AGENDA (subject to change)

Monday, June 22

- [ ] 6:00 AM to 9:00 AM **Apache Spark for Machine Learning and Data Science**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Apache Spark Tuning and Best Practices**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Building Better Data Pipelines for Apache Spark with Delta Lake**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Databricks Administration**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Introduction to Apache Spark Programming**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Introduction to Reinforcement Learning**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Unified Data Analytics for Managers**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **MLflow: Managing the Machine Learning Lifecycle**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Scaling Deep Learning with TensorFlow and Apache Spark**
  Track:Training Duration:180m
- [ ] 6:00 AM to 9:00 AM **Structured Streaming with Databricks**
  Track:Training Duration:180m
- [ ] 7:00 AM to 8:30 AM **What's New in Apache Spark 3.0**
  Track:Follow Along Training Duration:90m
- [ ] 10:00 AM to 1:00 PM **Apache Spark for Machine Learning and Data Science**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Apache Spark Tuning and Best Practices**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Building Better Data Pipelines for Apache Spark with Delta Lake**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Introduction to Apache Spark Programming**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Unified Data Analytics for Managers**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Machine Learning Deployment: 3 Model Deployment Paradigms, Monitoring and Alerting**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Model-Free Reinforcement Learning**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Scaling Deep Learning with TensorFlow and Apache Spark**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **SQL on Databricks**
  Track:Training Duration:180m
- [ ] 10:00 AM to 1:00 PM **Structured Streaming with Databricks**
  Track:Training Duration:180m
- [ ] 2:00 PM to 3:30 PM **Spark Meetup - Brew Talk and AI for Social Good**
  Speaker(s): Denny Lee, Databricks | Jules Damji, Databricks | Laurence Moroney, Google
  Track:Special Event Duration:90m
- [ ] 2:00 PM to 3:30 PM **What's New in Apache Spark 3.0**
  Speaker(s): Doug Bateman, Databricks
  Track:Follow Along Training Duration:90m

Tuesday, June 23

- [ ] 9:00 AM to 12:00 PM **Apache Spark for Machine Learning and Data Science**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Apache Spark Tuning and Best Practices**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Building Better Data Pipelines for Apache Spark with Delta Lake**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Certification Prep: Databricks Certified Associate Developer for Apache Spark 2.4**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Distributed Machine Learning in SparkR/sparklyr**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Introduction to Apache Spark Programming**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Introduction to Reinforcement Learning**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Unified Data Analytics for Managers**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **MLflow: Managing the Machine Learning Lifecycle**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Natural Language Processing at Scale**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Practical Problem-solving in Finance: Real-time Fraud Detection with Apache Spark**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Practical Problem-solving in Healthcare: Real-time Data Analytics with Apache Spark**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Practical Problem-solving in Manufacturing: Real-time Data Analytics with Apache Spark**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Scaling Deep Learning with TensorFlow and Apache Spark**
  Track:Training Duration:180m
- [ ] 9:00 AM to 12:00 PM **Structured Streaming with Databricks**
  Track:Training Duration:180m
- [ ] 10:00 AM to 11:30 AM **What's New in Apache Spark 3.0**
  Speaker(s): Doug Bateman, Databricks
  Track:Follow Along Training Duration:90m
- [ ] 1:00 PM to 4:00 PM **Apache Spark Tuning and Best Practices**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Building Better Data Pipelines for Apache Spark with Delta Lake**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Introduction to Delta Lake**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Unified Data Analytics for Managers**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Machine Learning Deployment: 3 Model Deployment Paradigms, Monitoring and Alerting**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Model-Free Reinforcement Learning**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Natural Language Processing at Scale**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Practical Problem-solving in Finance: Real-time Fraud Detection with Apache Spark**
  Track:Training Duration:180m
- [ ] 1:00 PM to 4:00 PM **Practical Problem-solving in Healthcare: Real-time Data Analytics with Apache Spark**
  Track:Training Duration:180m

☐ 1:00 PM to 4:00 PM **Practical Problem-solving in Retail: Real-time Data Analytics with Apache Spark**
Track:Training Duration:180m

☐ 1:00 PM to 4:00 PM **Scaling Deep Learning with TensorFlow and Apache Spark**
Track:Training Duration:180m

☐ 1:00 PM to 4:00 PM **Structured Streaming with Databricks**
Track:Training Duration:180m

☐ 2:00 PM to 3:30 PM **What's New in Apache Spark 3.0**
Track:Follow Along Training Duration:90m

Wednesday, June 24

☐ 6:30 AM to 7:00 AM **The Importance of Model Fairness and Interpretability in AI Systems**
Speaker(s): Francesca Lazzeri, Microsoft
Track:Data Science, Deep Learning and ML Duration:30m

☐ 6:30 AM to 7:00 AM **Responsible AI: Protecting Privacy and Preserving Confidentiality in Machine Learning and Data Analytics**
Speaker(s): Sarah Bird, Microsoft
Track:AI Use Cases Duration:30m

☐ 8:30 AM to 10:30 AM **Wednesday Morning Keynotes**
Speaker(s): Ali Ghodsi, Databricks | Matei Zaharia, Databricks | Brooke Wenig, Databricks | Reynold Xin, Databricks | Vishwanath Subramanian, Starbucks
Track:Keynote Duration:120m

☐ 10:30 AM to 11:00 AM **Unlock Mainframe and IBM i for the Data Lakehouse**
Speaker(s): Ashwin Ramachandran, Precisely
Track:Data Engineering and Architecture Duration:30m

☐ 10:30 AM to 11:00 AM **Modernize Your Data Warehouse and Data Lake to Databricks Delta with Informatica**
Speaker(s): Rodrigo Sanchez Bredee, Informatica
Track:Data Engineering and Architecture Duration:30m

☐ 10:30 AM to 11:00 AM **Your Tools Suck!! Re-Imagining Apache Spark Development**
Speaker(s): Raj Bains, Prophecy.io
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Building a Feature Store around Dataframes and Apache Spark**
Speaker(s): Jim Dowling, Logical Clocks AB | Fabio Buso, Logical Clocks AB
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Composable Data Processing with Apache Spark**
Speaker(s): Dilip Biswal, Adobe | Shone Sadler, Adobe, Inc.
Track:Developer Duration:30m

☐ 11:00 AM to 12:05 PM **Deep Dive into the New Features of Apache Spark 3.0**
Speaker(s): Xiao Li, Databricks | Wenchen Fan, Databricks
Track:Developer Duration:65m

☐ 11:00 AM to 11:30 AM **End-to-End Deep Learning with Horovod on Apache Spark**
Speaker(s): Thomas Graves, NVIDIA | Travis Addair, Uber, Inc.
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Disrupting Risk Management through Emerging Technologies**
Speaker(s): Badrish Davay, Capital One | Yuri Bogdanov, Capital One
Track:Data & ML Industry Use Cases Duration:30m

☐ 11:00 AM to 12:05 PM **Performant Streaming in Production: Preventing Common Pitfalls when Productionizing Streaming Jobs**
Speaker(s): Stefan van Wouw, Databricks | Max Thone, Databricks
Track:Data Engineering and Architecture Duration:65m

☐ 11:00 AM to 11:30 AM **Portable Scalable Data Visualization Techniques for Apache Spark and Python Notebook-based Analytics**
Speaker(s): Douglas Moore, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Power of Visualizing Embeddings**
Speaker(s): Pramod Singh, Bain & Company
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Productionizing Machine Learning with a Microservices Architecture**
Speaker(s): Yaron Haviv, Iguazio
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Running Apache Spark on Kubernetes: Best Practices and Pitfalls**
Speaker(s): Jean-Yves Stephan, Data Mechanics | Julien Dumazert, Data Mechanics
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Using AI to Support Proliferating Merchant Changes**
Speaker(s): Anurag Tangri, Visa | Jianhua Huang, Visa
Track:AI Use Cases Duration:30m

☐ 11:00 AM to 12:05 PM **Delta from a Data Engineer's Perspective**
Speaker(s): Palla Lentz, Databricks | Jake Therianos, Databricks
Track:Developer Duration:65m

☐ 11:00 AM to 11:30 AM **Accelerating MLFlow Hyper-parameter Optimization Pipelines with RAPIDS**
Speaker(s): John Zedlewski, NVIDIA
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Make the Most of Your Talent and Time When Working on AI and ML Projects - Automate the Rest**
Speaker(s): Ramesh Menon, Infoworks | Kevin Holder, Infoworks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Managing Business Risks of Large Scale Cloud Migrations**
Speaker(s): Paul-Scott Murphy, WANdisco | Madasamy Thandayuthapani , WANdisco
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Find and Protect Your Crown Jewels in Databricks with Privacera and Apache Ranger**
Speaker(s): Dr. Srikanth Venkat, Privacera
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Building A Data Platform For Mission Critical Analytics**
Speaker(s): Igor Alekseev, AWS | Sally Hoppe, AWS | Denis Dubeau, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 11:30 AM to 1:00 PM **Government Industry Forum**
Speaker(s): Chase Baker, FBI | Eileen M. Vidrine, U.S. Air Force | Rob Brown , USCIS
Track:Vertical Event Duration:90m

☐ 11:30 AM to 1:00 PM **Education Industry Forum**
Speaker(s): Colby Ford, Univ North Carolina Charlotte | Charlie Lindville, University of Illinois at Urbana-Champaign | Dong Wha, University of Maryland, Baltimore County (UMBC)
Track:Vertical Event Duration:90m

☐ 11:30 AM to 12:00 PM **Case Study and Automation Strategies to Protect Sensitive Data**
Speaker(s): Steve Touw, Immuta | Greg Galloway, Artis Consulting
Track:Data Engineering and Architecture Duration:30m

☐ 11:30 AM to 12:00 PM **Building a Better Delta Lake with Talend and Databricks**
Speaker(s): Michael Destein, Talend | Cameron Davie, Ää, Talend
Track:Data Engineering and Architecture Duration:30m

☐ 11:35 AM to 12:05 PM **Apache Spark NLP for Healthcare: Lessons Learned Building Real-World Healthcare AI Systems**
Speaker(s): Veysel Kocaman, John Snow Labs
Track:AI Use Cases Duration:30m

☐ 11:35 AM to 12:40 PM **AutoML Toolkit - Deep Dive**
Speaker(s): Daniel Tomes, Databricks | Ben Wilson, Databricks
Track:Data Science, Deep Learning and ML Duration:65m

☐ 11:35 AM to 12:05 PM **Data Driven Decisions at Scale**
Speaker(s): Jim Forsythe, Comcast
Track:Data Engineering and Architecture Duration:30m

☐ 11:35 AM to 12:05 PM **Deliver Dynamic Customer Journey Orchestration at Scale**
Speaker(s): Krish Kuruppath, Publicis Media-COSMOS | Sharad Varshney, Publicis Media-COSMOS
Track:Data & ML Industry Use Cases Duration:30m

☐ 11:35 AM to 12:05 PM **Enabling Push Button Productization of AI Models**
Speaker(s): Moty Fania, Intel Corporation
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Productionalizing Models through CI/CD Design with MLflow**
Speaker(s): Mary Grace Moesta, Databricks | Peter Tamisin, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Scaling up Deep Learning by Scaling Down**
Speaker(s): Nick Pentreath, IBM
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Tuning ML Models: Scaling, Workflows, and Architecture**
Speaker(s): Joseph Bradley, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Data Science Across Data Sources with Apache Arrow**
Speaker(s): Tomer Shiran, Dremio | Jacques Nadeau, Dremio
Track:Data Engineering and Architecture Duration:30m

☐ 11:35 AM to 12:05 PM **Redash: Open Source SQL Analytics on Data Lakes**
Speaker(s): Jesse Whitehouse, Redash | Francois Callewaert, Databricks
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **Data Microservices in Apache Spark using Apache Arrow Flight**
Speaker(s): Ryan Murray, Dremio
Track:Data Engineering and Architecture Duration:30m

☐ 12:10 PM to 12:40 PM **Benchmark Tests and How-Tos of Convolutional Neural Network on HorovodRunner Enabled Apache Spark Clusters**
Speaker(s): Jing Pan, eHealth Inc. | Wendao Liu, eHealth Inc.
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **High-Performance Analytics with Probabilistic Data Structures: the Power of HyperLogLog**
Speaker(s): Simeon Simeonov, Swoop
Track:Data Engineering and Architecture Duration:30m

☐ 12:10 PM to 12:40 PM **Ibis: Seamless Transition Between Pandas and Apache Spark**
Speaker(s): Li Jin, Two Sigma Investments
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Operationalize Apache Spark Analytics**
Speaker(s): Ivan Nardini, SAS Institure srl | Artem Glazkov, SAS Institute LLC
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Powering Interactive BI Analytics with Presto and Delta Lake**
Speaker(s): Kamil Bajda-Pawlikowski, Starburst
Track:Data Engineering and Architecture Duration:30m

☐ 12:10 PM to 12:40 PM **Securing Apache Spark Applications at Facebook**
Speaker(s): Abdulrahman Alfozan, Facebook | Ankur Pathela, Facebook
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **Tracing the Breadcrumbs: Apache Spark Workload Diagnostics**
Speaker(s): Cheng Lian, Databricks | Kris Mok, Databricks
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **The Killer Feature Store: Orchestrating Spark ML Pipelines and MLflow for Production**
Speaker(s): Nathan Buesgens, Accenture
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **How (Not) To Scale Deep Learning in 6 Easy Steps**
Speaker(s): Sean Owen, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Automating Federal Aviation Administration's (FAA) System Wide Information Management (SWIM) Data Ingestion and Analysis**
Speaker(s): Marcelo Zambrana, Microsoft | Sheila Stewart, Databricks | Mehdi Hashemipour, USDOT
Track:Data & ML Industry Use Cases Duration:30m

☐ 12:10 PM to 12:40 PM **Geosp.AI.tial: Applying Big Data and Machine Learning to Solve the World's Toughest Geospatial Intelligence Problems**
Speaker(s): Donald Polaski, Booz Allen Hamilton
Track:AI Use Cases Duration:30m

☐ 1:00 PM to 1:45 PM **Wednesday Afternoon Keynotes**
Speaker(s): Dr. Phillip Atiba Goff, Center for Policing Equity | Prof. Jennifer Chayes, UC Berkeley's Division of Computing, Data Science, and Society (CDSS) | Nate Silver, FiveThirtyEight.com
Track:Keynote Duration:45m

☐ 2:00 PM to 3:30 PM **Media and Entertainment Industry Forum**
Speaker(s): Dan Morris, Viacom | Stephen Layland, Tubi | Eric Wasserman, Jam City
Track:Vertical Event Duration:90m

☐ 2:30 PM to 3:00 PM **Hyperspace: An Indexing Subsystem for Apache Spark**
Speaker(s): Rahul Potharaju, Microsoft | Terry Kim, Microsoft
Track:Developer Duration:30m

☐ 2:30 PM to 3:00 PM **A Production Quality Sketching Library for the Analysis of Big Data**
Speaker(s): Lee Rhodes, Verizon Media
Track:Data Science, Deep Learning and ML Duration:30m

☐ 2:30 PM to 3:00 PM **An Approach to Data Quality for Netflix Personalization Systems**
Speaker(s): Preetam Joshi, Netflix | Vivek Kaushal, Netflix
Track:Data Engineering and Architecture Duration:30m

☐ 2:30 PM to 3:00 PM **Children Safety Retrieval (CENSER) System for Retrieval of Kidnapped Children from Brothels in India**
Speaker(s): Amarjot Singh, Skylark Labs LLC.
Track:AI Use Cases Duration:30m

☐ 2:30 PM to 3:35 PM **Deep Learning Enabled Price Action with Databricks and AWS**
Speaker(s): Kris Skrinak, AWS | Igor Alekseev, AWS
Track:Data Science, Deep Learning and ML Duration:65m

☐ 2:30 PM to 3:00 PM **Getting Started Contributing to Apache Spark – From PR, CR, JIRA, and Beyond**
Speaker(s): Holden Karau, Apple
Track:Developer Duration:30m

☐ 2:30 PM to 3:00 PM **How Azure and Databricks Enabled a Personalized Experience for Customers and Patients at CVS Health**
Speaker(s): Michelle Un, CVS Health | Raghu Nakka, CVS Health
Track:Data Science, Deep Learning and ML Duration:30m

☐ 2:30 PM to 3:00 PM **Koalas: Making an Easy Transition from Pandas to Apache Spark**
Speaker(s): Takuya Ueshin, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 2:30 PM to 3:35 PM **Productionizing Machine Learning Pipelines with Databricks and Azure ML**
Speaker(s): Trace Smith, ExxonMobil | Amirhessam Tahmassebi, Data Scientist
Track:Data Science, Deep Learning and ML Duration:65m

☐ 2:30 PM to 3:00 PM **Top Down Specialization Using Apache Spark**
Speaker(s): Macarious Abadeer, IQVIA
Track:Data Engineering and Architecture Duration:30m

☐ 2:30 PM to 3:00 PM **Zipline – A Declarative Feature Engineering Framework**
Speaker(s): Nikhil Simha, Airbnb
Track:Data Science, Deep Learning and ML Duration:30m

☐ 2:30 PM to 3:00 PM **Building Data Quality Audit Framework using Delta Lake at Cerner**
Speaker(s): Madhav Agni, Cerner
Track:Data & ML Industry Use Cases Duration:30m

☐ 3:05 PM to 3:35 PM **Accelerating Data Processing in Spark SQL with Pandas UDFs**
Speaker(s): Michael Tong, Quantcast
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **AI-Assisted Feature Selection for Big Data Modeling**
Speaker(s): Alvin Henrick, Clarify Health Solutions | Imran Qureshi, Clarify Health Solutions | Iman Haji, Clarify Health Solutions
Track:AI Use Cases Duration:30m

☐ 3:05 PM to 3:35 PM **Anomaly Detection at the Edge**
Speaker(s): Arun Kejariwal, Independent | Ira Cohen, Anodot
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:05 PM to 3:35 PM **Columbia Migrates from Legacy Data Warehouse to an Open Data Platform with Delta Lake**
Speaker(s): Bilal Obeidat, Databricks | Lara Minor, Columbia Sportswear
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **Operationalizing Big Data Pipelines At Scale**
Speaker(s): Brad May, Starbucks | Arjit Dhavale, Starbucks | Denny Lee, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **How Adobe Does 2 Million Records Per Second Using Apache Spark!**
Speaker(s): Yeshwanth Vijayakumar, Adobe, Inc.
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **Native Support of Prometheus Monitoring in Apache Spark 3.0**
Speaker(s): Dongjoon Hyun, Apple | DB Tsai, Apple
Track:Developer Duration:30m

☐ 3:05 PM to 3:35 PM **Preventing Abuse Using Unsupervised Learning**
Speaker(s): James Verbus, LinkedIn | Grace Tang, LinkedIn
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:05 PM to 3:35 PM **Simplify CDC Pipeline with Spark Streaming SQL and Delta Lake**
Speaker(s): Jun Song , Alibaba
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **Cloud and Analytics - From Platforms to an Ecosystem**
Speaker(s): Ming Yuan, Zurich North America | David Carlson, Zurich North America
Track:Data & ML Industry Use Cases Duration:30m

☐ 3:40 PM to 4:10 PM **A Thorough Comparison of Delta Lake, Iceberg and Hudi**
Speaker(s): Junjie Chen, Tencent | Junping Du , Tencent
Track:Data Engineering and Architecture Duration:30m

☐ 3:40 PM to 4:10 PM **Accelerating the ML Lifecycle with an Enterprise-Grade Feature Store**
Speaker(s): Mike Del Balso, Tecton.ai | Geoff Sims, Atlassian
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Accelerating Spark SQL Workloads to 50X Performance with Apache Arrow-Based FPGA Accelerators**
Speaker(s): Calvin Hung, WASAI Technology, Inc. | Weiting Chen, Intel Corporation
Track:Developer Duration:30m

☐ 3:40 PM to 4:10 PM **Advertising Fraud Detection at Scale at T-Mobile**
Speaker(s): Eric Yatskowitz, T-Mobile | Phan Chuong, T-Mobile
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Clinical Suspecting at Scale Using PySpark**
Speaker(s): Manas Ranjan Kar, Episource LLC
Track:Data Engineering and Architecture Duration:30m

☐ 3:40 PM to 4:10 PM **Building a Federated Data Directory Platform for Public Health**
Speaker(s): Mark Paul, Healthdirect Australia | Anshul Bajpai, Healthdirect Australia
Track:Data Engineering and Architecture Duration:30m

☐ 3:40 PM to 4:10 PM **Pandas UDF and Python Type Hint in Apache Spark 3.0**
Speaker(s): Hyukjin Kwon, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 3:40 PM to 4:10 PM **Scalable AutoML for Time Series Forecasting using Ray**
Speaker(s): Shengsheng Huang, Intel Corporation | Jason Dai, Intel Corporation
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Scaling Security Threat Detection with Apache Spark and Databricks**
Speaker(s): Josh Gillner, Apple
Track:Developer Duration:30m

☐ 3:40 PM to 4:10 PM **Tackling Scaling Challenges of Apache Spark at LinkedIn**
Speaker(s): Min Shen, LinkedIn | Zoe Lin, LinkedIn
Track:Data Engineering and Architecture Duration:30m

☐ 4:00 PM to 5:30 PM **Attendee Party**
Track:Special Event Duration:90m

Thursday, June 25

☐ 6:00 AM to 7:30 AM **Financial Services Industry Forum**
Speaker(s): Anurag Sehgal, Credit Suisse | Douglas Hamilton , NASDAQ | Junta Nakai , Databricks
Track:Vertical Event Duration:90m

☐ 9:00 AM to 10:30 AM **Thursday Morning Keynotes**
Speaker(s): Ali Ghodsi, Databricks | Clemens Mewald, Databricks | Lauren Richie, Databricks | Matei Zaharia, Databricks | Sue Ann Hong, Databricks | Rohan Kumar, Microsoft | Sarah Bird, Microsoft | Anurag Sehgal, Credit Suisse
Track:Keynote Duration:90m

☐ 11:00 AM to 12:05 PM **Introducing MLflow for End-to-End Machine Learning on Databricks**
Speaker(s): Sean Owen, Databricks
Track:Data Science, Deep Learning and ML Duration:65m

☐ 11:00 AM to 12:05 PM **Running Apache Spark Jobs Using Kubernetes**
Speaker(s): Yaron Haviv, Iguazio
Track:Data Engineering and Architecture Duration:65m

☐ 11:00 AM to 11:30 AM **Artificial Lawyers. Will Your Next Attorney be a Machine?**
Speaker(s): Fernando Ortega Gallego, Plain Concepts | Eduardo Matallanas de Avila, Plain Concepts
Track:AI Use Cases Duration:30m

☐ 11:00 AM to 11:30 AM **Consolidate Your Technical Debt With Spark Data Sources -Tools and Techniques to Integrate Native Code**
Speaker(s): Doug Carson, Illuminate Technologies
Track:Developer Duration:30m

☐ 11:00 AM to 11:30 AM **Continuous Delivery of ML-Enabled Pipelines on Databricks using MLflow**
Speaker(s): Michael Shtelma, Databricks | Thunder Shiviah, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Data Mesh in Practice: How Europe's Leading Online Platform for Fashion Goes Beyond the Data Lake**
Speaker(s): Max Schultze, Zalando SE | Arif Wider, Thoughtworks
Track:Data & ML Industry Use Cases Duration:30m

☐ 11:00 AM to 11:30 AM **Democratizing Data**
Speaker(s): Cindy Mottershead, Blackbaud | Shiran Algai, Blackbaud
Track:Developer Duration:30m

☐ 11:00 AM to 11:30 AM **Encryption and Masking for Sensitive Apache Spark Analytics Addressing CCPA and Governance**
Speaker(s): Alon Rosenthal, SecuPi Security Solutions Ltd. | Les McMonagle, SecuPi
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Everyday Probabilistic Data Structures for Humans**
Speaker(s): Yeshwanth Vijayakumar, Adobe, Inc.
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Evolution is Continuous, and so are Big Data and Streaming Pipelines**
Speaker(s): Gustav Ranby, Scania CV | Sarah Hantosi Albertsson, Scania CV
Track:Data Engineering and Architecture Duration:30m

☐ 11:00 AM to 11:30 AM **Filtering vs Enriching Data in Apache Spark**
Speaker(s): Gokul Prabagaren, Capital One
Track:Developer Duration:30m

☐ 11:00 AM to 11:30 AM **Text Extraction from Product Images Using State-of-the-Art Deep Learning Techniques**
Speaker(s): Rajesh Shreedhar Bhat, Walmart Labs
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Enabling Physics and Empirical-Based Algorithms with Spark Using the Integration of MATLAB in Databricks**
Speaker(s): Meaghan Kosmatka, Deere & Company | Arvind Hosagrahara, MathWorks, Inc.
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:00 AM to 11:30 AM **Scaling Value with Azure Databricks**
Speaker(s): Susana Lira-Gonzalez, Avanade | Luke Pritchard, Avanade
Track:Data & ML Industry Use Cases Duration:30m

☐ 11:30 AM to 1:00 PM **Retail and Consumer Packaged Goods Industry Forum**
Speaker(s): Ojas Nivsarkar, 7-Eleven | Brad Kent , LoyaltyOne | Saritha Ivaturi, Director of Data Systems
Track:Vertical Event Duration:90m

☐ 11:30 AM to 1:00 PM **Healthcare and Life Sciences Industry Forum**
Speaker(s): Dr. Binu Mathew , Mercy | Sanji Fernando, Optum | Joanne Hackett, PhD, IZY Capital / Genomics England
Track:Vertical Event Duration:90m

☐ 11:30 AM to 12:30 PM **Women in Unified Data Analytics Panel Discussion**
Speaker(s): Rama Assaf-Smith, Comcast | Ali Vanderveld, ShopRunner | Franziska Bell, Ph.D., Toyota Research Institute
Track:Special Event Duration:60m

☐ 11:30 AM to 12:00 PM **Scaling Quantitative Research on Sensitive Data**
Speaker(s): Slava Frid, Worldquant Predictive
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **AI Disruption of Quantitative Finance. From Forecasting, to Probability Density Estimation, to Generative Models, and to Optimization with Reinforcement Learning**
Speaker(s): Nima Nooshi, Databricks
Track:AI Use Cases Duration:30m

☐ 11:35 AM to 12:05 PM **Care and Feeding of Catalyst Optimizer**
Speaker(s): Rose Toomey, Bloomberg
Track:Developer Duration:30m

☐ 11:35 AM to 12:05 PM **Faster Data Integration Pipeline Execution using Spark-Jobserver**
Speaker(s): Sailee Jain, Informatica | Prabhakar Gouda, Informatica
Track:Developer Duration:30m

☐ 11:35 AM to 12:05 PM **Managing ADLS gen2 using Apache Spark**
Speaker(s): Jacek Tokar, Procter & Gamble
Track:Data Engineering and Architecture Duration:30m

☐ 11:35 AM to 12:05 PM **Parallelization of Structured Streaming Jobs Using Delta Lake**
Speaker(s): Oliver Lewis , Tubi Inc.
Track:Data Engineering and Architecture Duration:30m

☐ 11:35 AM to 12:05 PM **Saving Energy in Homes with a Unified Approach to Data and AI**
Speaker(s): Stephen Galsworthy, Quby | Erni Durdevic, Quby
Track:Impact and Innovation Duration:30m

☐ 11:35 AM to 12:05 PM **Using Apache Spark for Predicting Degrading and Failing Parts in Aviation**
Speaker(s): Christopher Miller, U.S. Navy | Matthew Proetsch, U.S. Navy
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Using Machine Learning Algorithms to Construct All the Components of a Knowledge Graph**
Speaker(s): Maureen Teyssier, Reonomy
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Wood Log Inventory Estimation using Image Processing and Deep Learning Technique**
Speaker(s): Tomas Vantuch, Stora Enso
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:35 AM to 12:05 PM **Zeus: Uber's Highly Scalable and Distributed Shuffle as a Service**
Speaker(s): Mayank Bansal, Uber, Inc. | Bo Yang, Uber, Inc.
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **Productionizing Deep Reinforcement Learning with Spark and MLflow**
Speaker(s): Patrick Halina, Zynga | Curren Pangler, Zynga
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Best Practices for Engineering Production-Ready Software with Apache Spark**
Speaker(s): Maximo Gurmendez, Montevideo Labs | Javier Buquet, Montevideo Labs
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **Building the Petcare Data Platform using Delta Lake and 'Kyte': Our Spark ETL Pipeline**
Speaker(s): George Claireaux , Mars | Kirby Prowting, Mars
Track:Data Engineering and Architecture Duration:30m

☐ 12:10 PM to 12:40 PM **Fine Tuning and Enhancing Performance of Apache Spark Jobs**
Speaker(s): Kira Lindke, IBM | Blake Becerra, IBM | Kaushik Tadikonda, IBM
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **How to Performance-Tune Apache Spark Applications in Large Clusters**
Speaker(s): Omkar Joshi, Uber, Inc.
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **On Improving Broadcast Joins in Apache Spark SQL**
Speaker(s): Jianneng Li, Workday
Track:Developer Duration:30m

☐ 12:10 PM to 12:40 PM **Scaling Production Machine Learning Pipelines with Databricks**
Speaker(s): Max Cantor, Conde Nast | James Evers, Conde Nast
Track:Data & ML Industry Use Cases Duration:30m

☐ 12:10 PM to 12:40 PM **The 2020 Census and Innovation in Surveys**
Speaker(s): Zack Schwartz, U.S. Census Bureau
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Using Apache Spark and Differential Privacy for Protecting the Privacy of the 2020 Census Respondents**
Speaker(s): Simson Garfinkel, U.S. Census Bureau
Track:Data Engineering and Architecture Duration:30m

☐ 12:10 PM to 12:40 PM **Using Bayesian Generative Models with Apache Spark to Solve Entity Resolution Problems (DeDup, Merging, Uniqueness) at Scale**
Speaker(s): Charles Adetiloye, MavenCode | Timo Mechler, MavenCode
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Using Machine Learning to Evolve Sports Entertainment**
Speaker(s): David Cunningham, DataFactory | Young Bang, Atlas Research
Track:Impact and Innovation Duration:30m

☐ 12:10 PM to 12:40 PM **Translating Models to Medicine an Example of Managing Visual Communications**
Speaker(s): Andrew Bauman, Seattle Childrens | James Hibbard, Seattle Childrens
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:10 PM to 12:40 PM **Battling Model Decay with Deep Learning and Gamification**
Speaker(s): Sinan Ozdemir, Directly
Track:Impact and Innovation Duration:30m

☐ 1:00 PM to 2:00 PM **Thursday Afternoon Keynotes**
Speaker(s): Kim Hazelwood, Facebook | Hany Farid, UC Berkeley
Track:Keynote Duration:60m

☐ 2:30 PM to 3:00 PM **Automated and Explainable Deep Learning for Clinical Language Understanding at Roche**
Speaker(s): David Talby, Pacific AI | Vishakha Sharma, Roche | Yogesh Pandit, Roche
Track:Data Science, Deep Learning and ML Duration:30m

☐ 2:30 PM to 3:00 PM **Building a Real-Time Feature Store at iFood**
Speaker(s): Daniel Galinkin, iFood
Track:Developer Duration:30m

☐ 2:30 PM to 3:35 PM **Deep Dive into GPU Support in Apache Spark 3.x**
Speaker(s): Robert Evans, NVIDIA | Jason Lowe, NVIDIA
Track:Developer Duration:65m

☐ 2:30 PM to 3:00 PM **Generative Hyperloop Design: Managing Massively Scaled Simulations Focused on Quick-Insight Analytics and Demand Modelling**
Speaker(s): Patryk Oleniuk, Virgin Hyperloop One | Sandhya Raghavan, Virgin Hyperloop One
Track:Impact and Innovation Duration:30m

☐ 2:30 PM to 3:00 PM **Using Databricks as an Analysis Platform**
Speaker(s): Anup Segu , YipitData
Track:Data Engineering and Architecture Duration:30m

☐ 2:30 PM to 3:00 PM **Lessons Learned from Modernizing USCIS Data Analytics Platform**
Speaker(s): Shawn Benjamin, U.S. Citizenship and Immigration Services (USCIS) Office of Information Technology | Prabha Rajendran, U.S. Citizenship and Immigration Services (USCIS) Office of Information Technology
Track:Data & ML Industry Use Cases Duration:30m

☐ 2:30 PM to 3:00 PM **Scalable Acceleration of XGBoost Training on Apache Spark GPU Clusters**
Speaker(s): Rong Ou, NVIDIA | Bobby Wang, NVIDIA
Track:Data Science, Deep Learning and ML Duration:30m

☐ 2:30 PM to 3:00 PM **Sputnik: Airbnb's Apache Spark Framework for Data Engineering**
Speaker(s): Egor Pakhomov, Airbnb
Track:Data Engineering and Architecture Duration:30m

☐ 2:30 PM to 3:00 PM **The Apache Spark File Format Ecosystem**
Speaker(s): Vinoo Ganesh, Veraset
Track:Developer Duration:30m

☐ 2:30 PM to 3:00 PM **Building Identity Graphs over Heterogeneous Data**
Speaker(s): Sudha Viswanathan, Walmart Labs | Saigopal Thota, Walmart Labs
Track:Developer Duration:30m

☐ 2:30 PM to 3:35 PM **Delta from a Data Engineer's Perspective**
Speaker(s): Palla Lentz, Databricks | Jake Therianos, Databricks
Track:Developer Duration:65m

☐ 3:05 PM to 4:10 PM **Adaptive Query Execution: Speeding Up Spark SQL at Runtime**
Speaker(s): Maryann Xue, Databricks | Ke Jia, Intel Corporation
Track:Developer Duration:65m

☐ 3:05 PM to 3:35 PM **Apache Spark vs Apache Spark: An On-Prem Comparison of Databricks and Open-Source Spark**
Speaker(s): Justin Hoffman, Booz Allen Hamilton
Track:Impact and Innovation Duration:30m

☐ 3:05 PM to 3:35 PM **Efficiently Building Machine Learning Models for Predictive Maintenance in the Oil & Gas Industry with Databricks**
Speaker(s): Daili Zhang, Halliburton | Varun Tyagi, Halliburton
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:05 PM to 3:35 PM **Geospatial Options in Apache Spark**
Speaker(s): Dan Corbiani, Pacific Northwest National Lab
Track:Developer Duration:30m

☐ 3:05 PM to 3:35 PM **Leveraging Apache Spark for Scalable Data Prep and Inference in Deep Learning**
Speaker(s): James Nguyen, Microsoft
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:05 PM to 3:35 PM **Optimize the Large Scale Graph Applications by using Apache Spark with 4-5x Performance Improvements**
Speaker(s): Fuwang Hu, PayPal | Grace (Jie) Huang
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **Productionizing Machine Learning with Apache Spark, MLflow and ONNX from the ground to cloud using SQL Server**
Speaker(s): Daniel Coelho, Microsoft
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:05 PM to 3:35 PM **SQL Performance Improvements at a Glance in Apache Spark 3.0**
Speaker(s): Kazuaki Ishizaki, IBM
Track:Developer Duration:30m

☐ 3:05 PM to 3:35 PM **Taming the Search: A Practical Way of Enforcing GDPR and CCPA in Very Large Datasets with Apache Spark**
Speaker(s): Miao Wang, Adobe, Inc. | Jun Ma, Adobe, Inc.
Track:Data Engineering and Architecture Duration:30m

☐ 3:05 PM to 3:35 PM **Geospatial Analytics at Scale: Analyzing Human Movement Patterns During a Pandemic**
Speaker(s): Joel McCune, ESRI | Jim Young, ESRI
Track:Data & ML Industry Use Cases Duration:30m

☐ 3:40 PM to 4:10 PM **Accelerating Apache Spark Shuffle for Data Analytics on the Cloud with Remote Persistent Memory Pools**
Speaker(s): Jian Zhang, Intel Corporation | Chendi Xue, Intel Corporation
Track:Developer Duration:30m

☐ 3:40 PM to 4:10 PM **All In - Migrating a Genomics Pipeline from BASH/Hive to Spark (Azure Databricks) - A Real World Case Study**
Speaker(s): Victoria Morris, Atrium Health
Track:Data & ML Industry Use Cases Duration:30m

☐ 3:40 PM to 4:10 PM **Automatic Forecasting using Prophet, Databricks, Delta Lake and MLflow**
Speaker(s): Perry Stephenson, Atlassian
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Bucketing 2.0: Improve Spark SQL Performance by Removing Shuffle**
Speaker(s): Jun Guo, Bytedance
Track:Developer Duration:30m

☐ 3:40 PM to 4:10 PM **Building Understanding Out of Incomplete and Biased Datasets using Machine Learning and Databricks**
Speaker(s): Luke Heinrich, Atlassian | Mike Dias, Atlassian
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Patterns and Anti-Patterns for Memorializing Data Science Project Artifacts**
Speaker(s): Derrick Higgins, Blue Cross / Blue Shield of Illinois | Sonjia Waxmonsky, Blue Cross / Blue Shield of Illinois
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Running Emerging AI Applications on Big Data Platforms with Ray On Apache Spark**
Speaker(s): Kai Huang, Intel Corporation | Jason Dai, Intel Corporation
Track:Developer Duration:30m

☐ 3:40 PM to 4:10 PM **Scaling Data and ML with Apache Spark and Feast**
Speaker(s): Willem Pienaar, GOJEK
Track:Data Science, Deep Learning and ML Duration:30m

☐ 3:40 PM to 4:10 PM **Vectorized Deep Learning Acceleration from Preprocessing to Inference and Training on Apache Spark in SK Telecom**
Speaker(s): Hongchan Roh, SK Telecom | Jason Dai, Intel Corporation
Track:Data Science, Deep Learning and ML Duration:30m

☐ 4:00 PM to 5:30 PM **Attendee Party**
Track:Special Event Duration:90m

Friday, June 26

☐ 9:00 AM to 10:30 AM **Friday Morning Keynotes**
Speaker(s): Adam Pazske, PyTorch | Amy Heineike, Primer
Track:Keynote Duration:90m

☐ 10:00 AM to 10:30 AM **Bring Satellite and Drone Imagery into your Data Science Workflows**
Speaker(s): Jason Brown, Astraea, Inc
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:00 AM to 10:30 AM **Building a Pipeline for State-of-the-Art Natural Language Processing Using Hugging Face Tools**
Speaker(s): Lysandre Debut, Hugging Face | Anthony Moi, HuggingFace
Track:Open Source Data and ML Tools Duration:30m

☐ 10:00 AM to 10:30 AM **Cloud-Native Apache Spark Scheduling with YuniKorn Scheduler**
Speaker(s): Li Gao, Lyft, Inc. | Weiwei Yang, Cloudera
Track:Developer Duration:30m

☐ 10:00 AM to 10:30 AM **Continuous Delivery of Deep Transformer-Based NLP Models Using MLflow and AWS Sagemaker for Enterprise AI Scenarios**
Speaker(s): Yong Liu, Outreach Corporation | Andrew Brooks, Outreach Corporation
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:00 AM to 10:30 AM **Democratizing PySpark for Mobile Game Publishing**
Speaker(s): Ben Weber , Zynga
Track:Data Engineering and Architecture Duration:30m

☐ 10:00 AM to 10:30 AM **Flash for Apache Spark Shuffle with Cosco**
Speaker(s): Aaron Feldman, Facebook
Track:Developer Duration:30m

☐ 10:00 AM to 10:30 AM **From Python to PySpark and Back Again - Unifying Single-host and Distributed Deep Learning with Maggy**
Speaker(s): Moritz Meister, Logical Clocks AB | Jim Dowling, Logical Clocks AB
Track:Data and ML Research Duration:30m

☐ 10:00 AM to 10:30 AM **Healthcare Claim Reimbursement using Apache Spark**
Speaker(s): Mohammed Salim Sayed, Optum
Track:Data Engineering and Architecture Duration:30m

☐ 10:00 AM to 10:30 AM **Leveraging Apache Spark to Develop AI-Enabled Products and Services at Bosch**
Speaker(s): Prasanth Lade, Bosch | Goktug Cinar, Bosch
Track:AI Use Cases Duration:30m

☐ 10:00 AM to 10:30 AM **Shparkley: Scaling Shapley with Apache Spark**
Speaker(s): Xiang Huang, Affirm | Cristine Dewar, Affirm
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:00 AM to 11:05 AM **Koalas: Pandas on Apache Spark**
Speaker(s): Niall Turbitt, Databricks
Track:Data Engineering and Architecture Duration:65m

☐ 10:00 AM to 10:30 AM **Enabling Scalable Data Science Pipeline with Mlflow at Thermo Fisher Scientific**
Speaker(s): Allison Wu, Data Scientist
Track:Data & ML Industry Use Cases Duration:30m

☐ 10:35 AM to 11:05 AM **Automated Testing For Protecting Data Pipelines from Undocumented Assumptions**
Speaker(s): Eugene Mandel, Superconductive | Abe Gong, Superconductive
Track:Data Engineering and Architecture Duration:30m

☐ 10:35 AM to 11:05 AM **Chromatic Sparse Learning**
Speaker(s): Vladimir Feinberg, Sisu Data
Track:Data and ML Research Duration:30m

☐ 10:35 AM to 11:05 AM **Is This Thing On? A Well State Model for the People**
Speaker(s): Tristan Arbus, Devon Energy
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:35 AM to 11:05 AM **Best Practices for Building Robust Data Platform with Apache Spark and Delta**
Speaker(s): Vini Jaiswal, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:35 AM to 11:05 AM **Patterns and Operational Insights from the First Users of Delta Lake**
Speaker(s): Dominique Brezinski, Apple
Track:Data Engineering and Architecture Duration:30m

☐ 10:35 AM to 11:05 AM **SparkCruise: Automatic Computation Reuse in Apache Spark**
Speaker(s): Abhishek Roy, Microsoft | Priyanka Gomatam, Microsoft
Track:Developer Duration:30m

☐ 10:35 AM to 11:05 AM **User Defined Aggregation in Apache Spark: A Love Story**
Speaker(s): Erik Erlandson, Red Hat, Inc.
Track:Developer Duration:30m

☐ 10:35 AM to 11:05 AM **Machine Learning Data Lineage with MLflow and Delta Lake**
Speaker(s): Richard Zang, Databricks | Denny Lee, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 10:35 AM to 11:05 AM **The Revolution Will be Streamed**
Speaker(s): R Tyler Croy, Scribd
Track:Data & ML Industry Use Cases Duration:30m

☐ 10:35 AM to 11:05 AM **Deploy and Serve Model from Azure Databricks onto Azure Machine Learning**
Speaker(s): Reema Kuvadia, Microsoft | Tao Li, Microsoft
Track:Data Engineering and Architecture Duration:30m

☐ 10:35 AM to 11:05 AM **Deep Learning at Scale with Apache Spark and Determined**
Speaker(s): Neil Conway, Determined AI | David Hershey, Determined AI
Track:Open Source Data and ML Tools Duration:30m

☐ 11:10 AM to 11:40 AM **Using Machine Learning to Optimize COVID-19 Predictions**
Speaker(s): Scott Black, Databricks | Denny Lee, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 11:10 AM to 12:15 PM **Advanced Natural Language Processing with Apache Spark NLP**
Speaker(s): David Talby, Pacific AI | Alex Thomas, Pacific AI
Track:Data Science, Deep Learning and ML Duration:65m

☐ 11:10 AM to 11:40 AM **Deploying Apache Spark Jobs on Kubernetes with Helm and Spark Operator**
Speaker(s): Tom Lous, GraphIQ, hired by Shell
Track:Developer Duration:30m

☐ 11:10 AM to 11:40 AM **From HDFS to S3: Migrate Pinterest Apache Spark Clusters**
Speaker(s): Daniel Dai, Pinterest | Xin Yao, Pinterest
Track:Data Engineering and Architecture Duration:30m

☐ 11:10 AM to 11:40 AM **How Intuit uses Apache Spark to Monitor In-Production Machine Learning Models at Large-Scale**
Speaker(s): Qingbo Hu, Intuit | Sumanth Venkatasubbaiah, Intuit
Track:Data Engineering and Architecture Duration:30m

☐ 11:10 AM to 11:40 AM **Memory Optimization and Reliable Metrics in ML Pipelines at Netflix**
Speaker(s): Vivek Kaushal, Netflix
Track:Developer Duration:30m

☐ 11:10 AM to 11:40 AM **Model Explanation and Prediction Exploration Using Spark ML**
Speaker(s): Iman Haji, Clarify Health Solutions | Imran Qureshi, Clarify Health Solutions | Alvin Henrick, Clarify Health Solutions
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:10 AM to 11:40 AM **Resource-Efficient Deep Learning Model Selection on Apache Spark**
Speaker(s): Yuhao Zhang, University of California, San Diego | Supun Nakandala, University of California, San Diego
Track:Data and ML Research Duration:30m

☐ 11:10 AM to 11:40 AM **Scaling Up AI Research to Production with PyTorch and MLFlow**
Speaker(s): Joe Spisak, Facebook
Track:Open Source Data and ML Tools Duration:30m

☐ 11:10 AM to 12:15 PM **Simplify and Boost Spark 3 Deployments with Hypervisor-Native Kubernetes**
Speaker(s): Enrique Corro, VMware | Justin Murray, VMware
Track:Data Engineering and Architecture Duration:65m

☐ 11:10 AM to 11:40 AM **Generalized SEIR Model on Large Networks**
Speaker(s): Amir Kermany, Databricks
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:10 AM to 11:40 AM **Rapid Response to Hospital Operations using Data and AI during COVID-19**
Speaker(s): Rohan D, ÃôSouza, KenSci | Tony Pastorino, Indiana University Health
Track:Data & ML Industry Use Cases Duration:30m

☐ 11:45 AM to 12:15 PM **Apache Pulsar: The Next Generation Messaging and Queuing System**
Speaker(s): Matteo Merli, Splunk | Karthik Ramasamy, Splunk
Track:Open Source Data and ML Tools Duration:30m

☐ 11:45 AM to 12:15 PM **Operationalizing Machine Learning at Scale at Starbucks**
Speaker(s): Balaji Venkataraman, Starbucks | Denny Lee, Databricks
Track:Developer Duration:30m

☐ 11:45 AM to 12:15 PM **DataSource V2 and Cassandra - A Whole New World**
Speaker(s): Russell Spitzer, DataStax
Track:Developer Duration:30m

☐ 11:45 AM to 12:15 PM **Designing the Next Generation of Data Pipelines at Zillow with Apache Spark**
Speaker(s): Derek Gorthy, Zillow Group | Nedra Albrecht, Zillow Group
Track:Data Engineering and Architecture Duration:30m

☐ 11:45 AM to 12:15 PM **From Idea to Model: Productionizing Data Pipelines with Apache Airflow**
Speaker(s): Daniel Imberman, Astronomer
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:45 AM to 12:15 PM **Presto on Apache Spark: A Tale of Two Computation Engines**
Speaker(s): Wenlei Xie, Facebook | Andrii Rosa, Facebook
Track:Data and ML Research Duration:30m

☐ 11:45 AM to 12:15 PM **Scoring at Scale: Generating Follow Recommendations for Over 690 Million LinkedIn Members**
Speaker(s): Emilie de Longueau, LinkedIn | Abdulla Al-Qawasmeh, LinkedIn
Track:Data Science, Deep Learning and ML Duration:30m

☐ 11:45 AM to 12:15 PM **Understanding and Improving Code Generation**
Speaker(s): Michael Chen, Workday
Track:Developer Duration:30m

☐ 11:45 AM to 12:15 PM **How R Developers Can Build and Share Data and AI Applications that Scale with Databricks and RStudio Connect**
Speaker(s): James Blair, RStudio | Rafi Kurlansik, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 11:45 AM to 12:15 PM **Improving Therapeutic Development at Biogen with UK Biobank Data, Databricks, and DNAnexus**
Speaker(s): John Ellithorpe, PhD, DNAnexus | David Sexton, Biogen | Frank Austin Nothaft, Databricks
Track:Data & ML Industry Use Cases Duration:30m

☐ 12:20 PM to 12:50 PM **Ray: Enterprise-Grade, Distributed Python**
Speaker(s): Dean Wampler, Anyscale
Track:Data Engineering and Architecture Duration:30m

☐ 12:20 PM to 12:50 PM **Building a Streaming Microservice Architecture: with Apache Spark Structured Streaming and Friends**
Speaker(s): Scott Haines, Twilio
Track:Data Engineering and Architecture Duration:30m

☐ 12:20 PM to 12:50 PM **Lessons from Building Large-Scale, Multi-Cloud, SaaS Software at Databricks**
Speaker(s): Jeff Pang, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 12:20 PM to 12:50 PM **Willump: Optimizing Feature Computation in ML Inference**
Speaker(s): Peter Kraft, Stanford University
Track:Data and ML Research Duration:30m

☐ 12:20 PM to 12:50 PM **Real-Time Forecasting at Scale using Delta Lake and Delta Caching**
Speaker(s): Rashmina Menon, GumGum | Jatinder Assi, GumGum
Track:Data Science, Deep Learning and ML Duration:30m

☐ 12:20 PM to 12:50 PM **Simplifying Disaster Recovery with Delta Lake**
Speaker(s): Zeashan Pappa, Privacera | Itai Weiss, Databricks
Track:Data Engineering and Architecture Duration:30m

☐ 12:20 PM to 12:50 PM **Fugue: Unifying Spark and Non-Spark Ecosystems for Big Data Analytics**
Speaker(s): Han Wang, Lyft, Inc. | Jintao Zhang, Lyft, Inc.
Track:Data Engineering and Architecture Duration:30m

☐ 12:20 PM to 12:50 PM **User Behavior Hashing for Audience Expansion**
Speaker(s): Praveen Pratury, Samsung Research America | Yingnan Zhu, Samsung Research America
Track:Data Science, Deep Learning and ML Duration:30m