How to Build Reliable Systems Under Unpredictable Conditions

Smart Approaches in Chaos Engineering, Observability, and Incident Management

steadybit

Speaker

Benjamin Wilms

CO-FOUNDER & CEO

steadybit

Setting the scene

What is the mission of software development?

Continuously improve & deliver a software solution that delivers value to its users.

Customers trust a system when it’s consistently good in quality and performance

Why is it so hard?

The complexity of today's systems is massive

The complexity of today's systems is massive

Definition: System

Definition: System

Hardware

Software People Process

Organisation Pipeline

...and more.

Improve the Reliability of your System

By unleashing the Power of Chaos Engineering

> *“Chaos Engineering is thoughtful, planned experiments that reveal the weakness in our sociotechnical systems before they appear in production.”*

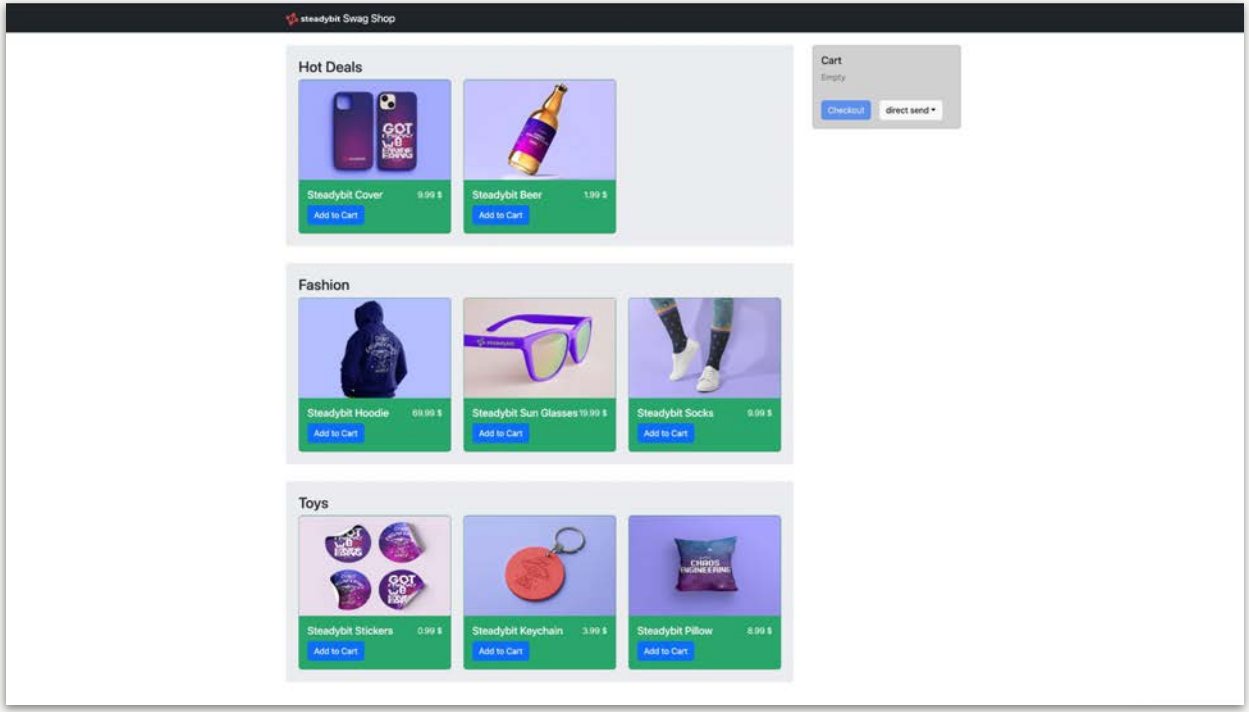
RUSS MILES
AUTHOR OF “LEARNING CHAOS ENGINEERING”

> *“Chaos Engineering is thoughtful, planned experiments that reveal the weakness in our **sociotechnical systems** before they appear in production.”*

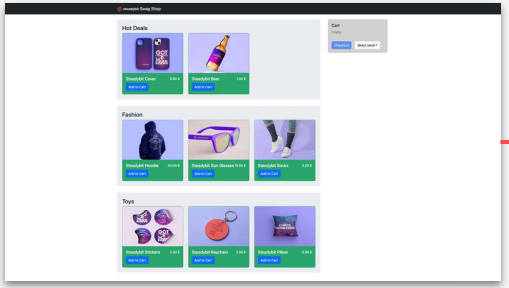
RUSS MILES
AUTHOR OF “LEARNING CHAOS ENGINEERING”

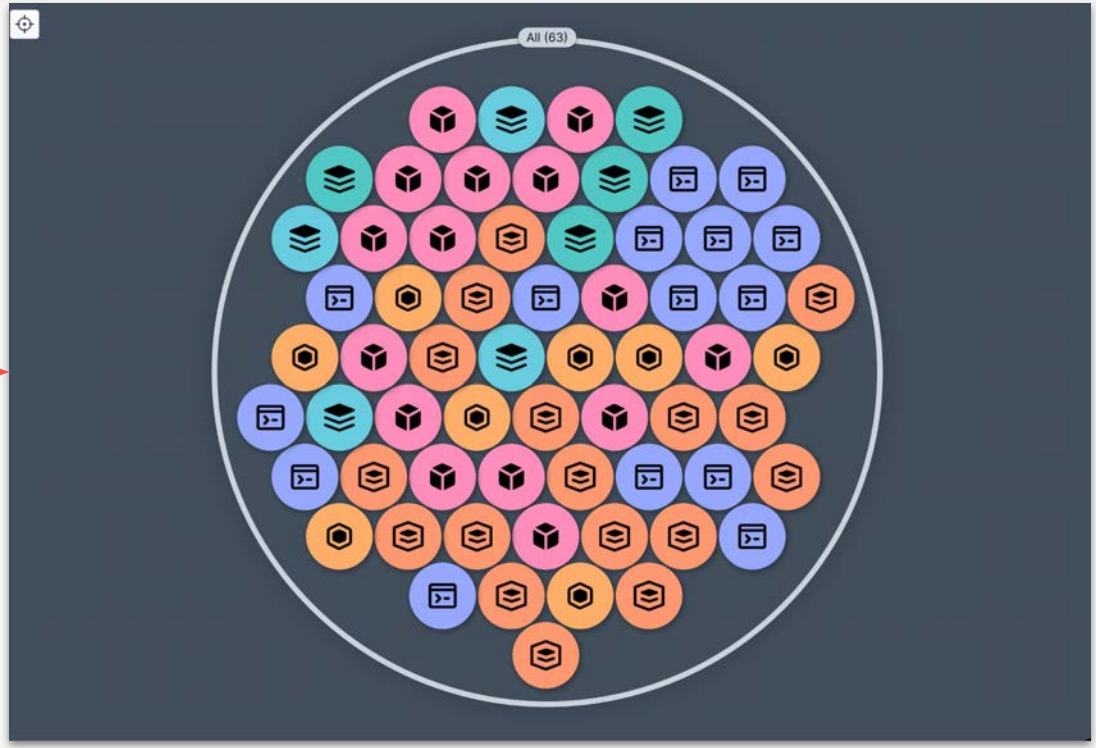
Exemplary result a sociotechnical system creates

Online shopping system

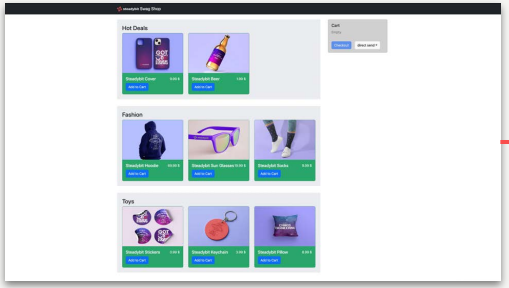


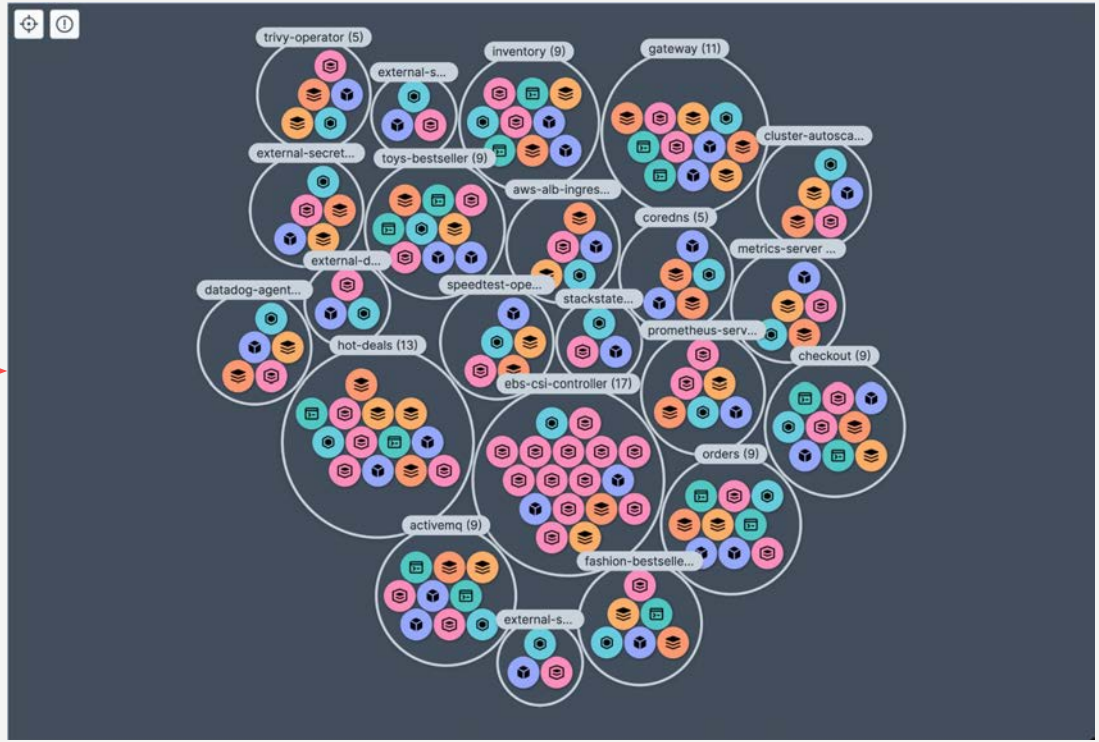
Technical view





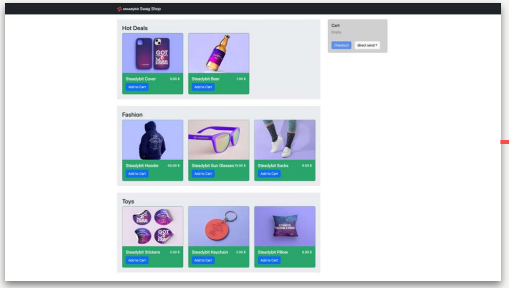
Technical view

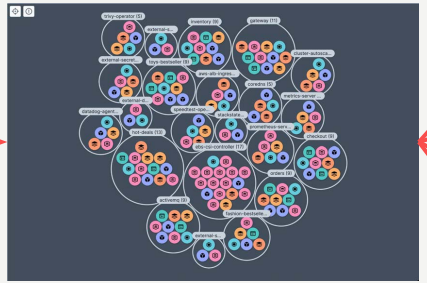




steadybit

Organizational view





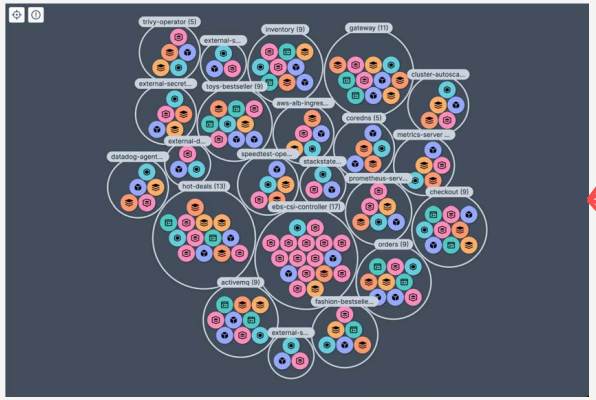
Team A

Team B

Team C

Team D

Organizational view



Team A

Team B

Team C

Team D

steadybit

Building Resilient Applications Part 1

steadybit

Increase Systemic Resilience...

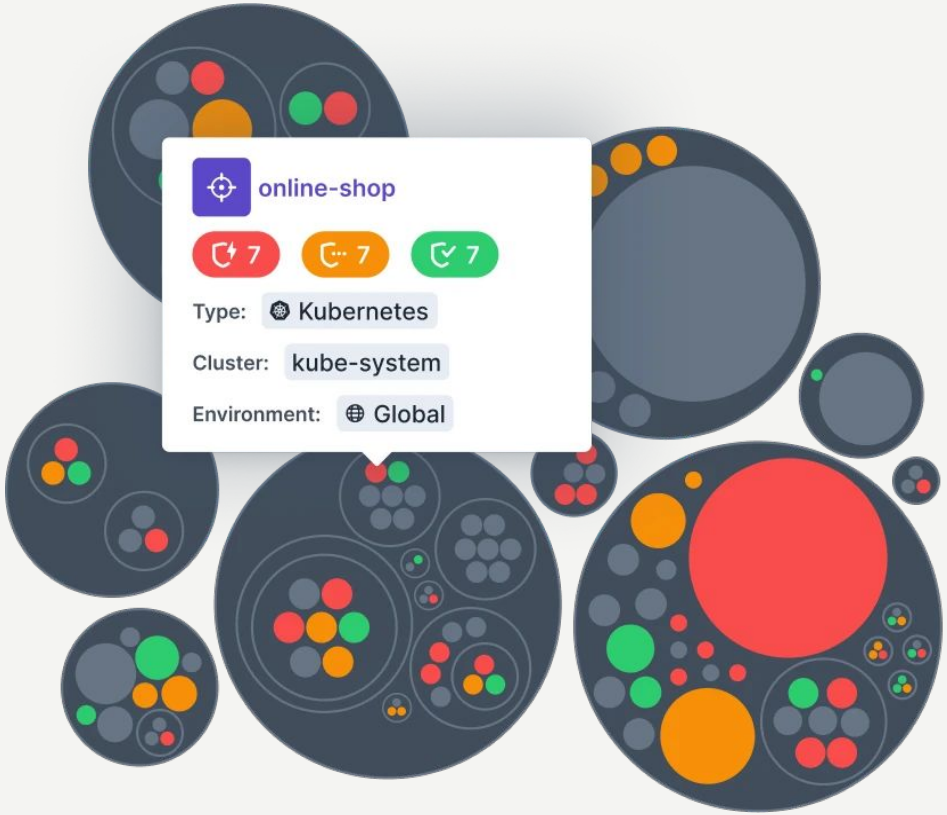
...with Steadybit.

steadybit

Uncover Risk

Identify and Address System Vulnerabilities

Advice analyzes system configurations to identify vulnerabilities and areas for improvement. It provides precise recommendations to enhance system robustness, including optimizing Kubernetes setups and ensuring effective redundancy.

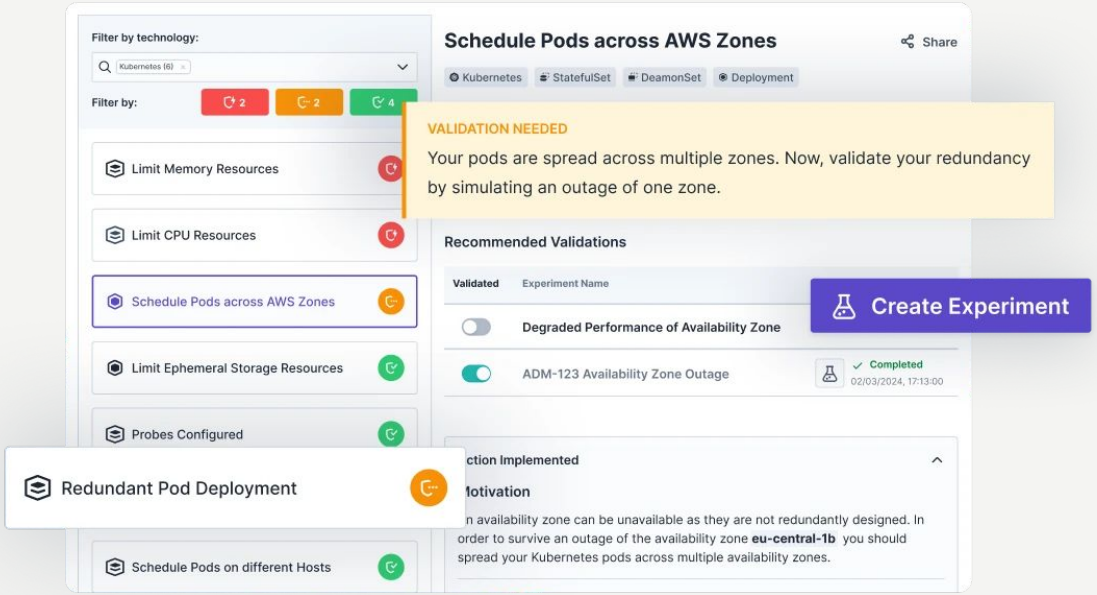


steadybit

Understand Impact

Verify your Resilience Strategy

Conduct targeted Chaos Engineering experiments to verify that the technical system withstands against potential disruptions.

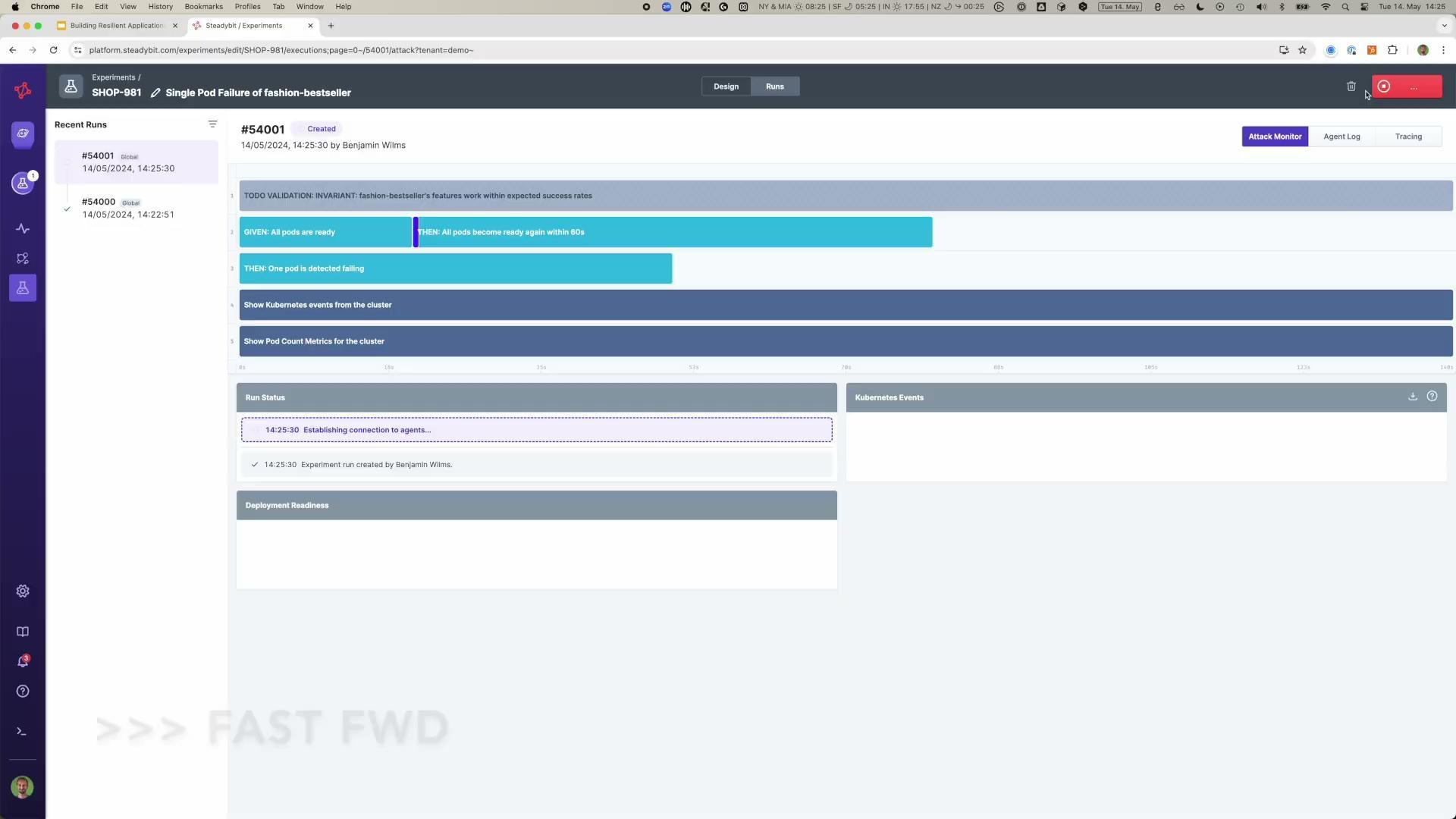


steadybit

Uncover Risk & Understand Impact

Demo Landscape & Reliability Advice

https://platform.steadybit.com/landscape/explore/997ca4ba-ea7a-4d9d-98aa-de12362bc73f?tenant=demo~


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Building Resilient Application
Steadybit / Experiments
platform.steadybit.com/experiments/edit/SHOP-981/executions;page=0~/54001/attack?tenant=demo~
Experiments /
SHOP-981
Single Pod Failure of fashion-bestseller
Design
Runs
Recent Runs
#54001 Global
14/05/2024, 14:25:30
#54000 Global
14/05/2024, 14:22:51
#54001
Created
14/05/2024, 14:25:30 by Benjamin Wilms
Attack Monitor
Agent Log
Tracing
TODO VALIDATION: INVARIANT: fashion-bestseller's features work within expected success rates
GIVEN: All pods are ready
THEN: All pods become ready again within 60s
THEN: One pod is detected failing
Show Kubernetes events from the cluster
Show Pod Count Metrics for the cluster
Run Status
14:25:30 Establishing connection to agents...
14:25:30 Experiment run created by Benjamin Wilms.
Kubernetes Events
Deployment Readiness
>>> FAST FWD

Building Resilient Applications Part 2

steadybit

Increase Systemic Resilience...

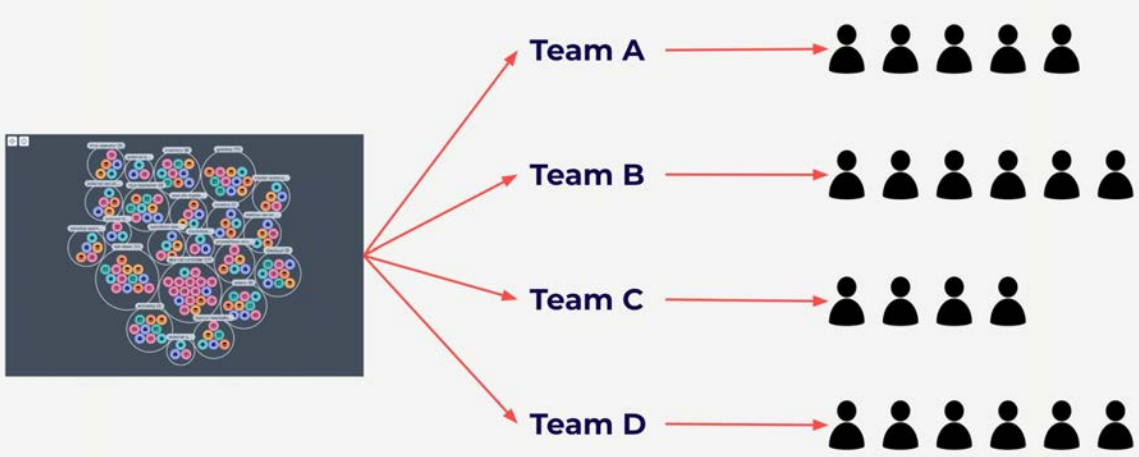
...with Steadybit.

steadybit

Strengthen Organisation

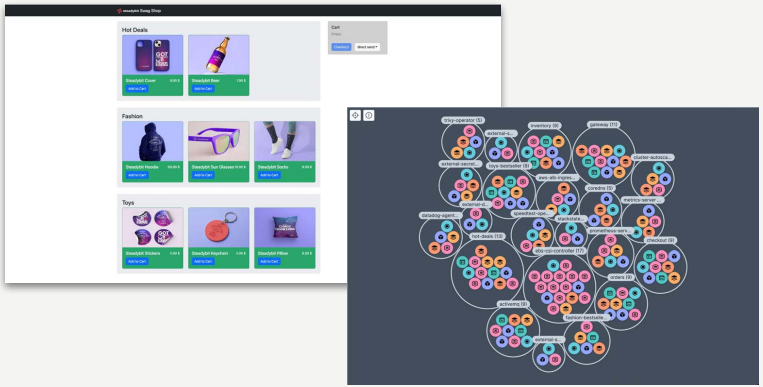
Verify interaction of all necessary players

After continuously improving and reviewing our technical system, we turn our attention to the sociotechnical system



Strengthen Organisation

Technical System



Organisational System

Team B

Team A

Team C

Team D

steadybit

Strengthen Organisation

Scenario

Latency spikes in backend service *Hot Deals* followed by total outage in the central *Gateway* service.

Hypothesis

Our monitoring recognises this scenario within 90s and reports it to the relevant team on call via PagerDuty. The incident is opened and acknowledged within 3 minutes.

The system normalises within 3 minutes and the incident and our on-call team determines that the resilient system has recovered and is working normally. The incident is resolved within 4 minutes..

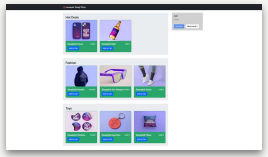
Strengthen Organisation

Interaction

Strengthen Organisation

Timeline

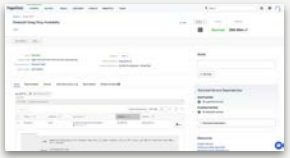
Latency spike followed by total outage



Monitoring **Event** within 90s

Incident **Triggered** in PagerDuty within 100s

Incident **Acknowledged** in PagerDuty within 3m



System **OK** after 3m

Incident **Resolved** in PagerDuty within 4m

0s — 4m

steadybit

It's not about testing how fast an individual person fixes an outage.

It’s about how good the organisation is at detecting and fixing faults and how processes are coordinated.

Strengthen Organisation

Demo scenario design in Steadybit

https://platform.steadybit.com/experiments/edit/SHOP-972/design?tenant=demo~&team=SHOP~

Experiments /
SHOP-972 Webinar Showcase

Design | Runs

00:04:59

Recent Runs

#53580 Global
03/05/2024, 17:18:19

#53580 Running

03/05/2024, 17:18:19 by Benjamin Wilms

Attack Monitor | Agent Log | Tracing

Run Status

17:18:25 Datadog Monitor Status = OK | Shopping UI Availability
17:18:25 Default system behaviour
17:18:25 Default system behaviour
17:18:25 Default system behaviour
17:18:25 Default system behaviour

HTTP Responses for https://api.pagerduty.com/incidents?urgency=high&statuses[]=triggered&service_ids[]=P3V6VIA&sort_by=created_at:DESC

Waiting for metrics...

HTTP Responses for https://api.pagerduty.com/incidents?urgency=high&statuses[]=acknowledged&service_ids[]=P3V6VIA&sort_by=created_at:DESC

Waiting for metrics...

HTTP Responses for https://api.pagerduty.com/incidents?urgency=high&statuses[]=resolved&service_ids[]=P3V6VIA&sort_by=created_at:DESC

Waiting for metrics...

PRE SYSTEM CHECK

Datadog Monitor Status

Recap

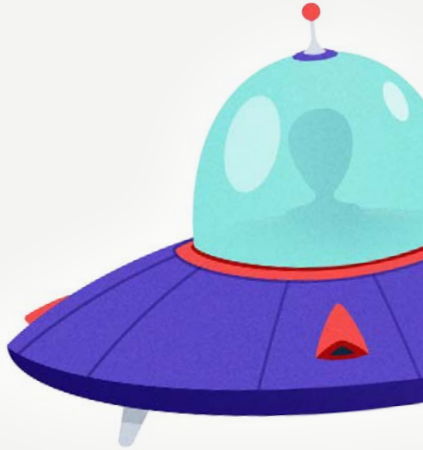
Contact details

Benjamin Wilms

steadybit.com

signup.steadybit.com

Thank you for your time.



steadybit