Desarrollo de un entorno de computacion distribuido basado en servicios virtuales

Álvaro López García 1

¹Instituto de Física de Cantabria - CSIC-UC Grupo de computación distribuida y Grid

> Facultad de Ciencias Tesis de Máster Máster en Computación



Esquema

- Antecedentes
 - e-Ciencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



Esquema

- Antecedentes
 - e-Ciencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- 3 Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



e-Ciencia y Grid

- «e-Science is about global collaboration in key areas of Science and the next generation of infrastructure that will enable it» – Dr. John Taylor.¹
- Término surgido a raiz de los avances en las comunicaciones entre computadores.
- Nuevos modelos de colaboración científica a nivel mundial.
- La realización de la e-Ciencia más concreta es el Grid.

miembro de la UK Office of Science and Technology

e-Ciencia y Grid

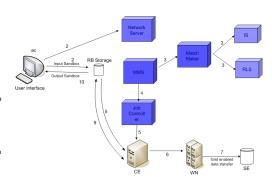
- Nuevo modelo de computación inspirado en el funcionamiento de la red eléctrica.
 - Agregación de un número indeterminado de recursos heterogéneos.
 - Políticas de acceso seguro.
 - Utilización eficiente.
 - Uso de estándares abiertos.
 - Acceso transparente para los usuarios.
- Amplia utilización en entornos científicos.
- Fomenta la colaboración.
- Permite abordar grandes problemas computacionales.
- Se utiliza un software de middleware (Globus, Condor, gLite...)



e-Ciencia y Grid Componentes Grid

Se basa en la agregación de servicios. Dentro de las infraestructuras basadas en gLite se encuentran.

- Ul User Interface. Nodo a través del que un usuario envía un trabajo al Grid.
- RB Resource Broker o WMS (Workload Management System). Encargado de distribuir los trabajos enviados a los diferentes nodos. Formado a su vez por el Network Server, el WMS, Job Controller y Match Maker (pueden estar el el mismo servidor o en diferentes).
- IS Information System. Sistema que agrega la información de los diferentes nodos Grid.
- CE Computing Element. Interfaz de un nodo Grid con el resto de la infraestructura. Distribuye los trabajos que lleguen a un Batch System local.
- WN Worker Node. Los nodos de cálculo del Batch System de un sitio Grid.
- SE Storage Element. Elemento encargado del almacenamiento y réplica de los datos.





Esquema

- 1 Antecedentes
 - e-Ciencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



Disponibilidad de servicios

- Dentro de la computación Grid ciertos componentes son críticos.
- Un fallo de estos puede tener consecuencias para el resto de la infraestructura y sus usuarios.
- Existen diferentes parámetros que afectan a la disponibilidad.

MTBF Tiempo medio entre fallos.

MTTR Tiempo medio de recuperación.

Disponibilidad
$$d = \frac{MTBF}{MTBF + MTTR}$$



Hardware.



Software.



Mortalidad temprana

Inicio del uso de un componente. Se detectan componentes defectuosos y se sustituyen.

Vida útil Período de uso en que el MTBF es aplicable y predecible.

Agotamiento La vida útil ha finalizado y la tasa de fallos aumenta.

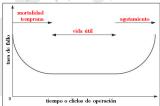
Etapas tempranas de un desarrollo. Se detectar los errores (bugs) y se solucionan.

Vida útil Se continuan elimnando errores, pero su tasa de aparición es menor.

Obsolescencia El software está obsoleto y se deja de utilizar. Los fallos que aparecen disminuyen.



Hardware.



Software.



Mortalidad temprana

Inicio del uso de un componente. Se detectan componentes defectuosos y se sustituyen.

Vida útil Período de uso en que el MTBF e aplicable y predecible.

Agotamiento La vida útil ha finalizado y la tasa de fallos aumenta.

Etapas tempranas de un desarrollo. Se detectar los errores (bugs) y se solucionan.

Vida útil Se continuan elimnando errores, pero su tasa de aparición es menor.

Obsolescencia El software está obsoleto y se deja de utilizar. Los fallos que aparecen disminuyen.



Hardware.



Software.



Mortalidad temprana Inicio del uso de un componente. Se detectan componentes defectuosos y se sustituyen.

Vida útil Período de uso en que el MTBF es aplicable y predecible.

Agotamiento La vida útil ha finalizado y la tasa o

Test e integración Etapas tempranas de un desarrollo. Se detectan

Vida útil Se continuan elimnando errores, pero su tasa de aparición es menor.

Obsolescencia El software está obsoleto y se deja de utilizar. Los fallos que aparecen disminuyen.



Hardware.



Software.



Mortalidad temprana

Inicio del uso de un componente. Se detectan componentes defectuosos y se

sustituyen.

Vida útil Período de uso en que el MTBF es

aplicable y predecible.

La vida útil ha finalizado y la tasa de Agotamiento

fallos aumenta



Hardware.



Software.



Mortalidad temprana Inicio del uso de un componente. Se detectan componentes defectuosos y se

sustituyen.

Vida útil Período de uso en que el MTBF es

aplicable y predecible.

La vida útil ha finalizado y la tasa de Agotamiento

fallos aumenta

Etapas tempranas de un desarrollo. Se detectan Test e integración los errores (bugs) y se solucionan.



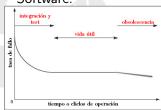
Disponibilidad de servicios

Diagramas de la bañera

Hardware.



Software.



Mortalidad temprana Inicio del uso de un componente. Se detectan componentes defectuosos y se

sustituyen.

Vida útil Período de uso en que el MTBF es

aplicable y predecible.

La vida útil ha finalizado y la tasa de Agotamiento

fallos aumenta

Etapas tempranas de un desarrollo. Se detectan Test e integración

los errores (bugs) y se solucionan.

Vida útil Se continuan elimnando errores, pero su tasa de

aparición es menor.



Hardware.



Software.



Mortalidad temprana Inicio del uso de un componente. Se detectan componentes defectuosos y se sustituyen.

Vida útil Período de uso en que el MTBF es

aplicable y predecible.

Agotamiento La vida útil ha finalizado y la tasa de

fallos aumenta.

Test e integración Etapas tempranas de un desarrollo. Se detectan los errores (bugs) y se solucionan.

to citores (bugs) y se solucionali.

Vida útil Se continuan elimnando errores, pero su tasa de aparición es menor.

Obsolescencia El software está obsoleto y se deja de utilizar. Los

fallos que aparecen disminuyen.



Disponibilidad de servicios

Recuperación

- El tiempo de recuperación se conoce como MTTR.
- Transcurre desde que el fallo se produce (no desde que se detecta) hasta que se soluciona.
- No es lo mismo recuperarse de un fallo hardware que de un fallo software.
- Es el parámetro de la disponibilidad más reducible.
 - Mediante hardware redundante.
 - Mediante la realización de backups, snapshots, etc.



Esquema

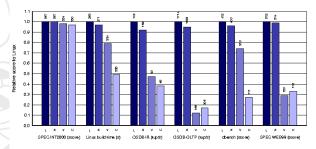
- Antecedentes
 - e-Giencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- 3 Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



- Técnología que permite la creación de diferentes ambientes de ejecución –llamados máquinas virtuales– dentro de una máquina dada –llamada máquina física–, mediante el uso de un software específico –VMM o hypervisor–.
 - Independientes entre sí.
 - Encapsulados y seguros.
- Separación entre máquinas virtuales y físicas.
- Diferentes tecnologías: A nivel de SO, completa, hardware, paravirtualización.
- Según la técnica utilizada habrá una pérdida de rendimiento mayor o menor.



Paravirtualización



- Permite rendimientos cercanos al SO nativo.
- Necesita la modificación del SO a virtualizar.
- Implementada por Xen.



Beneficios

Algunos de los beneficios de la virtualización:

Consolidación de servidores Varias máquinas virtuales por máquina física.

Beneficios

Algunos de los beneficios de la virtualización:

Consolidación de servidores Varias máquinas virtuales por

máquina física.

Separación máquina real/virtual Una máquina virtual se puede ejecutar en cualquier máquina física sin necesidad de modificaciones.

Beneficios

Algunos de los beneficios de la virtualización:

Consolidación de servidores Varias máquinas virtuales por

máquina física.

Separación máquina real/virtual Una máquina virtual se puede

ejecutar en cualquier máquina física sin necesidad de modificaciones.

Migración y balanceo Se pueden migrar las máquinas a servidores menos sobrecargados.

Beneficios

Algunos de los beneficios de la virtualización:

Consolidación de servidores Varias máquinas virtuales por máquina física.

Separación máquina real/virtual Una máquina virtual se puede

ejecutar en cualquier máquina física sin necesidad de modificaciones.

Migración y balanceo Se pueden migrar las máquinas a servidores menos sobrecargados.

Clonación y snapshotting Permite la vuelta hacia atrás en caso de fallo. Útil para hacer frente a fallos de software.

Beneficios

Algunos de los beneficios de la virtualización:

Consolidación de servidores Varias máquinas virtuales por

máquina física.

Separación máquina real/virtual Una máquina virtual se puede

ejecutar en cualquier máquina física sin necesidad de modificaciones.

Migración y balanceo Se pueden migrar las máquinas a servidores menos sobrecargados.

Clonación y snapshotting Permite la vuelta hacia atrás en caso de fallo. Útil para hacer frente a fallos de software.

Rapidez de despliegue La creación de nuevas máquinas virtuales es trivial.

Esquema

- 1 Antecedentes
 - e-Ciencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



Motivación de la companyación de

- El número de máquinas virtualizadas dentro del IFCA es cada vez mayor.
 - El instituto alberga diferentes servicios de proyectos Grid Europeos, algunos de los cuales se han virtualizado.
 - No solo hay que hacer frente a fallos hardware, sino que existen componentes software que es necesario actualizar y no son todo lo estables que deberían.
 - Se pueden aprovechar las ventajas de la virtualización para obtener un sistema más tolerante a fallos.



Motivación

- El número de máquinas virtualizadas dentro del IFCA es cada vez mayor.
- El instituto alberga diferentes servicios de proyectos Grid Europeos, algunos de los cuales se han virtualizado.
- No solo hay que hacer frente a fallos hardware, sino que existen componentes software que es necesario actualizar y no son todo lo estables que deberían.
- Se pueden aprovechar las ventajas de la virtualización para obtener un sistema más tolerante a fallos.



Motivación

- El número de máquinas virtualizadas dentro del IFCA es cada vez mayor.
- El instituto alberga diferentes servicios de proyectos Grid Europeos, algunos de los cuales se han virtualizado.
- No solo hay que hacer frente a fallos hardware, sino que existen componentes software que es necesario actualizar y no son todo lo estables que deberían.
- Se pueden aprovechar las ventajas de la virtualización para obtener un sistema más tolerante a fallos.



Motivación de la companyación de

- El número de máquinas virtualizadas dentro del IFCA es cada vez mayor.
- El instituto alberga diferentes servicios de proyectos Grid Europeos, algunos de los cuales se han virtualizado.
- No solo hay que hacer frente a fallos hardware, sino que existen componentes software que es necesario actualizar y no son todo lo estables que deberían.
- Se pueden aprovechar las ventajas de la virtualización para obtener un sistema más tolerante a fallos.



Esquema

- 1 Antecedentes
 - e-Giencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



- Las máquinas que proveen los servicios Grid serán virtualizadas.
 - nte de virtualización.
- Las imágenes de disco de dichas máquinas virtuales se almacenarán remotamente, utilizando herramientas que permitan la creación de snapshots.
 - Componente de almacenamiento.
- Un software supervisa el buen funcionamiento de la infraestructura.
 - mponente de monitorización.

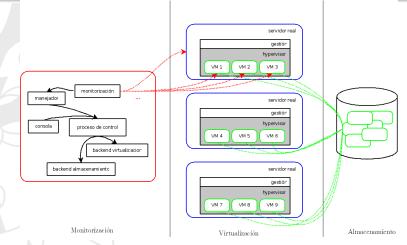
- Las máquinas que proveen los servicios Grid serán virtualizadas.
 - Componente de virtualización.
- Las imágenes de disco de dichas máquinas virtuales se almacenarán remotamente, utilizando herramientas que permitan la creación de snapshots.
 - Componente de almacenamiento.
- Un software supervisa el buen funcionamiento de la infraestructura.
 - mponente de monitorización

- Las máquinas que proveen los servicios Grid serán virtualizadas.
 - Componente de virtualización.
- Las imágenes de disco de dichas máquinas virtuales se almacenarán remotamente, utilizando herramientas que permitan la creación de snapshots.
 - Componente de almacenamiento.
- Un software supervisa el buen funcionamiento de la infraestructura.
 - mponente de monitorización

- Las máquinas que proveen los servicios Grid serán virtualizadas.
 - Componente de virtualización.
- Las imágenes de disco de dichas máquinas virtuales se almacenarán remotamente, utilizando herramientas que permitan la creación de snapshots.
 - Componente de almacenamiento.
- Un software supervisa el buen funcionamiento de la infraestructura.
 - Componente de monitorización.

Solución

Diseño





Esquema

- 1 Antecedentes
 - e-Giencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



Virtualización

- Las máquinas proveedoras de servicios se ejecutarán virtualizadas.
- Dos subcomponentes: hypervisor y capa de gestión.

hypervisor

- Es el software de virtualizacion utilizado.
- Las pruebas y el diseño se han realizado con Xen.

Gestión

- Para no depender de un hypervisor específico.
- Traduce llamadas genéricas a llamadas específicas de un hypervisor.
- libvirt



Almacenamiento

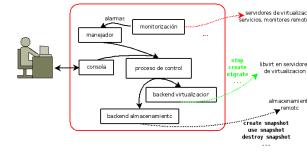
- Componente donde se almacenan las imágenes de disco de las máquinas virtuales.
- Accesible por todos los servidores de virtualización a través de una red de comunicaciones.
- Capacidad para realizar snapshots y/o clonaciones.
- Redundante y fiable.
- Dos subcomponentes:
 - frontend Hace accesibles las imágenes a los servidores (iSCSI, AoE, FC, etc.).
 - backend Proporciona la redundancia y el resto de funcionalidades (LVM, ZFS, GPFS, etc.).



Esquema

Partes

- Monitorización.
- Manejador de alertas.
- Software de control.
- Gestor
 - virtualización.
 - almacenamiento.
- Consola.





Monitorización y alertas

Monitorización

- Cualquier software de monitorización que permita gestionar alertas.
- El SW utilizado ha sido GNU/MON.
 - Plattform Agnostic.
 - Diseño modular, permite definir monitores de forma sencilla.
 - Alertas asociadas a monitores.

Alertas

- Manejador que recoge las alertas producidas por la monitorización y las pasa al SW de control.
- Independiente del SW utilizado.



Software de control

- Gestiona las alertas que recibe del manejador o de la consola.
- Guarda el estado actual y pasado del sistema.
- Se encarga de la toma de decisiones ante un fallo de una máquina fisica o virtual, utilizando los gestores de virtualización y almacenamiento.

Gestores

Virtualización Almacenamiento

- libvirt.
- Gestor que maneja de forma remota el almacenamiento.
- Realiza los snapshots, clonaciones, borrados, etc.
- Dependiente tanto del frontend como del backend.



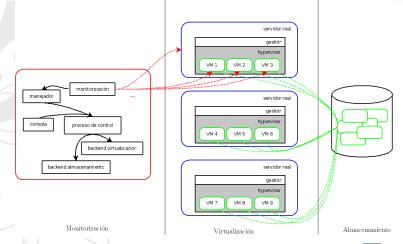
Esquema

- 1 Antecedentes
 - e-Ciencia y Grid
 - Disponibilidad de servicios
 - Virtualización
- 2 Motivación
- Solución
 - Concepto
 - Componentes
 - Virtualización
 - Almacenamiento
 - Monitorización y Gestión
 - Funcionamiento



Solución

Diseño





- MON detecta el fallo a través de un monitor y lanza una alerta.
 - El manejador pasa la alerta al programa de control.
 - Este comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - Ordena al almacenamiento realizar un snapshot de las mismas.
 - Busca los servidores de virtualización disponibles
 - Lanza las máquinas en ellos



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
 - Este-comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - Ordena al almacenamiento realizar un snapshot de las mismas
 - Busca los servidores de virtualización disponibles
 - 4 Lanza las máquinas en ellos.



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - Ordena al almacenamiento realizar un snapshot de las mismas
 - 3 Busca los servidores de virtualización disponibles
 - Lanza las máquinas en ellos



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - Ordena al almacenamiento realizar un snapshot de las mismas.
 - Busca los servidores de virtualización disponibles
 - Lanza las máguinas en ellos.



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Éste comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - 2 Ordena al almacenamiento realizar un snapshot de las mismas.
 - 3 Busca los servidores de virtualización disponibles
 - Lanza las máguinas en ellos.



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Éste comprueba que la alerta no esté duplicada y:
 - Busca que máquinas virtuales se estaban ejecutando.
 - 2 Ordena al almacenamiento realizar un snapshot de las mismas.
 - 3 Busca los servidores de virtualización disponibles.
 - Lanza las máguinas en ellos



- MON detecta el fallo a través de un monitor y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Éste comprueba que la alerta no esté duplicada y:
 - 1 Busca que máquinas virtuales se estaban ejecutando.
 - 2 Ordena al almacenamiento realizar un snapshot de las mismas.
 - 3 Busca los servidores de virtualización disponibles.
 - 4 Lanza las máquinas en ellos.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
 - El manejador pasa la alerta al programa de control.
 - Este comprueba que la alerta no esté duplicada y
 - 🕠 Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
 - Éste comprueba que la alerta no esté duplicada y:
 - 🐧 Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen
 anterior.
 - 2 Trata de recuperar con ella



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Séste comprueba que la alerta no esté duplicada y:
 - 1 Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - 1 Comprueba su máquina física. En caso de que falle lo gestiona.
 - Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla
 - 1 Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- 3 Éste comprueba que la alerta no esté duplicada y:
 - 1 Comprueba su máquina física. En caso de que falle lo gestiona.
 - 2 Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen
 anterior.
 - Trata de recuperar con ella.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Séste comprueba que la alerta no esté duplicada y:
 - 1 Comprueba su máquina física. En caso de que falle lo gestiona.
 - 2 Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - 2 Trata de recuperar con ella.



- MON detecta el fallo de una máquina virtual y lanza una alerta.
- 2 El manejador pasa la alerta al programa de control.
- Séste comprueba que la alerta no esté duplicada y:
 - 1 Comprueba su máquina física. En caso de que falle lo gestiona.
 - 2 Reinicia la máquina virtual.
 - 3 Si la máquina se recupera: Avisa a los administradores.
 - Si la máquina falla:
 - Interroga al almacenamiento para ver si hay una imagen anterior.
 - Trata de recuperar con ella.



Resumen

- Se ha implementado un sistema que es capaz de hacer frente a fallos hardware de forma autónoma.
- Permite recuperarse ante actualizaciones fallidas.
- Utilizando tecnología ya existente en el instituto.
- Mecanismos para facilitar la recuperación e investigación de los fallos.

Final

