# How Agencies Can Combat Fraud, Waste and Abuse

by Leveraging an Infrastructure Approach

*Data growth across government continues to accelerate at an astronomical rate, with the global pandemic doing little to slow the growth. Compounding the problem is widespread waste, fraud, and abuse, which cost agencies billions of taxpayer dollars, not to mention the toll it takes on constituent trust.*

*Learn how technology leaders from GAI, NVIDIA, Cloudera and Dell have teamed up to develop a holistic solution that combats fraud, waste, and abuse by marrying powerful processing capabilities with robust analytical tools.*

### The Challenge

The pandemic brought to the fore, issues of fraud, waste, and abuse as multiple agencies struggled to track and validate an unprecedented flow of public funding.

According to a New York University GovLab project, across government, "many datasets are still inaccessible outside of the government agency which owns them, limiting collaboration and undermining efforts to combine different datasets and produce new insights." While this partly relates to process — where and how data is collected and stored — it's overwhelmingly a technological problem.

Insufficient IT infrastructure directly impacts the federal mission when it comes to tackling fraud, waste, and abuse. Largely, this problem centers around data and the fact that most agencies and contractors don't have the software tools and compute power they need to organize, analyze and visualize data  in order to attack this problem.

It will be no easy task to tackle, either. The Government Accountability Office

(GAO) recently identified 36 operations that agencies will need to undergo transformation in order to keep up with exploding data growth, a list that includes high-risk areas affecting the nation's commerce, economy and security as well as the daily lives of U.S. constituents.

There are tools out there, however. Machine learning and artificial intelligence, for example, offer a powerful means for combating fraud and waste. According to the Brookings Institution, "Advanced algorithms can spot abnormalities and outliers that can be referred to human investigators to determine if fraud actually has taken place."

With accelerated processing married to powerful analytics, agencies could begin to tap this emerging potential. "Agencies will need to look at modernizing their infrastructure and adopting new architectures that can deal with humongous amounts of data and be able to accelerate the workloads using GPUs," Government Acquisitions Inc. (GAI) Chief Technology Officer Prem Jadhwani said. He notes that with an updated infrastructure data scientists will have the foundation they need to "analyze vast amounts of data and make effective changes in a timely manner to curb the fraud, waste, and abuse."

### An Infrastructure Solution

So, how can agencies begin to upgrade their infrastructure to introduce the powerful analytic capabilities investigators need to connect the dots across a sprawling base of transactional activities? In short, it requires specialized hardware optimized to support specialized processing.

NVIDIA, Cloudera, and Dell Technologies have combined their most sophisticated offerings to answer that call. The companies have joined

> ## "The NVIDIA and Cloudera solution offers deep analytic capabilities."
>
> **TARUN DAVE** | Solutions Engineer, Cloudera

forces to introduce a solution that leverages the NVIDIA RAPIDS Accelerator for Apache Spark integrated with Cloudera Data Platform (CDP) running on NVIDIA Certified Dell Technologies Servers, containing NVIDIA GPUs, optimized them for performance and security.

"The NVIDIA and Cloudera solution offers deep analytic capabilities. Its platform can process all kinds of data, structured and unstructured, as well as large volumes of data. Machine learning is included in the Cloudera Data Platform, with the NVIDIA GPU delivering acceleration for model training and for predicting outcomes faster than was possible before," said Cloudera Solutions Engineer Tarun Dave.

Dell's optimized servers support this effort, helping to determine whether data is immediately actionable or historically valuable. "To do that, you need processing power and reliability in your data center," said Dell Technologies Director of Business Development Bob Nicholson.

### CASE STUDY: IRS Tackles Fraud, Waste, and Abuse

*The challenge:* At the Internal Revenue Service (IRS), evidence of fraud lies buried in decades worth of data: a volume of information so vast, it exceeds even the most cutting-edge processing capabilities. One data scientist illustrated the challenges when she described trying to comb through a three-plus terabyte dataset looking for patterns that might indicate fraud. Even when she employed a large bank of CPU servers to process the information for a considerable period, the data refused to line up. The models, which ran on CPUs across hundreds of servers, struggled with the sheer volume of data. The team tried breaking down the data sets, server by server, but were forced to manually stitch data subsets together to make the solution work. Even with all this careful manual effort, it wasn't possible for the team to offer full visibility into real-time fraud detection.

*The solution:* Leveraging a combination of IT offerings, the IRS is
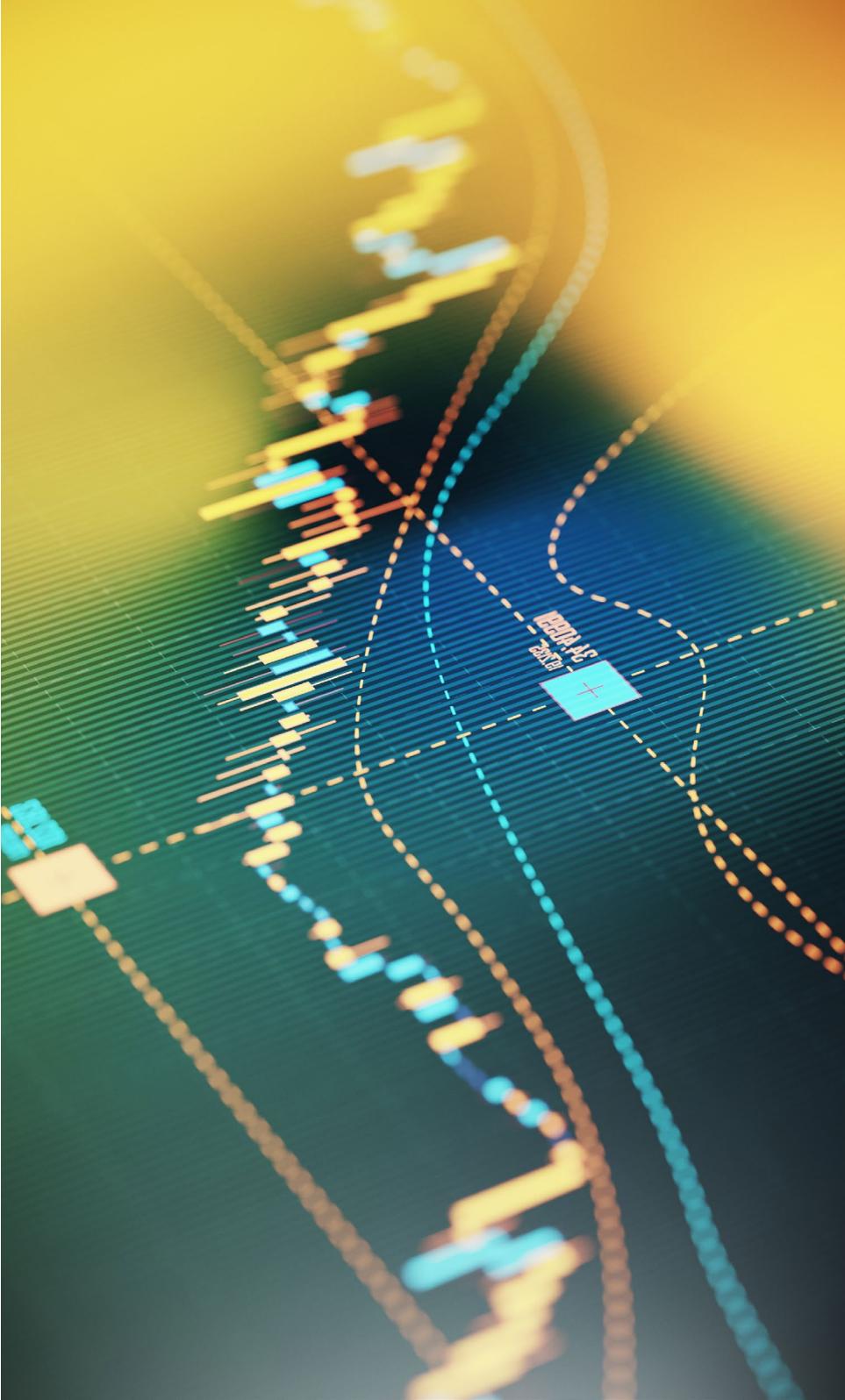
implementing high-powered AI tools, machine learning capabilities and applications capable of rapidly exposing attempts at fraud or identity theft. This infrastructure approach is helping the agency deliver more efficiently on its efforts to root out fraud, and to better protect taxpayers against risks such as identity theft. Moreover, with less time and manual effort needed to process and uncover the information, the IRS has vastly improved its data analysis capabilities while also cutting costs.

Enterprise data engineers at Cloudera and NVIDIA are tackling massive data bottlenecks with CDP Private Cloud and NVIDIA RAPIDS Accelerator for Apache Spark software. The combination of the Cloudera Data Platform and NVIDIA RAPIDS Accelerator for Apache Spark running on NVIDIA-Certified Systems — industry-standard servers accelerated with NVIDIA GPUs — enables faster, easier implementation of AI and machine learning at scale.

By adding GPUs to mainstream big data servers, "the Cloudera and NVIDIA integration will empower [the IRS] to use data-driven insights to power mission-critical use cases such as fraud detection," said Joe Ansaidi technical branch chief of the research and applied analytics and statistics division at the IRS.

*Ease of implementation:* The joint offering from NVIDIA and Cloudera fell into a sweet spot for the IRS, which had already recognized the powerful potential of GPUs in combination with a cloud-based data platform.

"The IRS used the Apache Spark workloads that they were already using, with graph analysis to interconnect people, individual to institution, to figure out the anomalies in their behaviors and patterns in support of fraud detection," Dave of Cloudera said.

"It was really almost a fortuitous set of circumstances," said Scott McClellan, NVIDIA senior director of product. "IRS couldn't meet the performance

requirements with their existing solution, they couldn't process those multiple terabytes of data. The pipeline to run that was just ridiculously long. NVIDIA gave them the means to accelerate the Spark data processing, to feed that ML pipeline, and Cloudera delivered a platform for scalable data processing, including analytics."

All this has helped address IRS' need to get a broader and deeper understanding of its data flows. "People are trying to hide their behaviors, so you need to trace them through multiple steps, multiple hops in a graph. To that end, IRS needs large-scale graph analytics," said William Benton, principal product architect at NVIDIA.

By developing workloads that use Apache Spark and graph analysis, the engineering teams created immense graphs with nodes and edges, connecting individuals to institutions and, subsequently, to larger entities spanning years and decades. AI bots and ML algorithms can quickly and repeatedly analyze these graphs to root out anomalies in behavior or patterns that signal potential fraud.

With an Apache Spark cluster of GPU-powered servers, the group can accelerate all its current jobs and run others previously thought impractical. Those jobs can tackle the big data sets the team has at its disposal. These same data sets used to take weeks or months to stitch together and analyze — if they could be processed on existing machines at all. Now, they can be processed in days, hours or even minutes, with testing revealing a ten-fold improvement in engineering and data science workflows and a 50% reduction in infrastructure costs.

It's not just the IRS that can benefit from an infrastructure approach to mitigating fraud, waste and abuse, of course. "Any place where there are payments or transactions, agencies can apply a similar infrastructure-based approach," Dave said. "Whether it's healthcare spending, social benefits, or just government acquisitions: This approach could apply across the broad scope of transactional mission sets."

## Looking Ahead

Federal technology leaders can begin today to pivot toward an infrastructure-based approach, as they seek to support their agencies' efforts around mitigating fraud, waste, and abuse. All the needed enterprise-grade components — from NVIDIA, Cloudera, and Dell Technologies — are readily available in the commercial marketplace and can be implemented by GAI with minimal customization.

As agencies look to pivot toward a more modernized approach, a good starting point is to identify the places where they may be falling short when it comes to fraud, waste, and abuse. For example, if there are vast tracts of unexplored data, or if attempts to parse the data fail often or require a significant amount of human intervention, modernization could help to move the needle.

Additionally, by working in tandem to identify risk areas and implement a technologically driven solution, mission leaders and technology executives together can change the federal narrative around fraud, waste, and abuse.

**With support from solution providers like GAI and industry leaders like NVIDIA, Cloudera, and Dell Technologies, agencies can leverage the best of what the commercial marketplace has to offer, often building seamlessly upon the strong IT capabilities they already have in place.**