

Cómo sacarle provecho a la plataforma Hadoop

Cloudera es una compañía con presencia en 20 países, proveedora de una plataforma de análisis y gestión de datos construida sobre Hadoop. A ella acuden 800 empresas usuarias para resolver distintas necesidades de las compañías.

En la era del *boom* del *big data*, las empresas sacan mayor provecho de la información para múltiples propósitos: desde poner a funcionar un servicio de taxis o una compañía de infraestructura o de seguridad gubernamental, hasta investigar distintas ciencias.

En esa tarea les ayudan empresas como Cloudera, que funciona en Palo Alto (California). Fue creada por miembros de Oracle, Yahoo!, Google y Facebook. Nació del proyecto de código abierto Hadoop, una plataforma de almacenamiento ilimitado y análisis de datos para multipropósito, fundada por Doug Cutting. Su misión es lograr que sus usuarios puedan sacarle mayor provecho a su información. Opera en 20 países y tiene 800 suscripciones de

clientes. Una de sus ramas importantes es la capacitación a través de cursos y talleres sobre cómo usar esta plataforma.

Mediante el uso de Hadoop ayudan a empresas como Uber, Google, Facebook, Ali Baba y Airbnb a aprovechar ampliamente los datos y para ello se apoyan en los más grandes computadores existentes para instrumentación, personalización y analítica avanzada.

Así lo explicó Joao Salcedo, ingeniero de sistemas Latam de Cloudera, quien dijo que *big data* existe ahora porque vivimos una explosión de información desde el momento en que nos levantamos y publicamos cosas en Facebook o pagamos impuestos. “Con esa estructura más instrumentada podemos empezar a experimentar más, y con un motor de recomen-

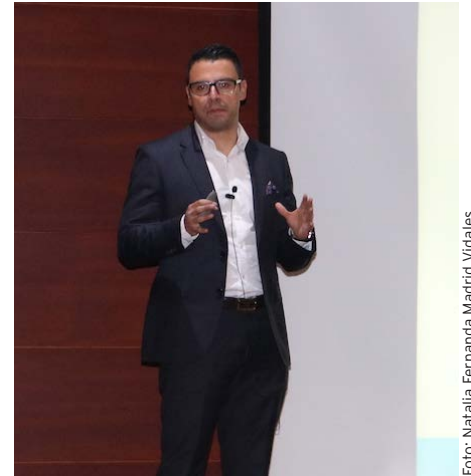
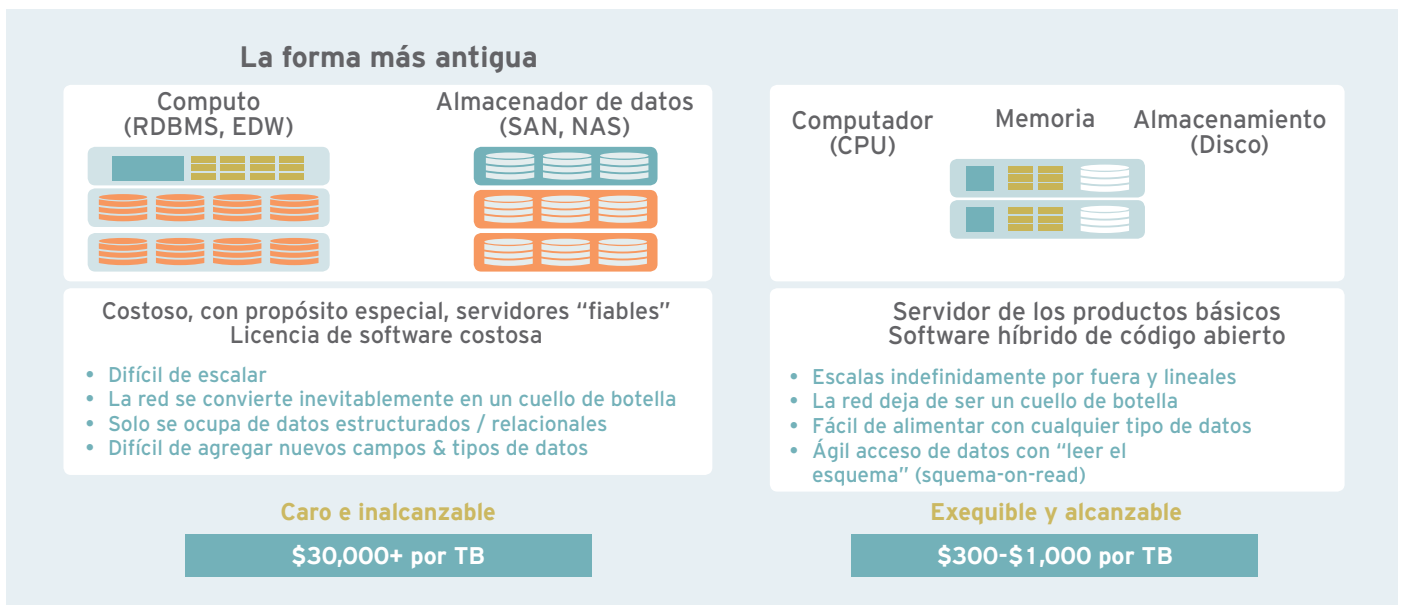


Foto: Natalia Fernanda Madrid Vidales

Joao Salcedo tiene una maestría en Ingeniería de Sistemas y Computación de Royal Melbourne Institute of Technology (Australia).

daciones se puede comenzar a monetizar ese conocimiento”, agregó.

En unión con Hadoop, Cloudera ayuda a generar esos usos de información personalizada con un equipo con conocimientos de estadística, programación y del negocio, que administra la plataforma de la



Why Hadoop vs Data Warehouse
Cuadro presentado por Joao Salcedo, ingeniero de sistemas Latam de Cloudera.

“Mediante el uso de Hadoop ayudan a empresas como Uber, Google, Facebook, Ali Baba y Airbnb a aprovechar ampliamente los datos”.

organización. Este grupo de especialistas debe saber correlacionar esos tres mundos y generar valor.

Ventajas de usar Hadoop para manejar datos

Joao Salcedo enumeró las siguientes razones por las cuales recomienda el uso de Hadoop:

- Las arquitecturas para manejar grandes volúmenes de información son insuficientes y no son flexibles. Hadoop es una plataforma más económica, *open source* (o de código abierto) y es fácil ingresar cualquier tipo de datos.

- Hadoop se usa, por ejemplo, en servicios financieros y en el Gobierno para predecir fraudes y riesgos; en telecomunicaciones para optimizar el trabajo de las redes, y en manufacturas para prever necesidades de mantenimiento, entre otros.

- Cuenta con *Data Warehouse Optimization*: cada compañía tiene la posibilidad de hacer que sus procesos operativos de información se integren en Hadoop donde esta información puede ser procesada a gran escala en procesos complejos de análisis. Hay un equipo de administración de la plataforma. ■

Por una tecnología más libre

Desarrollador de tecnologías de recuperación de información y motores de búsqueda, el ingeniero Doug Cutting ha sido promotor y creador de proyectos de software libre como Lucene, Nutch, Avro y Hadoop. Actualmente es el arquitecto jefe de Cloudera y fue el invitado especial al Foro.

Hace diez años, era difícil pensar que las empresas aceptarían usar una plataforma *open source* en tecnología para guardar datos, pero, gradualmente, se logró que la adoptaran porque es más barata, flexible y fácil de escalar. Así lo recordó Doug Cutting, un abanderado del código abierto, en su charla *The Future of Big Data and Why I Created Hadoop* (El futuro de *big data* y por qué creé Hadoop).

Cutting siempre quiso producir algo que se pudiera aprovechar sin pagar, sin licencias, y ayudar a solucionar diversos problemas. Le gustaba escribir software y que la gente lo usara porque descubrió que era una forma extraordinaria de trabajar: podía utilizar el mismo *software* en distintas compañías y, además, con este sistema de interacción con las personas, ellas lo ayudaban a mejorarlo cada vez más.

Su trayectoria incluye trabajos en varias compañías en el Silicon Valley. Duran-



Doug Cutting es egresado de la Universidad de Stanford, expresidente de la Junta Directiva de Apache Software Foundation. En el 2015, la empresa editorial estadounidense de libros de tecnología e informática O'Reilly le otorgó un premio de software libre.

te cinco años estuvo en el centro de investigación de Xerox Palo Alto (PARC), donde hizo aproximaciones a la recuperación de información. También integró el equipo de Apple cuando “no era exitoso” y en el de Excite desarrolló el núcleo del sistema de

recuperación de información de este motor de búsqueda.

En su conferencia se refirió especialmente a la creación de los motores de búsqueda, a sistemas para recuperar información, guardar y procesar datos, a partir de los cuales creó Hadoop, una multiplataforma de computación distribuida programada en Java. En ella hay muchos computadores organizados en clústeres, participantes de una infraestructura cooperativa conectada por una red de telecomunicaciones. Con Avro, (un proyecto para la serialización de datos e intercambio para Hadoop, que también desarrolló Cutting), *big data* puede ser desarrollado mediante programas escritos en distintos lenguajes.

Volviendo a la historia de estos desarrollos, Cutting explicó que “en el 2000 no eran una maravilla, pero el software libre de recuperación de información Lucene se volvió el más popular. No era la mejor tecnología, pero la gente la podía adaptar sin