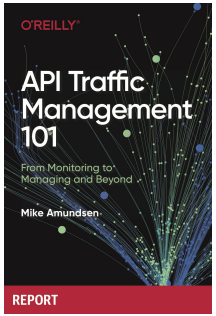
High Performing APIs

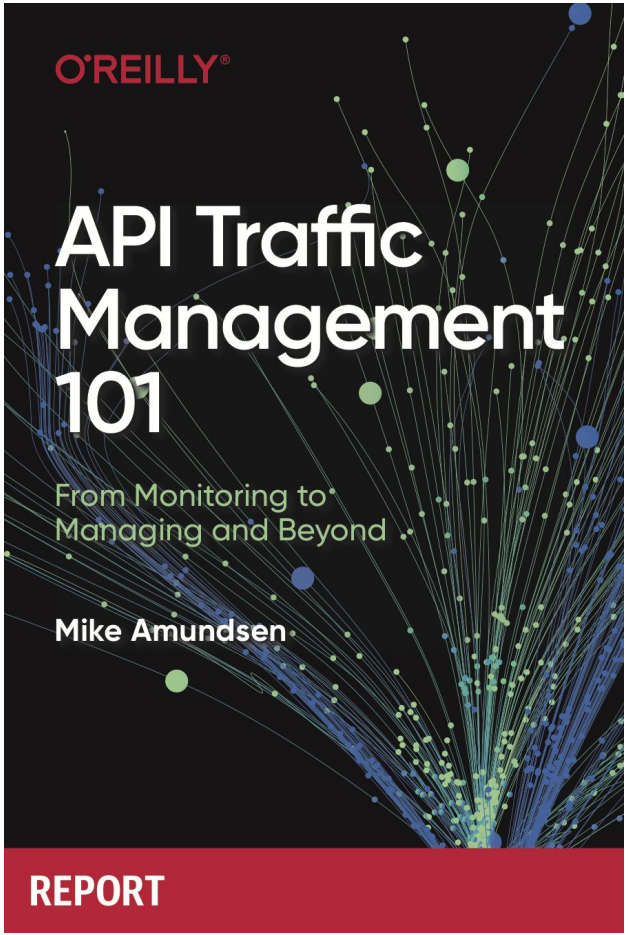
Architecting for speed at scale

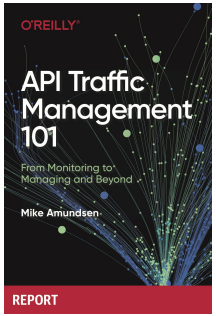
@mamund
Mike Amundsen
youtube.com/mamund




Mike Amundsen
@mamund

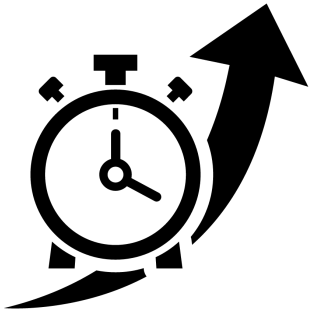
"Designed to provide you important insight into patterns and trends as well as pointers to specific tools and practices that you can use to build up your own experience and grow an API traffic management practice in your own company."

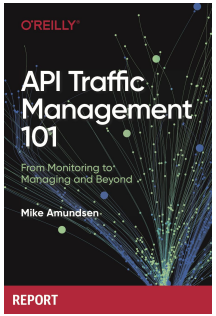




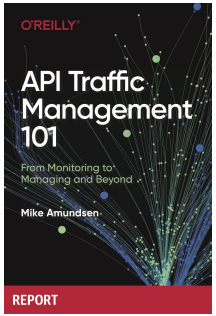
Overview

- The Performance Imperative
- Architecting for Performance
- Monitoring for Performance
- Managing for Performance



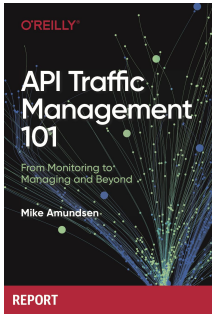


The Performance Imperative



The Performance Imperative

- Ecosystem Transformation
- API Call Volume
- Transaction Response Time



APIs are at the core of modern enterprises

75%

of organizations will be completely digitally transformed in the next decade. Those that do not transform will not survive *

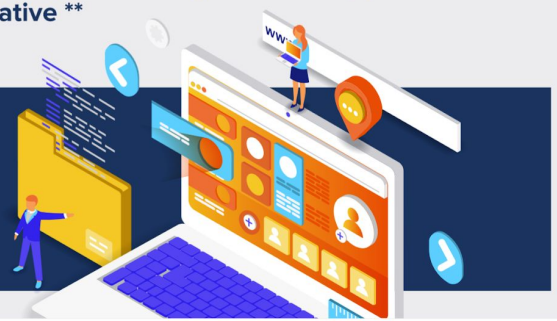
By 2022

90%

of all new apps will feature microservices architecture; 35% of all production apps will be cloud native **

APIs have a central role in both enabling the digital business and powering modern, microservices-based application architectures.

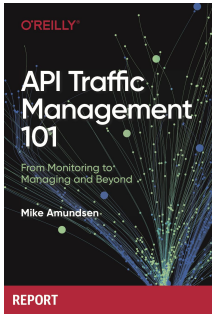
No organization can afford to ignore the pivotal role APIs hold in application and business modernization. Organizations that do not bring APIs at the core of their IT strategy will face substantial challenges to transform their technology and business foundations.

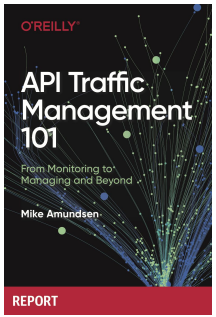


IDC ANALYZE THE FUTURE

*2019 IDC MaturityScape: Digital Transformation 1.0 **2019 IDC Futurescape

IDC #EUR145216019 | 3 | An IDC InfoBrief, Sponsored by NGINX

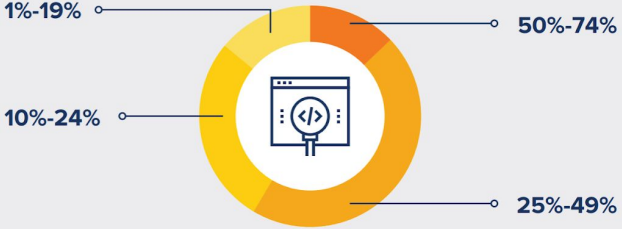




Acceleration in API Call Volumes Requires the Right Level of Management

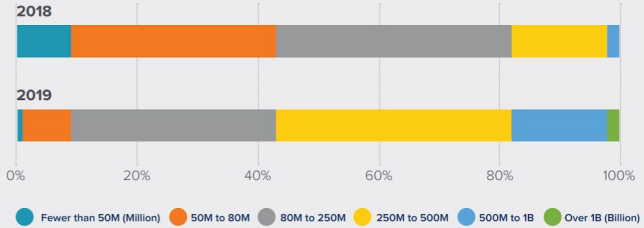
71% of organizations expect to see the volume of API calls increase in the next 2 years.

Expected increase in the volume of API calls:



Q. By how much will the number of API calls increase 2 years from now?

Q. Please estimate the number of total API calls on a monthly basis, today and 1 year ago?



As the volume of API calls is clearly growing, enterprises need to plan their ***API load balancing and traffic management*** position accordingly.

Top API Management components organizations will invest in over the next year

46% Management and monitoring of API

36% API traffic management

IDC | ANALYZE THE FUTURE

Source: IDC API Management Survey 2019, API Focused organizations N = 301

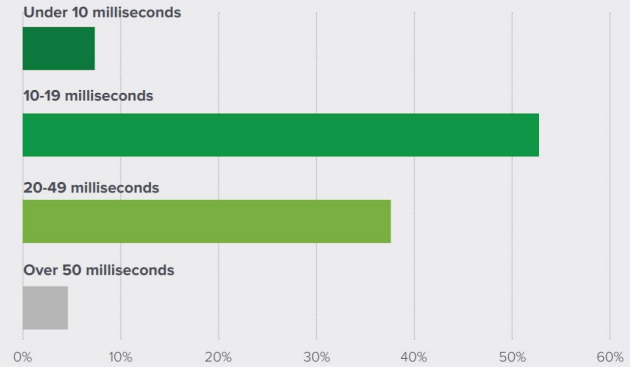
Performance is Critical for Successful API Programs

59%
of organizations expect a latency of
under 20 milliseconds

93%
of organizations expect a latency of
under 50 milliseconds

At enterprise volume and scale, ensuring the adequate mechanisms to manage and ensure performace is very important. Poor performance results in APIs not being adopted by API customers, which cascades into failed business opportunities and poor ROI.

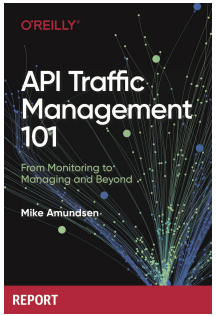
Q. Please indicate your performance expectations per average API call



IDC ANALYZE THE FUTURE

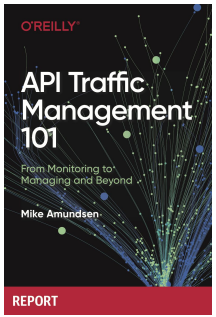
Source: IDC API Management Survey 2019, API Focused organizations N = 301

IDC #EUR145216019 | 14 | An IDC InfoBrief, Sponsored by NGINX

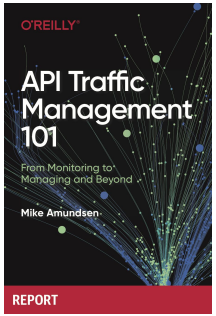


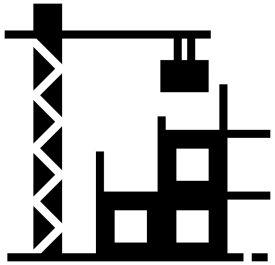
The Performance Imperative

- Transformation
  - 90% of companies will need to support microservice architectures
- Volume
  - Over 80% of respondents expected 250mil calls/month
- Response Time
  - About 60% of calls need to be within 20 ms

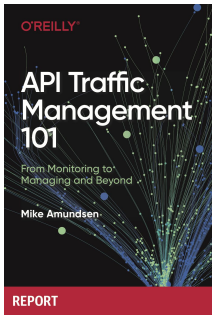


How do we meet these new demands?

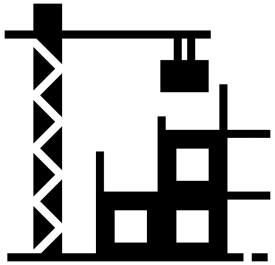




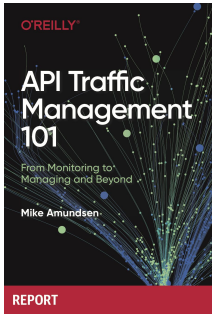
Architecting for Performance



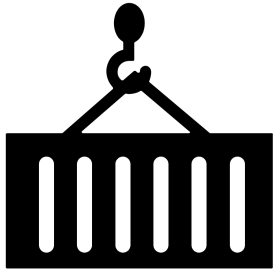
Architecting for Performance



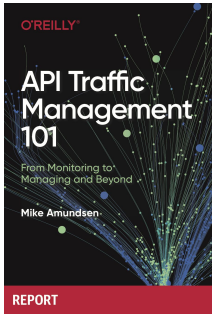
- Lift-and-Shift is not enough
- Redesigning Services
- Re-engineering Data
- Rethinking the Network



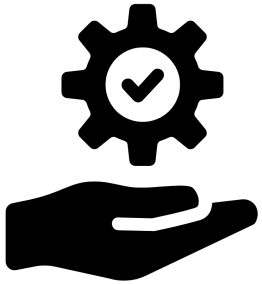
Architecting for Performance



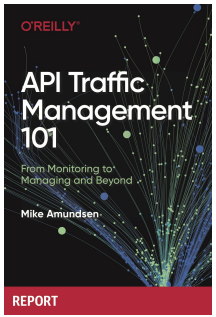
- Lift-and-Shift is not enough
  - Simply copying your on-prem to the cloud has limits
  - Adding distance & connections slows performance
  - Native storage and services operate under different rules
- Redesigning Services
- Re-engineering Data
- Rethinking the Network



Architecting for Performance

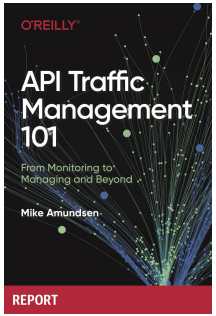


- Lift-and-Shift is not enough
- Redesigning Services
  - Make services smaller
  - Reduce wait states (async)
  - Build-in reversal and recovery
- Re-engineering Data
- Rethinking the Network

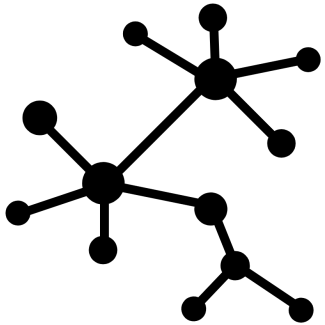


Architecting for Performance

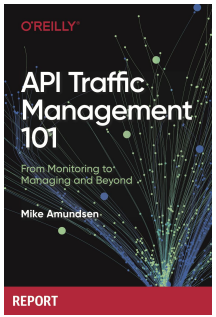
- Lift-and-Shift is not enough
- Redesigning Services
- Re-engineering Data
  - Cache results
  - Stage copies
  - Stream writes
- Rethinking the Network

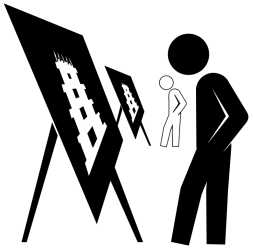


Architecting for Performance

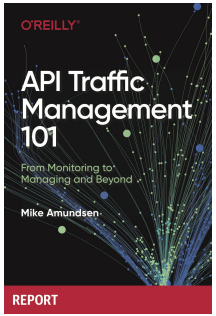


- Lift-and-Shift is not enough
- Redesigning Services
- Re-engineering Data
- Rethinking the Network
  - Decrease message size
  - Increase message volume
  - The return of RPC

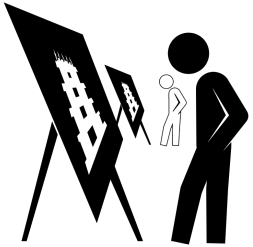




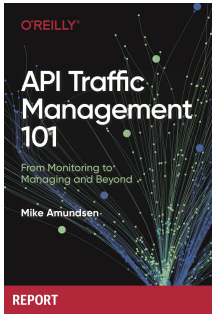
Monitoring for Performance

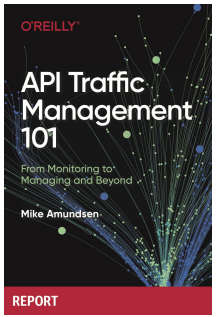


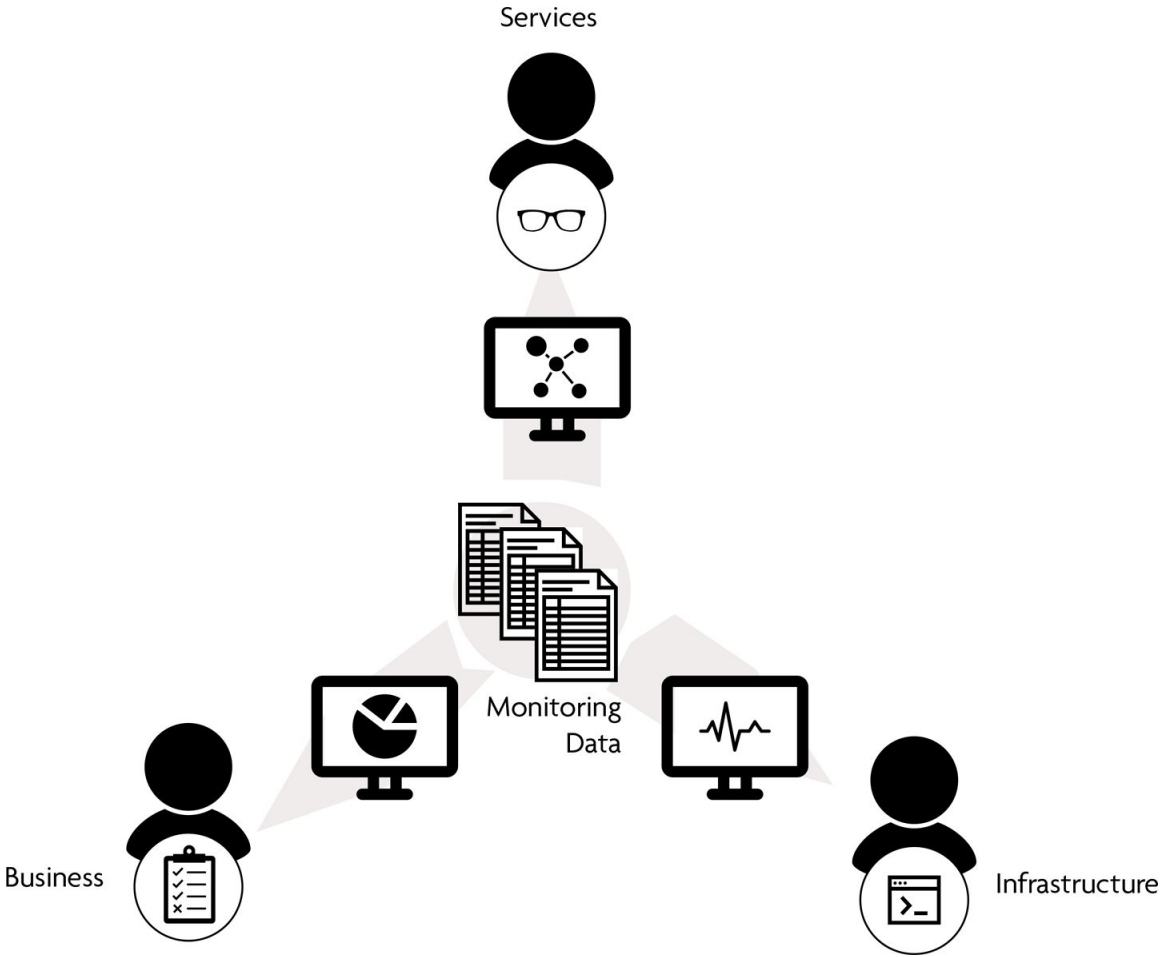
Monitoring for Performance



- Infrastructure
- Services
- Business

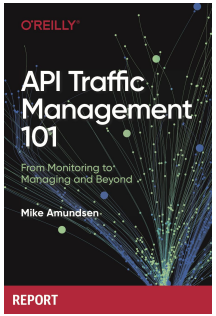




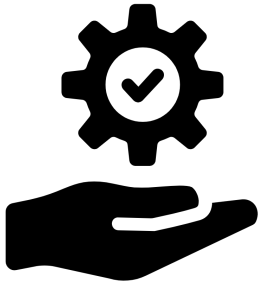


Monitoring for Performance

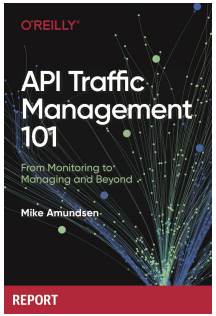
- Infrastructure
  - Machines and network connections
  - CPU, memory, bandwidth, saturation
- Services
- Business



Monitoring for Performance

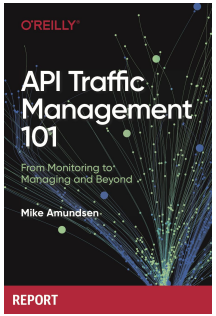


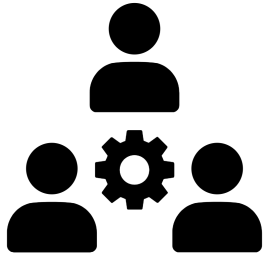
- Infrastructure
- Services
  - Microservices, ESBs, etc.
  - Latency, error rates, limits, etc.
- Business



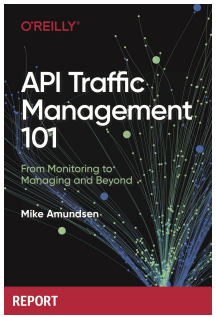
Monitoring for Performance

- Infrastructure
- Services
- Business
  - Users, transactions, etc.
  - Completed orders, new signups, etc.

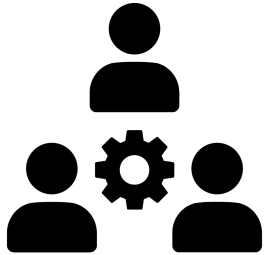




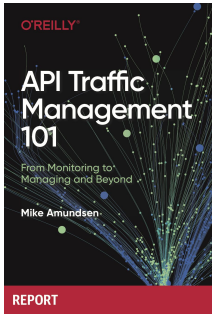
Managing for Performance

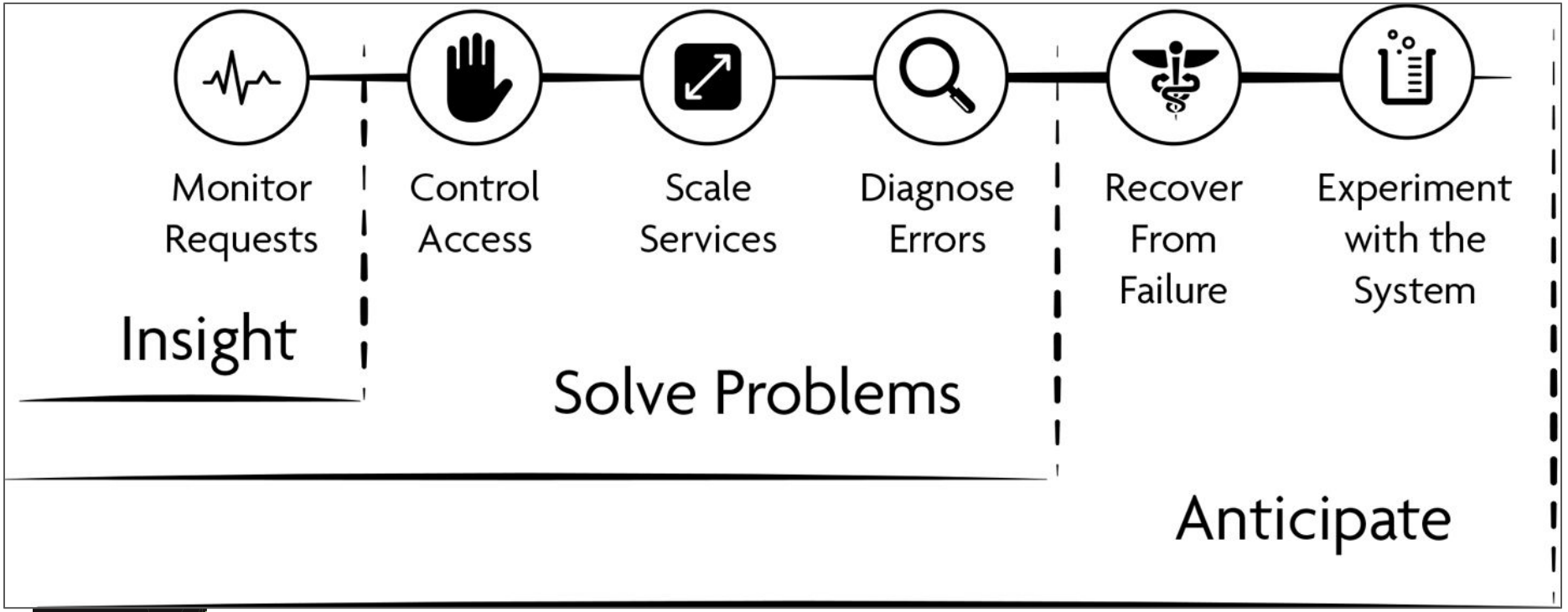


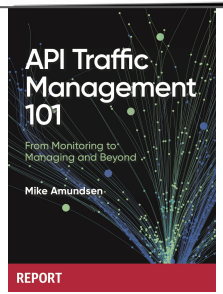
Managing for Performance



- Insight
- Solving Problems
- Anticipating Needs

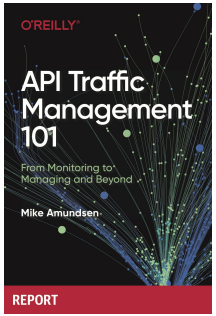




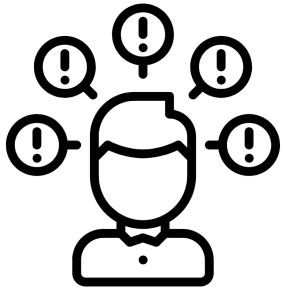


Managing for Performance

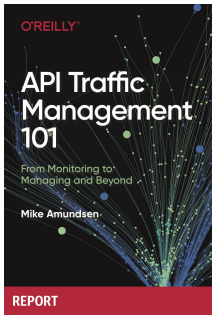
- Insights
  - Monitor traffic
  - Monitor builds
- Solving Problems
- Anticipating Needs



Managing for Performance

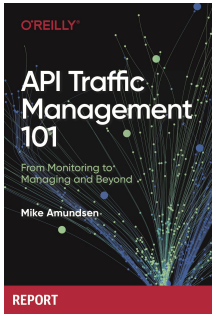


- Insights
- Solving Problems
  - Security Watch
  - Scaling Services
  - Diagnosing Errors
- Anticipating Needs

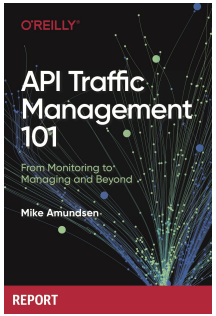


Managing for Performance

- Insights
- Solving Problems
- Anticipating Needs
  - Automating Recovery
  - Running Experiments

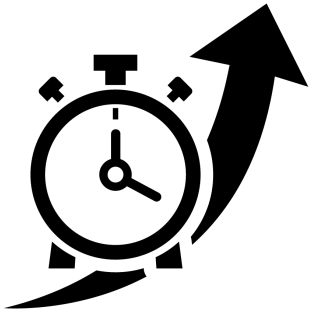


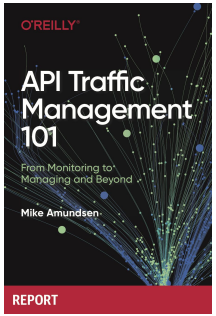
And So...



Overview

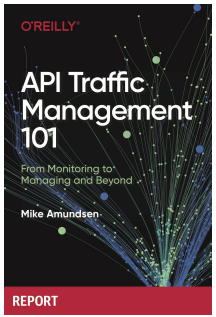
- The Performance Imperative
- Architecting for Performance
- Monitoring for Performance
- Managing for Performance



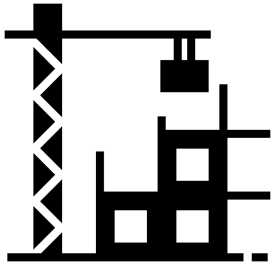


Overview

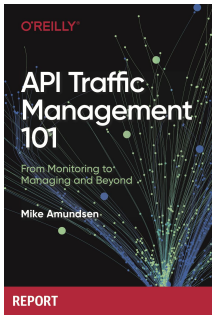
- The Performance Imperative
  - Prepare for call volumes to go up and transaction time to go down
- Architecting for Performance
- Monitoring for Performance
- Managing for Performance



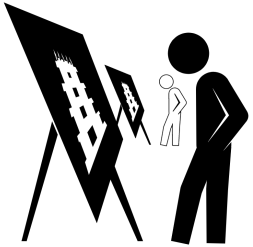
Overview



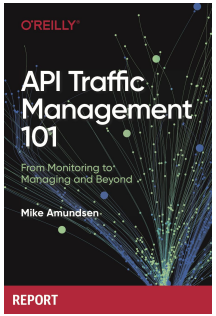
- The Performance Imperative
- Architecting for Performance
  - Redesign services, re-engineer data, rethink networks
- Monitoring for Performance
- Managing for Performance



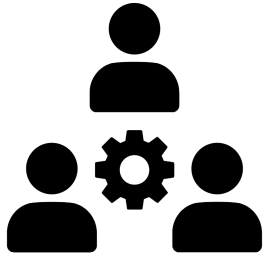
Overview



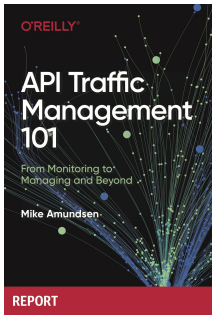
- The Performance Imperative
- Architecting for Performance
- Monitoring for Performance
  - Monitor infrastructure, services, and your business metrics
- Managing for Performance



Overview

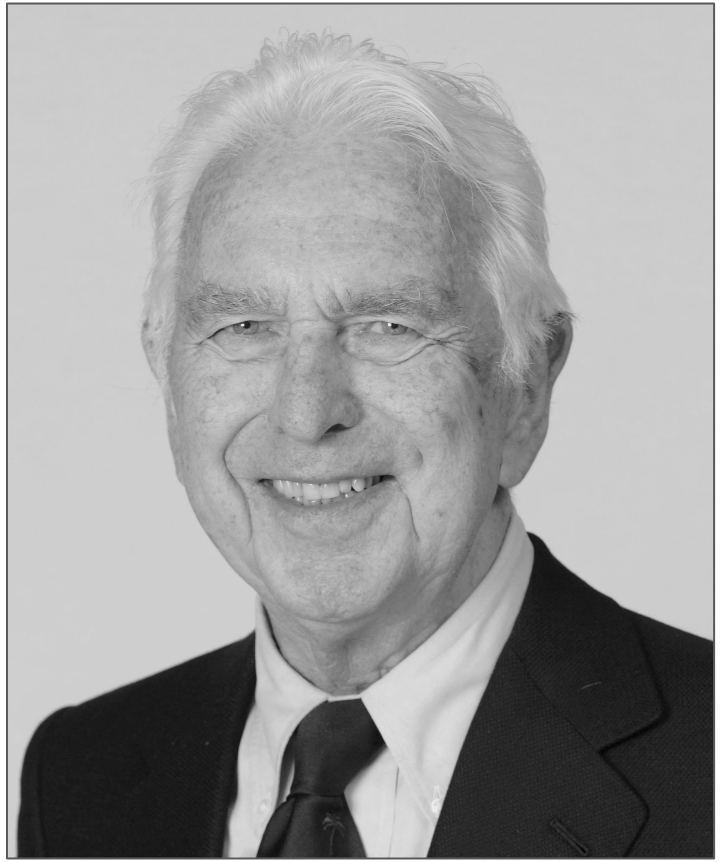


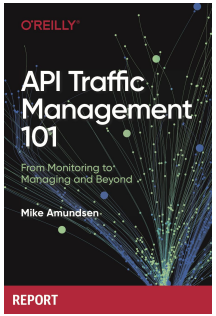
- The Performance Imperative
- Architecting for Performance
- Monitoring for Performance
- Managing for Performance
  - Manage traffic, resolve problems, and anticipate needs



"In life, change is inevitable. In business, change is vital."

-- Warren G. Bennis
On Becoming a Leader (1989)





http://g.mamund.com/api-traffic

High Performing APIs

Architecting for speed at scale

@mamund
Mike Amundsen
youtube.com/mamund