

Procesar enormes cantidades de datos exige máquinas poderosas que no suelen estar a disposición de grupos de investigación pequeños o medianos, pues carecen de presupuesto para montar y mantener un clúster que distribuya las cargas de trabajo en muchos equipos de cómputo. Y si lograran comprarlo, estaría subutilizado porque solo lo requieren por períodos cortos.

Para enfrentar esa dificultad, cada vez toma más fuerza la alternativa de software como servicio (SaaS), donde las aplicaciones científicas están en la nube y cualquier investigador puede arrendarlas cuando las requiere y pagar solo por lo que usa.

Los ingenieros Mario Villamizar, profesor del DISC (Uniandes), y Fernando Barraza, fundador de Omicsco (firma de desarrollo de software para ciencias de la vida), explicaron cómo sacarle provecho a esta solución.

#### e-Clouds, investigación de Los Andes

El profesor Villamizar centró su conferencia en e-Clouds, un proyecto de *cloud computing* para Bioinformática que se desarrolla en la Universidad. El fin es crear una nueva arquitectura que funcione en nubes públicas o privadas, donde los científicos encuentren aplicaciones listas para usar de forma fácil y ágil.



Fernando Barraza, fundador de Omicsco.

“En todo el proceso, sin que el investigador lo note, e-Clouds se encarga de ejecutar de forma transparente todas las tareas”.

A e-Clouds los investigadores entran a través de un portal Web, seleccionan las aplicaciones, configuran las tareas, indican los recursos de cómputo que necesitan, aprueban el costo y lanzan las tareas. Una vez se envían a ejecución, pueden monitorearlas y, al final, descargar los resultados y pagar por los recursos consumidos. En todo el proceso, sin que el investigador lo note, e-Clouds se encarga de ejecutar de forma transparente todas las tareas sobre los servicios y recursos de cómputo de un proveedor de nube privada o pública.

Para diseñar la solución se han considerado aspectos como que los servidores no pueden estar prendidos siempre debido a los costos asociados, que las aplicaciones se instalen bajo demanda, que los sistemas de almacenamiento sean altamente escalables, que se pueda medir y cobrar por el uso de los recursos, y que la solución sea segura.

Actualmente el proyecto se lleva a cabo en la plataforma de Amazon Web Services, pero se está trabajando en integrar e-Clouds a otras nubes públicas y privadas. Por el momento, no soporta la ejecución de aplicaciones paralelas y se están haciendo pruebas de usabilidad para facilitar su adopción en la comunidad científica colombiana. Los usuarios interesados en probar la solución pueden inscribirse en <http://ecloudshomepage.herokuapp.com>

#### Software para ciencias de la vida

El ingeniero de sistemas Fernando Barraza, fundador de Omicsco, se enfocó en

cómo proveer un software adecuado a las necesidades de los investigadores en Ciencias Biológicas que pueda usarse en la nube.

Entre ellos están:

- La información ha crecido enormemente gracias a la secuenciación genómica. El software como servicio (SaaS) busca resolver el problema de identificar únicamente los datos.
- Hay que conocer las necesidades del usuario y la jerga de los investigadores biólogos. Se deben responder las preguntas biológicas, construir aplicaciones a través de algoritmos y manejar técnicas computacionales. Las soluciones tienen que ser amigables, seguras, funcionales con múltiples fuentes de datos, debidamente soportadas y fáciles de instalar y mantener.
- Los desarrollos deben dirigirse a un mercado global. Hay que preguntarse por el modelo de negocio e identificar a quienes pueden ayudar a desarrollar las ideas.
- Es fundamental aclarar las preocupaciones de los investigadores sobre seguridad, privacidad y pérdida de control sobre los datos. ■



Mario Villamizar, profesor del DISC.

#### Más información

Mario Villamizar, ingeniero de sistemas, magíster en Ingeniería de Sistemas y Computación de la Universidad de los Andes, profesor y experto temático en cloud computing del DISC. [mj.villamizar24@uniandes.edu.co](mailto:mj.villamizar24@uniandes.edu.co)

## Descifrando los genomas del café y del arroz

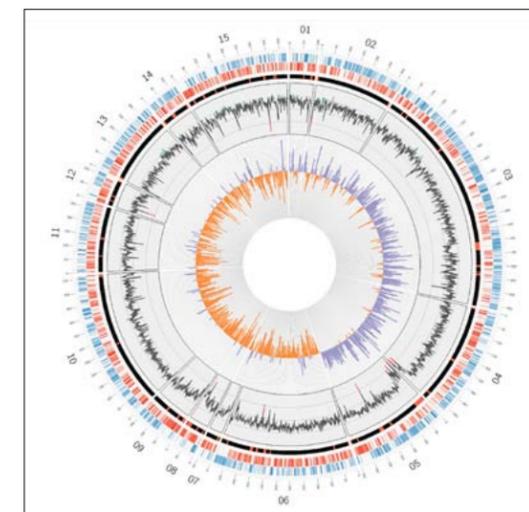
Dos expositores mostraron trabajos para encontrar genes de estos cultivos en Colombia. Estos pueden ser útiles para mejorar la productividad de las plantaciones y contribuir así a la seguridad alimentaria.

Al combinar los estudios de Genómica con Bioinformática se pueden mejorar las características agronómicas de plantas como el café y el arroz para hacerlas más productivas y resistentes a las plagas. Sobre este tipo de trabajo versaron las conferencias de Marco Aurelio Crisancho y Jorge Duitama, investigadores de Cenicafé en Chinchiná (Caldas) y del Centro Internacional de Agricultura Tropical (CIAT) en el Valle del Cauca.

En su exposición, Crisancho mostró resultados en secuenciación de genomas del insecto que produce la broca del café, uno de cuyos hallazgos patentaron Cenicafé y



En Cenicafé llevan a cabo estudios genéticos de la broca y la roya del café para mejorar la resistencia de las plantas a estas plagas.



Genoma de una bacteria patógena de plantas. Este consta de 5 millones de letras (nucleótidos) y tiene la información necesaria para producir 4000 proteínas diferentes. Los estudios de genómica permiten determinar en qué lugares está definida cada propiedad biológica de un organismo.

Además, se refirió a las investigaciones tendientes a diferenciar el café colombiano de otros en el mundo utilizando las tecnologías para hacer trazabilidad en la cadena productiva. Pretenden obtener una huella química y ojalá genética para que los consumidores tengan

la Federación Nacional de Cafeteros, pues podría aplicarse en la degradación de los azúcares de los desperdicios generados por la industria cafetera. También mencionó estudios con la Facultad de Ciencias de Los Andes acerca de la roya, un patógeno que produce la principal enfermedad de estos cultivos.

certeza de que el producto que reciben no solo es 100 % nacional, sino que conozcan la región de origen.

#### Devolverle los resultados al campo

El ingeniero Jorge Duitama destacó el papel de los bioinformáticos en las investigaciones sobre mejoramiento genético de

“Hay que desarrollar nuevo software porque las herramientas de apoyo a los estudiantes de Biología tienen problemas de calidad, de código y de soporte y la integración es muy difícil”.

Jorge Duitama

las plantas, pues es necesario desarrollar nuevo software porque las herramientas de apoyo a los estudiantes de Biología tienen problemas de calidad, de código y de soporte y la integración es muy difícil. Relató que en el CIAT idearon una solución para remediar esas deficiencias, que está publicada y es gratuita.

También expuso resultados de los estudios colaborativos con *Louisiana State University* para secuenciar genomas de las



Marco Aurelio Cristancho, investigador de Cenicafé, y Jorge Duitama, investigador del CIAT.

variedades de arroz índica (la que más se cultiva en Latinoamérica incluida Colombia) y japónica, con el fin de acelerar los procesos de mejoramiento vegetal.

El objetivo, enfatizó, es devolverle los resultados al campo con variedades mejoradas para que las plantas de arroz o yuca sean más resistentes a las enfermedades y dentro de 30 años se pueda generar alimento para los 9000 millones de habitantes que tendrá el planeta. ■

### Apoyos para transferir el conocimiento

El Centro de Bioinformática y Biología Computacional de Colombia (BIOS) y el Programa de Transformación Productiva del Ministerio de Comercio, Industria y Turismo son estrategias al servicio de los investigadores en Bioinformática.



En Manizales, BIOS está instalando un laboratorio de visualización en 3D con herramientas de Bioinformática y Biología Computacional. Permitirá, por ejemplo, hacer simulaciones en Biología de Sistemas o Dinámica Molecular. En la imagen, el Centro de Monitoreo.

Mauricio Rodríguez, director de BIOS, explicó que ese centro de desarrollo tecnológico quiere articular y acelerar la transferencia de conocimiento académico hacia el sector productivo. Para ello les ofrece gran capacidad computacional y desarrolla y optimiza soluciones tecnológicas como nuevos algoritmos, programas de software y plataformas de análisis, integración, almacenamiento y visualización de datos.

A su vez, Alejandro Delgado, gerente del Sector de Software y Tecnología de la Información del Programa de Transformación Productiva del MinComercio, señaló que la industria nacional de TI es muy dispersa y por su tamaño pequeño se le dificulta obtener financiación.

Para enfrentar esos retos, el Gobierno está creando clústers regionales para que, con trabajo cooperativo, las empresas brin-

den servicios más robustos. Estos se ubican en el Caribe, el Pacífico, Antioquia, Bogotá, Bucaramanga, Pasto y el Eje Cafetero (se enfoca en Bioinformática, por el momento para el agro y el sector cosmético).

Al mismo tiempo, impulsó la Ley de Garantías Mobiliarias, que facilita el acceso a créditos bancarios a las compañías de software y de TI, al permitirles ofrecer como garantía cualquier capital intelectual que pueda ser vendido o comprado en el mercado.



Mauricio Rodríguez, director general de BIOS, y Alejandro Delgado, de MinComercio.

## El amor por la ciencia debe fomentarse en la infancia

Los conferencistas del Foro de Bioinformática respondieron preguntas del público durante el panel que cerró la jornada. Foros ISIS sintetiza algunas de ellas.



*Cómo evitar el problema de las falsas conclusiones en los datos masivos?*

**María Mercedes Zambrano, Corpogen**

Una solución es que las empresas liberen los contenidos de sus bases de datos incluso si los experimentos no han sido exitosos. También debemos entrenar personas con mayor capacidad para interpretar la información con análisis muy específicos. Una cosa es manejarla y presentarla y otra entender su significado.

### Diego Riaño, Laboratorio de Ciencia y Tecnología de Bioetanol (Brasil)

En la última década, se volvió muy barato generar datos en Biología y eso es una revolución gigantesca. Antes éramos *data* pobres, ahora somos *data* ricos, pero no sabemos qué hacer con la información, la velocidad de producirla es mucho mayor que la capacidad de analizarla y en el afán de sacar conclusiones se cometen errores. El problema debe atacarse desde el colegio porque debemos generar conocimiento que nos ayude a vivir mejor.

### Fernando Barraza, Omicsco

Para abordar el problema se ha tratado de usar las tecnologías de la web semántica que buscan implementar mecanismos de inteligencia artificial y utilizar las ontologías y deontologías para tener representaciones únicas y entendimientos químicos de las bases de datos y hacerlas interoperables. Hay proyectos muy prometedores, por ejemplo en cáncer, no para saber qué tan ciertos son los datos, sino para mejorar la confianza que se puede tener sobre ellos.

### Jorge Duitama, CIAT

La calidad de los datos no puede manejarse como antes, uno por uno, porque es inviable económicamente. La estadística ayuda a producir conclusiones y a validar si son correctas. Siempre que se extrae información en un proceso biológico hay errores

y aunque no pueden encontrarse todos, sí hay forma de cuantificarlos, de medir la calidad en cada paso. Los ingenieros de sistemas debemos entender la Biología, pues a veces basta con comprender 2 o 3 conceptos básicos para darnos cuenta de si lo que estamos entregando tiene sentido.

### Silvia Restrepo, Facultad de Ciencias Uniandes

En la Maestría en Biología Computacional pretendemos que el biólogo y el ingeniero entiendan las herramientas. A veces, el estudiante se lanza con una de ellas como una caja negra, sin saber qué hay detrás, y nuestra labor de formadores es decirles: deténganse a ver qué están haciendo, pues así pueden interpretar mejor los datos.



Familiarizar a los niños con la ciencia es una prioridad.

“El reto en Colombia está en formar esas mentes inquietas no solo en posgrado, sino en pregrado e incluso en primaria”.

**María Mercedes Zambrano, Corpogen**

dos, económicos, agrícolas, con café, con caña, y así pueden formular un programa de desarrollo que empodere a los científicos locales para resolverlos. En nuestro currículo, en vez de enseñarles a los estudiantes las herramientas, tratamos de que aprendan tres cosas: adquirir tecnología, evaluarla y aplicarla a los problemas.

### Mauricio Rodríguez, BIOS

Muchos han tomado cursos de bioinformática y saben usar las herramientas, pero no han aprendido a mover la frontera del conocimiento, a generar nueva tecnología en Biología Computacional. Por eso, debemos ofrecerles a los estudiantes una opción de mercado, de desarrollo profesional. Una dificultad de formar bioinformáticos es que los biólogos no quieren estar sentados frente a un computador, sino en el campo, y no hay muchos ingenieros inclinados hacia la investigación científica; más bien quieren hacer un Tappsi o un Facebook, volverse millonarios de la noche a la mañana, y sueñan con fundar el próximo Silicon Valley aquí. Es a esa gente excepcional a la que debemos nutrir y llevar de la mano para que sean profesionales excelentes.

### María Mercedes Zambrano, Corpogen

Queremos promover que la gente se cuestione sobre el entorno y busque las herramientas bioinformáticas para solucionar problemas. El reto de Colombia es formar esas mentes inquietas no solo en posgrado, sino en pregrado e incluso en primaria.

*¿Cómo lograr que la información compleja sea asequible para los estudiantes de colegio y así fomentar su interés por la investigación?*

### Fernando Barraza, Omicsco

Para enfrentar el problema de la gran cantidad de información, deberíamos infundir en nuestros pequeños la cultura del *hacking*, no como quien rompe un sistema tecnológico para hacer algo malo, sino como el que está inquieto, no traga entero y quiere descubrir más allá. Esa es la filosofía que necesitamos para formar a los líderes, a los transformadores. ■

*¿Qué consejos puede darnos para fortalecer la investigación en Bioinformática en Colombia? ¿Qué estamos haciendo mal?*

### King Jordan, Georgia Tech

La colaboración debe ser en dos sentidos y son ustedes, con sus parámetros de investigación, quienes deben decidir dónde podemos encajar desde Georgia Tech. Con seguridad ustedes saben mucho mejor que yo cuáles son los problemas profun-