

Public Sector Guide

to Data + AI Summit

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

Open Your World to Data + AI

Hello!

We are thrilled to have you join us as we explore new innovations together at Data + AI Summit 2021. With 100,000 registrants and more than 200 sessions, we are excited to have you at our biggest event to date. Summit is virtual this year, which gives you the flexibility to explore different sessions from anywhere in the world.

This Public Sector Guide to Summit is designed to ensure you get the most from your experience. We have a stellar lineup of Public Sector sessions covering federal as well as state and local government. We are also featuring a series of interactive solution demos to help you get started innovating with AI.

In addition, we're hosting a Public Sector forum featuring Data + AI leaders from the Department of Defense; the Department of Veterans Affairs; Hennepin County, Minnesota; Centers for Medicaid & Medicare Services; and Booz Allen Hamilton. Our esteemed speakers will share their perspectives on the biggest issues facing government, including the fight against COVID-19, homeland security, and delivering better social services to citizens.

Take the time to virtually network and ask any Databricks team member questions. We are here to help and we love to discuss big ideas. And be sure to check out the [full agenda](#) for the most up-to-date information.

We look forward to a compelling week of Data + AI with you!



Kevin Davis

VP, Public Sector

Contents

This guide is designed to help Data + AI Summit 2021 attendees in the Public Sector industry plan their experience and identify breakout sessions and events of interest. It is intended to complement the full [Data + AI Summit](#) website and in-app experience.

○ Data + AI Summit 2021 Overview	4
○ Create Your Own Experience	4
○ All Things Public Sector at Summit	5
Public Sector Industry Forum	5
Public Sector Talks	6
Interactive Demos	12
○ Summit Checklist	13
○ Ask an Expert	14

Data + AI Summit 2021 Overview

Data + AI Summit is the global event for the data community, where 100,000 practitioners, leaders and visionaries come together to shape the future of data and AI. This guide outlines specific tracks relevant to the Public Sector that you may be interested in.

**Data + AI Summit is virtual this year
– and it's easy to [register here](#) ▶**

May 24–28, 2021

Create Your Own Experience

We have created three different Public Sector experiences so you can get the most from Summit:

- 1 PUBLIC SECTOR FORUMS**
Enjoy our industry-specific keynotes and panel discussions with thought leaders from some of the biggest government agencies and global brands
- 2 TECH TALKS**
Deep-dive into specific topics and technical sessions with a leading expert from your field who has tackled the data and AI problems you are looking to solve
- 2 LIVE DEMOS**
Join interactive Solution Accelerator demos for popular Public Sector use cases and AMA sessions with our Data + AI technical experts

All Things Public Sector at Summit

Use this section to schedule and design your Summit experience around the Public Sector and our **industry forum, talks** and **live demos**.

Public Sector Forum

WEDNESDAY, MAY 26

11:05 AM – 1:05 PM PT

Join us for our most popular industry event, featuring our Public Sector keynotes and a panel discussion with thought leaders from federal, state and local agencies, and systems integrators. Hear firsthand how they are unlocking the power of Data + AI to deliver on their mission objectives and better serve citizens.

KEYNOTE

Accelerating AI Initiatives to Achieve Mission Objectives

JOHN LARSON, Senior Vice President, Analytics, Booz Allen Hamilton

GREG LITTLE, Director, CFO Data Transformation, Department of Defense (DoD)

DEEPALI POWALE, Data Science Lead, Federal Bureau of Investigation (FBI)

PANEL

ANDREW LUCIUS, Principal Data Scientist, Hennepin County, Minnesota

DAVID FULLER, Chief, Data Analytics Division, the Department of Veterans Affairs (VA)

PAUL DONOHOE, Sr. Technical Advisor, Enterprise Database Group, Centers for Medicaid & Medicare Services (CMS)

Public Sector Tech Talks

We have a full agenda of Public Sector talks taking place at Summit. See which ones best fit what you're looking for.

WEDNESDAY, MAY 26

11:30 AM PT

Empower Splunk and Other SIEMs With the Databricks Lakehouse for Cybersecurity

GEORGE WEBSTER, Global Head of Cybersecurity Science and Analytics, HSBC

Cloud, cost, complexity and threat coverage are top of mind for every security leader. The lakehouse architecture has emerged in recent years to help address these concerns with a single unified architecture for all your threat data, analytics and AI in the cloud. In this talk, we will show how lakehouse is essential for effective cybersecurity and popular security use cases. We will also share how Databricks empowers the security data scientist and analyst of the future and how this technology allows cyber data sets to be used to solve business problems.

WEDNESDAY, MAY 26

12:05 PM PT

Creating Reusable Geospatial Pipelines

DAN CORBIANI, Data Scientist and Solutions Architect, Pacific Northwest National Lab

Geospatial pipelines in Apache Spark™ are difficult because of the diversity of data sets and the challenge of harmonizing on a single dataframe. We have worked over the past year to review different pipeline tools that allow us to quickly combine steps to create new workflows or operate on new data sets. We have reviewed Dagster, Apache Spark MLflow pipelines, Prefect, and our own custom solutions. The talk will cover the pros and cons of each solution and will include an actionable workflow implementation that any geospatial analyst can leverage. We will show how we can leverage a pipeline to run a traditional geospatial hot spot analysis. Interactive mapping within the Databricks platform will be demonstrated.

Public Sector Tech Talks

WEDNESDAY, MAY 26

3:50 PM PT

Improving Power Grid Reliability Using IoT Analytics

JIM BROWN, Director of Data Science, Neudesic

MINJIE YU, Data Scientist, DTE Energy

Society depends on reliable utility services to ensure the health and safety of our communities. Electrical grid failures have impact and consequences that can range from daily inconveniences to catastrophic events. In this session, we will demonstrate how we ingest hundreds of millions of quality measures each day from DTE's network of smart electric meters. This data is then further processed in Databricks to detect anomalies, apply graph analytics and spatially cluster these anomalies into "hot spots." Engineers and work management experts use a dashboard to explore, plan and prioritize diverse actions to remediate the hot spots. This allows DTE to prioritize work orders and dispatch crews according to impact on grid reliability. Because of this and other efforts, DTE has improved reliability by 25% year over year.

WEDNESDAY, MAY 26

5:00 PM PT

A Collaborative Data Science Development Workflow

NICK HALE, Senior R&D Specialist, Trillion Technology Solutions

Collaborative data science workflows have several moving parts, and many organizations struggle with developing an efficient and scalable process. Our solution consists of data scientists individually building and testing Kedro pipelines and measuring performance using MLflow Tracking. Once a strong solution is created, the candidate pipeline is trained on cloud-agnostic, GPU-enabled containers. If this pipeline is production worthy, the resulting model is served to a production application through MLflow.

Public Sector Tech Talks

THURSDAY, MAY 27

11:00 AM PT

Consolidating MLOps at One of Europe's Biggest Airports

FLORIS HOOGENBOOM, Lead Data Scientist, Royal Schiphol Group

SEBASTIAAN GRASDIJK, Senior Data Scientist, Royal Schiphol Group

At Schiphol Airport we run a lot of mission-critical machine learning models in production, ranging from models that predict passenger flow to computer vision models that analyze what is happening around the aircraft. Especially now, in times of COVID-19, it is paramount for us to be able to quickly iterate on these models by implementing new features, retraining them to match the new dynamics and, above all, monitoring them actively to see if they still fit the current state of affairs.

In this talk, we will take you through the way we rely on MLflow and how that enables us to release (sometimes) multiple versions of a model per week in a controlled fashion. With this setup, we are achieving the same benefits and speed that you'd find in a traditional software CI pipeline.

THURSDAY, MAY 27

11:35 AM PT

Considerations for Data Access in the Lakehouse

ZACHARY FRIEDMAN, Senior Product Manager, Immuta

Organizations are increasingly exploring lakehouse architectures with Databricks to combine the best of data lakes and data warehouses. Databricks SQL Analytics introduces new innovation on the "house" to deliver data warehousing performance with the flexibility of data lakes. The lakehouse supports a diverse set of use cases and workloads that require distinct considerations for data access. On the lake side, tables with sensitive data require fine-grained access control that is enforced across the raw data and derivative data products via feature engineering or transformations. Whereas on the house side, tables can require fine-grained data access such as row-level segmentation for data sharing, and additional transformations using analytics engineering tools. On the consumption side, there are additional considerations for managing access from popular BI tools such as Tableau, Power BI or Looker.

The product team at Immuta, a Databricks partner, will share their experience building data access governance solutions for lakehouse architectures across different data lake and warehouse platforms to show how to set up data access for common scenarios for Databricks teams new to SQL Analytics.

Public Sector Tech Talks

THURSDAY, MAY 27

3:15 PM PT

Entity Resolution Using Patient Records at CMMI

BRETT LUSKIN, Data Scientist, NewWave

The presentation will focus on the need for and deployment of a Databricks-enabled entity resolution capability at the Center for Medicare & Medicaid Innovation (CMMI) within the Centers for Medicare & Medicaid Services (CMS), the federal agency that is the nation's largest healthcare payer. CMMI builds innovation models that test healthcare delivery and payment systems, integrating and parsing huge data sets with multiple provenances and varied quality. We will explore the specific entity resolution use cases, the ML necessary for this data and the unique uses of Databricks by the federal government and CMS in providing this capability.

THURSDAY, MAY 27

4:15 PM PT

From Vaccine Management to ICU Planning: How CRISP Unlocked the Power of Data During a Pandemic

ANDY HANKS, Analytics Platform Owner, Chesapeake Regional Information System

STEVE DOWLING, Data Engineer, Slalom Consulting

Chesapeake Regional Information System for our Patients (CRISP) is a nonprofit healthcare information exchange (HIE) whose customers include states like Maryland and healthcare providers such as Johns Hopkins. CRISP's work supports the local healthcare community by securely sharing the kind of data that facilitates care and improves health outcomes.

When the pandemic started, the Maryland Department of Health reached out to CRISP with a request: Get us the demographic data we need to track COVID-19 and proactively support our communities. As a result, CRISP employees spent long hours attempting to handle multiple data sources with complex data enrichment processes. To automate these requests, CRISP partnered with Slalom to build a data platform powered by Databricks and Delta Lake.

Using the power of the Databricks Lakehouse Platform and the flexibility of Delta Lake, Slalom helped CRISP provide the Maryland Department of Health with near real-time reporting of key COVID-19 measures. With this information, Maryland has been able to track the path of the pandemic, target the locations of new testing sites, and ultimately improve access for vulnerable communities.

Public Sector Tech Talks

FRIDAY, MAY 28

10:30 AM PT

Building the Foundations of an Intelligent, Event-Driven Data Platform at EFSA

DAVID DE WILDE, Data Science and Engineering Consultant, Delaware Consulting CVBA

SEBASTIAAN LEYSEN, Data Platform Lead, Delaware Consulting CVBA

EFSA is the European agency that provides independent scientific advice on existing and emerging risks across the entire food chain. On March 27, 2021, a new EU regulation (EU 2019/1381) was enacted, requiring EFSA to significantly increase the transparency of its risk assessment processes for all citizens.

To help it comply with this new regulation, Delaware has been supporting EFSA in undergoing a large digital transformation program. We have been designing and rolling-out a modern data platform running on Azure and powered by Databricks.

FRIDAY, MAY 28

11:05 AM PT

Gender Prediction With Databricks AutoML Pipeline

SHARON XU, Senior Data Science Manager, AARP

As the nation's leading advocate for people aged 50+, each month AARP conducts thousands of campaigns made up of hundreds of millions of emails, mails and phone calls to over 37 million members and a broader universe of non-members. Missing information on demographics results in less accurate profiling and targeting strategies. For example, there are 1.5 Million active members and 15 Million expired members missing gender information in AARP's database. The Name gender model is a use case where AARP Data Analytics team utilized the Databricks Lakehouse platform to create a fully automated machine learning model. As a result, AARP dramatically (and accurately) improved the coverage of gender information from about 92.5% to 99.5%.

Public Sector Tech Talks

FRIDAY, MAY 28

11:40 AM PT

10 Things Learned Releasing Databricks Enterprise-Wide

JAKE KULAS, Senior Big Data Developer, Western Governors University

Implementing tools, let alone an entire Lakehouse Platform, like Databricks, can be quite the undertaking. Implementing a tool that you have not yet learned all the ins and outs of can be even more frustrating. Have you ever wished that you could take some of that uncertainty away? Four years ago, Western Governors University (WGU) took on the task of rewriting all our ETL pipelines in Scala/Python, as well as migrating our enterprise data warehouse into Delta Lake, all on the Databricks platform. Through this process, our team has had the chance and opportunity to learn while making a lot of mistakes. Taking a look back at those mistakes, there are a lot of things we wish we had known before opening the platform to our enterprise.

We would like to share with you 10 things we wish we had known before WGU started operating in our Databricks environment.

FRIDAY, MAY 28

11:40 AM PT

Build Real-Time Applications With Databricks Streaming

BENNIE HAELEN, Architect, Insight Digital Innovation

In this presentation, we will study a use case we implemented recently for a large, metropolitan fire department. Our company has already created a complete analytics architecture for the department based upon Azure Data Factory, Databricks, Delta Lake, Azure SQL and Azure SQL Server Analytics Services (SSAS). While this architecture works very well for the department, they would like to add a real-time channel to their reporting infrastructure.

Interactive Demos

Don't miss these live demos for our latest Public Sector Solution Accelerators. You'll have the opportunity to interact with Databricks Solutions Architects and ask them anything!

WEDNESDAY, MAY 26

11:30 AM – 12:00 PM PT

Predicting Opioid Misuse With Databricks and SQL Analytics

Every year, prescription opioid misuse results in unnecessary loss of life and places a massive financial burden on the healthcare system. Advanced analytics can be used to identify and flag anomalous opioid distribution patterns. Join this demo to learn how multiple data personas can collaborate using Databricks and SQL Analytics to ingest large volumes of pharma transaction data, identify statistical outliers, and build dashboards to distinguish and classify suspicious cases of potential opioid misuse.

Solutions Architect: **LUKE BILBRO**

THURSDAY, MAY 27

11:55 AM – 12:25 PM PT

Understanding and Predicting Student Success

In today's learning environment, more students than ever are learning through both in-person and digital means. This demo takes a look at the kinds of data often available to academic institutions and proposes a method for determining students who are at-risk of not matriculating, so that stakeholders can direct interventions and services to them to ensure the best educational outcomes.

Interactive Demos

THURSDAY, MAY 27

3:40 PM – 4:10 PM PT

Healthcare Claims Reporting: Healthcare Claims Analytics for Health and Human Services

As more government entities move to deliver value-based healthcare outcomes, analyzing the cost and complexity of healthcare services has never been more timely. This demo takes a look at integrating disparate types of healthcare encounter claims and transforming them into a patient-centric model, which can then be analyzed along different dimensions.

Solutions Architect: **COLIN RISLER**

THURSDAY, MAY 27

5:40 PM – 6:10 PM PT

Detecting Cyber Criminals Using ML, Threat Intel and pDNS Data

Have you ever wanted to see how Databricks technologies are used in Security Operations? This is a no-jargon demo for security practitioners. This demo will show you how to detect a remote access trojan – from data ingest to alerting.

You will see Databricks Autoloader, Delta, Spark, MLFlow in action. The demo covers day to day work of security practitioners: data ingest,, threat intel enrichments, ad hoc querying, writing analytics, alerting and collaboration.

For an extra deep dive, you can sign up for [Security Threat Detection with Databricks](#). Or [download the notebook](#) for this demo.

MONZY MERZA, VP, Security GTM

Summit Public Sector Checklist

Public Sector Forum

WEDNESDAY, MAY 26

- 11:05 AM – 1:05 PM PT: Keynote and Public Sector Panel

Public Sector Talks

WEDNESDAY, MAY 26

- 11:30 AM PT: Empower Splunk and other SIEMs With the Databricks Lakehouse for Cybersecurity, [HSBC](#)
- 12:05 PM PT: Creating Reusable Geospatial Pipelines, [Pacific Northwest National Lab](#)
- 3:50 PM PT: Improving Power Grid Reliability Using IoT Analytics, [Neudesic | DTE Energy](#)
- 5:00 PM PT: A Collaborative Data Science Development Workflow, [Trillion Technology Solutions](#)

THURSDAY, MAY 27

- 11:00 AM PT: Consolidating MLOps at One of Europe's Biggest Airports, [Royal Schiphol Group](#)
- 11:35 AM PT: Considerations for Data Access in the Lakehouse, [Immuta](#)
- 3:15 PM PT: Entity Resolution Using Patient Records at CMMI, [NewWave](#)
- 4:15 PM PT: From Vaccine Management to ICU Planning: How CRISP Unlocked the Power of Data During a Pandemic, Chesapeake Regional Information System for our Patients (CRISP)

FRIDAY, MAY 28

- 10:30 AM PT: Building the Foundations of an Intelligent, Event-Driven Data Platform at EFSA, [Delaware Consulting CVBA](#)
- 11:05 AM PT: Gender Prediction With Databricks AutoML Pipeline, [AARP](#)
- 11:40 AM PT: 10 Things Learned Releasing Databricks Enterprise-Wide, [Western Governors University](#)
- 11:40 AM PT: Build Real-Time Applications With Databricks Streaming, [Insight Digital Innovation](#)

Interactive Demos

- MAY 26, 11:30 AM: Predicting Opioid Misuse With Databricks and SQL Analytics
- MAY 27, 11:55 AM: Understanding and Predicting Student Success
- MAY 27, 3:40 PM: Healthcare Claims Reporting: Healthcare Claims Analytics for Health and Human Services
- MAY 27, 5:40 PM: Detecting Cyber Criminals With ML, Threat Intel and pDNS Data

Ask an Expert

Government solutions tour and live AMA

WEDNESDAY, MAY 26

2:30 PM – 3:30 PM PT

If you'd like to learn more as a follow up to any of the sessions, please reach out to your account manager to coordinate a meeting with:

- Public Sector leaders and subject matter experts
- Technical experts in Public Sector data
- Solutions Architects who have helped hundreds of customers

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

#DataAISummit