

Gestión de redundancia y alta disponibilidad

Breve descripción:

Garantizar la disponibilidad de los servicios de Tecnologías de la Información (TI), es fundamental para posibilitar la operación de las empresas, bien sea porque se prestará un servicio y se desea dotar de estas características o porque se van a adquirir servicios en la nube; por este motivo, es importante conocer los conceptos y términos a este proceso de gestión.

Abril 2024

Tabla de contenido

Introducción	3
1. Gestión de redundancia y alta disponibilidad	4
1.1. Clústeres	5
1.2. Ventajas de la gestión de redundancia y alta disponibilidad	7
2. Implantación de sistemas de alta disponibilidad.....	9
2.1. Indicadores de nivel de servicio	9
2.2. Métodos de cálculos de los indicadores.....	10
Síntesis	15
Material complementario	16
Glosario	17
Referencias bibliográficas	18
Créditos	19

Introducción

En el momento de contratar u ofrecer servicios en la nube, seguramente se debe asesorar o recibir el asesoramiento sobre un conjunto de características o configuraciones que permiten garantizar con mayor seguridad, la operación de los servicios que se van a contratar. No solo se trata de garantizar el almacenamiento de la información, sino la de aquellos servicios que la publican y la procesan.

El presente componente formativo, le permitirá conocer los conceptos relacionados con procesos de gestión de la redundancia y alta disponibilidad, de forma que se comprenda el contexto de los servicios que se ofrecen, y a su vez, también se trabajará la temática de la implantación de sistemas de alta disponibilidad.

Bienvenido.

1. Gestión de redundancia y alta disponibilidad

Para conocer sobre gestión de redundancia y alta disponibilidad, es importante identificar que la primera se centra en la capacidad de interconectar los “data centers” en caso de que el principal falle, mientras que la alta disponibilidad se centra en la capacidad de tener operando los sistemas sin interrupción. En ocasiones, se ofrecen servicios de alta disponibilidad (HA, de inglés “High Availability”) y para ello, es indispensable iniciar con el “backup” de los datos, porque la alta disponibilidad no conlleva copia de la información.

A continuación, se profundizará en estos conceptos:

Redundancia de “hardware”

Hace referencia a tener un componente del sistema duplicado. En esta instancia, se puede considerar que los proveedores de servicios en la nube tienen todos esos servicios, que ofrecen operando dentro de Centros de Procesos de Datos (CPD) o más conocidos como “Data Centers”, y las máquinas en algún momento terminan fallando, porque tienen un tiempo de vida estimado muy difícil de garantizar.

Su objetivo no es evitar fallas sino, que no afecten a los usuarios. La forma de lograr esto, es duplicando el dispositivo de manera que cuando falle la máquina los servicios que dependen de ese “hardware”, sean trasladados de forma automática a un dispositivo que no esté afectado, de tal forma que cuando dos dispositivos están publicados, se puede decir que están redundados.

Alta Disponibilidad: HA (“High Availability”)

Son las características que tiene una arquitectura tecnológica y que determina el grado con el que los servicios TI están activos, es decir, disponibles, para ser usados por

el cliente final. La alta disponibilidad garantiza que los sistemas estén disponibles a pesar de que existan fallos en las tecnologías que los soportan.

Operación de los servicios

La alta disponibilidad puede ser de dos formas, para poder mantener en operación los servicios:

Activo-Activo: ambos equipos, dispositivos o sistemas, están funcionando simultáneamente y generalmente se están repartiendo la carga de trabajo entre ellos.

Activo-Pasivo: solo uno de los dispositivos o servicios está en operación, mientras que el segundo equipo, solamente entra en operación, cuando el primero se ha “caído” o no está disponible. En este caso es posible utilizar una copia existente (de un momento determinado de tiempo) de los datos.

Como se puede inferir, para tener HA se debe tener como mínimo dos sistemas o dispositivos y ambos con idénticos sistemas y características; generalmente se aplica HA a servidores de aplicaciones y a sistemas gestores de base de datos.

En conclusión, el término redundancia, está vinculado a la implementación de alta disponibilidad, específicamente mediante la implantación de al menos un dispositivo adicional que se empleará como respaldo (según el modelo activo-pasivo) o balanceador de carga (en el caso activo-activo).

1.1. Clústeres

Cuando la HA es activo-activo, se puede tener mejor desempeño al existir balanceo de cargas, a este tipo de agrupación se le denomina clúster, que es el nombre que se le da a un sistema compuesto de más de una unidad de procesamiento, que trabajan de manera unificada con el mismo objetivo. Estas máquinas tienen en común

que realizan las mismas tareas, debido a que tienen una similitud en la configuración y los sistemas informáticos, de manera que pueden trabajar de forma simultánea.

La técnica del clúster se puede aplicar a:

- Máquinas virtuales.
- “Containers”.
- Arquitectura de computación.
- Redes.
- Computación en la nube.
- Banco de datos.

El funcionamiento del clúster requiere conocer el concepto llamado “nodo”, el cual es el nombre que atribuye a cada computador o unidad de procesamiento (máquina virtual o contenedor) agregado a un clúster, ya sea virtual o físico. Cuando se hace referencia a las tecnologías o herramientas necesarias para implementar los clústeres, se habla de “clustering”.

En un clúster los nodos se interconectan gracias a una tecnología o infraestructura de red, generalmente una ya conocida por la empresa, pensando en la facilidad de mantenimiento y control de los costos. Es así, como cada nodo desempeña la misma función que los demás nodos, siendo posible suprimir o adicionar nuevos cuando sea necesario, aún si el clúster se mantiene en funcionamiento sin interrupciones, de tal forma que estas operaciones de agregar o quitar no perjudican el proceso, porque las tareas se distribuyen automáticamente, sin ocasionar problemas.

Para que el sistema sea escalable es importante saber que no existe límite de nodos que se puedan operar en un solo clúster y esto es muy importante cuando las empresas están en proceso de crecimiento continuo.

Básicamente existen dos tipos de clúster o motivos por los que se configuran los clústeres:

Clúster de alto desempeño (“High Performance Computing Cluster”)

Se utiliza cuando el objetivo es resolver problemas que requieren de mucho procesamiento concurrente, es decir, simultáneo, permitiendo realizarlos en un tiempo útil y de manera que satisfaga la expectativa del usuario. Para lograr esto se hace uso de la técnica de balanceo de carga (“Load Balancing”), que consiste en repartir las tareas de manera similar o equilibrada (balanceada) entre los nodos.

Clúster de alta disponibilidad (“High Availability Computing Cluster”)

Su principal objetivo es mantener el sistema en funcionamiento, es decir, está centrado en saber responder automáticamente a fallos sin afectar la continuidad de la prestación del servicio. Para esto, es necesario dotar al clúster de herramientas control y monitoreo de fallas en la interconexión o en los nodos, redundancia de sistemas, sistemas sustitutos o de respaldo, entre otros.

1.2. Ventajas de la gestión de redundancia y alta disponibilidad

Existen varios motivos por lo que es importante la gestión de la redundancia y alta disponibilidad; a continuación, se relacionan los más representativos que contribuyen a adopción de estrategias para su gestión:

- **Eliminación de fallos**

Al ocurrir fallos en el “hardware”, en las redes, energía o de nivel físico, quedan resueltos de forma automática, debido a que las funciones son trasladadas a nodos no afectados.

- **Recuperación automática**

No requiere de la intervención humano–máquina, para que los sistemas se restablezcan o se reconfiguren, para salvaguardar los fallos.

- **Rendimiento**

Estas arquitecturas permiten que la empresa pueda crecer en operaciones y que los sistemas tecnológicos escalen las necesidades crecientes.

- **Seguridad**

Se puede garantizar el almacenamiento y disponibilidad de la información.

2. Implantación de sistemas de alta disponibilidad

La implantación requiere de algunos lineamientos, independiente de la tecnología de “clustering” o de almacenamiento, se centra en que sea servible para las tecnologías actuales y futuras dentro de una organización.

A continuación, conozca las siguientes definiciones:

Definición de objetivos

El objetivo es analizar las distintas opciones de alta disponibilidad, y con esto establecer su intencionalidad, según lo que se determine como importante, pero haciendo un análisis de riesgos en sistemas críticos.

Para su definición, se deben tener en cuenta elementos como el empleo de contenedores, el almacenamiento redundante, la duplicidad de las redes de telecomunicaciones, la virtualización de sistemas, la operación ininterrumpida y el balanceo de carga.

Análisis de opciones de alta disponibilidad

Son necesarios varios aspectos al momento de considerar la implementación de un sistema de alta disponibilidad o para adquirirlo desde un proveedor. Por eso se establecen algunos lineamientos, iniciando a partir de los indicadores de nivel de servicio.

2.1. Indicadores de nivel de servicio

El funcionamiento del “hardware” sin fallas es la base para dotar a un sistema una alta disponibilidad. Para saber cómo presentar el valor de alta disponibilidad a ofrecer, se debe tener en cuenta la norma IEEE 762/2006 que suministra el método del

cálculo de este indicador para sistemas eléctricos, pero es aplicable a todo sistema eléctrico o electrónico.

En dicha norma, se mencionan algunos indicadores como lo son: fiabilidad, disponibilidad, tiempo medio entre paradas (MTBF, “Mid Time Between failures”), duración de las paradas (MTTR, o “Mid Time To Repair”), número de paradas por mantenimiento y tiempo total perdido por mantenimiento.

Para poder realizar este cálculo, se necesita tener registrado en una tabla, un registro por cada parada de los servicios, con los siguientes datos:

- Servicio que ha parado.
- Motivo de la parada (mantenimiento, actualización, falla, etc.).
- Duración de la parada (en minutos o segundos).

2.2. Métodos de cálculos de los indicadores

A continuación, se describen cuatro métodos de cálculo:

- **Disponibilidad**

Es el cociente entre el tiempo que el sistema está disponible y el tiempo total de paradas de los servicios por servicios de mantenimiento programado y notificado o coordinado con el cliente. Para calcularlo, es necesario conocer el tiempo disponible como resta entre el tiempo total por las paradas de mantenimiento previamente programado y el tiempo por parada no programada. Una vez obtenido se divide el resultado entre el tiempo total del periodo considerado.

La disponibilidad es igual a las horas totales disponibles menos las horas totales de paradas por mantenimiento programado, sobre las horas totales.

- **Fiabilidad**

La fórmula es muy parecida a la anterior, con las diferencias que se sustituye en el numerador las horas de parada por mantenimiento por horas de parada por mantenimiento no programado.

La fiabilidad es igual a las horas totales disponibles menos las horas totales de paradas por mantenimiento NO programado, sobre las horas totales.

- **Tiempo medio entre paradas (TMEP)**

Es el tiempo medio que ha transcurrido entre dos paradas de mantenimiento, y se requiere para su cálculo en el numerador las horas totales del periodo, y en denominador, el número de paradas.

El TMEP es igual a las horas totales del periodo, sobre la cantidad de paradas.

- **Tiempo medio hasta puesta en marcha (TMPM)**

Representa el tiempo medio de duración de las diversas paradas ocurridas en el periodo e ítem analizado.

El TMPM es igual a las horas totales paradas, sobre la cantidad de paradas.

La implantación de sistemas de alta disponibilidad, requiere de un funcionamiento ininterrumpido y para garantizarlo, el “hardware” del sistema obligatorio debe exigirle a un sistema una alta disponibilidad, para lo cual es necesario mantenerlo en operación:

- **Condiciones ambientales adecuadas**

Para poder tener estas condiciones (temperatura, humedad, circulación de aire), se emplean instalaciones o sitios específicos para ubicar los equipos, donde se centraliza todo el funcionamiento.

- **Condiciones óptimas de suministro energético**

Para lo cual se suelen emplear técnicas de instalación tales como, duplicar las fuentes de alimentación de los equipos y duplicar los equipos de interconexión, incluyendo cableado.

Además de estas condiciones, es fundamental conocer la integridad, respaldo y los servidores redundantes; a continuación, se relaciona información importante para este proceso:

- **Integridad y respaldo de los datos**

Mediante la integridad de datos, se puede asegurar que no ha sido modificada por terceros. Para garantizar la integridad en los datos almacenados o transmitidos, se emplean técnicas como cifrado o firmas digitales.

Garantizar esto es un poco más difícil, generalmente es algún “malware” o alguna función interna de algún sistema operativo que tenga un bache de seguridad.

Sin embargo, existen herramientas para detectar este tipo de vulnerabilidades, como “Rootkit Hunger”, por citar un ejemplo.

- **Servidores redundantes**

Se evalúa la aplicación de clúster de servidores, máquinas virtuales o contenedores, y se selecciona la técnica y el número de nodos de acuerdo

con las características que se desea dotar al servicio, si estará centrada en el alto desempeño o alta disponibilidad.

- **Elección de tecnología y/o proveedor de servicio**

De acuerdo al análisis al tipo de nodos (máquinas virtuales, contenedores o equipos físicos), se ha de seleccionar o la tecnología a emplear o el proveedor en la nube que cumpla con las características deseadas.

Muchas veces el costo de licenciamiento o el coste de aprovisionamiento limitan lo deseado, en tal caso se recomienda dimensionar el alcance, pero considerar una estimación del crecimiento de manera que los sistemas se puedan escalar considerando los costes que esto conlleve.

- **Configuración de servicios y puesta en marcha**

En este punto del plan de implementación, independiente de la tecnología o el proveedor a seleccionar, es de vital importancia considerar el plan de capacitación técnica, de manera que la configuración seleccionada aproveche al máximo o el licenciamiento o el nivel de servicios que se han contratado, de otra forma es posible que se desplieguen con servicios con una configuración estándar o genérica que no explote las características contratadas o adquiridas.

- **Pruebas de carga**

Las pruebas de carga, miden el rendimiento del “software” y siguen dos tipos de procedimientos:

- a) **Pruebas de caja negra:** son pruebas sencillas que no se dotan de permisos de usuario para acceder, solo se puede probar lo que externamente se ve de un servicio, sin emplear las credenciales de acceso que permitan acceder a toda funcionalidad.

b) **Pruebas de caja blanca:** interesa verificar internamente las partes que componen un sistema, de manera que debe seguir una secuencia lógica, la primera sería acceder al sistema con credenciales de acceso válidas.

Se pueden hacer varias pruebas de carga según uno de estos dos métodos:

- **Pruebas reales**

Con determinado número de usuarios concurrentemente usando las funcionalidades del sistema.

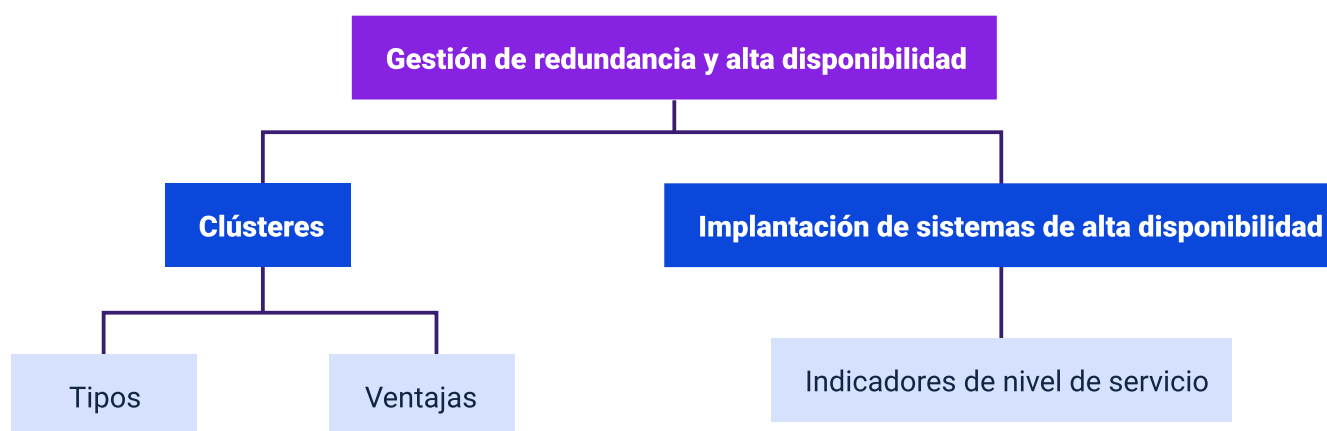
- **Pruebas sintéticas**

Se emplea un sistema o programa especialmente diseñado para tales fines como por ejemplo el sistema JMeter.

Es recomendable seguir unos lineamientos metodológicos previamente establecidos, y al momento de ser probados, se debe dotar a los sistemas de alta disponibilidad, tanto para servicios locales o servicios a contratar, para que sean desplegados en la nube.

Síntesis

A continuación, se presenta una síntesis de la temática estudiada en el componente formativo:



Material complementario

Tema	Referencia	Tipo de material	Enlace del recurso
Gestión de redundancia y alta disponibilidad	IBM Docs. (2021). Componentes de la alta disponibilidad.	Artículo	https://www.ibm.com/docs/es/i/7.5?topic=overview-components-high-availability

Glosario

Clúster: servidores unidos entre sí normalmente por una red de alta velocidad y que se comportan como si fuesen un único servidor.

Disponibilidad: cualidad o condición de disponibilidad.

Sistemas redundantes: son aquellos en los que se repiten aquellos datos o hardware de carácter crítico, que se quiere asegurar ante los posibles fallos que puedan surgir por su uso continuado.

Referencias bibliográficas

Apache Software Foundation. (2021). Apache JMeter™. <https://jmeter.apache.org>

Axioma B2B Marketing. (2021). Fórmulas de cálculo de indicadores de disponibilidad. <https://www.reporteroindustrial.com/blogs/Formulas-de-calculo-de-indicadores-de-disponibilidad+115450>

HP, Oracle, Cisco (1999). High Availability & Load Balancing Testing using Oracle Application, Cisco Local Director and HPWebQos.

IBM. (2021). Soluciones SAN.

OmniPlayer. (2019). Redundancia y alta disponibilidad. <https://omniplayer.com/es/administracion/redundancia-y-alta-disponibilidad>

Créditos

Nombre	Cargo	Centro de Formación y Regional
Milady Tatiana Villamil Castellanos	Líder del Ecosistema	Dirección General
Olga Constanza Bermudez Jaimes	Responsable de Línea de Producción	Centro de Servicios de Salud - Regional Antioquia
Peter Emerson Pinchao	Experto temático	Centro de Teleinformática y Producción Industrial - Regional Cauca
Ana Catalina Córdoba Sus	Evaluadora Instruccional	Centro de Servicios de Salud - Regional Antioquia
Blanca Flor Tinoco Torres	Diseñador de Contenidos Digitales	Centro de Servicios de Salud - Regional Antioquia
Luis Jesús Pérez Madariaga	Desarrollador Fullstack	Centro de Servicios de Salud - Regional Antioquia
Edgar Mauricio Cortés García	Actividad Didáctica	Centro de Servicios de Salud - Regional Antioquia
Jaime Hernán Tejada Llano	Validador de Recursos Educativos Digitales	Centro de Servicios de Salud - Regional Antioquia
Margarita Marcela Medrano Gómez	Evaluador para Contenidos Inclusivos y Accesibles	Centro de Servicios de Salud - Regional Antioquia
Daniel Ricardo Mutis Gómez	Evaluador para contenidos inclusivos y accesibles	Centro de Servicios de Salud - Regional Antioquia