

Cloud storage (agosto 2020)

Michael Sebastian Preciado Garzon, Estudiante de la Escuela Colombiana de Ingeniería Julio Garavito

Abstract—En la medida que la información aumenta antigua y actualmente era necesario un cambio en la forma en que abordábamos el tema de almacenamiento de datos, pasamos de almacenarla in situ a agruparla dentro de una red, para ahora tener servicios de almacenamiento a demanda, es por ello que nacieron arquitecturas para solventar esta necesidad y se crearon tecnologías como Network Attached Storage cuyo principal aporte fue separar los datos de los servidores, haciendo los datos accesibles por varios servidores a la vez, su desventaja principal fue que no permitía escalabilidad, también surgió Storage Area Network, al definir una red explícitamente para almacenamiento permitió una escalabilidad asombrosa, sin embargo debían suplirse los requerimientos de confiabilidad y eficiencia por lo que implementar una red de este tipo es costosa pues cada uno de los dispositivos de almacenamiento debe estar interconectado con fibra óptica, sobre estas arquitecturas se plantearon varios modelos de almacenamiento para suplir los servicios de almacenamiento en la nube, block storage, file storage y object storage, cada uno con sus limitantes y ventajas lo cual da variedad al catálogo de soluciones que pueden ser usadas dentro de la nube

Index Terms—Almacenamiento, File Storage, Block Storage, Object Storage, NAS, SAN

I. ANTECEDENTES

Para entender la importancia del almacenamiento en la nube y su arquitectura, debemos entender cuáles fueron los antecedentes y las necesidades que surgieron para llegar a los modelos que ahora dominan la nube, almacenamiento en bloques, almacenamiento en archivos y almacenamiento de objetos.

Uno de los factores que aunque eran baratos, a la hora de escalar y mantener se hacía costoso era el uso de almacenamiento de conexión directa(DAS), era el método tradicional de conectar al servidor físico, el dispositivo de almacenamiento para que de esa manera se aumentase su capacidad, sin embargo para un proveedor cloud no puede depender directamente de este modelo pues no puede estar disponible a demanda ni escalar tan fácilmente y el costo de hacerlo de esta manera es mayor que el de las ganancias, una alternativa es virtualizar los servidores de esta forma puede configurarse para que sea disponible a demanda sin embargo, seguimos con la desventaja de poder escalar en caso de que la solución crezca rápidamente

II. ALMACENAMIENTO DE ACCESO REMOTO

Ante las problemáticas que tenía almacenar localmente usando DAS, surgieron dos arquitecturas ambas con el objetivo de eliminar el acoplamiento que surgía al depender de los dispositivos de almacenamiento, son NAS y SAN, que son los acrónimos en inglés para Network Attached Storage y Storage Area Network respectivamente, que se basan en una solución de almacenamiento en red

A. Network Attached Storage

Es un dispositivo que administra varios dispositivos de almacenamiento y está conectado a una red, en la cual los diferentes equipos pueden acceder a este por medio de la ip asignada por el router, este dispositivo tiene varios compartimientos en los que se pueden conectar y extraer los discos duros y demás dispositivos de almacenamiento, en el mercado un NAS de 4 compartimento puede estar costando 1000 dólares estadounidenses, para acceder a este almacenamiento ya sea para escribir o leer archivos es usual el uso de protocolos Samba, FTP o NFS

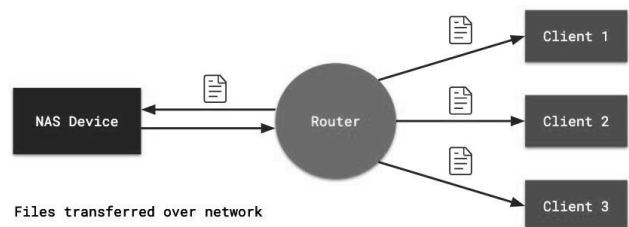


Fig. 1. Rol de un dispositivo NAS dentro de una red

B. Storage Area Network

La idea de implementar una SAN viene dada cuando, por ejemplo tecnologías NAS no son suficientes para el mantenimiento de la empresa, o cuando el crecimiento del negocio es exponencial como el caso de los unicornios o startups, es por ello que, la idea de implementar una red de dispositivos de almacenamiento, que operen de como si de un NAS se tratase, es atractiva, incluso, si se implementaran varias redes de este tipo para los diferentes casos de negocio, sería aún más ideal, es de ese estilo de implementaciones que definen lo que es una SAN, pero para lograr que un montón de dispositivos de almacenamiento interconectados entre si operen como si de

uno se tratase, es necesario que esa conexión sea rápida, es por ello que se usa la fibra óptica para lograr ese objetivo, a nivel de capa de transporte uno de los protocolos más usados es iSCSI pues permite la implementación de SCSI sobre redes TCP/IP

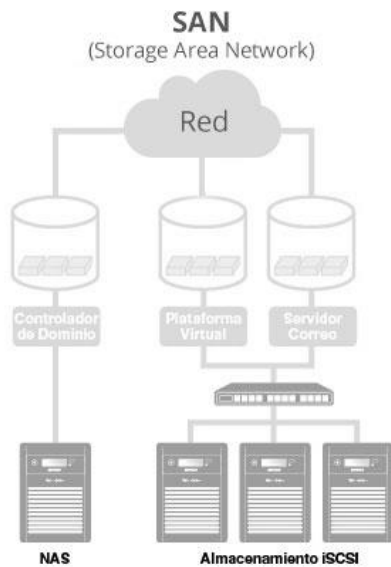


Fig. 1. Implementación de SAN en un entorno, nótese que puede existir un NAS dentro de la infraestructura, por lo que no es una situación en la que solo pueda usarse una o la otra

III. BLOCK STORAGE

La idea detrás del almacenamiento en bloque es simular el funcionamiento de un dispositivo de almacenamiento como un disco, en el que para escribir o leer sobre un tipo de dato, solo es necesario escribir la referencia en memoria, así pues, la idea es separar los datos en bloques haciéndolos indistinguibles, y cuando sea necesario hacer una operación sobre un conjunto de datos solo se van a los bloques que los componen a través de un identificador y se opera sobre ellos. Usualmente cuando se implementa un sistema de almacenamiento en bloque se hace sobre una arquitectura SAN pues el sistema decide donde poner los bloques de la forma más eficiente. Además, montar este sistema desacopla la relación de datos y entornos por lo que ya sean los usuarios o las aplicaciones pueden leer esta información y el sistema de almacenamiento en bloque vuelve a montar los bloques que los componen y los presenta a quien lo solicita.

Es efectivamente gracias a que los bloques tienen un identificador único, que implementar una solución de almacenamiento en bloque es tan eficiente por lo que los desarrolladores optan por esta opción cuando se necesita una rápida acción sobre el transporte de información, entre los casos comunes de uso se encuentra, implementaciones de bases de datos pues su eficiencia permite una baja latencia.

IV. FILE STORAGE

Lo más importante cuando se habla de file storage, es que su organización es jerarquizada, se compone enteramente de archivos y carpetas, las carpetas contienen archivos y subcarpetas, por lo que se empieza a ver como un árbol donde las hojas son los archivos y las ramas son subcarpetas, por lo que para acceder modificar o eliminar un archivo solo se requiere el path dentro del sistema de archivos

Si se quiere implementar un sistema de file storage, usualmente se hace sobre NAS pues el objetivo de un sistema jerarquizado es que sea centralizado, los casos de uso comunes van desde backups y centralizado de archivos en una local area network (LAN)

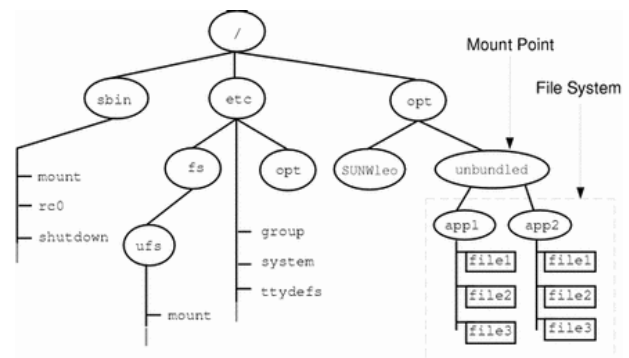


Fig. 1. Representación gráfica de un sistema de archivos Linux

V. OBJECT STORAGE

Ante el constante crecimiento de la información desestructurada tales como videos, imágenes, audios, etc. Era necesario un modelo que no fuera tan rígido como lo es el file storage y que no separe los datos pues no es conveniente separar un video o una canción, es por ello que, ante la necesidad de un nuevo modelo, nace el object storage.

La idea principal es que el acceso a este tipo de datos sea a partir de una petición http, el sistema está implementado a partir de un rest api, por lo que guardar, acceder, eliminar o modificar es bastante sencillo, gracias a que los datos están guardados de manera plana sobre el sistema de archivos, es fácilmente accesible gracias a los identificadores de los objetos, al eliminar todo rastro de estructura u organización sobre el sistema, deja escalar fácilmente, por lo que la única limitante para crecer son dispositivos de almacenamiento, puede estar montado sobre SAN o NAS ya depende más de los requerimientos de la solución

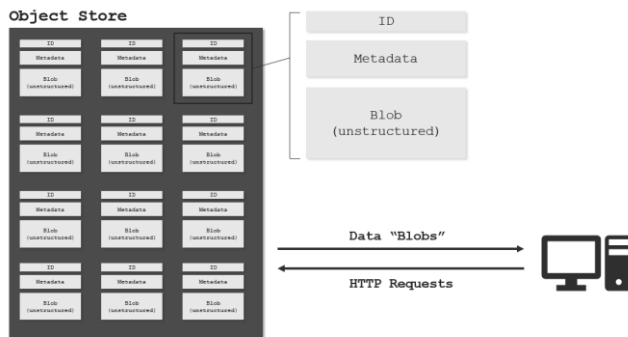


Fig. 1. Representación grafica de la implementación de un sistema de almacenamiento de objetos

Un objeto se define como una unidad de información que contiene un identificador, el contenido en sí, y la meta data asociada al contenido, todo este objeto es un repositorio de sí mismo en el sistema de almacenamiento, la consulta de este objeto se hace a partir de la petición al api rest llevando consigo la meta data para hacer más sencilla la búsqueda, no obstante, el sistema usa esta meta data para encontrar el identificador y así localizar el objeto

Los casos comunes de uso van por la ventaja de que permite manejar grandes volúmenes de información no estructurada por lo que no es difícil imaginar cuales, casos de negocio pueden aplicar a aplicaciones relacionadas, música, videos, logs, imágenes, por lo que no sería imposible pensar que YouTube, Reddit, Spotify manejen este modelo de almacenamiento

REFERENCIAS

- Es.wikipedia.org. n.d. Almacenamiento De Conexión Directa. [online] Available at: <https://es.wikipedia.org/wiki/Almacenamiento_de_conexi%C3%B3n_directa> [Accessed 21 August 2020].
- Es.wikipedia.org. n.d. Almacenamiento De Objetos (Informática). [online] Available at: <[https://es.wikipedia.org/wiki/Almacenamiento_de_objetos_\(inform%C3%A1tica\)](https://es.wikipedia.org/wiki/Almacenamiento_de_objetos_(inform%C3%A1tica))> [Accessed 21 August 2020].
- n.d. [online] Available at: <<https://www.redhat.com/en/topics/data-storage/file-block-object-storage>> [Accessed 21 August 2020].
- Ibm.com. n.d. Block Storage. [online] Available at: <<https://www.ibm.com/cloud/learn/block-storage>> [Accessed 21 August 2020].
- Ibm.com. n.d. File Storage. [online] Available at: <<https://www.ibm.com/cloud/learn/file-storage>> [Accessed 21 August 2020].
- Es.wikipedia.org. n.d. ISCSI. [online] Available at: <<https://es.wikipedia.org/wiki/ISCSI>> [Accessed 21 August 2020].
- Lavila, J., n.d. Comparativa SAN Vs. NAS - Infordisa. [online] Infordisa. Available at: <<https://www.infordisa.com/es/comparativa-san-vs-nas/>> [Accessed 21 August 2020].
- IONOS Digitalguide. n.d. Object Storage. [online] Available at: <<https://www.ionos.es/digitalguide/servidores/know-how/object-storage/>> [Accessed 21 August 2020].

- Ibm.com. n.d. Us-En_Cloud_Learnhub_Object-Storage-A-Complete-Guide. [online] Available at: <<https://www.ibm.com/cloud/learn/object-storage>> [Accessed 21 August 2020].
- Medium. n.d. File (NAS) Vs. Block (SAN) Vs. Object Cloud Storage. [online] Available at: <<https://medium.com/@zach.gollwitzer/file-nas-vs-block-san-vs-object-cloud-storage-c021d81fa3ff>> [Accessed 21 August 2020].