

El caso por Computación en la nube

Roberto L. Grossman

Universidad de Illinois en Chicago y Open Data Group

Para comprender las nubes y la computación en la nube, primero debemos comprender los dos tipos diferentes de nubes. El autor distingue entre nubes que brindan instancias de cómputo bajo demanda y aquellas que brindan capacidad de cómputo bajo demanda.

definición estándar, pero una buena descripción de trabajo es decir que *las nubes*, los grupos de computadoras distribuidas, brindan recursos y servicios bajo demanda a través de una red, generalmente Internet, con la escala y confiabilidad de un centro de datos. Este artículo ofrece una introducción rápida a la computación en la nube. Cubre varios tipos diferentes de nubes, describe las novedades de la computación en la nube y analiza algunas de las ventajas y desventajas que ofrecen las nubes.

Dos tipos diferentes pero relacionados de nubes son las que proporcionan *instancias* informáticas bajo demanda y las que proporcionan *capacidad* informática. Bajo demanda. Ambos usan máquinas similares, pero el primero está diseñado para escalar al proporcionar instancias informáticas adicionales, mientras que el segundo está diseñado para admitir aplicaciones de uso intensivo de datos o computación a través de la capacidad de escalamiento.

Servicios EC2 de Amazon (www.amazon.com/ec2) son un ejemplo de la primera categoría. Una pequeña instancia informática EC2 cuesta 0,10 USD por hora y ofrece la potencia informática aproximada de un procesador Opteron 2007 o Xeon 2007 de 1,0 a 1,2 GHz, con 1,7 Gbytes de memoria, 160 Gbytes

de espacio disponible en disco y rendimiento moderado de E/S. El sistema Eucalyptus (<http://eucalyptus.tus.cs.ucsb.edu>) es una opción de código abierto que proporciona instancias informáticas bajo demanda y comparte las mismas API que la nube EC2 de Amazon.

MapReduce de Google es un ejemplo de la segunda categoría. Un trabajo reciente da una idea de cómo funciona; en este caso, los investigadores realizaron una evaluación comparativa en un clúster que contenía aproximadamente 1800 máquinas, 1 cada una de las cuales tenía dos procesadores Intel Xeon de 2 GHz, memoria de 4 Gbytes y dos discos IDE de 160 Gbytes. Los investigadores utilizaron MapReduce en el clúster para ejecutar el punto de referencia de TeraSort (<http://research.microsoft.com/baron/TeraSortBenchmark>), cuyo objetivo es clasificar 1010 Registros de 100 bytes (aproximadamente 1 Tbyte de datos). La aplicación requirió aproximadamente 850 segundos para completarse en este clúster. El sistema Hadoop (<http://hadoop.apache.org/core>) es un sistema de código abierto que implementa MapReduce.

Las nubes que proporcionan instancias informáticas bajo demanda pueden usar estas instancias para proporcionar software como servicio (SaaS), como lo hace Salesforce.com con su producto, o para proporcionar una plataforma como servicio (PaaS), como lo hace Amazon con su producto EC2.

Computación en la nube

¿Qué hay de nuevo?

Ahora que hemos cubierto los conceptos básicos de la computación en la nube, es importante entender qué hay de nuevo en ella. Los servicios y recursos a pedido han estado disponibles a través de Internet durante algún tiempo, pero el mayor enfoque actual en la computación en la nube se debe a tres diferencias importantes:

- **Escala.** Algunas empresas que dependen de la computación en la nube tienen infraestructuras que escalan en varios (o más) centros de datos.
- **Simplicidad.** Antes de los servicios informáticos basados en la nube, escribir código para la informática distribuida y de alto rendimiento era relativamente complicado y, por lo general, requería trabajar con servicios basados en grid, desarrollar código que pasara mensajes explícitamente entre nodos y emplear otros métodos especializados. Aunque la simplicidad está en el ojo del espectador, la mayoría de las personas sienten que las API del servicio de almacenamiento basado en la nube y las API informáticas de estilo MapReduce son relativamente simples en comparación con los métodos anteriores.
- **Precios.** La computación en la nube a menudo se ofrece con un modelo de precios que le permite pagar sobre la marcha y solo por los servicios que necesita. Por ejemplo, si necesita 1000 instancias informáticas adicionales durante una hora, paga solo por estas 1000 instancias informáticas y solo por la hora que las usa. No se requieren gastos de capital.

El impacto ha sido revolucionario: al usar Google File System (GFS) y MapReduce, o Hadoop Distributed File System con su implementación de MapReduce, es relativamente fácil para un proyecto realizar un cálculo de más de 10 Tbytes.

de datos utilizando 1.000 nodos. Hasta hace poco, esto habría estado fuera del alcance de la mayoría de los proyectos. Varios otros cambios y mejoras también han elevado el perfil de la computación en la nube.

Nubes privadas vs. alojadas

La administración, el costo y la seguridad de las nubes dependen de si una organización elige comprar y operar su propia nube u obtener servicios en la nube de un tercero. Una *nube privada* está dedicada al uso interno de una sola organización; puede ser administrado por la propia organización o subcontratado a un tercero para que lo opere. De manera similar, una nube privada puede ser propiedad de la propia organización o arrendada por la organización. Por el contrario, una nube *pública o alojada* es gestionada por otro o

Organización que proporciona servicios en la nube a una variedad de clientes de terceros que utilizan los mismos recursos en la nube. Google, por ejemplo, utiliza GFS, MapReduce y BigTable internamente como parte de sus servicios de nube privada; en el momento de escribir este artículo, estos servicios no estaban disponibles para terceros. Por el contrario, los servicios en la nube alojados, como EC2, S3 y SimpleDB de Amazon, están abiertos a cualquier persona que tenga una tarjeta de crédito, incluso a las 3 a. m.

Es importante tener en cuenta que Google utiliza su nube privada para proporcionar aplicaciones alojadas basadas en la nube, como su correo electrónico y servicios de oficina, a usuarios externos regulares.

Precios elásticos basados en el uso

La computación en la nube generalmente se ofrece con un modelo basado en el uso en el que paga solo por los recursos de la nube que requiere una computación en particular. Los cómputos que requieren recursos adicionales simplemente solicítelos desde la nube (hasta la capacidad total de la nube). A veces, los términos *computación elástica* o *de utilidad* se utilizan para describir esta capacidad de una nube para proporcionar recursos adicionales cuando se requieren. S3 y EC2 de Amazon utilizan este modelo de precios.

Las organizaciones, por lo tanto, tienen varias opciones para obtener servicios en la nube, incluida la ejecución de sus propias nubes privadas o la compra de servicios en la nube de un tercero utilizando el modelo de precios elástico basado en el uso. Este tipo de fijación de precios también ofrece otras dos ventajas importantes:

- No requiere inversiones iniciales; en cambio, como un servicio bajo demanda, los usuarios pagan por la capacidad a medida que la necesitan.
- Permite a los usuarios acceder a la capacidad exactamente cuando la necesitan. Para las aplicaciones Web 2.0, esto significa que el modelo puede admitir 100 usuarios un día y 10 000 al día siguiente.

Para obtener una mejor comprensión de la informática de servicios públicos, supongamos que tiene el requisito de operar 100 servidores en el transcurso de tres años. Una opción es alquilarlos a 0,40 USD por hora de instancia, lo que costaría aproximadamente

años 100 servidores $0,40 \text{ USD} \times 24 \text{ horas} \times 365 \text{ días} \times 100 = 35280 \text{ USD}$ *

Otra opción es comprarlos. Supongamos que el costo de comprar cada servidor es de \$1500, que necesita

dos miembros del personal a \$100,000 por año para administrar los servidores, y que los servidores requieren 150 vatios cada uno, con un costo de electricidad de \$0.10 por kilovatio-hora, lo que eleva el costo anual para operar los 100 servidores a \$13,140. Esta opción costaría aproximadamente

$$100 \text{ servidores} \times \$2,500 \text{ por año} + \$100,000 \text{ salarios/año} + 3 \text{ años} = \$789,420.$$

Por lo tanto, si tuviera que ejecutar los servidores al 100 por ciento de utilización, comprar los 100 servidores es menos costoso. Sin embargo, si tuviera que ejecutarlos con una utilización del 75 por ciento o menos, usar un estilo de nube bajo demanda sería menos costoso.

Por supuesto, estos números son solo estimaciones y no he considerado todos los costos, pero incluso a partir de este ejemplo simple, está claro que usar un modelo de computación de utilidad de pago por uso es preferible para muchos casos de uso.

Algunas ventajas y desventajas

La computación en la nube brinda varios beneficios importantes sobre el modelo dominante actual en el que una empresa compra computadoras en un rack a la vez y las opera por sí misma. En primer lugar, el modelo de precios basado en el uso de la computación en la nube ofrece varias ventajas, que incluyen gastos de capital reducidos, una barrera de entrada baja y la capacidad de escalar según lo requiera la demanda, así como admitir aumentos breves en la capacidad. En segundo lugar, los servicios en la nube disfrutan de las mismas economías de escala que brindan los centros de datos. Al brindar servicios a la escala de un centro de datos, es posible brindar operaciones, continuidad comercial y seguridad de manera más eficiente que cuando se brindan estos servicios en un rack a la vez. Por esta razón, el costo unitario de los servicios basados en la nube suele ser menor que el costo si los servicios los proporcionara directamente la propia organización. Finalmente, las arquitecturas de computación en la nube han demostrado ser muy escalables; por ejemplo, los servicios de almacenamiento basados en la nube pueden administrar fácilmente un petabyte de datos, mientras que administrar esta cantidad de datos con una base de datos tradicional es problemático.

Por supuesto, la computación en la nube también tiene algunas desventajas. En primer lugar, debido a que los servicios en la nube suelen ser remotos (al menos para los servicios en la nube alojados), pueden sufrir los problemas relacionados con la latencia y el ancho de banda asociados con cualquier aplicación remota. En segundo lugar, porque los servicios en la nube alojados sirven múltiples

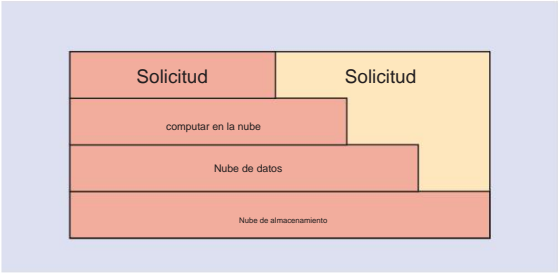


Figura 1. Modelo en capas. Algunas nubes que brindan capacidad informática bajo demanda usan capas de servicios, formando una pila de servicios en la nube.

varios clientes, pueden surgir varios problemas relacionados con varios clientes que comparten la misma pieza de hardware. Por ejemplo, si la aplicación de un usuario compromete el sistema, también puede comprometer las aplicaciones de otros usuarios que comparten el mismo sistema. Además, tener datos accesibles a terceros (como un proveedor de servicios en la nube) puede presentar problemas de seguridad, cumplimiento y reglamentarios.

Servicios en capas

Una *nube* de almacenamiento proporciona servicios de almacenamiento (basados en bloques o archivos); una *nube* de datos proporciona servicios de gestión de datos (basados en registros, columnas u objetos); y una *nube informática* proporciona servicios informáticos. A menudo, están en capas (servicios informáticos sobre servicios de datos sobre servicios de almacenamiento) para crear una pila de servicios en la nube que actúa como una plataforma informática para desarrollar aplicaciones basadas en la nube; ver la figura 1.

Computación paralela sobre nubes

En esencia, MapReduce es un estilo de programación paralela compatible con nubes de capacidad según demanda. Un buen ejemplo ilustrativo de cómo funciona algo como MapReduce es calcular un índice invertido en paralelo para una gran colección de páginas web almacenadas en una nube.

Supongamos que cada nodo *i* en la nube almacena páginas web *p*_{1,1}, *p*_{1,2}, *p*_{1,3}, ..., y que cada página web *p*_{1,1}, *p*_{1,2}, *p*_{1,3}, ... contiene palabras (términos) *w*_{1,1}, *w*_{1,2}, *w*_{1,3}, Una estructura básica pero importante en la recuperación de información es un índice invertido, es decir, una lista

$$(w_1; p_{1,1}, p_{1,2}, p_{1,3}, \dots)$$

$$(w_2; p_{2,1}, p_{2,2}, p_{2,3}, \dots)$$

$$(w_3; p_{3,1}, p_{3,2}, p_{3,3}, \dots),$$

donde la lista está ordenada por la palabra *w_j*, y asociada con cada palabra *w_j* hay una lista de todas las páginas web *p_i* que contienen esa palabra.

Computación en la nube

MapReduce utiliza un modelo de programación que procesa una lista de pares <clave, valor> para generar otra lista de pares <clave', valor'>. La lista inicial de pares <clave, valor> se distribuye entre los nodos de la nube. En la fase de *mapa*, cada página web *pi* se procesa de forma independiente en su nodo local para producir una lista de salida de varios pares clave-valor <*wj*, *pi*>, uno para la palabra *wj* en la página. Luego, una función de partición *h(wj)* asigna cada tecla (una palabra *wj* en este ejemplo) una máquina en la nube para su posterior procesamiento. Esto se denomina fase de *barajado* y, en general, los nodos involucrados en el cómputo envían datos a otros nodos involucrados en el cómputo según lo determinado por la función de partición *h(wj)*. En la siguiente fase, llamada fase de clasificación, cada nodo en la nube ordena los pares clave-valor <*wj*, *pi*> según la clave *wj*. En la fase final—denominada fase de *reducción*: los pares clave-valor con la misma clave *wj* se fusionan para crear la fase invertida

Los terceros pueden aprovechar las economías de escala para brindar un nivel de seguridad que podría no ser rentable para las empresas más pequeñas.

índice. Entonces, con MapReduce, el programador define una función Map y Reduce, mientras que el sistema proporciona las funciones Shuffle y Sort.³

Consideremos otro ejemplo: archivos de registro que describen el uso de recursos de una entidad. Es importante analizar los archivos de registro para identificar anomalías que indiquen si un recurso en particular se ha visto comprometido. Para archivos de registro pequeños, esto es fácil de hacer con una base de datos, pero, a medida que crece el tamaño de los archivos de registro, es difícil administrarlos solo con una base de datos. Sin embargo, las nubes pueden administrar fácilmente incluso colecciones muy grandes de archivos de registro, y los cálculos de estilo MapReduce pueden identificar fácilmente patrones anómalos indicativos de compromisos.

Seguridad

La seguridad es un área de la computación en la nube que presenta algunos desafíos especiales. Para las nubes alojadas, el primer desafío es simplemente que un tercero es responsable tanto de almacenar los datos como de protegerlos. En el lado positivo, los terceros pueden aprovechar las economías de escala para brindar un nivel de seguridad que podría no ser rentable.

para empresas más pequeñas, pero una desventaja es que dos o más organizaciones pueden compartir el mismo recurso físico y no ser conscientes de ello.

Para algunas aplicaciones en la nube, la seguridad aún es algo inmadura. Hadoop, por ejemplo, actualmente no tiene autenticación a nivel de usuario o controles de acceso, aunque ambos se esperan en una versión posterior. Afortunadamente, no existe ninguna dificultad técnica per se para proporcionar estas herramientas a las nubes. Sector,5 que también brinda capacidad de cómputo bajo demanda, ofrece autenticación, autorización y controles de acceso y, según lo medido por el punto de referencia TeraSort, es más rápido que Hadoop (<http://sector.sourceforge.net>).

Estándares, interoperabilidad y puntos de referencia

Las organizaciones que desarrollan aplicaciones basadas en la nube están interesadas en marcos que permitan que las aplicaciones se transfieran fácilmente de una nube a otra y que interactúen con diferentes servicios basados en la nube. Por ejemplo, con un marco de interoperabilidad adecuado, una aplicación en la nube podría cambiar de un proveedor a otro que ofrezca un menor costo o una mayor variedad de servicios en la nube.

Las API de Amazon (www.aws.amazon.com) se han convertido en el estándar de facto para las nubes que proporcionan instancias bajo demanda. Las aplicaciones basadas en la nube que usan esta API disfrutan de portabilidad e interoperabilidad; por ejemplo, Eucalyptus usa estas API y las aplicaciones que se ejecutan en el servicio EC2 de Amazon pueden, a su vez, ejecutarse en una nube de Eucalyptus. Desafortunadamente, para las nubes que brindan capacidad bajo demanda, la portabilidad y la interoperabilidad son mucho más problemáticas. Hadoop es, con mucho, el sistema más frecuente que proporciona capacidad bajo demanda, pero, por ejemplo, no es sencillo que una aplicación Hadoop MapReduce se ejecute en otra nube de capacidad bajo demanda escrita en C++.⁵

Si bien es posible que aún sea demasiado pronto para que surjan completamente los estándares, varias organizaciones están intentándolos, incluido un esfuerzo del Foro de interoperabilidad de computación en la nube (www.cloudforum.org/) y por Open Cloud Consortium (www.opencloudconsortium.org). Los marcos de trabajo basados en servicios para las nubes también han debutado recientemente; por ejemplo, Thrift es un marco de trabajo de software para el desarrollo escalable de servicios multilingües que se basa en un motor de generación de código (<http://incubator.apache.org/thrift>). El ahorro hace

Es más fácil para las aplicaciones basadas en la nube acceder a diferentes nubes de almacenamiento, como Hadoop y Sector. Un lenguaje común también podría ayudar al proporcionar una forma interoperable para que las aplicaciones accedan a los servicios informáticos en varias nubes diferentes; Hasta ahora, varias personas han intentado proporcionar un lenguaje para la programación paralela al estilo de MapReduce, incluidos algunos que amplían SQL de una manera que admite este estilo de programación, pero todavía no ha surgido ningún lenguaje como el claro ganador.

Un desafío estrechamente relacionado es crear un estándar que permita la interoperabilidad de diferentes nubes. Quizás la infancia de Internet podría guiar este tipo de esfuerzo; en ese momento, cualquier organización que quisiera una red configurada por su cuenta, por lo que enviar datos entre redes era bastante difícil. La introducción de TCP y los protocolos y estándares de Internet relacionados remedió esta situación, pero muchas empresas con productos de red se resistieron durante algún tiempo. Hoy en día, nos encontramos en una posición algo análoga: aunque los proveedores de servicios en la nube han retrocedido un poco en la estandarización, la capacidad de diferentes nubes para interoperar fácilmente permitiría una nueva clase interesante de aplicaciones.

Al igual que con los estándares y un lenguaje común, la computación en la nube aún no tiene puntos de referencia bien establecidos. El método más común para medir el rendimiento de la nube hasta la fecha es el punto de referencia de TeraSort. Para las nubes que proporcionan instancias bajo demanda, ha surgido un punto de referencia reciente llamado Cloudstone.6 Cloudstone es un conjunto de herramientas que consiste en una aplicación social Web 2.0 de código abierto, un conjunto de herramientas para generar cargas de trabajo, un conjunto de herramientas para monitorear el desempeño, y una metodología recomendada para calcular una métrica que cuantifique los dólares por usuario por mes que requiere una nube determinada.

Para las nubes que brindan capacidad según demanda, un punto de referencia reciente llamado MalStone (code.google.com/p/malgen/) también ha surgido. MalStone se basa en el ejemplo del archivo de registro de un cálculo de MapReduce que describí anteriormente. Incluye código para generar eventos sintéticos y un cálculo de MapReduce recomendado.

puting" se ha movido de un solo com
Con la computación en la nube, la "unidad de com
computadora o rack de computadoras a un centro de
datos de computadoras. Para decirlo simplemente, la unidad de com

poner es ahora el centro de datos. La computación en la nube no solo ha escalado la computación al centro de datos, sino que también ha introducido software, sistemas y modelos de programación que reducen significativamente la complejidad del acceso y uso de estos recursos.

Igual de importante, con modelos de precios elásticos basados en el uso, una persona u organización paga solo por las instancias informáticas o la capacidad informática que requiere y solo cuando las requiere.

En verdad, esto es revolucionario.



Referencias

1. J. Dean y S. Ghemawat, "MapReduce: Procesamiento de datos simplificado en grandes clústeres", *Comm. ACM*, vol. 51, núm. 1, 2008, págs. 107–113.
2. S. Ghemawat, H. Gobioff y S.-T. Leung, "El sistema de archivos de Google", *Proc. 19º ACM Symp. Principios de sistemas operativos*, ACM Press, 2003, págs. 29–43.
3. J. Dean y S. Ghemawat, "MapReduce: procesamiento de datos simplificado en grandes clústeres", *Proc. 6to sim. Diseño e implementación de sistemas operativos*, Usenix As soc., 2004, págs. 137–150.
4. F. Chang et al., "Bigtable: un sistema de almacenamiento distribuido para datos estructurados", *Proc. 7º sim. Diseño e implementación de sistemas operativos*, Usenix Assoc., 2006, págs. 205–218.
5. RL Grossman y Y. Gu, "Minería de datos mediante nubes de alto rendimiento: estudios experimentales mediante sector y esfera", *Proc. 14ª Conferencia Internacional ACM SIGKDD. Descubrimiento de conocimientos y minería de datos*, ACM Press, 2008, págs. 920–927.
6. W. Sobel et al., "Cloudstone: herramientas de medición y evaluación comparativa multiplataforma y multilingüe para la Web 2.0", *Proc. computación en la nube y sus aplicaciones*, 2008; www.cca08.org/papers.php.

Robert L. Grossman es profesor de matemáticas, estadística e informática en la Universidad de Illinois en Chicago y socio gerente de Open Data Group.

Sus intereses de investigación incluyen análisis y minería de datos, computación distribuida y redes de alto rendimiento.

Grossman tiene un doctorado en matemáticas aplicadas de la Universidad de Princeton. Póngase en contacto con él en grossman@uic.edu.

¿ Está interesado en escribir un artículo para la revista **IT Professional** Visite www.computer.org/itpro y haga clic en "escribir para profesionales"