Informatica

White paper

Inteligencia artificial para las empresas inteligentes basadas en datos

Cómo impulsan los nuevos avances en el ámbito de la gestión de datos las innovaciones de CLAIRE basadas en el aprendizaje automático

Acerca de Informatica

La transformación digital modifica las expectativas: mejor servicio, entrega más rápida, menores costes. Los negocios deben transformarse para seguir siendo relevantes y los datos tienen la respuesta.

Como líder mundial en gestión de datos de cloud empresariales, le brindamos ayuda para que encabece la marcha de forma inteligente, en cualquier sector, categoría o nicho. Informatica le aporta perspectiva para que aumente su agilidad, concrete nuevas oportunidades de crecimiento o incluso invente cosas nuevas. Al estar completamente centrados en todo lo relacionado con los datos, ofrecemos la versatilidad necesaria para alcanzar el éxito.

Le invitamos a explorar todo lo que puede ofrecerle Informatica y a desatar el poder de los datos para impulsar su próxima disrupción inteligente.

Contenidos

"Los líderes de datos y análisis se enfrentan a las complejidades de su panorama de datos. Nuestras predicciones sobre soluciones de gestión de datos reconocen los desarrollos clave y la creciente demanda de funcionalidades de cloud, arquitecturas de datos conectados, metadatos y la automatización de tareas rutinarias y no rutinarias mediante la aplicación de la inteligencia artificial".[1]

— Gartner

La importancia de la inteligencia artificial

La inteligencia artificial (IA) y el aprendizaje automático (ML) están impulsando las transformaciones digitales que ocurren en todos los sectores a nivel mundial. La IA es la prioridad para los directivos a la hora de diseñar estrategias para transformar sus empresas. Además, se ha convertido en un elemento esencial para la mejora de nuestra vida diaria, presente tanto en las películas que vemos como, incluso, en los automóviles que conducimos. La inteligencia artificial y el aprendizaje automático (IA/ML) son fundamentales para el descubrimiento de nuevas terapias en el ámbito de las ciencias de la salud, ya que reducen el fraude y los riesgos en el sector de los servicios financieros y ofrecen experiencias de cliente realmente personalizadas.

Para los líderes empresariales, la combinación IA/ML puede parecer cosa de magia; aunque sus ventajas potenciales son evidentes, es posible que no las entiendan o no sepan cuál es la mejor manera de sacar partido de estas potentes innovaciones. IA/ML es la tecnología que respalda muchas de las nuevas soluciones empresariales, ya sea para aplicaciones de próximas mejores acciones, seguimiento de la satisfacción del cliente, eficiencia de las operaciones y desarrollo de productos innovadores. El aprendizaje automático en general y, en concreto, el aprendizaje profundo, necesitan los datos. Para conseguir la precisión requerida, el ML necesita grandes cantidades de datos para formarse. Estos datos deben ser un reflejo fiel del estado actual de la empresa. Una IA basada en datos imprecisos o limitados tendrá un impacto muy negativo en las iniciativas empresariales, hasta el punto de que podría dar lugar a consecuencias contrarias a los resultados esperados.

Para que la IA sea eficaz, es decir, que se utilicen y formen las funciones correctas, debemos acceder a una amplia variedad de datos, dentro y fuera de la organización. Estos datos se deben recopilar de tal manera que sea posible crear y entrenar un modelo de ML. Para ello, es precisa la gestión de datos. No solo se trata de saber gestionar su mayor volumen y complejidad; es también una cuestión de confianza. ¿Los datos que se utilizan para "entrenar" el modelo proceden de los sistemas adecuados? ¿Se ha eliminado la información personal identificable (PII) y se cumple la normativa pertinente? ¿Hay transparencia y es posible demostrar el linaje de los datos que utiliza el modelo? ¿Es posible documentarse y estar preparado para mostrar a los reguladores o investigadores que no hay sesgo en los datos? Todo esto requiere un buen control y una base para la gestión de los datos. Sin una base sólida para la gestión de datos, la IA será incomprensible y poco fiable; en otras palabras, sin la gestión de datos, la IA puede ser una caja negra de consecuencias no deseadas.

La IA necesita datos

El éxito de la IA depende de la eficacia de los modelos diseñados por los científicos de datos para entrenarla y ampliarla. Y el éxito de esos modelos depende de la disponibilidad de datos fiables y oportunos.

¿Por qué los científicos de datos encargados de la creación de modelos de IA/ML necesitan datos de alta calidad? Pensemos, por ejemplo, en un modelo de predicción cuyo fin es anticipar el comportamiento de los consumidores. Una característica valiosa para ese modelo podría ser la ubicación del consumidor, indicada, por ejemplo, por el código postal. ¿Qué ocurriría si faltan esos datos del código postal, o bien son incompletos o inexactos? El comportamiento del modelo se verá afectado negativamente durante el entrenamiento y la implantación, lo que podría dar lugar a predicciones incorrectas y reducirá el valor de todo el esfuerzo. Por el contrario, un código postal exacto, completo y verificado ayudaría a predecir la segmentación del mercado, la clase de ingresos, la edad, la esperanza de vida, etc. de un consumidor, razón de más para hacer las cosas bien. Cabría esperar que la "IA explicable" se convierta en un mandato regulado, no solo en una opción. Sin el linaje y la trazabilidad basados en metadatos, no se pueden llevar a producción aplicaciones e ideas impulsadas por IA.

[1] Gartner, Predicts 2020: Data Management Solutions, Rick Greenwald, Donald Feinberg, Mark Beyer, Adam Ronthal, Melody Chien, 5 de diciembre de 2019.

La IA necesita una gestión de datos inteligente para determinar rápidamente cuáles son las características adecuadas para el modelo; transformar automáticamente los datos para satisfacer las necesidades del modelo de IA (ampliación de funciones, estandarización, etc.); eliminar datos duplicados y proporcionar datos maestros fiables sobre clientes, pacientes, socios y productos; y proporcionar el linaje de los datos de principio a fin, también dentro del modelo y de sus operaciones. El éxito de la IA depende de la eficacia de los modelos diseñados por los científicos de datos para entrenarla y ampliarla. Y el éxito de esos modelos depende de la disponibilidad de datos fiables y oportunos.

Los datos necesitan la IA

IA/ML también desempeña un papel fundamental a la hora de escalar las prácticas de gestión de datos. Debido a los enormes volúmenes de datos necesarios para la transformación digital, las organizaciones deben detectar y catalogar sus datos y metadatos más relevantes para certificar la importancia, el valor y la seguridad, así como para garantizar la transparencia. Deben depurar y dominar estos datos. Y deben además controlarlos y protegerlos de forma eficaz. Si los datos no se gestionan eficazmente (y a escala), los modelos de IA/ML tendrán el mismo destino que todas las iniciativas tradicionales de data warehousing de los últimos 30 años: utilizar datos de mala calidad y proporcionar información poco fiable.

Según una investigación reciente, se espera que el volumen total del tráfico de data centers alcance los 20,6 zettabytes en 2021, y que el número de dispositivos conectados y conexiones se multiplique hasta superar los 25 millones para el año 2022.[2] Todos estos datos se deben procesar y hacer que sean utilizables y fiables, además de conformes a las políticas de gobierno. También es necesario moverse con rapidez y responder a los cambios en la estrategia y los procesos de la empresa. El esfuerzo que implica la preparación de los datos para las iniciativas de transformación digital ha aumentado en complejidad, al tiempo que ha crecido el volumen de datos. Según LinkedIn, el puesto de científico de datos es uno de los más prometedores en EE. UU.[3] y la demanda de ingenieros de datos ha crecido en un 96 % de un año para otro.[4] Pero no basta con contratar más expertos para hacer frente al aumento en el volumen de datos.

No aplique un enfoque lineal ante un desafío exponencial

No podemos resolver estos desafíos simplemente designando más ingenieros y desarrolladores. No es un problema que se puede solventar solo a escala lineal con recursos humanos. Los enfoques tradicionales están plagados de ineficacias. Los proyectos se implementan en silos con poca visibilidad de los metadatos de principio a fin y con una automatización limitada. No hay aprendizaje, el procesamiento es costoso y los pasos de control y privacidad se repiten una y otra vez.
Ante este panorama, ¿cómo pueden las organizaciones seguir el ritmo del negocio, permitir el autoservicio, servir mejor a sus clientes, aumentar la eficacia operativa y acelerar la innovación?

[2] Cisco, Global Cloud Index Forecast y Complete Visual Networking Index Forecast.
[3] LinkedIn, "LinkedIn's Most Promising Jobs of 2019".
[4] Datanmi, "Data Engineering Continues to Move the Employment Needle".

Aquí es donde destaca la IA, ya que puede automatizar y simplificar las tareas relacionadas con la gestión de datos mediante la detección, la integración, la limpieza, el gobierno y el dominio de los datos. Los métodos de aprendizaje automático pueden aprender y asumir tareas rutinarias y repetitivas, liberando así a desarrolladores y usuarios para que puedan dedicarse a proyectos innovadores de mayor valor. La IA mejora la comprensión de los datos e identifica anomalías en la privacidad y la calidad de los mismos. También es un recurso ideal para desarrolladores, analistas, administradores y usuarios de negocio que les permite acelerar las tareas a través de la automatización y la ampliación con recomendaciones y próximas mejores acciones.

Es más eficaz cuando se utiliza para acelerar los procesos de principio a fin en todo el entorno de datos. Por esa razón consideramos que la IA es esencial para la gestión de datos y es el motivo por el que Informatica® ha centrado en gran medida sus inversiones de innovación en el motor CLAIRE®, nuestra funcionalidad de IA basada en metadatos. CLAIRE aprovecha todos los metadatos empresariales unificados para automatizar y ampliar las tareas de administración y gestión de datos rutinarias.

Cuatro ventajas principales de la IA para la gestión de datos
En general, la IA beneficia a los equipos de gestión de datos en cuatro aspectos principales: mejora la productividad de los profesionales de datos, aumenta la eficiencia de las operaciones, proporciona una experiencia de datos guiada de forma más inteligente y una comprensión más profunda, y acelera los procesos de gobierno de datos. A continuación, se muestran algunos ejemplos del actual potencial de la IA.

Productividad: un sistema de recomendaciones para la integración de datos ayuda a los ingenieros de datos a crear rápidamente mappings para extraer, transformar y entregarlos. Este sistema aprende de los mappings existentes, comprende el contenido empresarial de las bases de datos y los sistemas de archivos, y sugiere transformaciones adecuadas para estandarizar y limpiar los datos antes de entregarlos a los sistemas de destino y a los consumidores de datos.

Eficiencia: en una empresa típica, se ejecutan miles de procesos de integración de datos al día. La supervisión de estos procesos es en gran parte pasiva, con herramientas de administración que solo registran el tiempo, el espacio de CPU y la memoria consumidos. La IA puede aprender a partir de los valores históricos de los datos de series temporales de los archivos de registro y supervisión, y marcar automáticamente los valores atípicos, así como predecir problemas que podrían surgir si no se gestionan a tiempo.

Experiencia de datos: cuando una entidad del mundo real (por ejemplo, el registro de un paciente o una orden de venta) se almacena en una base de datos o en un conjunto de archivos, los datos se desglosan y distribuyen entre varias tablas o archivos, por lo que están optimizados para el almacenamiento y la mejora del rendimiento. La IA puede detectar las relaciones entre los datos y reconstituir la entidad original rápidamente. Los usuarios no tienen que recordar ni consultar documentación obsoleta sobre relaciones de clave principal/externa y unir manualmente los diversos conjuntos de datos. Además, la IA puede identificar conjuntos de datos similares y hacer recomendaciones basadas en patrones de uso, calidad de datos y colaboración de fuentes colectivas.

Gobierno de datos: un paso habitual pero tedioso en el gobierno de datos es asociar términos empresariales con elementos de datos físicos para establecer el contexto empresarial y la relevancia de los elementos de datos, y hacer que estos sean comprensibles para los usuarios. En muchos casos, la IA puede vincular automáticamente los términos empresariales con datos físicos mediante una combinación de técnicas de procesamiento de lenguaje natural (NLP) y una identificación de nivel empresarial. Esto puede reducir drásticamente la monotonía de esta tarea propensa a errores. En esta era del cloud, conviene apuntar que este enfoque también tiene validez para las aplicaciones SaaS. Es posible recopilar los metadatos de las aplicaciones SaaS, como Salesforce o Workday, y añadirlos al catálogo empresarial.

Gestión de datos basados en IA: un ejemplo del sector de la banca
Para ilustrar por qué la IA necesita la gestión de datos y viceversa, veamos un ejemplo del sector de la banca.

Mediante la aplicación de la IA a volúmenes crecientes de datos para el análisis avanzado, predictivo y en tiempo real, los bancos pueden:
- ofrecer servicios más personalizados que aumenten los índices de retención de clientes;
- reducir las transacciones fraudulentas en el punto de venta;
- mejorar los resultados de los inversores de los consumidores y, al tiempo, reducir el coste de los asesores financieros;
- reducir el coste del cumplimiento de las normativas relacionadas con el proyecto.

En el ámbito de la gestión de datos, la IA puede detectar y catalogar automáticamente todos los tipos de datos relevantes, como ERP, CRM, aplicaciones en cloud y web, archivos de máquinas y registros, datos de terceros, etc. Esto les da a los científicos de datos una ventaja para acceder a todos los datos que necesitan para ejecutar cientos de experimentos en la búsqueda de patrones que revelen información relacionada con el comportamiento del consumidor, actividades fraudulentas, oportunidades de inversión asociadas con la propensión al riesgo del consumidor y más.

La IA, en lo que se refiere a la gestión de datos, puede enriquecer automáticamente una visión integral de los clientes y las personas de interés (POI) mediante la identificación de relaciones entre los datos de clientes y la correspondencia de información con personas específicas. Esto ayuda a las organizaciones a interactuar mejor con sus clientes, ofrecerles ofertas más relevantes y proporcionarles una experiencia perfecta a través de diversos canales, ya sea online, en dispositivos móviles o por teléfono. Una visión integral de las POI ayuda a los bancos a identificar patrones y redes de actividad fraudulenta mucho más rápido, lo que podría ahorrar millones.

Además, la IA puede automatizar y guiar la integración de datos y las tareas de calidad de datos para combinar y limpiar datos de cientos de fuentes de datos, lo que aumenta la potencia predictiva de los modelos y algoritmos analíticos. Se ha demostrado que disponer de más datos de mayor calidad, en combinación con IA/ML y análisis avanzados, generan resultados significativos, como la mejora de las próximas mejores ofertas y la identificación de fraudes.

La IA también potencia el gobierno de datos, lo que garantiza que las políticas no solo se documenten, sino que realmente se apliquen. Esto ayuda a los profesionales de la seguridad de la información a cumplir con las regulaciones de privacidad de datos, como el Reglamento General de Protección de datos (RGPD), la Ley Sarbanes-Oxley (SOX), Basilea II y Basilea III, entre otras.

Informatica CLAIRE: la "inteligencia" de Intelligent Data Management Cloud

Este es el enfoque que plantea Informatica para aumentar la productividad en la gestión de datos con el aprendizaje automatizado:

1. Intelligent Data Management Cloud™: ofrecemos una plataforma de gestión de datos nativa del cloud e integrada de principio a fin para favorecer la máxima productividad. Esta plataforma unificada facilita la gestión unificada de la conectividad, los metadatos y las operaciones, de manera que acelera tanto el desarrollo como la implantación de proyectos de gestión de datos nuevos. La plataforma dispone de un conjunto potente y lógico de funcionalidades que permiten gestionar datos procedentes de fuentes del entorno local, del cloud, multicloud y entornos híbridos. El nombre de esta plataforma de datos unificada es Intelligent Data Management Cloud.

Se trata de una plataforma modular: se empieza con una sola herramienta y se añaden otras según las necesidades.

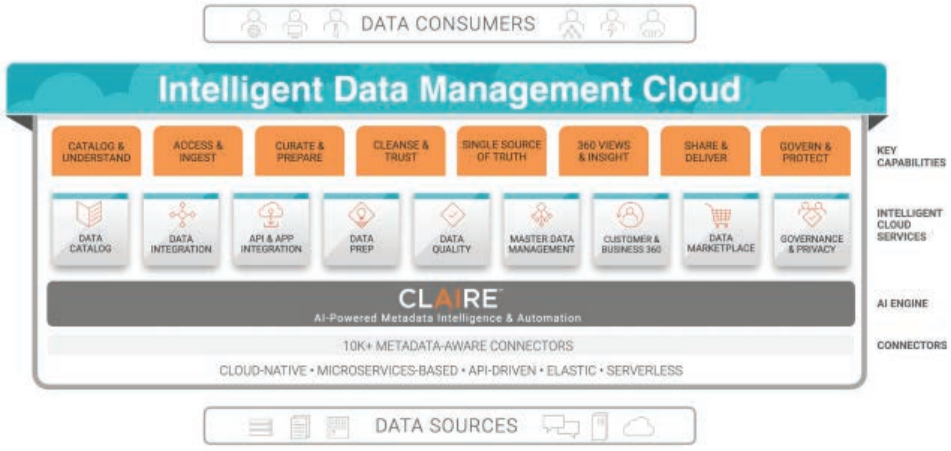

Figura 1: Intelligent Data Management Cloud integra las funcionalidades de gestión de datos con la conectividad compartida, la información operacional y la inteligencia tanto de datos como de metadatos.

2. Metadatos: hace mucho que Informatica goza de reconocimiento como líder en la gestión de metadatos técnicos y de negocio, pero, ahora, ha ampliado sus funcionalidades en este aspecto mediante la recopilación de un abanico mayor de metadatos de toda la empresa, por ejemplo, los siguientes:
   - Metadatos técnicos, como tablas de bases de datos, información de columnas, estadísticas de perfiles de datos, secuencias de comandos y linaje de datos
   - Metadatos de negocio, que capturan contexto sobre los datos, como su significado, su pertinencia y su relevancia para diferentes funciones y procesos de negocio
   - Metadatos operativos sobre la ejecución de procesos y sistemas para responder preguntas como: cuándo se actualizaron por última vez los datos, la última ejecución del proceso de carga o los datos a los que más se accede
   - Metadatos de uso sobre la actividad de los usuarios, entre otros, los conjuntos de datos consultados, los resultados de búsquedas en los que se hace clic o las calificaciones u observaciones aportadas

Este conjunto más amplio de metadatos resulta fundamental para el aprendizaje automatizado. Proporciona conjuntos de datos que sirven para "entrenar" los algoritmos del aprendizaje automático, de modo que se ajusten como sea preciso para arrojar resultados más certeros.

3. Inteligencia: Informatica ofrece con CLAIRE una combinación integrada de metadatos y aprendizaje automatizado (o inteligencia artificial).

   Los metadatos recopilados por Intelligent Data Management Cloud suponen una mina de información que permite a los algoritmos de CLAIRE obtener información sobre el panorama de datos de la empresa. Gracias a esos conocimientos, CLAIRE puede sugerir recomendaciones inteligentes, automatizar el desarrollo y la supervisión de los proyectos de gestión de datos y adaptarse a los cambios experimentados tanto en la empresa como fuera de ella. CLAIRE es el motor de la inteligencia de todas las funcionalidades de gestión de datos de Intelligent Data Management Cloud.

CLAIRE resulta útil para una amplia variedad de usuarios:

- Los ingenieros de datos cuentan con numerosas tareas de implementación parcialmente o incluso totalmente automatizadas
- Los analistas de datos localizan y preparan los datos que precisan con mayor facilidad
- Los usuarios de negocio identifican al instante los datos que se deben someter a los controles de cumplimiento y el gobierno de datos prescritos
- Los especialistas en datos interpretan los datos con más rapidez
- Los administradores de datos visualizan la calidad de datos con mayor facilidad
- Los profesionales de seguridad y privacidad de datos detectan el uso indebido de los datos, protegen los datos sensibles y demuestran la aplicación de los controles adecuados de manera más sencilla
- Los administradores y los operadores disfrutan de las facilidades del mantenimiento predictivo y la optimización del rendimiento de los procesos de gestión de datos

A continuación se ofrecen algunos ejemplos actuales de uso de la inteligencia que proporciona CLAIRE.

CLAIRE para la catalogación de datos

Detectar y comprender los datos que se tienen es el primer paso en cualquier iniciativa basada en datos. CLAIRE ofrece un motor de detección basado en el aprendizaje automático para analizar y catalogar activos de datos en toda la empresa. Un catálogo de datos inteligente respaldado por CLAIRE puede ayudar a los científicos de datos, analistas e ingenieros de datos a encontrar y recomendar los datos que necesitan, lo que reduce significativamente el tiempo dedicado a la detección y preparación de los datos.

Detección de relaciones avanzada

Una tarea clave de catalogación y modelado de datos es documentar las relaciones entre conjuntos de datos. CLAIRE utiliza técnicas de aprendizaje automático para identificar automáticamente claves primarias, claves únicas y uniones entre conjuntos de datos estructurados. Esto reduce meses de esfuerzos de documentación a minutos. CLAIRE mejora continuamente su capacidad de identificar relaciones al incluir humanos en el proceso de conservación de los datos; por ejemplo, los usuarios pueden aceptar o rechazar las relaciones deducidas y CLAIRE aprende de estas acciones.

Por ejemplo, un analista de datos de un banco que crea un informe sobre qué clientes responderán con mayor probabilidad a una campaña de marketing debería ser capaz de encontrar información sobre los productos existentes y los préstamos para todos los clientes. Sin embargo, debido a la naturaleza aislada de los datos en toda la empresa, es difícil encontrar tales conjuntos de datos entre los distintos departamentos y almacenes de datos. CLAIRE utiliza uniones documentadas en las bases de datos, conexiones realizadas con otras herramientas como BI y ETL, y estadísticas derivadas de los valores de datos para inferir y recomendar conexiones al analista de datos. Esto ayuda a ampliar el análisis del usuario y utiliza toda la información disponible para identificar la audiencia objetivo adecuada para la campaña.

CLAIRE combina varias técnicas para la detección de claves y uniones. En el caso de las claves, las estadísticas para la creación de perfiles, como la exclusividad, los valores nulos, los metadatos de columnas (por ejemplo, nombres de columnas que contienen "ID") y otros se combinan para la detección de claves primarias y únicas. La inferencia de uniones y claves de unión utiliza después una combinación de técnicas de aprendizaje automático, como el análisis de firmas de columnas, para identificar uniones a escala en muchos conjuntos de datos potenciales.

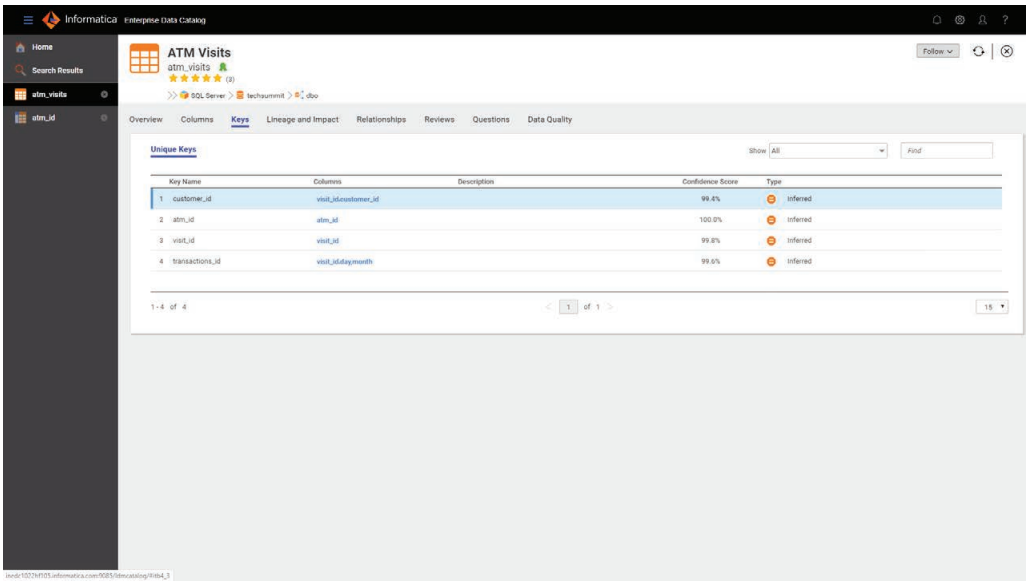

Figura 2: Detección de claves únicas a través de la inferencia mediante técnicas de aprendizaje automático.

Semejanza de datos inteligente

CLAIRE utiliza técnicas de aprendizaje automatizado como la agrupación en clústeres para detectar datos similares entre miles de bases de datos y conjuntos de archivos. La semejanza de datos inteligente es una de las funcionalidades clave que se utilizan para varios fines, como identificar los datos, detectar los duplicados, combinar en entidades de negocio campos de datos individuales, propagar etiquetas por distintos conjuntos de datos o recomendar conjuntos de datos a los usuarios.

La semejanza de datos calcula hasta qué punto se parecen los datos de dos columnas. Un enfoque de comparación forzosa de todos los pares de columnas de una configuración empresarial (por ejemplo, unos cien millones de columnas) sería prohibitivo desde el punto de vista del procesamiento. La semejanza de datos, en cambio, aplica técnicas de aprendizaje automatizado para agrupar las columnas similares e identificar las probables coincidencias.

Este proceso funciona en varias etapas. Primero, las columnas se agrupan en clústeres en función de sus características. Luego, se calcula la superposición de datos con valores únicos en cada uno de los clústeres. Por último, se seleccionan los pares más probables para calcular la semejanza de datos mediante los índices de Bray-Curtis y Jaccard.

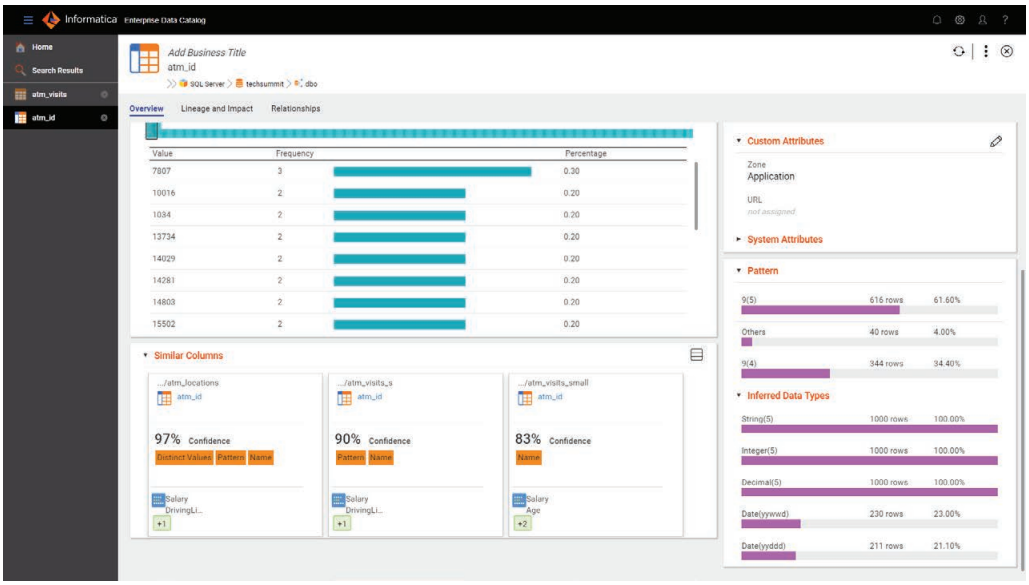

Figura 3: Identificación de columnas similares mediante la agrupación en clústeres y los coeficientes Bray-Curtis y Jaccard.

Detección inteligente de dominios con etiquetas

CLAIRE puede clasificar los campos de datos aplicando etiquetas semánticas a cada una de las columnas. Dichas etiquetas semánticas se denominan dominios de datos.

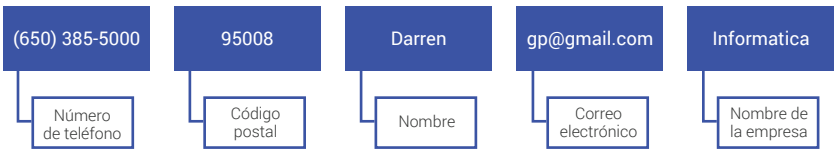

Figura 4: CLAIRE clasifica automáticamente los campos de datos y aplica las etiquetas semánticas.

Por lo general, las etiquetas semánticas se aplican evaluando reglas basadas en expresiones regulares, tablas de referencia u otra lógica compleja de codificación manual. No obstante, resulta tedioso realizar la definición y el mantenimiento de miles de reglas así.

CLAIRE se sirve del concepto de las etiquetas para simplificar en gran medida el proceso consistente en detectar y etiquetar los campos de datos. En el caso de las columnas que no estén clasificadas todavía, basta con que el usuario proporcione una etiqueta (por ejemplo, “fecha de pago de reclamaciones”) que señale de manera sencilla su contenido. El sistema aprende por asociación y, a continuación, propaga esa etiqueta a todas las columnas semejantes de forma automática. Se trata de una técnica parecida al “reconocimiento facial” que permite etiquetar a usuarios de Facebook en una fotografía, pero aplicada a los datos, es decir, equivaldría a etiquetar a las mismas personas en millones de fotos.

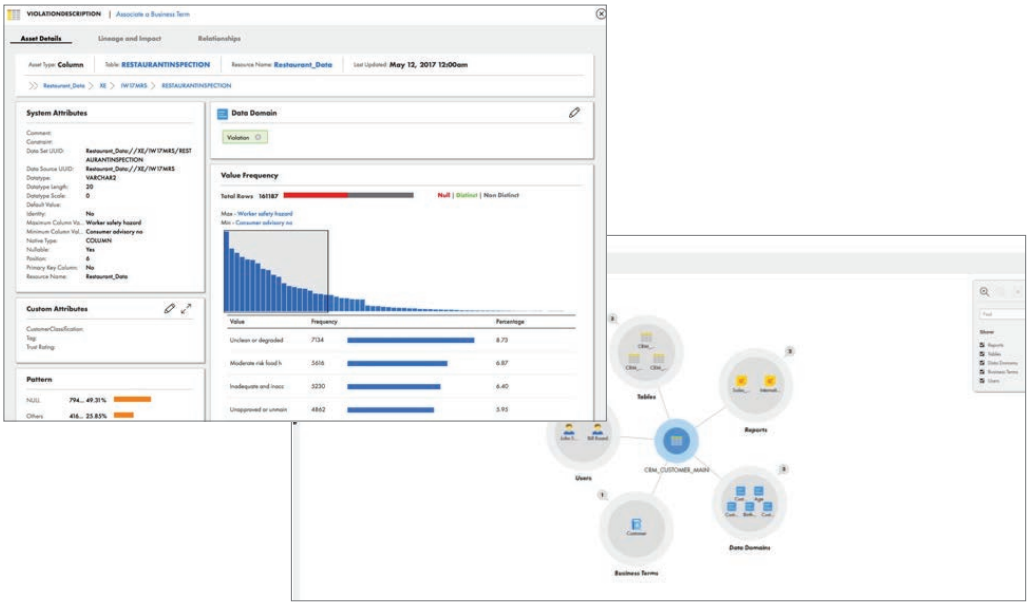

Figura 5: Clasificación automática de los datos.

Detección inteligente de entidades

Una vez identificados los dominios de las columnas, CLAIRE ensambla esos campos individuales en entidades de negocio de carácter más general. En el ejemplo siguiente se muestra la creación de una entidad denominada Orden de compra combinando los campos identificados como Cliente y Producto. La detección de entidades aprende el método empleado por los usuarios para ensamblar campos de datos dispares en sus procesos de análisis o integración de datos y aplica esos conocimientos para derivar entidades a partir de todos los datos empresariales.

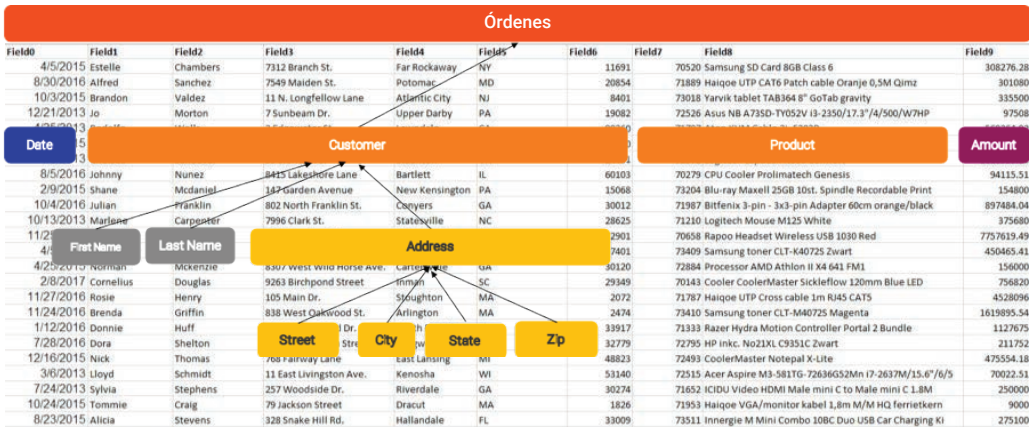

Figura 6: Combinación de dominios de datos para detectar entidades a partir de tablas y archivos.

CLAIRE para análisis

La inteligencia y la automatización impulsadas por CLAIRE aceleran de manera significativa los procesos y la información analíticos, aumentan la disponibilidad de datos y optimizan la preparación de los datos para los análisis. CLAIRE mejora la productividad de los procesos de ingeniería de datos con recomendaciones de canales de datos y la capacidad de analizar automáticamente datos complejos y multiestructurados.

Recomendaciones para la transformación

Cierre el bucle de diseño y mejore la productividad de los ingenieros de datos con la creación automatizada de mappings de integración de datos con predicciones para la próxima transformación y expresiones. Cuando una organización opta por recibir recomendaciones basadas en CLAIRE, se analizan los metadatos anónimos de los canales de datos de la organización y se aplica la inteligencia artificial y el aprendizaje automático para ofrecer recomendaciones de diseño. Estos metadatos se utilizan para generar recomendaciones de transformación y expresión. CLAIRE mejora con cada uso, cada vez que se acepta o rechaza una recomendación. Esto acelera el desarrollo, automatiza tareas repetitivas y permite a más tipos de usuarios conectar e integrar datos rápidamente.

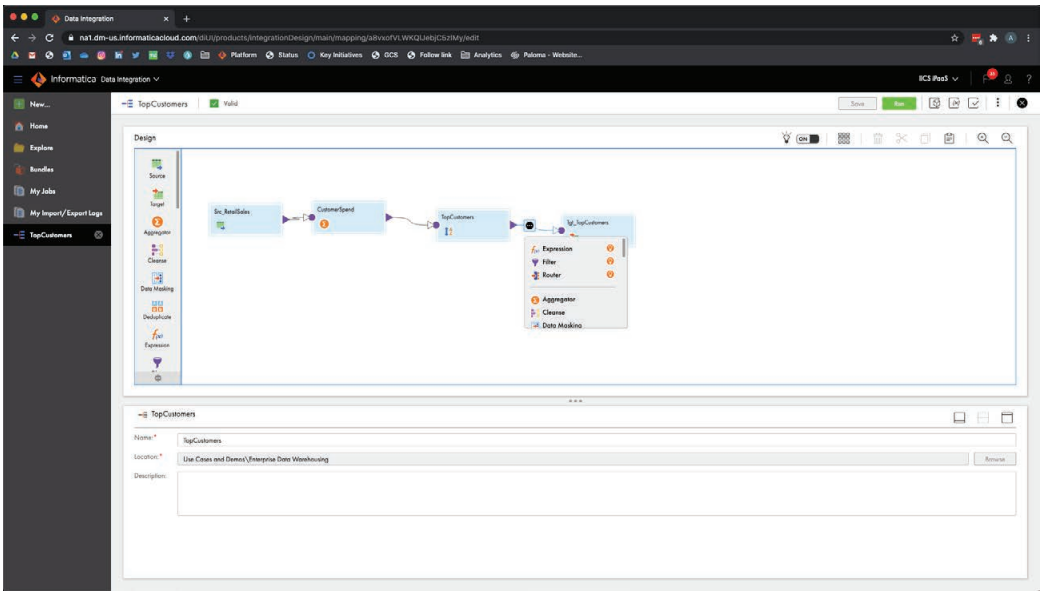

Figura 7: CLAIRE recomienda las próximas transformaciones al crear canales de datos.

Ejecución optimizada de procesos a escala

CLAIRE aplica diversos métodos de optimización para aumentar el rendimiento de la integración en todo el canal de datos. Un optimizador inteligente decide cuál es el mejor motor de procesamiento para ejecutar una carga de trabajo de Big Data en función de las características de rendimiento; se presentan recomendaciones de asignación a los ingenieros de datos según las actividades anteriores del usuario, y un optimizador basado en el coste junto con la heurística cambia de forma inteligente el orden de unión de un canal de datos con el fin de optimizar el rendimiento. Estos son solo algunos ejemplos de cómo CLAIRE optimiza los canales de datos.

Recomendaciones para unir columnas

CLAIRE sugiere automáticamente unir columnas (claves de unión) cuando un usuario selecciona la acción de combinar dos conjuntos de datos. De esta forma, ahorra cientos de horas de esfuerzo manual a los analistas de datos para intentar determinar la mejor manera de fusionar conjuntos de datos en uno compuesto para su análisis. CLAIRE comienza por las relaciones de clave principal/externa (Pk-Fk) definidas en los sistemas de origen originales (p. ej., bases de datos relacionales como Oracle) de los conjuntos de datos que se importaron al data lake. Si se unen los mismos conjuntos de datos en otros proyectos, la información de esta columna de unión también se utilizará para hacer recomendaciones. CLAIRE procesa y clasifica toda esta información para sugerir las mejores columnas para unir entre dos conjuntos de datos. Además, según un muestreo de los conjuntos de datos, también se muestra el porcentaje de superposición de datos entre las columnas sugeridas.

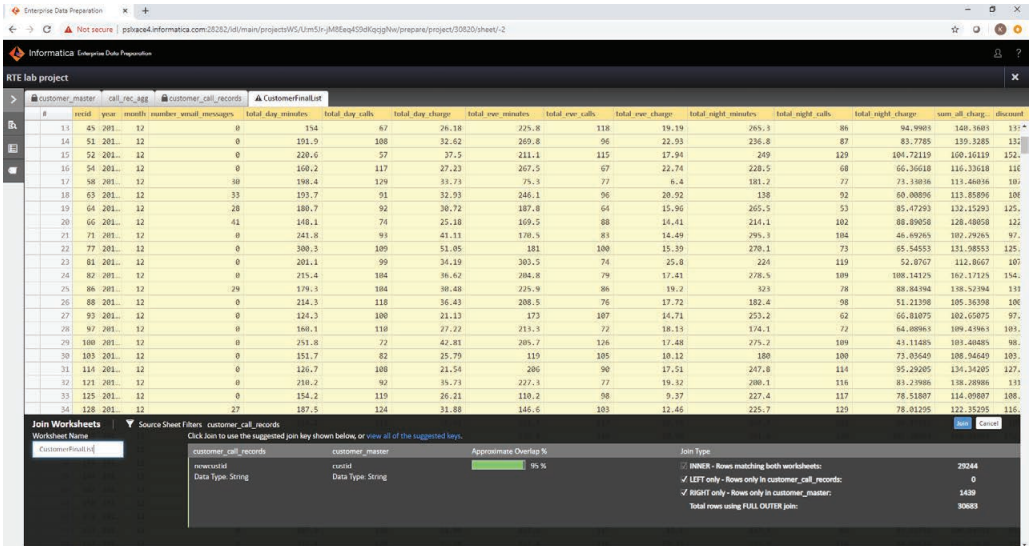

Figura 8: Sugerencias automáticas de unión de columnas al combinar dos conjuntos de datos.

Recomendaciones de visualización de Apache Zeppelin

Informatica Enterprise Data Preparation utiliza Apache Zeppelin para ver las hojas de trabajo en forma de bloc de notas con gráficos y tablas. Cuando el usuario abre el bloc de notas de una publicación, puede ver las recomendaciones de visualización de CLAIRE. Cuando el usuario abre el bloc de notas por primera vez después de su publicación, ve visualizaciones del histograma de columnas numéricas derivadas. Si la publicación no contiene columnas numéricas derivadas, el usuario ve una consulta de tabla "SELECT * FROM" en el primer párrafo del bloc de notas.

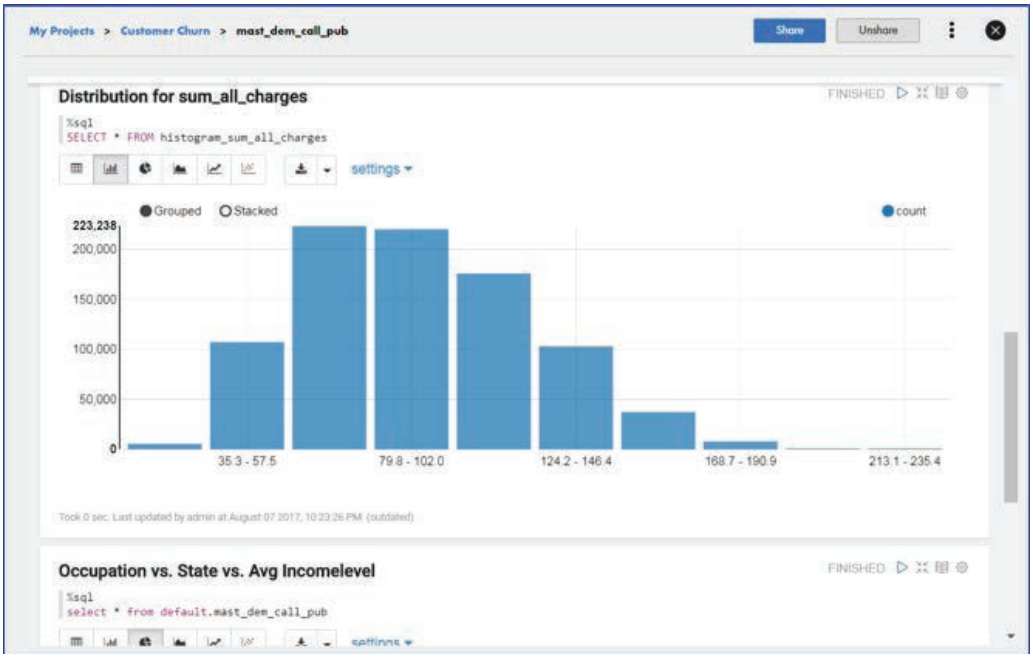

Figura 9: Visualización recomendada en el bloc de notas de Apache Zeppelin.

Recomendaciones de datos inteligentes

CLAIRE ofrece a los analistas de datos y a los especialistas en datos sugerencias sobre los conjuntos de datos que les conviene usar en sus proyectos. Analiza los conjuntos de datos seleccionados por los usuarios y propone conjuntos parecidos, pertinentes o complementarios. Gracias a las recomendaciones de datos inteligentes, los usuarios se ahorran repetir el mismo trabajo que quizá hayan realizado colegas con anterioridad. Las recomendaciones incluyen lo siguiente:

- Una versión preparada de los datos idénticos (datos sustituibles)
- Otra tabla que contiene los registros del mismo tipo (datos susceptibles de unión)
- Una tabla susceptible de incorporación a fin de enriquecer los datos con atributos adicionales (datos susceptibles de incorporación)

En las recomendaciones de datos, se emplean técnicas de filtrado basadas en el contenido para ofrecer sugerencias sobre conjuntos de datos adicionales. Se tienen en consideración varias características (condiciones), por ejemplo, el linaje, la calificación de los usuarios y la semejanza de los datos. Para puntuar la equivalencia de los distintos conjuntos de datos, se tienen en cuenta varias medidas de semejanza. Según la puntuación obtenida, se recomiendan conjuntos de datos con propiedades similares. Para recomendar elementos complementarios, se consulta el gráfico de metadatos en busca de conjuntos de datos que usuarios diferentes suelan utilizar juntos.

Detección inteligente de estructuras

Se genera y recopila una cantidad cada vez mayor de datos en multitud de máquinas, empresas y aplicaciones, en formato no estructurado o no relacional. Estos tipos de datos se caracterizan no solo por los grandes volúmenes, sino también por su velocidad, variedad y variabilidad.
El "desfase de datos" es un término que ahora se utiliza comúnmente para describir la fluctuación en el formato, el ritmo y el contenido de los datos en estos nuevos tipos de datos.

La tecnología Informatica Intelligent Structure Discovery (ISD) basada en el motor CLAIRE se ha diseñado para automatizar el proceso de incorporación de archivos de modo que las empresas puedan detectar y analizar archivos complejos. ISD proporciona compatibilidad de serie con una variedad de formatos de archivo de datos, como los flujos de clics, el registro de IoT, CSV, delimitado por texto, XML, JSON, Excel, ORC, Parquet, Avro, PDF y los archivos de tablas de Word. CLAIRE puede obtener automáticamente la estructura de estos archivos, lo que facilita su comprensión y el trabajo con ellos. Al analizar los archivos con un enfoque basado en el contenido, se adapta a los cambios frecuentes en los archivos sin que el procesamiento de estos se vea afectado.

ISD aplica un algoritmo genético para automatizar el reconocimiento de patrones en los archivos. En este enfoque, se emplea el concepto de "evolución" para mejorar los resultados. Cada candidato propuesto posee una serie de propiedades que se pueden modificar automáticamente y comprobar posteriormente para determinar si se trata de la solución más acertada. Las estructuras resultantes se puntúan conforme a distintos factores, como la cobertura de entrada y los dominios derivados. Las estructuras que reciben la mayor puntuación entran entonces en una fase de "mutación" durante la cual se les aplican varios cambios, por ejemplo, combinar subestructuras para averiguar si sube su puntuación. El proceso finaliza cuando se confirma la idoneidad de la estructura para los datos.

ISD también emplea mecanismos personalizados de reconocimiento de entidades con nombre (NER) y comprensión del lenguaje natural (NLU) con tecnología ML para identificar campos y tipos de campo, lo que simplifica las integraciones y permite que las aplicaciones externas utilicen ISD como un motor de NER/NLU subyacente. Por ejemplo, ISD se utiliza para detectar información PII en la carga útil de API entrante y saliente, y facilita el cumplimiento de las normativas, además de fortalecer la seguridad de la empresa. ISD también se utiliza en casos de uso de detección, incorporación y transmisión de datos.

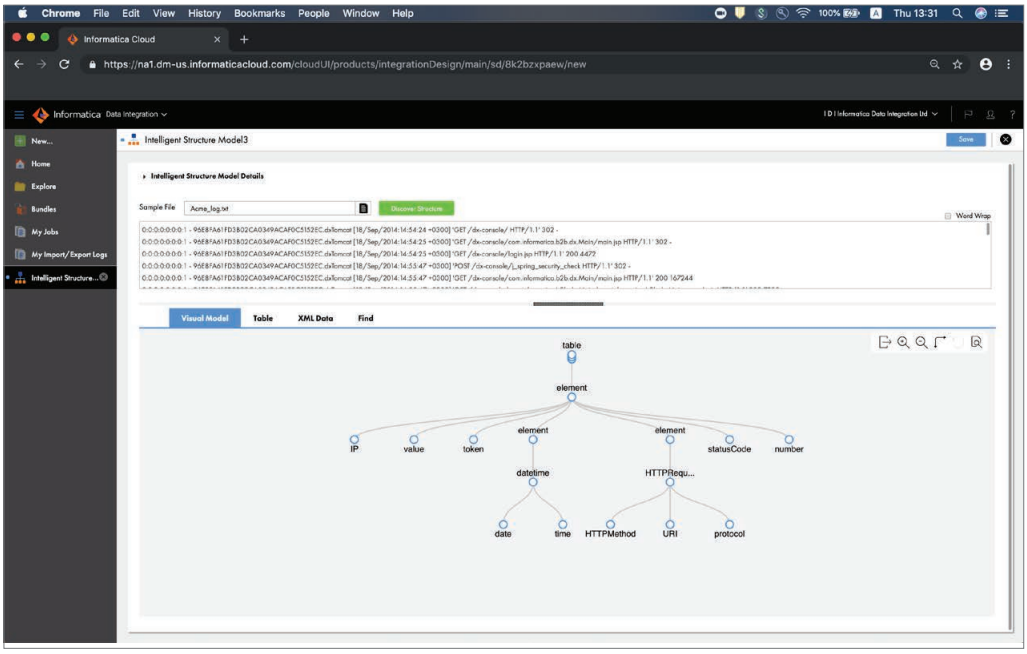

Figura 10: Búsqueda inteligente de estructuras en archivos de datos no estructurados.

CLAIRE para la gestión de datos maestros

La inteligencia y la automatización basada en el motor CLAIRE que utiliza tecnología avanzada de IA y ML enriquecen y mejoran la precisión de la visión integral del negocio para clientes, productos, proveedores y otros dominios. Se emplean diversas técnicas combinadas de IA/ML que van desde algoritmos deterministas, heurísticos y probabilísticos hasta la correspondencia de síntesis contextual y la correspondencia de entidades de aprendizaje activa para que la correspondencia de registros sea rápida, escalable y muy precisa, y para el enriquecimiento de los datos maestros.

Correspondencia de síntesis

La síntesis es una técnica de correspondencia de última generación que aborda, por ejemplo, la correspondencia entre clientes potenciales y clientes existentes, y la correspondencia de interacciones y datos no estructurados con los clientes, además de detectar relaciones no obvias. Para ello, utiliza "atributos contextuales", aprendizaje automático, NLP y una combinación de correspondencia basada en probabilidades y reglas declarativas.

Los atributos demográficos (como el nombre, la dirección y el número de teléfono), los atributos de interacción (correo electrónico, chats web, etc.) y los atributos contextuales (cuándo, qué, dónde, quién) son muy útiles a la hora de establecer correspondencias entre los datos relacionados con el cliente con un nivel de confianza determinado. NLP puede extraer los "atributos contextuales" de texto no estructurado para proporcionar muchos más puntos de datos que se utilizarán en el proceso de correspondencia. Un algoritmo de ML puede ser muy eficaz para la correspondencia cuando se aplica un enfoque de entrenamiento supervisado en el que los administradores de datos y expertos en la materia etiquetan un conjunto de pares coincidentes seleccionado correctamente como coincidencias o no coincidencias. Estos pares etiquetados forman un conjunto de entrenamiento que se utiliza para generar un algoritmo de correspondencia.

La síntesis genera una visión integral del cliente que combina datos demográficos, de cuenta, de transacción, de interacción y no estructurados. Los algoritmos de correspondencia tradicionales combinan registros para proporcionar una perspectiva única del cliente, mientras que la correspondencia de síntesis muestra todos los datos de los clientes en un gráfico. Los datos se relacionan entre sí con los niveles de confianza, por lo que es posible ofrecer varias perspectivas de un cliente.

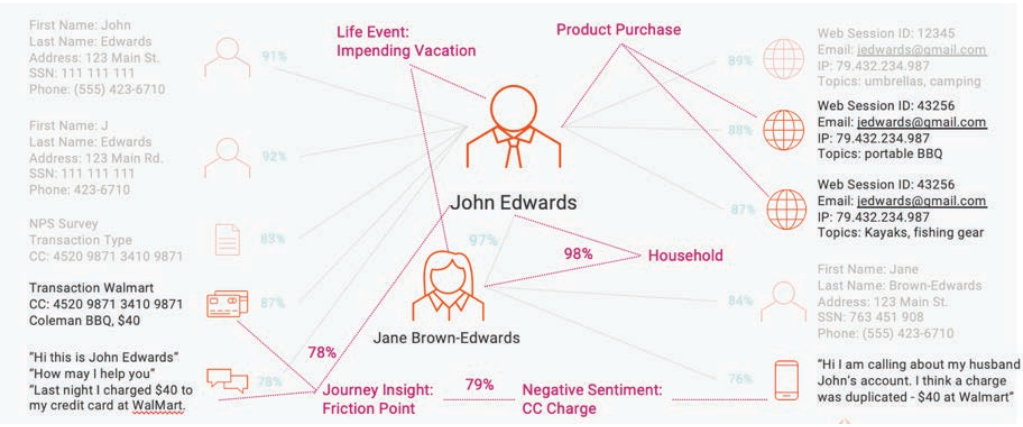

Figura 11: La correspondencia de síntesis y el razonamiento generan información útil que se almacena como parte de Customer 360.

Correspondencia de identidades
La correspondencia de identidades NAME3 de CLAIRE condensa más de 30 años de entrenamiento y ajuste mediante una variedad de técnicas como la generación inteligente de claves para fines de indexación y bloqueo, la estabilización de texto semántica y la comparación de datos de las partes y de ubicación, reglas de estabilización de texto y edición de listas para 80 grupos de población, y ponderación inteligente de la importancia de las funciones para diferentes propósitos. Estas poderosas técnicas permiten la indexación y el bloqueo en varios campos, reglas definidas por el cliente de correspondencia y anticorrespondencia, y reglas de correspondencia y anticorrespondencia definidas por la implementación que complementan otras reglas de IA.

Correspondencia de entidades
La correspondencia de entidades encuentra registros de datos que se refieren a la misma entidad del mundo real (por ejemplo, clientes, productos, etc.). Los registros de datos pueden ser no estructurados (por ejemplo, la información del cliente oculta en un chat web) y estructurados. La clasificación de correspondencia compara un par coincidente y determina si hay realmente una correspondencia, una posible correspondencia o si no existe correspondencia con un nivel de confianza. Hay técnicas que utilizan reglas configuradas por humanos (es decir, reglas declarativas) o reglas de IA (es decir, una configuración obtenida por aprendizaje automático). Los mejores resultados se obtienen al combinar estas dos técnicas.

Las reglas declarativas, creadas por expertos en la materia, complementan las potentes reglas de IA que CLAIRE aplica en forma de un clasificador de bosque aleatorio aprendido. CLAIRE utiliza el aprendizaje activo supervisado (en contraposición con el aprendizaje de fuentes colectivas o de múltiples usuarios) para acelerar el proceso de entrenamiento de IA, en el que se presentan microlotes de 10 o 20 pares coincidentes para etiquetar a un usuario (es decir, como correspondencia, posible correspondencia o no correspondencia). Una vez etiquetados, CLAIRE vuelve a entrenar el clasificador de bosque aleatorio y determina cuáles son los próximos mejores pares coincidentes para presentarlos a un usuario en este proceso de etiquetado iterativo. También utiliza los pares etiquetados para inferir las reglas de bloqueo (es decir, para eliminar falsas correspondencias), realizar bloqueos, entrenar modelos y llevar a cabo la correspondencia de entidades.

CLAIRE aplica una combinación de similitudes/comparaciones de cadenas como Jaccard, reglas declarativas derivadas del perfilado de datos, conjuntos de datos estabilizados (archivos de población, apodos, comparaciones semánticas, etc.) y reglas definidas por el usuario que gestionan las excepciones. Estas reglas declarativas abordan las brechas y excepciones, y ayudan a acelerar el proceso de entrenamiento de aprendizaje activo (es decir, reducen la cantidad de pares coincidentes necesarios para el aprendizaje), agilizar la creación de funciones de reglas de IA y mejorar la precisión de las correspondencias. Por ejemplo, cada vez que el nombre, la fecha de nacimiento y el SSN tienen un alto nivel de similitud, la regla lo reconoce como una coincidencia. Esta combinación de reglas declarativas y de IA acelera el proceso de entrenamiento y mejora la precisión de las correspondencias.

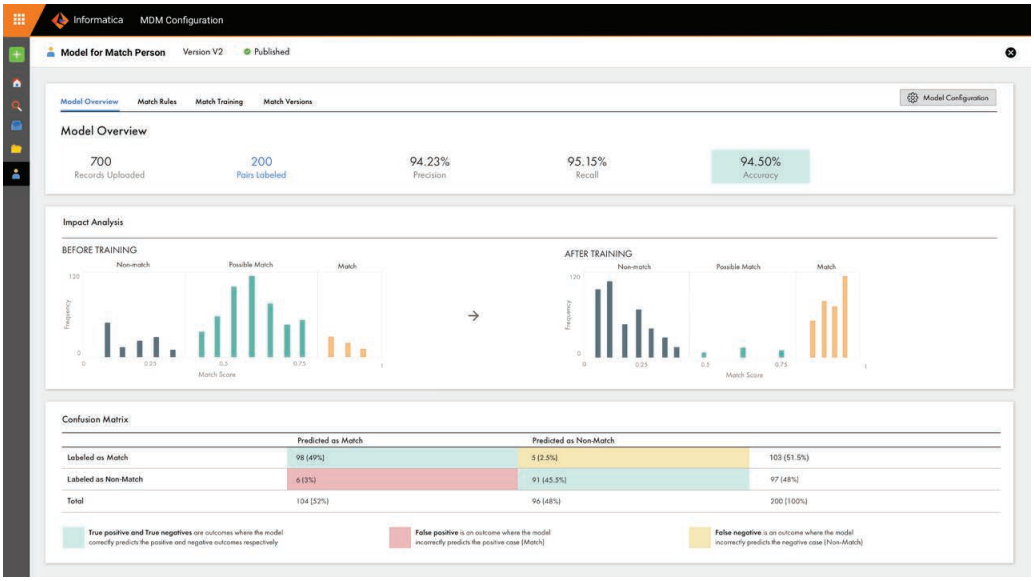

Figura 12: Correspondencia de entidades

CLAIRE para fines de gobierno de datos y cumplimiento

La inteligencia artificial y el aprendizaje automático son esenciales para automatizar las tareas de gobierno de datos más exigentes hoy en día: encontrar datos, medir su calidad y permitir la colaboración para facilitar su gobierno. CLAIRE genera automáticamente reglas basadas en políticas (por ejemplo, de calidad de datos) y vincula la semántica del negocio con metadatos técnicos, además de dirigir a los usuarios a los datos más relevantes y fiables en función de sus necesidades empresariales.

Enriquecimiento automático de la calidad de los datos

CLAIRE aplica un enfoque de NLP basado en Stanford NER para analizar y extraer entidades de texto no estructurado. Por lo general, para extraer entidades de cadenas (por ejemplo, el código de un producto), los usuarios terminan creando reglas de análisis utilizando tablas de referencia y expresiones regulares. El volumen de datos, la complejidad y los patrones crecen continuamente; crear todas las reglas posibles para establecer correspondencias con cada nueva entrada no es práctico ni escalable. En su lugar, CLAIRE utiliza modelos entrenados previamente para identificar y extraer entidades basándose en Stanford NER.

CLAIRE utiliza el aprendizaje automático para clasificar el texto entrante, por ejemplo: el idioma, el tipo de producto y el caso de soporte técnico. La metodología de aprendizaje automático utilizada se conoce como aprendizaje supervisado con modelos Naive-Bayes y Max Entropy (regresión logística multinomial). El aprendizaje supervisado se utiliza para entrenar modelos y asignar etiquetas. Posteriormente, el modelo entrenado se puede implementar durante el procesamiento de datos para etiquetar, dirigir y procesar diferentes clases de información, por ejemplo, para tratar "problemas del motor" independientemente de los "problemas de configuración" con significados similares y distinguir entre usos de palabras con varios significados. CLAIRE automatiza el etiquetado y la clasificación de imágenes utilizando modelos de NLP y ML para la clasificación de productos y la extracción de metaetiquetas de imagen.

Una gran empresa mundial del sector de la sanidad asignó un empleado a tiempo completo para asociar 21 000 activos técnicos con 6000 términos empresariales, un proceso que duró dos meses. Con Axon Data Governance y Enterprise Data Catalog, CLAIRE automatizó la asignación de 18 000 activos técnicos con una precisión del 99 % en 8 minutos.

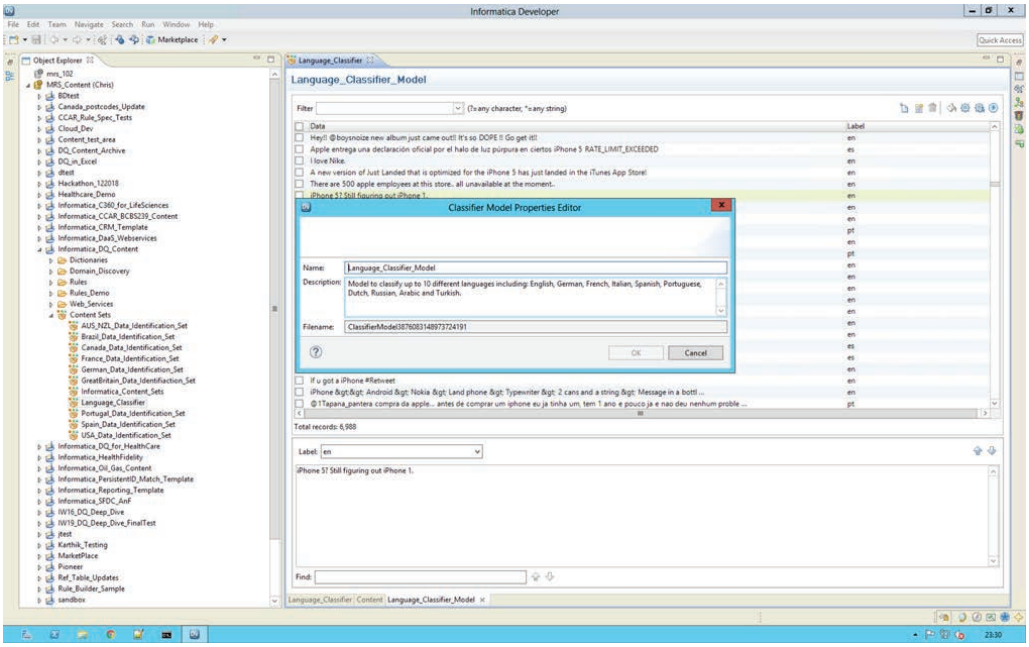

Figura 13: El NLP de aprendizaje automático clasifica el texto y extrae entidades.

Asociación automática de términos empresariales con conjuntos de datos físicos

El gobierno de datos requiere la documentación de artefactos empresariales, definiciones, partes interesadas, procesos y políticas, entre otros. Para obtener una visión verdaderamente alineada, es esencial que los usuarios puedan asociar definiciones y visiones empresariales a las implementaciones técnicas subyacentes en todos sus datos. Por lo general, esta tarea es lenta, laboriosa y propensa a errores, ya que depende de personas clave para comunicarse y alinear manualmente manifestaciones técnicas una por una, tarea que puede tardar días, semanas o incluso meses en completarse.

Informatica Axon Data Governance, mediante una estrecha integración con Informatica Enterprise Data Catalog, puede acortar el proceso. CLAIRE proporciona a los usuarios recomendaciones de elementos de datos pertinentes y apropiados para vincularlos a medida que se completan los análisis de metadatos. Esto agiliza la tarea de buscar, validar y vincular elementos de datos, lo que permite a los administradores de datos y al departamento encargado del gobierno de datos centrarse en tareas más críticas. A medida que progresan las implementaciones, el proceso se puede automatizar completamente.

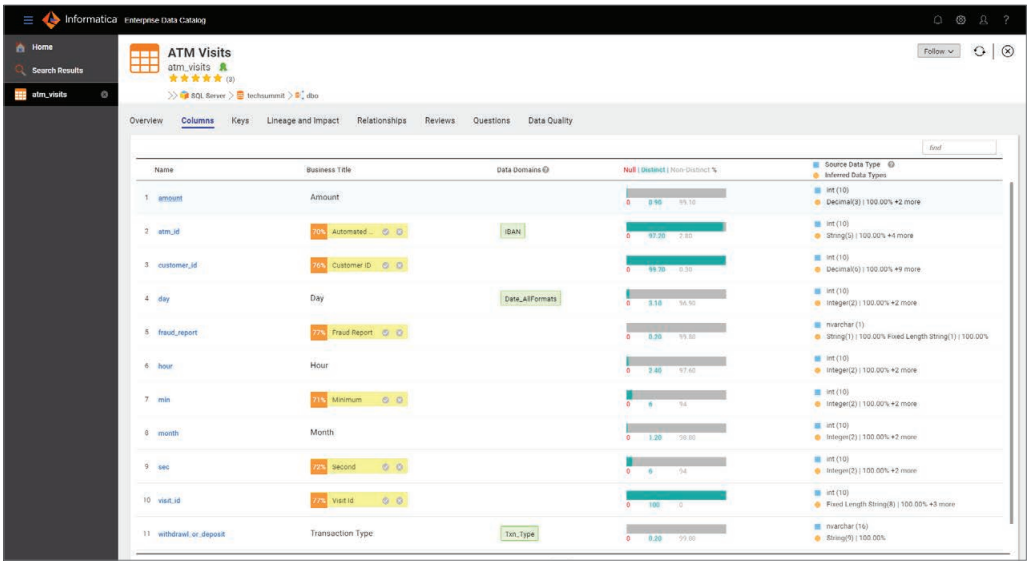

Figura 14: Asociación automática de términos empresariales con conjuntos de datos físicos.

Evaluación automática de la calidad de los datos

Un indicador clave del rendimiento (KPI) en el ámbito del gobierno de datos es la calidad de los datos de un sistema que respalda un proceso, sustenta políticas, etc. Los responsables del gobierno de datos deben garantizar que estos sean completos, precisos, coherentes y válidos, entre otras cosas. En resumen, deben ser lo suficientemente buenos y fiables como para respaldar las operaciones empresariales. A medida que aumentan las implementaciones de gobierno de datos, la evaluación de la calidad de un número creciente de sistemas y campos del panorama de los datos, desde bases de datos a data lakes, se convierte en una tarea que consume cada vez más tiempo.

A través de CLAIRE, Axon Data Governance, en coordinación con Informatica Data Quality e Informatica Enterprise Data Catalog, puede automatizar la aplicación de análisis de la calidad de los datos en toda la empresa, lo que supone un ahorro de miles de horas de trabajo. El equipo de gobierno de datos vincula reglas de calidad de datos para varias dimensiones de calidad de datos a términos empresariales y elementos de datos críticos. A continuación, el sistema subyacente genera las comprobaciones de datos necesarias en los distintos sistemas e informa de los resultados al departamento de gobierno.

Esta automatización es posible gracias a la combinación de tres elementos de información clave:

1. Conocimiento de los elementos empresariales esenciales y de las reglas de calidad de datos que Axon requiere
2. Reglas de calidad de datos portátiles y ejecutables, así como un motor de ejecución flexible de Informatica Data Quality
3. Detalles de metadatos de activos de datos físicos de Enterprise Data Catalog

CLAIRE combina esta información para generar trabajos de ejecución de reglas de calidad de datos en Informatica Data Quality en función de los activos de datos físicos de Enterprise Data Catalog. CLAIRE también mantiene el contexto de usuario empresarial de Axon para garantizar que los resultados se muestren en los paneles correctos y en visiones agregadas para su uso por parte del departamento de gobierno de datos.

La automatización permite escalar los programas de gobierno más rápido que nunca, eliminando así miles de horas de trabajo manual para crear evaluaciones de calidad de datos y vincularlas una por una al contexto de gobierno. CLAIRE también garantiza la evaluación automática de la calidad de cualquier activo físico nuevo identificado. Además, se detectan nuevos dominios utilizando un clasificador o la extracción de entidades con nombre en reglas de calidad de datos.

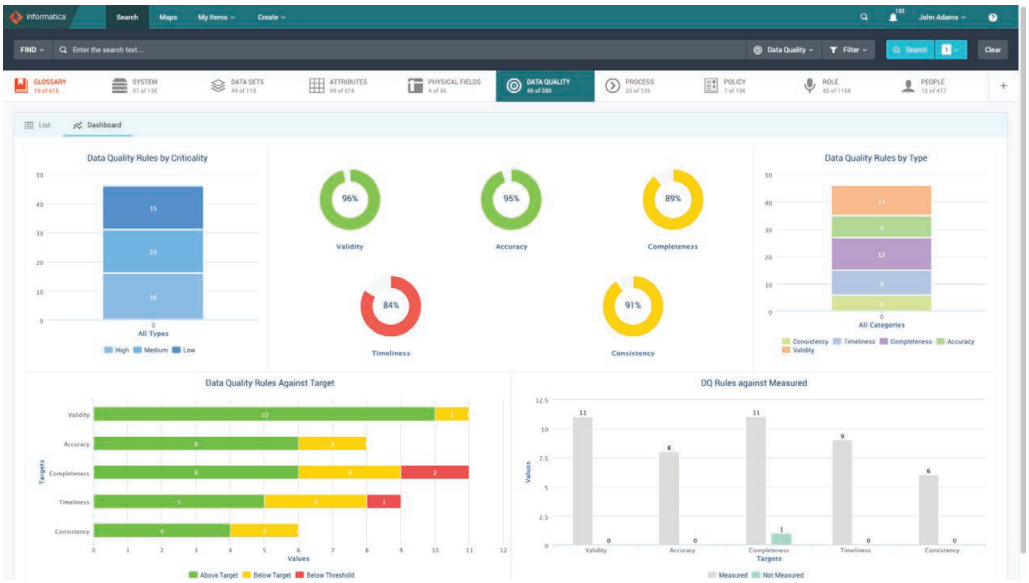

Figura 15: Las evaluaciones automáticas de calidad de todos los datos de la empresa ahorran miles de horas de trabajo manual.

Identificación y regla de calidad de datos asistida por NLP y ML

La calidad de los datos es un imperativo clave para cualquier programa de gobierno de datos pero, en las implementaciones de mayor tamaño, puede haber muchas reglas de calidad de datos. Para ayudar a los administradores de datos a identificar las reglas correctas que se deben utilizar, CLAIRE no solo les facilita tal identificación, sino que genera las reglas que faltan.
Un usuario de Axon Data Governance puede especificar su requisito de regla en texto sin formato (por ejemplo: "los identificadores de clientes deben tener ocho caracteres y comenzar por la letra C") y solicitar la ayuda de CLAIRE. Mediante técnicas de ML y NLP, CLAIRE analiza los requisitos de los usuarios y los convierte en representaciones técnicas. Basándose en esa representación, así como en los metadatos asociados (por ejemplo: nombre de términos de glosario), CLAIRE busca en las reglas de Informatica Data Quality e identifica posibles candidatas. Después, el usuario podrá elegir entre una regla coincidente ya existente o bien (si ninguna es adecuada) solicitar a CLAIRE que genere una nueva regla de calidad de datos.

Si no se encuentra ninguna regla aplicable, CLAIRE generará automáticamente una nueva regla de calidad de datos para satisfacer el requisito en el repositorio de Informatica Data Quality y la vincularla al contexto de Axon Data Governance. Además, CLAIRE asocia automáticamente reglas de calidad de datos a perfiles de cloud basados en fuentes de Salesforce y Microsoft Common Data Model (CDM). A medida que los usuarios creen nuevos perfiles basándose en objetos principales de una de estas fuentes, CLAIRE sugerirá automáticamente las reglas de calidad de datos que considere óptimas para aplicar a la evaluación.

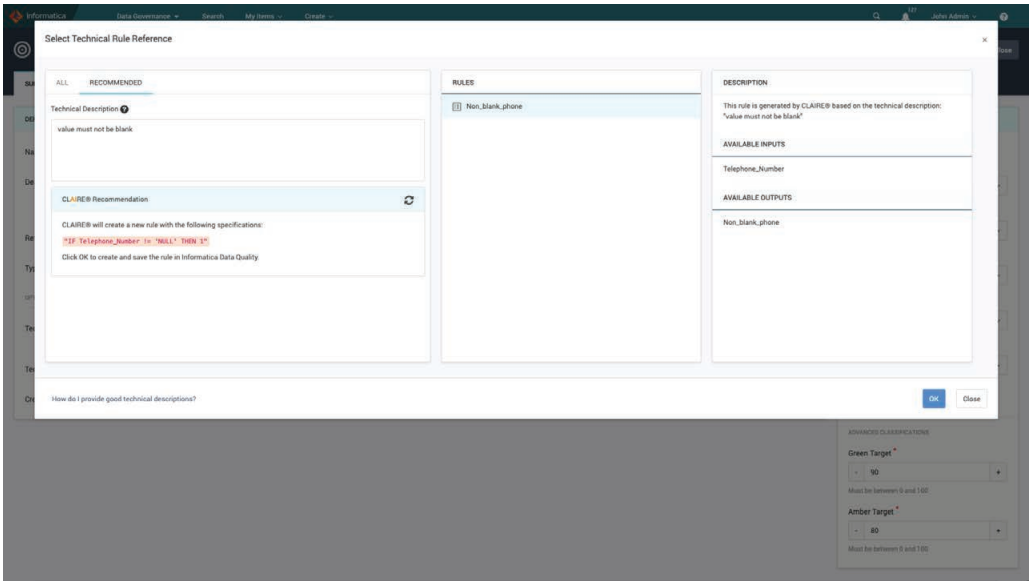

Figura 16: Identificación automática de reglas de calidad de datos con NLP.

CLAIRE para la protección y privacidad de los datos

Con soluciones inteligentes de privacidad de datos respaldadas por CLAIRE, las organizaciones pueden obtener una visión completa de toda la empresa y realizar un análisis de la información personalmente identificable (PII) incluida en sus activos de datos. La automatización basada en IA permite identificar datos personales y sensibles, comprender el movimiento de los datos, vincular identidades, analizar los riesgos y solucionar problemas.

Asignación de identidades del registro de usuarios

CLAIRE determina la correlación de identidades con los datos sensibles que proporciona el mapeo de datos para la creación de informes sobre el acceso de los usuarios de los datos y el cumplimiento de los requisitos de privacidad. CLAIRE evalúa y clasifica los datos que, combinados, pueden servir para identificar a los sujetos. Además de la correspondencia exacta, se utilizan varias técnicas avanzadas, incluido el reconocimiento de entidades con nombre (NER), para mejorar los resultados que suelen obtenerse cuando se combinan los datos de diversas fuentes.

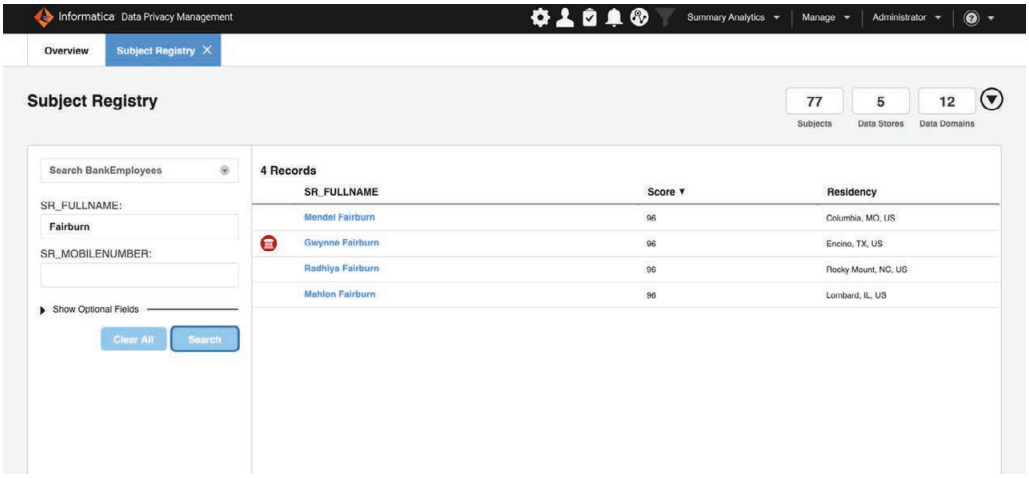

Figura 17: Asignación de identidades en el registro de usuarios para crear informes sobre el acceso de los usuarios de los datos y el cumplimiento de los requisitos de privacidad.

Movimiento y asignación de datos sensibles

CLAIRE aprovecha y amplía las capacidades de linaje mencionadas anteriormente para identificar también cómo proliferan los datos sensibles en los repositorios a fin de respaldar el cumplimiento de los requisitos de seguridad y privacidad. CLAIRE determina el movimiento ascendente y descendente, así como los metadatos relacionados, como el tipo de datos específico, el proceso, el estado de protección y la ubicación de los datos, con el fin de evaluar si se han producido infracciones. Por ejemplo, podría ocurrir una infracción si los datos personales se trasladan de su origen a un destino más allá de los límites geográficos, o si los datos incorporados para procesos de facturación se trasladan a otros departamentos o ubicaciones para utilizarlos en procesos de marketing que puedan infringir las normas de privacidad. Las infracciones se pueden notificar a los responsables de políticas o procesos para su corrección.

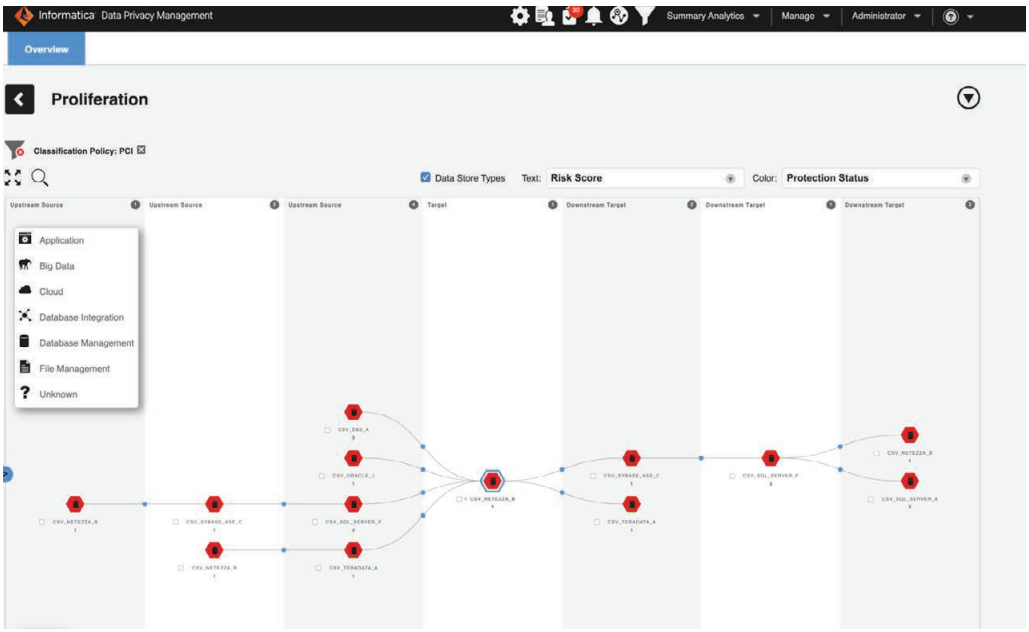

Figura 18: Identificación y seguimiento del movimiento de datos sensibles en los repositorios.

Planes de simulación de riesgo

Las normativas sobre privacidad exigen cada vez más que las organizaciones cuenten con planes de protección de datos. CLAIRE puede ayudar a las empresas a simular el impacto de estos planes de protección con el fin de garantizar un mayor retorno de la inversión y facilitar los procesos presupuestarios. CLAIRE evalúa las técnicas de protección aplicadas a uno o más dominios de datos y después calcula el cambio en la puntuación de riesgo, la exposición de los datos sensibles y el coste del riesgo residual para cada uno de los almacenes de datos seleccionados, así como el impacto total para la organización del uso de un modelo de utilidad esperado.

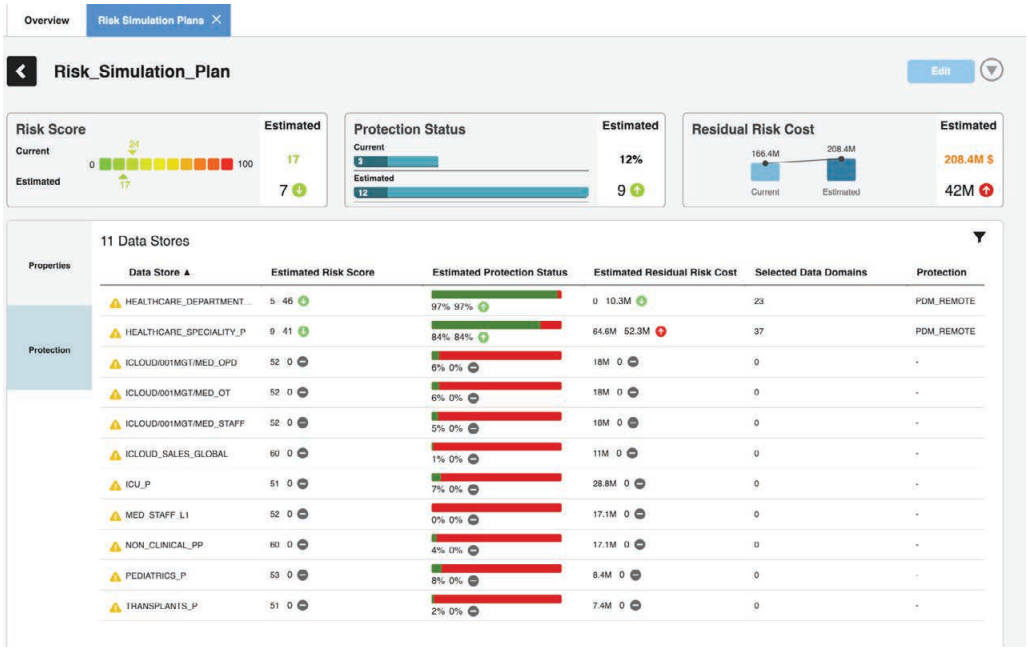

Figura 19: CLAIRE evalúa las técnicas de protección aplicadas a los dominios de datos para determinar el nivel de riesgo.

Detección inteligente de anomalías

CLAIRE aplica dos enfoques, de estadística y de aprendizaje automatizado, para detectar los datos atípicos y las anomalías en los datos. La funcionalidad de análisis de comportamiento de los usuarios detecta los patrones de conducta de los usuarios que pueden poner en riesgo la organización por favorecer el uso indebido de los datos. El análisis de comportamiento de los usuarios detecta los ataques de suplantación, apropiación de credenciales y escalado de privilegios.

El análisis del comportamiento de los usuarios aplica el aprendizaje automatizado sin supervisión a un modelo multidimensional de actividades del usuario, el cual incluye el número de almacenes de datos a los que accede, el número de solicitudes que presenta y el número de registros afectados en los distintos sistemas. A este modelo se le aplica el análisis de componentes principales para reducir la dimensionalidad. La técnica de reducción y agrupación iterativas equilibradas mediante jerarquías se aplica a la agrupación jerárquica en clústeres sin supervisión con el fin de hallar usuarios cuya conducta haya variado durante un período determinado. Para validar las conductas anómalas, se emplean métodos de detección de datos atípicos basados en la distancia y la densidad y, para confirmar que los objetos señalados con los dos primeros métodos son realmente datos atípicos en el sistema del clúster, se ejecuta la prueba estadística de Grubbs.

Figura 20: Análisis del comportamiento de los usuarios para detectar automáticamente anomalías en los usuarios que puedan indicar un uso indebido de los datos.

Protección de datos de API en tiempo real

Proteja los datos sensibles (p. ej., PII) en tiempo real mediante la identificación de filtraciones de datos personales en las API, el bloqueo y el enmascaramiento de datos. La solución de gestión de API de Informatica incorpora bibliotecas de protección de datos para bloquear datos sensibles en llamadas API entrantes y salientes, lo que minimiza el riesgo de exposición de los datos sensibles.

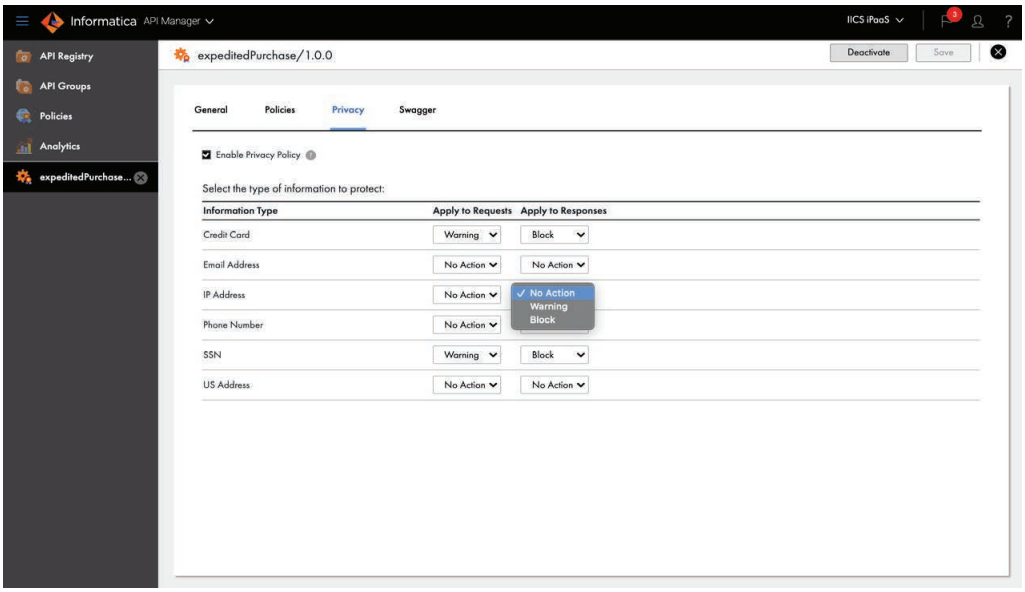

Figura 21: Bloqueo del acceso a datos sensibles en llamadas API entrantes y salientes.

CLAIRE para DataOps

Con CLAIRE, las organizaciones pueden acelerar los canales de procesamiento de datos al automatizar muchos aspectos de la gestión de datos para lograr una integración y entrega continuas en relación con DataOps.

Análisis detallado y predictivo para entornos de gestión de datos

El análisis operativo ayuda a comprender el uso actual de los proyectos y recursos existentes y a planificar la capacidad necesaria en el futuro. Ofrece parámetros para la creación de modelos de contracargo (chargeback) a la vez que da soporte a varias líneas de negocio en una única plataforma de gestión de datos. Sobre la base de la observación continua de las tendencias de uso de recursos, se ofrecen proyecciones de procesamiento de volumen de datos para facilitar la planificación de la capacidad. CLAIRE lleva todo esto a un nivel superior al ofrecer capacidad de ampliación automática de los recursos de tiempo de ejecución de gestión de datos.

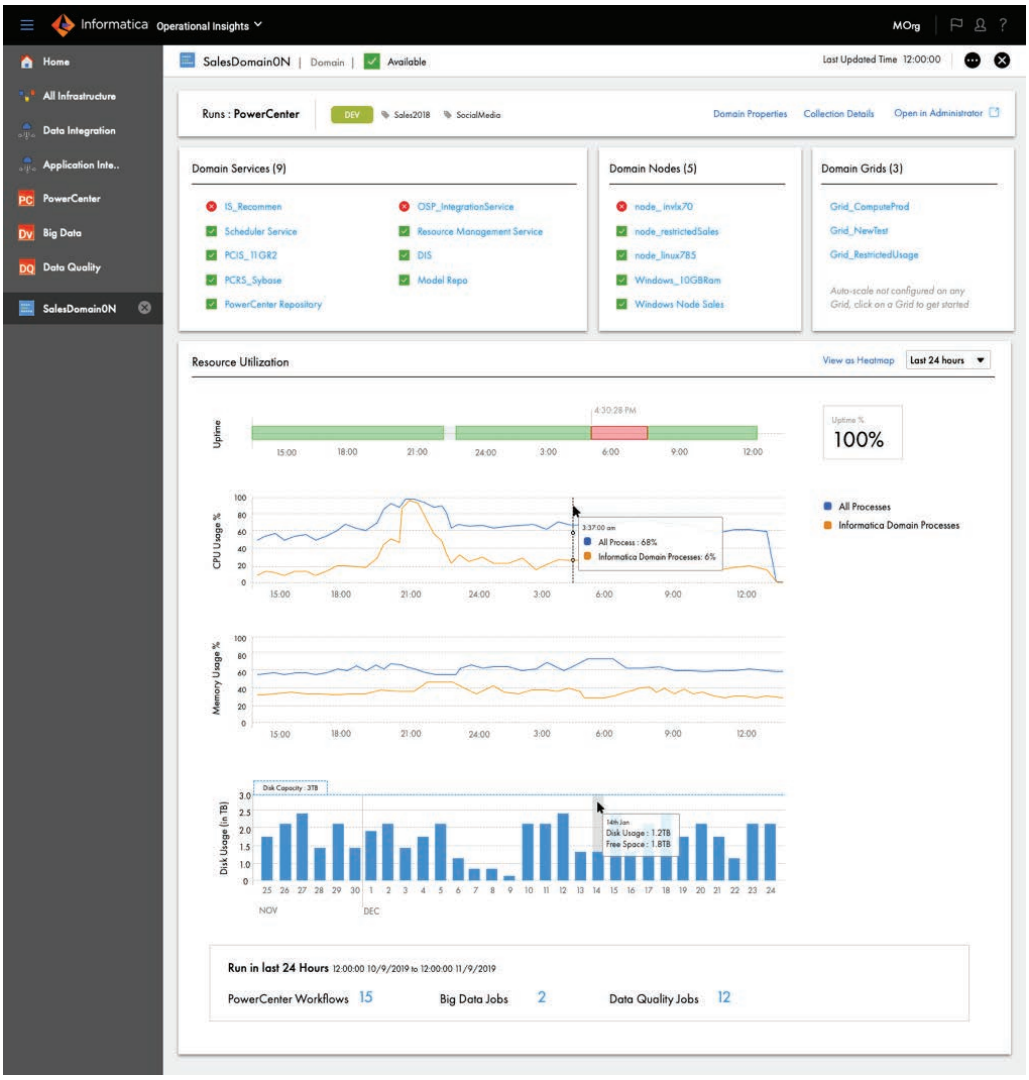

Figura 22: Uso de recursos de información operativa para procesos de dominio de Informatica.

Detección de anomalías en ejecuciones de trabajo

CLAIRE detecta automáticamente las anomalías relacionadas con los tiempos de ejecución del trabajo, los datos procesados, los datos cargados, los recursos consumidos, el rendimiento, etc. La detección automática de estas anomalías ayuda al departamento de TI a solucionar los problemas de forma proactiva en trabajos de integración de datos antes de que afecten a los procesos empresariales posteriores. Se utiliza el algoritmo Seasonal Hybrid ESD para detectar anomalías en el comportamiento de ejecución del trabajo. Este algoritmo considera la estacionalidad (carga máxima de fin de mes, temporada de vacaciones, etc.) y elimina los trabajos con anomalías esperadas inducidas por los ciclos empresariales.

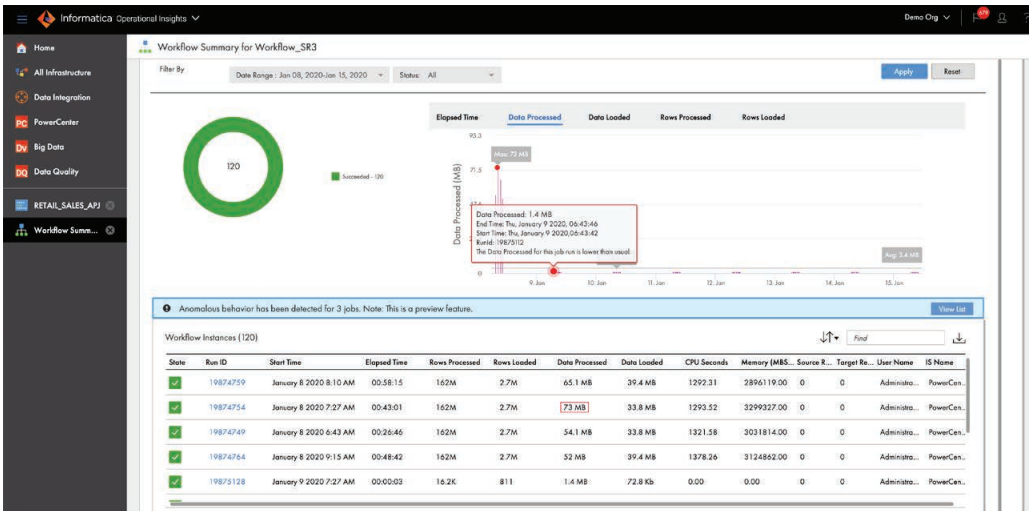

Figura 23: CLAIRE detecta automáticamente las anomalías relacionadas con trabajos de Informatica y el procesamiento de datos.

El futuro de CLAIRE

A medida que CLAIRE evolucione, seguirá aumentando la productividad y la eficiencia, permitiendo que los líderes de datos aprovechen la automatización inteligente para obtener mejor información de forma más rápida, así como para lograr una gestión de datos más eficaz. Entre las capacidades futuras se incluyen las siguientes:

1. **Integración automática:** integrar de forma automática los datos recién vertidos a los procesos de integración de datos; identificar datos, buscar patrones de integración que procesen datos semejantes, y transformar y trasferir automáticamente los datos sobre la base de los conocimientos adquiridos a partir de millones de mappings y acciones de usuarios existentes.
2. **Ayuda para el desarrollo:** ofrecer a los usuarios recomendaciones y sugerencias de acciones siguientes durante el proceso de desarrollo, por ejemplo:
   - Finalización automática de transformaciones
   - Recomendaciones de plantillas
   - Sugerencias de tipo de enmascaramiento de datos sensibles
   - Sugerencias de calidad de datos para la limpieza y la estandarización
   - Optimizaciones automáticas del rendimiento
3. **Mapping automático:** detectar entidades de datos maestros en toda la empresa y asignarlas de forma automática al modelo de datos maestros aplicando las transformaciones y las reglas de calidad necesarias

4. **Reparación automática:** gestionar sin dificultad problemas de sistema externos, como la falta de memoria o los relacionados con la potencia de procesamiento, por ejemplo, añadir potencia de procesamiento (ráfagas al cloud) para gestionar los picos de datos
5. **Ajuste automático:** predecir la programación o los recursos informáticos basándose en la información histórica, los volúmenes de datos actuales y los recursos de sistema disponibles y ajustarlos de modo que satisfagan los criterios de rendimiento
6. **Protección automática:** detectar automáticamente los datos sensibles y enmascararlos antes de que abandonen cualquier zona segura

Conclusión

Las actuales estrategias de negocio centradas en los datos se basan en los datos. Para que tengan éxito, es indispensable fomentar la competencia en la gestión de datos, de modo que se aproveche todo su potencial. En vista de todos los retos que la gestión de datos plantea en circunstancias normales, es imposible que los enfoques tradicionales satisfagan los requisitos actuales, mucho menos los futuros. Es posible utilizar los datos de forma revolucionaria estandarizándolos en una plataforma de gestión de datos de principio a fin que aproveche el potencial de los datos, los metadatos, el aprendizaje automatizado y la inteligencia artificial para aumentar la productividad de todos sus usuarios: técnicos, operacionales y de negocio y, en particular, de los usuarios de negocio de autoservicio.

Póngase en contacto con nosotros para aprender a sacar más partido a los datos con CLAIRE e Intelligent Data Management Cloud.

Informatica

Informatica Local Headquarters José Echegaray 8, Parque Alvia - Edif. 3, PB 3, 28232 Las Rozas de Madrid Tel: +34 91 787 61 40
Número gratuito en Estados Unidos: 1.800.653.3871

IN09_0521_03328