

El Futuro del Data Management

Por Fiona McNeil y Ron Agresta



Introducción

La mayoría de las organizaciones coinciden en que tener los datos correctos, y gestionarlos bien, es absolutamente crítico para el éxito de las iniciativas que impulsan la innovación y el aumento de los ingresos. Predecir, y aprovechar, la próxima gran tendencia es solamente una suposición educada sin una base sólida que contenga datos confiables finamente pulidos que sirvan como los bloques de construcción.

Si bien existe consenso respecto a la importancia que tiene la gestión de datos, no hay un acuerdo en lo que implica. En términos generales, la gestión de datos abarca las áreas de descubrimiento, acceso, calidad, integración, federación, gobierno y dominio. Cada una de estas áreas requiere de especialistas que entienden los datos que tienen a la mano, sus propiedades inherentes y el correspondiente contexto de negocio que los datos soportan.

Sin embargo, es evidente que la disponibilidad de expertos en la gestión de datos no sigue el ritmo de las necesidades organizacionales de contar con datos precisos para seguir siendo competitivas. El surgimiento de las tecnologías de big data ha hecho que se adquieran y almacenen cantidades cada vez mayores de datos, algo común - datos que siguen necesitando organizarse, limpiarse y que se extraiga de ellos las gemas ocultas del conocimiento. Esto ha creado la necesidad de nuevas metodologías de gestión de datos que no dependan de tecnólogos tan especializados, sino de nuevos enfoques que escalarán la experiencia actual al tamaño de los datos - dondequiera que éstos residan.

Entonces, ¿cómo está evolucionando la industria de la gestión de datos para seguir apoyando las necesidades actuales, y dentro de las restricciones actuales, al tiempo de construir entornos para satisfacer las necesidades futuras?

Estrategias Actuales

En un esfuerzo por responder mejor al mismo tiempo que nadar en los datos albergados en los lagos de Hadoop y en otras plataformas de big data, el procesamiento más generalizado ha reducido con éxito el tiempo necesario para ejecutar las tareas de gestión de datos bien establecidas. Basada en los entornos existentes, y ofreciendo una mayor fluidez que el procesamiento histórico, la ejecución de las rutinas ahora se está realizando más cerca a la fuente de los datos. Este ha sido el beneficio de aprovechar el hardware existente, reduciendo el movimiento de los datos y simplificando el acceso, acelerando el procesamiento y el tiempo de la CPU.

Iniciado por los avances en arquitectura, el procesamiento grid distribuye las cargas de trabajo en el hardware conectado a la red, aprovechando la ventaja de los procesadores colectivos que ejecutan el código en compartimientos, y que suman los resultados finales. La ejecución en la memoria eleva los datos dentro de la memoria para el procesamiento (el cual también puede distribuirse en las máquinas habilitadas con grid), llegando a los datos almacenados en disco sólo cuando es necesario. Más recientemente, el procesamiento en la base de datos ha ganado popularidad, ya que lleva la ejecución del código a los repositorios existentes, ejecutándose como un proceso integrado. En muchas organizaciones, tiene sentido hacer todo lo anterior - para recibir los beneficios de desarrollar rutinas de gestión de datos una vez y simplemente desplegadas de diferentes formas con poco o ningún retrabajo adicional.

Los flujos de datos - con alto volumen y alto rendimiento - también van en aumento, lo que aumenta el tamaño de las colecciones de centros de datos actuales. Una nueva fuente de datos detallados con marcas de tiempo para humanos y máquinas por igual, se espera que desborde incluso los lagos de datos más grandes. Y aunque la infraestructura de nube ofrece un procesamiento elástico

y de alta disponibilidad, aún hay un costo gradual que en algún punto necesitará reducirse. El procesamiento de estos datos al integrar la gestión de datos en flujo, y las funciones analíticas en sitio, las agregaciones de nube, o incluso a la orilla de la red, enfoca los recursos de almacenamiento en sólo los datos que vale la pena conservar en primer lugar. Así, las organizaciones están modernizando gradualmente los sistemas existentes para satisfacer las nuevas necesidades en lugar de invertir más en la tecnología legada.

Los Cambios Necesarios a las Estrategias Actuales

Si bien las nuevas estrategias de gestión de datos incluyen mover las transformaciones a los datos y no al contrario - y darle a los procesos de gestión de datos sobrecargados espacio para respirar -, incluso esto tiene sus límites. Las proyecciones de crecimiento de los datos muestran que estamos en el trayecto hacia la gestión de big data completa.

Tampoco se trata únicamente de tecnología. Un número cada vez mayor de usuarios de negocio expertos en tecnología están trabajando. Esta gente insiste en trabajar directamente con los datos que alimentan los reportes de marketing, las proyecciones de ventas y las predicciones de ingresos de los productos. A menudo se dejan enredar con soluciones de "freeware" apenas integradas que pueden ofrecer una gratificación instantánea, pero que después se agregan al caos de los usuarios de datos de diferentes tipos que trabajan con los mismos datos. Los especialistas en gestión de datos ahora tienen a su cargo el control de un número explosivo de transmisiones de datos, así como soportar un mayor número de consumidores de datos de todos los niveles de habilidades.

Las tecnologías de big data, la analítica avanzada y los algoritmos de código abierto han dado a los profesionales de la gestión de datos la esperanza de contener la marea de datos. Las modernas tecnologías altamente interactivas basadas en la web han facilitado llevar la preparación de datos de autoservicio a las masas. Encontrar la fuente de datos correcta y darle forma para satisfacer las necesidades de los proyectos deben ser un ejercicio de fácil uso. Con el fin de ayudarles a afinar y entender los datos, es pertinente que a los usuarios se les presenten las sugerencias pertinentes. Los datos deben combinarse fácilmente para liberar nuevos conocimientos. Las alertas son necesarias cuando se hacen cambios que afecten el trabajo de los usuarios o resulten en un elemento de datos crítico que haga falta

en los resultados. Sólo algunos usuarios pueden ahora lograr lo que en el pasado podría haber necesitado una plantilla completa de profesionales de la integración de datos.

Los procesos de descubrimiento de datos, de análisis y de transformación complejos y sofisticados, y los servicios de micro analítica que tienen que ocultarse del usuario promedio sustentan esta experiencia de usuario simplificada y ágil. Aún están disponibles para los usuarios técnicos todas las poderosas herramientas necesarias para los proyectos de gestión de datos empresariales. Pero sus colegas en la parte del negocio necesitan tener los medios para ser autosuficientes, dejando más tiempo para que el personal de gestión de datos se enfoque en las amplias actividades de automatización, afinación, auditoría y gobierno estratégico.

Lo que Depara el Futuro

En el futuro está casi garantizado que la gestión de datos avance para no sólo satisfacer todos los requerimientos de autoservicio de los usuarios tanto sofisticados como novatos, pero también para automatizar aún más la disciplina, con la interacción humana enfocada en la innovación y aplicación iniciales.

Los sistemas autoajustables guiados por algoritmos de optimización podrán ajustar automáticamente el procesamiento para maximizar la infraestructura de cómputo empresarial para que operen de conformidad con los acuerdos de nivel de servicio individualizados. Los datos serán más que un activo; serán la constitución celular de las organizaciones - con conocimiento autónomo llevando las moléculas que albergan las acciones necesarias para transformar, ajustar, ejecutar y aplicar instrucciones prescriptivas definidas por humanos. Los sistemas de gestión de datos se auto organizarán, buscarán y descubrirán nuevos conocimientos aprendidos por las máquinas, proponiendo nuevos elementos y combinaciones de datos de fuentes internas y externas, anticipándose así a las necesidades organizacionales. Las estructuras de datos serán dinámicas y contextualmente relevantes, y los metadatos - no sólo el procesamiento - serán elásticos. Ya vislumbramos los inicios de estas tendencias de las tecnologías innovadoras de hoy.

La gestión de datos del futuro se orientará al contexto de su uso, predecirá las necesidades futuras y se automatizará - liberando a los usuarios para que se dediquen a supervisar.