Análisis de incidencias

Curso de escalabilidad v2, día 5

Qué veremos hoy

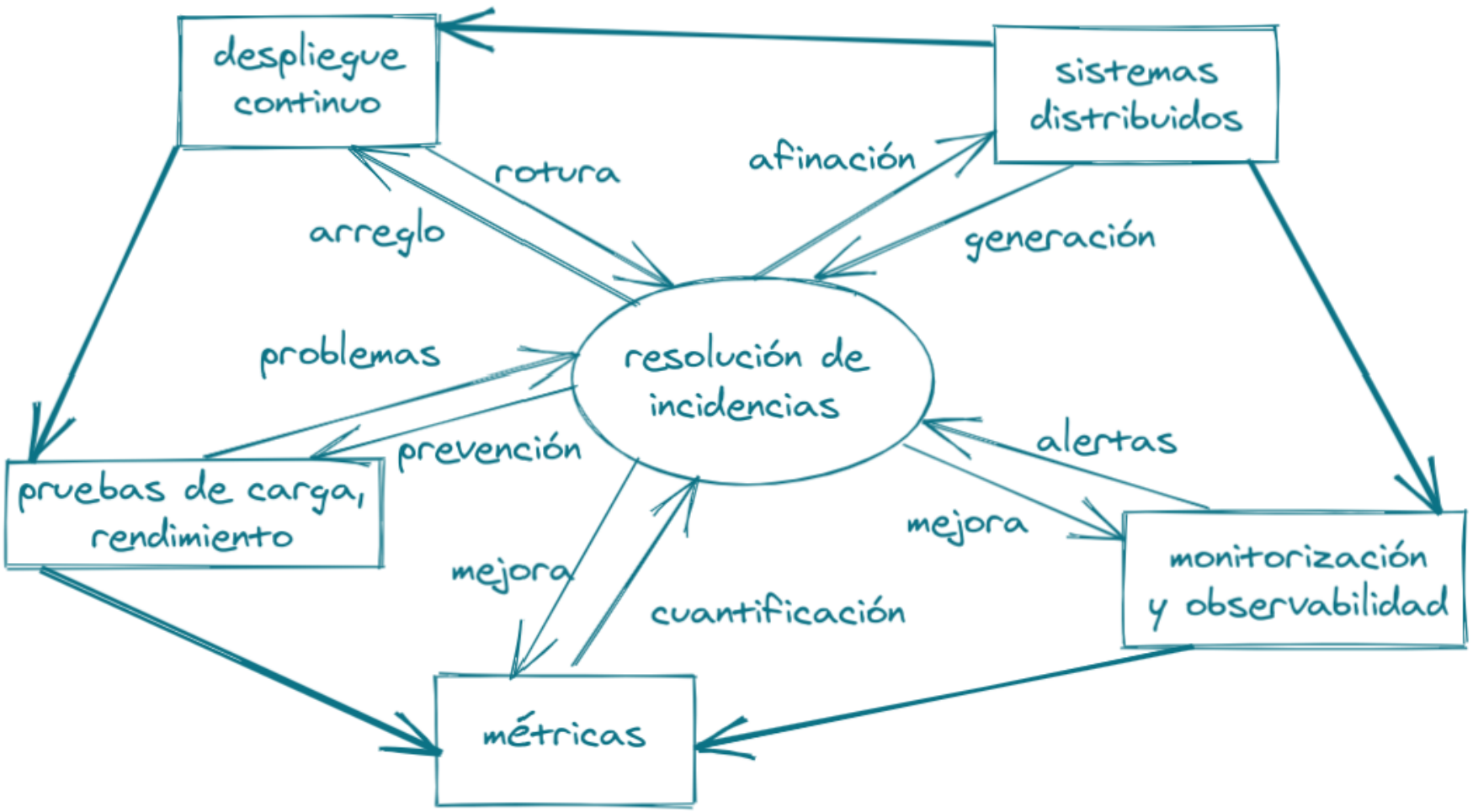
Investigación de causa raíz. Los cinco porqués

Postmortems sin culpa

Tareas post-incidente

Actitudes y expectativas. Liderazgo

Un hueco con forma de incidencias



¿Qué es una incidencia?

Un problema con impacto negativo en el servicio

Idealmente, resulta en una o más alertas

Requiere atención fuera de lo ordinario

Con consecuencias en usuarios o negocio

Ejercicio: Explosión de Chernóbil

Revisa la documentación del incidente

Intenta pensar cuál es la causa raíz

¿Cómo la resolverías?

□

¡Suspense!

Investigación de causa raíz

Los cinco porqués

Hay que preguntar como un niño

No detenerse en la primera causa

Pero, ¿**por qué cinco porqués**?

Hay que seguir hasta que esté todo claro

Problema: ¿causa raíz?

En un sistema complejo
los fallos no son por una única causa

Siempre hay que buscar múltiples fallos
y luego arreglarlos todos

Bucea hasta que tengas completa seguridad
de que has entendido el problema

Raíz de causas

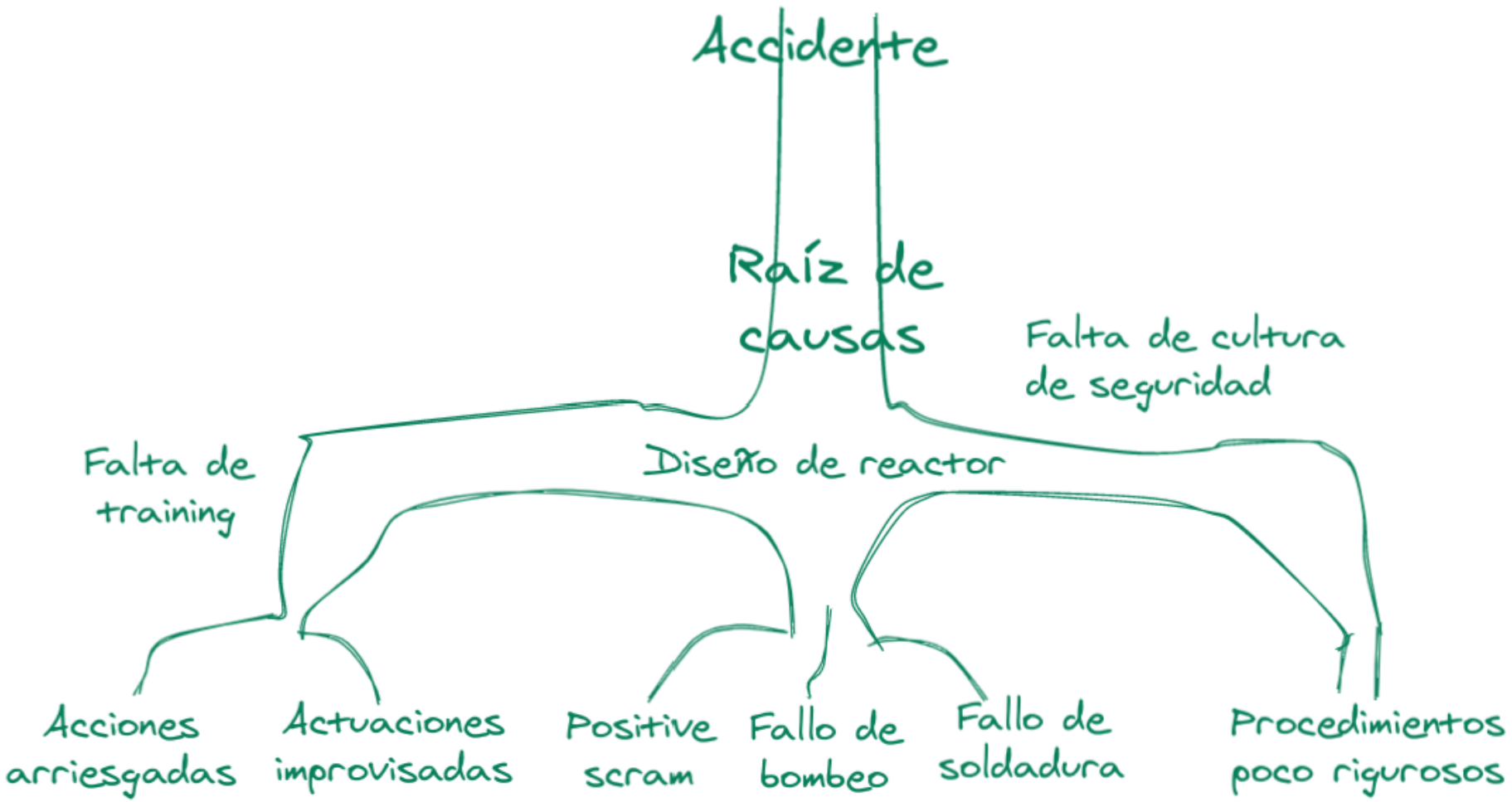
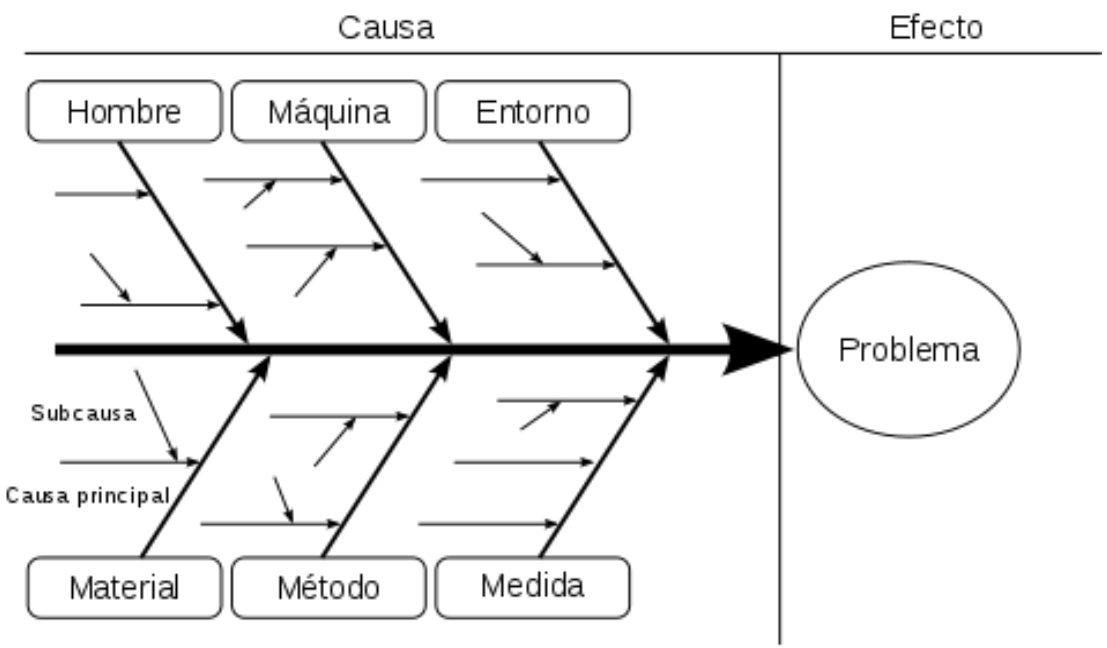


Diagrama de Ishikawa

O de espina de pescado



Finding the root cause of a failure is like finding the root cause of a success

John Allspaw

Ejercicio: Colisión USS John McCain

Revisa la documentación del incidente

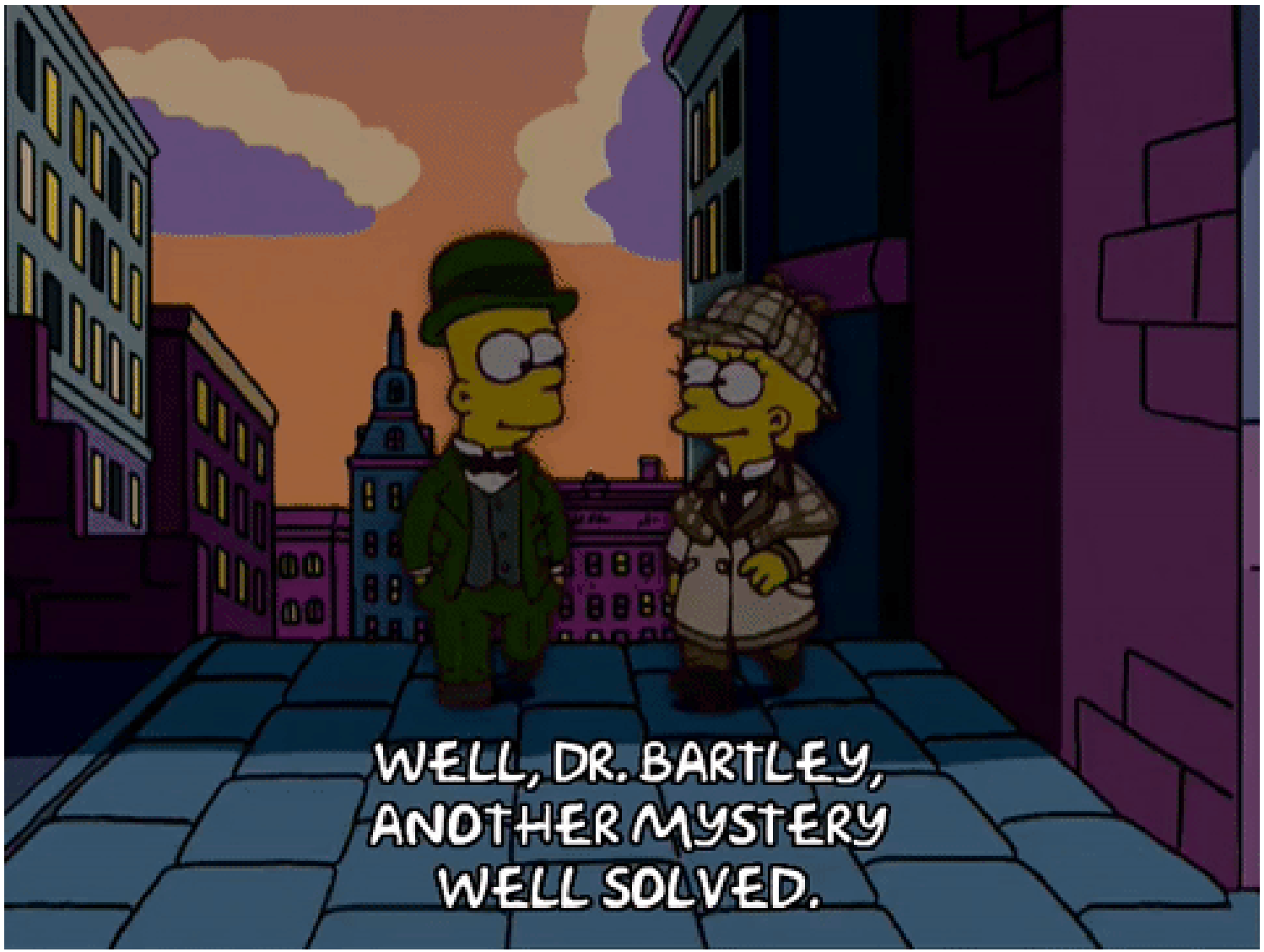
Revisa las consecuencias del mismo

¿Quién tuvo la culpa?

¿Qué medidas fueron acertadas y cuáles no?

□

Mystery solved!



Situaciones emocionales

Los ejercicios que estamos viendo son fuertes

¿Es posible un punto de vista desapasionado?

Estas industrias han tenido décadas (o siglos) para aprender

Saber hacer un análisis desapasionado es la clave

Postmortems sin culpa

¡Error humano!

Human error is not a cause, it is an effect.

John Allspaw: Outages, Post Mortems, and Human Error 101

Si se castiga a la gente por ser honrada sobre lo que ha pasado, los empleados pronto aprenderán que el coste personal de hablar no compensa los posibles beneficios personales. La mejora en seguridad de un sistema se basa en la información.

United States Forest Service
en Blameless Post Mortems

Segundas historias

First Stories	Second Stories
Human error is seen as cause of failure	Human error is seen as the effect of systemic vulnerabilities deeper inside the organization
Saying what people should have done is a satisfying way to describe failure	Saying what people should have done doesn't explain why it made sense for them to do what they did
Telling people to be more careful will make the problem go away	Only by constantly seeking out its vulnerabilities can organizations enhance safety

Behind Human Error, via Blameless Postmortems

Línea temporal (*Timeline*)

Habla con las partes implicadas

Asegúrate de entender las consecuencias

Narra los sucesos en orden

Crea una historia
Captura los detalles
Contrástala

Labor de equipo

Revisión de código

Responsabilidad compartida

Ejercicio: Accidente del Challenger

Revisa la documentación del accidente

¿Crees que fue una buena investigación?

¿Crees que las recomendaciones fueron útiles?

□

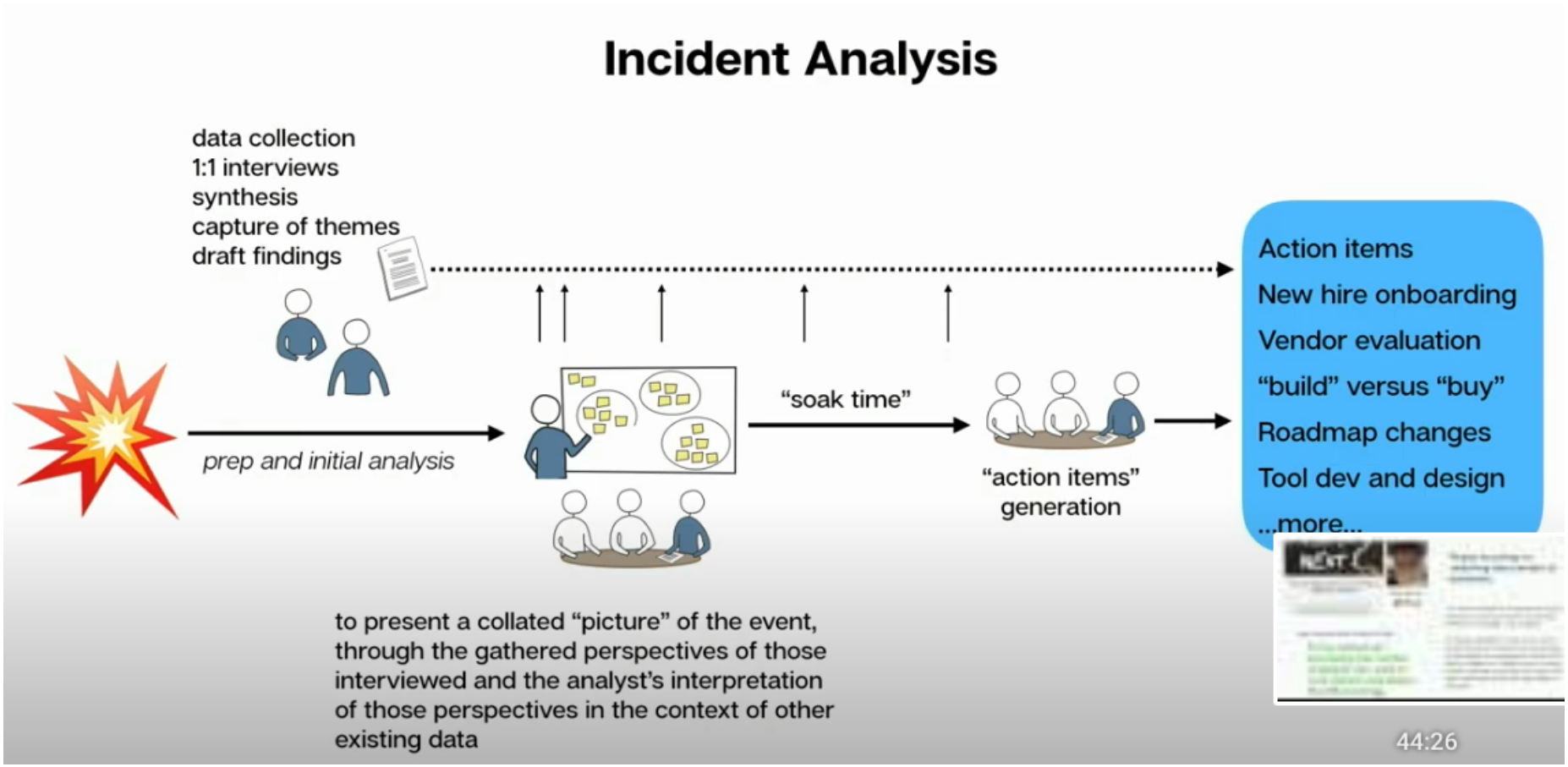
Excellent!

Facilitación del análisis

If the incident analyst participated in the incident, they will inevitably have a deeper understanding and bias towards the incident that will be impossible to remove in the process of analysis.

Ryan Kitchens, Netflix

Tareas post-incidente



Fuente: John Allspaw - What The Industry Misses About Incidents and What You Can Do

¿Cómo hacer un informe?

Un simple email puede bastar

Que no sea un tostón

Que sea fácil de leer

Con gráficos y datos

Informe público

Para enviar a la organización

Todo el mundo puede aprender

Amplía la audiencia a los afectados

Un fallo público suele aceptarse mejor

Planifica medidas

Resolución: resuelve problemas

Mitigación: reduce consecuencias

Prevención: evita incidentes

Monitorización: alerta de futuros problemas

Checklists

Revisa los procedimientos

Elabora una checklist

Cada procedimiento delicado debería tener su checklist elaborado en frío

Checklist para análisis de incidentes

Mantén scripts

El siguiente paso es automatizar las tareas

Un script permite ejecutar múltiples tareas

Los scripts deben estar repositados y catalogados

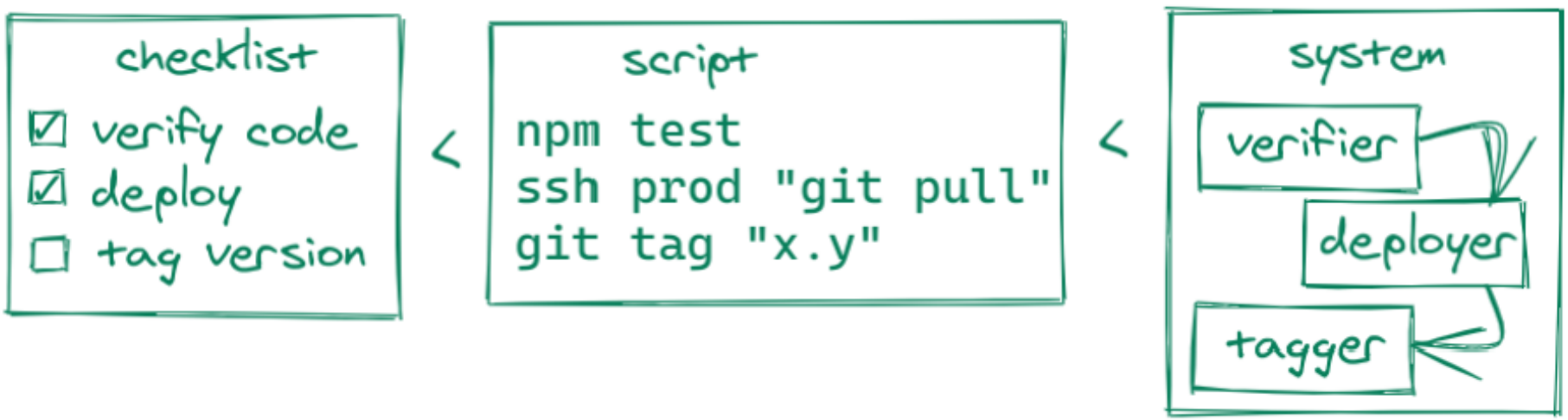
Crea sistemas resistentes

Un sistema robusto resiste los embates del azar

Intenta que tus sistemas se recuperen solos

Mejor un aviso el lunes que una alerta el sábado

Usa lo que más convenga



What you want to avoid at all cost is incident-driven development. Testing should be deeply enshrined in the genes of the organization. We want to do this in as structured a way as possible.

Edwin Hakkennes:

Automation makes testing better and more fun

Mal: Apagando fuegos todo el día

¡No confundir con IDD!

Bien: Incident-Driven Development

Los incidentes marcan el camino a seguir

Son incógnitas desconocidas: ¡acéptalo!

Intenta resolverlos a fondo

¿Qué puedo hacer para que no se repita?

Ejercicio: Accidente de Santiago

Lee la documentación del accidente

¿Cuáles fueron las causas?

¿Qué se podía haber hecho mejor?

□

We can do better!

Liderazgo

Consider listening to what an incident has to teach you. It's your job to figure out what that is.

John Allspaw, Incidents as we Imagine Them Versus How They Actually Are

Actitudes y expectativas

Cómo actúes en una crisis marcará el tono

La comunicación es lo más importante

Explica claramente qué buscas

Intenta sacar lo mejor de cada cual

*The problems begin when the pressure to **fix** outweighs the pressure to **learn**.*

Todd Conklin, via John Allspaw

Riesgo

Fija las **expectativas de riesgo**

Sin riesgo no hay beneficio
Sin riesgo no hay mejora

El único software "estable" es el que no cambia

Pablo Almunia: IEDGE – Proyectos de TI: Riesgos

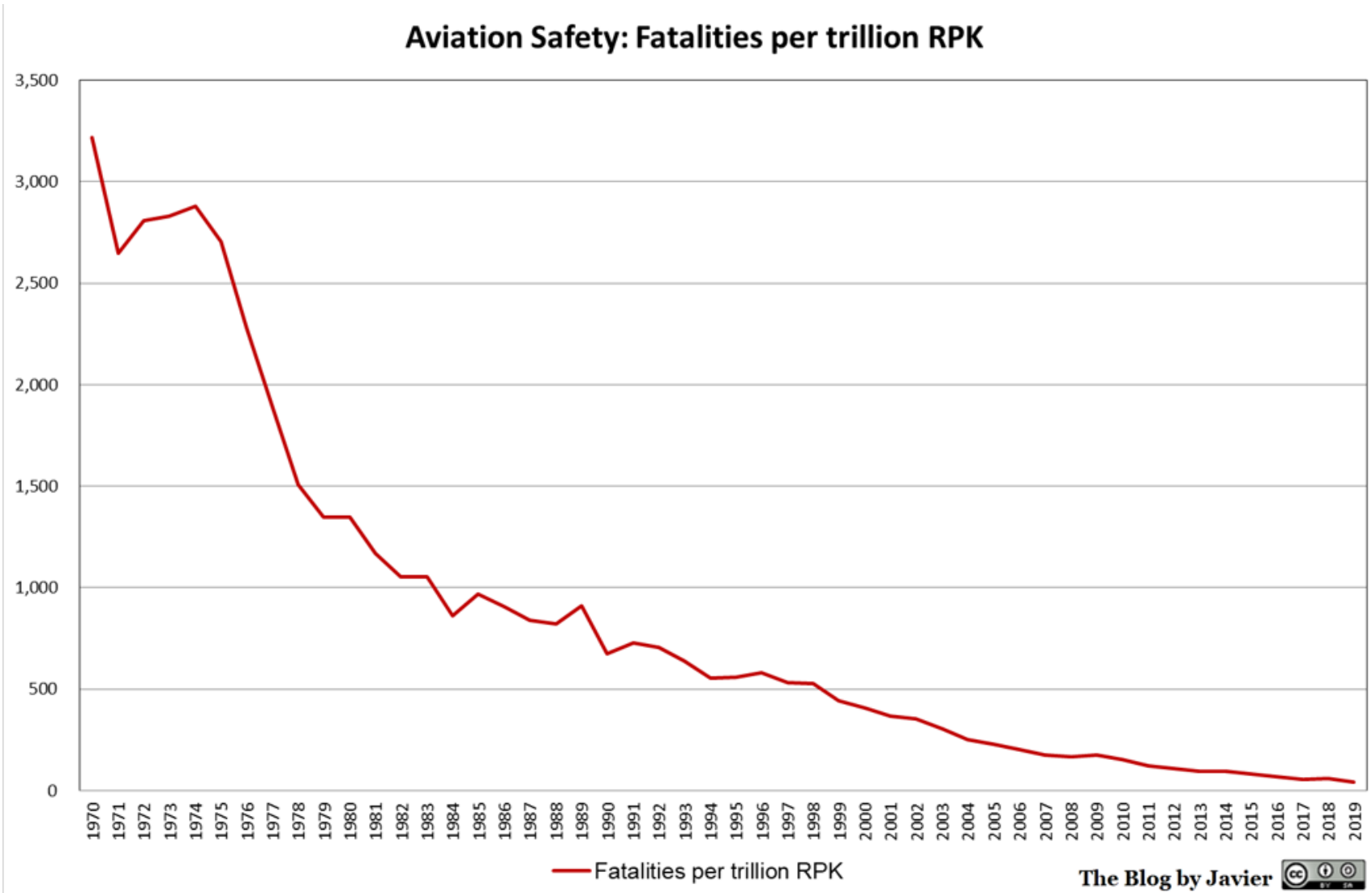
Calibración del riesgo

Según el tipo de sistema

Riesgo menor
Riesgo económico
Riesgo de impacto a clientes
...
Riesgo de vidas humanas

Elabora planes de contingencia si es necesario

¡Sí se puede!


Aviation Safety: Fatalities per trillion RPK
3,500
3,000
2,500
2,000
1,500
1,000
500
0
1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
Fatalities per trillion RPK
The Blog by Javier

Inversión

¿Cuánto tiempo del equipo estás dispuesto o dispuesta a dedicar a mejorar la estabilidad?

¿Prefieres pagar en horas de trabajo
o en "sangre" (alertas fuera del trabajo)?

Nota: la "sangre" no es gratis

Los incidentes afectan al desarrollo

Las planificaciones sufren

A no ser que tengas un equipo aparte
y ni por ésas

¡Y así es como debería ser!

Deja que guíen las mejoras del sistema

Ejercicio: Boeing 737 MAX

Revisa este vídeo de los incidentes

Pregúntale al ingeniero residente

Elabora un análisis del incidente

Sigue el checklist

□

Busted!

Bibliografía

John Allspaw: Blameless PostMortems and a Just Culture

The New Yorker: The Checklist

SRE Book*: Postmortem Culture: Learning from Failure

Ryan Kitchens: Characteristics of Next-Level Incident Reports in Software

Jacob Scott: Awesome Tech Postmortems

Recapitulemos

Sistemas distribuidos. Escalado y replicación

Pruebas de carga. Optimización de rendimiento

Uso de métricas. Incógnitas desconocidas

Monitorización y observabilidad

Análisis de incidencias