UNIVERSIDAD DE GRANADA

HARDWARE Y SOFTWARE DE GOOGLE, TWITTER, FACEBOOK (Y OTROS GRANDES SISTEMAS)

Servidores Web de Altas Prestaciones
Horas de trabajo: 10 horas

Grupo 9
Juan Pablo García Sánchez
Elena Ortiz Moreno

Contents

1 Introducción

En la asignatura de **Servidores Web de Altas Prestaciones** hemos estudiado como crear servidores de alta calidad, qué es imprescindible que tengan, cómo estructurar una granja de servidores y configurarla para que actúe como NFS o compartan base de datos entre otros. Pero los recursos a los que podemos acceder los estudiantes no son ni de lejos comparables a los que tienen acceso empresas como **Facebook** y **Google**.

En este trabajo nos hemos propuesto estudiar el hardware y software que usan estas compañías para ver que herramientas utilizan para mantener la seguridad y fiabilidad de sus servidores y buscar similitudes con lo que nosotros hemos usado en nuestras prácticas.

2 Preliminares

Las empresas sobre las que vamos a hablar en este trabajo son sobre Google y Facebook.

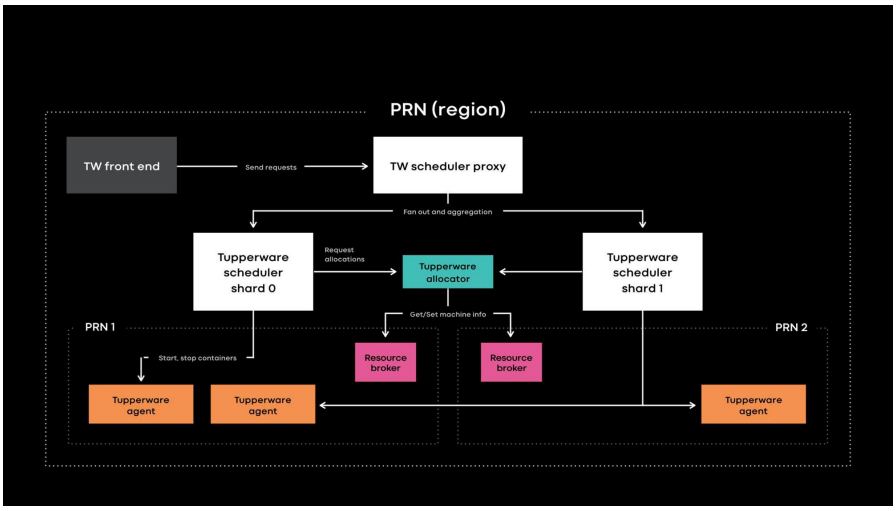
Facebook es una red social fundada por Mark Zuckerberg y se considera una de las empresas más importantes de tecnología junto a Google, Amazon, Apple y Microsoft. En enero contaba con 2.249 millones de usuarios [1] y también es dueña de Whatsapp, Instagram, Oculus VR, Giphy y Mapillary. [2] Es por ello que consideramos que debe tener unos servidores considerablemente grandes para mantener todos los datos que pueden existir en unas redes sociales como Facebook, Whatsapp e Instagram, y por eso hemos decidido añadir Facebook a este trabajo.

Google es una empresa dedicada a sacar productos y servicios relacionados con Internet y software. Su producto más conocido es el motor de búsqueda del mismo nombre, aunque cuenta con muchos más servicios como Google Drive, Gmail, Google Maps, Google Earth, Google Street View, Hangouts, Meets, Blogger, Youtube, Stadia, Google Play, Google Chrome, Google Home, entre otros. [3] Al contar con tantísimos servicios basados en Internet, lo hemos añadido también al trabajo para ver como consiguen tener todos esos servicios funcionando para millones de usuarios y como manejan los fallos y caídas para que no afecte la experiencia de los usuarios.

3 Análisis de Facebook

Los servidores de Facebook [4] se componen de varios centros de datos ubicados alrededor del mundo funcionando continuamente. Los componentes físicos pueden fallar debido a varias razones como el desgaste, el uso por encima de sus capacidades o el entorno.

Para predecir estos fallos y minimizar las interrupciones, Facebook usa **Twine** [5]: un sistema para la administración de los clústers que les permitió pasar de usar máquinas personalizadas para distintas cargas de trabajo a usar una infraestructura ubicua compartida donde cualquier máquina puede ejecutar cualquier carga de trabajo. Los desarrolladores de aplicaciones despliegan sus aplicaciones que consisten en varios contenedores normalmente ejecutando el mismo código y Twine ubica estos contenedores a lo largo de la red.

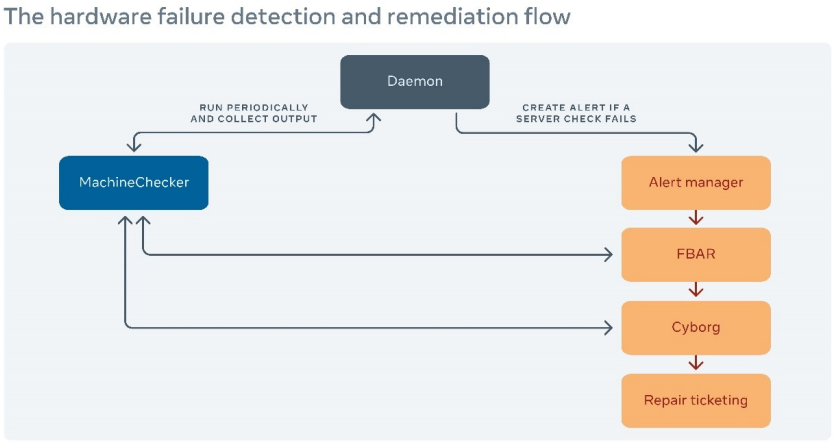


En la imagen se ve la arquitectura de Twine. En el ejemplo se utiliza el centro de datos PRN, que se subdivide en 2 edificios que son PRN1 y PRN2. El *front end* proporciona APIs para la UI, CLI y otras herramientas para interactuar con Twine sin conocer los detalles internos. El Twine Scheduler es el responsable de administrar los trabajos y el *lifespan* de cada contenedor. Este planificador puede ser tanto global como regional y el proxy sirve para ocultar los detalles internos. El allocator asigna los contenedores a cada servidor. El agente es un Daemon que funciona en cada servidor para preparar y demoler los contenedores.

Facebook también ha construido sistemas que les permiten:

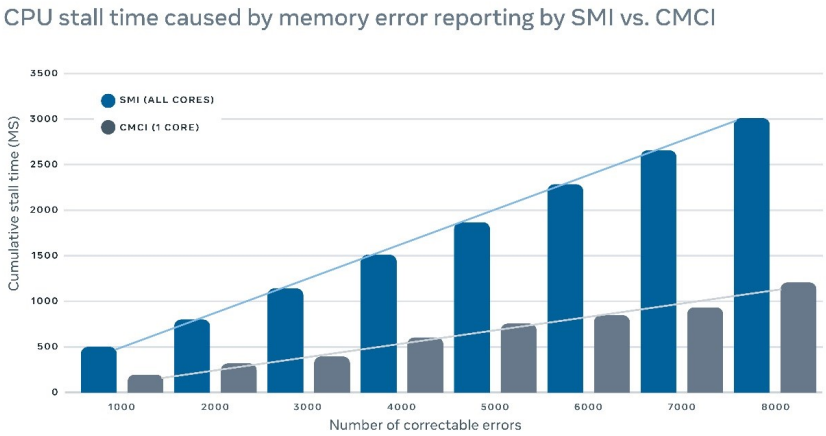
- Detectar y solucionar problemas.
- Monitorizar y remediar eventos del hardware sin perjudicar gravemente el rendimiento de las aplicaciones.
- Usar metodología predictiva para reparaciones.
- Automatizar análisis de la raíz de los problemas para encontrar fallos de hardware/sistema y resolver los problemas rápidamente.

3.1 Detectar y solucionar problemas



Periódicamente se ejecuta una herramienta llamada **MachineChecker** en cada servidor para encontrar fallos en el hardware o la conectividad. Entonces MachineChecker manda un aviso a **FBAR** (Facebook Auto-Remediation), un sistema dedicado a manejar estas alertas y arreglar los errores de los que es avisado. Si FBAR no puede solucionar el problema, entonces **Cyborg** intentará arreglar los fallos en su lugar. Cyborg puede ejecutar código de bajo nivel como actualizaciones de firmware o kernel. Si no puede solucionarlo, avisa a un técnico a través de un ticket que manda al sistema de reparaciones.

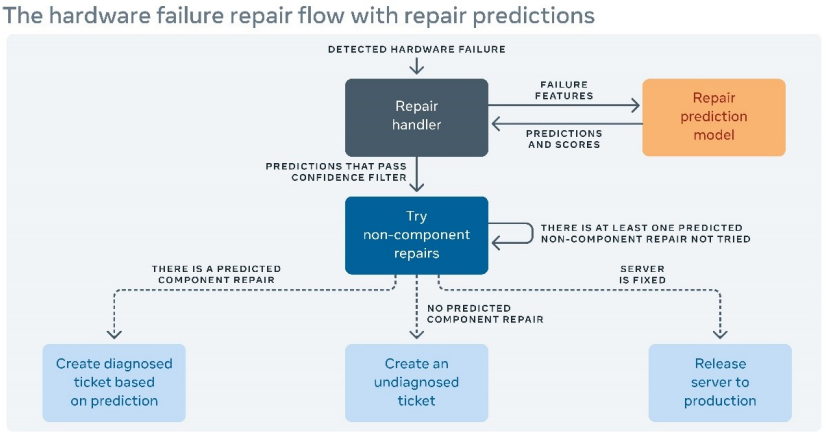
3.2 Monitorizar y remediar eventos del hardware



MachineChecker detecta los errores del hardware leyendo los logs de los servidores. Normalmente, cuando un error de hardware ocurre, se manda una señal de interrupción a la CPU para poder manejar y logear el error. Esto provoca un impacto negativo en el servidor, ya que una pausa de milisegundos puede ser devastador para servicios que son sensibles a la latencia. Varias interrupciones en distintos ordenadores podrían producir un efecto en cascada en el rendimiento.

Por eso Facebook usa un mecanismo híbrido de **SMI** y **CMCI**. Mientras que SMI para el CPU por completo para arreglar errores, CMCI solo para un core y permite que el resto siga funcionando normalmente mientras ese core intenta solucionar el error.

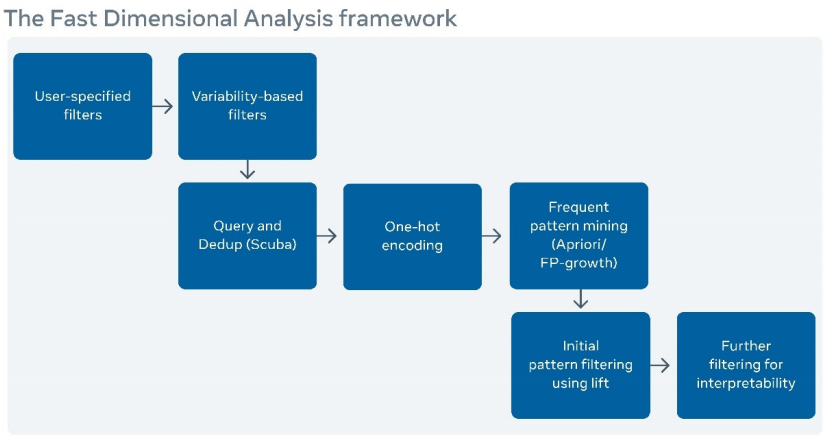
3.3 Metodología predictiva para reparaciones



Al añadir nuevo hardware o cambiar la configuración del software, puede que aparezcan nuevos errores que no están abordados por el sistema automatizado de reparación. En ese lapso de tiempo entre que se implementan los nuevos cambios y se incorporan las nuevas reglas de remediación, es posible que se manden tickets de reparación clasificados como **"no diagnosticado"** si el sistema no sugiere como reparar el error (porque no sabe cómo hacerlo) o **"mal diagnosticado"** si la sugerencia para reparar no es efectiva. Esto provoca más tiempo con el sistema caído mientras los técnicos diagnostican el problema ellos mismos.

Para evitar esto, Facebook creó un framework de *machine learning* que aprende sobre como ha solucionado los errores en el pasado e intenta predecir que reparaciones serán necesarias para los tickets "no diagnosticados" y "mal diagnosticados". Con el cálculo del coste y beneficio de las predicciones correctas e incorrectas, pueden implementar un *threshold* de confianza y priorizar el orden de las reparaciones para optimizar el sistema.

3.4 Automatizar análisis de la raíz de los problemas



A parte de los logs del servidor que guardan los reinicios, los errores out-of-memory, etc., también hay herramientas de software y logs en el sistema. Pero debido a la escala y la complejidad hacen difícil examinar todos los logs en busca de correlaciones entre ellos. Es por ello que Facebook implementó el sistema **RCA** (Root-cause-analysis) que organiza los millones de logs para encontrar correlaciones entre todas sus columnas.

Con el uso de **Scuba** para la pre-agregación de los datos han conseguido mejorar la escabilidad del algoritmo de minería de datos **FP-Growth**.

4 Análisis de Google

La infraestructura de Google proporciona implementación segura de servicios, almacenamiento seguro de datos, comunicaciones seguras entre servicios, comunicación segura y privada con los clientes a través de Internet y operaciones seguras por parte de los administradores. Google utiliza esta infraestructura para construir sus servicios de Internet, como Gmail, G Suite, etc. [6] [7]

4.1 Capas de seguridad de la infraestructura de Google

A continuación, se representarán en una tabla los distintos niveles de la infraestructura y los métodos que usa Google para garantizar la seguridad en cada uno de ellos.

<table>
<tr><th colspan="4">SEGURIDAD OPERACIONAL</th></tr>
<tr><td>Detección de intrusión</td><td>Reducir riesgo interno</td><td>Dispositivos y credenciales de empleados seguros</td><td>Desarrollo de software seguro</td></tr>
<tr><th colspan="4">COMUNICACIÓN DE INTERNET</th></tr>
<tr><td colspan="2">Front-end de Google</td><td colspan="2">Protección DoS</td></tr>
<tr><th colspan="4">SERVICIOS DE ALMACENAMIENTO</th></tr>
<tr><td colspan="2">Cifrado en reposo</td><td colspan="2">Eliminación de datos</td></tr>
<tr><th colspan="4">IDENTIDAD DE USUARIO</th></tr>
<tr><td colspan="2">Autenticación</td><td colspan="2">Protección de abuso de login</td></tr>
<tr><th colspan="4">DESPLIEGUE DE SERVICIO</th></tr>
<tr><td>Gestión de acceso de datos de usuario final</td><td>Cifrado de la comunicación entre servicios</td><td>Gestión de acceso entre servicios</td><td>Identidad de servicio, integridad y aislamiento</td></tr>
<tr><th colspan="4">INFRAESTRUCTURA HARDWARE</th></tr>
<tr><td>Pila de arranque e identidad de las máquinas seguras</td><td colspan="2">Diseño y procedencia del hardware</td><td>Seguridad de las instalaciones físicas</td></tr>
</table>

En los siguientes apartados se va a describir cada capa, desde la capa de menor nivel hasta la de nivel superior.

4.1.1 Seguridad en infraestructura hardware

En este apartado se describe cómo protege Google las capas inferiores de su infraestructura.

Seguridad de las instalaciones físicas

Google dispone de sus propios centros de datos [8] con múltiples capas de seguridad física. El acceso a estos datos se limita a una pequeña parte del personal de la empresa.

Para proporcionar esta seguridad a su infraestructura utilizan tecnologías como identificación biométrica, detección de metales, cámaras, barreras para vehículos y sistemas de detección de intrusos basados en láser.

Además, Google aloja parte de sus servidores en centros de terceros, sumando así tanto la protección que proporcionan estos como su propia protección.

Diseño y procedencia del hardware

Los centros de datos mencionados anteriormente están formados por miles de máquinas conectadas a una red local, diseñadas ambas por Google. También diseñan chips personalizados, que ayudan a identificar y autentificar de forma segura los elementos hardware de Google.

Asegurar la pila de arranque y la identidad de la máquina

Para un arranque correcto, Google utiliza firmas criptográficas en componentes de bajo nivel como la BIOS, el cargador de arranque, el kernel y la imagen del sistema operativo base. Estas firmas se pueden validar en cada arranque o actualización.

Cada máquina del centro de datos tiene su propia identificación, utilizada para identificar llamadas de API desde y hacia servicios de administración de bajo nivel.

Google ha automatizado las actualizaciones de sus pilas de software para detectar y diagnosticar problemas de hardware y software.

4.1.2 Seguridad en el despliegue del servicio

La infraestructura está diseñada fundamentalmente para ser multiarrendataria.

Identidad, integridad y aislamiento del servicio

Google garantiza un buen control de acceso en la capa de aplicación para la comunicación entre servicios.

Cada servicio que se ejecuta en la infraestructura tiene una identidad de cuenta de servicio vinculada. A un servicio se le dan credenciales criptográficas que puede utilizar para probar su identidad al hacer o recibir llamadas a métodos remotos (RPC) a otros servicios.

El código fuente de Google se almacena en un repositorio central, debiendo ser revisado, registrado y probado. Dichas revisiones de código necesitan la inspección y aceptación de por lo menos un ingeniero que no sea el creador, y el sistema hace cumplir que las modificaciones del código a cualquier sistema tienen que ser aprobadas por los propietarios de ese sistema. Dichos requisitos limitan la capacidad de un interno o contrincante para

hacer modificaciones maliciosas en el código fuente y además dar un rastro forense a partir de un servicio hasta su fuente.

Google cuenta con una variedad de técnicas de aislamiento y sandbox para proteger unos servicios de otros. Estas técnicas integran la división usual de usuarios de Linux, ámbitos limitados basados en el lenguaje y el kernel y la virtualización de hardware.

Gestión de acceso entre servidores

El dueño del servicio puede controlar el acceso especificando qué servicios se pueden comunicar con él.

Ese servicio se puede configurar con la lista blanca de las identidades de cuenta de servicio permitida y esta restricción de acceso se aplica automáticamente por la infraestructura.

Los ingenieros de Google que entran a los servicios además reciben identidades personales, por lo cual los servicios tienen la posibilidad de configurarse de igual manera para permitir o denegar sus accesos.

La infraestructura da un sistema de flujo de trabajo de administración de identidades para estas identidades internas, incluidas las cadenas de aprobación, el registro y la notificación. Por ejemplo, estas identidades se pueden asignar a equipos de control de ingreso por medio de un sistema que posibilita el control de dos partes, donde un ingeniero puede plantear un cambio a un grupo que otro ingeniero (que además es administrador del grupo) debería aprobar.

Cifrado de la comunicación entre servicios

Para proveer dichos beneficios de seguridad a otros protocolos de la capa de aplicación como HTTP, se encapsulan en los mecanismos de RPC de la infraestructura. Esencialmente, esto otorga un aislamiento de la capa de aplicación y quita cualquier dependencia de seguridad de la ruta de red.

Para protegerse contra los servicios que pueden estar tratando de acceder a los enlaces WAN privados, la infraestructura encripta automáticamente todo el tráfico RPC de infraestructura que pasa por la WAN entre centros de datos, sin requerir ninguna configuración explícita del servicio.

Gestión de acceso de datos de usuario final

El Servicio de contactos se puede configurar para que solo se permitan solicitudes de RPC del Servicio de Gmail.

Dentro del alcance de este permiso, el servicio de Gmail podría pedir los contactos de cualquier usuario en cualquier momento.

Ya que el servicio de Gmail ejecuta una solicitud de RPC al servicio de Contactos en nombre de un usuario final en especial, la infraestructura otorga una capacidad para que el servicio de Gmail presente un “ticket de permiso de usuario final” como parte del

RPC. Este ticket prueba que el servicio de Gmail está atendiendo en la actualidad una solicitud en nombre de aquel usuario final en especial.

La infraestructura da un servicio de identidad de usuario central que emite dichos "tickets de permiso de usuario final". Un inicio de sesión de usuario final es verificado por el servicio central de identidad que después emite una credencial de usuario.

Si la credencial del usuario final se verifica de manera correcta, el servicio de identidad central devuelve un "ticket de permiso de usuario final" de corta duración que se puede utilizar para los RPC involucrados con la solicitud.

4.1.3 Seguridad en almacenamiento de datos

En este punto se explicará cómo implementa el almacenamiento seguro de datos en la infraestructura.

Cifrado en reposo

La infraestructura de Google aporta pluralidad de servicios de almacenamiento, tales como BigTable y Spanner, y un servicio central de gestión de claves. La mayor parte de las aplicaciones de Google acceden indirectamente por medio de dichos servicios de almacenamiento al almacenamiento físico.

En la capa de aplicación la ejecución del cifrado posibilita a la infraestructura aislarse de posibles amenazas en los niveles más bajos de almacenamiento, como el firmware de disco malicioso. Antes de que un dispositivo de almacenamiento cifrado retirado del servicio logre abandonar físicamente las instalaciones de Google, se limpia por medio de un proceso de diversos pasos que incluye dos verificaciones independientes.

Eliminación de datos

La eliminación de datos en Google constantemente empieza con la marcación de datos específicos como "programados para eliminación" en vez de borrar los datos por completo. Luego de haber sido marcado, los datos se eliminan según las políticas concretas del servicio.

4.1.4 Seguridad en la comunicación de Internet

Como se dijo previamente, la infraestructura se basa en un gran grupo de máquinas físicas que permanecen interconectadas por medio de LAN y WAN, y la seguridad de la comunicación entre servicios no es dependiente de la seguridad de la red.

Servicio de front-end de Google

Cuando se quiere que un servicio esté disponible en Internet, puede registrarse con un servicio denominado Google Front End (GFE), que garantiza que las conexiones TLS utilicen certificados correctos y aplica protección contra ataques de Denegación

de Servicio. El GFE después reenvía las demandas del servicio usando el protocolo de seguridad RPC.

Protección de denegación de servicio (DoS)

La infraestructura de Google tiene protecciones contra DoS y capas que reducen aún más el riesgo.

Después de que su red troncal proporcione una conexión externa a uno de sus centros de datos, pasa a través de varias capas de equilibrio de carga de hardware y software.

Dichos equilibradores de carga proporcionan datos sobre el tráfico entrante a un servicio DoS central que se ejecuta en la infraestructura. Una vez que el servicio DoS central detecta que se está produciendo un ataque DoS, puede configurar los balanceadores de carga para que eliminen o aceleren el tráfico asociado con el ataque. En la siguiente capa, las instancias de GFE además reportan datos acerca de las solicitudes que permanecen recibiendo al servicio DoS central, incluida la información de la capa de aplicación que los balanceadores de carga no poseen.

Autenticación de usuario

Después de la protección DoS, la siguiente capa de protección viene de su servicio de identidad central.

Este servicio principalmente se presenta a los usuarios finales como la página de inicio de sesión de Google. Además de pedir un nombre de usuario y una contraseña básicas, el servicio además reta inteligentemente a los usuarios para obtener información adicional basada en factores de peligro, como si han iniciado sesión desde el mismo dispositivo o una localización parecida en el pasado. Los usuarios además tienen la opción de ocupar segundos factores como OTP o claves de seguridad resistentes al phishing al iniciar sesión.

4.1.5 Seguridad operacional

Para operar en la infraestructura de forma segura se creará un programa de infraestructura seguro, se protegerán las máquinas y credenciales de los empleados y se defenderán las amenazas a la infraestructura tanto del personal interno como de factores externos.

Desarrollo de software seguro

Se ofrecen bibliotecas para evitar ciertos errores de seguridad.

Además, Google cuenta con herramientas automatizadas para identificar automáticamente errores de seguridad integrados en fuzzers, herramientas de estudio estático y escáneres de seguridad web. Aparte de lo explicado, se hacen verificaciones finales de seguridad manual, como revisiones de implementación y diseño en profundidad para las funcionalidades de mayor riesgo. Estas revisiones son llevadas a cabo por un equipo que incluye profesionales en seguridad web, criptografía y seguridad del sistema operativo. Las revisiones además tienen la posibilidad de ofrecer como consecuencia novedosas propiedades de la biblioteca de seguridad y nuevos fuzzers que después se pueden utilizar en otros productos futuros.

Mantener seguros los dispositivos y credenciales de los empleados

Para conservar seguros los dispositivos y credenciales de los empleados se hace una gran inversión. También se monitorean las actividades internas en busca de acciones ilícitas. Se hace una gran inversión en la supervisión de los dispositivos que los empleados usan para operar en la infraestructura de la compañía.

5 Conclusiones

Los servidores de Facebook y Google son muchísimo más grandes y cubren muchas más peticiones que los que nosotros hicimos en las prácticas de la asignatura, por lo que no es ninguna sorpresa que la infraestructura que estas empresas utilizan sea mucho más compleja y costosa.

En **Facebook** vemos que se centran sobretodo en la detección prematura de errores, su rápida solución y en evitar tener el servidor caído el menor tiempo posible. Cuentan con un sistema Twine para predecir fallos, tres sistemas para detectarlos y solventarlos (MachineChecker, FBAR y Cyborg) y un framework de machine learning que aprende como ha solucionado los problemas en ocasiones pasadas para saber como volver a resolverlas en casos futuros. Además, utilizan CMCI (parar un solo core de las CPUs) para minimizar el tiempo que está el servidor caído. Como trabajan con datos y logs enormes, utilizan un sistema RCA y metodos de minería de datos para agrupar y encontrar rápidamente los datos que buscan.

En **Google** se centran en la seguridad de sus servidores. Lo separan en seguridad de hardware, del código fuente, de la base de datos, de las conexiones y de las operaciones. Aumentan la seguridad a base del uso de identificación, encriptación de datos y comunicación, DoS, protocolo de eliminación de datos y monitorización.

La mayoría de técnicas usadas por Google son factibles para un estudiante que quiera usarlas (monitorización, encriptación, identificación), pero son mucho más costosas en servidores de tal calibre. Por otro lado, Facebook ha personalizado casi todos las herramientas que utiliza y los ha creado a medida para sus servidores. Aún así, hemos visto que ambas empresas crean varias capas para aumentar mucho la seguridad, fiabilidad y eficiencia de sus servidores.

Bibliography

[1] Marketing 4 e-commerce. https://marketing4ecommerce.net/cuales-redes-sociales-con-mas-usuarios-mundo-ranking/. Accessed: 14-06-2021.

[2] Wikipedia. https://es.wikipedia.org/wiki/Facebook_(empresa). Accessed: 14-06-2021.

[3] Wikipedia. https://es.wikipedia.org/wiki/Google. Accessed: 14-06-2021.

[4] Engineering FB. https://engineering.fb.com/2020/12/09/data-center-engineering/how-facebook-keeps-its-large-scale-infrastructure-hardware-up-and-running/. Accessed: 14-06-2021.

[5] Engineering FB. https://engineering.fb.com/2020/11/11/data-center-engineering/twine-2/. Accessed: 14-06-2021.

[6] Google Cloud. https://cloud.google.com/compute/docs/cpu-platforms?hl=es-419. Accessed: 14-06-2021.

[7] Plataforma de google. https://es.wikipedia.org/wiki/Plataforma_de_Google. Accessed: 14-06-2021.

[8] Google Centros de datos. https://www.google.com/intl/es-419/about/datacenters/. Accessed: 14-06-2021.