

Lección 2

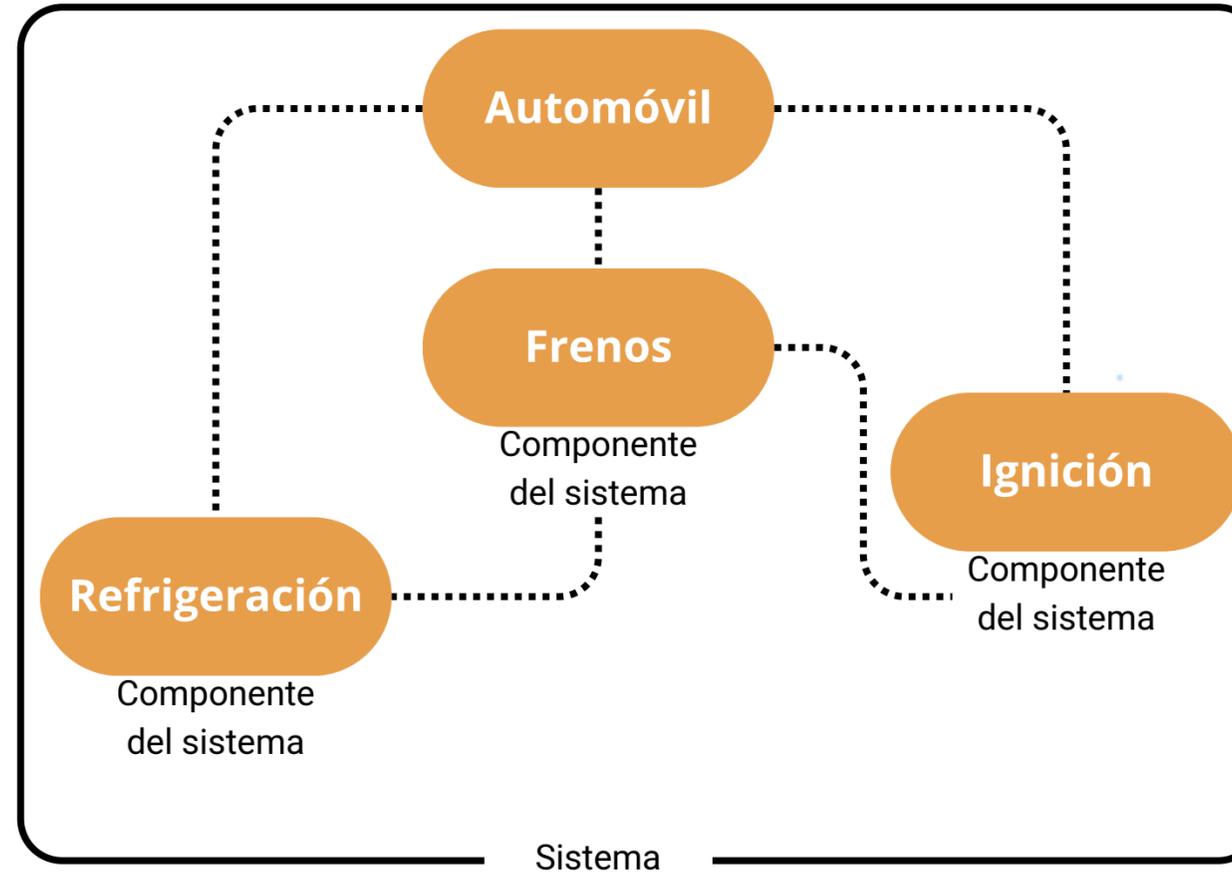
Fiabilidad y disponibilidad



En palabras de Werner Vogels, CTO de Amazon, “Todo falla, en todo momento”. Una de las prácticas recomendadas identificadas en el Marco de AWS Well-Architected es planificar los errores (o el tiempo de inactividad de las aplicaciones o cargas de trabajo). Una forma de hacerlo es diseñar las aplicaciones y cargas de trabajo para que resistan los errores. Hay dos factores importantes que los arquitectos de la nube tienen en cuenta a la hora de diseñar arquitecturas que resistan los errores: la fiabilidad y la disponibilidad.

Fiabilidad

- Una medida de la capacidad de su sistema para proporcionar
- funcionalidad cuando lo desea el usuario.
- En el sistema se incluyen todos los componentes del sistema: hardware, firmware y software.
- La probabilidad de que todo su sistema funcione según lo previsto durante un período específico.
- Tiempo promedio de errores (MTBF) = tiempo total en servicio/cantidad de errores

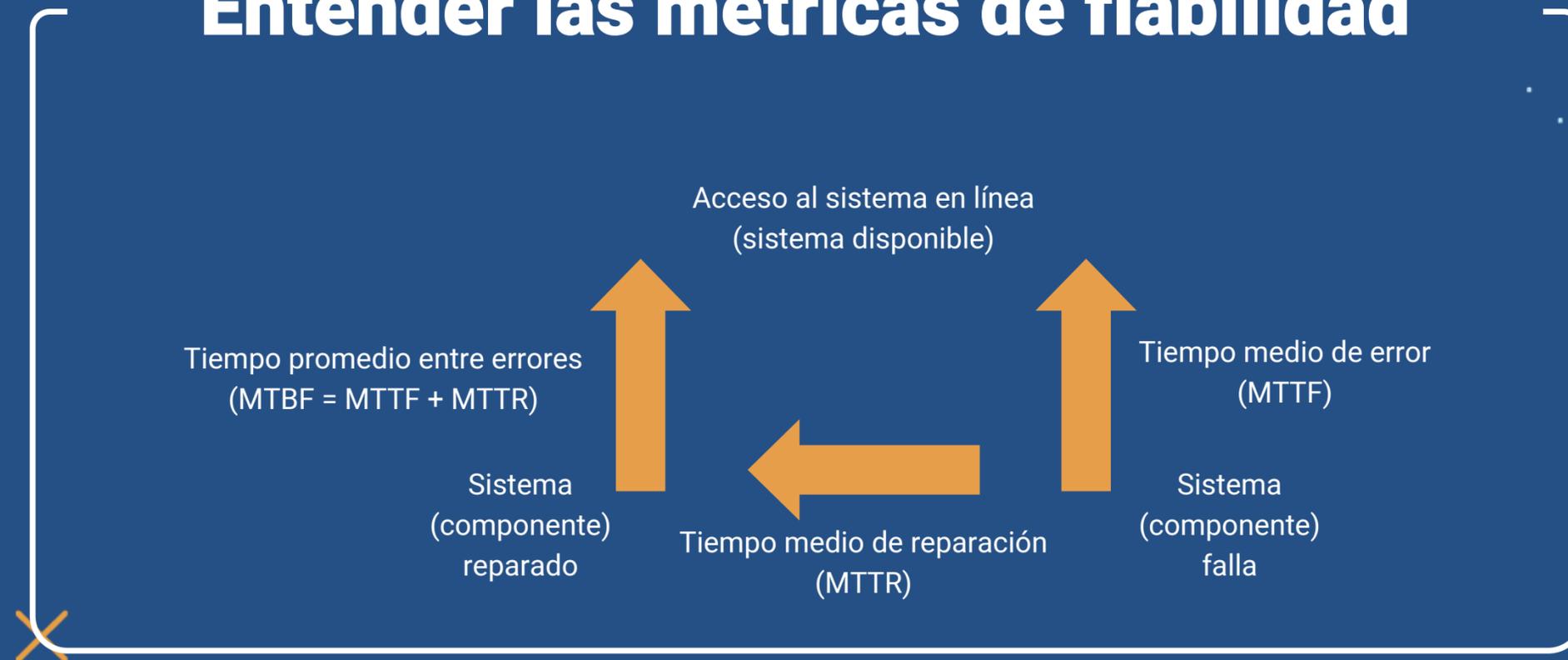


La fiabilidad es una medida de la capacidad del sistema para proporcionar funcionalidad cuando el usuario lo desea. Como “todo falla, en todo momento”, hay que pensar en la fiabilidad en términos estadísticos. La fiabilidad es la probabilidad de que todo un sistema funcione según lo previsto durante un periodo específico. Tenga en cuenta que en un sistema se incluyen todos sus componentes, como hardware, firmware y software. El error de los componentes del sistema afecta su disponibilidad.

Para entender la fiabilidad, es útil considerar el ejemplo familiar de un automóvil. El automóvil es el sistema. Cada componente del automóvil (por ejemplo, refrigeración, encendido y frenos) deben trabajar juntos para que el automóvil funcione correctamente. Si intenta arrancar el automóvil y el encendido falla, no podrá conducir a ninguna parte: el automóvil no está disponible. Si el encendido falla repetidamente, el automóvil no se considera fiable.

Una forma habitual de medir la fiabilidad es utilizar mediciones estadísticas, como el tiempo medio entre errores (MTBF). El MTBF es el tiempo total en servicio sobre el número de errores.

Entender las métricas de fiabilidad





Supongamos que tiene una aplicación que pone en línea el lunes a mediodía. Se dice que la aplicación está disponible. Funciona normalmente hasta que falla el viernes a mediodía. Por lo tanto, el tiempo hasta el error (o el tiempo que la aplicación está disponible) es de 96 horas. Desde el viernes a mediodía hasta el lunes a mediodía, usted se dedica a diagnosticar por qué falló en la aplicación y a repararla, momento en el que la vuelve a poner en línea. Por tanto, el tiempo de reparación es de 72 horas.



A continuación, vuelve a suceder: la aplicación falla el viernes a mediodía, usted pasa desde el viernes a mediodía hasta el lunes a mediodía reparándola y la vuelve a poner en línea el lunes a mediodía.

Digamos que este ciclo de error-reparación-restauración ocurre cada semana. Ahora puede calcular la media de estas cifras. En este ejemplo, nuestro tiempo medio hasta el error (MTTF) es de 96 horas y nuestro tiempo medio hasta la restauración (MTTR) es de 72 horas. El tiempo medio entre errores (MTBF) es de 168 horas (o 1 semana), que es la suma de MTTF y MTTR.

Disponibilidad

- Tiempo normal de operación/tiempo total
- Un porcentaje de tiempo de actividad (por ejemplo, 99,9 por ciento) a lo largo del tiempo (por ejemplo, 1 año)
- Número de 9: cinco 9 significan un 99,999 por ciento de disponibilidad

Como acaba de aprender, los errores de los componentes del sistema afectan a la disponibilidad del sistema.

Formalmente, la disponibilidad es el porcentaje de tiempo que un sistema funciona con normalidad o realiza correctamente las operaciones que se esperan de él (o tiempo de funcionamiento normal sobre el tiempo total). La disponibilidad se reduce cada vez que la aplicación no funciona con normalidad, incluidas las interrupciones programadas y no programadas.





La disponibilidad también se define como el porcentaje de tiempo de actividad (es decir, el tiempo que un sistema está en línea entre errores) durante un periodo de tiempo (normalmente 1 año).

Una abreviatura común para referirse a la disponibilidad es número de nueves. Por ejemplo, cinco nueves significa un 99,999 por ciento de disponibilidad.

Alta disponibilidad

- Un sistema puede soportar cierto grado de degradación y todavía estar disponible.
- El tiempo de inactividad se minimiza.
- Se requiere mínima intervención humana.

Un sistema de alta disponibilidad es aquel que puede soportar cierto grado de degradación sin dejar de estar disponible. En un sistema de alta disponibilidad, el tiempo de inactividad se minimiza al máximo y se requiere una mínima intervención humana.





Un sistema de alta disponibilidad puede verse como un conjunto de recursos compartidos por todo el sistema que cooperan para garantizar los servicios esenciales. La alta disponibilidad combina software con hardware estándar para minimizar el tiempo de inactividad al restaurar de forma rápida los servicios esenciales cuando falla un sistema, un componente o una aplicación. Los servicios se restauran con rapidez, con frecuencia en menos de 1 minuto.

Niveles de disponibilidad

Disponibilidad	Interrupción máxima (por año)	Categoría de la aplicación
99 %	3 días 15 horas	Trabajos de procesamiento por lotes, extracción, transferencia y carga de datos
99,9 %	8 horas, 45 minutos	Herramientas internas como administración del conocimiento, seguimiento del proyecto
99,95 %	4 horas, 22 minutos	Comercio en línea, punto de venta
99,99 %	52 minutos	Distribución de video, sistemas de difusión
99,999 %	5 minutos	Transacciones de ATM, sistemas de telecomunicaciones

Los requisitos de disponibilidad varían. El tiempo de interrupción aceptable depende del tipo de aplicación. A continuación se muestra una tabla de objetivos comunes de diseño de disponibilidad de aplicaciones y la duración máxima de interrupción que puede producirse en un año sin dejar de cumplir el objetivo. La tabla contiene ejemplos de los tipos de aplicaciones habituales en cada nivel de disponibilidad.

Para la accesibilidad: capas de disponibilidad con interrupciones máximas por año y por categorías de aplicación. Las capas van del 99 al 99,999 por ciento de disponibilidad.

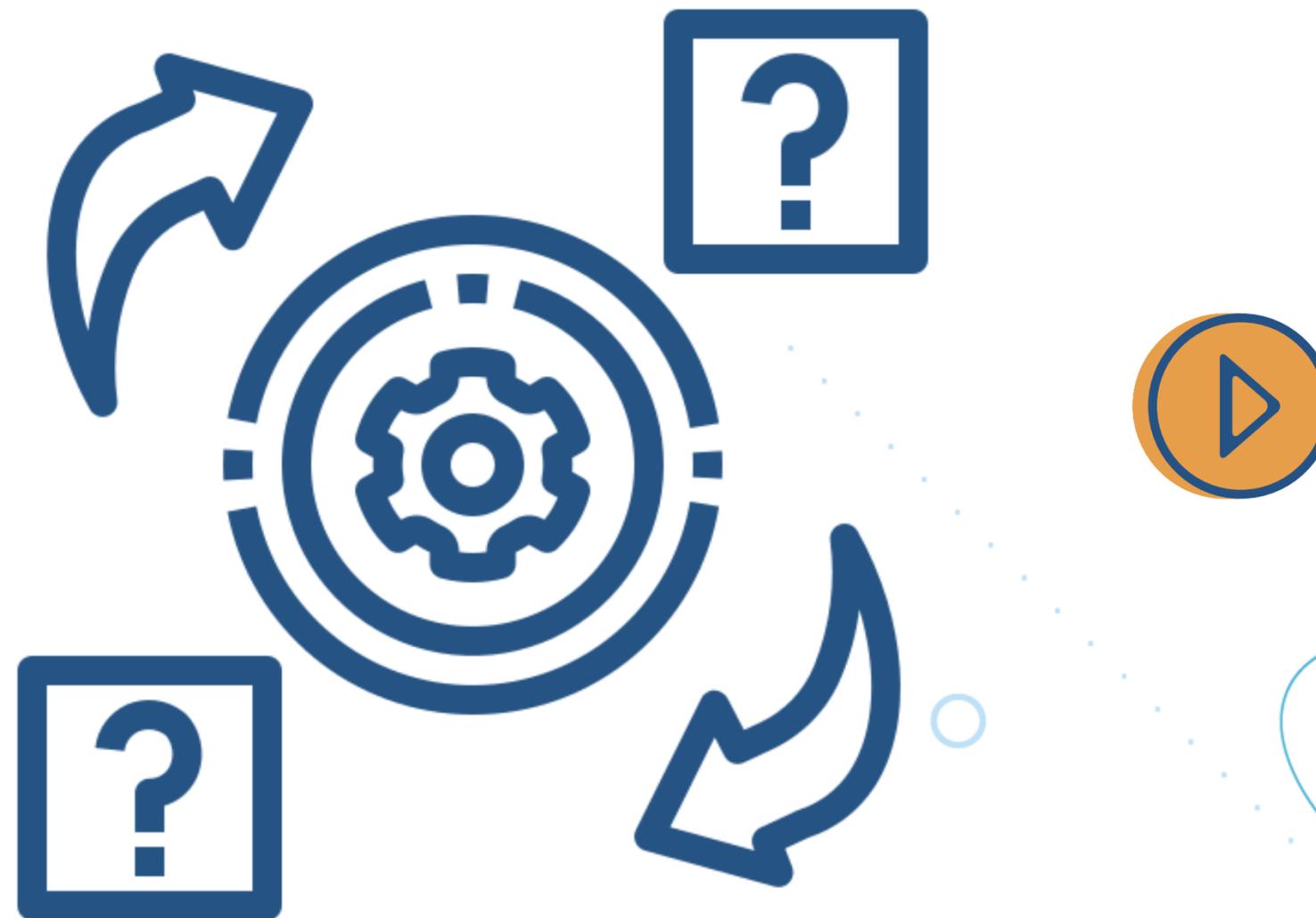
Factores que influyen la disponibilidad

Características de las políticas de control de servicios (SCP)

- Le permiten controlar a qué servicios pueden acceder los usuarios de IAM en cuentas miembro
- El administrador local no puede anular los SCP
- Las políticas de IAM definidas en cuentas individuales aún se aplican

Ejemplos de usos de SCP

- Crear una política que bloquee el acceso al servicio o acciones específicas
 - Ejemplo: impedir que los usuarios deshabiliten AWS CloudTrail en todas las cuentas miembro
- Crear un política que permita el acceso completo a servicios específicos
 - Ejemplo: permitir el acceso completo a Amazon ec2 Y cloudWatch
- Crear una política que imponga el etiquetado de los recursos



Aunque no siempre se pueden predecir los eventos que podrían interrumpir la disponibilidad de una aplicación, puede incorporar la disponibilidad al diseño de la arquitectura. Hay tres factores que determinan la disponibilidad general de la aplicación:

- Tolerancia a errores hace referencia a la redundancia integrada de los componentes de una aplicación y la habilidad de la aplicación de permanecer operativa, incluso si alguno de sus componentes falla. La tolerancia a errores se basa en un hardware especializado para detectar errores en un componente del sistema (como un procesador, una tarjeta de memoria, una fuente de alimentación, un subsistema de E/S o un subsistema de almacenamiento) y cambiar instantáneamente a un componente de hardware redundante. El modelo tolerante a errores no aborda errores de software, que es el motivo más frecuente de tiempo de inactividad.



- La escalabilidad es la capacidad de su aplicación para adaptarse a aumentos en las necesidades de capacidad, permanecer disponible y funcionar dentro de los estándares requeridos. No garantiza la disponibilidad de la aplicación, pero contribuye su disponibilidad.
- La recuperabilidad es la capacidad de restablecer el servicio rápidamente y sin pérdida de datos si se produce un desastre y hace que sus componentes no estén disponibles o destruye los datos.



Tenga en cuenta que la mejora de la disponibilidad suele conllevar un aumento de los costos. Cuando considere cómo aumentar la disponibilidad del entorno, es importante compensar los costos de la mejora con los beneficios para los usuarios.

Algunos puntos clave de esta lección:

- Fiabilidad es una medida de la capacidad de su sistema para proporcionar funcionalidad cuando lo desea el usuario que puede medirse en términos de MTBF.
- La disponibilidad es el porcentaje de tiempo que un sistema funciona con normalidad o realiza correctamente las operaciones que se esperan de él (o tiempo de funcionamiento normal sobre el tiempo total).
- Los tres factores que influyen la disponibilidad de las aplicaciones son la tolerancia a errores, la escalabilidad y la recuperación.
- Puede diseñar las cargas de trabajo y las aplicaciones para que tengan una alta disponibilidad, pero hay que considerar los costos.



INICIO