

---

Jeff Feng, administrador de productos de Tableau

# La visión de Tableau sobre los big data

Vivimos en una nueva era. Los datos constituyen la “materia prima” clave que impulsa a las empresas de hoy en día y que nos conduce a la nueva revolución industrial. Los nuevos procesos de fabricación de los siglos XVIII y XIX transformaron completamente la forma en que se fabricaban los bienes. De manera similar, en la era de los big data se está redefiniendo el modo en que se generan, analizan y consumen los datos.

La ironía es que los big data son tanto una promesa como una amenaza. Los recursos de datos actúan cada vez más como diferenciadores clave entre las empresas inmensamente rentables y las que luchan por mantenerse a flote. Sin embargo, manejar la magnitud, la variedad y el crecimiento masivos de los datos resulta difícil y costoso para los sistemas de administración de bases de datos.

Por lo tanto, las empresas recurren a las tecnologías de big data, como las bases de datos de Hadoop, Spark y NoSQL, para satisfacer sus necesidades de datos, que evolucionan rápidamente. Usan tanto modelos en instalaciones físicas (in situ) como modelos en la nube para implementar estas tecnologías. Además, en las bases de datos de análisis rápidas y en los almacenes de datos se incorporan conceptos de Hadoop para enriquecer sus ofertas o, directamente, se crean conectores a Hadoop. A medida que el panorama de los big data evoluciona y se consolida, se vislumbra un aspecto consistente: las empresas deben ser capaces de usar la misma herramienta de análisis para acceder a sus datos, ya sean cuantiosos o escasos, dondequiera que se encuentren.

## Índice

La estrategia de Tableau para los big data.....	3
¿Cómo trabaja Tableau con los big data? .....	5
Casos prácticos: Tableau y los big data .....	7
Resumen.....	8
Acerca del autor .....	8



## La estrategia de Tableau para los big data

Tableau emprendió una **misión** para ayudar a los usuarios a ver y comprender sus datos. Creemos fervientemente que la democratización de los datos conducirá al éxito de esta misión: “las personas que conocen los datos deberían ser las que hagan preguntas acerca de ellos”. Los trabajadores del conocimiento deberían poder acceder fácilmente a sus datos dondequiera que se encuentren. También tener la capacidad de analizar y descubrir información acerca de sus datos sin necesitar asistencia de la escasa élite (los científicos de datos y los desarrolladores de TI).

La visualización de los datos es importante, independientemente de la cantidad de estos, porque los traduce en información y acción. El método para la visualización de los big data tiene una importancia especial porque el costo de almacenamiento, preparación y consulta de estos datos es mucho más alto. En consecuencia, las organizaciones deben usar fuentes de datos con una buena arquitectura y aplicar de manera rigurosa las prácticas recomendadas para permitir que los trabajadores del conocimiento consulten los big data directamente. En los últimos años, se innovó mucho en materia de big data, por lo que existen muchas opciones con distintas ventajas. Tableau tiene la intención de admitir todas las plataformas de big data que sean pertinentes para sus usuarios y ayudarlos a mantener conversaciones con sus datos en tiempo real.

Para hacer realidad esta visión sobre los big data, Tableau se concentra en seis pilares:

1. **Acceso generalizado a las plataformas de big data:** Parte de nuestra visión consiste en permitir el análisis de los big data dondequiera que estén. Hoy en día, **Tableau** admite más de 40 fuentes de datos diferentes y acepta muchas otras, mediante sus opciones de extensibilidad. A medida que surjan nuevas fuentes de datos y estas se vuelvan valiosas para nuestros usuarios, continuaremos incorporándolas en nuestros productos para reducir los problemas de acceso a los datos.

Entre nuestros conectores reconocidos para el ecosistema de big data se incluyen:

- **Hadoop:** Cloudera Impala y Hive; Hortonworks Hive; MapR Hive; Amazon EMR con Impala y Hive; Pivotal HAWQ; IBM BigInsights.
- **NoSQL:** MarkLogic, Datastax.
- **Spark:** Apache Spark SQL.
- **Nube:** Amazon Redshift, Google BigQuery.
- **Datos operativos:** Splunk.
- **Bases de datos de análisis rápidas:** Actian Vectorwise y ParAccel; Teradata Aster; HP Vertica; SAP Hana; SAP Sybase; Pivotal Greenplum; EXASOL EXASolution.

2. **Visualización de autoservicio de big data para usuarios profesionales:** Los usuarios profesionales pueden visualizar sus datos con operaciones de arrastrar y soltar, sin necesidad de escribir códigos de SQL o Java ni trabajos MapReduce complejos. Tableau simplifica el análisis de datos y permite a los usuarios descubrir información visual acerca de ellos más rápido que nunca.

3. **Arquitectura de datos híbrida para la optimización del rendimiento de las consultas:** Tableau puede conectarse en tiempo real a fuentes de datos o llevar los datos a la memoria. La conectividad en tiempo real funciona de maravilla para la conexión a grandes conjuntos de datos y rápidos motores de consulta interactiva. No obstante, también podemos incrementar y acelerar las fuentes de datos más lentas creando una extracción de los datos y llevándola a nuestro motor de datos en memoria.
4. **Combinación de datos para el análisis de distintas fuentes de datos:** Los datos distribuidos suelen ser un desafío más grande que los big data. No es común que un analista cuente con datos bien empaquetados en un único lugar. Por el contrario, los datos normalmente están distribuidos en tecnologías y plataformas dispares. Tableau permite a los usuarios trascender las fuentes de datos mediante la **combinación** de big data con otras fuentes de datos (p. ej., archivos de Excel, Salesforce, MySQL), lo que permite que las organizaciones conserven sus recursos de datos donde residan.
5. **Rendimiento general de las consultas a la plataforma:** A medida que los volúmenes de datos crecen, Tableau continúa invirtiendo **en mejoras de rendimiento clave**, que favorecen las conversaciones con los datos en tiempo real. Recientemente, se incluyeron funcionalidades como consultas paralelas, fusión de consultas y almacenamiento en caché externo de consultas. Ahora, Tableau también saca partido de la vectorización para los procesadores que la admiten.
6. **Interfaz visual eficaz y homogénea para los datos:** Tableau cuenta con herramientas de análisis que le permiten filtrar datos, hacer pronósticos y realizar análisis de línea de tendencia con acciones simples. Además, interpreta las acciones del usuario y selecciona la mejor manera de representar los datos de acuerdo a las prácticas recomendadas de visualización. Cuando uno se conecta a los datos, Tableau proporciona una interfaz única para la visualización de estos, que es coherente para todas las fuentes de datos.

Nuestra visión se adapta bien al modo en que evoluciona el panorama general de los datos.

En la actualidad, lo que sucede normalmente es que muchos clientes hacen frente a un conjunto heterogéneo de tecnologías de big data. Tecnologías como Hadoop y Spark se convirtieron en parte de la arquitectura de datos, junto con los almacenes de datos, debido a su capacidad de almacenar y procesar datos. Al mismo tiempo, los clientes corrigen el tamaño de sus almacenes de datos para ajustarlos a sus implementaciones de Hadoop. Frecuentemente, se prefieren las bases de datos NoSQL antes que las bases de datos relacionales para el back-end de las aplicaciones debido a sus modelos de datos flexibles, su baja latencia y su diseño específico para las aplicaciones. Por último, las fuentes de datos en la nube son ubicuas, ya que los sistemas ERP y CRM en la nube se convirtieron en las opciones preferidas para la administración de procesos de negocios, y el modelo de consumo de “pago por uso” cada vez resulta más popular para el almacenamiento en la nube y el procesamiento de datos. Con back-ends tan diversos y flexibles, el usuario necesita una herramienta de front-end como Tableau para obtener la flexibilidad necesaria para conectarse a diferentes plataformas de big data, fuentes de datos en la nube y bases de datos relacionales, y disponer así de las capacidades necesarias para analizar sus datos.

## ¿Cómo trabaja Tableau con los big data?

En el corazón de Tableau se encuentran VizQL y el motor de datos. VizQL es una tecnología patentada que permite al usuario crear una representación visual de sus datos de inmediato y le proporciona una respuesta visual instantánea. Con VizQL, los usuarios tienen una única herramienta de visualización que sirve para producir una amplia gama de resúmenes gráficos, como gráficos de barras, gráficos de líneas y mapas, con todas las acciones. Por otro lado, el motor de datos es una representación de los datos comprimida, en memoria y basada en columnas, que se integra en la tecnología de “conexión en tiempo real” de Tableau. Esta tecnología envía a la base de datos consultas SQL perfectamente optimizadas y específicas para la plataforma, lo que permite a Tableau visualizar enormes cantidades de datos en tiempo real, sin necesidad de mover los datos.

En las siguientes secciones, analizaremos cómo trabaja Tableau con los big data con respecto al acceso a los datos y la seguridad de estos, y cómo habilita otras funcionalidades específicas en Hive.

### **Acceso a los datos**

La clave para trabajar con los big data reside en contar con un modelo de conexión elegante. Nuestros conectores reconocidos para big data se basan en el protocolo ODBC y sacan partido de las funciones específicas para bases de datos mediante la optimización de las consultas SQL que enviamos:

### **Conexiones basadas en SQL**

Tableau interactúa con Hadoop, bases de datos NoSQL y Spark mediante SQL. El SQL que genera Tableau está estandarizado de acuerdo con el estándar ANSI SQL-92. Usar SQL es conveniente porque es extremadamente compacto (una sola expresión). Está estandarizado y tiene código abierto. No depende de bibliotecas y es muy completo y expresivo. Por ejemplo, con SQL se pueden expresar funciones, criterios, resúmenes, agrupaciones, operaciones de unión y operaciones de anidado.

### **ODBC**

Tableau usa controladores que aprovechan el estándar de programación de conectividad abierta de bases de datos (ODBC) en calidad de capa traductora entre interfaces de datos SQL y de tipo SQL proporcionadas por las plataformas de big data. Para Hadoop, esto incluye interfaces como el lenguaje de consulta de Hive (HiveQL), Impala SQL, BigSQL y Spark SQL. Para lograr el mejor rendimiento posible, optimizamos de manera personalizada el SQL que generamos y trasladamos las agregaciones, los filtros y otras operaciones de SQL a las plataformas de big data.

### **Interfaces NoSQL**

Como lo indica su nombre, las bases de datos NoSQL (“no solo SQL”) pueden contener datos modelados en formatos no relacionales, además de los formatos relacionales. También pueden admitir interfaces de tipo SQL. Hoy en día, Tableau admite los conectores de MarkLogic y DataStax Enterprise como conectores reconocidos que usan interfaces de tipo SQL. Con MarkLogic, podemos conectarnos a textos completos o hacer búsquedas complejas en conjuntos de datos relacionales heredados o en datos estructurados. Con DataStax Enterprise y Cassandra, se admite una interfaz Hive ODBC, que saca partido de HiveQL para acceder al almacén de filas particionadas de Cassandra.

## Seguridad de datos

Solo se pueden implementar análisis visuales independientes y de autoservicio en el nivel organizativo cuando se tratan los aspectos de seguridad, como la autenticación y el acceso a los datos. Actualmente, estamos trabajando en múltiples lanzamientos para permitir el acceso seguro a los datos de fuentes de datos relacionadas con big data.

Ofrecemos **autenticación Kerberos** o LDAP para que los usuarios de Tableau Desktop puedan conectarse de manera segura a clústeres de HiveServer2 con Cloudera Hadoop, Hortonworks Hadoop o MapR Hadoop. Además, **admitimos el inicio de sesión único y acceso delegado con Kerberos** para cuando desee conectarse a Cloudera Impala. Esto amplía la compatibilidad anterior que hacía uso de Active Directory, SAML y el sistema de autenticación incorporado en Tableau. Para los usuarios, esto implica una mejor experiencia porque, si iniciaron sesión en su equipo local, no deberán volver a iniciar sesión en Tableau Server ni en ninguna fuente de datos en tiempo real de Impala. Para los administradores de TI, la compatibilidad de Tableau con Apache Sentry garantiza la protección de los datos confidenciales, ya que los usuarios solo verán los datos que estén autorizados a ver. Gracias al esfuerzo conjunto de Tableau y Cloudera para admitir la delegación de usuarios para **Impala**, somos capaces de garantizar que los usuarios puedan conectarse a Impala como a una fuente de datos en tiempo real, a través de autenticación de back-end automatizada y estable. En el futuro, esperamos ampliar el inicio de sesión único y el acceso delegado con Kerberos a muchas otras fuentes de datos.

## Funcionalidades especiales de Hadoop Hive

Hadoop se convirtió en una tecnología de big data casi ubicua. Además, amplía en gran medida el procesamiento de datos que se puede realizar en la capa de almacenamiento, en comparación con las bases de datos tradicionales. En consecuencia, Tableau brinda una variedad de funcionalidades únicas para la conexión con **Hadoop Hive**. Entre estas funcionalidades se incluyen:

- **Procesamiento de XML:** Tableau proporciona numerosas funciones definidas por los usuarios (UDF) para el procesamiento de datos XML con **XPath**. Estas funciones permiten a los usuarios extraer contenido, realizar análisis simples y filtrar los datos XML.
- **Procesamiento web y de textos:** además de los operadores XPath, el lenguaje de consulta Hive ofrece **varias formas** de trabajar con datos de texto y elementos web comunes, entre las que se incluyen:
  - **Objetos JSON:** permite recuperar elementos de datos a partir de cadenas que contienen objetos JSON.
  - **URL:** permite extraer componentes de una dirección URL, como el tipo de protocolo o el nombre del host, o recuperar el valor asociado con cierta clave de consulta de una lista de parámetros de claves/valores.
  - **Datos de texto:** permite encontrar y reemplazar texto de Hive desde Tableau.

- **Extracciones, transformaciones y cargas (ETL) al instante:** el SQL personalizado permite a los usuarios definir sus conexiones de datos con filtrado previo, agregación previa y condiciones de unión complejas.
- **SQL inicial:** el SQL inicial permite a los usuarios especificar un conjunto de declaraciones SQL que deben ejecutarse inmediatamente después de que se establezca una conexión de datos. Normalmente, esto se hace para optimizar características de rendimiento o desarrollar una lógica de procesamiento de datos personalizada.
- **Análisis personalizado con funciones UDF y MapReduce:** Tableau permite implementar funciones UDF, funciones agregadas definidas por usuarios (UDAF) y expresiones SQL arbitrarias de Hive mediante el uso de funciones “de paso”. Generalmente, estas funciones se diseñan como archivos de Java (JAR), que pueden copiarse en el clúster de Hadoop. Los usuarios también pueden ejercer un control explícito sobre la ejecución de operaciones de MapReduce en SQL personalizado.

## Casos prácticos: Tableau y los big data

Las organizaciones que recién comiencen a usar big data descubrirán que existen dos casos principales para usar Tableau y sus activos de datos: la exploración de datos y la visualización de datos.

### Exploración de datos

Las organizaciones recopilan y almacenan toda clase de datos, frecuentemente sin ningún análisis previamente definido, con la esperanza de que esos datos proporcionen información útil en el futuro. Los datos, como los registros web, los registros de servidor, las secuencias de clics, la información de sensor y la información de redes sociales, se recopilan en plataformas de datos, como Hadoop, en vez de descartarse. Este tipo de enfoque proporciona flexibilidad y alienta la experimentación en el análisis. Tableau puede explorar y visualizar las tendencias generales de los datos antes de comprometer recursos significativos en el procesamiento exhaustivo de todos los datos.

En EMC, se usa Tableau para extraer los datos del sensor de medición de energía en Hadoop. El ingeniero de soluciones Tom Hudgins comparte lo siguiente: “Tengo aproximadamente 70 mil millones de filas en la base de datos que analizamos con Tableau. Usamos Tableau para analizar datos de medidores inteligentes: información proveniente de los hogares y las empresas acerca del uso de la energía. Las empresas que, al encontrarse con una montaña de información, puedan detectar los detalles relevantes que conducen al descubrimiento, el conocimiento y otros factores que no habían considerado antes serán las que tengan más posibilidades de alcanzar el éxito y evadir el fracaso”.

### Visualización de datos

El enfoque de una organización sobre la visualización de datos debería centrarse en la optimización del rendimiento, después de decidir qué conjunto de análisis debería habilitarse o procesarse exhaustivamente. Encontrar el equilibrio entre la generalidad y el nivel de detalle de los big data

es fundamental para propiciar los dashboards que respondan adecuadamente y las conversaciones en tiempo real con los datos. Tanto los administradores de TI como los usuarios de Tableau tienen herramientas a su disposición para dar a sus datos el nivel de detalle adecuado. Además, seleccionar una plataforma de procesamiento de datos que admita el análisis interactivo es igualmente importante. Entre las prácticas recomendadas para maximizar el rendimiento se incluyen:

- El aprovechamiento de un rápido motor de consultas interactivas;
- La personalización del rendimiento de su conexión para consultas en tiempo real;
- La optimización de sus extracciones mediante resumen, filtrado y muestreo;
- El uso de prácticas recomendadas relacionadas con bases de datos, como el diseño de particiones.

El equipo de Rosenblatt Securities es un ejemplo de una organización que realizó iteraciones y optimizó su enfoque de big data. “Con un equipo de solo 5 personas, hicimos cosas con Tableau que hubieran llevado una cantidad significativa de tiempo a 50 personas”, afirmó Scott Burrill, socio y director ejecutivo. “Hacemos análisis de predicción para 800 valores con el fin de determinar, en tiempo real, si nos encontramos en condiciones de entrar en el mercado o debemos salir de él. Podemos hacer análisis derivados de cientos de miles de campos diferentes y podemos tomar esos análisis, visualizarlos, obtener información, actuar en consecuencia y contar historias muy rápidamente. Tableau abrió un nuevo mundo para nosotros, ya que podemos actuar basándonos en información. Antes, teníamos que sacar muestras; ahora, podemos ver poblaciones de datos completas”.

## Resumen

Llegó la era de los big data: los volúmenes de datos se incrementan y las organizaciones llevan sus infraestructuras de datos a Hadoop, Spark, NoSQL y rápidas bases de datos de análisis para admitir los nuevos datos estándares. Gracias a Tableau y su capacidad de desarrollar el potencial de usuarios de negocios comunes, como los de EMC y los de Rosenblatt Securities, la información visual de los big data está llegando a todas las personas.

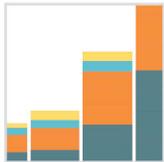
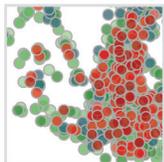
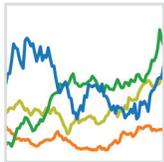
## Acerca del autor

### **Jeff Feng - Administrador de productos de Tableau Software (@jtfeng)**

Jeff Feng es administrador de productos en Tableau Software. Se encarga de la hoja de ruta, las estrategias y el desarrollo de nuevas funcionalidades de los productos relacionados con big data, para transformar la manera en que las personas trabajan con los datos. Antes de unirse a Tableau, Jeff fue asesor de administración en McKinsey & Co., donde prestaba asesoramiento a empresas de alta tecnología (incluidas en la lista Fortune 500) acerca de sus estrategias de negocios, tecnología y productos. También fue administrador de programas en Apple, donde participó en el lanzamiento del iPhone 4. Jeff tiene una maestría (MBA) en administración de empresas de la Escuela de Administración y Dirección de Empresas Sloan del MIT. Asimismo, obtuvo una maestría (MS) y una licenciatura (BS) en ingeniería electrónica de la Universidad de Illinois en Urbana-Champaign.

## Acerca de Tableau

Tableau ayuda a las personas a ver y comprender datos. Tableau ayuda a todas las personas a analizar, visualizar y compartir información rápidamente. Más de 26 000 cuentas de clientes obtienen resultados rápidos con Tableau, en la oficina o en cualquier otro lugar. Además, decenas de miles de personas usan Tableau Public para compartir datos en sus blogs y sitios web. Para ver la forma en que Tableau puede ayudarlo, descargue la versión de prueba gratuita en [www.tableau.com/es-es/trial](http://www.tableau.com/es-es/trial).



### Informes relacionados

5 prácticas recomendadas para Tableau y Hadoop

7 sugerencias para tener éxito con los big data

Fostering a Data-Driven Culture:

A Special Report from The Economist Intelligence Unit and Tableau (Creación de una cultura impulsada por los datos: un informe especial de Economist Intelligence Unit y Tableau)

Big Data: The Next Industrial Revolution (Big data: la próxima revolución industrial)

Tableau Software y los big data

Aberdeen Group: Maximizing the Value of Analytics and Big Data (Aberdeen Group: cómo maximizar el valor del análisis y de los big data)

Ver todos los informes

### Recursos adicionales

- Descargar versión de prueba gratuita
- Demostración del producto
- Capacitación y tutoriales
- Comunidad y soporte
- Historias de clientes
- Soluciones