Spark Summit North America 201806
全部PPT下载[共147个]

为期三天的 Spark Summit 在美国时间 2018-06-04 ~ 06-06 于旧金山的 Moscone Center 举行，不少人已经注意到，今年的会议已经更名为 Spark+AI, 去年 12 月份时，Databricks 在他们的博客中就已经提到过，2018 年的会议将包括更多人工智能的内容，某种意义上也代表着 Spark 未来的发展方向。作为大数据领域的顶级会议，Spark Summit 2018 吸引了全球近 2000 位技术大咖参会。本次会议议题超过了170多个，有超过一半的议题为机器学习及深度学习。会议的全部日程请参见：https://databricks.com/sparkaisummit/north-america/schedule。

如果想及时了
解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

GitHub 下载地址：https://github.com/397090770/spark-summit-north-america-2018-06
CSDN 下载：https://download.csdn.net/download/w397090770/10485708 (分卷 1)、
https://download.csdn.net/download/w397090770/10484033 (分卷 2)
，为了避免伸手党，CSDN
的文件设置了解压密码（解压密码为不带www的本博客域名，或关注微信公众号 iteblog_hadoop 回复 spark_summit_201806 获取），共需要 2 积分下载。
本站 FTP 下载：https://www.iteblog.com/sparksummit/

全部可下载的PPT

本博客整理了共 147 个 PPT，已经全部上传到 GitHub
供大家下载：（GitHub）：进入GitHub下载本次会议全部PPT

1. 99 Problems but Databricks + Apache Spark Ain't One
2. A Deep Dive into Stateful Stream Processing in Structured Streaming

3. A Machine Learning Approach to Time-Sensitive Data Analysis

4. A Tale of Three Deep Learning Frameworks TensorFlow, Keras, and Deep Learning Pipelines

5. Accelerated Spark on Azure Seamless and Scalable Hardware Offloads in the Cloud

6. Accelerating Data Analysis of Brain Tissue Simulations with Apache Spark

7. Accelerating Inference in the Data Center

8. Accelerating Real Time Analytics with Spark Streaming and FPGAaaS

9. AI as a Service, Build Shared AI Service Platforms Based on Deep Learning Technologies

10. Alchemist An Apache Spark = MPI Interface

11. An End-to-End Spark-Based Machine Learning Stack in the Hybrid Cloud

12. An Update on Scaling Data Science Applications with SparkR in 2018

13. Analytics Zoo - Building Analytics and AI Pipeline for Apache Spark and BigDL

14. Analyzing Blockchain Transactions in Apache Spark

15. Apache Spark Acceleration Using Hardware Resources in the Cloud, Seamlessl

16. Apache Spark and Machine Learning Boosts Revenue Growth for Online Retailers

17. Apache Spark at Apple

18. Apache Spark Based Hyper-
Parameter Selection and Adaptive Model Tuning for Deep Neural Networks

19. Apache Spark Data Source V2

20. Apache Spark for Library Developers

21. Apache Spark-Based Stratification Library for Machine Learning Use Cases

22. Apply Hammer Directly to Thumb; Avoiding Apache Spark and Cassandra AntiPatterns

23. Automated Debugging of Big Data Analytics in Apache Spark Using BigSift

24. Automating and Productionizing Machine Learning Pipelines for Real-Time Scoring

25. Automobile Route Matching with Dynamic Time Warping Using PySpark

26. Avoiding Performance Potholes - Scaling Python for Data Science on Spark

27. Azure Databricks Customer Experiences and Lessons

28. Bighead - Airbnb's End-to-End Machine Learning Platform

29. Bring Your Own Models—Machine Learning as a Service

30. Bringing an AI Ecosystem to the Domain Expert and Enterprise AI Developer

31. Building a Scalable Record Linkage System with Apache Spark, Python 3, and Machine Learning

32. Building Deep Reinforcement Learning Applications on Apache Spark with Analytics Zoo using BigDL

33. Building Intelligent Applications, Experimental ML with Uber's Data Science Workbench

34. Building Machine Learning Algorithms on Apache Spark Scaling Out and Up

35. Building Real-
Time Data Pipeline for Diabetes Medication Recommender System Using Databricks

36. Cardinality Estimation through Histogram in Apache Spark 2.3

37. Cloud Computing Was Built for Web Developers—What Does v2 Look Like for Deep Learning

38. Cloud Cost Management and Apache Spark

39. Cognitive Database An Apache Spark-Based AI-Enabled Relational Database System

40. Conquering Hadoop and Apache Spark with Operational Intelligence

41. Continuous Processing in Structured Streaming

42. Conversational Artificial Intelligence

43. Create a Loyal Customer Base by Knowing Their Personality Using AI-

Based Personality Recommendation Engine
44. Data Science and Enterprise Engineering
45. Deep Credit Risk Ranking
46. Deep Dive into Spark SQL with Advanced Performance Tuning
47. Deep Learning for Domain-Specific Entity Extraction from Unstructured Text
48. Deep Learning for Natural Language Processing Using Apache Spark and TensorFlow
49. Deep Learning for Recommender Systems
50. Deep Learning-Based Opinion Mining for Bitcoin Price Prediction
51. Deploying and Monitoring Heterogeneous Machine Learning Applications
52. Deploying MLlib for Scoring in Structured Streaming
53. Deploying Real-Time Decision Services Using Redis
54. Detecting Mobile Malware with Apache Spark
55. Distributed Inference on Large Datasets Using Apache MXNet and Apache Spark
56. DLoBD An Emerging Paradigm of Deep Learning Over Big Data Stacks
57. Dynamic Class-Based Spark Workload Scheduling and Resource Using YARN
58. Dynamic Healthcare Dataset Generation, Curation & Quality with PySpark
59. Dynamic Priorities for Apache Spark Application's Resource Allocations
60. Efficiently Triaging CI Pipelines with Apache Spark - Mixing 52 Billion EventsDay of Streaming with 40 TBHour of Batch Processing
61. Enabling Composition in Distributed Reinforcement Learning with Ray RLlib
62. Enterprise Data Governance and Compliance at Scale
63. Extending Apache Spark APIs Without Going Near Spark Source or a Compiler
64. Extending Spark SQL API with Easier to Use Array Types Operations
65. Fact Store at Scale for Netflix Recommendations
66. Fiducial Marker Tracking Using Machine Vision
67. Flare and TensorFlare Native Compilation for Spark and TensorFlow Pipelines
68. From Genomics to NLP – One Algorithm to Rule Them All
69. From Prototyping to Deployment at Scale
70. HIPAA Compliant Deployment of Apache Spark on AWS
71. Horovod Uber's Open Source Distributed Deep Learning Framework for TensorFlow
72. How Apache Spark Changed the Way We Hire People
73. How Azure Databricks helped make IoT Analytics a reality
74. How Neural Networks See Social Networks
75. How to Rebuild an End-to-End ML Pipeline with Databricks and Upwork
76. How to Use Millions of Mobile Activity Logs to Understand Our Customers, in Real Time
77. Hunt For Lunar Ice AI Lunar Crater Detector
78. Image Similarity Detection at Scale Using LSH and Tensorflow
79. Implementing AutoML Techniques at Salesforce Scale
80. Insights from Building the Future of Drug Discovery with Apache Spark
81. Integrating Existing C++ Libraries into PySpark
82. Interactive Deep Learning in Cloud via MMLSpark
83. Large Scale Feature Aggregation Using Apache Spark
84. Large Scale Fuzzy Name Matching with a Custom ML Pipeline in Batch and Streaming
85. Large-Scaled Telematics Analytics in Apache Spark
86. Lightning-Fast Analytics for Workday Transactional Data
87. Machine Learning for the Apache Spark Developer

88. MacroBase Efficient Explanation On Big Data
89. Managing Thousands of Spark Workers in Cloud Environment
90. Matchmaking Audiences to Content
91. Meltdown, Spectre and Apache Spark™ Performance
92. Merchant Churn Prediction Using SparkML at PayPal
93. Metrics-Driven Tuning of Apache Spark at Scale
94. Migrating Apache Hive Workload to Apache Spark - Bridge the Gap
95. Model Parallelism in Spark ML Cross-Validation
96. Moment-Based Estimation for Hierarchical Models in Apache Spark
97. Moving eBay's Data Warehouse Over to Apache Spark – Spark as Core ETL Platform at eBay
98. Near Real-Time Netflix Recommendations Using Apache Spark Streaming
99. Nouns are Better than N-Grams
100. Operation Tulip - Using Deep Learning Models to Automate Auction Processes
101. Operationalizing Edge Machine Learning with Apache Spark
102. Operationalizing Machine Learning—Managing Provenance from Raw Data to Predictions
103. Optimizing Apache Spark Throughput Using Intel Optane and Intel Memory Drive Technology
104. Overview of Apache Spark 2.3 - What's New
105. Pandas UDF-Scalable Analysis with Python and PySpark
106. Pharmacy Claims Fraud Detection Using Apache Spark
107. Predictive Maintenance at the Dutch Railways
108. Productionizing H2O Models with Apache Spark
109. Productionizing Spark ML Pipelines with the Portable Format for Analytics
110. Programming by Examples
111. Real-Time Attribution with Structured Streaming and Databricks Delta
112. Real-Time In-Flight Drone Route Optimization with Apache Spark
113. Risk Management Framework Using Intel FPGA, Apache Spark, and Persistent RDDs
114. Scalable Monitoring Using Prometheus with Apache Spark Clusters
115. Scale a Near Real-Time AI System by 4X and Beyond with Apache Spark
116. Scaling Machine Learning at Booking.com with H2O Sparkling Water and FeatureStore
117. Separating Hype from Reality in Deep Learning
118. Serverless Machine Learning on Modern Hardware Using Apache Spark
119. SOS - Optimizing Shuffle IO
120. Spark + AI Helps the FDA Protect the Nation
121. Spark from Notebook to Cloud Native Application
122. Spark NLP Extending Spark ML to Deliver Fast, Scalable & Unified Natural Language Processing
123. Spark SQL Adaptive Execution Unleashes The Power of Cluster in Large Scale
124. Sparser-Faster Parsing of Unstructured Data Formats in Apache Spark
125. State of the Art Natural Language Processing
126. Strava Labs - Exploring a Billion Activity Dataset from Athletes with Apache Spark
127. Streaming Trend Discovery Real-Time Discovery in a Sea of Events
128. The Rise Of Conversational AI with David Low
129. Theory Meets Reality—Large Scale Frequent Pattern Mining with Apache Spark in the Real World
130. Threat Detection and Response at Scale

131. Time Series Forecasting Using Recurrent Neural Network and Vector Autoregressive Model
132. Training neural networks with low precision floats
133. Transparent GPU Exploitation on Apache Spark
134. TuneIn How to Get Your HadoopSpark Jobs Tuned While You're Sleeping
135. Understanding Parallelization of Machine Learning Algorithms in Apache Spark™
136. Using AI to Build a Self-Driving Query Optimizer
137. Using AI to Deliver a Device as a Service
138. Using Apache Spark to Predict Installer Retention from Messy Clickstream Data
139. Using Apache Spark to Tune Spark
140. Using BigDL on Apache Spark to Improve the MLS Real Estate Search Experience at Scale
141. Using Spark-Solr at Scale Productionizing Spark for Search
142. Virtualizing Apache Spark and Machine Learning
143. When Apache Spark meets TiDB
144. Which Data Broke My Code Inspecting Spark Transformations
145. Whirlpools in the Stream
146. Why is My Stream Processing Job Slow
147. Zipline - Airbnb's Machine Learning Data Management Platform