

Healthcare and Life Sciences 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 the Data + AI Summit 2021! With 100,000 registrants and more than 200 sessions, we are excited to have you for 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 Health and Life Sciences Guide to Data + AI Summit is designed to ensure you get the most out of your experience. We have a stellar lineup of Health and Life Sciences sessions led by GlaxoSmithKline, Sanofi, Takeda, Centers for Medicare & Medicaid, and many others in the industry. We are also featuring a series of interactive solution demos to help you get started innovating.

In addition, we're hosting a Health and Life Sciences Forum featuring Data + AI leaders from Humana, The Department of Veteran's Affairs, Providence St. Joseph's Health, and Aetion. Our esteemed list of speakers will share their perspectives on the biggest issues facing the industry including health equity, displacement of care, real-world evidence, and digital engagement.

Take the time to virtually network and ask any Databricks team members questions: We are here to help and 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!



Mike Sanky

Global Industry Lead,
Health and Life Sciences

Contents

This guide is designed to help Data + AI Summit 2021 attendees in the Healthcare and Life Sciences 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 Healthcare and Life Sciences at Summit	5
Healthcare and Life Sciences Industry Forum	5
Healthcare and Life Sciences 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 Healthcare and Life Sciences 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 Healthcare and Life Sciences experiences so that you can get the most from Summit:

- 1 HEALTHCARE AND LIFE SCIENCES FORUMS**
Enjoy our industry-specific keynotes and panel discussions with thought leaders from some of the biggest 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 Healthcare and Life Sciences use cases and AMA sessions with our data + AI technical experts

All Things Healthcare and Life Sciences at Summit

Use this section to schedule and design your Summit experience around Healthcare and Life Sciences consisting of our **industry forum, talks** and **live demos**.

Healthcare and Life Sciences Forum

THURSDAY, MAY 27

11:00 AM – 1:00 PM PT

Join us for our most popular industry event, featuring our Healthcare and Life Sciences keynotes and panel discussion with thought leaders from some of the biggest global brands. Hear first-hand how they are unlocking the power of data + AI to accelerate drug discovery and improve patient care. Discussion topics include health equity, novel uses of real-world data, improving patient engagement and new healthcare delivery models.

KEYNOTE

The Opportunity for Real-world Data

CAROLYN MAGILL, CEO, AETION

PANEL

DEEPAK SADAGOPAN, SVP, Population Health Informatics, Providence St. Joseph's Health

DR. JOHN S. SCOTT, Physician Informaticist, The Department of Veterans Affairs

SLAWEK KIERNER, SVP, CDAO, Humana

CAROLYN MAGILL, CEO, Aetion

Healthcare and Life Sciences Tech Talks

We have a full agenda of Healthcare and Life Sciences talks taking place at Data + AI Summit. See which ones best fit what you're looking for.

WEDNESDAY, MAY 26

11:30 AM PT

RWE & Patient Analytics Leveraging Databricks – A Use Case

HARINI GOPALAKRISHNAN, Director, Sanofi

MARTIN LONGPRE, Architect, Sanofi

Gaining insights and knowledge from real-world health data (RWD) has been an area of continued opportunity for pharma organizations. This session dives into what real-world data and real-world evidence is and how it is generated, what value it drives for life sciences in general and what kind of analytics are performed. The team at Sanofi will discuss some considerations and challenges related to data security, privacy, and industrialization of a big data platform hosted in the cloud. They'll also share how they use Databricks to perform big data ingestion including advantages over native AWS Batch/Glue and examples of some of the advanced analytics use cases that leveraged Databricks for RWE.

WEDNESDAY, MAY 26

3:15 PM PT

Advanced Natural Language Processing With Apache Spark NLP

DAVID TALBY, Chief Technology Officer, John Snow Labs

This hands-on deep-dive session uses the open-source Apache Spark NLP library to explore advanced NLP in Python. Apache Spark NLP provides state-of-the-art accuracy, speed, and scalability for language understanding by delivering production-grade implementations of some of the most recent research in applied deep learning.

You'll edit and run executable Python notebooks as we walk through these common NLP tasks: document classification, named entity recognition, sentiment analysis, spell checking and correction, grammar understanding, question answering, and translation.

Healthcare and Life Sciences Tech Talks

WEDNESDAY, MAY 26

4:25 PM PT

FlorenceAI: Reinventing Data Science at Humana

DAVID MACK, Cognitive and Machine Learning Principal, Humana

Humana strives to help communities and individual members achieve their best health – no small task in the past year! Rethinking their existing operations and reimagining what a collaborative ML platform for hundreds of data scientists might look like, they built FlorenceAI to automate and accelerate the delivery lifecycle of data science solutions at scale. In this presentation, they will walk through an end-to-end example of how to build a model at scale on FlorenceAI and deploy it to production. Tools highlighted include Azure Databricks, MLFlow, AppInsights, and Azure Data Factory.

WEDNESDAY, MAY 26

4:25 PM PT

Drug Repurposing Using Deep Learning on Knowledge Graphs

ALEX THOMAS, principal data scientist, Wisecube

VISHNU VETTRIVEL, Founder and CTO, Wisecube

Discovering new drugs is a lengthy and expensive process. This means that finding new uses for existing drugs can help create new treatments in less time and with less time. The difficulty is in finding these potential new uses. How do we find these undiscovered uses for existing drugs? In this talk, we discuss how to unify the available structured and unstructured data sets into a knowledge graph and use deep learning techniques to predict latent relationships.

WEDNESDAY, MAY 26

5:00 PM PT

Wizard Driven AI Anomaly Detection With Databricks in Azure

NAOMI KADUWELA, Head of Kavi Labs, Kavi Global

Fraud is prevalent in every industry, and growing at an increasing rate, as the volume of transactions increases with automation. The National Healthcare Anti-Fraud Association estimates \$350B of fraudulent spending. At the same time as fraud and anomaly detection use cases are booming, the skills gap of expert data scientists available to perform fraud detection is widening.

The Kavi Global team will present a cloud native, wizard-driven AI anomaly detection solution, enabling Citizen Data Scientists to easily create anomaly detection models to automatically flag Collective, Contextual, and Point anomalies, at the transaction level, as well as collusion between actors.

Healthcare and Life Sciences Tech Talks

THURSDAY, MAY 27

11:35 AM PT

Empowering Real Time Patient Care Through Spark Streaming

ARNAV CHAUDHARY, Manager, Takeda

Takada's Plasma Derived Therapies (PDT) business unit has recently embarked on a project to use Spark Streaming on Databricks to empower how they deliver value to their Plasma Donation centers. As patients come in and interface without clinics, we store and track all of the patient interactions in real time and deliver outputs and results based on said interactions. The current problem with our existing architecture is that it is very expensive to maintain and has an unsustainable number of failure points. In this talk, we will discuss how we use Spark Streaming and Delta Lake enables us to replace our existing ETL processes into a purely stream driven architecture.

THURSDAY, MAY 27

11:35 AM PT

How Machine Learning and AI Can Support the Fight Against COVID-19

DMITRY SOSHNIKOV, Director, Microsoft

FRANCESCA LAZZERI, Director, Microsoft

In this session, we show how to leverage CORD dataset, containing more than 400000 scientific papers on COVID and related topics, and recent advances in natural language processing and other AI techniques to generate new insights in support of the ongoing fight against this infectious disease.

THURSDAY, MAY 27

3:15 PM PT

Entity Resolution Using Patient Records at CMMI

BRETT LUSKIN, Data Scientist, NewWave

The presentation will discuss 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 government agency that is also the nation's largest healthcare payer. CMMI builds innovation models that test healthcare delivery and payment systems, integrating and parsing huge datasets with multiple provenance and quality. We will explore the specific entity resolution use cases, the ML necessary for this data and the unique uses of Databricks for the federal government and CMS in providing this capability.

Healthcare and Life Sciences Tech Talks

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 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.

FRIDAY, MAY 28

10:30 AM PT

Towards Personalization in Global Digital Health

AFRICA PERIANEZ, Founder and CEO, benshi.ai

The rapid expansion of mobile phone usage in low-income and middle-income countries has created unprecedented opportunities for applying AI to improve individual and population health. We aim to do so by providing personalized predictions and recommendations to support diagnosis to medical care teams and frontline workers, as well as to nudge patients through personalized incentives towards an improvement in disease treatment management and general wellness. This talk will focus on how we overcome technical challenges to build a state-of-the-art machine learning platform, on Databricks, that serves to improve global health outcomes.

Healthcare and Life Sciences Tech Talks

FRIDAY, MAY 28

10:30 AM PT

Drug and Vaccine Discovery: Knowledge Graph + Apache Spark

JOHN HUNTER, Director, GlaxoSmithKline

GSK has set out to build the world's largest medical knowledge graph to provide our scientists access to the world's medical knowledge, and also enable machine learning to infer links between facts. To power RDF sub-graphing, GSK has developed a set of open-source libraries codenamed "Project Bellman" that enable Sparql queries over partitioned RDF data in Apache Spark. These tools are used by both GSK's Ai/ML team to discover gene to disease mappings, and GSK's scientists to query over the world's medical knowledge.

FRIDAY, MAY 28

11:05 AM PT

Managing R&D Data on Parallel Compute Infrastructure

ANDREW BROWN, Manager, ZS Associates

Clinical genomic analytics pipelines using Databricks and the Delta Lake for the benefit of loading individual reads from raw sequencing or base-call files have significant advantages over more traditional methods. Analysis pipelines that perform genomic mapping to purpose-built reference data artifacts persisted to tables allows for enhanced performance that is magnitudes greater than previous mapping methods. These scalable, reproducible, and potentially open sourced methods have the ability to transform bioinformatics and R&D data management/governance.

FRIDAY, MAY 28

11:40 AM PT

Automatic ICD-10 Code Assignment to Consultations

JOINAL AHMED, Data Scientist, Halodoc Technologies**NIRAV KUMAR**, Director of Data Science, Halodoc Technologies

This session aims at constructing a machine learning model for ICD-10 coding, where the model automatically determines the corresponding diagnosis codes solely based on free-text medical notes. By applying Natural Language Processing (NLP) and Recurrent Neural Network (RNN) architecture, this model can significantly reduce human resources in coding time compared with a professional coder.

Interactive Demos

Don't miss these live demos for our latest Healthcare and Life Sciences solution accelerators with the opportunity to interact and ask Databricks Solution Architects 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.

Solution Architect: **LUKE BILBRO**

WEDNESDAY, MAY 26

12:30 PM - 1:00 PM PT

Building a Health Lakehouse

Healthcare and life sciences organizations generate petabytes of structured and unstructured data. By aggregating this data into a health lakehouse, organizations can create a single longitudinal view of patient health. In this demo we demonstrate how to build an OMOP common data model in delta and SQLA

Healthcare and Life Sciences, Technical Lead: **AMIR KERMANY**

THURSDAY, MAY 27

2:40 PM - 3:10 PM PT

Solving the n+1 Problem in Genomics With Glow & Delta Lake

Glow is an open-source library for aggregating and mining genomics variant data on population scales. Here we will demonstrate how to continuously integrate new genomes into Delta Lake, optimize them for querying, and perform bioinformatics analytics.

Genomics Technical Lead: **WILLIAM BRANDLER**

Interactive Demos

THURSDAY, MAY 27

3:40 PM – 4:10 PM PT

Healthcare Claims Analytics for Health and Human Services

As more government entities move to deliver value-based healthcare outcomes, analyzing the cost and complexion 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.

Genomics Technical Lead: **COLIN RISLER**

THURSDAY, MAY 27

5:10 PM – 5:40 PM PT

Burning Through Electronic Health Records in Real Time With Smolder

Smolder is an open-source library for ingesting EHR data in HL7 format in real-time. In this demo we show how to create a bronze delta layer directly from streams of HL7 messages.

Solutions Architect: **RYAN DECOSMO**

FRIDAY, MAY 28

11:50 AM – 12:20 PM PT

Improving Drug Discovery With QSAR Models at Scale

Modern drug discovery sifts through petabytes of structured data to deliver positive drug outcomes against protein targets. In the past, this process leveraged specialty hardware and low level programming languages to achieve practical performance. In this demo, we demonstrate a paradigm shift from the older message passing interface (MPI) structure of in-silico drug discovery to a distributed MapReduce data flow paradigm. Essentially, we demonstrate how we can harness common cloud technologies to scale and speed up modern drug discovery.

Solutions Architect: **MARK LEE**

Summit Healthcare and Life Sciences Checklist

Healthcare and Life Sciences Forum

THURSDAY, MAY 27

- 11:00 AM – 1:00 PM PT: Keynote and Healthcare and Life Sciences Panel

Healthcare and Life Sciences Talks

WEDNESDAY, MAY 26

- 11:30 AM PT: RWE & Patient Analytics Leveraging Databricks – A Use Case, [Sanofi](#)
- 3:15 PM PT: Advanced Natural Language Processing With Apache Spark NLP, [John Snow Labs](#)
- 4:25 PM PT: FlorenceAI: Reinventing Data Science at Human, [Humana](#)
- 4:25 PM PT: Drug Repurposing Using Deep Learning on Knowledge Graphs, [Wisecube](#)
- 5:00 PM PT: Wizard Driven AI Anomaly Detection With Databricks in Azure, [Kavi Global](#)

THURSDAY, MAY 27

- 11:35 AM PT: Empowering Real Time Patient Care Through Spark Streaming, [Takeda](#)
- 11:35 AM PT: How Machine Learning and AI Can Support the Fight Against COVID-19, [Microsoft](#)
- 3:15 PM PT: Entity Resolution Using Patient Records at CMML, [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: Towards Personalization in Global Digital Health, [benshi.ai](#)
- 10:30 AM PT: Drug and Vaccine Discovery: Knowledge Graph + Apache Spark, [GlaxoSmithKline](#)
- 11:05 AM PT: Managing R&D Data on Parallel Compute Infrastructure, [ZS Associates](#)
- 11:40 AM PT: Automatic ICD-10 Code Assignment to Consultations, [Halodoc Technologies](#)

Interactive Demos

- MAY 26, 11:30 AM: Predicting Opioid Misuse with Databricks and SQL Analytics
- MAY 26, 12:30 PM: Building a Health Lakehouse
- MAY 27, 2:40 PM: Solving the n+1 Problem in Genomics With Glow & Delta Lake
- MAY 27, 3:40 PM: Healthcare Claims Analytics for Health and Human Services
- MAY 27, 5:40 PM: Burning Through Electronic Health Records in Real Time With Smolder
- MAY 28, 11:50 AM: Improving Drug Discovery With QSAR Models at Scale

Ask an Expert

Healthcare and life sciences solutions tour and live AMA

THURSDAY, MAY 27

4:10 PM – 5:10 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.

- Healthcare and Life Sciences leaders and subject matter experts
- Technical Experts of Data in Healthcare and Life Sciences
- Solutions Architects who have helped hundreds of customers

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

#DataAISummit