

“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

licencia, podía devolverla a los fabricantes para resolver los problemas y no necesitaba abogados para usarlo”.

En el 2004 trabajó el sistema Nutch, un motor de búsqueda de código abierto basado en Lucene. Pero solo era empleado por académicos, conectado a Google, y la idea era que pudiera usar algo de código abierto. Para apoyar las búsquedas de ese motor, Cutting ideó Hadoop, cuyo nombre proviene del elefante de juguete de su hijo. Comenzó con cinco máquinas y luego 25 que podía dejar corriendo todo el día, pero se perdía información.

En el 2006 ingresó a Yahoo! adonde encontró ingenieros e investigadores con sus mismos intereses, que le proporcionaron más computadoras y un equipo que pasó de tres a doce personas. “Así pude contar con cientos de computadores para poder hacer este *software* viable”. Con su grupo puso a producir un sistema clúster de almacenamiento y análisis basado en Hadoop para las operaciones de misión crítica de negocio de la compañía.

En el 2007 llegó a su meta de tener miles de computadores trabajando en para-

“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 y descubrió que era una forma extraordinaria de trabajar”.

lelo (algo que no se había hecho y era más barato) y un *software* gratis. Antes, la gente pagaba por almacenar datos, por lo que muchas personas comenzaron a buscar el *software* para empresas y fábricas.

En el 2009, se retiró de Yahoo! para vincularse a Cloudera, una plataforma libre con soporte comercial que se emplea no solo para manejar datos, sino para procesar e interpretar los que generan sensores de las máquinas (luego hay que analizarlos para el mejoramiento de la empresa). En esa época era común que las organizaciones usaran *software* de tecnología Enterprise para ejecutar esas tareas, pero esta es muy costosa

e inflexible y a menudo no querían mezclarse con quienes operaban *open source*.

Era el momento en que, recuerda, se diseñaba la aplicación del *software* y luego se alimentaba con datos, lo cual no es productivo; es mejor introducir la información y hacer experimentos, mirar si funciona el programa para luego empezar a usarlo. A esta nueva manera de trabajar Cutting la llama “la transformación digital de los negocios”. Y en ella, los ingenieros que desarrollaron el *open source* pueden ayudar a quienes controlan la evolución del sistema en las empresas. Es una manera de mantener este proyecto constantemente vivo. ■

Big data contra el lavado de activos

Luis Edmundo Suárez, director de la Unidad de Información y Análisis Financiero (UIAF), habló de cómo el manejo de enormes volúmenes de datos les ha permitido mejorar los resultados en sus investigaciones.

La lucha contra el lavado de activos es global porque es un problema que está afectando todas las economías. Según las Naciones Unidas los ingresos de los criminales se acercan a 2.6 trillones de dólares, de los cuales introducen a la economía mundial 1.9 trillones de dólares, una cifra superior al Producto Interno Bruto (PIB) de cualquier país de América Latina.

Para luchar contra este flagelo se han conformado agrupaciones a nivel mundial y regional que comparten información so-

bre redes de tráfico de drogas, armas o terrorismo. Se destaca el grupo Egmont, con sede en París, que regula el tema y congrega unidades de inteligencia financiera de 149 países.

De los 2.6 trillones de ingresos producto de delitos se logra interceptar cerca de 0.20 %, explicó Luis Edmundo Suárez durante el Foro. Es decir 20 centavos de cada 100 dólares. Y en Colombia, de acuerdo con una investigación del Banco Mundial hecha entre 1983 y el 2013, hubo un marcado ascenso de 1999 al 2000 y los ingresos generados por bandas de crimen co-