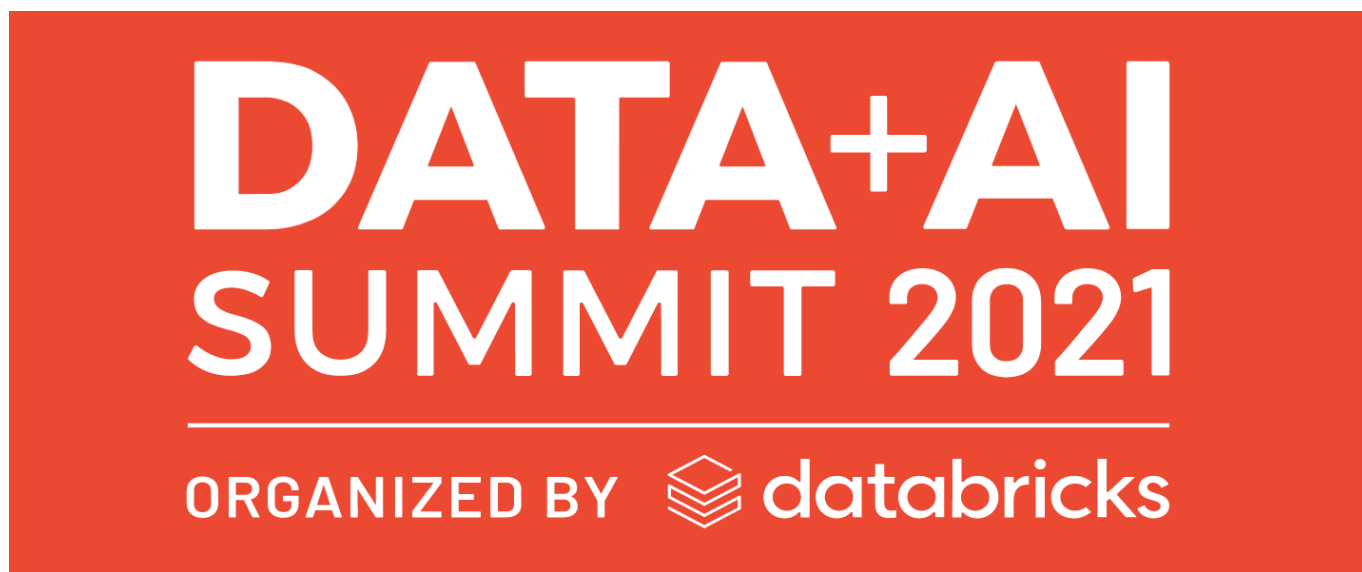


Data + AI Summit 2021 全部超清 PPT 下载

Data + AI Summit 2021 于2021年05月24日至28日举行。本次会议是在线举办的，一共为期五天，第一、二天是培训，第三天到第五天是正式会议。本次会议有超过200个议题，演讲嘉宾包括业界、研究和学术界的专家，会议涵盖来自从业者的技术内容，他们将使用 Apache Spark™、Delta Lake、MLflow、Structured Streaming、BI和SQL分析、深度学习和机器学习框架来解决棘手的数据问题。会议的全部日程

请参见：<https://databricks.com/dataaisummit/north-america-2021/sessions>



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：过往记忆大数据

按照惯例，这次会议的 KeyNote 部分数砖发布了一些新产品，比如 Delta Sharing、Delta Live Tables、Unity Catalog 等等。本次会议有些干货大家可以看下的。在接下来的几天，本公众号也会对一些比较有意思的议题进行介绍，敬请关注本公众号。

本次会议的议题范围具体如下：

- Apache Spark™, Delta Lake, MLflow, PyTorch, TensorFlow, Transformers 等最佳实践和用户案例；
- 数据工程，包括流架构
- 使用数据仓库（data warehouse）和数据湖（data lakes）进行 SQL 分析和 BI；
- 数据科学，包括 Python 生态系统；
- 机器学习和深度学习应用

下载途径

关注微信公众号 过往记忆大数据 或者 Java与大数据架构 并回复 9977 获取。

可下载PPT

下面议题提供 PPT 下载，共186个。

- [10 Things Learned Releasing Databricks Enterprise Wide](#)
- [5 Critical Steps to Clean Your Data Swamp When Migrating Off of Hadoop](#)
- [A Collaborative Data Science Development Workflow](#)
- [A Fast Decision Rule Engine for Anomaly Detection](#)
- [A High Performance Mutable Engagement Activity Delta Lake](#)
- [A Practical Enterprise Feature Store on Delta Lake](#)
- [Accelerate Data Science Initiatives: Databricks & Privacera](#)
- [Accelerate Your ML Pipeline with AutoML and MLflow](#)
- [Accelerating Data Ingestion with Databricks Autoloader](#)
- [Advanced Model Comparison and Automated Deployment Using ML](#)
- [Advanced Natural Language Processing with Apache Spark NLP](#)
- [Advanced SQL For Data Scientists](#)
- [AI Data Acquisition and Governance: Considerations for Success](#)
- [AI Modernization at AT&T and the Application to Fraud with Databricks](#)
- [AI-Driven Personalized Email Marketing](#)
- [Analytics-Enabled Experiences: The New Secret Weapon](#)
- [Anomaly Detection at Scale!](#)
- [Architecting Agile Data Applications for Scale](#)
- [Architect's Open-Source Guide for a Data Mesh Architecture](#)
- [Automated Background Removal Using PyTorch](#)
- [Automated Metadata Management in Data Lake – A CI/CD Driven Approach](#)
- [Automatic ICD-10 Code Assignment to Consultations](#)
- [Automating Data Quality Processes at Reckitt](#)
- [Auto-Train a Time-Series Forecast Model With AML + ADB](#)
- [Best Practices for Enabling Speculative Execution on Large Scale Platforms](#)
- [Bootstrapping of PySpark Models for Factorial A/B Tests](#)
- [BOTS TESTING BOTS: From manual to automated testing for conversational AI](#)
- [Bridging the Completeness of Big Data on Databricks](#)
- [Bring Your Own Container: Using Docker Images In Production](#)
- [Brokering Data: Accelerating Data Evaluation with Databricks White Label](#)
- [Build Large-Scale Data Analytics and AI Pipeline Using RayDP](#)
- [Build Real-Time Applications with Databricks Streaming](#)
- [Building a Data Science as a Service Platform in Azure with Databricks](#)
- [Building A Product Assortment Recommendation Engine](#)
- [Building an ML Platform with Ray and MLflow](#)
- [Building Data Quality pipelines with Apache Spark and Delta Lake](#)
- [Building Data Science into Organizations: Field Experience](#)
- [Building End-to-End Delta Pipelines on GCP](#)
- [Building Lakehouses on Delta Lake with SQL Analytics Primer](#)
- [Building Source of Truth Place Data at Scale](#)
- [Building the Artificially Intelligent Enterprise](#)

- [Building the Foundations of an Intelligent, Event-Driven Data Platform at EFSA](#)
- [Scaling AI At H&M](#)
- [Catch Me If You Can: Keeping Up With ML Models in Production](#)
- [ChakraView – A 360° Approach to Data Quality](#)
- [Change Data Feed in Delta](#)
- [Choose Your Weapon: Comparing Spark on FPGAs vs GPUs](#)
- [CI/CD in MLOps – Implementing a Framework for Self-Service Everything](#)
- [Code Once Use Often with Declarative Data Pipelines](#)
- [Commercializing Alternative Data](#)
- [Comprehensive View on Intervals in Apache Spark 3.2](#)
- [Configuration Driven Reporting On Large Dataset Using Apache Spark](#)
- [Considerations for Data Access in the Lakehouse](#)
- [Consolidating MLOps at One of Europe's Biggest Airports](#)
- [Conversational AI with Transformer Models](#)
- [Creating an 86,000 Hour Speech Dataset with Apache Spark and TPUs](#)
- [Creating Reusable Geospatial Pipelines](#)
- [Credit Card Fraud Detection Using ML In Databricks](#)
- [Customer Experience at Disney+ Through Data Perspective](#)
- [Data Discovery at Databricks with Amundsen](#)
- [Data Distribution and Ordering for Efficient Data Source V2](#)
- [Data Quality With or Without Apache Spark and Its Ecosystem](#)
- [Data Security at Scale through Spark and Parquet Encryption](#)
- [Databricks: A Tool That Empowers You To Do More With Data](#)
- [Deep Dive into the New Features of Apache Spark 3.1](#)
- [Degrading Performance? You Might be Suffering From the Small Files Syndrome](#)
- [Delight: An Improved Apache Spark UI, Free, and Cross-Platform](#)
- [Delivering Insights from 20M+ Smart Homes with 500M+ Devices](#)
- [Delta Lake Streaming: Under the Hood](#)
- [Democratizing Data Quality Through a Centralized Platform](#)
- [Detecting Anomalous Behavior with Surveillance Analytics](#)
- [DevOps for Databricks](#)
- [Drifting Away: Testing ML Models in Production](#)
- [Drug and Vaccine Discovery: Knowledge Graph + Apache Spark](#)
- [Drug Repurposing using Deep Learning on Knowledge Graphs](#)
- [Effective AIOps with Open Source Software in a Week](#)
- [Efficient Distributed Hyperparameter Tuning with Apache Spark](#)
- [Efficient Large-Scale Language Model Training on GPU Clusters](#)
- [Empower Splunk and other SIEMs with the Databricks Lakehouse for Cybersecurity](#)
- [Empowering Real Time Patient Care Through Spark Streaming](#)
- [Empowering Zillow's Developers with Self-Service ETL](#)
- [Entity Resolution Using Patient Records at CMMI](#)
- [Experimentation to Industrialization: Implementing MLOps](#)
- [Extending Machine Learning Algorithms with PySpark](#)
- [FlorenceAI: Reinventing Data Science at Humana](#)
- [From Chatbots to Augmented Conversational Assistants](#)
- [From Vaccine Management to ICU Planning: How CRISP Unlocked the Power of Data](#)

During a Pandemic

- [FrugalML: Using ML APIs More Accurately and Cheaply](#)
- [Fully Utilizing Spark for Data Validation](#)
- [Funnel Analysis with Apache Spark and Druid](#)
- [Gain 3 Benefits with Delta Sharing](#)
- [Gender Prediction with Databricks AutoML Pipeline](#)
- [Getting Started with Databricks SQL Analytics](#)
- [Giving Away The Keys To The Kingdom: Using Terraform To Automate Databricks](#)
- [Graph-Powered Machine Learning](#)
- [Growing the Delta Ecosystem to Rust and Python with Delta-RS](#)
- [How Adobe uses Structured Streaming at Scale](#)
- [How Machine Learning and AI Can Support the Fight Against COVID-19](#)
- [How to Build a ML Platform Efficiently Using Open-Source](#)
- [How to use Apache TVM to optimize your ML models](#)
- [How We Optimize Spark SQL Jobs With parallel and sync IO](#)
- [How We Scaled Bert To Serve 1+ Billion Daily Requests on CPU](#)
- [Hybrid Apache Spark Architecture with YARN and Kubernetes](#)
- [Hyperspace for Delta Lake](#)
- [Image Processing on Delta Lake](#)
- [Importance of ML Reproducibility & Applications with MLflow](#)
- [Improving Apache Spark for Dynamic Allocation and Spot Instances](#)
- [Improving Power Grid Reliability Using IoT Analytics](#)
- [Infrastructure Agnostic Machine Learning Workload Deployment](#)
- [Intro to Delta Lake](#)
- [Introducing Delta Live Tables: Make Reliable ETL Easy on Delta Lake](#)
- [Intuitive & Scalable Hyperparameter Tuning with Apache Spark + Fugue](#)
- [Jeeves Grows Up: An AI Chatbot for Performance and Quality](#)
- [Keeping Identity Graphs In Sync With Apache Spark](#)
- [KFServing, Model Monitoring with Apache Spark and a Feature Store](#)
- [Koalas: How Well Does Koalas Work?](#)
- [Large Scale Geospatial Indexing and Analysis on Apache Spark](#)
- [Large Scale Lakehouse Implementation Using Structured Streaming](#)
- [Learn to Use Databricks for Data Science](#)
- [Learn to Use Databricks for the Full ML Lifecycle](#)
- [Machine Learning CI/CD for Email Attack Detection](#)
- [Machine Learning with PyCaret](#)
- [Magnet Shuffle Service: Push-based Shuffle at LinkedIn](#)
- [Managing Millions of Tests Using Databricks](#)
- [Managing R&D Data on Parallel Compute Infrastructure](#)
- [Massive Data Processing in Adobe Using Delta Lake](#)
- [Migrating ETL Workflow to Apache Spark at Scale in Pinterest](#)
- [Migrating Your Data Platform At a High Growth Startup](#)
- [Misusing MLflow To Help Deduplicate Data At Scale](#)
- [MLCommons: Better ML for Everyone](#)
- [MLflow Model Serving](#)
- [Model Monitoring at Scale with Apache Spark and Verta](#)

- [Modelling Customer Lifetime Revenue for Subscription Business](#)
- [Modernizing to a Cloud Data Architecture](#)
- [Modularized ETL Writing with Apache Spark](#)
- [Monitor Apache Spark 3 on Kubernetes using Metrics and Plugins](#)
- [Natural Language Query and Conversational Interface to Apache Spark](#)
- [NLP-Focused Applied ML at Scale for Global Fleet Analytics at ExxonMobil](#)
- [Northwestern Mutual Journey – Transform BI Space to Cloud](#)
- [Object Detection with Transformers](#)
- [Observability for Data Pipelines With OpenLineage](#)
- [Offer Recommendation System with Apache Spark at Burger King](#)
- [Optimizing the Catalyst Optimizer for Complex Plans](#)
- [PandasUDFs: One Weird Trick to Scaled Ensembles](#)
- [Phar Data Platform: From the Lakehouse Paradigm to the Reality](#)
- [Play Head Time Analysis On OTT Video At Scale](#)
- [Portable UDFs: Write Once, Run Anywhere](#)
- [Predicting Optimal Parallelism for Data Analytics](#)
- [Processing Large Datasets for ADAS Applications using Apache Spark](#)
- [Productionalizing Machine Learning Solutions with Effective Tracking, Monitoring, and Management](#)
- [Productionizing Machine Learning in Our Health and Wellness Marketplace](#)
- [Productionizing ML Model Using MLflow Model Serving](#)
- [Radical Speed for SQL Queries on Databricks: Photon Under the Hood](#)
- [Raven: End-to-end Optimization of ML Prediction Queries](#)
- [Real-world Strategies for Debugging Machine Learning Systems](#)
- [Redis + Apache Spark = Swiss Army Knife Meets Kitchen Sink](#)
- [Re-imagine Data Monitoring with whylogs and Spark](#)
- [Role of Data Accessibility During Pandemic](#)
- [RWE & Patient Analytics Leveraging Databricks – A Use Case](#)
- [Sawtooth Windows for Feature Aggregations](#)
- [Scaling and Modernizing Data Platform with Databricks](#)
- [Scaling and Unifying SciKit Learn and Apache Spark Pipelines](#)
- [Scaling AutoML-Driven Anomaly Detection With Luminaire](#)
- [Scaling Online ML Predictions At DoorDash](#)
- [Scaling Privacy in a Spark Ecosystem](#)
- [Scaling your Data Pipelines with Apache Spark on Kubernetes](#)
- [Semantic Image Logging Using Approximate Statistics & MLflow](#)
- [Simplify Data Conversion from Spark to TensorFlow and PyTorch](#)
- [Speed up UDFs with GPUs using the RAPIDS Accelerator](#)
- [SQL Analytics Powering Telemetry Analysis at Comcast](#)
- [Stage Level Scheduling Improving Big Data and AI Integration](#)
- [Structured Streaming Use-Cases at Apple](#)
- [Superworkflow of Graph Neural Networks with K8S and Fugue](#)
- [Tensors Are All You Need: Faster Inference with Hummingbird](#)
- [The Critical Missing Component in the Production ML Stack](#)
- [The Function, the Context, and the Data—Enabling ML Ops at Stitch Fix](#)
- [The Rise of Vector Data](#)

- [The Rise of ZStandard: Apache Spark/Parquet/ORC/Avro](#)
- [Towards Personalization in Global Digital Health](#)
- [Unified MLOps: Feature Stores & Model Deployment](#)
- [Video Analytics At Scale: DL, CV, ML On Databricks Platform](#)
- [Weekday Demand Sensing at Walmart](#)
- [What's New with Databricks Machine Learning](#)
- [Why APM Is Not the Same As ML Monitoring](#)
- [Wizard Driven AI Anomaly Detection with Databricks in Azure](#)
- [You Can Do It in SQL](#)

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: [【】（）](#)