

Retail and Consumer Goods 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 Retail and Consumer Goods (CPG) Guide to Summit is designed to ensure you get the most from your experience. We have organized a stellar lineup of Retail and CPG sessions with industry leaders, including Apple, H&M, Walmart Labs, Anheuser-Busch, Stitch Fix 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 Retail and CPG breakout session, featuring a keynote from Reckitt, as well speakers from Shell, Mars, Gousto, Albertsons and H&M.

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!



Rob Saker

RVP and Global Industry Leader,
Retail and Consumer Goods

Contents

This guide is designed to help Data + AI Summit 2021 attendees in the Retail and Consumer Goods 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 Retail and Consumer Goods at Summit	5
Retail and Consumer Goods Industry Forum	5
Retail and Consumer Goods Talks	6
Interactive Demos	11
○ Summit Checklist	12
○ Ask an Expert	13

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 Retail and Consumer Goods 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 Retail and Consumer Goods 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 retail, CPG and beyond
- 2 RETAIL AND CONSUMER GOODS 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 Retail and Consumer Goods use cases and AMA sessions with our data and AI technical experts

All Things Retail and Consumer Goods at Summit

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

Retail and Consumer Goods Forum

WEDNESDAY, MAY 26

11:00 AM – 1:00 PM PT

Join us for our most popular industry event, featuring our Retail and Consumer Goods keynotes and panel discussion with thought leaders from some of the largest brands in the world. Hear firsthand how they are unlocking the power of data and AI to win market share and reduce costs by better understanding their consumers.

PANEL

ALBERTO ROSSI, Global Head of Retail Data and Analytics, Shell

DEEPAK JOSE, Head of Business Strategy and Analytics, Mars

COLLEEN QIU, VP, Head of Data Science, Albertsons

ERROL KOOLMEISTER, Head of AI Foundation, H&M

ROBERT BARHAM, Director of Data, Gousto

Retail and Consumer Goods Tech Talks

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

WEDNESDAY, MAY 26

11:30 AM PT

Data Distribution and Ordering for Efficient Data Source V2

ANTON OKOLNYCHYI, Software Engineer, Apple

This talk is going to be useful for developers and data engineers who either develop their own data sources or work with existing ones in Apache Spark™. The presentation will start with an overview of the Catalog API introduced in Spark 3, followed by its benefits and current limitations compared to the old Data Source API. The main focus will be on an extension to the Catalog API developed in SPARK-23889, which lets implementations control how Spark distributes and orders incoming records before passing them to the sink.

WEDNESDAY, MAY 26

11:30 AM PT

Scaling AI at H&M

BJORN HERTZBERG, Lead Data Scientist, H&M

This session is a continuation of "Apply MLOps at Scale" at Data + AI Summit Europe 2020 and "Automated Production Ready ML at Scale" at Spark + AI Summit Europe 2019. In this session, you'll learn how H&M is continuing to evolve and develop their AI platform in order to democratize and accelerate AI usage across the full H&M group, including speed to production, data abstraction, feature store, pipeline orchestration, etc.

Retail and Consumer Goods Tech Talks

WEDNESDAY, MAY 26

12:05 PM PT

Data Security at Scale Through Spark and Parquet Encryption

GIDON GERSHINSKY, Lead Systems Architect, Apple | **TIM PERELMUTOV**, Data Engineer, Apple

Big data presents new challenges for the protection of privacy and the integrity of sensitive information. A straightforward application of traditional file encryption and MAC techniques can't cope with the staggering volumes of data, flowing in modern analytic pipelines. Apple addresses these challenges by leveraging the new capabilities in the Apache Parquet format. In this talk, Gidon Gershinsky and Tim Perelmutov will outline the challenges of protecting the privacy of data at scale and describe the Apache Parquet encryption technology security approach. We will give a quick intro to usage of Apache Parquet encryption API in pure Java and in Apache Spark applications. We will also discuss the roadmap of the community work on new encryption features and on deeper integration with Apache Spark and other analytic frameworks. Finally, we will show a demo of the Apache Parquet modular encryption in action, sharing our learnings from using it at scale.

WEDNESDAY, MAY 26

12:05 PM PT

Automating Data Quality Processes at Reckitt

RICHARD CHADWICK, Data Engineering Consultant, Cervello | **KAROL SAWICZ**, IT Business Analyst, RB

Reckitt is a fast-moving consumer goods company with a portfolio of famous brands and over 30,000 employees worldwide. With that scale, small projects can quickly grow into big data sets, and processing and cleaning all that data can become a challenge. In this talk, we discuss our approach to maintaining high-quality data products and share examples of how we automate data quality processes.

WEDNESDAY, MAY 26

5:00 PM PT

AI-Driven Personalized Email Marketing

XI LIANG, Data Science Lead, Endeavour Drinks

Endeavour Groups Limited (EGL) is Australia's largest drinks retailer with over 2,000 venues across a wide range of brands. The EGL team has built a personalization marketing engine using Databricks that performs segmentation of EGL's 4.5 million loyalty members. This session will dive into how EGL has been able to develop effective brand campaigns that target the right people at the right time with the right messages to millions of their loyalty members.

Retail and Consumer Goods Tech Talks

WEDNESDAY, MAY 26

5:00 PM PT

Offer Recommendation System With Apache Spark at Burger King

KAI HUANG, Software Engineer, Intel | **LUYANG WANG**, Sr. Manager Data Science, Burger King

Offers are important sales drivers in the fast-food industry. Being able to segment customers based on their offer preferences and assign the best offer sets to each segment is critical for a customer-centric recommendation system to enhance the user's shopping experience.

At Burger King, we have developed our own offer recommendation system that leverages pretrained BERT and Inception models on Apache Spark to extract feature representations directly from offer descriptions and images, followed by Spark MLlib to form the user segmentations. In this session, we'll discuss our offer recommendation system in detail.

MAY 27, THURSDAY

11:00 AM PT

Structured Streaming Use-Cases at Apple

KRISTINE GUO, Developer, Apple | **LIANG-CHI HSIEH**, Software Engineer, Apple

Structured streaming plays an important role in current data infrastructure. In response to tremendous streaming requirements, we have actively worked on developing structured streaming in Spark in the past few months. In this talk, Kristine Guo and Liang-Chi Hsieh will detail some of the issues that arose when applying structured streaming and what was done to address them.

MAY 27, THURSDAY

11:35 AM PT

Building a Product Assortment Recommendation Engine

ETHAN DUBOIS, Senior Software Engineer, AB InBev | **JUSTIN MORSE**, Staff Data Scientist, Anheuser Busch

Amid the increasingly competitive brewing industry, the ability of retailers and brewers to provide optimal product assortments for their consumers has become a key goal for business stakeholders. At AB InBev, we approach this selection problem through a two-step method rooted in statistical learning techniques. In this talk, we will share how we use Databricks to scale to over 100,000 brick-and-mortar retailers across the United States and online platforms.

Retail and Consumer Goods Tech Talks

THURSDAY, MAY 27

11:35 AM PT

ChakraView – A 360° Approach to Data Quality

KEERTHIKA THIYAGARAJAN, Developer, Flipkart | SHANKAR MANIAN, Director, Flipkart

The availability of high-quality data is central to the success of any organization in the current era. As every organization ramps up their collection and storage of data, its usefulness largely depends on confidence in its quality. For the Financial Data Engineering team at Flipkart, where the bar for data quality is 100% correct and complete, this problem takes on a whole different dimension. In this presentation, we discuss how we developed a comprehensive data quality framework with the assumption that the people interested and involved in fixing these issues are not necessarily data engineers. Our framework has been developed to be largely config-driven with pluggable logic for categorization and cleaning. We will then talk about how it helped achieve scale in fixing the data quality issues and helped reduce many repeated issues.

THURSDAY, MAY 27

11:35 AM PT

The Rise of ZStandard: Apache Spark/Parquet/ORC/Avro

DONGJOON HYUN, Software Engineer, Apple | PANG WU, Software Engineer, Apple

Zstandard is a fast compression algorithm that you can use in various ways within Apache Spark. In this talk, we will summarize the evolution of Apache Spark in this area and cover the following:

- Optimizing Spark local disk IO
- Event log compression
- The benefits of using Zstandard data file compression
- How Zstandard can be used to serialize/deserialize MapStatus data instead of Gzip

THURSDAY, MAY 27

3:50 PM PT

Modularized ETL Writing With Apache Spark

NEELES SHALIAN, Software Engineer, Stitch Fix

Apache Spark has been an integral part of Stitch Fix's compute infrastructure. Over the past five years, it has become our de facto standard for most ETL and heavy data processing needs, and has expanded our capabilities in the data warehouse. This talk focuses on ETL writing in Stitch Fix and describes these modules that help our data scientists on a daily basis.

Retail and Consumer Goods Tech Talks

THURSDAY, MAY 27

5:00 PM PT

Weekday Demand Sensing at Walmart

DIVYA HINDUPUR, Data Scientist, Walmart Labs | **JAY KAKKAR**, Data Scientist, Walmart Labs

The SMART Forecasting team at Walmart Labs has built an innovative, cloud-agnostic, scalable platform to improve Walmart's ability to predict customer demand while improving in-stocks and reducing food waste. Over a period of two years, all of Walmart's key U.S. departments as well as those in Canada and Mexico have adopted our forecasting solution with planned extensions to other Walmart-operated international markets. Over 500M store-item combinations are forecasted every week for the next 52 weeks. We continue to enhance our modeling suite for COVID-19 impact, pricing in international markets, and weekend sales corrections. We will present a general overview of our scaled forecasting solution, followed by a concrete use case for in-week adjustments, which provides consistent business value for produce and is currently in the process of being scaled out to more Walmart departments.

FRIDAY, MAY 28

10:30 AM PT

Automated Background Removal Using PyTorch

OLEKSANDER MIROSHNYCHENKO, Data Scientist, GlobalLogic | **SIMONA STOLNICU**, Data Scientist, Levi9

Wehkamp is an online department store with more than 500,000 daily visitors. The wide variety of products presented on the Wehkamp website aims to meet the needs of many customers. An important aspect of any customer visit to the website is a qualitative and accurate visual experience of the products. To achieve this, thousands of product photos, especially of fashion garments, are processed in the local photo studio. Since these images' backgrounds are highly varied, background removal is one of the steps in the processing pipeline. In our presentation, we'll describe our automated end-to-end pipeline, which uses machine learning models for removing the background in images, and share how to create an efficient pipeline for deep learning image processing within the Databricks environment.

Interactive Demos

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

WEDNESDAY, MAY 26

5:50 PM – 6:20 PM PT

Fine-Grained Time Series Forecasting at Scale

Learn how retailers and manufacturers are cost-effectively generating millions of item – and location-specific forecasts on a daily basis.

Solution Architects: **KELLY O'MALLEY, BRYAN SMITH**

THURSDAY, MAY 27

3:10 PM – 3:40 PM PT

Segmentation in the Age of Personalization

Explore a structured approach to building and analyzing segments that enable the organization to effectively engage their customers.

Solutions Architect: **BRYAN SMITH**

FRIDAY, MAY 28

10:50 AM – 11:20 AM PT

Personalizing the Customer Experience With Recommendations

Explore how recommenders can be used in a variety of ways to deliver personalized customer experiences.

Solutions Architect: **BRYAN SMITH**

Summit Retail and Consumer Goods Checklist

Retail and Consumer Goods Forum

WEDNESDAY, MAY 26

- 11:00 AM – 1:00 PM PT: Keynote + Retail and Consumer Goods Panel

Retail and Consumer Goods Talks

WEDNESDAY, MAY 26

- 11:30 AM PT: Data Distribution and Ordering for Efficient Data Source V2, [Apple](#)
- 11:30 AM PT: Scaling AI at H&M, [H&M](#)
- 12:05 PM PT: Data Security at Scale Through Spark and Parquet Encryption, [Apple](#)
- 12:05 PM PT: Automating Data Quality Processes at Reckitt, [Cervello](#) | [RB](#)
- 5:00 PM PT: AI-Driven Personalized Email Marketing, [Endeavour Drinks](#)
- 5:00 PM PT: Offer Recommendation System With Apache Spark at Burger King, [Intel](#) | [Burger King](#)

THURSDAY, MAY 27

- 11:00 AM PT: Structured Streaming Use-Cases at Apple, [Apple](#)
- 11:35 AM PT: Building a Product Assortment Recommendation Engine, [AB InBev](#) | [Anheuser Busch](#)
- 11:35 AM PT: ChakraView – A 360° Approach to Data Quality, [Flipkart](#)
- 11:35 AM PT: The Rise of ZStandard: Apache Spark/Parquet/ORC/Avro, [Apple](#)
- 3:50 PM PT: Modularized ETL Writing With Apache Spark, [Stitch Fix](#)
- 5:00 PM PT: Weekday Demand Sensing at Walmart, [Walmart Labs](#)

FRIDAY, MAY 28

- 10:30 AM PT: Automated Background Removal Using PyTorch, [GlobalLogic](#) | [Levi9](#)

Interactive Demos

- MAY 26 5:50 PM: Fine-Grained Time Series Forecasting at Scale
- MAY 27 3:10 PM: Segmentation in the Age of Personalization
- MAY 28 10:50 AM: Personalizing the Customer Experience With Recommendations

Ask an Expert

Ask Me Anything: Cross-Industry live AMA

FRIDAY, MAY 28

9:10 AM - 10:10 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:

- Retail and Consumer Goods leaders and subject matter experts
- Technical experts of data in Retail and Consumer Goods
- Solutions Architects who have helped hundreds of customers

DATA+AI
SUMMIT 2021
FORMERLY SPARK+AI SUMMIT

#DataAISummit