

Servicios centrales de datos: componentes básicos para generar valor comercial

WHITE PAPER:

CALIDAD DE DATOS



Servicios centrales de datos: componentes básicos para establecer valor en los negocios

2

INTRODUCCIÓN

EL USO EFICAZ DE INFORMACIÓN DE NEGOCIOS VA MÁS ALLÁ DE LAS DEMANDAS DE VALIDACIÓN ESTRUCTURAL DE DATOS O DE DOMINIOS DE VALORES PARA LAS PRUEBAS DE MEMBRESÍA. A MEDIDA QUE MÁS DECISIONES SON DISTRIBUIDAS A TRAVÉS DE LA JERARQUÍA ORGANIZACIONAL, Y QUE SE PROCESA Y ENVÍA UNA MAYOR CANTIDAD DE INFORMACIÓN A LOS USUARIOS FINALES, SURGE LA NECESIDAD DE REPRESENTAR DE MANERA VISUAL MUCHOS ASPECTOS DE LAS TRANSACCIONES DE NEGOCIOS COMO TIPO, MOMENTO, DURACIÓN, Y DE MANERA CRÍTICA, LA UBICACIÓN DE DICHAS TRANSACCIONES. AUMENTAR EL VALOR Y LA UTILIDAD DE LA INFORMACIÓN DE NEGOCIOS, TANTO PARA LAS APLICACIONES OPERATIVAS COMO ANALÍTICAS, REQUIERE UNA COMBINACIÓN DE PRÁCTICAS SÓLIDAS DE ADMINISTRACIÓN DE DATOS, TALES COMO LA GOBERNANZA DE DATOS, EL ASEGURAMIENTO DE LA CALIDAD DE LOS DATOS, LA MEJORA DE DATOS Y, CADA VEZ MÁS, CAPACIDADES DE LOCALIZACIÓN INTELIGENTE TALES COMO GEOCODIFICACIÓN, MAPEADO Y ENRUTAMIENTO.

ESTANDARIZAR LAS PRÁCTICAS SÓLIDAS DE ADMINISTRACIÓN DE DATOS AL DESARROLLAR O REDISEÑAR LA ARQUITECTURA DE LA EMPRESA DISMINUYE EL RIESGO Y AUMENTA LA CONSISTENCIA; LA CONFIANZA ORGANIZACIONAL EN LOS DATOS SIMPLIFICA EL DESARROLLO DE APLICACIONES Y AUMENTA LA FIABILIDAD DE LOS DATOS. SI BIEN SE HA ESPECULADO MUCHO EN LO QUE SE REFIERE A CONCEPTOS TALES COMO LA INTEGRACIÓN DE DATOS MAESTROS COMO UNA OPORTUNIDAD PARA MEJORAR LA CALIDAD DE LOS DATOS, ESTE DOCUMENTO EXPLORA CÓMO UN CONJUNTO CONSOLIDADO DE SERVICIOS DE VALOR AGREGADO PARA DATOS QUE CUBRE EL ESPECTRO DE LAS MEJORES PRÁCTICAS DE ADMINISTRACIÓN DE DATOS PUEDE REDUCIR LOS COSTOS OPERATIVOS Y AUMENTAR LA EFICIENCIA, AL TIEMPO QUE MEJORA LA CALIDAD Y LA UTILIDAD DE LOS DATOS DE LA EMPRESA. UNA CAPA ÚNICA DE SERVICIOS DE DATOS REDUCIRÁ LA COMPLEJIDAD DE SU MARCO DE ADMINISTRACIÓN; ESTANDARIZAR LA FORMA EN QUE SE USAN LAS HERRAMIENTAS DE DATOS Y LAS MANERAS EN QUE SE APLICAN LAS REGLAS GARANTIZARÁ CONFIABILIDAD Y VALORES PREDECIBLES PARA LA INFORMACIÓN.

A MEDIDA QUE MÁS ORGANIZACIONES REALIZAN LA TRANSICIÓN HACIA UNA ARQUITECTURA ORIENTADA A LOS SERVICIOS, LA PERFECTA INTEGRACIÓN DE LOS DATOS SE CONVIERTE EN UN COMPONENTE FUNDAMENTAL DE LAS APLICACIONES EMPRESARIALES

Impulsores del valor comercial y confiabilidad de la información

Más allá del consenso de la alta gerencia sobre la expectativa general por la información de alto valor, existen determinados eventos que pueden dictar una necesidad inmediata de mejoras en el valor y la utilidad de los datos. Fusiones y adquisiciones corporativas, cambios en los requisitos para el cumplimiento de regulaciones, disminución de la cantidad de clientes, aumentos perceptibles en la actividad de los centros de llamadas, migraciones de sistemas, mejoras en los procesos de negocios de atención al cliente, introducción de sistemas de administración de las relaciones con los clientes (CRM) o de planificación de recursos empresariales (ERP) y gestión de datos maestros (MDM) son todos ejemplos de situaciones de negocios que dependen de ajustes organizacionales en las prácticas de administración de datos para su éxito.

Cada una de esas situaciones demuestra una necesidad de instituir procesos de negocios sólidos y prácticas de administración de la información correspondientes para mejorar la experiencia del cliente y mantener, al mismo tiempo, economías de escala. En vez de utilizar la tecnología al azar, en múltiples oportunidades, para un supuesto problema que se espera poder resolver de alguna manera, un enfoque más racional busca el retorno de la inversión tecnológica que se obtiene cuando las prácticas necesarias para la administración de la información reciben el soporte de servicios centrales de datos que se implementan una vez pero pueden ser utilizados en múltiples oportunidades de manera consistente.

Este enfoque combina prácticas fundamentales de administración de la información con servicios que mejoran el valor y la utilidad de la información de negocios, tales como la aseguración de la calidad de los datos, la mejora de datos y la localización inteligente. Activar esos servicios centrales de datos requiere:

- Reforzar los procesos de producción de información que dan soporte a múltiples consumidores de datos,
- Estandarizar las herramientas y técnicas en toda la empresa,
- Estandarizar la definición y el uso posterior de las reglas de negocios, y
- Estandarizar los servicios de análisis espacial/inteligencia de localización asociados con operaciones y transacciones.

Visualizar y estandarizar el catálogo de servicios de valor agregado para datos eliminará la situación en la que los mismos productos, o productos competidores, se compran en múltiples ocasiones, o cuando distintos grupos implementan las mismas técnicas sin conocer los planes del resto. Sin embargo, este enfoque en última instancia fortalecerá y reutilizará la inversión en herramientas de administración de datos para obtener múltiples retornos de la inversión.

Servicios centrales de datos

Si bien los criterios que caracterizan las capacidades como servicios “centrales” no están completamente definidos, pueden resumirse en términos del apoyo que brindan a las mejores prácticas de administración de la información en la empresa. Por ejemplo, uno puede esperar que cualquier servicio que mejore la interoperabilidad de los procesos de negocios, reduzca la complejidad del desarrollo de aplicaciones, estandarice técnicas, unifique la definición y aplicación de reglas de negocios, o mejore la calidad y utilidad de la información de los clientes pueda considerarse como “servicio central”.

Pero para garantizar un cumplimiento predecible de las buenas prácticas de administración de información, una lista razonable debe incluir al menos estos servicios de integración de datos, calidad de datos y localización inteligente:

- Acceso e integración de datos
- Creación de perfiles de datos
- Análisis sintáctico
- Estandarización
- Resolución de identidades
- Vinculación y combinación de registros
- Limpieza de datos
- Mejora de datos
- Geocodificación
- Mapeado y análisis espacial
- Enrutamiento
- Revisión y control de datos

Servicios centrales de datos: componentes básicos para establecer valor en los negocios

Acceso e integración de datos

Las técnicas utilizadas para el desarrollo del almacenamiento de datos son hoy omnipresentes, particularmente en la medida en que ha crecido la necesidad de interoperabilidad entre sistemas. A medida que más personas desean rediseñar sus entornos para permitir que se compartan sus datos, o incluso la consolidación de conjuntos de datos almacenados en forma aislada, existe una creciente necesidad de servicios subyacentes de integración de datos para proporcionar la capacidad de conexión que permita el acceso con una integración perfecta desde distintas fuentes y su envío a múltiples destinatarios. La capacidad de los servidores remotos para acceder a los datos en sus fuentes originales permite la implementación y el uso eficiente de las mejores prácticas para su administración, que pueden ponerse en práctica mediante los servicios centrales de datos que se describen en este trabajo. Los servicios de integración dependen de la habilidad para conectarse a infraestructuras y entornos de datos de uso habitual, sumados a los servicios de análisis sintáctico y estandarización para transformar los datos extraídos a formatos adecuados para su envío.

En la empresa moderna, el uso compartido, la federación y la sincronización de datos están convirtiéndose rápidamente en la norma. A medida que más organizaciones realizan la transición hacia una arquitectura orientada a los servicios, la perfecta integración de los datos se convierte en un componente fundamental de las aplicaciones empresariales.

Creación de perfiles de datos

Una evaluación objetiva de la calidad de los datos es un proceso de análisis y descubrimiento, basado a su vez en una revisión objetiva de los valores de los datos que componen el conjunto en estudio. Un análisis estadístico cuantitativo ayudará a los analistas de datos a detectar datos con potenciales fallas, como valores atípicos, valores faltantes, elementos de datos que no cumplen con los patrones definidos o los dominios de valores, o que son inconsistentes respecto de otros elementos de datos. La creación de perfiles de datos es un conjunto de algoritmos para el análisis estadístico y la evaluación de la calidad de los valores en un conjunto de datos, así como para explorar las relaciones que existen entre

los elementos de datos o sus conjuntos. La creación de perfiles de datos permite identificar fallas en los datos, y brinda una forma de comunicar esas instancias a los expertos en cada tema, cuyo conocimiento del negocio puede confirmar la existencia de problemas en los datos. La herramienta de creación de perfiles de datos revisará todos los valores en una columna y proporcionará una distribución de frecuencias de los valores de esa columna, sugiriendo el tipo y los usos potenciales de cada columna. El análisis entre columnas puede exponer dependencias incorporadas en los valores, mientras que los análisis entre tablas exploran conjuntos de valores superpuestos para identificar relaciones con claves externas entre las entidades. La creación de perfiles de datos puede ayudar a descubrir reglas de negocios incorporadas en los conjuntos de datos, que pueden usarse para la inspección y el control continuos.

Para parafrasear a Lord Kelvin, lo que no puede medirse no puede mejorarse. La creación de perfiles de datos brinda un mecanismo concreto para analizar, evaluar, cuantificar y medir la calidad de los datos a través de distintas dimensiones, tales como integridad, validez y razonabilidad. La creación de perfiles se reconoce hoy como el primer paso en la mayoría de los proyectos de migración e integración de datos y de administración de datos maestros, para comprender los problemas potenciales que existen en cada conjunto de datos y ayudar a diseñar transformaciones que limpiarán esos datos al moverlos a su destino.

Análisis sintáctico

En muchas instancias, se espera que los valores de datos cumplan con ciertos tamaños y formatos predefinidos. Sin embargo, pequeñas variaciones o ambigüedades en las representaciones de los valores pueden confundir tanto a las personas como a las aplicaciones automatizadas. Por ejemplo, considere estos diferentes valores: {California, CA, Calif., US-CA, Cal, 06}. Unos valores usan caracteres, algunos cifras, y otros, signos de puntuación o caracteres especiales. En su mayor parte, una persona los leería y reconocería que todos representan el mismo valor conceptual: el estado de California. Sin embargo para automatizar el proceso de detección de su precisión o para investigar si existen registros duplicados, los valores deben ser analizados sintácticamente para detectar los segmentos que los componen y luego transformarlos a un formato estándar.

UN CONJUNTO CONSOLIDADO DE SERVICIOS DE DATOS DE VALOR AGREGADO PUEDE REDUCIR LOS COSTOS OPERATIVOS Y AUMENTAR LA EFICIENCIA, AL TIEMPO QUE MEJORA LA CALIDAD Y LA UTILIDAD DE LOS DATOS EMPRESARIALES

Describir los distintos componentes y patrones de formato utilizados para representar un objeto de datos (el nombre de una persona, la descripción de un producto, etc.) es parte del servicio central de análisis sintáctico, que se utiliza para determinar si un valor cumple con un patrón reconocible como parte del proceso de evaluación, correspondencia y limpieza. El análisis sintáctico basado en patrones permite entonces el reconocimiento automático y la posterior estandarización de componentes significativos de los valores, tales como códigos postales, centrales telefónicas en números de teléfono, o las distintas partes del nombre de una persona. Este tipo de servicio incorpora, por necesidad, la habilidad de definir conjuntos de patrones, listas de valores de datos que pertenecen a dominios definidos de valores, o expresiones regulares y gramáticas.

Cuando la misma información se representa muchas veces en formas distintas resulta difícil determinar si registros distintos (ya sea en el mismo conjunto de datos o entre conjuntos diversos) se refieren a la misma entidad. Pero incluso antes de que puedan aplicarse algoritmos más complejos de vinculación de registros, es necesario analizar sintácticamente el texto en los campos de datos críticos para identificar y posteriormente separar y reordenar los símbolos clave que contienen la información de identificación. Junto con la estandarización, el análisis sintáctico ayudará a reducir la varianza entre muchos registros y a facilitar su correspondencia y vinculación.

Estandarización

El análisis sintáctico usa patrones definidos, expresiones regulares o gramáticas, administrados por un motor de reglas junto con búsquedas en tablas, para distinguir entre los valores válidos e inválidos de los datos. Cuando se reconocen los patrones, otras reglas y acciones pueden iniciarse para transformar los datos de entrada en una forma que permita un uso más eficaz, ya sea para estandarizar la representación (asumiendo una representación válida) o para corregir los valores (si se identifican errores conocidos). Para continuar con nuestro ejemplo, cada uno de nuestros valores para el estado de California {California, CA, Calif., US-CA, Cal, 06} puede estandarizarse según el código de dos caracteres del Servicio Postal de los Estados Unidos: CA.

El servicio de estandarización se agrega al servicio de análisis sintáctico, que puede combinarse con una librería de dominios de datos para dividir los valores de los datos en múltiples componentes y reorganizar esos componentes

en un formato normalizado. La estandarización también puede cambiar palabras completas por sus abreviaturas, o abreviaturas a palabras completas, transformar sobrenombres en una forma de nombres estándar, traducir entre idiomas (p. ej., del español al inglés), y corregir errores comunes de escritura.

La estandarización proporciona dos tipos diferentes de valores. Primero, al mapear las representaciones variadas en patrones bien definidos, el proceso de estandarización puede ayudar a normalizar distintas representaciones y limpiar eficazmente los datos con problemas. Segundo, mapear un valor a una representación estandarizada (incluso sin limpiarla específicamente) reduce la varianza en la estructura de los valores y mejora la vinculación de los registros para propósitos de deduplicación y consolidación de datos maestros. Estandarizar sobrenombres (como "Bob," "Rob," o "Bobby" a "Robert") aumentará la probabilidad de que los registros que representan la misma entidad sean vinculados entre sí, y eso aumenta significativamente la resolución unificada necesaria para un programa exitoso de calidad de datos y administración de datos maestros.

Resolución de identidades

Debido a que los sistemas operativos han crecido en forma orgánica hasta convertirse en un conjunto de aplicaciones empresariales, no es extraño que múltiples instancias de datos en distintos sistemas tengan formas distintas de referirse a la misma entidad del mundo real. Por otro lado, el deseo de consolidar y vincular los datos sobre los mismos conceptos de negocios con un alto nivel de confianza puede convencer a alguien de que un registro puede no haberse creado todavía para una entidad de mundo real, cuando en realidad sí existe. Estos dos problemas representan en última instancia el mismo desafío central: comparar datos de identificación entre un par de registros para determinar su similitud o distinguir las entidades representadas en esos registros.

Ambas cuestiones se abordan mediante un proceso llamado resolución de identidades, en el cual el grado de similitud entre dos registros cualesquiera es evaluado, la mayor parte de las veces según una aproximación ponderada a la correspondencia entre un conjunto de valores de atributos entre los dos registros. Si la puntuación está por encima de un límite específico, se entiende que los dos registros coinciden, y se presentan al cliente final como probables representaciones de la

Servicios centrales de datos: componentes básicos para establecer valor en los negocios

misma entidad. La resolución de identidades se usa para reconocer si sólo pequeñas variaciones sugieren una conexión entre registros diferentes y los valores pueden ser limpiados, o si suficientes diferencias entre los datos sugieren que los dos registros verdaderamente representan entidades diferentes. Como comparar cada registro contra todos los demás requiere cálculos intensivos, muchos servicios de datos utilizan algoritmos avanzados para bloquear en conjuntos más pequeños a los registros que más probablemente contengan coincidencias y así reducir el tiempo de cálculo.

La resolución de identidades es un componente crítico de la mayoría de las aplicaciones de calidad de datos, administración de datos maestros e inteligencia de negocios. El deseo de “ver al cliente desde todos los ángulos” o de lograr un catálogo completo de productos se basa en la habilidad para encontrar todo los registros que contienen información sobre cada entidad única y a resolverlos en una vista unificada. El análisis sintáctico y la estandarización son pasos preliminares, pero la técnica clave es la resolución de identidades.

Vinculación, combinación y consolidación de registros duplicados

La resolución de identidades proporciona los cimientos para un servicio central de datos más sofisticado: el análisis y la eliminación de registros duplicados. Si se identifican registros similares dentro del mismo conjunto de datos probablemente haya duplicaciones que se puedan limpiar y/o eliminar. Si se identifican registros semejantes en distintos conjuntos de datos puede existir una vinculación entre ellos, lo que facilita la combinación de registros similares para propósitos de limpieza de los datos y para permitir iniciativas de administración de datos maestros o de integración de los datos del cliente.

La resolución de identidades conduce al proceso de combinación y consolidación para establecer la vista única del cliente. Si bien algunas aplicaciones de negocios pueden ser capaces de proceder examinando un conjunto de registros que contiene información, automatizar el proceso de combinación seleccionará los mejores valores de datos entre esos registros y permitirá la creación de una “mejor copia” única de los datos, que permite apoyar tanto procesos operativos como analíticos para mejorar la forma de hacer negocios.

Limpieza de datos

Después de determinar que un valor de datos no cumple con las expectativas del usuario final, el analista de datos puede tomar distintos caminos para solucionarlo. Un enfoque es el de transformar los datos a una forma que cumpla con el nivel de aceptabilidad del usuario de negocios. Este servicio de limpieza de datos se suma a los servicios de análisis sintáctico, estandarización y mejora de datos, así como a la resolución de identidades y la vinculación de registros. Al realizar el análisis sintáctico de los valores y detectar los patrones de error conocidos, el servicio de limpieza de datos puede aplicar reglas de transformación para imputar valores de datos, corregir nombres o direcciones, eliminar datos extraños o no significativos, e incluso combinar registros duplicados.

Si bien una de las mejores prácticas en cuanto a calidad de datos es detectar la fuente de los errores y eliminar las causas que originan la insuficiente calidad de los datos, frecuentemente el equipo de calidad de datos no tiene el control administrativo para modificar las aplicaciones de producción existentes. Por lo tanto, una alternativa que puede ser necesaria para garantizar la más alta calidad de los datos es limpiar los datos directamente, y al menos brindar información adecuada para todos los consumidores posteriores de esos datos, lo cual es posible gracias al servicio de limpieza de datos.

Mejora de datos

Un enfoque distinto para aumentar el valor de los datos tiene que ver con su mejora, que se genera al agregar información adicional a los registros ya existentes. La mejora de los datos, la cual se construye sobre el análisis sintáctico, la estandarización y la vinculación de registros, es un proceso de perfeccionamiento de los datos que agrega información proveniente de conjuntos de datos de terceros (como la estandarización de nombres, la importación de datos demográficos, importación de datos psicográficos y añadidos por listas de hogares). El servicio de mejora se simplifica significativamente cuando el proveedor se ha asociado con concentradores de datos, cuya información puede utilizarse tanto como sistema de registros contra el cual se compara las instancias de datos, así como un recurso que puede ser agregado. Una actividad típica de mejora de datos, estandarización de direcciones y limpieza, depende del análisis sintáctico, la estandarización y la disponibilidad de información de direcciones de terceros.

LA CREACIÓN DE PERFILES DE DATOS SE RECONOCE HOY COMO EL PRIMER PASO EN LA MAYORÍA DE LOS PROCESOS DE MIGRACIÓN DE DATOS, INTEGRACIÓN DE DATOS Y ADMINISTRACIÓN DE DATOS MAESTROS

Hay pocas industrias que no se benefician del uso de fuentes de datos adicionales para agregar y mejorar la información de la empresa. La mejora de datos es un proceso que optimiza las decisiones de negocios en los contextos operativos y analíticos. La estandarización y corrección de direcciones para reducir los costos de envío y mejorar las tasas de respuesta directa, los agregados geodemográficos que brindan entendimiento sobre los perfiles y la segmentación de los clientes, datos de crédito para evaluar y reducir el riesgo, información histórica (como registros de servicio mecánico o eventos médicos) pueden proporcionar mayor comprensión al diagnóstico y cuidado preventivo de la información, entre otros.

Localización geográfica y geocodificación

La geocodificación es un proceso de localización inteligente para traducir una ubicación conceptual a coordenadas específicas en un mapa sobre la superficie terrestre. Un geocodificado es un par de coordenadas para la latitud y longitud de la ubicación, y el servicio de ubicación geográfica (o “geoubicación”) proporciona un par de coordenadas “latlong” cuando se le entrega una dirección. Al combinarse con el servicio de mejora de datos, los valores pueden determinarse (geocodificarse) a partir del nombre de un lugar, una dirección, un código postal, o la intersección de dos calles. Este servicio también funciona en forma inversa si usted le proporciona valores de latitud/longitud.

Muchas industrias dependen de la localización geográfica para brindar soporte a procesos de negocios de evaluación de riesgos, especialmente en los sectores de finanzas y seguros. Los negocios de venta minorista de todo tipo y tamaño continuamente están revisando ubicaciones para determinar los lugares óptimos para nuevas tiendas, sucursales, etc. Es así como la combinación de mayor movilidad y un aumento en la capacidad de proporcionar información “en cualquier momento y lugar” sugiere que habrá una necesidad creciente de las aplicaciones de negocios por ser “conscientes de la ubicación”.

Mapeado y análisis espacial

A menudo se espera que las aplicaciones que registran la ubicación de las interacciones de negocios ubiquen las posiciones de la información espacial (p. ej., una

ubicación, un punto de interés, una ruta o una dirección) en un mapa. El servicio de mapeado permite la edición, el análisis espacial y la visualización de información espacial en un mapa. A su vez, el usuario final puede querer considerar ubicaciones adicionales o puntos de interés relativamente próximos a posiciones identificadas, y un servicio de directorio combina el análisis espacial con datos provistos para realizar búsquedas de proximidad en categorías definidas (como una guía de restaurantes o una lista de parques públicos) y proporciona un listado de los puntos de interés más cercanos.

El análisis espacial brinda una revisión estadística y cuantitativa de los fenómenos y las actividades en el espacio, y permite la derivación de conocimiento sobre el contexto geográfico a partir de la información existente. Conocer dónde reside una persona o donde sucede un evento brinda un vector de datos; el análisis espacial permite que uno responda a preguntas más interesantes, tales como:

- ¿Cuántos clientes similares están dentro de un radio determinado?
- ¿Cuáles son las tasas de mortalidad por enfermedad en las distintas partes del país?
- ¿Cuáles publicaciones de venta minorista cubren mejor a una región específica?
- ¿Cuántas ejecuciones de hipotecas han tenido lugar en un vecindario específico?
- ¿A cuántos clientes se asiste en cada zona postal?
- ¿Cómo varió la mediana del ingreso en una región durante un determinado período?

Todo evento de negocios o transacción sucede en algún lugar; el análisis espacial permite buscar patrones basados en la localización y usar esos resultados para mejorar el desempeño del negocio.

Enrutamiento

El enrutamiento es un servicio de localización inteligente para planificar rutas entre dos puntos, determinar el camino óptimo entre muchos puntos y crear matrices y

Servicios centrales de datos: componentes básicos para establecer valor en los negocios

8

polígonos de tiempo de viaje. El servicio de enrutamiento puede calcular el camino según distintos criterios, tales como la distancia y el tiempo, y puede brindar resultados optimizados para la menor distancia de viaje o el menor tiempo de conducción. Con el servicio de enrutamiento, las organizaciones pueden mantener altos niveles de servicio y eficiencia en los servicios de despacho, identificar zonas de captación y mejorar la toma de decisiones.

Desde un punto de vista operativo, el servicio de enrutamiento puede ayudar a optimizar la logística y las entregas, y mejorar el servicio al cliente. Encontrar las rutas óptimas reducirá los costos energéticos y puede permitir una mejor programación para aumentar la tasa de entregas. Simultáneamente, disminuir los tiempos de entrega mejorará el servicio a los clientes y aumentará su satisfacción.

Inspección y control

Debido a que la creación de perfiles de datos evidencia potenciales reglas de negocios utilizadas en y a través de distintos procesos de negocios, el analista de datos puede documentarlas y confirmar su criticidad junto con los expertos en esos temas. Esas reglas describen esencialmente las expectativas del usuario final de los datos, y pueden utilizarse para medir y controlar su validez. La auditoría y control de reglas definidas sobre la calidad de los datos proporciona una evaluación proactiva del cumplimiento de las expectativas, y los resultados de esas auditorías pueden alimentar medidas sobre la calidad de los datos que se visualicen en tableros de comando y cuadros de datos.

El servicio de auditoría y control se suma a los servicios de análisis sintáctico y creación de perfiles de datos para validar los datos en forma proactiva respecto a un conjunto de reglas de negocios definidas (o descubiertas). Esta inspección de datos puede avisar a los analistas cuando las instancias de datos no cumplen con las expectativas definidas sobre su calidad, así como crear líneas de base respecto a las cuales pueden compararse las inspecciones en curso sobre la calidad de la información.

Ya hemos visto que la creación de perfiles de datos permite el análisis y descubrimiento de problemas potenciales con los datos que pueden tener impacto sobre el éxito de los negocios. Pero al brindar soporte a la inspección y el control de los datos, la creación de perfiles cumple con otra necesidad crítica de negocios para la aseguración continua de la calidad de la información. El servicio de creación de perfiles de datos impulsa su gobernanza operativa y entrega información a los encargados de los datos sobre fallas que pueden revisarse y corregirse antes de que ocurra un impacto negativo sobre el negocio.

Capas de servicios de información de negocios

Cada vez más organizaciones buscan oportunidades para reducir los costos del diseño, desarrollo y mantenimiento de tecnología, tanto la desarrollada internamente como la adquirida de manera externa. Combinar los servicios centrales de datos con productos de información agregada en servicios estandarizados de información de negocios permitirá el uso de distintas aplicaciones en toda la empresa. Estos servicios suelen ocuparse de necesidades habituales en diversas aplicaciones de negocios, y proporcionar versiones estandarizadas ayudará a reducir los costos y a mejorar la eficiencia operativa. Aquí hay algunos ejemplos comunes:

- **Validación de la calidad de los datos** – Cuando los datos de uso habitual se obtienen o generan a través de más de un canal (por ejemplo los datos de clientes ingresados a través de un sitio web o transcritos por un representante del centro de llamadas), existe la posibilidad de variaciones en la escritura, el orden de los nombres, o la ausencia de títulos y tratamientos honoríficos. Las reglas de validación aplicadas en forma inconsistente pueden llevar al ingreso duplicado de datos, o impedirán que se identifiquen registros en los conjuntos de datos existentes.

PITNEY BOWES BUSINESS INSIGHT BRINDA UNA SOLUCIÓN EMPRESARIAL QUE FACILITA CENTRALIZAR EL CONTROL Y MEJORAR EL DESEMPEÑO

Para solucionar las posibles inconsistencias, un servicio de validación de la calidad de los datos puede prepararse combinando los servicios de análisis sintáctico, estandarización, limpieza y creación de perfiles. Este servicio de validación de la calidad de los datos puede aplicar conjuntos comunes de estandarización de datos y reglas de validación a múltiples canales de obtención de datos en toda la organización. Estandarizar las restricciones y transformaciones para la calidad de los datos en todos sus puntos de captura e intercambio reducirá significativamente los riesgos de duplicación o incongruencias, y conducirá a datos consistentes en toda la empresa.

- **Evaluación de riesgos sobre la ubicación de propiedades** – Toda industria que depende de evaluar los riesgos relacionados con ubicaciones y propiedades se beneficiará de un servicio estandarizado de evaluación de riesgos de ubicación. Por ejemplo, las aseguradoras de propiedad y contra accidentes evalúan los riesgos asociados con la emisión de una póliza de seguros sobre una propiedad, incluidos los riesgos de catástrofes naturales como huracanes o terremotos, el crimen en la zona, y la disponibilidad y eficacia de los servicios de emergencia. Los emisores de hipotecas están interesados en la clasificación de las propiedades (p. ej. vivienda unifamiliar o condominio), si están ocupados por sus dueños, los valores de propiedades comparables, y las cesaciones de pagos y ejecuciones de hipotecas en el vecindario. Pero una vez más, sin embargo, distintas aplicaciones de negocios dentro de la misma organización pueden aplicar sus reglas de evaluación de riesgos en formas diferentes, utilizando técnicas y herramientas distintas.

Un servicio estandarizado para la evaluación de riesgos de ubicación de propiedades incorporaría el análisis sintáctico de direcciones y la mejora de datos para proporcionar una dirección estandarizada. Esa dirección estandarizada puede enviarse a los servicios de codificación geográfica y mapeado combinados con conjuntos de datos agregados que contienen atributos de riesgos para la propiedad, zonas de catástrofes naturales y otros atributos relacionados con la ubicación geocodificada.

- **Responsabilidad operativa de los datos** – En muchas organizaciones existe un creciente interés por

operacionalizar políticas y procedimientos para la gobernanza de datos mediante la introducción del rol de responsable de datos, quien se encarga de garantizar que la información cumple con las necesidades de los trabajadores que la utilizarán posteriormente. Las expectativas sobre la calidad de los datos son definidas por los usuarios finales de la información, pero debido a que los conjuntos de datos creados en un contexto son utilizados por múltiples procesos de negocios, los requisitos sobre la calidad de los datos recolectados deben imponerse a través de las numerosas etapas de procesamiento y múltiples procesos de un extremo a otro.

Los servicios de responsabilidad operativa de datos pueden usarse para instituir inspecciones sobre la calidad de los datos aplicadas en forma consistente, y para controlar los procesos de negocios en curso. Las reglas de calidad de los datos pueden definirse una vez e implementarse utilizando servicios de creación de perfiles de datos para validar los registros individuales o conjuntos de ellos, en el mismo lugar o cuando se transfieren entre actividades. Un servicio habitual de responsabilidad de datos inspeccionará automáticamente los datos para controlar su validez, y notificar al responsable cuando las expectativas sobre su calidad no se cumplen. Además, pueden recolectarse estadísticas acumulativas y presentarse a través de un tablero de comando o cuadros de datos.

- **Creación de perfiles demográficos** – Un aspecto del marketing dirigido tiene que ver con la combinación de datos individuales confiables e información demográfica geográfica. Como la gente de hábitos similares tiende a mantenerse junta, entender las características generales asociadas con la ubicación de un potencial cliente brinda información sobre los enfoques para dirigir el marketing.

Dado el nombre y domicilio de un cliente potencial, un servicio de creación de perfiles demográficos puede analizar sintácticamente y estandarizar su nombre, empleando la resolución de identidades para compararlo con clientes conocidos o identificar partes relacionadas con un conjunto de datos sobre clientes conocidos. A continuación, la dirección puede ser analizada sintácticamente y estandarizada. A su vez, los datos agregados asociados con la ubicación mapeada pueden usarse para agregar información

Servicios centrales de datos: componentes básicos para establecer valor en los negocios

10

demográfica valiosa relacionada con la región en la que se ha mapeado esa ubicación. Combinar los datos demográficos con el registro del cliente potencial proporcionará al equipo de ventas las recomendaciones para la dirección del marketing según el perfil que se haya identificado.

- **Administración de impuestos jurisdiccionales** – Otro aspecto del cumplimiento y la administración de riesgos tiene que ver con la determinación de las tasas de impuestos jurisdiccionales. Si bien este tipo de servicios es importante para el análisis de riesgo de propiedades, también es necesario para organizaciones con responsabilidades impositivas corporativas asociadas con distintas ubicaciones.

Este tipo de servicio combina los servicios de estandarización de direcciones y limpieza de datos, de geocodificación y localización, y de provisión de datos agregados para brindar la información correcta sobre los impuestos estatales, del condado, del distrito, del municipio y especiales asociados con una dirección, para estar al tanto de las jurisdicciones impositivas continuamente cambiantes y de las exigencias municipales y jurisdiccionales.

Consideraciones

Estos ejemplos son tan solo algunos de los servicios de negocios que pueden diseñarse mediante los servicios centrales de datos. De manera interesante, tienden a exhibir estas tres características como subproducto de la combinación de los servicios centrales de datos:

- Servicios de análisis sintáctico, estandarización e inspección y control de datos para normalizar y limpiar los datos,

- La combinación de datos limpios y localización inteligente para preparar los datos para su mejora, y
- La mejora y el aumento del valor a través de los agregados de información provenientes de conjuntos de datos agrupados.

Si existe un acuerdo en que las prácticas sólidas de administración de datos, como la gobernanza de datos, el aseguramiento de la calidad de los datos y la mejora de los datos pueden mejorar los resultados finales, entonces vale la pena considerar la forma de implementar servicios de información de negocios estandarizados mediante un sistema de servicio central de datos. Una evaluación sobre las necesidades de información de negocios resaltarán los servicios de negocios que, si se estandarizan a través de toda la empresa, mejorarán la calidad y consistencia y disminuirán los costos operativos.

Una reflexión final: pueden existir muchas formas de diseñar e implementar los servicios centrales de datos, que van desde su producción interna, pasando por la adquisición de conjuntos de productos, hasta su completa tercerización. Pero si es claro que los servicios críticos de información de negocios dependen de una combinación de calidad de datos, localización inteligente y conjuntos de datos con valor agregado, entonces debería incorporarse la calidad de esos aspectos en cualquier decisión de compra o producción. En otras palabras, los criterios para la evaluación de soluciones deben incorporar el grado en que los productos soportan los servicios centrales de calidad de datos, los servicios de localización inteligente y la integración con los conjuntos de datos de valor agregado.

PITNEY BOWES BUSINESS INSIGHT: SU FUENTE DE
RESPUESTAS Y SOLUCIONES



ESTADOS UNIDOS

One Global View
Troy, NY 12180
teléfono: +1.518.285.6000
1.800.327.8627
fax: +1.518.285.6070
la.sales@pb.com
www.pbinsight.com

LATINOAMÉRICA Y EL CARIBE

OFICINA MIAMI
15100 NW 67th Avenue, Suite 403
Miami Lakes, FL 33014
U.S.A.
teléfono: +1.305.825.8381
fax: +1.305.826.1246
la.sales@pb.com

OFICINA MÉXICO
San Francisco 1393
Colonia Del Valle
Delegación Benito Juárez
03210 México, DF
México
teléfono: +52.55.5636.0949
la.sales@pb.com

OFICINA BRASIL
Rua Dom Aguirre, 438-3o andar
04671-903 São Paulo, SP
Brasil
teléfono: +55.11.5698.5555
la.sales@pb.com
www.pbinsight.com.br