5. User Experience

Responsive and fast-loading applications are crucial for an exceptional user experience, they ensure that users can quickly access content and interact with features without frustrating delays.

In essence, the speed and responsiveness of an application directly impact its usability and the overall perception of the brand.

SEO is closely linked with front end application performance, so much so that Google have invested heavily in encouraging company's to focus on core web vitals that quantify interactivity and perceived load times.

Possible impacts of no focus on front end metrics and the UX:

Performance Issues: Undetected performance issues, such as slow page load times during peak periods.

Customer Dissatisfaction: Frustrated customers leading to abandoned sessions and decreased engagement.

Revenue Loss: Significant drop in revenue due to poor customer experience and reduced conversion rates.

6. Process and Culture

Establish key performance indicators (KPIs) to measure workload health and performance, identifying metrics that align with business goals.

Use monitoring solutions to pinpoint critical performance areas, and define a process to evaluate new services and configurations.

Regularly review and load test your workload to identify and address performance bottlenecks.

Utilize automation to proactively remediate issues, and keep your workload and services up-to-date with the latest cloud features.

Possible implications of not embedding performance into organizational process:

Undetected Bottlenecks: Critical performance bottlenecks go unnoticed, causing intermittent outages and slowdowns.

Inefficient Resource Use: Inefficient use of cloud resources and higher operational costs.

Unexpected Failures: Unprepared for traffic spikes, resulting in unexpected system failures during high-demand periods.

Reputation Damage: Frequent outages and poor performance damage the business's reputation, leading to a loss of customer trust and loyalty.

Delayed Issue Resolution: Longer downtime and slower recovery times due to manual intervention.

Security Vulnerabilities: Exposure to security vulnerabilities, increasing the risk of data breaches and compliance issues.

Common Performance KPIs

Identifying KPIs for workload health and performance aligns teams on priorities and defines successful outcomes.

Sharing these metrics across departments ensures visibility and alignment on expectations and business impact.

EXAMPLE KPIs	ASSOCIATED AREA	BENEFIT
% CPU, disk and memory utilization	Infrastructure	Efficient provisioning
Network latency and bandwidth	Networking and CDN	Fast delivery
Error rate	Application	Stability, security and consistent service
Response time	Application	Usability and efficient delivery
Throughput and capacity (request, user, transactions etc)	Multi-tier	Measure limitations
Cache hits	Application	Efficient delivery, resource optimization
Core Web vitals	User experience	User retention and SEO

Performance Efficiency Maturity Model

CRAWL



Basic Understanding: Teams have a fundamental understanding of performance efficiency principles.

Manual Processes: Performance tuning and optimization are done manually.

Limited Monitoring: Basic monitoring tools are in place, but they may not provide comprehensive insights.

Reactive Approach: Issues are addressed as they arise, often after they impact performance.

WALK



Intermediate Knowledge: Teams have a deeper understanding and start applying best practices.

Automated Processes: Some performance optimization processes are automated.

Enhanced Monitoring: More advanced monitoring tools are used, providing better insights and alerts.

Proactive Approach: Teams start to anticipate performance issues and address them before they impact users.

RUN



Expert Knowledge: Teams are experts in performance efficiency and continuously update their knowledge.

Governance: There are performance specialists in the organization responsible for maintaining quality, communicating targets and championing any wins

Proactive Remediation Automation: KPIs are combined with monitoring and alerting systems, to proactively address performance-related issues

Comprehensive Monitoring: Sophisticated monitoring and analytics tools provide real-time insights and predictive analytics.

Predictive Approach: Teams use data and analytics to predict and prevent performance issues, ensuring optimal performance at all times.

Performance Optimization in the SDLC - CI / CD

- **Monitor:** Continuously track application performance metrics to detect issues early.
- **Analyse:** Examine performance data to identify bottlenecks and root causes.
- **Prioritize:** Rank performance issues based on their impact and urgency.
- **Optimize:** Implement changes to improve application performance and efficiency.
- **Validate:** Test the application to ensure performance improvements are effective.
- **Release:** Deploy the optimized application to production, ensuring performance standards are met.



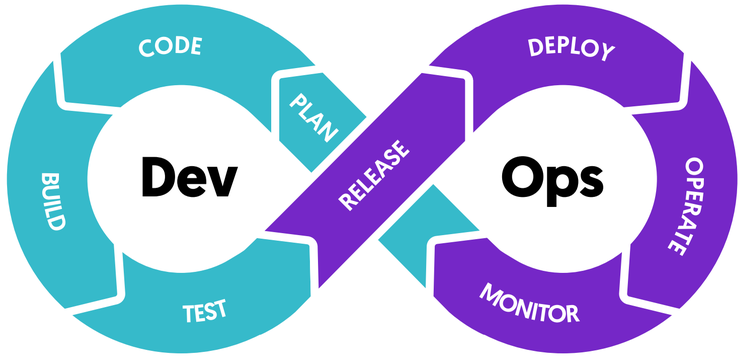
Performance Optimization in the SDLC - CI / CD CONT.

Integration Testing:
Ensuring different modules work together as expected.

API Testing:
Assessing API performance, reliability, and security.

CI/CD:
Running core tests automatically with each code change.

Load Generation:
Stress and soak testing the full system, multiple endpoints



Shift Left
Ensures software meets design

Shift Right
Ensures performance, resilience and reliability

Production Monitoring:
Continuously tracking live application performance.

User Feedback Collection:
Engaging with users to gather experience feedback.

A/B Testing: Comparing performance by splitting traffic between different versions.

Feature Toggling: Enabling or disabling features selectively in the live environment.

Solutions

Some more detail on how we do it...

Native tools such as Azure Monitor and AWS CloudWatch provide visibility on cloud infrastructure, but with more complex environments additional tooling is required for full end to end application transparency, observability and automating processes.

8West have experimented, licenced and partnered with some of the best providers to manage a suite of tools that cover browser monitoring, security and error checking, code level APM, Infrastructure monitoring and much more.

Monitoring Partners

DATADOG

SpeedCurve

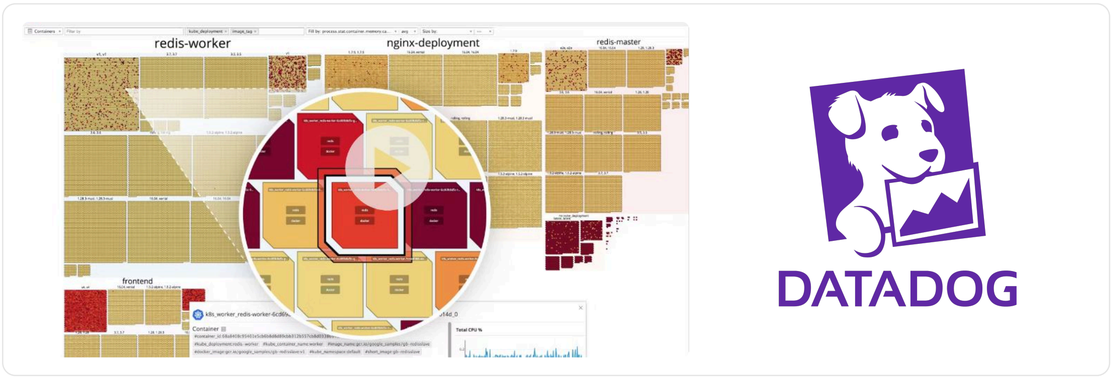
new relic

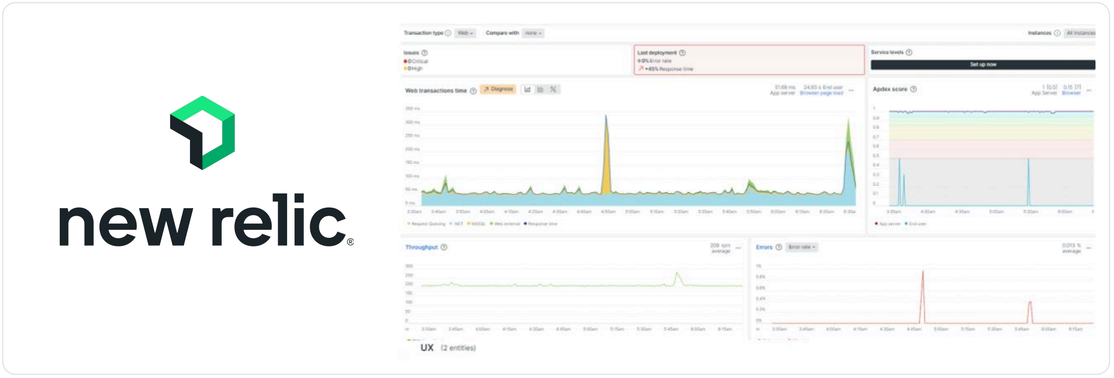
catchpoint

Full Stack Monitoring

Monitoring production environments and utilizing using automated alerting thresholds help react quickly to system failures with remedial actions, reducing stress on engineering teams. The goal of course is to never trigger such production alerts by performance testing the KPI's against pre-production environments.

Observability tools are essential for tracking how traffic impacts architecture across multiple tiers. Third-party vendors like **Datadog** and **New Relic** excel in diagnosing application performance issues. These tools trace requests through multiple tiers, facilitating the debugging of issues within the pipeline, from external services and databases to critical infrastructure.

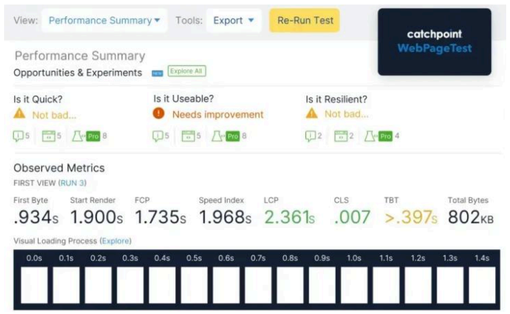




UX Monitoring and Front End Synthetics

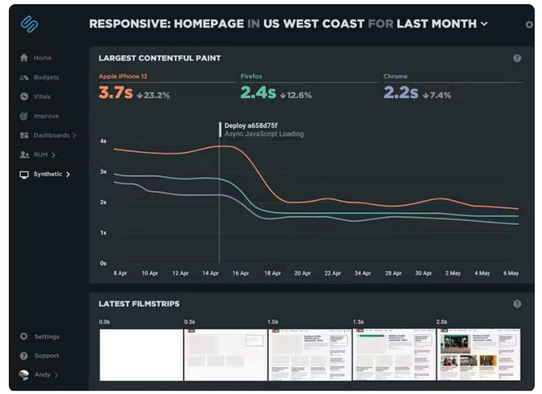
SpeedCurve is a synthetic and RUM tool that measures and analyses user centric metrics focusing on SEO optimization. This is driven is by Google and primarily concerned with core web vitals.

We use it to track UX for key pages with a history of up to a year. It can generate automated reports and threshold alerts and flags milestone moments such as core optimization initiatives and major releases so that we can keep tabs on user experience over time.





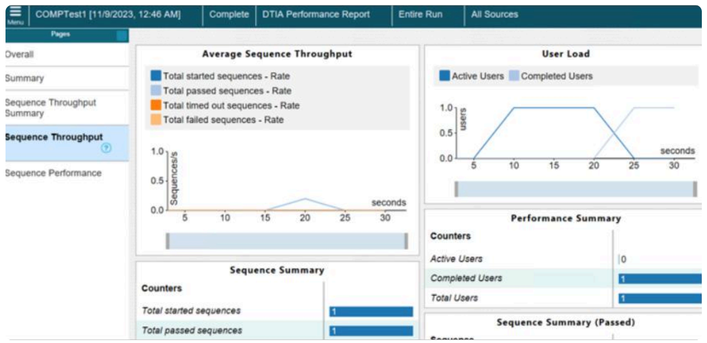




Load and API Performance Testing

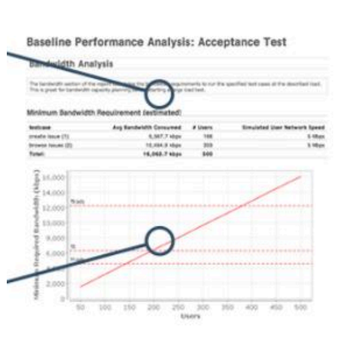
8West have hands on experience with a range of enterprise tools such as Akamai **CloudTest**, **HCL OneTest** and **WP load tester**. These tools can be expensive to licence sometimes running into six figures. There are more cost effective options such as Gatling, but they don't do as much of the heavy lifting so scripting and configuration time is greatly increased and the repository more difficult to maintain.

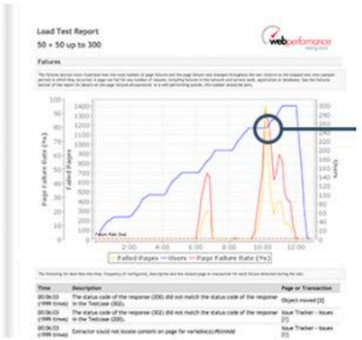
8West hold a loyalty licence for WP load tester which uses a combination of GUI configuration and code snippets to create and maintain scripts efficiently but with flexibility so we can create, execute and maintain the repo at no extra cost to the client.











Load and API Performance Testing CONT.

For any significant configuration or code change the goal is:

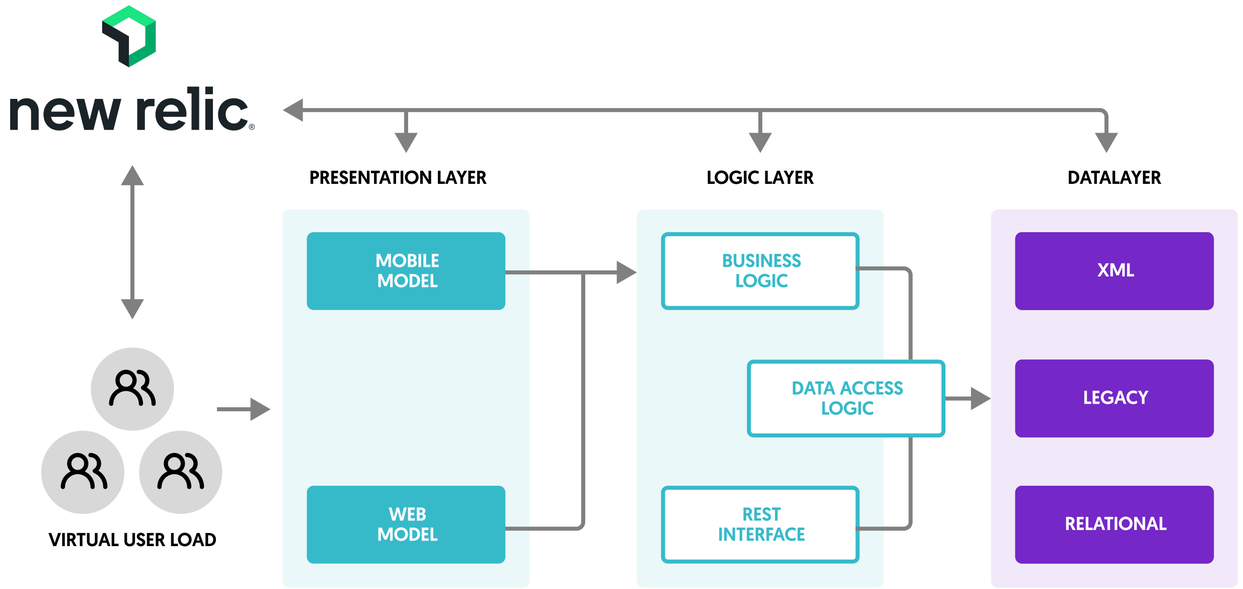
To identify any bottlenecks or slowdowns in the API request pipeline such as external service calls and DB transactions before it hits production.

Ensure the error rate for these APIs is acceptable

Verify autoscaling or test infrastructure capacity when implementing new configurations.

Sanity check the applications before anticipated high load events

Verify the application loads efficiently for the end user



Strategy

How 8 West Can Help

8 West have a cloud engineering team comprised of certified cloud architects, dev-ops specialists and performance engineers. We can provide the expertise to help ironclad your application.



- Dedicated specialists to engage with teams, drive collaboration and provide regular monitoring and reports
- Carefully curated tool suites where we have partnered with the best providers to deliver visibility and transparency
- Performance audits
- Integrated performance testing
- Management of automated processes and AI tooling that will save time and money
- The implementation of a roadmap for long-term success
- Incremental establishment of performance best practices

Step 1 - Monitor and Analyze

ENGAGEMENT

The first practical step we take is to install our partners premium insights tools (if none already installed) to gain centralized visibility into the application (New Relic / Datadog / SpeedCurve)

This will help with:

- Mapping current performance
- Resource utilization analysis
- UX recommendations and audit
- Transaction tracing
- Automation opportunities

**Note: the typical timeline for this step varies depending on the complexity of the architecture*

OUTPUT

- High level weekly report on cost optimization savings
- Monthly report on resource utilization, cost analysis, KPI targets and budget, architectural recommendations
- Management of the tool suite
- Managed system for tracking the ticketing and implementation of engineering improvements
- Systematic adoption of FinOps best practices as laid out in the roadmap from step 2

TYPICAL TIMELINE: 1 WEEK

Step 2 - Test and Optimize

ENGAGEMENT

- Prioritize front end initiatives to improve UX and CWV
- Setup suitable pre-production test environment
- Develop load simulation scripts and configure load tests
- Configure automated alerting and reporting
- Benchmark and baseline performance

OUTPUT

- Clear roadmap for optimization opportunities
- Establish goals and KPI's
- Implemented automated solutions
- Testcase repository created
- Benchmarked performance metrics

TYPICAL TIMELINE: 2 WEEKS

Step 3 - Ongoing Engagement within Dev-Ops

ENGAGEMENT

Performance is now a consideration within the release cycles and dev-ops structure. There is a focus on end to end optimization, everything from user device to data storage is monitored, analysed, tested and optimized.

8West manage the tooling, testing and reporting and provide guidance from the wider cloud engineering team for architectural improvements and cloud services going forward.

OUTPUT

- Release cycle performance testing
- UX monitoring and reporting
- Test repository maintenance
- Cloud architecture and services updates
- Automated monitoring and alerting

TIMELINE - ONGOING QUARTERLY MAINTENANCE CONTRACT

Ready to Take the Next Step?

Simply arrange a date and time for your no obligation introductory meeting with our cloud engineering team.

Already have specific needs in other areas? We can arrange for you to meet with one of the specialists in our other core areas of competency:

- Architectural services
- Security services
- FinOps services

OR Request a summary of our services and tech talent rate card

8 west
CONSULTING

LET US KNOW HOW WE CAN HELP

or request our rate card and service listing

info@8west.ie