

Financial Services Guide

to Data + AI Summit

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

Open Your World to Data + AI

Hello!

You're invited to open your world to the new innovations at Data + AI Summit 2021. With 100,000 registrants and more than 200 sessions, we're 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 Financial Services Guide to Data + AI Summit is designed to ensure you get the most from your experience. We have organized a stellar lineup of financial services sessions with industry leaders, including Capital One, J.P. Morgan, Northwestern Mutual, S&P Global and many more. We are also featuring a series of interactive solution demos to help you get started innovating with AI.

In addition, we have put together a financial services breakout session, featuring a keynote from the CDO of Northwestern Mutual, as well speakers from Intuit, ABN AMRO and KX.

Take the time to virtually network and ask any Databricks team members, aka Bricksters, questions – we are excited to help you! 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!



Junta Nakai

RVP and Global Lead Industry
Leader, Financial Services

Contents

This guide is designed to help Data + AI Summit 2021 attendees in the financial services industry plan their experience and identify breakout sessions and events of interest. It's 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 Financial Services at Summit	5
Financial Services Industry Forum	5
Financial Services Talks	6
Interactive Demos	9
○ Summit Checklist	11
○ Ask an Expert	12

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 financial services 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 four different financial services experiences so that you can get the most from Summit:

1

MAINSTAGE KEYNOTE

A thought leader from your industry will speak on our mainstage about topics ranging from data and AI to finance and beyond

2

FINANCIAL SERVICES FORUMS

Enjoy our industry-specific keynotes and panel discussions with thought leaders from some of the biggest global brands

3

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

4

LIVE DEMOS

Join an interactive Solution Accelerator demo for popular financial services and AMA sessions with our data and AI experts

All Things Financial Services at Summit

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

Mainstage Keynote

THURSDAY, MAY 27

1:00 PM PT

KEYNOTE

MANUELA VELOLSO, Head AI Research, JP Morgan

Dr. Veloso will speak about the intersection of machine learning and finance. She is the head of J.P. Morgan AI Research and a professor of computer science at Carnegie Mellon University, where she previously led the Machine Learning Department.

Financial Services Forum

THURSDAY, MAY 27

6:00 AM – 8:00 AM PT

Join us for our most popular industry event, featuring financial services keynotes and a panel discussion with thought leaders from some of the biggest global brands. Hear first-hand how they are unlocking the power of data and AI to minimize risk and create engaging and prosperous services for their customers.

KEYNOTE

What Do Insurance, Real Estate and Baseball Have in Common?

DON VU, Chief Data Officer, Northwestern Mutual

PANEL

MARCEL KRAMER, Head of Data Engineering, ABN AMRO

PRIYANK PATWA, M&G, Head of ESG & Research Transformation

BHARATH RAMARATHINAM, Principal Engineer, Fintech Platforms, Intuit

SPEAKER

CONOR TWOMEY, Manager Director-North America, KX Systems

Financial Services Tech Talks

We have a full agenda of financial services talks taking place at Data + AI Summit. See which ones best fit what you're looking for – or explore a new area of interest.

WEDNESDAY, MAY 26

3:15 PM PT

Commercializing Alternative Data

JAY BHANKHARIA, S&P Global | SRINIVASA PODUGU, S&P Global

An end-to-end walk-through of how S&P Global ingests, structures and links data to make it more usable and then builds out Sandbox workspaces using the unified analytics platform for clients to conduct research and analysis on the data.

WEDNESDAY, MAY 26

3:50 PM PT

Empowering Zillow's Developers With Self-Service ETL

DEREK GORTHY, Zillow | YUAN FENG, Zillow

Learn why Zillow created two separate interfaces to meet the needs of different user groups.

Additionally, the talk will cover:

- What degree of abstraction from the orchestration, deployment, processing and other ancillary tasks was chosen for each user group
- How they leveraged internal services and packages, including their Apache Spark™ package – Pipeler – to democratize the creation of high-quality, reliable pipelines within Zillow

WEDNESDAY, MAY 26

4:45 PM PT

Productionalizing Machine Learning Solutions With Proper Mon

SUMANTH VENKATASUBBAIAH, Intuit | PANKAJ RASTOGI, Intuit

Intuit products increasingly rely on AI solutions to drive in-product experiences and customer outcomes (a realization of Intuit's AI-driven expert platform strategy). In order to provide complete confidence to Intuit customers through reliable and predictable experiences, we need to ensure the health of all AI solutions by continuously monitoring, managing and understanding them within Intuit products. In this Summit talk, we will share the different open source solutions we evaluated and present the challenges of building this central system. We will also share the overall architecture and the internals of the system.

Financial Services Tech Talks

THURSDAY, MAY 27

12:10 PM PT

Credit Card Fraud Detection Using ML in Databricks

BADRISH DAVAY, Capital One | **NEIL ALLEN**, Capital One

Fraud detection is the task of determining whether a card has been used by the cardholder. One of the methods for recognizing fraudulent card usage is to leverage machine learning (ML) models. In order to more dynamically detect fraudulent transactions, one can train ML models on data sets that include credit card transaction information as well as the card and demographic information of the owner of the account. Capital One aims to achieve this goal while leveraging Databricks, which will be the focus of discussion.

THURSDAY, MAY 27

3:50 PM PT

Scaling Privacy in a Spark Ecosystem

AARON COLCORD, Northwestern Mutual | **DON BOSCO DURAI**, Privacera

Privacy has become one of the most critical topics in data today. More than just a matter of how we ingest and consume data, it deals with the important factors in protecting your customer's rights while balancing business needs. In our session, we will bring Privacera's CTO, Don Bosco Durai, together with Aaron Colcord of Northwestern Mutual to detail an important use case in privacy and then show how to effortlessly scale privacy with a focus on business needs.

THURSDAY, MAY 27

3:50 PM PT

Using MLflow to Help Deduplicate Data at Scale

ROBIN OLIVA-KRAFT, Intuit | **MAYA LIVSHITS**, Intuit

At Intuit, we have a lot of data – and a lot of duplicate data collected over decades. So we built a rule-based, self-serve tool to identify and merge duplicate records. It takes experimentation and iteration to get deduplication right for hundreds of millions of records, and spreadsheet-based tracking just wasn't enough. We now use MLflow to automatically capture execution notes, rule settings, weights, key validation metrics, etc., all without requiring end-user action. In this talk, we'll discuss our use case and why MLflow is useful outside its traditional MLOps use cases.

Financial Services Tech Talks

THURSDAY, MAY 27

5:00 PM PT

Large-Scale Lakehouse Implementation Using Structured Streaming

TOMASZ MAGDANSKI, Asurion

Business leads, executives, analysts and data scientists rely on up-to-date information to make business decisions, adjust to the market, meet the needs of their customers or run effective supply chain operations. Come hear how Asurion used Delta, Structured Streaming, AutoLoader and SQL Analytics to improve production data latency from day-minus-one to near real-time. Asurion's technical team will share battle-tested tips and tricks you only get with a certain scale. Asurion's data lake executes 4,000+ streaming jobs and hosts over 4,000 tables in a production data lake on AWS.

FRIDAY, MAY 28

11:40 AM PT

Fully Utilizing Spark for Data Validation

KEVIN KHO, Prefect

In this talk, we'll show how data validation is possible with a library called Fugue, an open-source framework that lets users port native Python code or pandas code to Spark. We will show an interactive demo of how to extend Pandera (or any other pandas-based data validation library) to a Spark workflow. There is also a deficiency in the current frameworks that we'll address in the demo. With big data, there is a need to apply different validation rules for each partition. For example, data that encompasses a lot of geographic regions may have different acceptable ranges of values (think of currency). Since the current frameworks are designed to apply a validation rule to the whole DataFrame, this can't be done. Using Fugue and Pandera, we can apply different validation rules to each partition of data.

Interactive Demos

Don't miss these live demos of our latest Financial Services Solution Accelerators. You'll have the opportunity to interact with Databricks solutions architects and ask them anything!

WEDNESDAY, MAY 26

12:00 PM – 12:30 PM PT

Combining Rules-Based Systems With AI to Combat Fraud

Building a fraud prevention framework often goes beyond just creating a highly accurate machine learning model due to an ever-changing landscape and customer expectations. Oftentimes, it involves a complex ETL process with a decision science setup that combines a rules engine with an ML platform.

Solutions Architect: **SRI GHATTAMANENI**

WEDNESDAY, MAY 26

4:50 PM – 5:20 PM PT

Modernizing an Investment Data Platform and Time Series Processing

As investors gain access to and trade alternative assets such as cryptocurrency, trading volumes have skyrocketed and created new data challenges. Scalable AI (such as volatility forecasting) is difficult to achieve due to high maintenance costs and the lack of a blueprint for scale and, hence, trading profitability.

Lead Solutions Architect: **RICARDO PORTILLA**

Interactive Demos

THURSDAY, MAY 27

12:25 PM – 12:55 PM PT

Understanding YOLO, Stonks and Diamond Hands

Alternative data always offers a competitive advantage to those brave enough to mine it. Retail investment strategies are widely discussed on public forums and could be used by institutional investors as a leading indicator to near-term events. In this solution centered around NLP, we demonstrate how sentiment analysis and entity resolution can be used to extract machine-readable insights from online public discussions.

Solutions Architects: **JOSH SEIDEL, MILOS COLIC AND ASHLEY TRAINOR**

FRIDAY, MAY 28

11:20 AM – 11:50 AM PT

Understanding Customer Spending Behavior With Geospatial Clustering

As consumers become more digitally engaged, large FSIs often have access to real-time GPS coordinates of every purchase made by their customers. With around 40 billion card transactions processed in the United States every year, retail banks have a lot of data they can leverage to better understand the transaction behaviors of customers opting into GPS-enabled banking applications.

Financial Services, Technical Director: **ANTOINE AMEND**

Summit Financial Services Checklist

Financial Services Forum

THURSDAY, MAY 27

- 6:00 AM – 8:00 AM PT: Keynote and Financial Services Panel

Financial Services Talks

WEDNESDAY, MAY 26

- 3:15 PM PT: Commercializing Alternative Data, [S&P Global](#)
- 3:50 PM PT: Empowering Zillow's Developers With Self-Service ETL, [Zillow](#)
- 4:45 PM PT: Productionizing Machine Learning Solutions With Proper Mon, [Intuit](#)

THURSDAY, MAY 27

- 12:10 PM PT: Credit Card Fraud Detection Using ML in Databricks, [Capital One](#)
- 3:50 PM PT: Scaling Privacy in a Spark Ecosystem, [Northwestern Mutual and Privacera](#)
- 3:50 PM PT: Using MLflow to Help Deduplicate Data at Scale, [Intuit](#)
- 5:00 PM PT: Large-Scale Lakehouse Implementation Using Structured Streaming, [Asurion](#)

FRIDAY, MAY 28

- 11:40 AM PT: Fully Utilizing Spark for Data Validation, [Prefect](#)

Interactive Demos

- MAY 26, 12:00 PM PT: Combining Rules-Based Systems With AI to Combat Fraud
- MAY 26, 4:50 PM PT: Modernizing Investment Data Platform and Time Series Processing
- MAY 27, 12:25 PM PT: Understanding YOLO, Stonks and Diamond Hands
- MAY 28, 11:20 AM PT: Understanding Customer Spending Behavior With Geospatial Clustering

Ask an Expert

Financial services solutions tour and live AMA

THURSDAY, MAY 27

10:55 AM – 11:55 AM 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:

- Financial services leaders and subject matter experts
- Technical experts of data in financial services
- Solutions architects who have helped hundreds of customers

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

#DataAISummit