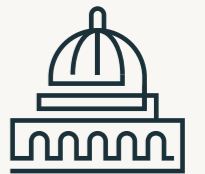


EBOOK

Driving Innovation and Transformation in the State and Local Government With Data + AI

Empowering state and local governments to deliver better public services to their communities



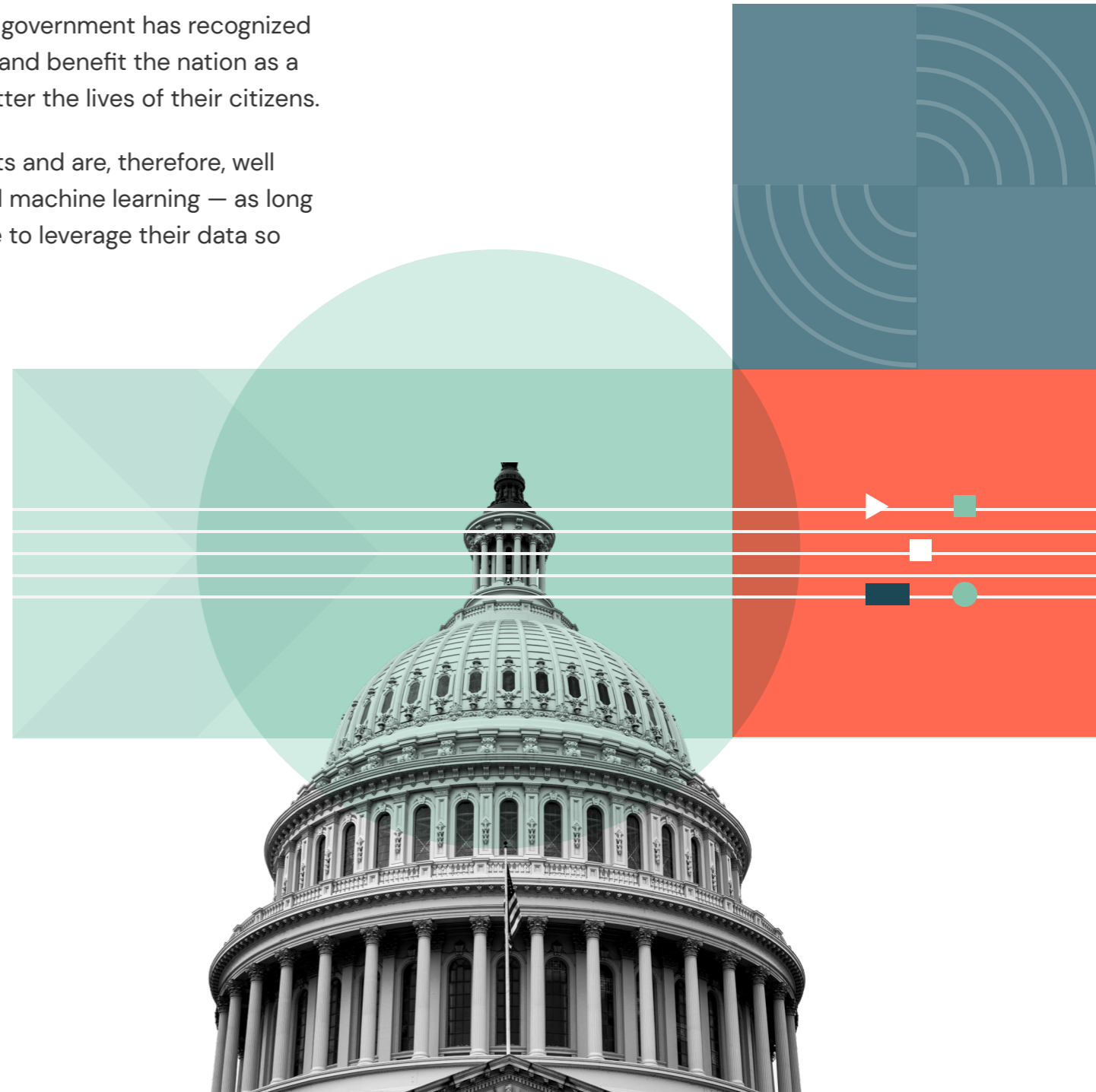
Contents

Making a Difference: Leveraging Data in State and Local Government	03
Recognizing the Opportunity for Data and AI	04
Challenges to State and Local Success	06
The Databricks Lakehouse Platform: Empowering State and Local Government Agencies to Better Serve Their Citizens	08
Customer Story — The City of Spokane: Promoting Government Transparency and Accountability With Data	11
Conclusion	12

Making a Difference: Leveraging Data in State and Local Government

The proliferation of data has helped fuel the digital transformation of a wide array of industries, from healthcare and life sciences to banking and financial services. Just as the federal government has recognized how analytics and machine learning (ML) can make their agencies more efficient and benefit the nation as a whole, state and local agencies also see how analytics and ML can be used to better the lives of their citizens.

These agencies are much closer to actionable data than their federal counterparts and are, therefore, well positioned to lead the way on technology initiatives leveraging data analytics and machine learning — as long as the barriers they face are removed. Today's local governments need to be able to leverage their data so they can make the most positive impact possible.



Recognizing the Opportunity for Data and AI

The opportunities for state and local government agencies to leverage data analytics and AI — and have an immediate impact on their citizens — are massive.

Building a smarter and more innovative state and local government starts with unlocking the power of data and machine learning. In order to realize innovation, state and local agencies must be focused on modernizing their data analytics capabilities to better serve and protect their communities.

This includes enabling better recommendations for citizens seeking health and social services, advancing smarter police, fire and medical responses, using data to improve the efficiencies and resilience of public utilities while lowering costs for consumers, reducing wasteful public spending by identifying potential tax and entitlement fraud and waste, improving budget forecasting, and even reducing the impact of traffic on commuters in local areas. The list goes on.

Thanks to the federal prioritization of AI — with spending expected to reach \$4.3 billion in fiscal year 2023, according to a recent report from Bloomberg Government — state and local leaders expect to take advantage of this investment to help their local municipalities and communities. In fact, a recent Gartner study of CIOs¹ at all levels of government has confirmed that the top technologies they will be focusing on are AI, data analytics and the cloud.

As an example of the potential impact, a recent study by Deloitte shows that government agencies can save upward of \$41 billion from data-driven automation and AI.

¹ [Gartner Survey Finds Government CIOs to Focus Technology Investments on Data Analytics and Cybersecurity in 2019](#)

[Administration Projects Agencies Will Spend \\$1 Billion on Artificial Intelligence Next Year](#)



Common use cases

Across state and local government agencies, data and AI are producing insights and the predictive capabilities that provide better social services more efficiently, improve the delivery and quality of healthcare services, and support a more secure and sustainable energy and transportation infrastructure.



HEALTH AND HUMAN SERVICES

Enhance disease surveillance and the delivery and quality of health and human services to our communities

- Citizen recommendations for social service and health insurance benefits
- Social determinants of health (SDOH) analytics
- Disease surveillance for outbreak prediction and resource deployment
- Insurance claim fraud detection



COMMERCE

Proactively detect anomalies with machine learning to mitigate risk and prevent fraudulent activity

- Tax and revenue forecasting
- Tax fraud and collection
- Process and operations management
- Grants management



ENERGY

Plan for the efficient delivery of resources and a sustainable energy grid with data analytics

- Security of energy infrastructure
- Smarter energy management
- Energy exploration
- Electrical grid reliability



LAW ENFORCEMENT

Empower criminal justice departments with all data sources and CJIS-compliant predictive analytics to mitigate criminal activities

- Fleet maintenance and logistics
- Prediction of high crime areas
- Predictive models for courts and for incarceration management
- Video and social media surveillance analytics



PUBLIC AUTHORITY/UTILITIES

Leverage data analytics to transform local infrastructure, improve operational efficiency and reduce costs

- Predictive maintenance
- Customer 360
- Sentiment analysis
- Fraud prevention



TRANSPORTATION

Streamline logistics through analytics to improve traffic flow and predict maintenance issues for public transportation systems

- Airport logistics
- Highway planning
- Road traffic optimization
- Predictive maintenance

Challenges to State and Local Success

The interest in big data and artificial intelligence in state and local governments is steadily increasing, which is great news for citizens. But the ability to move from interest to innovation is blocked by challenges that are technology- and personnel-driven. Many local governments and their agencies need to upgrade their legacy IT infrastructure in order to use data and AI to its full potential. Additionally, many of the same governments have a wide variety of data scattered throughout their departments but lack the resources needed to bring them together to create the data streams that AI projects require. Also restraining the adoption of data and AI in government is the lack of data scientists and subject matter experts required to determine problems that AI can solve for a given department, agency or use case.

The “state” of state infrastructure is a barrier

It’s no secret that technology in state and local governments has lagged behind the private sector. Government systems are traditionally built with on-premises data warehouses and data marts that are highly complex to maintain, costly to scale as compute is coupled with storage, limiting from a data science perspective and lacking support for the growing volumes of unstructured data. This inhibits data-driven innovation and roadblocks the use of AI as local government agencies search for data science tools to fill the gaps.

Infrastructure becomes harder and more expensive to maintain as it ages. As time passes, these environments become more complex due to their need for specialized patches and updates that keep these systems available while

essentially doing nothing to address their poor interoperability, slow processing speeds and inability to scale — which are critical limitations in today’s more data-intensive use cases.

Maintaining these systems requires a massive investment of both time and money compared to using modern cloud-based systems. For the technical teams that are tasked with trying to integrate any of these legacy systems with third-party tooling or services, this often requires significant customization, and even then, there is still a chance that the final integration won’t be successful. Legacy systems also keep personnel from spending their energy and resources on emerging technologies such as AI. In 2015 alone, states collectively spent almost \$30 billion on IT² — with the majority of that going to maintain legacy systems.

And data reliability is a big concern, as data replication occurs across data marts when various data teams try to access and explore the data, creating data management and governance challenges. Without a single source of truth, teams struggle with data inconsistencies, which can result in inaccurate analysis and model performance that is only compounded over time.

State and local governments must be focused on establishing infrastructure that can handle the requirements of a data-driven organization with a complex data ecosystem. A modern IT infrastructure is critical for boosting reliability, performance and IT efficiency. Improvements to existing infrastructure are required, as are investments in technologies that will help these entities achieve their missions in areas such as data analytics and AI — and help them secure access to cloud-based applications and services.

Siloed data can severely limit program effectiveness

Being close to data collected at state and local levels should be a benefit to the government agencies at those levels. But due to the fragmented nature of the data itself, gaining actionable insights is complicated and frustrating. State and local governments depend upon a complex mix of legacy systems, databases, monolith applications (both on-premises and cloud-based), sensor data, citizen-facing websites and portals, and much more for collecting and storing data — all in structured and unstructured formats, and subject to a variety of privacy standards and governance models.

This siloed data drastically limits productivity because agency data team members are unable to get a 360-degree view of all the data. This can lead to interruptions in service delivery, decisions driven by bad data and more. Data quality is greatly affected by these silos, as the full value of the data locked into them is not accessible. Priority must be given to developing reliable and performant data pipelines that will take the data that exists in these siloed sources and migrate it into one centralized location, creating clean, secure and compliant data for government decision-makers.

Doing nothing will result in limited or incomplete analysis at best, and at worst, flawed analyses that can negatively impact decision-making that directly affects citizens.

Lack of collaboration exacerbates the skills gap

The skills gap that exists within the public sector, in general, is a challenge that will be very difficult to overcome. The demand for data science, data engineering and data analysis skills is outpacing supply in both the public and private spheres. With the competition for these skills being very high, the public sector will be well behind in bringing in the number of qualified individuals required to push forward data-focused initiatives, including AI and machine learning, in a way that beneficially impacts their local areas. Collaboration is an efficient and inexpensive way to overcome the challenges posed by siloing data team members and acts as a force multiplier of the high-value skills that exist in the public sector. But enabling collaboration within and across government agencies and departments is difficult.

The factors that limit collaboration in the public sector are various:

- Regulatory and compliance resistance to sharing data due to privacy and security concerns
- Insufficient understanding of the data required — and available — for advanced analytics
- Misalignment between project interests and the expectations around data sharing
- Lack of a standard toolset for data teams to leverage
- Organizational silos

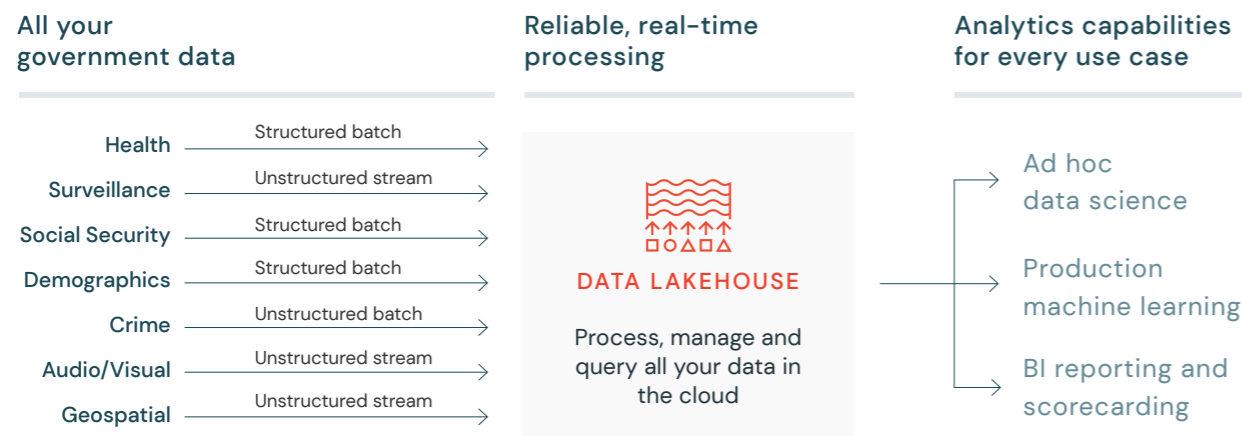
The Databricks Lakehouse Platform: Empowering State and Local Government Agencies to Better Serve Their Citizens

Building a smarter and more innovative state or local government starts by unlocking the power of data, analytics and AI. Driven by the increasing demands of citizens who require access to better services faster, governments are being called upon to tackle challenges ranging from population healthcare to crime prevention.

Databricks provides state and local government agencies with the Lakehouse Platform, which combines the best of data warehouses and data lakes — to store and manage all their data for all their analytics workloads. Lakehouse offers a single solution for all major data workloads, whether structured or unstructured, and supports use cases from streaming analytics to business intelligence, data science and AI.

The Databricks Lakehouse Platform has three unique characteristics that address the biggest challenges that state and local governments face:

- **Simple:** The Lakehouse Platform is architected to support all data workloads — structured, unstructured, streaming and batch — in one common platform, whether statewide or within a locale. Have all your data at your fingertips with this simplified approach to data management.
- **Open:** Built on top of the world’s most successful open source data projects, the Lakehouse Platform unifies your data ecosystem with open standards. This means there are no proprietary data formats or lock-ins, and you have the flexibility to leverage existing investments in tooling.
- **Collaborative:** Most importantly, it’s collaborative, so data engineers, analysts and data scientists are able to work together much more easily, which helps states upskill their data teams and more effectively leverage resources across teams.

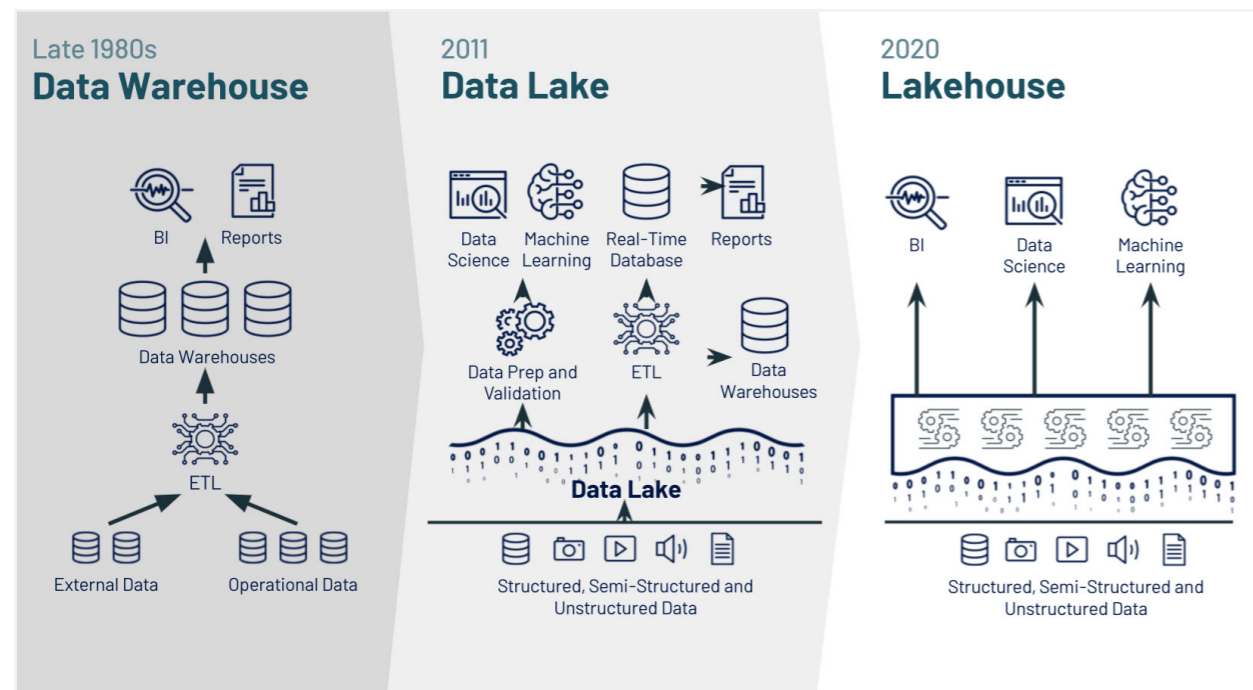


State and local agencies that are powering impactful innovations with Databricks Lakehouse

 <p>City of Spokane</p> <p>Enabling multiple BI and AI use cases through the modernization of their entire data analytics infrastructure</p>	 <p>DC Water Authority</p> <p>Using predictive analytics to ensure quality control for the city's water supply for over 700,000 residents and nearly 20 million annual visitors</p>	 <p>Center for Medicare and Medicaid Services</p> <p>Unlocking population-scale data insights to improve public health for those who have complex care needs</p>
--	---	--

Managing government data with a unified approach

Databricks makes it possible to aggregate and process the massive collections of diverse and sensitive government data that exist in silos, both structured and unstructured. This capability provides a pathway for state and local departments to overcome the significant infrastructure challenges they are experiencing. The Databricks Lakehouse leverages Delta Lake to unify the diverse and very large amounts of data these agencies are working with. Delta Lake is an open format, centralized data storage layer that delivers reliability, security and performance — for both streaming and batch operations.



The Lakehouse Platform combines the best elements of data lakes and data warehouses — delivering data management and performance typically found in data warehouses with the low-cost, flexible object stores offered by data lakes

With a unified data foundation for business intelligence, data science and machine learning, government agencies can add reliability, performance and quality to existing data lakes while simplifying data engineering and infrastructure management with automation to simplify the development and management of data pipelines.

Break down the institutional silos limiting collaboration

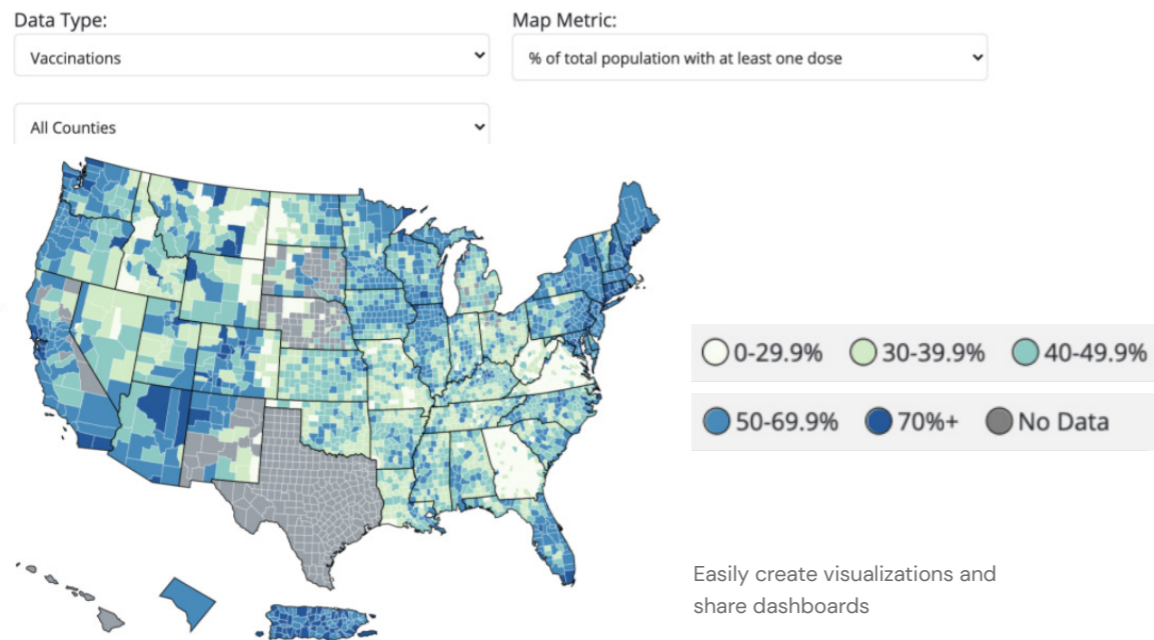
Collaborate across engineering, data science and machine learning teams with support for multiple languages, built-in data visualizations, automatic versioning and operationalization with jobs. With support for multiple languages (R, Python, SQL and Scala) and the ability to easily share notebooks and work with peers across data teams, government organizations are empowered to get the most from their data regardless of their data science background and expertise.

Through built-in visualizations, analytics teams can quickly discover new insights and share interactive dashboards with other teams. Databricks notebooks can also serve as an operations tool, allowing teams to schedule notebooks to automatically run machine learning and data pipelines at scale, create multistage pipelines using notebook workflows, set up alerts and quickly access audit logs for easy monitoring and troubleshooting.

Realize insights through business analytics across agencies

The Databricks Lakehouse Platform includes the business intelligence capability Databricks SQL, which allows data analysts and users to query and run reports against all of an agency’s unified data. Databricks SQL integrates with BI tools, like Tableau and Microsoft Power BI, and complements any existing BI tools with a SQL-native interface that allows data analysts and data scientists to query data directly within Databricks.

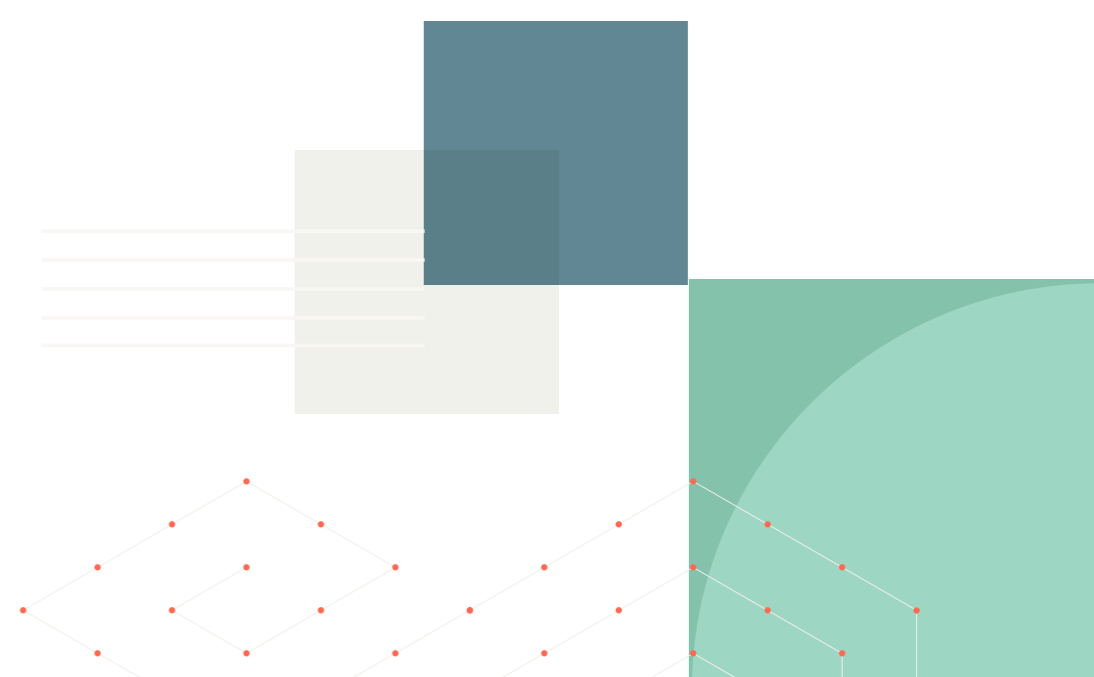
Additionally, with Databricks SQL, a data team can transform insights from real-world data into powerful visualizations designed for machine learning. Visualizations can be turned into interactive dashboards to share insights with peers across agencies, policy makers, regulators and decision-makers.



Ensure data security and compliance at scale

Databricks is fully aware of the sensitivity of the data that many federal agencies are responsible for. From national security and defense data to individual health and financial information to national infrastructure and energy data — all of it is critical. Data needs to be protected at every level of the platform through fine-grained access control and deep integration with cloud-provider access control mechanisms. Databricks Lakehouse is a massively secure and scalable multicloud platform running millions of machines every day. It is independently audited and compliant with FedRAMP security assessment protocols on the Azure cloud and can provide a HIPAA-compliant deployment on both AWS and Azure clouds.

The platform’s administration capabilities include tools to manage user access, control spend, audit usage and analyze activity across every workspace, all while seamlessly enforcing user and data governance, at any scale.





CUSTOMER STORY: CITY OF SPOKANE

Promoting Government Transparency and Accountability With Data

The city of Spokane in Washington state is committed to providing information that promotes government transparency and accountability, and it understands firsthand the challenges of data quality. The city of Spokane deals with an enormous amount of data critical to many of its operations, including financial reports, city council meeting agendas and minutes, issued and pending permits, as well as map and geographic information system (GIS) data for road construction, crime reporting and snow removal.

To abstract their entire ETL process and achieve consistent data through data quality and to master data management services, the city of Spokane leveraged DQLabs and Azure Databricks. They merged a variety of data sources, removed duplicate data and curated the data in Azure Data Lake Storage (ADLS).

Legacy IT architecture limits city-scale data processing

With their legacy architecture, it was nearly impossible for the city of Spokane to obtain operational analytics and real-time reports. They needed a method of

publishing and disseminating city data sets from various sources for analytics and reporting purposes through a central location that could efficiently process data to ensure data consistency and quality.

Leveraging Azure Databricks to improve data quality

DQLabs uses Azure Databricks to process and de-identify both streaming and batch data in real time for data quality profiling. This data is then staged and curated for machine learning models using PySpark MLlib.

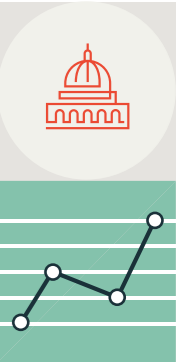
Refined data are published as golden views for downstream analysis, reporting and analytics. Thanks to DQLabs and Azure Databricks, this process is fast and efficient, putting organizations like the city of Spokane in a leading position to leverage their data for operations, decision-making and future planning.

“DQLabs and Azure Databricks enable us to deliver a consistent source of cleansed data to address concerns for high-risk populations and to improve public safety and community planning.”

Eric Finch | Chief Innovation and Technology Officer, City of Spokane

80% improvement in data quality

50% lower total cost of ownership of data operations



Conclusion

State and local governments have the opportunity to greatly improve the day-to-day lives of their citizens by taking advantage of data analytics and AI. The Databricks Lakehouse Platform will unify data, analytics and AI workloads, making these organizations more data-driven and giving policy makers access to deeper, more meaningful insights for decision-making. It will also eliminate data silos and increase communication and collaboration across agencies to ensure the best results for all citizens.



About Databricks

Databricks is the data and AI company. More than 5,000 organizations worldwide — including Comcast, Condé Nast, H&M and over 40% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark,™ Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on [Twitter](#), [LinkedIn](#) and [Facebook](#).

Get started with a free trial of Databricks and start building data applications today

[START YOUR FREE TRIAL](#)

To learn more, visit us at:

dbricks.co/state-local-government

