

Resumo do Traballo de Fin de Máster

Contexto

La copia de seguridad es un componente indispensable en cualquier sistema de información. Este elemento permite recuperar un sistema en caso de desastres o errores tanto software como humanos. Los sistemas de copias de seguridad son, de por sí, complejos. Esta complejidad se acentúa todavía más en el caso de que se deban implementar estrategias de copias de seguridad para grandes volúmenes de datos.

CERNBox es un sistema utilizado por los usuarios del CERN (físicos, administración y asociados) para almacenar, sincronizar y compartir ficheros. Dada su creciente importancia y criticidad en el día a día de la organización, CERNBox necesita un sistema de copias de seguridad moderno, fiable escalable y fácil de administrar. CERNBox almacena actualmente 7 petabytes de información distribuidos en 17,000 cuentas de los colectivos mencionados actualmente, cuya protección se debe implementar.

Objetivo

El objetivo principal del trabajo es la implementación e implantación de un mecanismo de copia de seguridad para CERNBox. Actualmente, existe un proyecto (liderado por mi) en cuyo marco se ha desarrollado una versión inicial (prototipo) del sistema. Esta solución debe ser revisada, replanteada y completada con funcionalidades adicionales. Entre las funcionalidades adicionales se destacan las siguientes:

- Verificar idoneidad de la arquitectura actual con fines de mejorar el despliegue, re-usabilidad, seguridad y mantenimiento del mismo.
- Verificar idoneidad del sistema de base de datos, explorando otras soluciones que puedan ser optimas para la solución actual.
- Incluir nuevas funcionalidades y componentes del sistema como la verificación automática de *backups* y la posibilidad de almacenar las copias de seguridad en diferentes destinos con una granularidad a nivel de cada trabajo.

Metodología/Equipamiento

El proyecto se desarrollará siguiendo una metodología ágil (tipo SCRUM) y el equipamiento lo conformaran tanto medios personales (laptop, equipo de sobremesa), como los recursos proporcionados por el CERN. Particularmente se contempla hacer uso de la siguiente infraestructura:

- Hasta 20 maquinas virtuales con configuración 4 cores y 8G de RAM, 40Gb de almacenamiento local, configurables utilizando Puppet.
- Una base de datos MySQL para el prototipo actual.
- Hasta 300 terabytes de espacio en S3 para la realización de pruebas con el prototipo. Para la implantación del sistema se necesitaría una capacidad de almacenamiento de 7 petabytes.

Aspectos técnicos

Para la implementación de este trabajo será necesario combinar y usar adecuadamente las siguientes tecnologías:

- Un lenguaje de programación de tipo *script* (como Python) capaz de procesar los argumentos indicados en la interfaz de línea de comandos y que permita hacer una implementación basada en objetos. Estas características permitirán mejorar la re-usabilidad del software, así como el empaquetado y el despliegue.
- Conexión y acceso concurrente a base de datos.
- Operaciones sobre sistemas de ficheros locales y distribuidos.
- Configuración de servicios de sistema.
- Sistemas de monitoreo y gestión de registros.

Ourense, 18 de Setembro de 2019

O/A alumno/a

Vº e Pr. O/A titor/a do TFM

Asdo: Roberto Valverde Cameselle

Asdo: José Ramón Méndez Reboredo
