

#### CASE STUDY

# Public Safety Analytics Company Accelerates Data Analytics with Dremio

#### At a Glance

#### Challenge

The public safety analytics company faced significant hurdles accessing and standardizing data from legacy law enforcement systems that were 10-20 years old. These disparate systems created multiple isolated data sources with inconsistent formats, while strict security requirements added complexity. As a startup with limited resources, they needed a cost-effective approach that wouldn't require a large engineering team.

#### **Solution**

The company built their platform with Dremio at its core, implementing it on Kubernetes in the cloud with Amazon S3 as their data lake using Apache Iceberg format. They created Pythonbased ETL processes to standardize data, used Airflow for orchestration, and integrated Elasticsearch for powerful user-facing search capabilities while maintaining strong security controls.

#### **Results**

- Eliminated need for 2 additional engineering hires (~\$300,000/year)
- Scaled from 1 to 20+ customers with lean data team
- Processing 4TB of customer data daily with excellent performance
- Transformed 8-hour data retrieval processes into instant queries
- Enhanced data standardization
  across different departments

### **The Challenge**

A data analytics company focused on public safety faced significant challenges when building their data analytics platform for law enforcement agencies. Police departments struggle with accessing and analyzing data stored in legacy systems that are often 10-20 years old, creating major barriers to effective data use. These aging systems create substantial friction points that hamper operational effectiveness and analytical capabilities.

Access to data was slow and difficult, with law enforcement agencies often waiting hours or days to retrieve critical information from legacy record management systems. Multiple isolated data sources existed with inconsistent formats, making it nearly impossible to conduct cross-system analysis without extensive manual work. Strict security requirements for sensitive law enforcement data added another layer of complexity, requiring careful handling and robust access controls. There was also an urgent need for rapid standardization across departments to enable information sharing while maintaining data security.

The company needed a solution that could help them normalize data across departments, make it quickly searchable, and provide valuable insights while maintaining the highest security standards. As a startup with limited resources, they required a cost-effective approach that wouldn't demand an oversized engineering team or excessive infrastructure investments.

## Why Dremio?

When evaluating data platforms, the company had several key requirements that guided their technology selection. Speed was essential - they needed the ability to quickly explore and search through large volumes of public safety data without lengthy processing times. Flexibility was equally important, as they needed to connect to various source systems that couldn't be predicted in advance, given the diversity of systems used by different law enforcement agencies.

Accessibility was another critical factor - they wanted an open-source or free-to-try component that would allow them to test and validate the solution before making a significant investment. Security was nonnegotiable, requiring robust controls to manage sensitive law enforcement data while meeting compliance requirements. Finally, scalability was vital - they needed an architecture that could grow with their business without proportional resource increases, allowing them to onboard new clients efficiently.



"The biggest appeal early on was quickly being able to understand the product and try the product and the flexibility of it in that we could operate in different clouds, connect to different databases," explained their Co-Founder and Head of Engineering.

The company ultimately chose Dremio after carefully evaluating several options. Dremio's ability to query data in place within S3 without having to move it into another format was particularly attractive, as it simplified their data architecture and reduced storage costs. The robust security controls for sensitive data access addressed their compliance requirements. The simple, intuitive user interface wasn't overwhelming, making it accessible to team members with varying technical backgrounds. Native integration with Apache Iceberg tables provided advanced data management capabilities. Perhaps most importantly, Dremio could scale effectively as their data and customer base grew, supporting their business expansion plans.

## **The Solution**

The company implemented Dremio as a core component of their data platform, which extracts, normalizes, and analyzes public safety data from various legacy systems. Their comprehensive architecture leverages several technologies working in concert to deliver a powerful, flexible solution.

At the foundation is their data lake, built on Amazon S3 and storing data in Apache Iceberg format for optimal management and performance. Orchestration is handled by Airflow, which manages their data workflows and ensures reliable processing of information from diverse sources. Dremio serves as the query engine, providing semantic data representation and query capabilities that make the data accessible and analyzable. The user-facing search experience is powered by Elasticsearch, delivering fast and relevant results to end users.

"We run Dremio in Kubernetes. We run everything in Kubernetes. And so our Dremio implementation is basically just a Helm chart. We pick the node count, pick the CPUs, pick the RAM, hit enter. Now it's running," said the Co-Founder, highlighting the simplicity of their deployment approach.

Their data pipeline uses Python scripts to collect data from various law enforcement systems, standardize it according to common models, and store it in S3 as leeberg tables. Dremio then provides semantic representation of this data, enabling them to handle it as if it were transformed without actually duplicating storage. This ELT (Extract, Load, Transform) approach gives them significant benefits in terms of cost and maintenance, reducing the overhead associated with managing duplicate copies of large datasets. The company processes approximately 4 terabytes of customer data daily through this architecture with excellent performance, demonstrating the scalability and efficiency of their Dremio-based solution.

## The Impact

Implementing Dremio has delivered substantial benefits to the company, transforming both their business operations and technical capabilities. The impact spans from financial savings to enhanced product functionality, creating competitive advantages in their market.

From a business perspective, the company achieved significant staffing efficiencies that directly affected their bottom line. As their Co-Founder explained, "I'd start with staffing... the easiest cost comparison that I could make in the early days is you can eliminate two employees that don't yet exist, and you would have to hire otherwise... there's 300,000 a year that you have to actually get two people to manage what not having a tool in Dremio actually incurs."

Customer onboarding accelerated dramatically, allowing the company to grow from their first client to more than 20 customers while maintaining a lean data engineering team. This scaling efficiency provided a crucial competitive advantage in their growth phase. Their product capabilities expanded significantly as well, with the ability to quickly answer complex questions about public safety data that would take weeks with traditional methods, enhancing the value proposition for their customers.

The technical benefits were equally impressive. The Co-Founder noted their remarkable operational efficiency: "We have two data engineers...and we all do more than just deal with data pipelines. We would have had to have double the staff three years earlier." Data standardization improved substantially, with Dremio helping them normalize information across different law enforcement departments despite varying source formats.

Data exploration became vastly simpler for their team. "The fact that it's like, let's go look at Dremio, run a query. Great. There it is. There's the data we're looking for. Instead of an eight-hour long battle of trying to find what file to pull," explained the Co-Founder. Their operational reliability remained excellent even as data volumes grew: "Right now, we're managing somewhere around four terabytes of customer data every day, and operations are as smooth as butter."

Through these improvements, the company now enables law enforcement agencies to quickly search and analyze their data, understand events without reading lengthy reports, share information securely with other departments, and make data-driven decisions for community safety.



The impact extends beyond technical metrics to real-world outcomes in public safety operations.

## **Looking Ahead**

The company continues to evolve their Dremio implementation as they scale, with several strategic initiatives planned to enhance their platform capabilities. They are moving toward a more event-driven data flow to provide fresher, more current data to their users, enabling near real-time insights for law enforcement operations.

Integration plans include connecting additional data systems beyond their current scope, incorporating records

management, citation data, and incident data to provide a more comprehensive view of public safety information. They're exploring opportunities to enrich their datasets with correlated information like weather data, adding contextual dimensions that can reveal new insights about patterns and trends.

The company is also focused on enhancing data exploration capabilities for their engineering team, making it easier to investigate patterns and develop new analytical features. These ongoing improvements demonstrate their commitment to leveraging Dremio as a foundation for continued innovation in public safety analytics.

#### **ABOUT DREMIO**

Dremio is the intelligent lakehouse platform for the business, serving hundreds of global enterprises, including Maersk, Amazon, Regeneron, NetApp, and S&P Global. Based on open-source technologies like Apache Iceberg and Apache Arrow, Dremio provides an open lakehouse architecture enabling the fastest time to insight and platform flexibility at a fraction of the cost. Learn more at <u>www.dremio.com</u>.

Dremio and the Narwhal logo are registered trademarks or trademarks of Dremio, Inc. in the United States and other countries. Other brand names mentioned herein are for identification purposes only and may be trademarks of their respective holder(s). © 2025 Dremio, Inc. All rights reserved.