Building Resilient Apps
Architecting for Continuous Improvement
Resilience in AWS Well-Architected

Operational excellence

Reliability
Operational excellence pillar

- Perform operations as code
- Make frequent, small, reversible changes
- Refine operations procedures frequently
- Anticipate failure
- Learn from all operational failures
Resilience ‘in’ the cloud

ACHIEVING CONTINUOUS RESILIENCE
Where failures happen

Code deployments & configuration
e.g. bad deployment, cred expiration

Core infrastructure
e.g. data center failure, host failure

Data and state
e.g. data corruption

Dependencies
e.g. infrastructure, external APIs
Building a resilient foundation

- Applications
- Data
- Network
- AWS Infrastructure
The three pillars of observability

Logs
Metrics
Traces

Answer the questions you didn't know you needed to ask
The cycle of monitoring

Instrumentation

Questions

Improved operations

Logs, metrics, traces

Alarms, dashboards
Metrics: What to measure?

Can we get stats around
How much log data
Is collected on average
Per day?
Example of what to measure?

- Which product are we looking up?
- Who called the API?
- Did we find the item in the local cache?
- Did we find the item in the remote cache?
- How long did it take to read from the cache?
- How full is the local cache?
- How long did the query take?
- Did the query succeed?
- How long did it take to populate the caches?
- Were they full and did they evict other items?
- How big was the product info object?
- What was the response code from the server?
- What was the latency?

Diagram:

- Customer
- Load balancer
- Remote cache
- Remote database
- Cache
- Local cache
Different types of metrics

**Health metrics**
“Am I failing?”

**Diagnostic metrics**
“What is happening in my system?”
Implementing essential metrics

IN ORDER TO DETECT, INVESTIGATE, AND RESPOND TO IMPACT

- Business metrics
- Impact assessment metrics
- Operational health metrics

© 2023, Amazon Web Services, Inc. or its affiliates.
Business metrics

Payments – Transactions per second
Retail – Order per second
Streaming – Stream starts per second
Media/Audio – Playback event started

All are related to business KPIs and performance
Operational health metrics

Latency
API Error Rates
CPU Utilization
Disk Usage/IOPs

All are related to internal systems
Differential observability

Service View

Customer View

Healthy

Unhealthy

Normal Operations

Gray Failure

Over-reaction

Hard Failure

Healthy

Unhealthy
Gray Failure Example
Detection Tools

**Amazon CloudWatch Contributor Insights**

- Visualize metrics from the top-N contributors and see the total number of unique contributors
- Can find top talkers, bad hosts, heaviest network users, or resources with the most errors
- Can be used with CloudWatch metric math
- Compare logs against rules you define

**CloudWatch composite alarms**

- Monitor the state of other alarms
- Can also monitor other composite alarms
- Simple logic alarm expressions, AND, OR, NOT

```plaintext
(ALARM("CPUUtilizationTooHigh")
 OR
 ALARM("DiskReadOpsTooHigh")
)
AND OK("NetworkOutTooHigh")
```
Detecting a Gray Failure Impacting a Single AZ

Controller: Home
Action: Index
AZ ID: use1-az1
Threshold: Availability < 99.9%

Controller: Products
Action: List
AZ ID: use1-az1
Threshold: Availability < 99.9%

Controller: Home
Action: Index
AZ ID: use1-az1
Threshold: Latency > 100ms

Controller: Products
Action: List
AZ ID: use1-az1
Threshold: Latency > 150ms

**az1-availability**
ALARM(A) OR ALARM(B)

**az1-latency**
ALARM(A) OR ALARM(B)

**use1-az1-impact**
(ALARM(az1-availability) OR ALARM(az1-latency)) AND NOT
(ALARM(az2-availability) OR ALARM(az2-latency)) OR
ALARM(az3-availability) OR ALARM(az3-latency))

**not-single-instance-use1-az1**
INSIGHT_RULE_METRIC(“5xx-errors-use1-az1”, “UniqueContributors”) >= 2

**use1-az1-isolated-impact**
ALARM(use1-az1-impact) AND ALARM(not-single-instance-use1-az1)
Reducing Failover Time from 30 Minutes to 3 Minutes Using Amazon CloudWatch with Thomson Reuters

Customer Story

Challenge

With over 38,000 employees in more than 100 countries, Thomson Reuters needs to provide personnel with secure access to company systems, no matter where they are.

Solution

Thomson Reuters improved its single sign-on (SSO) solution using Amazon CloudWatch and Amazon Route 53.

Results

• 1.5-hour reduction in recovery point objective time
• 27-minute reduction in failover time
• Enhanced security
• Increased availability
• Saved labor time
Thank you!

[speaker name]
[speaker role]
[speaker team]
linkedin.com/in/{speaker handler}

Please complete the survey