

Mejores prácticas para migrar de On-Premise (local) a Cloud

¿Cuáles son las Mejores prácticas para migrar de On-Premise (local) a [Cloud](#) ? Antes de explicar las mejores prácticas para migrar de on-premise a la nube, comencemos con las razones principales por las que debería hacer este traspaso: rentabilidad , rendimiento y seguridad.

Eficiencia de costo

En última instancia, el motor subyacente para pasar del almacén de datos empresarial heredado en las instalaciones a la nube es la rentabilidad. El costo abarca no solo el costo directo de poseer o licenciar el equipo del centro de datos, sino también el costo del servicio, mantenimiento y administración de terceros, capacitación y políticas. Estos costos indirectos pueden representar una proporción significativa de los gastos de administrar centros de datos en las instalaciones.

Veamos más de cerca el costo de mantenimiento, gestión y capacitación.

Mantener la infraestructura informática local requiere: espacio, construcción, electricidad, aire acondicionado, adquisición de hardware, sistemas operativos, redes, seguridad física y más. En cambio, puede transferir estos costos a un proveedor de servicios en la nube para ahorrar dinero drásticamente.

Administrar un entorno local también requiere una gran experiencia y experticia. Con la tendencia imparable hacia la nube, el costo de encontrar y entrenar talento de ingeniería se volverá más difícil y más costoso. Desde la contratación de especialistas, hasta la gestión de proyectos de cualquier

instalación de software nueva, hasta el monitoreo y la respuesta a incidentes. Es una molestia tremenda que le cuesta a su organización mucho más que los recursos informáticos reales. Por el contrario, es el negocio de los proveedores de servicios en la nube ejecutar y administrar centros de datos en la nube en un mercado de crecimiento agresivo, lo que le ahorra este dolor de cabeza.

Actuación

La desventaja de invertir todo en ser dueño de su propio entorno local es la inflexibilidad. Es un compromiso a largo plazo, que impide que su organización obtenga la última y mejor tecnología. Financieramente, los costos iniciales combinados de gasto de capital de compra del equipo y los costos de depreciación a medida que envejece, también se acumulan muy rápidamente.

Por otro lado, la infraestructura de servicio en la nube le brinda flexibilidad y el mejor equipo de la industria: los procesadores Intel más nuevos, la memoria RAM más optimizada, las redes petabit, el almacenamiento SSD, las matrices de GPU e incluso el hardware programable en la nube. Esto viene con un amplio ecosistema de plataformas de software: bases de datos y almacenes administrados, plataformas de orquestación, almacenamiento en caché, sistemas de colas y casi cualquier cosa que necesite para cualquier arquitectura de software. El impacto en su negocio es prácticamente inmediato, especialmente si opera en un mercado competitivo donde los líderes se mueven rápidamente.

Seguridad



Se requiere que los proveedores de servicios en la nube cumplan con los más altos estándares de seguridad en la industria, establecidos por instituciones de salud, financieras y gubernamentales.



ienen una infraestructura de datos en las instalaciones están preocupados por la seguridad. La realidad es que es difícil estar más seguro y en control que alojar su centro de datos dentro de su propio perímetro. Finalmente, la seguridad en las instalaciones es tan confiable como las políticas de seguridad implementadas y aplicadas. Esta sobrecarga de la política no es insignificante y no se puede ignorar.

Los proveedores de servicios en la nube, como AWS, Google Cloud Platform y Snowflake, entienden perfectamente que miles de clientes les confían la seguridad de sus datos. En consecuencia, se requiere que los proveedores de servicios en la nube cumplan con los más altos estándares de seguridad en la industria, establecidos por instituciones de salud, financieras y gubernamentales. Esto permite que las certificaciones como SOC2, ISO27001, HIPAA y PCI sean mucho más fáciles de obtener cuando se ejecuta en una infraestructura informática que ya está certificada. La autenticación, autorización, registro y auditoría están integrados en todas las plataformas en la nube.

Las medidas de seguridad como los cortafuegos de red y

aplicación, la protección DDOS y la gestión de identidad se estandarizan, prueban y se ponen a disposición para su instalación y configuración. Aunque los entornos de nube tienen una capa de computación completamente segura y la capa de computación virtual (o lógica) tiene todos los mecanismos de seguridad de nivel empresarial, aún depende de usted implementar políticas de seguridad adecuadas para su arquitectura, como lo haría para una configuración local . Dado que no tiene el 100% de control de su entorno de nube, presenta más riesgos que en las instalaciones.

Un plan de migración propuesto

Un error común sobre la migración a la nube es que será un viaje de una sola vez. Pero la realidad es que el proceso de migración de la infraestructura de datos a la nube debería ocurrir gradualmente. Una migración exitosa debe sentirse lo más fluida posible para la organización, de modo que el trabajo no se vea interrumpido. El viaje se parece un poco a la historia real sobre el Citigroup Center de 59 pisos en Nueva York , que se sometió a refuerzos de infraestructura críticos realizados durante la noche durante varios meses, para que el público y los ocupantes del edificio no se dieran cuenta. Del mismo modo, al planificar una importante migración de datos. El objetivo es garantizar que todo se haga paso a paso, al tiempo que se minimiza el tiempo de inactividad y las interrupciones para los usuarios. Es por eso que debe estar preparado para un proceso de migración que puede llevar desde varios meses hasta un año.

Paso 1 – Migrar datos existentes

El primer paso es crear una copia inicial de sus datos existentes en un almacén de datos en la nube. Este proceso requerirá elegir el almacén de datos en la nube adecuado para su organización y luego hacer una copia inicial de todos sus

datos.

Hay dos desafíos principales en este paso. El primero es experimentar y elegir la infraestructura adecuada para su organización. Para hacer esto, puede seleccionar un conjunto de datos más pequeño y migrarlo a varios almacenes de datos diferentes para compararlos. El segundo desafío es copiar todos sus datos existentes. Esto podría ser cientos de terabytes de datos, una cantidad que no se transfiere fácilmente a través de Internet a la nube. Google y Amazon proporcionan varios medios para superar este desafío, transportando físicamente sus discos duros, camiones y USPS.

Tenga en cuenta que copiar los datos sin procesar es solo una parte de esta migración inicial. Debe verificar el formato y el esquema de los datos que exporta desde su almacén, y luego importar el esquema de sus datos al almacén de datos en la nube antes de la carga real. Por último, para continuar con el paso 2, debe marcar el punto en el tiempo de la instantánea exportada y utilizarlo al configurar un mecanismo de replicación en curso.

Paso 2: configurar la replicación en curso

D
e
s
p
u
é
s
d
e
e
x
p
o
r
t



Cualquier solución de sincronización debe ser comparada para latencia y confiabilidad, ya que estos parámetros son cruciales para el éxito de la migración de la organización a la nube.



ar una primera instantánea de su almacén de datos local y copiarla en su almacén de datos en la nube, el siguiente paso sería configurar un proceso de sincronización continuo. La replicación continua es más complicada que una operación de copia única, ya que en realidad es una serie de operaciones de copia incrementales.

Cada operación requiere capturar los cambios en los datos y su esquema, y aplicar esos cambios al almacén de datos en la nube. Algunos cambios, como los datos eliminados o los tipos de columnas alterados, pueden requerir soluciones personalizadas para aplicarse en el almacén de datos en la nube. Más desafíos técnicos se describirán más adelante en esta publicación.

Cualquier solución de sincronización debe ser comparada para latencia y confiabilidad, ya que estos parámetros son cruciales para el éxito de la migración de la organización a la nube. Puede construir esta sincronización usted mismo o usar un servicio de canalización de datos para manejar la replicación continua de datos y esquemas. Una vez que este nivel básico está asegurado, puede migrar el resto de su

infraestructura un componente tras otro.

Paso 3 – Migrar BI

Según nuestra experiencia, la primera funcionalidad central que una organización estará dispuesta a mover a la nube es su infraestructura de análisis. Dado que es un componente importante de la infraestructura de datos de una organización, pero no el núcleo absoluto, es un objetivo de migración ideal y menos riesgoso. Además, hay varios equipos de análisis de datos en cada organización, y será más fácil comenzar con un grupo pequeño e innovador para liderar el cambio. Encuentre este grupo y trabaje con ellos para configurar una herramienta de [BI](#) que funcione con su almacén de datos en la nube.

Algunos ejemplos de herramientas de BI son Tableau , Looker , Periscope , Chartio e incluso el código abierto Redash . Con un equipo de analistas líderes y un almacén de datos en la nube totalmente sincronizado, puede trabajar en portar informes y paneles que demostrarán el valor del almacén de datos en la nube. Si se hace correctamente, esto creará un efecto dominó en su organización, lo que hará que más consumidores de datos deseen pasar a la nube.

Paso 4: migre sus aplicaciones de datos heredados

En este punto de la migración, su organización debería estar alborotada con un nuevo almacén de datos en la nube que ejecuta 5 veces más informes, cada uno al menos 10 veces más rápido que el almacén de datos local. Con tales vientos soplando en sus velas, puede completar la migración de todos los equipos de BI para usar las herramientas de BI en la nube y luego priorizar la migración de sus herramientas de informes personalizadas y aplicaciones de datos.

Esto traerá más desafíos técnicos, ya que puede descubrir que

los controladores ODBC deben reemplazarse y las consultas deben ajustarse o reescribirse. También pueden ser necesarios cambios en el modelo de datos para aprovechar al máximo las ventajas de rendimiento de su almacén de datos en la nube. Confiar en un mecanismo de sincronización de datos que pueda adaptarse a las transformaciones hará que la implementación de estos cambios en el modelo de datos sea más dinámica y flexible.

Es mejor trabajar con los diversos equipos de ingeniería de datos en su organización para instruirlos sobre el uso del nuevo almacén de datos en la nube y las herramientas de migración que utilizó. Comparta la mayor parte de la experiencia de migración para potenciar más agentes de cambio en la organización. Implemente las aplicaciones recién migradas progresivamente y cree una red de seguridad de reversión para minimizar la interrupción de la eficiencia de su organización.

Paso 5: migre sus procesos ETL heredados

Una vez que todas las aplicaciones de datos heredadas se han migrado a su nube, el último paso es apuntar sus procesos ETL a su almacén de datos en la nube. En algunos casos será solo un cambio de configuración, y en otros puede requerir una reescritura completa. Afortunadamente, en el espacio de ETL en la nube, hay varios servicios integrales que pueden ahorrarle gran parte del trabajo. Completar este paso esencialmente hará que su almacén de datos local no esté sincronizado y, por lo tanto, quede obsoleto.

¡Felicidades!

Los desafíos que nos esperan

A pesar de dividir el plan de migración en lo que parecen ser cinco pasos simples, la verdad es que el viaje no ha terminado. Hay muchos desafíos técnicos que tienes que navegar. Exploraremos algunos de ellos aquí:

Desafío 1: volver a planificar su modelo de datos

Los almacenes de datos en la nube admiten un tipo diferente de esquemas, lo que tiene sus ventajas y desventajas. Los tipos de datos pueden ser diferentes. AWS Redshift es compatible con PostgreSQL y, por lo tanto, tiene el conjunto de tipos de datos más familiar. BigQuery, por otro lado, usa STRING en lugar de VARCHAR, y emplea los tipos RECORD (objetos semiestructurados) y REPEATED (arrays). Snowflake presenta VARIANT, OBJECT y ARRAY, para admitir datos semiestructurados.



T
a
m
b
i
é
n
h
a
y
u
n
a
m
p

lio conjunto de tipos de datos eclécticos que no son compatibles (BLOB, coordenadas geográficas, etc.). Además, los almacenes de datos en la nube fomentan un enfoque de

desnormalización de esquemas para un mayor rendimiento. El mayor almacenamiento requerido para contener datos desnormalizados es relativamente barato. En contraste, ejecutar JOINS en tablas almacenadas en servidores distribuidos es muy costoso y no conducirá a las mejoras de rendimiento deseadas. Para aumentar la complejidad de estas diferencias, no es suficiente ajustar el modelo de datos una vez. Es necesario mantener los dos modelos de datos en sincronización continua a medida que cambian con el tiempo. El modelo de datos seguramente cambiará durante el período de la migración.

Desafío 2 – Seguridad

Anteriormente describimos las ventajas de confiar en la seguridad de nivel gubernamental de los grandes proveedores de la nube. Pero este tipo de seguridad se aplica principalmente a los niveles de hardware e [IaaS \(infraestructura como servicio\)](#). Contrariamente a este nivel de seguridad, la facilidad de la nube y la presión para hacer una migración rápida y confiable también traerá algo de promiscuidad. Es muy fácil simplemente proporcionar a todos los desarrolladores y consumidores de datos permiso para cualquier recurso en la nube, posponiendo la tediosa gestión de permisos para «hacernos un futuro». Cuanto más se posponga, más difícil será imponer permisos de seguridad más estrictos.

Nuestro consejo es tener esto en cuenta justo después de que la migración gane crédito e ímpetu de la organización. Este es el punto donde los permisos pueden salirse de control, pero también es una oportunidad para mapear todos los grupos y roles que requerirán acceso a recursos en la nube y planificar políticas de seguridad en consecuencia. Estudie detenidamente las herramientas de su proveedor de la nube para administrar roles y permisos, ya que son muy completas y le permitirán definir casi cualquier política que desee. Si se configura y audita correctamente, es otra gran victoria en el largo viaje

de migración.

Desafío 3: conectar aplicaciones de datos personalizadas a su almacén de datos

Otro obstáculo es ajustar las interfaces que utilizan sus aplicaciones de datos personalizadas para conectarse a su almacén de datos. En otras palabras, aunque los controladores ODBC / JDBC son compatibles y se mantienen activamente, apenas se comportan de manera idéntica entre sí. Cambiar el controlador de la base de datos de una aplicación puede requerir varios ajustes de consulta. Algunos cambios serán obvios en el primer uso, ya que las declaraciones SQL pueden provocar errores visibles. Otros cambios son menos obvios, ya que los diferentes controladores ODBC pueden hacer conversiones de datos leves. Ejemplos de tales conversiones pueden ser la precisión de los números de coma flotante, el formato de zona horaria de las marcas de tiempo y la forma en que se manejan los valores NULL.

Estos cambios solo aparecerán como discrepancias en los datos y requerirán pruebas más rigurosas para detectarlos. La ventaja de enfrentar este tipo de obstáculos es la mayor cantidad de ingenieros involucrados. Cada aplicación generalmente tiene un mantenedor, por lo que una fuerza de trabajo más grande puede superar los obstáculos. Mientras se comparta el conocimiento de estos ajustes, toda la organización puede superarlos más rápido.

Desafío 4: escribir y usar procedimientos almacenados

Por último, una característica de almacenamiento de datos local muy común que a menudo se pasa por alto y que sus competidores en la nube no tienen en cuenta es la capacidad de escribir y usar procedimientos almacenados . Los principales almacenes de datos en la nube, Snowflake , Redshift y BigQuery

, admiten funciones definidas por el usuario (definidas en Python, SQL o JavaScript), pero para muchos esto no es suficiente. La capa de procedimientos almacenados en un almacén de datos es como un depósito de aplicaciones de datos en miniatura, que ahorran mucho trabajo y ayudan a preservar el conocimiento específico de la organización. La alternativa común es utilizar una plataforma separada para programar consultas parametrizadas u orquestación de tareas. Hay opciones de código abierto como Luigi y Airflow. así como alternativas comerciales basadas en la nube.

Leer también: [Colocation vs On premise, ¿cuál es el adecuado para Ud?](#); [Tipos de métodos de integración de sistemas, TI](#); [IaaS versus On premise, comparación de costos del servidor en las instalaciones](#) ; [Servicios en la nube](#)