Warehouse-scale Computing (WSC) diseñados por Google

Carlos Estevez, Rubén Duarte, Sebastián Fonseca.
Universidad industrial de Santander.
Escuela de Ingeniería de Sistemas e Informática, Arquitectura de computadores.
Bucaramanga, Colombia.
2017

In the present paper a study of the Warehouse-scale computer was carried out, in order to determine its approach, its differences and similarities with the datacenters, which have been their applications and how it has been growing in these years responding to the demand of the market. In addition, an explanation of the main components and characteristics of the WSC model implemented by Google will be presented, and a parallel will be made with the SC3UIS system, which although not a WSC, is an HPC-SC.

En el presente trabajo se realizó un estudio del *Warehouse-scale computer,* con el fin de determinar su enfoque, cuáles son sus diferencias y similitudes con los *datacenters*, cuáles han sido sus aplicaciones y como ha ido creciendo en estos años respondiendo a la demanda del mercado. Además, se presentará una explicación de los principales componentes y características del modelo WSC implementado por Google, y se realizara un paralelo con el sistema de SC3UIS, que si bien no es un WSC, si es un HPC-SC.

1. INTRODUCCIÓN

Las últimas tendencias en computación, específicamente en la investigación y el desarrollo en materia de procesamiento y almacenamiento de grandes cantidades de datos del lado de los servidores han transformado el paisaje de los desarrollos tecnológicos, grandes empresas de tamaño mundial con importantes cantidades de recursos han impulsado nuevas tecnologías y nuevos negocios con miras a la solución de problemas y a suplir las necesidades de las personas. Es precisamente gracias a los WSC's que en los últimos años grandes empresas tecnológicas han logrado alcanzar sus objetivos en la búsqueda de materializar conceptos tales como el SaaS (*software as a service*, software como un servicio), que solo es concebible con la existencia de grandes infraestructuras. La idea de los WSC puede sintetizarse en: computadores de tamaño industrial dedicados a prestar servicios especializados a sus clientes. No debe confundirse con los tradicionales *datacenters* que son extensas 'granjas' de servidores en donde cada uno simplemente presta su servicio como una infraestructura en la cual se alojan los programas para cada cliente, y en donde un servidor puede ser compartido por diferentes entidades, cada WSC realiza solo una tarea y exactamente el mismo servicio es prestado a muchos clientes, más adelante estas ideas se profundizaran.

Por el momento las grandes compañías están utilizando la idea del WSC para mejorar su almacenamiento, ya que estos son más que un simple *datacenter* en donde las instalaciones tienen requisitos ambientales específicos, hardware heterogéneo, software de sistema, mientras que los WSC's son más integrales, tienen un sistema informático diseñado para ejecutar servicios masivos de Internet, hardware homogéneo, un conjunto común de recursos gestionados centralmente, y diseño integrado de instalaciones y maquinaria informática, entre otras características [7].

2. ESTADO DEL ARTE

Alojar aplicaciones web a gran escala y servicios en la nube es uno de los enfoques de WSC, ya que el costo de construcción y operación de los *datacenters* oscila entre decenas y cientos de millones de dólares. A medida que la computación se mueve hacia la nube, es sumamente importante aprovechar eficientemente los beneficios que proporciona los WSC's [2].

Otro de los usos que se le da a los WSC es el *cloud computing* que ha surgido como una forma rentable de satisfacer las demandas de las empresas y los consumidores, para esto se están construyendo un mayor número de WSC's que apoyan este cambio de paradigma [3].

Una aplicación bastante importante para destacar del WSC es que en la actualidad proporciona la infraestructura de computación y almacenamiento para grandes empresas tales como Netflix [4].

A medida que gran parte de la computación del mundo continúa moviéndose hacia la nube, la demanda de computación WSC se eleva. Como tal, es cada vez más importante que el rendimiento y la utilización de las máquinas alojadas en WSC se maximicen. Aunque Google ha diseñado y desplegado una de las infraestructuras de *datacenter* más grandes y más avanzadas del mundo, observamos que no está libre de ineficiencias [5].

3. MARCO TEORICO

Datacenter – Centro de procesamiento de datos:
Se denomina centro de procesamiento de datos (CPD) a aquella ubicación donde se concentran los recursos necesarios para el procesamiento de la información de una organización.
También se conoce como centro de cómputo' en Hispanoamérica y en España como centro de cálculo, centro de datos (por su equivalente en inglés, data center), centro de proceso de datos o centro de informática.
Dichos recursos consisten esencialmente en unas dependencias debidamente acondicionadas, computadoras y redes de comunicaciones [9].

Computacion en la nube - *cloud computing*:
En este tipo de computación todo lo que puede ofrecer un sistema informático se ofrece como servicio,2 de modo que los usuarios puedan acceder a los servicios disponibles "en la nube de Internet" sin conocimientos en la gestión de los recursos que usan. La computación en la nube son servidores desde Internet encargados de atender las peticiones en cualquier momento. Se puede tener acceso a su información o servicio, mediante una conexión a internet desde cualquier dispositivo móvil o fijo ubicado en cualquier lugar. Sirven a sus usuarios desde varios proveedores de alojamiento repartidos frecuentemente por todo el mundo. [9] [10] [11]

PUE (*Power Usage Effectiveness*) [6] es una medida de la eficiencia energética de un centro de datos, da a conocer que cantidad de energía es la que realmente es usada para el procesamiento de datos, y que cantidad es usada en otros destinos, tales como refrigeración

$$PUE = \frac{Potencia\ total\ de\ las\ instalaciones}{Potencia\ de\ los\ equipo\ de\ TI}$$

Nodo: Puede referirse a simples computadores, sistemas multiprocesador o estaciones de trabajo. En redes cada una de las máquinas es un nodo, y cuando hablamos de internet, cada servidor constituye también un nodo. Un *cluster* puede estar conformado por nodos dedicados o por nodos no dedicados . En un *cluster* con nodos dedicados, los nodos no disponen de teclado, ratón ni monitor y su uso está exclusivamente dedicado a realizar tareas relacionadas con el clúster, mientras que un *cluster* no dedicado sucede lo contrario.

4. WSC's DE GOOGLE

Siendo la competencia directa de marcas como Apple, Microsoft y AT&T en la disputa del título de empresa más valiosa del mundo según revistas como Forbes[13], manteniéndose año a año en lo más altos de los escalafones en dichos rankings y ubicándose en el primer puesto de esta categoría según conteos como el de Millward Brown no es de sorprender que Google, marca valorada en 82,500 millones de dólares según Forbes y 229,198 millones de dólares según Millward Brown[14] sea una compañía que se pueda permitir ostentar el título de innovadora tecnológica, de imponer tendencias en su ámbito y de marcar la hoja de ruta de cualquier *startup* que quiera participar en el mercado de la tecnología. Para lograr tales hazañas Google se vale de la constante innovación y un invariable mejoramiento de su infraestructura

tecnológica y es precisamente allí donde se centra el foco de nuestro interés, en la tecnología y diseño subyacente para la implementación de los *warehouse-scale computers* de Google.

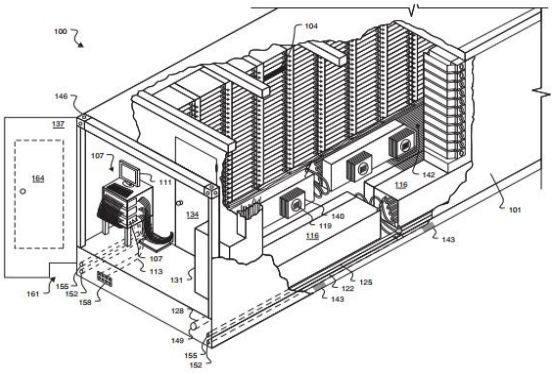
Millones de usuarios de la internet que diariamente usan servicios alojados en la nube ignoran parcial o totalmente la compleja red de edificaciones y cables que permite el desarrollo de dichas actividades, pues consideran casi mágica la posibilidad de comunicarse con alguien al otro lado del mundo o de ver en vivo la presentación de sus cantantes favoritos desde cualquier lugar del planeta, incluso el presente artículo fue posible mediante la colaboración de los autores usando dichos servicios y es desde el origen menos esperado que Google sorprende con su innovación pues dichos servicios son prestados mediante el uso de contenedores de carga, a continuación, describiremos detalladamente el estado de esta temática actualmente y sus avances, presentados por el gigante de la tecnología.

A. Contenedores

Junto a Google otro gigante se suma a dicha idea, Microsoft implementa también la construcción de sus WSC mediante el uso de contenedores de carga, esto con miras al modularidad, así la tarea de los WSC se limita a proveer de redes, energía y agua fría a cada módulo, a su vez cada módulo suministra redes, energía y agua fría a cada servidor. Cada WSC de Google cuenta con 45 de estos contenedores, un contenedor tiene capacidad

para 1160 servidores, por lo que 45 contenedores tienen espacio para 52.200 servidores. (El WSC materia del presente artículo tiene unos 40.000).

(a)



(b)

Fig. 1 Vista transversal de los contenedores usado por Google.

(a)Vista real de los contenedores

(b)Distribución de los equipos en el interior

B. Enfriamiento y potencia en los WSC de Google

El enfriamiento de los dispositivos corre por cuenta de ductos en el piso de los contenedores, además los racks en los que se ubican los servidores están fijos en el techo, lo que permite el flujo de aire que se muestra en la figura 2, el aire frío entra, es aspirado hacia los aparatos y por la geometría del contenedor el aire caliente puede salir sin mezclarse con el frio, dichos módulos se mantienen a temperaturas de 27°C, esta temperatura es cuidadosamente calibrada mediante el uso de calefacción o enfriamiento según sea necesario para evitar daños en los servidores, lo cual aporta para una mayor eficiencia energética.

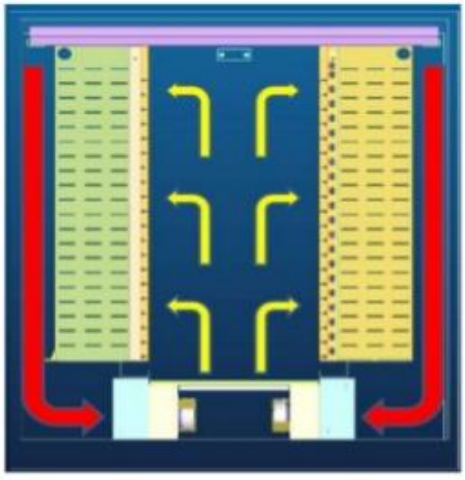


Fig. 2 Flujo del aire dentro de los contenedores.

En la figura 3 puede observarse un servidor diseñado por Google para sus WSC, a cada *motherboard* se suministra cerca de 12 volts de electricidad lo cual permite a Google una eficiencia energética de 92%, correspondiente a la categoría *gold* en fuentes de alimentación.

En cada estante se ubica una UPS lo que evita el sobredimensionamiento en capacidad.

Una medida destacable en esta implementación de Google son los valores de efectividad en el uso de energía (PUE, *power usage effectiveness*), pues la compañía logró una medida de 1.3, tal logro se atribuye a la detallada medición de uso energético (mediante dispositivos llamados *current transformers*, transformadores de corriente) que permite a los ingenieros ajustar el uso de corriente a través del tiempo. Las estadísticas de dicha medida pueden detallarse en reportes presentados por la empresa y que aquí se presentan en la figura 4.

C. Servidores en los WSC's de Google

En la figura 3 puede apreciarse una fotografía de los servidores usados por Google. Cada servidor tiene dos *sockets*, cada uno posee un procesador dual-core AMD Opteron corriendo a 2.2 GHz.

En la fotografía pueden observarse además 8 DIMMS, y generalmente este tipo de servidores cuentan con 8 GB de DDR2 DRAM, una novedad importante es que el bus de la memoria fue *downclocked* (a 533 MHz desde el estándar de 666 MHz), ya que dicha velocidad tiene poco impacto en el desempeño, pero un impacto mayor en el uso de energía.

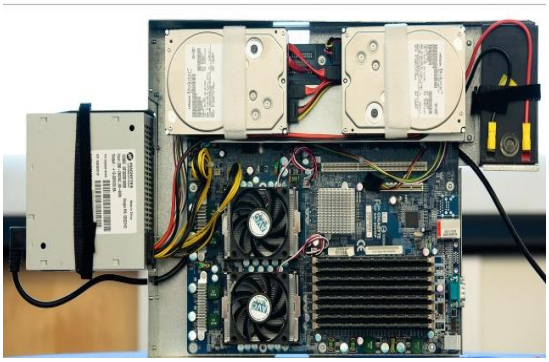


Fig. 3 Servidores usados por los WSC de Google.

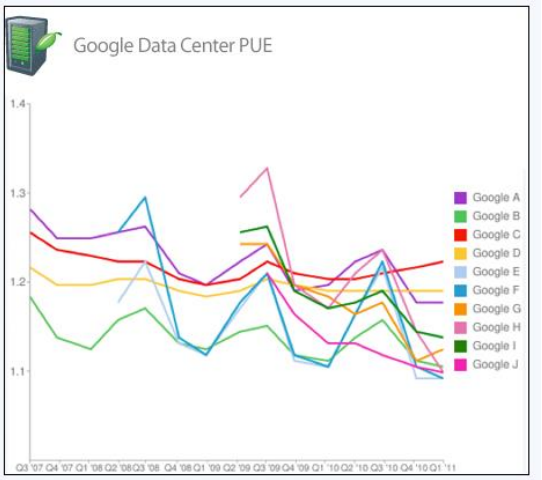


Fig. 4 Mediciones de la PUE de 10 WSCs de Google a través del tiempo.

El servidor básico tiene una sola tarjeta de red (NIC) para un enlace Ethernet de 1 Gbit/seg además de dos unidades de disco SATA. Pero estas prestaciones pueden escalarse por ejemplo con una bandeja que alberga 10 discos SATA, y una segunda tarjeta de red (ya que un servidor puede saturar una conexión de 1 Gbit/seg) y así un nodo básico se convierte en un nodo de almacenamiento.

Google desplegó 40.000 servidores en lugar de 52.200 para los 45 contenedores ya que el nodo de almacenamiento ocupa dos ranuras en el *rack*. En este WSC específico, la proporción era de aproximadamente dos nodos de cálculo para cada nodo de almacenamiento, pero esa proporción varía en cada WSC de Google. Así las cosas, este WSC contaba con alrededor de 190.000 discos (en 2007), o un promedio de casi 5 discos por servidor.

D. Redes en los WSC's de Google

Los 40.000 servidores se dividen en 'grupos' de más de 10.000 unidades cada uno (estos grupos son llamados *clusters*) los *switches* en cada *rack* cuentan con 48 puertos *Ethernet*, 40 para conectar los servidores y 8 para conexiones entre *switches*, cada *cluster* cuenta con 480 puerto, algunos a velocidades de 1 Gbit/sec para conectar los *switches* y algunos con velocidades de 10 Gbit/sec para interconexiones de los *routers*, cada WSC usa dos *routers* para conectarse al mundo exterior y permitir la redundancia, así si uno de ellos falla el WSC completo no quedará deshabilitado. Para aprovechar lo mejor posible el ancho de banda a

cada puerto en un *switch* se conectan varias máquinas, generalmente en una tasa de 20:1. pero para aplicaciones con mayor demanda de recursos de red esta relación puede reducirse hasta 5:1.

E. Monitoreo y reparaciones en los WSCs de Google

Un solo operario es a la vez responsable de hasta 1000 servidores, esto es posible gracias a que Google implementa software para escanear y diagnosticar constantemente sus dispositivos, muchos de los inconvenientes que son rutinarios pueden ser solucionados automáticamente, luego el servidor es reiniciado y el *software* instalado nuevamente. Las máquinas que necesitan reparaciones son ubicadas por lotes, y cada máquina posee un *ID,* si el diagnóstico dado por el *software* es confiable la máquina es reparada sin una revisión manual, si no hay un diagnóstico o este es desconfiable entonces se procede a una revisión manual. El objetivo aquí es nunca tener más del 1% de los servidores en reparación, pues esto aumenta los costos, el tiempo promedio para la reparación de un nodo es una semana, aunque los técnicos tardan mucho menos, pero hay que tener en cuenta que por ejemplo cuando se corre un diagnóstico y un reinicio esto puede durar incluso solo minutos, pero las pruebas de trabajo para verificar que el nodo reparado funciona correctamente pueden tardar horas.

5. PARALELO

Ahondemos en el tema del WSC original, en el año 2005 GOOGLE tomo esta iniciativa y la llevó a la realidad, inicialmente en un almacén que tenía 75000 pies cuadrados de área, y usó, 30 contenedores de 40 pies de largo cada uno para realizar la implementación al interior de cada uno de estos, ya que uno de sus objetivos era el aprovechamiento del espacio, GOOGLE, decidió apilar pilas de 2 contenedores, haciendo más eficiente el uso del espacio, dichos contenedores contenían 2 filas principales de 29 racks, cada uno de estos, contenía a su vez 20 servidores. Estas filas de racks, estaban enfrentadas dentro del contenedor con el fin de hacer fluir una corriente de aire frío por en medio de las 2 filas y sacar el calor de los mismos. Como la idea consistía en poner todo dentro del contenedor, los sistemas de respaldo de energía a su vez fueron introducidos dentro de este, para el WSC de GOOGLE, el manejo de energía era crucial para que fuese eficiente su uso, por lo tanto decidieron usar una fuente de 12V lo cual les permitía reducir las pérdidas de potencia y a su vez usar baterías de ácido como sistema de respaldo de energía. De estos contenedores solo salían las conexiones de red y de alimentación, en la arquitectura inicial se usaron por cada rack procesadores AMD Opteron Dual Core a 2.2Ghz , 8 Gb de memoria RAM y usaba 2 HDD.[28]

Al comparar el sistema WSC de GOOGLE con GUANE-1, primero debemos evidenciar, que las arquitecturas, son diferentes, y los propósitos de los mismos, también lo son, pues el WSC de GOOGLE proporciona una plataforma robusta para ejecutar grandes volúmenes de solicitudes de los clientes, sin necesidad de realizar cálculos complejos, los cuales necesitarían ciertos cambios en la arquitectura para poder realizarlos de manera que su costo de ejecución sea pagable, a diferencia de WSC, GUANE-1 está un solo cluster en el Parque Tecnológico Guatiguará de la UIS, este está compuesto por 16 nodos HP Proliant SL390s G7, estos nodos usan como procesador el Intel XEON E5645 2,40GHz (12 núcleos), estos nodos implementan 104 Gb de memoria RAM, 8 GPU NVIDIA TESLA M2050 y poseen conexión Gigabit a internet.[28]

Al analizar al detalle los tipos de arquitectura en los sistemas mencionados, podemos observar, que el servicio GUANE-1 es una plataforma de cálculo científico, donde es prioritario el tiempo de procesamiento de los datos, y el volumen que se van a procesar, por esta razón, estos racks poseen una gran cantidad de memoria RAM y su procesador tiene varios núcleos y cuenta con el apoyo al momento de procesar los datos de la tecnología de la GPU, NVIDA CUDA, la cual permite optimizar el tiempo de ejecución para ciertos procesos.[29]

6. CONCLUSION

Como una conclusión se puede decir que en cierto sentido los *warehouse-scale computing* son simples, son sólo unos pocos cientos de servidores unidos a través de una red local. En realidad, la construcción de una plataforma de computación a gran escala rentable que tiene los requisitos de fiabilidad y programación necesarios para la próxima generación de cargas de trabajo de *cloud computing*.

Podemos también identificar que los WSC nop son comparables con los datacenters pues tienen implementaciones diferentes.

REFERENCIAS

[1] Mohney Warehouse-scale computers (Yes, really),Green Data Center News, http://www.greendatacenternews.org/articles/750666/warehouse-scale-computers-yes-really-by-doug-mohne/

[2] Mars, J., Tang, L., Hundt, R., Skadron, K., & Soffa, M. L. (2011, December). Bubble-up: Increasing utilization in modern warehouse scale computers via sensible co-locations. In Proceedings of the 44th annual IEEE/ACM International Symposium on Microarchitecture (pp. 248-259). ACM.

[3] Yeo, S., & Lee, H. H. S. (2012). Simware: A holistic warehouse-scale computer simulator. Computer, 45(9), 48-55.

[4] Delimitrou, C., & Kozyrakis, C. (2013). The netflix challenge: Datacenter edition. IEEE Computer Architecture Letters, 12(1), 29-32.

[5] Tang, L., Mars, J., Zhang, X., Hagmann, R., Hundt, R., & Tune, E. (2013, February). Optimizing Google's warehouse scale computers: The NUMA experience. In High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on (pp. 188-197). IEEE.

[6] Patterson, D., & Hennessy, J. L. (2012). Computer architecture: a quantitative approach. Elsevier.

[7] Holzle_WarehouseScaleComputers_2010-08-27-08-19

[18] Barroso, L. A., Clidaras, J., & Hölzle, U. (2013). The datacenter as a computer: An introduction to the design of warehouse-scale machines. Synthesis lectures on computer architecture, 8(3), 1-154.

[19] Armbrust, Michael; Fox, Armando; Griffith, Rean; Joseph, Anthony D.; Katz, Randy; Konwinski, Andy; Lee, Gunho; Patterson, David et al. (1 de abril de 2010). «A View of Cloud Computing». Commun. ACM 53 (4): 50-58. ISSN 0001-0782. Consultado el 1 de marzo de 2016

[20] William Y. Chang, Hosame Abu-Amara, Jessica Feng Sanford, Transforming Enterprise Cloud Services, London: Springer, 2010, pp. 55-56

[21] Ko, Ryan K. L. Ko; Kirchberg, Markus; Lee, Bu Sung (2011). «From System-Centric Logging to Data-Centric Logging – Accountability, Trust and Security in Cloud Computing». Proceedings of the 1st Defence, Science and Research Conference 2011 – Symposium on Cyber Terrorism, IEEE Computer Society, 3–4 August 2011, Singapore.

[22] Jimmy Lin and Chris Dyer. Data Intensive Text Processing with MapReduce.

[23] Forbes staff. Las marcas más valiosas del mundo en 2016. *Forbes* [en línea]. 11 de mayo de 2016. [fecha de consulta: 31 Enero 2017]. Disponible en: <http://www.forbes.com.mx/las-marcas-mas-valiosas-del-mundo-2016/#gs.sQHBQgo>.

[24] Forbes staff. Las marcas más valiosas del mundo en 2016. *Forbes* [en línea]. 8 de junio de 2016. [fecha de consulta: 31 Enero 2017]. Disponible en: <http://www.forbes.com.mx/las-10-

marcas-mas-valiosas-del-mundo-2016/#gs.KRoEDkA >.

[25] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Kowinski, et al. *Above the Clouds: A Berkeley View of Cloud Computing* [pdf]. Febrero 10 de 2009. [fecha de consulta: 31 Enero 2017]. Disponible en:

<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf >.

[26] L. A. Barroso, J. Clidaras, U. Hölzle.(2013). *The Datacenter as a Computer. An Introduction to the Design of Warehouse-Scale Machines* [ebook]. Madison:Morgan & Claypool. [fecha de consulta: 31 Enero 2017]. Disponible en:

<http://web.eecs.umich.edu/~mosharaf/Readings/DC-Computer.pdf >.

[27] (2007). *THE GREEN GRID DATA CENTER POWER EFFICIENCY METRICS: PUE AND DCiE* [pdf]. [fecha de consulta: 31 Enero 2017]. Disponible en:

<http://www.premiersolutionsco.com/wp-content/uploads/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf >.

[28]Supercomputador GUANE http://www.sc3.uis.edu.co

[29] NVIDIA TESLA M 2050 http://www.nvidia.es/object/product_tesla_M2050_M2070_es.html