

A man with glasses and a dark blue shirt is standing at a desk in a large warehouse, looking at a computer monitor. The warehouse is filled with tall metal shelving units stacked with cardboard boxes. In the background, another person is visible working at a desk. A green metal cart is parked on the right side of the aisle. The overall scene is brightly lit with overhead industrial lights.

alteryx

# Analítica de autoservicio en Google Cloud Platform:

Cinco lecciones de preparación de datos  
para garantizar el éxito

## Deseas abordar los desafíos de la analítica empresarial por tu cuenta y ser más productivo e innovador, pero necesitas los datos para hacerlo.

En el panorama actual de datos dinámicos y complejos, las soluciones de almacenes de datos tradicionales e inteligencia empresarial ya no son suficientes. Incluso cuando accedes a almacenes de datos tradicionales con soluciones como Tableau o Qlik, igual sigues trabajando en hojas de cálculo y necesitando asistencia de TI.

### **La respuesta es una solución de analítica de autoservicio.**

En esta publicación, se describen cinco lecciones que Alteryx ha aprendido ayudando a organizaciones como la tuya a adoptar la analítica de autoservicio. Se centra en Google Cloud Platform, un conjunto de analítica inteligente e integral que ofrece a los analistas la libertad de trabajar con cualquier formato de datos, a cualquier escala o velocidad.





# Lección n.º 1


## Descartar la reubicación en la nube

Cuando se mueve la analítica a la nube, muchas organizaciones adoptan un enfoque de reubicación: las aplicaciones de analítica se instalan en una máquina virtual en la nube. Si bien este enfoque facilita el mantenimiento de la infraestructura, no mejora la forma en que se proporciona la analítica, ni aumenta sus beneficios comerciales o reduce sus costos operativos. El proceso y las herramientas siguen siendo los mismos, y los usuarios no ven un valor agregado real en lo que respecta al autoservicio.

**Para ser verdaderamente de autoservicio, las soluciones de analítica deben ser nativas de la nube. Los procesos de adquisición, almacenamiento, organización y generación de informes de datos deben estar diseñados para integrarse de manera nativa en sistemas creados exclusivamente para la nube.**

Deben dar asistencia a entornos en la nube dinámicos, elásticos, escalables y cada vez más contenerizados y orientados a proporcionar microservicios. Las organizaciones exitosas, en lugar de limitarse a recrear las plataformas de datos monolíticas que tenían en las instalaciones, adoptan la agilidad en la nube y utilizan sistemas de código abierto, como Kubernetes o Docker, para organizar los contenedores y sus flujos de datos.

Google Cloud Platform resulta atractiva para la analítica de autoservicio, ya que proporciona un conjunto de analítica inteligente, nativo y sin servidores. Cada componente se activa de forma fácil, escala dinámicamente o reduce la asignación de recursos según el uso. No es necesario planificar las operaciones de mantenimiento. Las organizaciones pueden centrarse en los datos y en el valor que proporcionan a su empresa con la libertad de aprovechar cualquier componente analítico, utilizar los recursos de manera flexible y controlar los costos.



## Lección n.º 2 ETL y autoservicio no son una buena combinación

Las tecnologías de integración de datos heredadas, como extraer, transformar y cargar (ETL), se pueden utilizar con el fin de crear los pipelines para mover los datos almacenados en las instalaciones y en otras nubes a Google Cloud Platform. De manera alternativa, las nuevas tecnologías de integración y flujo de datos dedicadas a la nube, como StreamSets, Fivetran o Google Cloud Data Fusion, son eficaces en la transmisión de datos entre sistemas y están disponibles para la analítica en Google Cloud Platform.

**Puede resultar tentador continuar utilizando ETL después de que se haya llenado un almacén de datos o un lago de datos, pero ETL no es compatible con la analítica de autoservicio.**

Se trata de una tecnología bastante técnica y compleja dirigida a los ingenieros de datos que trabajan con transformaciones recurrentes y en tiempo real, bien definidas y en movimientos de datos a gran escala. Se necesita una preparación de datos adicional para que los profesionales expertos en datos empresariales puedan acceder a ellos y utilizarlos en su analítica.

Google Cloud Platform ofrece una solución de preparación de datos como parte de su conjunto de analítica inteligente: Cloud Dataprep by Trifacta. Evalúa la calidad de los datos, los refina, estandariza y limpia, además, combina los datos y maneja diversos cálculos de datos.

Cloud Dataprep by Trifacta está diseñado para ayudar a una gran cantidad de profesionales de datos (ingenieros de datos, analistas de datos, analistas comerciales y otros profesionales basados en datos) a desbloquear un lago de datos o almacén de datos mediante la interacción con el contenido de los datos para refinarlo de manera iterativa e integrarlo a fin de alimentar la analítica posterior basada en la empresa.

# Lección n.º 3 El autoservicio no está disponible para todos

Cuando más personas tienen la posibilidad de utilizar la analítica de autoservicio, más personas la utilizarán. Sin embargo, eso no significa que los datos estén disponibles para todos. Para empoderar a un mayor número de usuarios a que pueda utilizar el autoservicio, es necesario tener un control minucioso para:

- Evitar que los datos proliferen fuera de control
- Cumplir con los requisitos normativos
- Mantener la confianza en los datos utilizados para tomar decisiones empresariales

**Es importante encontrar el equilibrio adecuado entre proteger los activos de datos y permitir a los usuarios colaborar y obtener valor de los datos como más les convenga.**



Las organizaciones deben encontrar el equilibrio adecuado entre proteger los activos de datos (desde las perspectivas de gobernanza y seguridad) y permitir a los usuarios colaborar y obtener valor de los datos como más les convenga. Existen tres formas de lograr este equilibrio.

### Reducir los silos

Puedes detener la proliferación de los silos de datos cuando se centralizan el almacenamiento y el procesamiento de datos en la nube con una escalabilidad prácticamente ilimitada y, además, cuando se autoriza a los usuarios finales a llevar sus propios datos a la nube. Los usuarios recopilan extractos de datos, ejecutan sus propias rutinas de preparación y crean sus informes en la nube y desde ella en lugar de extraer datos y duplicarlos en hojas de cálculo.

### Utilizar un catálogo de datos

Los recursos compartidos, como un catálogo central o un glosario, que administran definiciones de datos, metadatos y conocimiento sobre el linaje de los datos ayudan a los usuarios a encontrar datos más rápido y permiten a las organizaciones controlar las fuentes de datos y supervisar su ciclo de vida. Las soluciones de aprendizaje automático (ML) e inteligencia artificial (IA) automatizan la colección y administración de metadatos y conocimiento relacionado sobre los datos.

### Rastrear y documentar el linaje de los datos

El linaje de datos (es decir, cómo se han utilizado y transformado los datos y quién lo ha hecho), es importante para la generación de informes regulatorios y auditorías. También es importante para los encargados de la toma de decisiones, quienes deben comprender el historial de los datos detrás de la analítica, las visualizaciones y las recomendaciones prescriptivas, así como el impacto de los nuevos requisitos en el pipeline de datos para producir el análisis.

## Lección n.º 4

# Revisar los roles y las responsabilidades

La preparación de datos de autoservicio invita a las organizaciones a revisar los roles y las responsabilidades de los principales stakeholders en la cadena de valor del análisis de datos y a reconsiderar el rígido proceso de pipeline de datos como un flujo más flexible y ágil. Cada stakeholder aprovecha la preparación de datos de autoservicio de manera diferente para sus necesidades, a la vez que adopta un diseño, comunicación y marco de colaboración comunes para proporcionar analítica.

**La preparación de datos de autoservicio presenta nuevas formas de colaboración entre ingenieros de datos, analistas de datos, científicos de datos, arquitectos de datos y ejecutivos y administradores empresariales centrados en la analítica.**

**LOS ANALISTAS DE DATOS**, además de los analistas comerciales, los administradores de proyectos y otros puestos relacionados, suelen estar en contacto estrecho con los usuarios de la empresa. A menudo exploran datos sin procesar con el fin de descubrir lo que puede ser útil para responder preguntas de la empresa. Su objetivo es obtener respuestas de la manera más rápida y fácil posible.

**Cloud Dataprep by Trifacta** es ideal para los analistas de datos porque es fácil de implementar y no requiere una gran experiencia técnica en codificación. Los analistas de datos pueden participar desde el principio en el proceso de preparación de datos explorando y creando prototipos de datos que se adapten a sus necesidades comerciales. Pueden colaborar con los ingenieros de datos para organizar el pipeline de datos integral cuando la automatización es necesaria para la sostenibilidad y la repetibilidad.




**LOS INGENIEROS DE DATOS** diseñan, crean y administran el procesamiento de datos y la arquitectura de datos para apoyar la analítica y la ciencia de datos. Están estrechamente involucrados en la transformación de los datos, incluida la exploración y elaboración de perfiles de los datos sin procesar. Unos de los principales objetivos de los ingenieros de datos es la optimización y automatización de los procesos relacionados con los datos para poder administrar un mayor número de ellos.

**Cloud Dataprep by Trifacta** es ideal para los ingenieros de datos porque pueden hacer operativos y monitorear los diversos flujos de datos que diseñan ellos mismos o los analistas de datos. Facilita que los ingenieros de datos colaboren con todos los stakeholders para comprender los requisitos de la infraestructura de datos y proporcionar orientación a los usuarios para mejorar la forma en que exploran, analizan y consumen datos, además de realizar modelos de ellos.

**LOS CIENTÍFICOS DE DATOS** aplican conocimientos y habilidades especializados para diseñar y realizar modelos de algoritmos aprovechando el aprendizaje automático y la inteligencia artificial. No obstante, hasta un 80 % de su tiempo se consume con tareas rutinarias de preparación de datos, lo que les deja poco tiempo para la innovación.

**Cloud Dataprep by Trifacta** es ideal para los científicos de datos porque es una solución de preparación de datos nativa de la nube. Simplifica las tareas rutinarias de preparación de datos, lo que permite delegarlas en recursos más fácilmente disponibles y menos costosos.





## Lección n.º 5

# La calidad de los datos es responsabilidad de todos

---

Todos los stakeholders necesitan saber que pueden confiar en los datos que impulsan su analítica, sus visualizaciones y sus algoritmos. El resultado de cualquier modelo predictivo o insight de analítica es tan bueno como los datos que lo alimentan.

**Todas las organizaciones deberían comprometerse a hacer que los datos limpios sean una prioridad.**

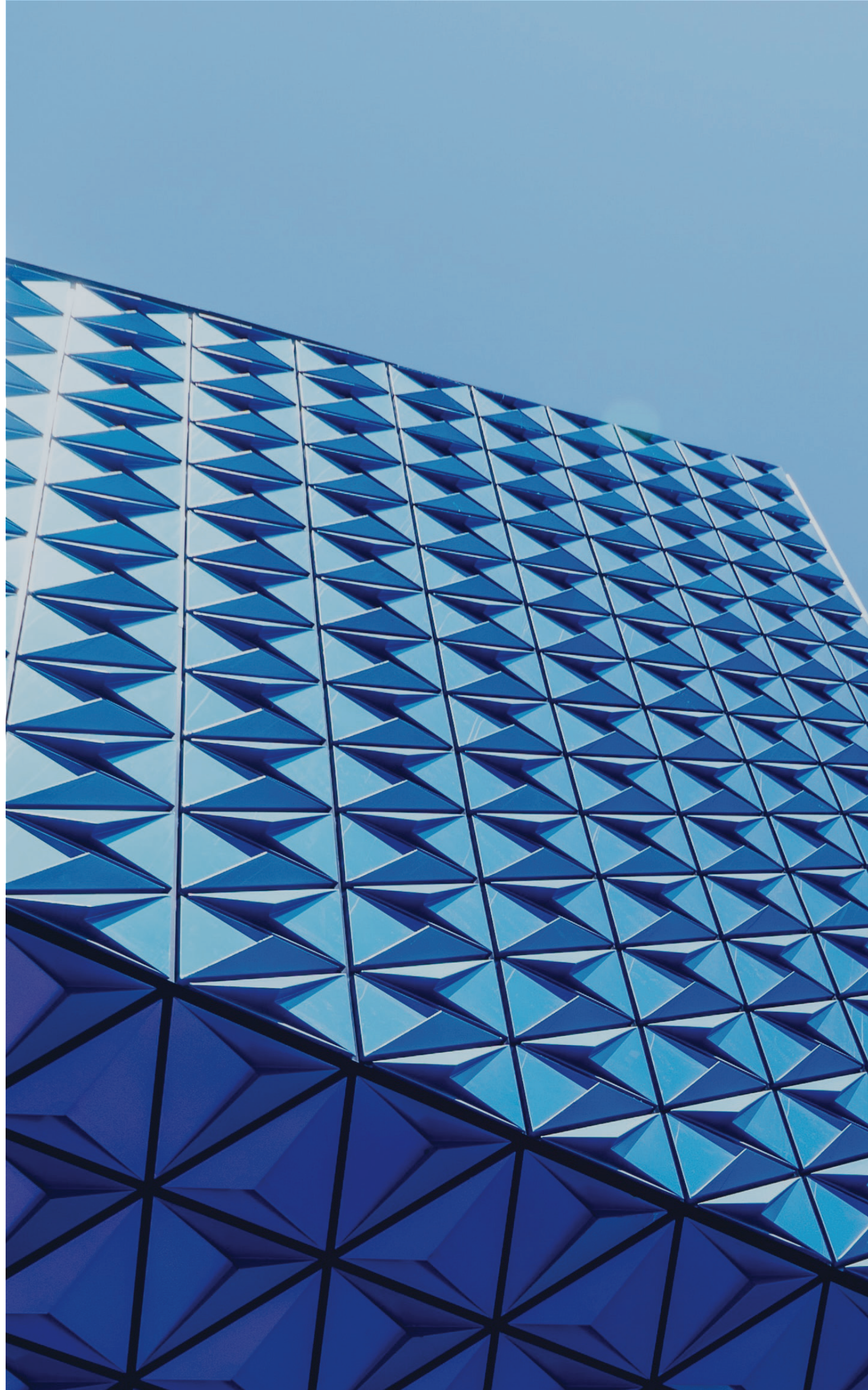
Aunque sería ideal que la calidad de los datos pudiera abordarse a nivel corporativo, no es realista. El volumen de datos y la enorme variedad de fuentes de datos (tanto internas como de terceros), los tipos y el contexto dificultan la evaluación de la calidad de los datos, la corrección de fallas y el monitoreo de nuevos datos para detectar problemas de calidad. Los nuevos repositorios de datos, como los lagos de datos en la nube y los almacenes de datos modernos, presentan desafíos más complejos en relación con la calidad de datos.

Es más eficaz abordar la calidad de los datos en la etapa de preparación de datos. Cloud Dataprep by Trifacta mejora la exactitud, la coherencia y la integridad de los datos mediante la aplicación de aprendizaje automático e inteligencia artificial para automatizar los procedimientos de limpieza de datos. La automatización maneja la escala de repositorios de datos muy grandes e identifica rápidamente los valores de datos que parecen estar incorrectos o no válidos, que faltan o no coinciden. Los valores atípicos que requieren una inspección más detallada se marcan automáticamente.

Las rutinas automatizadas de elaboración de perfiles y

limpieza de datos detectan inconsistencias en todas las fuentes que se integran en un almacenamiento de datos en la nube, destacan probables duplicaciones de datos y recomiendan cómo corregir visualmente problemas de calidad de datos mediante transformaciones automatizadas sin código.

A medida que se incorporan nuevos datos estructurados, no estructurados o semiestructurados en la nube, la calidad de los datos se valida continuamente. “Validación continua” significa que los usuarios no tienen que esperar hasta el final del proceso de validación para ver y probar los resultados, lo que sería un retraso incompatible con las actuales metodologías de desarrollo ágiles.



## Conclusión

Alteryx ha ayudado a muchas organizaciones a tener éxito con la analítica de autoservicio en la nube. De esta experiencia, hemos aprendido cinco lecciones clave:

**1.** No funciona el uso de un enfoque simple de reubicación para mover la analítica local a la nube.

Selecciona una plataforma en la nube que ofrezca componentes de analítica diseñados específicamente para la nube, como Google Cloud Platform, con su conjunto de analítica inteligente e integral nativo y sin servidores.

**2.** Las tecnologías ETL pueden suavizar la fricción resultante del movimiento de datos locales a repositorios en la nube, pero ETL es demasiado técnico e inflexible para la analítica de autoservicio. Una mejor alternativa es adoptar soluciones de preparación de datos de autoservicio, como Cloud Dataprep by Trifacta, para empoderar a todos los usuarios técnicos y empresariales a aprovechar los datos que necesitan para la analítica.

**3.** El autoservicio no es está disponible para todos y tiene sus reglas. Para que la analítica de autoservicio tenga éxito, las organizaciones deben encontrar el equilibrio adecuado entre una gobernanza y apertura cuidadosas.

Comprender el linaje de datos y utilizar un glosario de datos empresariales son fundamentales para el éxito.

**4.** Adoptar nuevos procesos ágiles que destaquen la colaboración y un lenguaje común entre todos los stakeholders a fin de transformar los datos en un activo refinado para la analítica.

**5.** La calidad de los datos es responsabilidad de todos. Cloud Dataprep by Trifacta ofrece una evaluación automática de la calidad de los datos y una validación y resolución continuas con el fin de transmitir confianza en los datos utilizados para la analítica de autoservicio.



## ACERCA DE ALTERYX

Alteryx, la empresa de automatización analítica, se centra en permitir que cada persona transforme los datos en innovación. Alteryx unifica analítica, ciencia de datos y automatización de procesos empresariales en una plataforma integral para acelerar la transformación digital y dar forma al futuro de la automatización de analítica. Organizaciones de todos los tamaños y todas partes del mundo confían en Alteryx para entregar resultados comerciales de alto impacto y mejorar rápidamente las capacidades de sus colaboradores modernos. Para obtener más información, visita [www.alteryx.com](http://www.alteryx.com).

Alteryx es una marca comercial registrada de Alteryx, Inc. Todos los demás productos y marcas pueden ser marcas comerciales o marcas comerciales registradas de sus respectivos dueños.

### Alteryx Analytics Cloud

Perfila, transforma y entrega tus datos, y conéctate a ellos con pipelines de datos automatizados de manera rápida y eficiente con Alteryx como parte de Alteryx Analytics Cloud. Con código simple o sin código, tenemos lo que necesitas. Explora de forma visual la calidad de los datos, acelera la preparación de datos para la analítica y crea e implementa pipelines de datos en segundos.

[MÁS INFORMACIÓN >](#)

### Alteryx Machine Learning Platform

Escala la ciencia de datos en toda tu empresa con el aprendizaje automático automatizado (AutoML) y la ingeniería de características, que empoderan por igual a los expertos en negocios y a los científicos de datos para acelerar los insights.

[MÁS INFORMACIÓN >](#)

### Alteryx Designer

Automatiza cada paso de analítica, incluidas la preparación de datos, la combinación, la generación de informes, la analítica predictiva y la ciencia de datos. Accede a cualquier fuente de datos, archivo, aplicación o tipo de datos, y experimenta la simplicidad y la eficacia de una plataforma de autoservicio con más de 300 bloques de creación de automatización. Comienza a crear resultados interactivos hoy.

[MÁS INFORMACIÓN >](#)

### Alteryx Intelligence Suite

Independientemente de que seas un principiante o un experto, podrás crear innovadores modelos de aprendizaje automático en pocos minutos. Alteryx Intelligence Suite está diseñado para ayudarte a encontrar insights ocultos en tus datos. Desde el análisis de datos estructurados hasta la búsqueda de insights ocultos en textos y documentos, Alteryx Intelligence Suite te permite resolver los problemas más complejos de tu empresa.

[MÁS INFORMACIÓN >](#)

### Alteryx Server

Escala, comparte y controla los procesos, modelos y datos de flujos de trabajo de analítica con Alteryx Server. Crea flujos de trabajo de analítica con Alteryx Designer y publícalos en Alteryx Server para programar la generación automatizada de informes y resultados.

El control de datos incorporado, la seguridad administrada centralmente y la alta disponibilidad permiten a las organizaciones escalar la analítica tanto a través de los diferentes departamentos como en toda la organización.

[MÁS INFORMACIÓN >](#)