

An open source approach to data analytics

Thanks to advances in storage and analysis, agencies can harness all their data for more effective decision-making



Howard Levenson

Databricks

For the past 40 years, agencies have used data warehouses to collect and analyze their data. Although those warehouses worked well, they were limited in what they could do. For instance, they could only handle structured data, but by some estimates, 90% of agencies' data is unstructured and in the form of text, images, audio, video and the like.

Furthermore, proprietary data warehouses can show agencies what has happened in the past but can't predict what might happen in the future. To achieve the government's goal of evidence-based decision-making, agencies need to be able to tap into all their data and predict what might come next.

Reliable, secure ways to house and share data

Agencies can do predictive analytics of their structured, semi-structured and unstructured data with the help of machine learning and a concept Databricks pioneered called the lakehouse. The Databricks Lakehouse Platform combines the reliability, governance and performance of data warehouses with the openness, flexibility and machine learning support of data lakes.

In addition, Databricks' Delta Lake is an open format storage layer that delivers reliability, security and performance for data lakes, and Delta Sharing is a secure, open protocol, built on the Delta format, that enables organizations to share data regardless of where the data lives.

Powerful results from predictive analytics

The backlog to receive a work visa from U.S. Citizenship and Immigration Services can be very long, and the interview process can be a bottleneck. Data analysis showed that 3% of people missed their appointments, but by using predictive analytics capabilities in Databricks, USCIS could identify who was likely not to show up.

Now USCIS can schedule interviews on days that are more convenient for applicants or remind people of their appointments through various channels. As a result, USCIS has reduced the number of missed appointments and improved the agency's overall efficiency.

In another example, an average of 20 veterans a day have committed suicide in recent years. To address the crisis, the Department of Veterans Affairs has reviewed medications

and mental health factors and is now able to predict which veterans are at risk. By leveraging predictive analytics, the VA is getting veterans the help they need and reducing the number of suicides.

Powerful results like those inspire me and my colleagues at Databricks to continue supporting government agencies with innovative open source solutions. ■

Howard Levenson is regional vice president at Databricks.

Better government begins with better data and AI

By modernizing data management, government agencies open new opportunities for delivering better services and meeting mission objectives. The Databricks Lakehouse Platform unifies all your data and makes it available for analytics, machine learning and AI.

Carah.io/Databricks-eBook

