

LAS 10
TENDENCIAS PRINCIPALES DE
BIG DATA

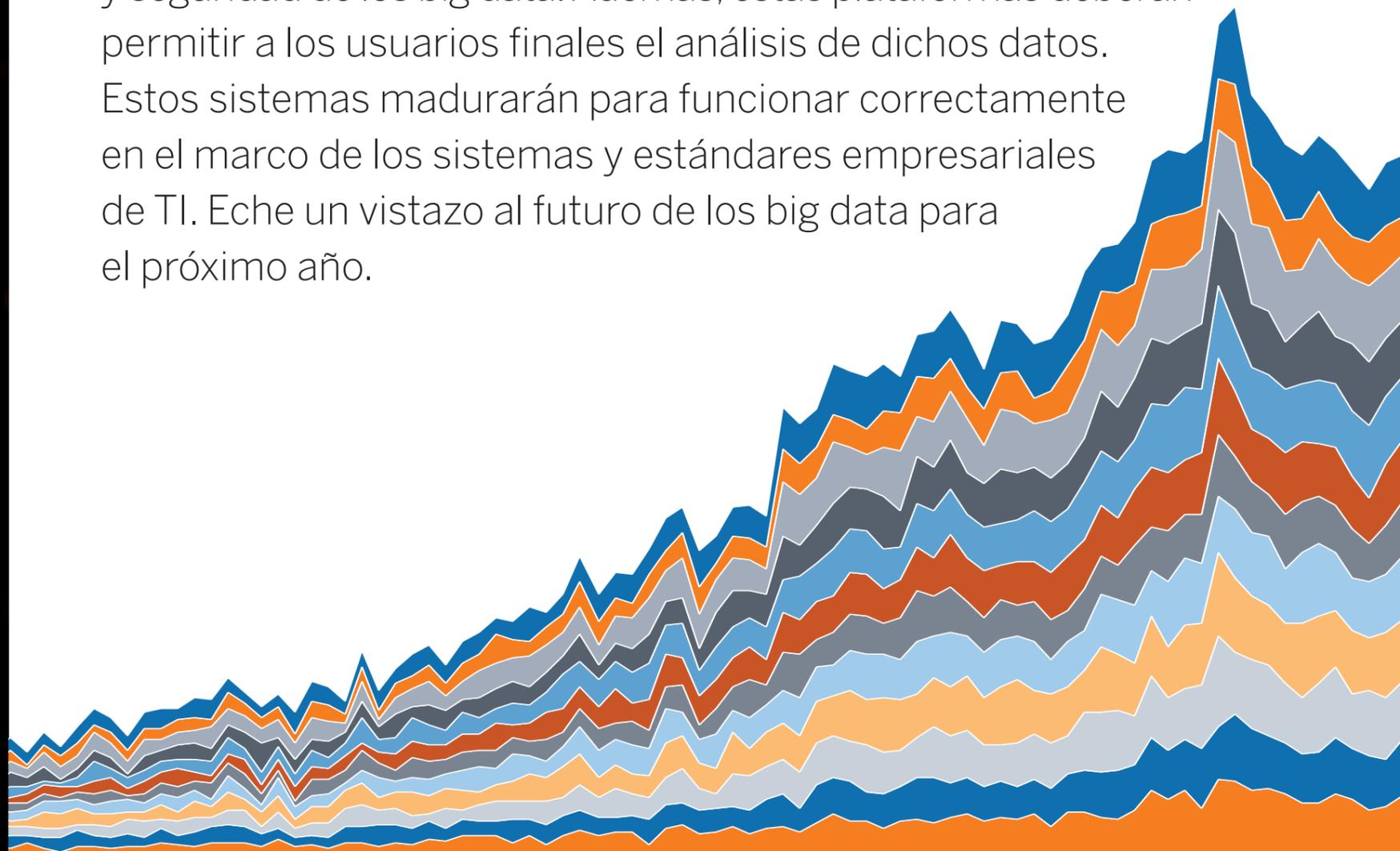




Las 10 tendencias principales de big data

Una cantidad cada vez mayor de organizaciones almacena y procesa datos de todo tipo de formatos y tamaños, además de extraer valor de ellos. Continuará el crecimiento de los sistemas que admiten grandes volúmenes de datos, tanto estructurados como no estructurados. El mercado exigirá plataformas que faciliten a los responsables de los datos las tareas de gobierno y seguridad de los big data. Además, estas plataformas deberán permitir a los usuarios finales el análisis de dichos datos. Estos sistemas madurarán para funcionar correctamente en el marco de los sistemas y estándares empresariales de TI. Eche un vistazo al futuro de los big data para el próximo año.

En Tableau, cada año analizamos lo más destacado del sector. El debate ofrece como resultado una lista con las tendencias principales de los big data para el año siguiente. Estas son nuestras predicciones.



BIG DATA

1

Los big data se vuelven rápidos y simples: las opciones se expanden para inyectar rapidez a Hadoop

Es posible implementar el aprendizaje automático y realizar análisis de opiniones con Hadoop, pero la primera pregunta que la gente suele formular es la siguiente: ¿cuál es la velocidad del SQL interactivo? Al fin y al cabo, SQL es el medio utilizado por los usuarios corporativos que desean acceder a datos de Hadoop para generar dashboards de KPI más rápidos e iterativos, además de realizar análisis exploratorios.

Esta necesidad de velocidad potenció la adopción de bases de datos más rápidas, como [Exasol](#) y [MemSQL](#), almacenes basados en Hadoop, como [Kudu](#), y tecnologías que permiten realizar consultas más ágiles. Con el uso de motores SQL en Hadoop ([Apache Impala](#), [Hive LLAP](#), [Presto](#), [Phoenix](#) y [Drill](#)) y tecnologías OLAP en Hadoop ([AtScale](#), [Jethro Data](#) y [Kyvos Insights](#)), estos aceleradores de consultas están desdibujando aún más las líneas divisorias entre los almacenes de datos tradicionales y el mundo de los big data.

Los big data ya no solo se hospedan en Hadoop: las herramientas diseñadas específicamente para Hadoop se vuelven obsoletas

En años anteriores, con la ola de los big data, vimos surgir varias tecnologías para satisfacer las necesidades de análisis en Hadoop. Sin embargo, las empresas con entornos complejos y heterogéneos ya no desean adoptar un punto de acceso a BI aislado, para una sola fuente de datos (Hadoop). Las respuestas a sus preguntas están en el interior de un conjunto de fuentes, desde sistemas de registros y almacenes de datos en la nube, hasta datos estructurados y no estructurados incluidos en fuentes de Hadoop u otros tipos. Casualmente, incluso las bases de datos relacionales se están adaptando para los big data. Por ejemplo, SQL Server recientemente incorporó compatibilidad con JSON.

Los clientes exigirán análisis de todo tipo de datos. Las plataformas compatibles con múltiples tipos de datos y fuentes prosperarán, mientras que las diseñadas para ser utilizadas solo con Hadoop y que no sirvan para diferentes casos de uso se quedarán en el camino. La [adquisición de Platfora](#) es un indicio precursor de esta tendencia.



Las organizaciones aprovechan los mares de datos desde un comienzo para obtener el máximo rendimiento

Un mar de datos es como un embalse artificial. En primer lugar, se construye una pared contenedora (se genera un clúster). Luego, se llena de agua (datos). Una vez confeccionado el mar, se comienza a utilizar el agua (los datos) con diversos propósitos, como generar electricidad, producir agua potable o proporcionar actividades de recreo (análisis predictivos, aprendizaje automático, seguridad cibernética, etc.).

Hasta ahora, incorporar agua al mar era un fin en sí mismo. Esto cambiará, ya que la justificación de negocio para utilizar Hadoop se volverá más estricta. Para obtener respuestas con mayor rapidez, las organizaciones exigirán un uso iterativo y ágil del mar. Realizarán un análisis detenido de los resultados de negocios antes de invertir en personal, datos e infraestructura. A su vez, esto promoverá una asociación más estrecha entre la **empresa y TI**. Las plataformas de autoservicio tendrán un mayor reconocimiento. Serán consideradas como las herramientas que permiten aprovechar los recursos de los big data.

4

La arquitectura evoluciona hacia un marco en el que la “talla única” deja de ser el estándar

Hadoop ya no solo es una plataforma de procesamiento por lotes para casos de uso científicos. Se ha convertido en un motor multipropósito para realizar análisis ad hoc. Incluso se está utilizando para generar informes operativos sobre volúmenes de trabajo diarios (de lo que se encargaban tradicionalmente los almacenes de datos).

Las organizaciones están reaccionando ante estas necesidades híbridas buscando diseños de arquitectura informática para casos de uso específicos. Antes de comprometerse con una estrategia de datos, investigarán un gran número de factores, entre los que se incluyen roles de usuarios, preguntas, volúmenes, frecuencia de acceso, velocidad de datos y nivel de agregación. Estas arquitecturas modernas se adaptarán a las necesidades. Combinarán las mejores herramientas de preparación de datos de autoservicio con Hadoop Core y plataformas de análisis de usuario final. De esta forma, podrán reconfigurarse según evolucionen las necesidades. En última instancia, la flexibilidad de estas arquitecturas será el factor decisivo entre las opciones tecnológicas.



5

Ni volumen ni velocidad: la variedad será lo que impulse las inversiones en big data

Gartner define los big data con las “tres uves”: grandes volúmenes, alta velocidad y gran variedad de activos de información. A medida que estos tres componentes crecen, la variedad se convierte en el factor más decisivo a la hora de evaluar una inversión en big data. Esto se puede ver en los resultados de una **encuesta reciente** realizada por New Vantage Partners. Esta tendencia continuará en aumento a medida que las empresas quieran integrar más fuentes y centrarse en la “**larga cola**” de los big data. JSON sin esquemas, tipos anidados de otras bases de datos (relacionales y NoSQL), datos estructurados (Avro, Parquet, XML)... Los formatos de datos se multiplican, por lo que los conectores son cada vez más fundamentales. Las empresas seguirán evaluando las plataformas de análisis según su capacidad para proporcionar conectividad directa y en tiempo real a esta diversidad de fuentes.

Spark y el aprendizaje automático dinamizan el mundo de los big data

Apache Spark, un antiguo componente del ecosistema de Hadoop, se está convirtiendo en la plataforma de big data preferida entre las empresas. En una *encuesta* realizada a arquitectos de datos, administradores de TI y analistas de BI, casi el 70 % de ellos priorizó Spark sobre el tradicional MapReduce. Este último realiza procesos por lote y no se presta a aplicaciones interactivas ni a procesamiento secuencial en tiempo real.

Estas grandes capacidades de cómputo en big data pusieron en primer plano a las plataformas que ofrecen aprendizaje automático con cálculo exhaustivo, inteligencia artificial y algoritmos gráficos. En particular, el aprendizaje automático de Microsoft Azure se ha destacado, gracias a su facilidad de uso, adaptada para principiantes, y la integración simplificada con plataformas existentes de Microsoft. Poner el aprendizaje automático a disposición de las masas conducirá a la creación de más modelos y aplicaciones que generen petabytes de datos. A medida que las máquinas aprendan y los sistemas se vuelvan más inteligentes, toda la atención se centrará en los proveedores de software de autoservicio. Existirá curiosidad por ver cómo facilitan el acceso de los datos al usuario final.

7

La confluencia entre IoT, la nube y los big data crea nuevas oportunidades para el análisis de autoservicio

Cualquier objeto cotidiano estará dotado de un “sensor” que enviará información. La IoT (Internet de las cosas) está generando volúmenes masivos de datos estructurados y no estructurados. Además, una parte cada vez mayor de estos **datos se está implementando en servicios en la nube**. En su mayoría, los datos son heterogéneos y están hospedados en diferentes sistemas, tanto relacionales como de otros tipos, que abarcan desde clústeres de Hadoop hasta bases de datos NoSQL. Mientras que las innovaciones en los servicios de almacenamiento y administración aceleraron el proceso de captura, el acceso a los datos y su comprensión todavía suponen un importante desafío final. Como resultado, se observa un aumento en la demanda de herramientas analíticas que combinen una amplia variedad de fuentes de datos hospedados en la nube y que se conecten con ellas sin problemas. Dichas herramientas permiten que los negocios exploren y visualicen cualquier tipo de datos, sin importar dónde estén almacenados. Esto los ayuda a descubrir las oportunidades ocultas en su inversión de IoT.

La preparación de datos de autoservicio se vuelve una tendencia dominante a medida que los usuarios finales comienzan a dar forma a los big data

Hacer que los datos de Hadoop sean accesibles para los usuarios corporativos es uno de los desafíos más grandes de la actualidad. El surgimiento de plataformas de análisis de autoservicio ha allanado este camino. Sin embargo, los usuarios corporativos desean reducir aún más los tiempos y la complejidad de preparar datos para el análisis. Y esto se vuelve particularmente importante a la hora de procesar diversos tipos y formatos de datos.

La agilidad de las herramientas de preparación de datos de autoservicio no solo permite preparar datos de Hadoop directamente en la fuente, sino que también los ponen a disposición en forma de instantáneas. Así se permite una exploración más rápida y sencilla. En relación con esto, hemos sido testigos de numerosas innovaciones producidas por empresas dedicadas a la preparación de big data para el usuario final, como [Alteryx](#), [Trifacta](#) y [Paxata](#). Estas herramientas les están abriendo el paso tanto a los [escépticos](#) como a los [rezagados en la adopción de Hadoop](#). Por ello, se continuarán afianzando.

Los big data crecen: Hadoop se incorpora a los estándares empresariales

Una tendencia en aumento es la transformación de Hadoop en una parte fundamental del entorno de TI empresarial. Veremos un crecimiento de la inversión en los componentes de seguridad y gobierno que rodean a los sistemas empresariales. Apache Sentry proporciona un sistema para aplicar una autorización detallada y basada en roles sobre los datos y los metadatos almacenados en un clúster de Hadoop. [Apache Atlas](#), creado como parte de la iniciativa de gobierno de datos, permite que las organizaciones apliquen una clasificación homogénea en todo su ecosistema de datos. [Apache Ranger](#) proporciona una herramienta de administración de seguridad centralizada para Hadoop.

Estos son los tipos de funcionalidades que los clientes corporativos esperan de sus plataformas de sistemas de administración de bases de datos relacionales (RDBMS). Actualmente, están a la vanguardia de las tecnologías emergentes de big data y contribuyen a simplificar la adopción empresarial.

El incremento de los catálogos de metadatos contribuye al descubrimiento de big data valiosos para el análisis

Durante mucho tiempo, las empresas desechaban datos, porque no los podían procesar en su totalidad. Con Hadoop, pueden procesar grandes cantidades de datos. Sin embargo, estos no suelen estar organizados de manera sencilla.

Mediante las herramientas de autoservicio, los catálogos de metadatos pueden ayudar a los usuarios a descubrir y comprender datos relevantes y dignos de análisis. Esta necesidad de los clientes generó una oportunidad que fue aprovechada por [Alation](#) y [Waterline](#). Ambas organizaciones utilizan el aprendizaje automático para automatizar el trabajo de encontrar datos en Hadoop. Catalogan archivos con etiquetas, encuentran relaciones entre activos de datos e incluso proporcionan sugerencias de consultas a través de interfaces de usuario que permiten realizar búsquedas. Esto ayuda tanto a los consumidores como a los administradores de datos en la reducción del tiempo necesario para encontrar datos, hacer consultas precisas y confiar en los resultados. En el próximo año, se observará un mayor conocimiento y una mayor demanda del descubrimiento de autoservicio. Este crecerá como una extensión natural del análisis de autoservicio.



Acerca de Tableau

Integrar la visualización de datos en sus programas y procesos de comercio minorista es más fácil de lo que imagina.

Tableau Software ayuda a las personas a ver y comprender los datos, independientemente de su volumen o el número de sistemas en los que se almacenen. Conéctese rápidamente a dashboards de datos, combínelos, visualícelos y compártalos. Sin problemas, desde su PC o iPad. Cree y publique dashboards con actualizaciones de datos automáticas y compártalos con colegas, socios o clientes, sin necesidad de tener conocimientos de programación. Inicie una prueba gratuita hoy mismo.

[TABLEAU.COM/ES-ES/TRIAL](https://tableau.com/es-es/trial)