

EXPERTOS LÍDERES EN MANEJO DE DATOS:

Quiénes son los científicos que sientan las bases de nuestro universo digital

En junio se realizará por primera vez en Chile SIGMOD/PODS, la conferencia más importante del mundo dedicada al estudio y manejo de datos digitales a gran escala. De este evento ha emergido la innovación aplicada que hoy permite que contemos con motores de búsqueda, comercio en línea, almacenamiento en la nube, redes sociales y algoritmos de IA.

Uno de los temas que se investigan en la actualidad es el uso responsable de datos en general, y en particular en IA.

ARTURO CATALÁN A.

Cada vez que una persona utiliza un cajero automático da inicio a una serie de procesos que ocurren a una velocidad vertiginosa. Parten con una consulta o requerimiento en cualquier lugar del país o del mundo; por ejemplo, el retiro de dinero. Esta solicitud llega a una base de datos a través de un sistema que, en segundos, encuentra la información del cliente, determina sus saldos y define si es posible o no que la persona se aleje de ese cajero automático con dinero en efectivo en su billetera.

Esta acción, que se repite miles de millones de veces al día a nivel global, no podría ocurrir a esta escala, ni con tanta velocidad, de no ser por Edgar Codd, matemático y científico de la computación que comenzó a abordar el desafío de cómo hacer consultas de manera eficiente sobre grandes volúmenes de datos. La clave estaba en el modo de organizar la información y su primer estudio sobre el tema lo presentó en un encuentro que reunía a investigadores que compartían, ya en aquel entonces, su preocupación en torno a los datos digitales.

¿El año? 1970. ¿El encuentro? Una de las primeras versiones de la conferencia que hoy se conoce como SIGMOD/PODS, la más importante a nivel internacional en el área de estudio y manejo de datos digitales a gran escala y que por primera vez en su historia se realiza en América Latina,

siendo Chile la sede de este evento que, en junio de 2024, reunirá a más de 800 científicos de todo el mundo. Durante cinco días de trabajo intensivo, abordarán a nivel teórico y aplicado los problemas actuales y emergentes de los grandes volúmenes de datos, en áreas tan diversas como blockchain, información en la nube, redes sociales e inteligencia artificial, entre muchas más.

"No es por azar que se escogió a Chile como sede. Es una señal del liderazgo del país en este campo, conseguido por una comunidad de académicos y académicas reconocidos a nivel global por sus aportes, e investigadores jóvenes que ya están siendo premiados por sus trabajos. Es desde Chile de donde han surgido investigaciones que hoy posibilitan, por ejemplo, extraer información que permite entender el comportamiento en redes sociales en tan solo segundos", señala Pablo Barceló, director del Instituto de Ingeniería Matemática y Computacional de la Universidad Católica (IMC UC), investigador del Instituto Milenio Fundamentos de los Datos (IMFD) y del Centro Nacional de Inteligencia Artificial (Cenia), codirector del comité organizador de SIGMOD/PODS 2024 y uno de los artífices de que esta conferencia se haga en Chile.

Nayat Sánchez, directora de Iria Chile y codirectora del comité organizador de SIGMOD/PODS 2024, enfatiza que el esfuerzo de traer esta conferencia a Chile no ha sido menor: "Que lo hayamos logrado es un reconocimiento al trabajo de científicos que son considerados como líderes en la región y en el mundo. Ese prestigio ganado por años ha he-



Nayat Sánchez, directora de Iria Chile.



Pablo Barceló, director de IMC UC.

cho que nuestro país sea sede y es un hito que nos permitirá impulsar aún más el posicionamiento de Chile en datos e inteligencia artificial".

Los expertos que llegan al país no solo se dedican a la investigación en universidades y centros del más alto prestigio, también trabajan para compañías tecnológicas que siguen de cerca el avance en el estudio de sistemas de bases de datos. Entre ellas se encuentran Apple, Huawei y Google, que además son sponsors oficiales de la conferencia.

LOS "NOBEL" DE LA COMPUTACIÓN

Las investigaciones de Edgar Codd en los 70 abordaron lo que él bautizó como bases de datos rela-

cionales, sistemas que fueron creando y evolucionando para llegar a ser aplicados en prácticamente todas las industrias y sectores que manejan grandes volúmenes de información: la banca, los sistemas de compras en línea, los registros de salud, la gestión de inventarios en el retail, y muchísimos más. En el mundo del manejo de datos en grandes volúmenes hay un antes y un después de Codd, y el impacto de su investigación le hizo merecedor en 1981 del Turing Award, también conocido como el "Nobel" de la computación.

Pero una década después, cuando ya la cantidad de datos almacenados crecía sin parar, emergió un nuevo problema. Cuando una persona está haciendo una transferencia bancaria desde su celular y repentinamente se

quedó sin batería, ¿qué pasa con el dinero? Jim Gray, en los 80, comenzó a investigar en torno a una premisa que pareciera sencilla, pero que tecnológicamente implicaba un desafío colosal: cualquier movimiento de datos, como transferir dinero de una cuenta bancaria a otra, debe garantizar la integridad de la base en la que se alberga la información. Es decir, independiente de lo que ocurra (como un corte de energía), el dinero o se mueve a la cuenta de destino o se queda en la cuenta original. Al sistema que garantiza que esto ocurra lo bautizó como R y, al igual que sucedió con los trabajos de Codd, siguió creciendo y evolucionando, estando en el origen de sistemas de bases de datos como Oracle, Microsoft SQL Server, PostgreSQL y MySQL para manejar información que requiere actualización permanente, preservando la seguridad e integridad de los datos. Por este hito, Jim Gray recibió el Turing Award en 1998.

El último laureado fue Michael Stonebraker, que ahondó en el área de bases de datos relacionales creando nuevos sistemas de manejo que revolucionaron el mercado, especialmente con Postgres, un modelo que permitió, por una parte, trabajar con datos más complejos, diversos e interconectados, y que creaban sostenidamente; y, por otra, hacer nuevos tipos de consulta a la base de datos. Este sistema de bases de datos ha sido y está siendo utilizado por empresas y servicios como Instagram, TripAdvisor, Uber y Spotify, entre muchos otros. Por su innovación en este terreno, Stonebraker recibió el Turing Award en 2015.

DATA FAIRNESS

En los 90, la aparición de la World Wide Web hizo crecer los usuarios y servicios de internet de manera imparable, lo que generó nueva investigación en torno a los mecanismos para intercambiar y extraer información desde datos a gran escala. En este campo fue decisivo el trabajo del mexicano Héctor García-Molina, académico de la U. de Stanford y mentor de dos estudiantes de doctorado cuyo proyecto es un pilar del universo digital como lo conocemos hoy: Sergei Brin y Larry Page, creadores de Google.

"Los 2000 es la década en la que se acuña el término *big data* y se comienzan a estudiar sus desafíos, como el volumen, la velocidad de crecimiento y la variedad en los tipos de datos. Esto impulsa el trabajo en nuevos métodos para el uso de datos en áreas como la astronomía o la genómica", explica Pablo Barceló.

Desde 2010 en adelante, la irrupción de poderosos algoritmos de inteligencia artificial (IA), que utilizan como base grandes repositorios de datos, ha hecho que parte de la comunidad de científicos que se reúnen en torno a SIGMOD/PODS se esté enfocando en el gran reto del sesgo en la IA.

"Se ha hecho urgente la necesidad de desarrollar investigación en el uso de datos y en IA de manera responsable, que se haga cargo de los riesgos que tiene la manipulación de grandes volúmenes de información para tomar decisiones. Muchos de los trabajos presentados en SIGMOD/PODS abordan de manera teórica y aplicada estos desafíos", concluye Nayat Sánchez.

EL MERCURIO

