

INTRODUCCIÓN A LA COMPUTACION EN LA NUBE Y BIG DATA (1)

© Ing. Carlos Ormella Meyer

En los últimos años, el interés por la **Computación en la Nube (Cloud Computing)**, tanto para uso personal como para negocios, ha mostrado una tendencia al alza debido, entre otras cosas, a la flexibilidad que proporciona, y a que permite reducir las limitaciones de las arquitecturas informáticas convencionales, beneficiando a una empresa en términos de espacio, tiempo, energía, costo y procesos de negocio.

La computación en la Nube es un entorno para habilitar un acceso ubicuo, conveniente y a demanda de un reservorio compartido (pool) de recursos configurables de computación, tales como redes, servidores, almacenamiento, aplicaciones y servicios.

Estos recursos son administrados por un **Proveedor de Servicios en la Nube (CSP)**, que pone estos servicios a disposición de los clientes y usuarios finales que pueden obtenerlos y liberarlos en forma rápida con esfuerzo mínimo de gestión o interacción con el **CSP**.

Por otra parte, como cualquier sistema de computación, la seguridad de la información debe ser tomada muy en cuenta, ya que la Seguridad y Confidencialidad de los datos son de suma importancia para las organizaciones que buscan utilizar la computación en la Nube.

De hecho, la encriptación de la información manejada se vuelve de mayor necesidad frente a la cada vez mayor “captación” por terceros de información y datos en este escenario, con fines de marketing en muchos casos pero también con otras intenciones.

Modelos de Servicios de Computación en la Nube

La computación en la Nube se caracteriza por una arquitectura constituida por tres modelos de servicios: **Software como Servicio (SaaS)**, **Plataforma como Servicio (PaaS)** e **Infraestructura como Servicio (IaaS)**.

A continuación comentamos las principales características propias de cada modelo de servicio.

1) Software como Servicio (SaaS)

SaaS ofrece capacidad para que el cliente pueda usar las aplicaciones del proveedor que corren en la infraestructura de la Nube.

En esta modalidad, los clientes, en lugar de comprar e instalar las aplicaciones o software en sus computadores, rentan al **CSP** aplicaciones completas que ya están implementadas y gestionadas por el proveedor de la Nube, lo que les permite utilizar la aplicación sin necesidad de mantenerla o administrarla.

En este tipo de servicio el cliente obtiene el derecho de uso a demanda, es decir conforme sus necesidades, de aplicaciones específicas, y la gestión de datos de la aplicación, tales como copias de seguridad y el intercambio de datos entre los clientes.

Las aplicaciones **SaaS** son aplicaciones para múltiples arrendatarios, lo cual significa que son compartidas por múltiples clientes, aunque son lógicamente únicas para cada cliente.

Ejemplos típicos de aplicaciones **SaaS** son las herramientas de procesamiento de Word en línea y los servicios Web de entrega de contenido.

2) Plataformas como Servicio (PaaS)

PaaS proporciona al cliente la capacidad de instalar, en la infraestructura de la Nube, aplicaciones que haya adquirido o incluso propias, y que se hayan creado usando lenguajes, bibliotecas, servicios y herramientas soportadas por el proveedor.

Para todo eso, el proveedor de la Nube ofrece una plataforma para el entorno de desarrollo donde los clientes puedan ejecutar sus propias aplicaciones, sin que tengan necesidad de crear o configurar la infraestructura subyacente.

El servicio incluye la plataforma de gestión de la configuración, la plataforma de desarrollo y herramientas de desarrollo como la **Interfaz de Programación de Aplicaciones (API)**, que incluso se puede configurar en forma remota.

Por lo tanto, los clientes pueden ejecutar sus aplicaciones sin tener conocimientos especializados de administración, así como también crear y desplegar sus aplicaciones Web sin tener que instalar ninguna herramienta en sus computadores.

Todo esto implica que los clientes pueden usar las herramientas del proveedor de la Nube **PaaS** y los recursos de ejecución para desarrollar, probar, implementar y administrar las aplicaciones.

3) Infraestructura como Servicio (IaaS)

IaaS proporciona al cliente la capacidad de usar el procesamiento, almacenamiento, redes y otros recursos fundamentales de computación, por medio de los cuales el cliente puede instalar y correr cualquier tipo de software, tanto sistemas operativos como aplicaciones.

Para ello el servicio **IaaS** ofrece **Máquinas Virtuales** (ver Cuadro 1 – Máquina Virtual, VM), así como otro hardware básico y el sistema operativo en la red.

Con una Nube **IaaS** se obtienen: el acceso a los computadores virtuales, el almacenamiento accesible desde la red, y los componentes de la infraestructura de red, tales como firewalls y servicios de configuración.

Cuadro 1

Máquina Virtual, VM

Una **máquina virtual** es un software que simula a una computadora y puede ejecutar programas como si fuese una computadora real.

Este proceso se llama Virtualización de hardware o simplemente **Virtualización**, y el software correspondiente se denomina **Hipervisor** que asume la funcionalidad de un monitor de Máquina Virtual.

Este proceso da lugar a dos formas de Máquina Virtual, de modo tal que ambas situaciones dan lugar a dos tipos de servicios diferentes.

Una consiste en que el software de virtualización corra sobre el Sistema Operativo del anfitrión soportado por el hardware real, permitiendo que la máquina física subyacente se multiplique en varias Máquinas Virtuales, cada una ejecutando su propio Sistema Operativo soportado por el hardware virtual que produce el Hipervisor.

La otra posibilidad es que el **Hipervisor** apoyado directamente en el hardware real permita trabajar con pilas de diferentes Sistemas Operativos con soporte de un hardware virtual.

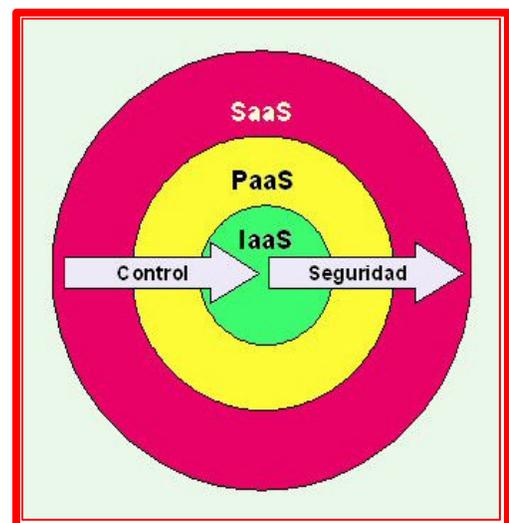


Figura 1

De esta manera entonces, con el arrendamiento de un servicio **IaaS**, los clientes pueden disponer y utilizar la última tecnología de infraestructura y no tienen que preocuparse con la actualización de la tecnología.

Comparaciones básicas de los tres tipos de Servicios

Estos tres modelos de servicios de computación en la Nube pueden ser representados por medio de círculos concéntricos (Ludwig & Coetzee), como se muestra en la **Figura 1**.

En el centro, **IaaS** utiliza sólo un subconjunto de la funcionalidad y la infraestructura de una Nube.

De los tres modelos, **IaaS** es el más comparable a un entorno informático convencional, y le da al cliente de la Nube el mayor control sobre los datos de infraestructura y la Nube.

En el círculo externo, en cambio, **SaaS** abarca la mayor parte de lo que puede ser ofrecido por una Nube. En este caso, el cliente de la Nube tiene menor control sobre la arquitectura de Nube, porque ésta ya está configurada con una mínima participación del usuario final.

Sin embargo, y a estar de dicha arquitectura, esto también significa que **SaaS** ofrece un mayor nivel de seguridad de la información en comparación con **IaaS**, debido a que un proveedor de la Nube confiable puede demostrar que ha implementado y configurado correctamente medidas de seguridad adecuadas para el caso.

Modelos de implementación de una Nube.

Los tres modelos de servicio comentados pueden ser alojados e implementados de diferentes maneras, determinado así la forma de acceso al sistema, lo cual influye en el nivel de control y seguridad correspondientes.

NIST (Instituto Nacional de Estándares y Tecnología) - de USA - define cuatro modelos de implementación que son:

- 1) Nube Privada:** La infraestructura de la computación en la Nube es utilizada exclusivamente por una única organización.
- 2) Nube Pública:** La Nube se pone a disposición de un grupo mucho más amplio de usuarios.
- 3) Nube Comunitaria:** La Nube está disponible sólo para un grupo limitado de organizaciones.
- 4) Nube Híbrida:** La Nube comprende una mezcla de organizaciones privadas, comunitarias y/o Nubes Públicas vinculadas entre sí.

Cualquiera de los tres modelos de servicios vistos antes se puede implementar utilizando uno de los cuatro modelos de implementación.

En este punto conviene destacar la combinación que ha surgido últimamente entre la computación en la Nube y los procesos de **Big Data** o **datos masivos** (ver **Cuadro 2 - Big Data**).

En relación a lo anterior se introduce el concepto de analítica de big data (ver **Cuadro 3 - Analítica de Big Data**).

La **Analítica de Big Data** se define como el proceso de examinar grandes conjuntos de datos constituidos por una variedad de formatos diferentes, para descubrir patrones ocultos,

correlaciones desconocidas, tendencias de mercado, preferencias de los clientes y otras informaciones útiles para los negocios.

Todo esto implica que, en el entorno de big data, el análisis en tiempo real de grandes conjuntos de datos que se reciben a gran velocidad y en cualquier tipo de formato, requiere una plataforma adecuada para almacenar esos datos a través de clusters de servidores distribuidos, así como también un método adecuado para procesar dichos datos de múltiples orígenes.

El análisis de datos en esas condiciones requiere mucho tiempo y por lo tanto mayores costos para albergarlos en una base de datos relacional tradicional, amén de las restricciones generadas por los datos no estructurados.

En cambio, las exigencias comentadas antes se satisfacen sin mayores complicaciones en la computación en la Nube.

Pasa que los data warehouses tradicionales conformados por bases de datos relacionales no pueden manejar bien datos semiestructurados y no estructurados, especialmente en tiempo real con actualizaciones frecuentes e incluso constantemente.

Cuadro 2

BIG DATA

Big Data o **Datos Masivos** se refiere a la gestión y análisis de enormes volúmenes de **datos**, producidos tanto por dispositivos como por personas (el caso de las redes sociales), su transformación en **información** útil para los negocios que produzca a su vez el **conocimiento** necesario para optimizar la toma de decisiones.

Significativamente este proceso constituye la base del **Business Intelligence, BI**, es decir, la cualidad de transformar los **datos** en información, y la información en conocimiento, de forma tal que se pueda optimizar el proceso de toma de decisiones en los negocios.

La tecnología **Big Data** se puede definir mediante cinco indicadores llamados **5v**, aunque los tres primeros son los más mencionados.

- 1) **Volumen**, grandes cantidades de datos, donde se manejan totales de datos medidos en **Terabytes** (10^{12}), o en **Petabytes** (10^{15}), y hasta en **Exabytes** (10^{18}).
- 2) **Variedad**, datos de tres tipos diferentes:
 - a) **Estructurados**: Bases de datos relacionales tradicionales.
 - b) **Semiestructurados**: Registros internos, del servidor Web, seguimiento de clics en Internet, etc.
 - c) **No e estructurados** Datos que no están almacenados en una base de datos tradicional, porque de hecho no se los puede almacenar adecuadamente en estructuras de datos relacionales.
Aquí se incluyen imágenes, video, audio, etc., así como otros de tipo texto, por ejemplo, los generados en las redes sociales, foros, e-mails, archivos de Powerpoint y Word, etc..
- 3) **Velocidad**, con que se reciben los datos
- 4) **Veracidad**, como resultado de un análisis que la establece, así como la validación de su utilidad para mejorar la toma de decisiones.
- 5) **Valor**, que deben tener los datos, de forma tal que la información realmente tenga valor para los negocios.

En este caso, se necesita recurrir a programas especiales como Hadoop para almacenar grandes conjuntos de datos a través de clusters de servidores distribuidos, y MapReduce, para coordinar, combinar y procesar datos de múltiples orígenes.

Y aquí es donde la **Análítica de Big Data** se asocia generalmente con la **Computación en la Nube**.

Efectivamente, los servicios de computación en la Nube pueden proporcionar todos esos recursos y facilidades, ofreciendo a los procesos de **big data** una infraestructura eficiente y efectiva en costo para soportar las tecnologías correspondientes,

De esta manera, las empresas que cuentan con una Nube Privada pueden agregar la analítica de **big data**, o usar una Nube Pública, o bien usar una Nube Híbrida que proteja los datos sensibles en la Nube Privada, y que también aproveche los recursos y aplicaciones externas propias de las Nubes Públicas.

Cuadro 3

ANALITICA DE BIG DATA

Se puede definir la **Análítica de Datos** o **Data Analytics** como la ciencia de examinar datos en bruto con el fin de lograr conclusiones acerca de la información.

La **Análítica de Datos** se distingue del Data Mining por el alcance y foco del análisis.

El Data Mining se basa en revisar en un *data warehouse* grandes conjuntos de datos usando un software especial para identificar y descubrir patrones y establecer interrelaciones ocultas.

El Data Warehouse, por su parte, es un almacén de datos estructurados históricos, que incluyen la copia de las transacciones para consulta y análisis.

El **Data Analytics**, a diferencia del Data Mining se enfoca en la **inferencia**, o sea el proceso de deducir una conclusión basada exclusivamente en conocimiento ya existente por parte del usuario.

El **Data Analytics** se usa para describir el **Proceso Analítico en Línea (OLAP)** y el análisis de la **Gestión de las Relaciones con los Clientes (CRM)** en los call centers.

Su uso se viene difundiendo especialmente en bancos y compañías de tarjetas de crédito para prevenir el fraude o identificar robos.

Finalmente, el **Big Data Analytics** como extensión del **Data Analytics** es el proceso de examinar grandes *volúmenes* de datos que contienen una *variedad* de tipos de datos, que se generan a alta *velocidad* (big data), para poner en claro patrones ocultos, correlaciones desconocidas, tendencias de mercado, y preferencias de clientes, así como otros tipos de información útiles para los negocios.

(1) Extraído del primer Módulo de Estudio del Curso a Distancia “**Seguridad en el Ciberespacio**” del autor.