

CASE STUDY

Leveraging Public Sector Data: Tackle Operational Inefficiencies, Crime, and Other Critical Initiatives

Governments are known for creating and maintaining large volumes of data – ranging from healthcare records to satellite imagery. At today's estimates, U.S. federal agencies alone currently store an average of three petabytes of data. With more devices coming online, government agencies find themselves with more data than they can handle. Meanwhile, they are under more pressure than other organizations to use that data for immediate responsiveness and efficiency on every front, whether that means detecting fraud, improving citizens' lives, or executing combat missions.

It has been reported that 2.5 quintillion bytes of data are produced each day, and 90% of the data in the world today has been created in the last two years alone.¹ The data comes from everywhere and typically extends across multiple applications as well as multiple data sets. Today, stored structured data is integrated with Big Data from the cloud, social media, and Big Data stores in order to feed business use cases across multiple applications. For the government, the same data is often applied across multiple applications concurrently. For example, one agency could be running a financial report to assess the efficiency of procurement, while another office might use the same data for budgeting. The use cases vary, but the same data will be required.

When it comes to managing that data, however, government agencies have faced the same longstanding issue as most organizations do today: how to effectively manage and deal with massive amounts of data, both structured and unstructured. It is reported that public sector data analysts spend almost half of their time collecting and organizing data but less than a third of their time is spent gleaning actionable insights from it.²

As agencies are quickly adopting cloud technologies, interoperability is fundamental in supporting high-velocity data analysis that is required of these agencies. With that, maintaining a 360-degree view across their data sets is not only critical, but required.³



BUNDESPOLIZEI



Despite all of the challenges of managing their data, 76% of respondents of a recent research with iGov believe that Big Data could benefit their organizations and 88% of them see one of the key benefits is gaining deeper insights into their citizens.

When all data is stored in one central, secure location, it will become possible to give government officials across the organization access to the right data, thereby reducing errors and inefficiencies within the government and ensuring that the correct information is used. This will enable government officials to have access to the most up-to-date information on their citizens. A fundamental aspect to making this work is the seamless integration of all data sources by eliminating manual coding or complexity. By introducing this ease of use framework, government organizations can more quickly scale their operations and increase efficiency and productivity while driving down costs. Pentaho's Data Integration and Business Analytics platform has been shown to accelerate the design and deployment of Big Data analytics by up to 15 times compared to hand-coding techniques. Read on to see how government agencies have better leveraged modern data tools to improve efficiency, maximize the value of sensor data, and fight critical initiatives such as cybersecurity.

Improving Operational Efficiency

83% of Federal IT officials say Big Data can save roughly \$380 billion or more from the federal budget, or about \$1,200 per American.⁴ Despite all of the challenges of managing their data, 76% of respondents of a recent research with iGov believe that Big Data could benefit their organizations and 88% of them see one of the key benefits is gaining deeper insights into their citizens.

Since agencies rely on public funds to run their operations, reducing costs is always important. Compiling and analyzing Big Data offers governments a way to do this. For example, governments can use Big Data to track fiscal spending and audit expenditures to identify areas for reduction or improvement.

A great example is one of the world's largest finance and accounting operations, responsible for overseeing and distributing payments and other benefits to active and retired U.S. military personnel. Between issuing payments

and the other responsibilities assigned to the agency (selling military equipment, issuing travel vouchers, etc.), the agency handles tens of thousands of transactions every day and processes over a hundred million dollars' worth of payment transactions to over six million different accounts annually. This agency wanted to set an example of creating organizational efficiency and wanted to drive continual improvements by giving their field offices the ability to create and run their own business intelligence reports—without requiring assistance from their IT departments. By leveraging Pentaho to do this, and eliminating the need to build their own ETL (extract, transform, load) solution, the organization was able to save years of development work and millions of tax payer dollars.

FINRA, a non-governmental organization, is responsible for overseeing more than five petabytes of data that is created each day from 4,250 brokerage firms. Searching for fraud, insider trading, and compliance violations, FINRA was challenged with retaining at least seven years of data and maintaining a scalable architecture to accommodate a projected 6x data growth. In addition, FINRA needed to reduce investigative search times. By putting data in the hands of their analysts, Pentaho helped FINRA reduce query time from hours to 90 seconds or less with self-service analytics. In 2015, FINRA detected \$96.2 million in misconduct.⁵

Her Majesty's Revenue and Customs (HMRC), the UK Tax Authority, has been working on a multi-year project to improve how they manage their data — by shifting to an open source framework to future proof their environment and eliminate vendor lock-in. The agency has grown to 70,000 full time employees with separate, discrete lines of businesses relying on different datasets to perform their job and meet arbitrary key performance indicators (KPIs). They turned to Pentaho to centralize their data architecture, eliminate the need for third party consulting services and provide call center managers faster access to operational reporting. After implementing Pentaho, HMRC has already saved roughly 900 man hours a day with self-service analytics.

By putting data in the hands of their analysts, Pentaho helped FINRA reduce query time from hours to 90 seconds or less with self-service analytics. In 2015, FINRA detected \$96.2 million in misconduct.⁵

Lastly, a government healthcare agency was processing medical claim data up to one terabyte per day. While the daily volume of data wasn't overwhelming, the processing time was. It was taking 27 hours to process a day's worth of data. The agency was relying on hand coding to blend various data sources together, but the analysis was incomplete and prone to human error. An incomplete view of the organization's data only compounded the agency's problem as they struggled to provide answers to simple questions that were being asked by the public. The team of analysts quickly realized they needed access to the underlying data to modify the queries, data sets, and filters and understand and address the anomalies within their data sets. By consolidating and blending the organization's disparate data sources and providing self-service analytics directly to the business users, the agency was able to increase productivity and provide a 360-degree view of their constituents. With Pentaho's automated, codeless environment, the agency was able to reduce their development time by 15x compared to handcoding. However, it was the agency's ability to reduce its processing times from 27 hours to under 45 minutes that transformed the agency's business processes and drove real efficiency. Not only could the agency provide timely reports to their constituents, they were able to trust the underlying data that fueled these reports. The agency can now focus on their true mission: improving patient care.

Providing Proactive Intelligence with the Internet of Things (IoT)

The Internet of Things is viewed as one of the biggest technology disruptions in today's market. With over 30 billion connected things today, the explosion of IoT devices is generating a massive amount of data that organizations need to effectively integrate, manage and analyze to provide business value.

Marquis Software provides a 360-view of an inmate solution to state department corrections facilities. Marquis has been providing software to state departments since the early 1990s; however, they only recently extended their offering to provide geospatial intelligence. With this solution, prisons can now physically track the location of their prisoners and guards to determine that safety matters are addressed appropriately and quickly. By tracking the movement of prisoners, guards can anticipate where and when a confrontation might occur before it happens. This information is blended with inmate statistics, such as race, gang affiliations and medical disabilities, so that state correction facilities can maintain a 360-view of their prisoners despite prison overcrowding and staff limitations.

In another IoT use case, Caterpillar Marine helped their clients, including government agencies, to become more operationally efficient by leveraging sensor data. One of Caterpillar Marine's customers was spending \$35,000 every two years to clean the hulls of its drive-on ships, which are used to transport vehicles. An analysis quickly discovered that dirty hulls were costing the organization \$1.3 million per vessel in lost fuel efficiency every two years, or over \$10 million for a fleet of eight.

Government entities can blend data from criminal activity, land use, service requests, intelligence sensors, and a host of other things to drive better decision making. For example, the Navy can use network sensors from weapons systems so that missiles, drones, and ship-based weapons systems can locate and destroy approaching targets at distances "beyond-the-horizon." With all of that data, government agencies can then create a single view of their operations — ranging from soldier, sailor, airman, marine, or coastguard to taxpayer, inmate, or patient.

Using Big Data to Fight Security Threats

Traditionally, security threats have been synonymous with tighter law enforcement as in the case with Bundespolizei, the German Federal Police, who wanted a more complete picture of their armed forces for mission planning purposes that resulted in blending over 20 streams of incoming data sources such as entry numbers at the borders, asylum caseloads and criminal offences.

However, today's security threats are more sophisticated than ever. It goes beyond just the protection of one's borders to include cyberthreats. Traditional methods of handling cyber security require custom scripting and blending of multiple data sources that can be very slow and time-intensive. The average analyst takes 205 days to detect an attack due to the huge volume and variety of data to go through.⁶ This isn't just simple big data, but all kinds of messy data from all kinds of sources, including application, network, firewall logs, web access data, HCM and CRM—all which are difficult to blend. It takes time to clean and blend the data, and critical decisions, such as threat detection, can't wait on legacy systems and processes. Threat detection needs to be extended to government entities that are monitoring what is happening within their borders and monitoring potential malicious activities that could result in a cyber threat. Governments can collect, process, and analyze the data from government networks, as well as public data, to protect their country from being attacked and identify insider threats.

A large energy organization was challenged with parsing and analyzing enormous log volumes coming from different sources and locations. Since the data came in many formats, it took weeks to blend the data in order to detect certain patterns and activities. By turning to Pentaho, this organization leveraged natural language processing capabilities and R algorithms to deploy at scale on Hadoop, thereby detecting cyberattacks faster and reduce development time 10x so that the data could immediately be processed and analyzed.

In another cybersecurity example, a large telecommunications organization was unable to detect suspicious network activity but found an unexpected spike in outbound connections from a system with sensitive data. After leveraging Pentaho to orchestrate the data transformations behind the 'blacklist' and 'whitelist' analysis for anomaly detection, the organization was able to analyze potential threats in a timely fashion. They have already seen a huge reduction in IT resources and time dedicated to anomaly detection.

LEARN MORE

Over 1,500 customers, including a number of government and public sector organizations, rely on Pentaho to drive their strategic decisions. Learn more at www.pentaho.com.

¹<https://cloudtweaks.com/2015/03/surprising-facts-and-stats-about-the-big-data-industry>

²<http://www.govtech.com/opinion/Is-the-Government-Hoarding-Too-Much-Data.html>

³<http://www.csoonline.com/article/2134455/strategic-planning-erm/big-data-still--a-new-frontier--for-most-of-the-public-sector.html>

⁴<https://www.splunk.com/pdfs/fact-sheets/sap-public-sector-big-data-report-final-2.pdf>

⁵<https://www.yahoo.com/news/finra-orders-wall-st-brokerages-return-96-2-213759183--sector.html>

⁶<http://www.economist.com/news/business/21677639-business-protecting-against-computer-hacking-booming-cost-immaturity>