Perspectiva de disponibilidad y resistencia

USAC - Análisis y diseño de sistemas 2 - 1er semestre 2016
Ing. Ricardo Morales

Descripción

Calidad deseada

- La habilidad del sistema de estar total o parcialmente operacional y cuando sea requerido manejar efectivamente fallas que puedan afectar la disponibilidad del sistema

Aplicabilidad

- Cualquier sistema que tiene requerimientos complejos o extendidos de disponibilidad, procesos de recuperación complejos o un alto perfil (visible al público)

Descripción (II)

Temas de interés

- Clases de servicio
- Tiempo de baja planeado
- Tiempo de baja no planeado
- Tiempo de reparación
- Recuperación de desastres

Actividades

- Capturar requerimientos de disponibilidad
- Crear planificación de escalabilidad
- Estimar disponibilidad de la plataforma
- Estimar disponibilidad funcional
- Evaluar contra los requerimientos
- Trabajar nuevamente la arquitectura

Descripción (III)

Tácticas de arquitectura

- Seleccionar hardware tolerante a fallos
- Usar cluster de alta disponibilidad y balanceo de carga
- Aplicar soluciones de disponibilidad de software
- Seleccionar o crear software tolerante a fallos
- Diseñar para fallas
- Permitir replicación de componentes
- Reutilizar recursos y resultados
- Relajar consistencia transaccional
- Identificar soluciones de backup y recuperación de desastres

Problemas

- Un único punto de falla
- Fallas en cascada
- No disponibilidad por sobrecarga
- Requerimientos de disponibilidad demasiado ambiciosos
- Detección no efectiva de errores
- Sobre estimación de resistencia de componentes
- Requerimientos de disponibilidad globales no vistos
- Tecnologías incompatibles

Descripción (IV)

- El horario de disponibilidad de un sistema depende del negocio al que da soporte un sistema, de manera que sistemas usado globalmente tendrán un requerimiento de disponibilidad de 24 horas.
- Es importante considerar los costos del negocio de no estar disponible, para compararlos con los costos de tener una alta disponibilidad
- Esta perspectiva es importante para cualquier sistema que tenga requerimientos complejos de disponibilidad y resistencia o que es visible al público de alguna forma

Aplicabilidad a vistas

- Al considerar disponibilidad y resistencia es poco probable que resulte en muchos cambios a la vista de contexto, aunque puede causar considerar como es afectada la disponibilidad del sistema por la disponibilidad de sistemas externos, lo que podría cambiar la forma de interactuar con ellos

Contexto

- La disponibilidad es un tema de interés clave para el usuario porque puede impactar la habilidad del negocio de operar efectivamente
- A veces se pueden requerir cambios funcionales para soportar requerimientos de disponibilidad, tales como la habilidad de operar en modo fuera de línea cuando una red de comunicaciones no está disponible

Funcional

Aplicabilidad a vistas (II)

- Una consideración clave para la disponibilidad es el conjunto de procesos y sistemas para backup y recuperación
- Los sistemas deben tener respaldos de manera que puedan ser recuperados en un tiempo razonable si ocurre un desastre
- Los backups no deben impactar la disponibilidad en línea

Información

- Características como replicación de hardware y failover en el sistema pueden implicar cambios en el modelo de concurrencia

Concurrencia

Aplicabilidad a vistas (III)

Desarrollo

- El enfoque para alcanzar disponibilidad puede imponer restricciones de diseño a los módulos de software
- Por ejemplo, todos los subsistemas deben soportar comandos de inicio, detener y pausar para alinearse a la estrategia de failover

Deployment

- La disponibilidad y resistencia pueden tener un gran impacto en el ambiente de deployment
- Los requerimientos de disponibilidad pueden obligar un ambiente de producción tolerante a fallas o un sitio separado para recuperación de desastres que pueda ser activado rápidamente

Operacional

- Procesos y mecanismos que permitan la identificación y recuperación de problemas en el ambiente de producción pueden ser requeridos
- También puede existir necesidad para sitios de recuperación de desastres separados geográficamente

Temas de interés

- Clases de servicio
  - Cuando se piensa acerca de tiempo no disponible, no se debe restringir a un modelo de servicio binario disponible/ no disponible. A menudo es apropiado considerar diferentes niveles de servicio, en un espectro desde servicio completo a ningún servicio
- Tiempo de baja planeado
  - En la práctica, todo sistema computacional requiere un tiempo de baja ocasional para instalar hardware, actualizaciones de software o sistema operativo, o hacer tareas fuera de línea como backup
  - Ese es el tiempo planeado de baja, porque ocurre de acuerdo a una planificación que se alinea con los requerimientos del negocio (generalmente por las noches o fines de semana)

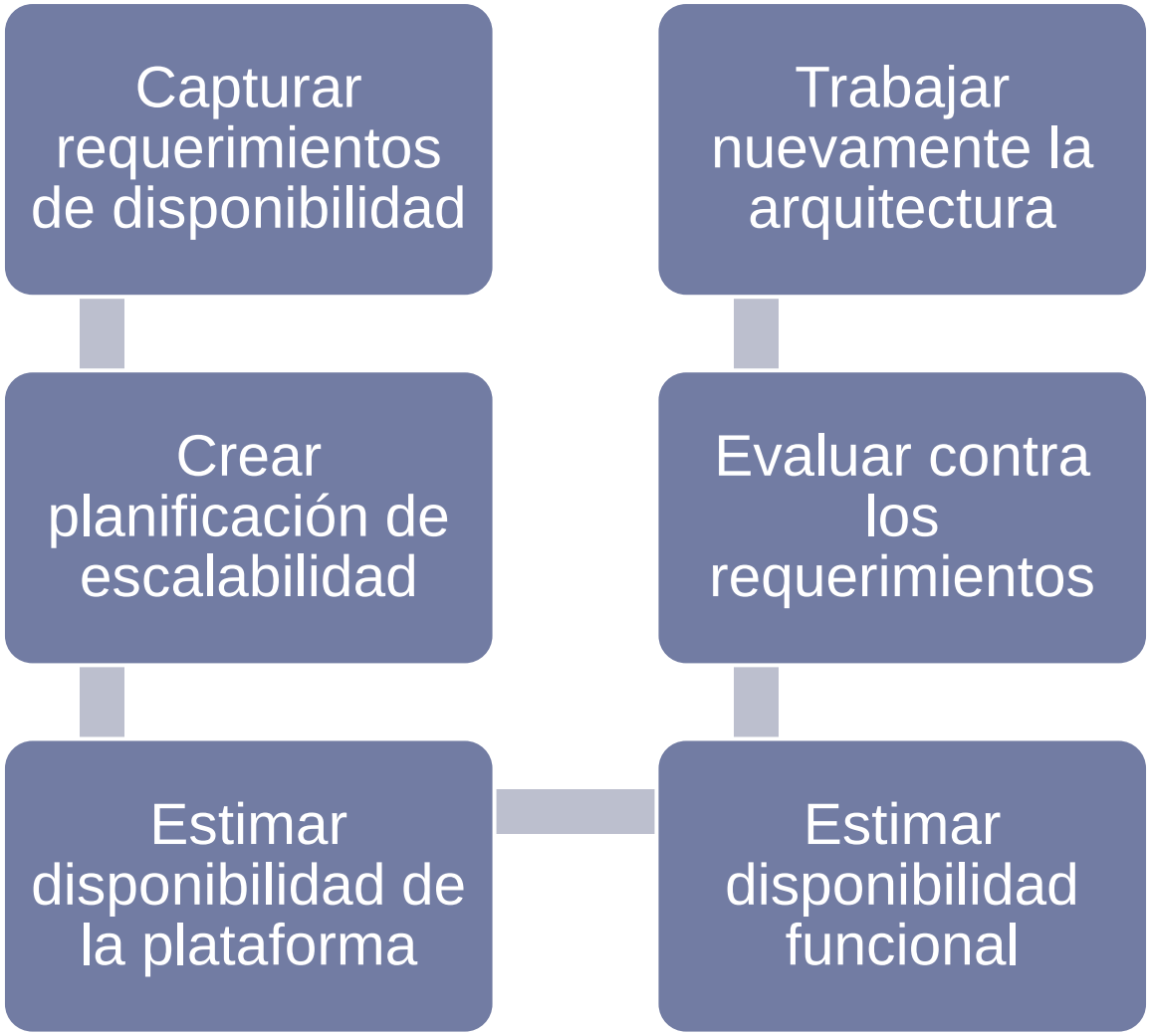
Temas de interés (II)

- Tiempo de baja no planeado
  - Generalmente ocurre por una falla de hardware o software que hace al sistema no disponible
  - Es difícil predecir el tiempo y frecuencia, pero se debe hacer un análisis para identificar si es necesario establecer procedimientos de contingencia (manuales o semiautomáticos) que permitan al negocio continuar
- Tiempo de reparación
  - Una falla es solo la mitad del problema, en cuanto a disponibilidad, la otra mitad es cuanto tiempo toma rectificar la falla
  - En el caso del hardware puede implicar cambiar un componente que ha fallado, lo que implica reiniciar parte o todo del sistema
  - En el caso de software, el tiempo de reparación es mas difícil de estimar porque probablemente se requiera algún análisis del problema, reparación, pruebas y publicación

Temas de interés (III)

- Recuperación de desastres
  - Si un sistema critico falla completamente o su ambiente físico operacional no está disponible, un proceso completo de recuperación de desastres es requerido para restaurar el servicio
  - Esto puede implicar la recreación de todo el ambiente del sistema, incluyendo hardware, red de comunicaciones y plataforma de software, así como software de aplicación y sus datos
  - Aunque el ambiente puede ser reemplazado, los datos no, de manera que el modelo debe contemplar la recuperación de datos en un desastre, cuantos datos puede perder el negocio y que tanto tiempo puede sobrevivir el negocio sin datos recuperados o reparados
  - El personal de administración de infraestructura también es un recurso a considerar

Actividades



Escenarios de recuperación de incidentes

Incident	Impact	Remedial Action	Time to Repair
Hardware (non-disk) failure	Reduced availability (throughput affected)	Replace the faulty component, and possibly reconfigure the hardware or the operating system	1 hour
Single disk failure	Performance degradation during automated recovery after failed disk is replaced in disk array	Failed disk must be replaced to allow automatic data recovery by disk array	1 hour to replace disk and recover
Disk array failure	Total unavailability (service offline)	Replace the faulty disk array, and possibly recover data from backup and/or other means	6 hours including restore
Nontransient network failure	Temporary service outage, in-flight transactions aborted, some loss of throughput when using backup network	Switch over to the standby network, possibly with reduced bandwidth	5 minutes
Operating system crash	Temporary service outage, normal availability within 5–10 minutes	Reboot, although the crash may be symptomatic of some other problem (such as faulty memory or disk) that needs to be addressed	5 minutes for reboot
Data corruption	Affected accounts and transactions unavailable, failure of end-of-period reconciliation	Recover data from backup and possibly other means (such as replaying transaction logs)	6 hours
Application software failure	Variable, from temporary outage (10 minutes) to total outage due to corruption	Restart, depending on the application and the nature of the failure; may be necessary also to recover data in some way	Not quantified

Tácticas de arquitectura

- Seleccionar hardware tolerante a fallos
  - Las plataformas de computación tolerantes a fallas pueden continuar operando aún si falla un componente de hardware. Esto se implementa con la redundancia o duplicación de hardware
  - En una plataforma tolerante a fallas, la disponibilidad es tan buena como lo sea el componente mas débil
- Usar cluster de alta disponibilidad y balanceo de carga
  - Un cluster de alta disponibilidad es una técnica para protegerse contra fallas al duplicar toda una máquina
  - Dos o mas computadoras idénticas (nodos) son desplegadas en paralelo y comparten la carga, si alguna falla, la carga se distribuye en las restantes
  - A través de una técnica, llamada balanceo de carga, se ejecuta la distribución de carga en los nodos, buscando que los nodos sean utilizados lo mas posible

Tácticas de arquitectura (II)

- Logs de transacciones
  - Es posible que los backups permitan recuperar datos hasta cierto punto del tiempo, debido al momento en que fue realizado el backup
  - Para recuperar transacciones adicionales, podría ser de utilidad un log de transacciones que es aplicado a los datos recuperados del backup
- Aplicar soluciones de disponibilidad de software
  - Algunos sistemas críticos alcanzan la tolerancia a fallos cl multiplexar (ejecución de la aplicación varias veces y usar un sistemas de votos en tiempo real para detectar inconsistencias en las salidas)
  - El costo de ese nivel de sofisticación no siempre es permisible para un proyecto

Tácticas de arquitectura (III)

- Seleccionar o crear software tolerante a fallos
  - El software puede ser escrito para que se reconfigure a si mismo en condiciones cambiantes, por ejemplo, obteniendo mas recursos (como memoria) cuando aumenta la carga o deshabilitar ciertas funciones si fallan, ofreciendo un servicio parcial
- Diseñar para fallas
  - Por muy confiable que sea el software, existe la probabilidad de una falla, por lo que se deben tener procesos en el software para reconocer fallas y recuperarse de ellas
- Permitir replicación de componentes
  - Para que una replicación de componentes funcione bien, se debe considerar eso en su diseño
  - La replicación de un componente no diseñado para eso puede causar que se congele, duplique datos o incluso corrompa datos

Tácticas de arquitectura (IV)

- Identificar soluciones de backup y recuperación de desastres
  - Cada sistema que maneja información persistente debe incluir mecanismos de backup de la información para que pueda ser recuperada en caso de falla
  - Tradicionalmente se ha utilizado mecanismos de cinta magnética para backup, aunque la velocidad y cantidad de información a almacenar pueden ser un problema