

Introducción a la Gestión de Datos

Qué es y por qué es importante



Usted ha investigado lo suficiente para saber que la gestión de datos es el primer paso importante para manejar el big data o iniciar cualquier proyecto de analítica. Pero no está muy convencido en admitir que aún no entiende las diferencias entre la gestión de datos maestros y la federación de datos.

O tal vez conoce perfectamente estos términos, y probablemente siente que se los ha estado explicando a su jefe o a sus unidades de negocio una y otra vez.

De cualquier manera, hemos creado el manual básico que usted ha estado buscando. Imprímalo, publíquelo en el tablero de anuncios de su equipo, o compártalo con su madre para que ella pueda entender lo que usted hace. Y recuerde, una estrategia de gestión de datos nunca se debe enfocar en una de estas áreas. Necesita considerarlas todas.

Data Quality

¿qué es? Data Quality es la práctica de asegurarse de que los datos sean precisos y útiles para un propósito determinado. Al igual que la gestión de la calidad ISO 9000 en el sector de manufactura, la calidad de los datos debe aprovecharse en cada paso del proceso de gestión de datos. Esto comienza desde el momento en que se tiene acceso a los datos, a través de varios puntos de integración con otros datos, e incluye el punto antes de que sean publicados, reportados o se haga referencia a ellos en otro destino

¿por qué es importante? Es muy sencillo almacenar datos, pero ¿cuál es el valor de esos datos si son incorrectos o no son útiles. Un ejemplo sencillo es un archivo con el texto "CALLE PRINCIPAL 123 CualquierPueblo, AZ 12345" contenido en él. Una computadora puede almacenar esta información y proporcionársela a un usuario, pero sin ayuda, puede determinar que este registro es una dirección, que es parte de la dirección, es el estado, o si el correo enviado a la dirección llegará a su destino. Corregir un registro único es fácil, ¡pero no lo es si intenta realizar este proceso en cientos, miles o incluso millones de registros! Es mucho más rápido utilizar una solución de calidad de datos que pueda estandarizar, analizar y verificar de una forma consistente y automática. Al hacerlo en cada paso, pueden eliminarse los riesgos, como es enviar correo a la dirección errónea del cliente.

Data Integration

¿qué es? Una vez que usted ha tenido acceso a los datos, ¿qué puede hacer con ellos? Un siguiente paso muy común es combinarlos con otros datos para presentar resultados unificados. La integración de datos es el proceso que define los pasos para hacer esto, y las herramientas de integración de datos ayudan a diseñar y automatizar los pasos que realizan este trabajo. Las herramientas de integración de datos más comunes son conocidas como ETL, que es el acrónimo de extraer, transformar y cargar, y ELT, que a su vez es el acrónimo de extraer, cargar y transformar. Actualmente, la integración de datos no se limita a los movimientos entre las bases de datos. Con la disponibilidad de los servidores en la memoria, usted podría estar cargando los datos directamente en la memoria, lo que evita totalmente a las bases de datos tradicionales.

¿por qué es importante? La integración de datos es lo que le permite a las organizaciones crear combinaciones de datos que son más útiles para tomar decisiones. Por ejemplo, un conjunto de datos podría incluir una lista de todos los nombres y sus direcciones. Otro conjunto de datos podría ser una lista de la actividad en línea y los nombres de los clientes. Por sí mismo, cada conjunto de datos es relevante y puede decirle algo importante. Pero cuando usted integra elementos de ambos conjuntos de datos, puede comenzar a responder preguntas como, "¿Quiénes son mis mejores clientes?" "¿Cuál es la próxima mejor oferta?" Combinar información clave de cada conjunto de datos le permitiría crear la mejor experiencia del cliente.

Data Federation

¿qué es? Data federation es un tipo especial de integración de datos. Los tipos de integración de datos ETL y ELT combinan datos y los almacenan en otro sitio para usarlos, en el pasado era en un data mart o en un almacén de datos. Pero, ¿y si usted únicamente quiere ver los resultados combinados sin la necesidad de moverlos o almacenarlos con anticipación? La federación de datos ofrece la capacidad de hacer eso, lo que le permite a tener acceso a los datos combinados en el momento que se solicite.

¿por qué es importante? Si bien muchas herramientas de integración de datos ETL y ELT pueden trabajar muy rápidamente, sus resultados sólo pueden representar una imagen de lo que sucedió en un punto en el tiempo cuando el proceso se realizó. Con la federación de datos, se genera un resultado basado en la forma en que lucen los datos en el momento en que se solicita el resultado. Esto permite tener una vista de la información más oportuna y mucho más precisa. Imagine que usted está comprando un regalo para un ser querido en una tienda. Cuando paga, recibe una oferta de otro producto que complementa el regalo que usted eligió y resulta ser algo que su ser querido disfrutaría. Aún mejor - el producto está en existencia en la misma tienda. Gracias al análisis en tiempo real de los datos de la próxima mejor oferta y de los datos de ubicación, la tienda mejora su experiencia de compra al hacerle llegar una oferta conveniente y relevante en el momento correcto en el lugar correcto.

Data Governance

¿qué es? Data Governance es el ejercicio de la autoridad para tomar decisiones sobre los procesos que gestionan los datos de su organización. O para decirlo en otras palabras, es asegurarse de que su estrategia de datos esté alineada con su estrategia de negocio.

¿por qué es importante? El gobierno de datos comienza haciendo preguntas y desarrollando políticas alrededor de las respuestas: ¿Cómo utiliza su organización los datos? ¿Cuáles son las limitaciones con las que usted trabaja? ¿Cuál es el entorno regulatorio? ¿Quién tiene la responsabilidad de los datos? Una vez que las respuestas a estas preguntas se conocen, pueden definirse las reglas para aplicarlas. Ejemplos de tales reglas podrían estar definiendo a qué datos pueden acceder los usuarios, qué usuarios pueden cambiar los datos versus simplemente verlos, y cómo se manejan las excepciones a las reglas. Se pueden utilizar entonces herramientas de gobierno de datos para controlar y gestionar las reglas, perfilar cómo se manejan, y generar reportes para propósitos de auditoría.

El aspecto de auditabilidad de esto es probablemente el más vital, pues los líderes de las organizaciones tienen que refrendar la precisión de los reportes financieros para los consejos de gobierno, los accionistas, los clientes y las instancias gubernamentales. Es una gran responsabilidad y si no se maneja correctamente, corre el riesgo de ser censurado, exponerse a recibir fuertes sanciones e incluso a una acción legal.

Master Data Management

¿qué es? Master data management (MDM) es un conjunto de procesos y tecnologías que definen, unifican y gestionan todos los datos que son comunes y esenciales para todas las áreas de una organización. Estos datos maestros son gestionados típicamente desde un solo lugar, a menudo denominado un centro de gestión de datos maestros, el cual actúa como un punto de acceso común para publicar y compartir estos datos críticos a lo largo de la organización de una manera consistente.

¿por qué es importante? Fácil: Porque asegura que diferentes usuarios no estén utilizando diferentes versiones de los datos comunes y esenciales de la organización. Sin MDM, un cliente que compra un seguro de una compañía podría seguir recibiendo ofertas para comprar un seguro de la misma aseguradora. Esto sucede cuando la información gestionada por la base de datos de relaciones con los clientes y la base de datos de marketing que no están ligadas, lo que genera dos registros distintos de la misma persona - y un cliente confuso e irritado.

Con la gestión de datos maestros, todos los sistemas organizacionales y las fuentes de datos pueden vincularse y gestionarse de forma consistente continuamente para asegurar que los datos maestros utilizados por la organización sean siempre consistentes y precisos. En el mundo del big data, MDM también puede automatizar la manera en que se utilizan ciertas fuentes de datos, qué tipos de modelos analíticos se aplican, en qué contexto se aplican y las mejores técnicas de visualización para sus datos