

Cloudera usará sus capacidades para analizar datos de Inova

Inova Translational Medicine Institute se ha asociado con Cloudera para avanzar en iniciativas de Machine Learning basadas en el genoma y en salvar vidas, tal y como anuncia un comunicado de prensa que explica que el Instituto de Investigación analiza los datos del genoma para avanzar en la medicina de precisión, con el objetivo de predecir y prevenir enfermedades.



Según el acuerdo, el instituto mencionado está utilizando **Cloudera Enterprise** para analizar de forma segura colecciones masivas de datos clínicos y genómicos a velocidades y escalas sin precedentes, para fomentar innovaciones más rápidas en investigación en medicina traslacional.

Como parte del **Centro Inova de Salud Personal** (ICPH, por sus siglas en inglés), el equipo de científicos, investigadores, analistas y colaboradores de ITMI aplica algoritmos de **Machine Learning** en terabytes de información clínica y genómica para identificar los vínculos genéticos con las enfermedades. De ese modo, **pueden hacer descubrimientos a partir de la información de los datos y, en colaboración con el médico tratante**, desarrollan planes de tratamiento personalizados para los pacientes. Este enfoque también se conoce como medicina de precisión y tiene el poder de ayudar a los pacientes a vivir vidas más largas y saludables.

El Instituto recoge datos clínicos de miles de pacientes nacidos en más de 110 países. Explican los expertos que "el ADN de una sola persona contiene seis mil millones de bits de información. Mapear los códigos de ADN individuales en las secuencias del genoma, ayuda a los científicos a determinar la causa de las enfermedades y a descubrir nuevos tratamientos. Como parte de este proceso, **ITMI también está reuniendo lo que se espera sea una de las bases de datos de secuencias del genoma más grandes del mundo** que también se encontrarán conectados a la historia del paciente en un sistema de atención médica".

La plataforma de Cloudera permitió a ITMI optimizar su análisis de datos genómicos para su descubrimiento. Este análisis genómico de datos **permite a un científico de bioinformática estudiar las correlaciones genómicas de personas con condiciones como la artritis, enfermedades autoinmunes o cáncer**. En el pasado, dado el tamaño masivo de genomas enteros, este proceso podía tomar a ITMI alrededor de dos meses para lograrlo. Utilizando Cloudera, ITMI puede realizar análisis de datos de extremo a extremo en una semana. En el futuro, ITMI espera hacer estos análisis de datos en sólo unas horas.

Mientras que un genoma equivale a más de tres mil millones de pares de bases de ADN, ITMI actualmente rastrea aproximadamente 9.000 genomas completamente secuenciados, con

intención de llegar a los 15.000 en el futuro. La moderna base de datos analítica de Cloudera, impulsada por **Apache Impala** (incubating), permite acceder a analíticas SQL de alto rendimiento para el Big Data.