

Tendencias

Top Diez

BIG DATA

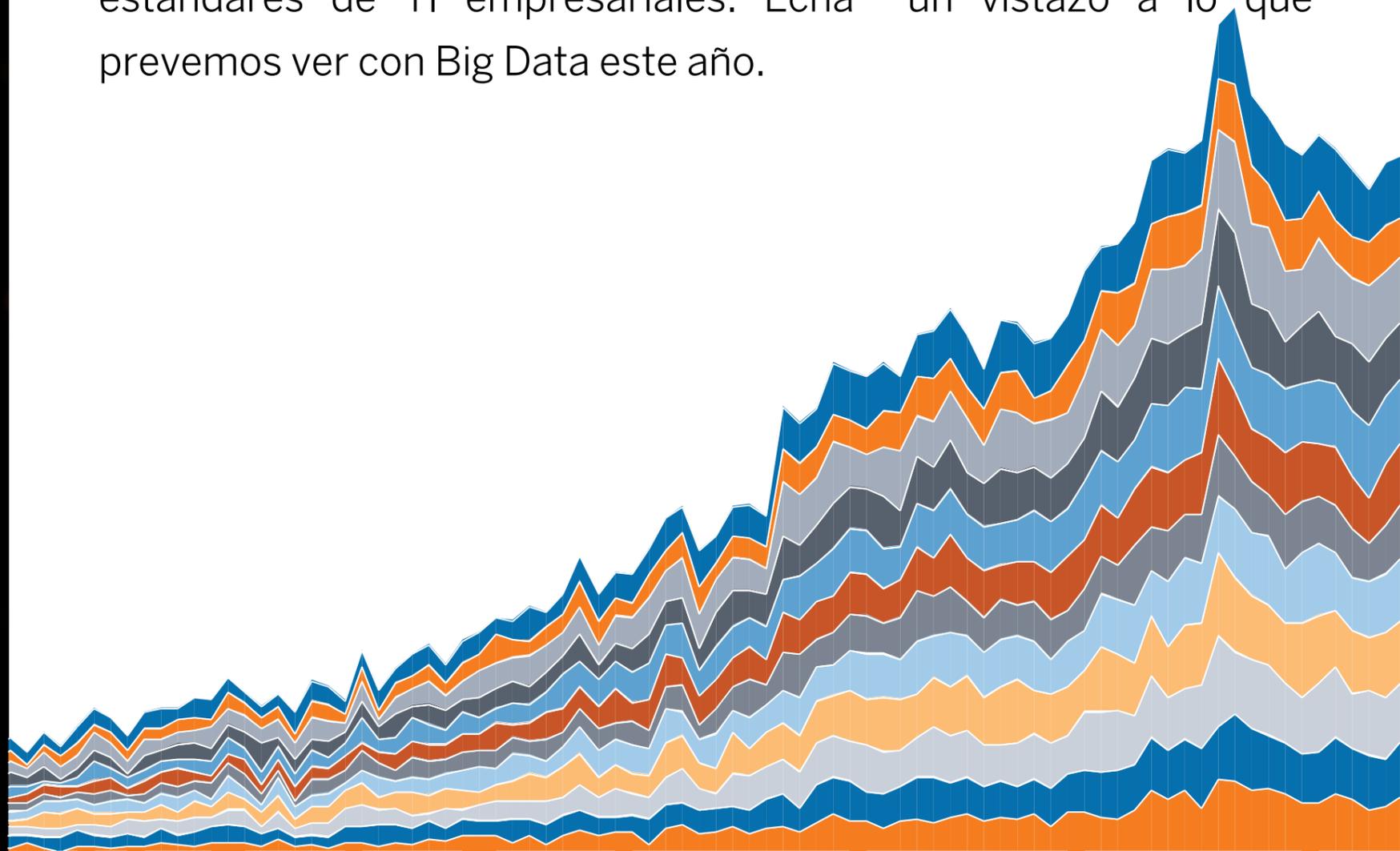




# Las 10 tendencias de Big Data

Más organizaciones almacenan, procesan y extraen valor de datos de todas las formas y tamaños. Sistemas que soportan grandes volúmenes de datos estructurados y no estructurados continuarán para levantarse. El mercado exigirá plataformas que ayuden a los datos los custodios gobiernan y protegen el big data mientras empoderan el fin usuarios para analizar esos datos. Estos sistemas madurarán para operar dentro de los sistemas y estándares de TI empresariales. Echa un vistazo a lo que prevemos ver con Big Data este año.

Cada año en Tableau, comenzamos una conversación sobre lo que está pasando en la industria. La discusión impulsa una lista de las principales tendencias de big data para el año. Estas son nuestras predicciones.



# BIG DATA

1

## El Big Data se vuelve rápido y accesible: Las opciones se amplían para acelerar Hadoop

Seguro puede realizar aprendizaje automático y realizar análisis de opiniones en Hadoop, pero la primera pregunta que la gente a menudo hace es: ¿Qué tan rápido es el SQL interactivo? SQL, después de todo, es el conducto a los usuarios empresariales que desean utilizar los datos de Hadoop para paneles de KPI más rápidos y repetibles como análisis exploratorio.

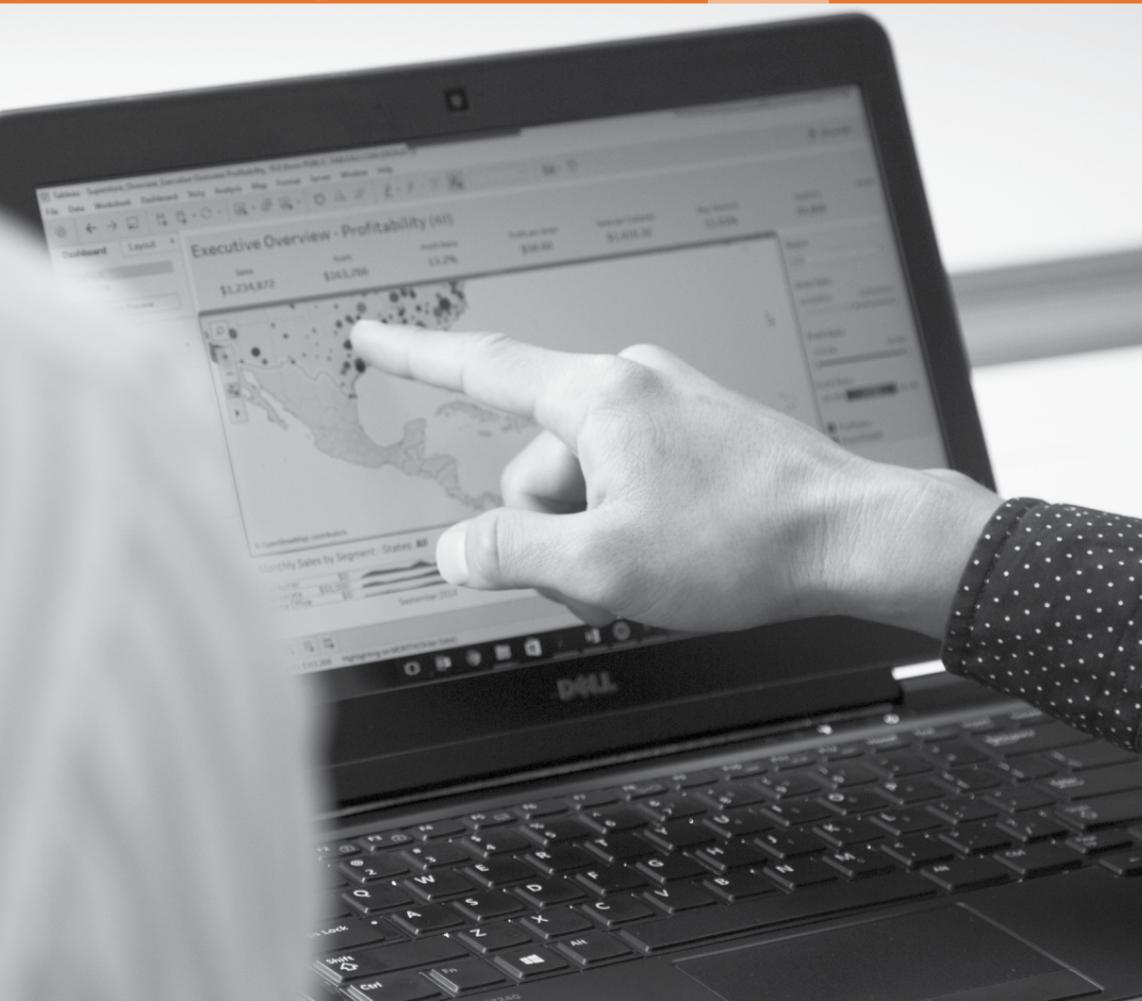
Esta necesidad de velocidad ha impulsado la adopción de bases de datos más rápidas como Exasol y MemSQL, Tiendas basadas en Hadoop como Kudu y tecnologías que permiten hacer consultas más rápidas.

El uso de motores SQL-onHadoop (Apache Impala, Hive LLAP, Presto, Phoenix, and Drill) y OLAP-on-Hadoop (AtScale, Jethro Data, and Kyvos Insights), son aceleradores de consultas difuminando las líneas entre los almacenes tradicionales y el mundo del big data.

# Big Data ya no solo Hadoop: Herramientas diseñadas específicamente para Hadoop se vuelven obsoletos

En años anteriores, vimos surgir varias tecnologías con la ola de los big-data para satisfacer la necesidad de análisis en Hadoop. Pero empresas con entornos complejos y heterogéneos no ya no quieren adoptar un punto de acceso de BI aislado solo para una fuente de datos (Hadoop). Las respuestas a sus preguntas están enterradas en un host de fuentes que van desde sistemas de registro hasta almacenes en la nube, a datos estructurados, no estructurados y fuentes que no son de Hadoop. (Por cierto, incluso las bases de datos relacionales se están preparando para el big data. SQL Server, por ejemplo, recientemente añadió soporte JSON.)

Los clientes exigirán análisis sobre todos los datos. Las Plataformas que son independientes de los datos y de la fuente prosperarán, mientras que los que son diseñados específicamente para Hadoop y no se implementa en todos los casos de uso, caerán por el camino. El exit of Platfora servirá en comienzo como indicador de esta tendencia.



## Apalancamiento de las organizaciones data lake desde el principio para proveer valor.

Un Data Lake es como un lago artificial creado por el hombre. Primero contruyes la cavidad (construir un clúster), a continuación, dejas que se llene de agua (datos). Una vez que esta construido el lago, se empieza a llenar de agua (datos), es como generar varios propósitos, beber y recrear (análisis predictivo, ML, ciberseguridad, etc.).

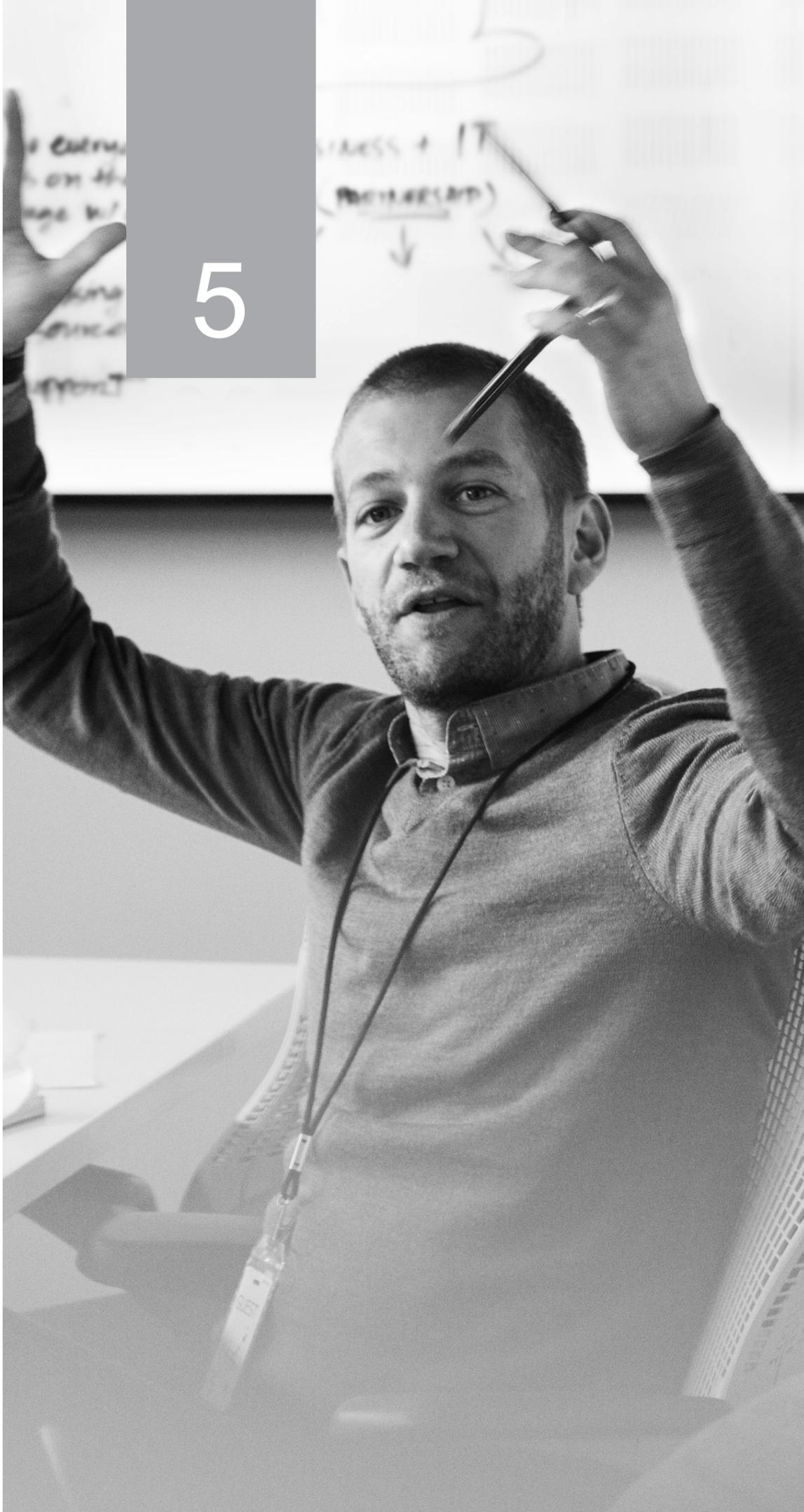
Hasta ahora, llenar el lago ha sido un fin en sí mismo. Éste cambiará a medida que se apriete la justificación empresarial de Hadoop. Las organizaciones exigirán un uso repetible y ágil del lago para respuestas más rápidas. Considerarán cuidadosamente los negocios antes de invertir en personal, datos e infraestructura. Esto fomentará una asociación más fuerte entre el negocio e IT. Y las plataformas de autoservicio obtendrán un reconocimiento más profundo como la de aprovechar los activos de big data.

# 4

## Arquitecturas maduras para rechazar modelos de talla única en todos los marcos

Hadoop ya no es sólo una plataforma de procesamiento por lotes para casos de uso de ciencia de datos. Tiende a convertirse en un motor multiusos para el análisis ad hoc. Incluso se está utilizando para operaciones donde se debe informar sobre las cargas de trabajo diario, que tradicionalmente maneja los almacenes de datos.

Las organizaciones están respondiendo a estas necesidades híbridas mediante la búsqueda de casos de uso específico. Investigarán una serie de factores, incluyendo usuarios, preguntas volumen, la frecuencia de acceso, la velocidad de los datos y el nivel de agregación antes de una estrategia de datos. Estas arquitecturas de referencia modernas estarán impulsadas por las necesidades. Ellos combinarán las mejores herramientas de preparación de datos de autoservicio, Hadoop Core y plataformas para el análisis del usuario final de manera que se puedan reconfigurar a medida que evolucionen esas necesidades. La flexibilidad de estas arquitecturas impulsará en última instancia las opciones tecnológicas.



5

## Variedad, no volumen o velocidad, impulsa las inversiones en big-data

Gartner define el big data como las tres V: Gran **Volumen**, **Velocidad** alta, gran **Variedad** activos de información (high-volume, highvelocity, high-variety information assets). Mientras que las tres V's se están convirtiendo en el mayor impulsor de las inversiones del big-data, como se ve en los resultados de una reciente encuesta de New Vantage Partners. Esta tendencia seguirá creciendo a medida que las empresas busquen integrar más fuentes y centrarse en el "Long Tail" de los big data.

Desde JSON sin esquemas hasta tipos anidados en otras bases de datos (relacional y NoSQL), a datos no planos (Avro, Parquet, XML), estos formatos de datos se multiplican y los conectores se están convirtiendo en algo crucial. Las empresas seguirán evaluando las plataformas de análisis basados en su capacidad para proporcionar conectividad en vivo y en directo para estas fuentes dispares.

# Spark y machine learning iluminan los big data

**Apache Spark**, un componente del ecosistema de Hadoop que se está convirtiendo ahora en la plataforma de big data de elección para las empresas. En una encuesta de arquitectos de datos, Gerentes de TI y analistas de BI, casi el 70% de los encuestados favorecieron a Spark por encima de MapReduce, que está orientado a lotes y no se presta aplicaciones interactivas o procesamiento de transmisiones en tiempo real.

Estas capacidades de big-compute-on-big-data han elevado plataformas con aprendizaje automático intensivo en informática, IA, y en algoritmos de gráficos. Microsoft Azure ML en particular ha tomado fuerza gracias a su fácil integración con plataformas de Microsoft existentes. La apertura de ML a las masas conducen a la creación de más modelos y aplicaciones que generen petabytes de datos. A medida que las máquinas aprenden y los sistemas se vuelven inteligentes, todos los ojos estarán en los proveedores de software de autoservicio para ver cómo hacer que estos datos sean accesibles para el usuario final.

# La convergencia del IoT, la nube, y el big data crean nuevas oportunidades para el análisis de autoservicio

Todo tendrá un sensor que envía información de vuelta a la nave nodriza. IoT está generando grandes volúmenes de datos estructurados y no estructurados, y una proporción cada vez mayor de estos datos es que se implementa en los servicios en la nube. Los datos son a menudo heterogéneos y viven en múltiples sistemas relacionales y no relacionales, desde clústeres de Hadoop hasta bases de datos NoSQL. Mientras que las innovaciones en el almacenamiento y los servicios gestionados, han acelerado el proceso de captura, el acceso y la comprensión de los datos, en sí todavía plantea todo un desafío significativo. Como resultado, va en creciente demanda las herramientas analíticas que se conectan a la perfección y combinan una amplia variedad de fuentes de datos en la nube. Estas herramientas permiten a las empresas explorar y visualizar cualquier tipo de datos almacenados en cualquier lugar, ayudándoles a descubrir oportunidades ocultas en su inversión en IoT.

## La preparación de datos de autoservicio se convierte en corriente principal a medida que los usuarios finales comienzan a dar forma a los big data.

Hacer que los datos de Hadoop sean accesibles a los usuarios empresariales es uno de los mayores desafíos de nuestro tiempo. El auge de la analítica de las plataformas de autoservicio, han mejorado este viaje. Pero los usuarios de negocios quieren reducir aún más el tiempo y la complejidad de la preparación de datos para el análisis, que es especialmente importante cuando se trata de un variedad de tipos de datos y formatos.

Las herramientas ágiles de preparación de datos de autoservicio no solo permiten datos de Hadoop prepararse en la fuente, si no que también hacen que los datos estén disponibles instantáneamente para una exploración más rápida y sencilla. Hemos visto un anfitrión innovación en este espacio de empresas enfocadas en la preparación de datos de usuarios finales para big data como Alteryx, Trifacta y Paxata. Estas herramientas están reduciendo las barreras de entrada para **Late Hadoop** y seguirán ganando terreno.

# El big data crece: Hadoop se suma a los estándares empresariales

Estamos viendo una creciente tendencia de Hadoop, convirtiéndose en una parte central del panorama de TI empresarial. Inversiones en seguridad y componentes de gobernanza se levantarán en torno a los sistemas empresariales. **Apache Sentry** proporciona un sistema para hacer cumplir una autorización basada en roles para datos y metadatos almacenados en un clúster de Hadoop. **Apache Atlas**, creado como parte de iniciativa de gobierno de datos, permite a las organizaciones aplicar clasificación de datos coherente en todo el ecosistema de datos. **Apache Ranger** proporciona administración de seguridad centralizada para Hadoop.

Los clientes están empezando a esperar este tipo de capacidades de sus plataformas RDBMS de nivel empresarial. Estas capacidades están avanzando a la vanguardia de los big data de tecnologías emergentes, eliminando así otra barrera para la adopción empresarial.

# El aumento de los catálogos de metadatos ayuda a las personas a encontrar análisis dignos de big data.

Durante mucho tiempo, las empresas tiraron datos porque tenían demasiado que procesar. Con Hadoop, pueden procesar muchos datos, pero estos no están generalmente organizados de una manera que se puedan encontrar.

Los catálogos de metadatos pueden ayudar a los usuarios a descubrir y comprender los datos relevantes que vale la pena analizar, utilizando herramientas de autoservicio. Este vacío en las necesidades del cliente lo están cubriendo empresas como Alation y Waterline, que utiliza el aprendizaje automático para automatizar el trabajo de la búsqueda de datos en Hadoop. Ellos categorizan archivos mediante etiquetas, descubren relaciones entre activos de datos e incluso proporcionan consultas y sugerencias a través de interfaces que el usuario puede buscar. Esto ayuda tanto a los consumidores de datos como a los administradores, reduciendo tiempo y ayudando a encontrar y consultar con precisión los datos. El próximo año veremos más conciencia y demanda de descubrimiento del autoservicio, el cual crecerá como una extensión natural del autoservicio analítico.

## Acerca de Tableau

Integrar la visualización de datos en sus programas y procesos minoristas es más fácil de lo que cree.

Tableau Software ayuda a las personas a ver y comprender los datos sin importar su tamaño o sistemas en los que se almacena. Conecte, mezcle, visualice y comparta rápidamente paneles de datos con un experiencia perfecta desde el PC hasta el iPad. Crear y publicar paneles con paneles automáticos actualizaciones de datos y compártelas con colegas, socios o clientes, sin habilidades de programación Obligatorio. Comience una prueba gratuita hoy mismo.

## Acerca de Solex

Socio estratégico de Tableau, que apoya a las empresas a ver y comprender sus datos

En **Solex** seleccionamos las mejores prácticas recomendadas por **Tableau Software** y la experiencia en **analítica de negocios** en más de 150 clientes, para crear una **estrategia empresarial y construir cultura impulsada por datos** confiables y seguros, logrando de esta manera que la **compañía se convierta en data driven** que toma **decisiones basadas en sus datos**.

Si tiene más consultas sobre las funcionalidades o **beneficios de Tableau** o si desea **implementar una solución de Inteligencia de negocios**, comuníquese con nosotros a través del [formulario de contacto](#) o escribanos al [Whatsapp](#) (+57 316 4576123).