

TRAVEL REPORT FORM

출장보고서

결 재	선임연구 원	팀장	팀장	원장
	06/14	06/14	대결 06/14	06/14
협 조	한정민	이현록	김선태	정홍상

I. Reporter 보고자

Department 소속	Position 직위(직급)	Name 성명	Travel Date 날짜	Note 비고
기후예측본부	선임연구원	한정민	2017/06/04~09	

II. Major Activities 주요업무 수행내용

1. Main Contents and Activities 주요내용 및 활동

- 일시: 2017년 6월 4일 ~ 9일
- 장소: Moscone West Convention Center, San Francisco, CA. USA
- 목적: Spark Summit2017(Data Science and Engineering at Scale) 참석
- 내용
 - : 빅데이터 분산처리 관련 최신 기술 동향 파악 및 분야별 적용사례 청취
 - : 기후정보처리시스템의 분산처리 기술 적용을 위한 선진 지식 습득
 - : 오픈소스 기반의 분산형 데이터 프로세싱 툴에 대한 활용 및 기술설명 청취.
 - : 시스템 전문가, 개발자, 데이터 과학자 등 IT, 정보처리 분야를 선도하는 Global기업들의 시스템 활용 전시회 참가

2. Relevance to APEC Climate Center's Activities 결론 및 소감

- Apache Spark Summit 2017는 developer, data science Machine Learning, spark ecosystem, spark experience and use cases, enterprise, streaming research, technical deep dives로 구성되어 진행되었음.
- 여러 주제 중에 개발자 섹션 위주로 참석하였으며 Apache Spark를 확장한 22개의Use Cases에 대한 설명, Machine Learning 적용 사례, Spark SQL's의 Deep Dive 최적화, TensorFlow와 spark의 Clustering 등 약 30개의 내용 발표를 청취함.
- APCC의 CLIPs는 대용량 기후자료에서 사용자들이 원하는 지역, 기간, 변수에 대해 추출해 주는 서비스로 현재, 단일서버의 분산 메시지 큐 방식으로 운영중에 있으나 사용자의 증가, 추출조건의 복잡도, 대용량 파일의 처리에는 queue에 overHead가 발생하여 자료처리 시간을 지연시킬 문제를 안고 있음.

그리하여, Spark Summit에서 다양한 적용 기법 및 사례를 기반으로 분산, 병렬 처리 적용의 위
험요소를 사전에 식별하고 장기적인 시스템 구축 계획 수립, CLIPs의 멀티서버 클러스터링 및 병
렬 프로세싱 기술 적용으로 빠른 결과를 제공할 수 있는 성능 개선을 하고자 함.

또한, CLIPs의 공간정보 검색(선형, 박스형, 폴리곤, 원형, 포인트형)을 위해 NoSQL기반의
MongoDB를 활용하여 다양한 방식으로 변수를 추출하여 CLIPs의 서비스 기능 확대에 도움이 될
수 있을 것으로 판단됨.

III. References 참고사항

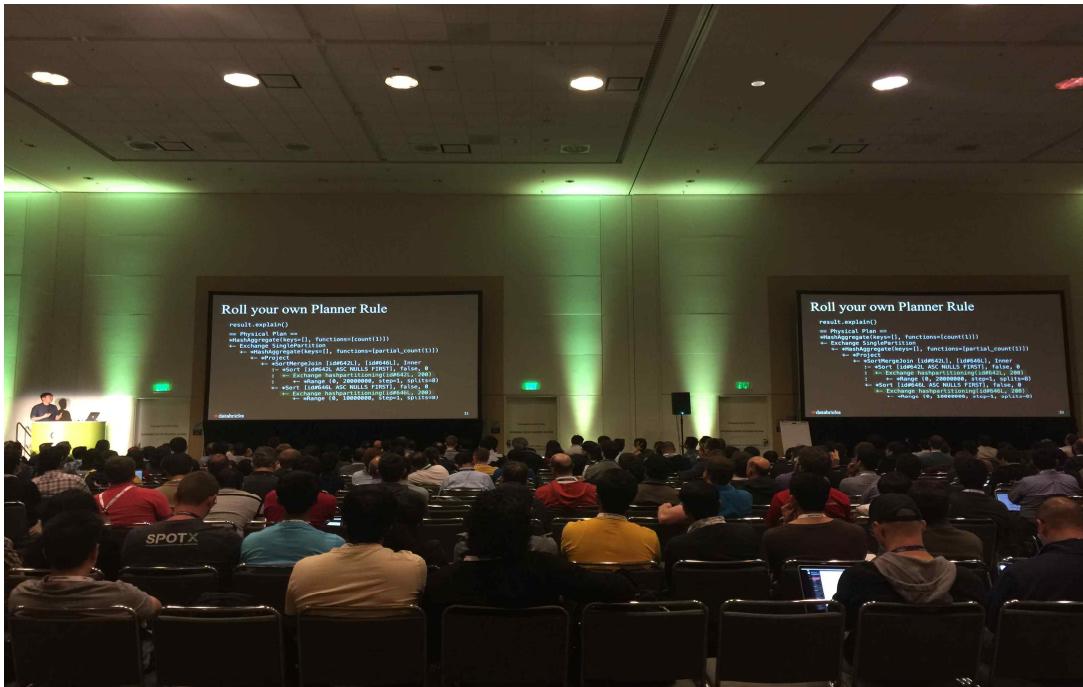
(with attachment of any information or report in case of attendance of conferences,
workshops and meetings) 학술대회, 워크숍, 회의 등 참석 시 관련 정보 및 문서 첨부

1. Presented and Collected Materials 주요 수집자료

- Spark Summit 참가 개발자 연락처 및 관련 자료 수집

2. Suggestions and Remarks 건의사항

3. 발표회장 사진



< 그림 1 Spark Summit 2017 발표회장 (developer 분야) 사진 >

[SPARK SUMMIT HOME \(/\)](#)[SUBSCRIBE](#)[PAST EVENTS ▾ \(/EAST-2017\)](#)

SPARK SUMMIT 2017

DATA SCIENCE AND ENGINEERING AT SCALE

JUNE 5 - 7, 2017 | SAN FRANCISCO

[Home](#)[Schedule](#)[Training](#)[REGISTER NOW \(HTTPS://PREVALENTDESIGNEVENTS.COM/SPARKSUMMIT/SS17/\)](https://prevalentdesignevents.com/sparksummit/ss17/)[Venue](#)[Hackathon](#)[Sponsors](#)Jump to: [Spark Training ▾](#) [Developer Day ▾](#) [Enterprise Day ▾](#)

DAY 1 • MONDAY, JUNE 5 • SPARK TRAINING

7:00 AM [Registration](#)

	TRAINING ROOM 1	TRAINING ROOM 2	TRAINING ROOM 3	TRAINING ROOM 4	TRAINING ROOM 5
9:00 AM	Training: Data Science With Apache Spark 2.x (https://spark-summit.org/2017/events/training-data-science-with-apache-spark-2x/) (3 hours)	Training: Exploring Wikipedia 2 With Apache Spark 2.x (https://spark-summit.org/2017/events/training-exploring-wikipedia-2-with-apache-spark-2x/) (3 hours)	Training: Apache Spark Intro for Machine Learning and Data Science (https://spark-summit.org/2017/events/training-apache-spark-intro-for-machine-learning-and-data-science/) (3 hours)	Training: Apache Spark Intro for Data Engineering (https://spark-summit.org/2017/events/training-apache-spark-intro-for-data-engineering/) (3 hours)	Just Enough Scala for Spark (https://spark-summit.org/2017/events/just-enough-scala-for-spark/) (3 hours)

12:00 PM [Lunch](#)

	TRAINING ROOM 1	TRAINING ROOM 2	TRAINING ROOM 3	TRAINING ROOM 4	TRAINING ROOM 5
1:00 PM	Training: Data Science With Apache Spark 2.x (https://spark-summit.org/2017/events/training-data-science-with-apache-spark-2x/) (4 hours)	Training: Exploring Wikipedia 2 With Apache Spark 2.x (https://spark-summit.org/2017/events/training-exploring-wikipedia-2-with-apache-spark-2x/) (4 hours)	Training: Apache Spark Intro for Machine Learning and Data Science (https://spark-summit.org/2017/events/training-apache-spark-intro-for-machine-learning-and-data-science/) (4 hours)	Training: Apache Spark Intro for Data Engineering (https://spark-summit.org/2017/events/training-apache-spark-intro-for-data-engineering/) (4 hours)	Just Enough Scala for Spark (https://spark-summit.org/2017/events/just-enough-scala-for-spark/) (4 hours)

6:00 PM [Meetup](#)

DAY 2 • TUESDAY, JUNE 6 • DEVELOPER DAY

7:00 AM [Registration](#)

9:00 AM **Keynotes** (<https://spark-summit.org/2017/events/keynotes/>)

10:30 AM **Break**

ROOM 1 ROOM 2 ROOM 3 ROOM 4 ROOM 5 ROOM 6 ROOM 7 ROOM 8 ROOM 9

11:00 AM

DEVELOPER

A Developer's View into Spark's Memory Model (<https://spark-summit.org/2017/events/a-developers-view-into-sparks-memory-model/>)
Sameer Agarwal (<https://spark-summit.org/2017/speakers/sameer-agarwal/>) (Databricks)
(30 minutes)

MACHINE LEARNING

Challenging Web-Scale Graph Analytics with Apache Spark (<https://spark-summit.org/2017/events/challenging-web-scale-graph-analytics-with-apache-spark/>)
Xiangrui Meng (<https://spark-summit.org/2017/speakers/xiangrui-meng/>) (Databricks)
(30 minutes)

SPARK ECOSYSTEM

Analyzing IOT Data in Apache Spark Across Data Centers and Cloud with NetApp Data Fabric and NetApp Private Storage (<https://spark-summit.org/2017/events/analyzing-iot-data-in-apache-spark-across-data-centers-and-cloud-with-netapp-data-fabric-and-netapp-private-storage/>)
Karthikeyan Nagalingam (<https://spark-summit.org/2017/speakers/karthikeyan-nagalingam/>) (NetApp)
Nilesh Bagad (<https://spark-summit.org/2017/speakers/nilesh-bagad/>) (NetApp)
(30 minutes)

SPARK EXPERIENCE AND USE CASES

Scaling Up: How Switching to Apache Spark Improved Performance, Realizability, and Reduced Cost on a Very Large Scale ML Application (<https://spark-summit.org/2017/events/scaling-up-how-switching-to-apache-spark-improved-performance-realizability-and-reduced-cost-on-a-very-large-scale-ml-application/>)
Kexin Xie (<https://spark-summit.org/2017/speakers/kexin-xie/>) (Salesforce.com)
Yacov Salomon (<https://spark-summit.org/2017/speakers/yacov-salomon/>) (Salesforce)
(30 minutes)

ENTERPRISE

Spark Compute as a Service @ Paypal (<https://spark-summit.org/2017/events/spark-compute-as-a-service-paypal/>)
Prabhu Kasinathan (<https://spark-summit.org/2017/speakers/prabhu-kasinathan/>) (PayPal)
(30 minutes)

STREAMING

SSR: Structured Streaming on R for Machine Learning (<https://spark-summit.org/2017/events/ssr-structured-streaming-on-r-for-machine-learning/>)
Felix Cheung (<https://spark-summit.org/2017/speakers/felix-cheung/>) (Microsoft)
(30 minutes)

RESEARCH

Scaling Genetic Data Analysis with Apache Spark (<https://spark-summit.org/2017/events/scaling-genetic-data-analysis-with-apache-spark/>)
Jonathan Bloom (<https://spark-summit.org/2017/speakers/jonathan-bloom/>) (Broad Institute)
Timothy Poterba (<https://spark-summit.org/2017/speakers/timothy-poterba/>) (Broad Institute of MIT and Harvard)
(30 minutes)

SPONSORED SESSIONS

TBA
(30 minutes)

TECHNICAL DEEP DIVES

Data Science Deep Dive: Spark ML with High Dimensional Labels (<https://spark-summit.org/2017/events/tba-2/>)
Michael Zargham (<https://spark-summit.org/2017/speakers/michael-zargham/>) (Cadent)
(30 minutes)

11:40 AM

DEVELOPER

TensorFlowOnSpark: Scalable TensorFlow Learning on Spark Clusters (<https://spark-summit.org/2017/events/tensorflowonspark-scalable-tensorflow-learning-on-spark-clusters/>)

Andy Feng (<https://spark-summit.org/2017/speakers/andy-feng/>) (Yahoo)

Lee Yang (<https://spark-summit.org/2017/speakers/lee-yang/>) (Yahoo)

(30 minutes)

MACHINE LEARNING

Needle in the Haystack—User Behavior Anomaly Detection for Information Security (<https://spark-summit.org/2017/events/needle-in-the-haystack-user-behavior-anomaly-detection-for-information-security/>)

Ping Yan (<https://spark-summit.org/2017/speakers/ping-yan/>) (Salesforce.com)

Wei Deng (<https://spark-summit.org/2017/speakers/wei-deng/>) (Salesforce)

(30 minutes)

SPARK ECOSYSTEM

Apache Kylin: Speed Up Cubing with Apache Spark (<https://spark-summit.org/2017/events/apache-kylin-speed-up-cubing-with-apache-spark/>)

Luke Han (<https://spark-summit.org/2017/speakers/luke-han/>) (Kyligence, Inc.)

Shaofeng Shi (<https://spark-summit.org/2017/speakers/shaofeng-shi/>) (Kyligence Inc)

(30 minutes)

SPARK EXPERIENCE AND USE CASES

Incremental Processing on Large Analytical Datasets (<https://spark-summit.org/2017/events/incremental-processing-on-large-analytical-datasets/>)

Prasanna Rajaperumal (<https://spark-summit.org/2017/speakers/prasanna-rajaperumal/>) (Uber)

(30 minutes)

ENTERPRISE

Using SparkML to Power a DSaaS (Data Science as a Service) (<https://spark-summit.org/2017/events/using-sparkml-to-power-a-dsaas-data-science-as-a-service/>)

Kiran Muglurmuth (<https://spark-summit.org/2017/speakers/kiran-muglurmuth/>) (Comcast)

Sridhar Alla (<https://spark-summit.org/2017/speakers/sridhar-alla/>) (Comcast)

(30 minutes)

STREAMING

Structured-Streaming-as-a-Service with Kafka, YARN, and Tooling (<https://spark-summit.org/2017/events/structured-streaming-as-a-service-with-kafka-yarn-and-tooling/>)

Jim Dowling (<https://spark-summit.org/2017/speakers/jim-dowling/>) (KTH Royal Institute of Technology)

(30 minutes)

RESEARCH

Lazy Join Optimizations Without Upfront Statistics (<https://spark-summit.org/2017/events/lazy-join-optimizations-without-upfront-statistics/>)

Matteo Interlandi (<https://spark-summit.org/2017/speakers/matteo-interlandi/>) (UCLA)

(30 minutes)

SPONSORED SESSIONS

TBA

(30 minutes)

TECHNICAL DEEP DIVES

Data Science Deep Dive: Spark ML with High Dimensional Labels-continues (<https://spark-summit.org/2017/events/tba-4/>)

Michael Zargham (<https://spark-summit.org/2017/speakers/michael-zargham/>) (Cadent)

(30 minutes)

12:20 PM

DEVELOPER
Hive Bucketing in Apache Spark (<https://spark-summit.org/2017/events/hive-bucketing-in-apache-spark/>)
Tejas Patil (<https://spark-summit.org/2017/speakers/tejas-patil/>) (Facebook)
(30 minutes)

MACHINE LEARNING
Random Walks on Large Scale Graphs with Apache Spark (<https://spark-summit.org/2017/events/random-walks-on-large-scale-graphs-with-apache-spark/>)
Min Shen (<https://spark-summit.org/2017/speakers/min-shen/>) (LinkedIn)
(30 minutes)

SPARK ECOSYSTEM
Building a Unified Data Pipeline with Apache Spark and XGBoost (<https://spark-summit.org/2017/events/building-a-unified-data-pipeline-with-apache-spark-and-xgboost/>)
Nan Zhu (<https://spark-summit.org/2017/speakers/nan-zhu/>) (Microsoft)
(30 minutes)

SPARK EXPERIENCE AND USE CASES
How to Productionize Your Machine Learning Models Using Apache Spark MLlib 2.x (<https://spark-summit.org/2017/events/how-to-productionize-your-machine-learning-models-using-apache-spark-mllib-2x/>)
Richard Garriss (<https://spark-summit.org/2017/speakers/richard-garriss/>) (Databricks)
(30 minutes)

ENTERPRISE
How Apache Spark and AI Powers UberEats (<https://spark-summit.org/2017/events/how-apache-spark-and-ai-powers-ubereats/>)
Chen Jin (<https://spark-summit.org/2017/speakers/chen-jin/>) (Uber)
Xian Xing Zhang (<https://spark-summit.org/2017/speakers/xian-xing-zhang/>) (Uber Technologies)
(30 minutes)

STREAMING
The Top Five Mistakes Made When Writing Streaming Applications (<https://spark-summit.org/2017/events/the-top-five-mistakes-made-when-writing-streaming-applications/>)
Mark Grover (<https://spark-summit.org/2017/speakers/mark-grover/>) (Cloudera)
Ted Malaska (<https://spark-summit.org/2017/speakers/ted-malaska/>) (Blizzard, Inc.)
(30 minutes)

RESEARCH
Running Apache Spark on a High-Performance Cluster Using RDMA and NVMe Flash (<https://spark-summit.org/2017/events/running-apache-spark-on-a-high-performance-cluster-using-rdma-and-nvme-flash/>)
Patrick Stuedi (<https://spark-summit.org/2017/speakers/patrick-stuedi/>) (IBM)
(30 minutes)

SPONSORED SESSIONS
TBA
(30 minutes)

TECHNICAL DEEP DIVES
TBA
(30 minutes)

12:50 PM Lunch

ROOM 1 ROOM 2 ROOM 3 ROOM 4 ROOM 5 ROOM 6 ROOM 7 ROOM 8 ROOM 9

2:00 PM

DEVELOPER

Apache Spark MLlib's Past Trajectory and New Directions (<https://spark-summit.org/2017/events/apache-spark-mllibs-past-trajectory-and-new-directions/>)

Joseph Bradley (<https://spark-summit.org/2017/speakers/joseph-bradley/>) (Databricks)
(30 minutes)

MACHINE LEARNING

Extending Spark Machine Learning: Adding Your Own Algorithms and Tools (<https://spark-summit.org/2017/events/extending-spark-machine-learning-adding-your-own-algorithms-and-tools/>)

Holden Karau (<https://spark-summit.org/2017/speakers/holden-karau/>) (IBM)
Seth Hendrickson (<https://spark-summit.org/2017/speakers/seth-hendrickson/>) (IBM)
(30 minutes)

SPARK ECOSYSTEM

Building Data Product Based on Apache Spark at Airbnb (<https://spark-summit.org/2017/events/building-data-product-based-on-apache-spark-at-airbnb/>)

Jingwei Lu (<https://spark-summit.org/2017/speakers/jingwei-lu/>) (Airbnb)
Liyin Tang (<https://spark-summit.org/2017/speakers/liyin-tang/>) (Airbnb)
(30 minutes)

SPARK EXPERIENCE AND USE CASES

Building a Versatile Analytics Pipeline on Top of Apache Spark (<https://spark-summit.org/2017/events/building-a-versatile-analytics-pipeline-on-top-of-apache-spark/>)

Mikhail Chernetsov (<https://spark-summit.org/2017/speakers/mikhail-cheretsov/>) (Grammarly)
(30 minutes)

ENTERPRISE

Herding Cats: Migrating Dozens of Oddball Analytics Systems to Apache Spark (<https://spark-summit.org/2017/events/herding-cats-migrating-dozens-of-oddball-analytics-systems-to-apache-spark/>)

John Cavanaugh (<https://spark-summit.org/2017/speakers/john-cavanaugh/>) (HP)
(30 minutes)

STREAMING

Real-Time Machine Learning Analytics Using Structured Streaming and Kinesis Firehose (<https://spark-summit.org/2017/events/real-time-machine-learning-analytics-using-structured-streaming-and-kinesis-firehose/>)

Caryl Yuhas (<https://spark-summit.org/2017/speakers/caryl-yuhas/>) (Databricks)
Myles Baker (<https://spark-summit.org/2017/speakers/myles-baker/>) (Databricks)
(30 minutes)

RESEARCH

Matrix Factorizations at Scale: a Comparison of Scientific Data Analytics on Spark and MPI Using Three Case Studies (<https://spark-summit.org/2017/events/matrix-factorizations-at-scale-a-comparison-of-scientific-data-analytics-on-spark-and-mpi-using-three-case-studies/>)

Michael Mahoney (<https://spark-summit.org/2017/speakers/michael-mahoney/>) (UC Berkeley)
(30 minutes)

SPONSORED SESSIONS

TBA

(30 minutes)

TECHNICAL DEEP DIVES

TBA

(30 minutes)

2:40 PM

DEVELOPER

Informational Referential Integrity Constraints Support in Apache Spark (<https://spark-summit.org/2017/events/informational-referential-integrity-constraints-support-in-apache-spark/>)

Ioana Delaney (<https://spark-summit.org/2017/speakers/ioana-delaney/>) (IBM)

Suresh Thalamati (<https://spark-summit.org/2017/speakers/suresh-thalamati/>) (IBM)

(30 minutes)

MACHINE LEARNING

Fuzzy Matching on Apache Spark (<https://spark-summit.org/2017/events/fuzzy-matching-on-apache-spark/>)

Jennifer Shin (<https://spark-summit.org/2017/speakers/jennifer-shin/>) (8 Path Solutions)

(30 minutes)

SPARK ECOSYSTEM

Extending the R API for Spark with sparklyr and Microsoft R Server (<https://spark-summit.org/2017/events/extending-the-r-api-for-spark-with-sparklyr-and-microsoft-r-server/>)

Ali Zaidi (<https://spark-summit.org/2017/speakers/ali-zaidi/>) (Microsoft)

(30 minutes)

SPARK EXPERIENCE AND USE CASES

Best Practices for Using Alluxio with Apache Spark (<https://spark-summit.org/2017/events/best-practices-for-using-alluxio-with-apache-spark/>)

Gene Pang (<https://spark-summit.org/2017/speakers/gene-pang/>) (Alluxio)

Haoyuan Li (<https://spark-summit.org/2017/speakers/haoyuan-li/>) (Alluxio)

(30 minutes)

ENTERPRISE

Scaling Data Science Capabilities with Apache Spark at Stitch Fix (<https://spark-summit.org/2017/events/scaling-data-science-capabilities-with-apache-spark-at-stitch-fix/>)

Derek Bennett (<https://spark-summit.org/2017/speakers/derek-bennett/>) (Stitch Fix)

(30 minutes)

STREAMING

A Practical Approach to Building a Streaming Processing Pipeline for an Online Advertising Platform (<https://spark-summit.org/2017/events/a-practical-approach-to-building-a-streaming-processing-pipeline-for-an-online-advertising-platform/>)

Amit Ramesh (<https://spark-summit.org/2017/speakers/amit-ramesh/>) (Yelp)

Yifan Wang (<https://spark-summit.org/2017/speakers/yifan-wang/>) (Yelp)

(30 minutes)

RESEARCH

Apache Spark on Supercomputers: A Tale of the Storage Hierarchy (<https://spark-summit.org/2017/events/apache-spark-on-supercomputers-a-tale-of-the-storage-hierarchy/>)

Costin Iancu (<https://spark-summit.org/2017/speakers/costin-iancu/>) (Lawrence Berkeley National Laboratory)

Nicholas Chaimov (<https://spark-summit.org/2017/speakers/nicholas-chaimov/>) (University of Oregon)

(30 minutes)

SPONSORED SESSIONS

TBA

2:40 PM (15 minutes)

SPONSORED SESSIONS

TBA

2:55 PM (15 minutes)

TECHNICAL DEEP DIVES

TBA

(30 minutes)

3:20 PM	DEVELOPER Tricks of the Trade to be an Apache Spark Rock Star (https://spark-summit.org/2017/events/tricks-of-the-trade-to-be-an-apache-spark-rock-star/) Ted Malaska (https://spark-summit.org/2017/speakers/ted-malaska/) (Blizzard, Inc.) (30 minutes) MACHINE LEARNING Assigning Responsibility for Deteriorations in Video Quality (https://spark-summit.org/2017/events/assigning-responsibility-for-deteriorations-in-video-quality/) Henry Milner (https://spark-summit.org/2017/speakers/henry-milner/) (Conviva) Oleg Vasilyev (https://spark-summit.org/2017/speakers/oleg-vasilyev/) (Conviva) (30 minutes) SPARK ECOSYSTEM HDFS on Kubernetes, Lessons Learned (https://spark-summit.org/2017/events/hdfs-on-kubernetes-lessons-learned/) Kimoon Kim (https://spark-summit.org/2017/speakers/kimoon-kim/) (Pepperdata) (30 minutes) SPARK EXPERIENCE AND USE CASES Experiences Migrating Hive Workload to SparkSQL (https://spark-summit.org/2017/events/experiences-migrating-hive-workload-to-sparksql/) Jie Xiong (https://spark-summit.org/2017/speakers/jie-xiong/) (Facebook) Zhan Zhang (https://spark-summit.org/2017/speakers/zhan-zhang/) (Facebook) (30 minutes) ENTERPRISE Transforming B2B Sales with Spark powered Sales Intelligence (https://spark-summit.org/2017/events/transforming-b2b-sales-with-spark-powered-sales-intelligence/) Songtao Guo (https://spark-summit.org/2017/speakers/songtao-guo/) (LinkedIn) Wei Di (https://spark-summit.org/2017/speakers/wei-di/) (LinkedIn) (30 minutes) STREAMING Easy, Scalable, Fault-Tolerant Stream Processing with Structured Streaming in Apache Spark (https://spark-summit.org/2017/events/easy-scalable-fault-tolerant-stream-processing-with-structured-streaming-in-apache-spark/) Tathagata Das (https://spark-summit.org/2017/speakers/tathagata-das/) (Databricks) (30 minutes) RESEARCH Flare: Scale Up Spark SQL with Native Compilation and Set Your Data on Fire! (https://spark-summit.org/2017/events/flare-scale-up-spark-sql-with-native-compilation-and-set-your-data-on-fire/) Tiark Rompf (https://spark-summit.org/2017/speakers/tiark-rompf/) (Purdue University) (30 minutes) SPONSORED SESSIONS TBA 3:20 PM (15 minutes) SPONSORED SESSIONS TBA 3:35 PM (15 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)
3:50 PM	Break

ROOM 1 ROOM 2 ROOM 3 ROOM 4 ROOM 5 ROOM 6 ROOM 7 ROOM 8 ROOM 9

4:20 PM

DEVELOPER

Improving Python and Spark Performance and Interoperability with Apache Arrow (<https://spark-summit.org/2017/events/improving-python-and-spark-performance-and-interoperability-with-apache-arrow/>)

Julien Le Dem (<https://spark-summit.org/2017/speakers/julien-le-dem/>) (Dremio)

Li Jing (<https://spark-summit.org/2017/speakers/li-jing/>) (Two Sigma Investments, LP)
(30 minutes)

MACHINE LEARNING

Multi-Label Graph Analysis and Computations Using GraphX (<https://spark-summit.org/2017/events/multi-label-graph-analysis-and-computations-using-graphx/>)

Qiang Zhu (<https://spark-summit.org/2017/speakers/qiang-zhu/>) (LinkedIn)

Qingbo Hu (<https://spark-summit.org/2017/speakers/qingbo-hu/>) (LinkedIn)
(30 minutes)

SPARK ECOSYSTEM

More Algorithms and Tools for Genomic Analysis on Apache Spark (<https://spark-summit.org/2017/events/more-algorithms-and-tools-for-genomic-analysis-on-apache-spark/>)

Ryan Williams (<https://spark-summit.org/2017/speakers/ryan-williams/>) (Mount Sinai School of Medicine)
(30 minutes)

SPARK EXPERIENCE AND USE CASES

Lessons Learned from Managing Thousands of Production Apache Spark Clusters Daily (<https://spark-summit.org/2017/events/lessons-learned-from-managing-thousands-of-production-apache-spark-clusters-daily/>)

Henry Davidge (<https://spark-summit.org/2017/speakers/henry-davidge/>) (Databricks)

Josh Rosen (<https://spark-summit.org/2017/speakers/josh-rosen/>) (Databricks)
(30 minutes)

ENTERPRISE

GoDaddy Customer Success Dashboard Using Apache Spark (<https://spark-summit.org/2017/events/godaddy-customer-success-dashboard-using-apache-spark/>)

Baburao Kamble (<https://spark-summit.org/2017/speakers/baburao-kamble/>) (GoDaddy)
(30 minutes)

STREAMING

Dynamic DDL: Adding Structure to Streaming Data on the Fly (<https://spark-summit.org/2017/events/dynamic-ddl-adding-structure-to-streaming-data-on-the-fly/>)

David Winters (<https://spark-summit.org/2017/speakers/david-winters/>) (GoPro)

Hao Zou (<https://spark-summit.org/2017/speakers/hao-zou/>) (GoPro)
(30 minutes)

RESEARCH

Microservices and Teraflops: Effortlessly Scaling Data Science with PyWren (<https://spark-summit.org/2017/events/microservices-and-teraflops-effortlessly-scaling-data-science-with-pywren/>)

Eric Jonas (<https://spark-summit.org/2017/speakers/eric-jonas/>) (UC Berkeley)

Shivaram Venkataraman (<https://spark-summit.org/2017/speakers/shivaram-venkataraman/>) (UC Berkeley)
(30 minutes)

SPONSORED SESSIONS

TBA

4:20 PM (15 minutes)

SPONSORED SESSIONS

TBA

4:35 PM (15 minutes)

TECHNICAL DEEP DIVES

TBA

(30 minutes)

3:00 PM	DEVELOPER Building Robust ETL Pipelines with Apache Spark (https://spark-summit.org/2017/events/building-robust-etl-pipelines-with-apache-spark/) Sameer Agarwal (https://spark-summit.org/2017/speakers/sameer-agarwal/) (Databricks) (30 minutes) MACHINE LEARNING Visualization of Enhanced Spark Induced Naive Bayes Classifier (https://spark-summit.org/2017/events/visualization-of-enhanced-spark-induced-naive-bayes-classifier/) Barry Becker (https://spark-summit.org/2017/speakers/barry-becker/) (ESI Group) (30 minutes) SPARK ECOSYSTEM Spark HBase Connector: Feature Rich and Efficient Access to HBase Through Spark SQL (https://spark-summit.org/2017/events/spark-hbase-connector-feature-rich-and-efficient-access-to-hbase-through-spark-sql/) Bikas Saha (https://spark-summit.org/2017/speakers/bikas-saha/) (Hortonworks) Weiqing Yang (https://spark-summit.org/2017/speakers/weiqing-yang/) (Hortonworks) (30 minutes) SPARK EXPERIENCE AND USE CASES From Python Scikit-learn to Scala Apache Spark—The Road to Uncovering Botnets (https://spark-summit.org/2017/events/from-python-scikit-learn-to-scala-apache-spark-the-road-to-uncovering-botnets/) Avi Aminov (https://spark-summit.org/2017/speakers/avi-aminov/) (Akamai Technologies) (30 minutes) ENTERPRISE TBA (30 minutes) STREAMING Building Continuous Application with Structured Streaming and Real-Time Data Source (https://spark-summit.org/2017/events/building-continuous-application-with-structured-streaming-and-real-time-data-source/) Arijit Tarafdar (https://spark-summit.org/2017/speakers/arijit-tarafdar/) (Microsoft) Nan Zhu (https://spark-summit.org/2017/speakers/nan-zhu/) (Microsoft) (30 minutes) RESEARCH Speeding up Spark with Data Compression on Xeon+FPGA (https://spark-summit.org/2017/events/speeding-up-spark-with-data-compression-on-xeon-fpga/) David Ojika (https://spark-summit.org/2017/speakers/david-ojika/) (University of Florida) (30 minutes) SPONSORED SESSIONS TBA 5:00 PM (15 minutes)
	SPONSORED SESSIONS TBA 5:15 PM (15 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)
5:40 PM	DEVELOPER Behavior-Driven Development (BDD) Testing with Apache Spark (https://spark-summit.org/2017/events/behavior-driven-development-bdd-testing-with-apache-spark/) Aaron Colcord (https://spark-summit.org/2017/speakers/aaron-colcord/) (FIS Global) Zachary Nanfelt (https://spark-summit.org/2017/speakers/zachary-nanfelt/) (FIS) (30 minutes) MACHINE LEARNING The Key to Machine Learning is Prepping the Right Data (https://spark-summit.org/2017/events/the-key-to-machine-learning-is-prepping-the-right-data/) Jean Georges Perrin (https://spark-summit.org/2017/speakers/jean-georges-perrin/) (Zaloni) (30 minutes) SPARK ECOSYSTEM Building a Large Scale Recommendation Engine with Spark and Redis-ML (https://spark-summit.org/2017/events/building-a-large-scale-recommendation-engine-with-spark-and-redis-ml/) Dvir Volk (https://spark-summit.org/2017/speakers/dvir-volk/) (Redis Labs) Shay Nativ (https://spark-summit.org/2017/speakers/shay-nativ/) (Redis Labs) (30 minutes) SPARK EXPERIENCE AND USE CASES Apache Spark and Citizen Science: Using eBird Data to Predict Bird Abundance at Scale (https://spark-summit.org/2017/events/apache-spark-and-citizen-science-using-ebird-data-to-predict-bird-abundance-at-scale/) Tom Auer (https://spark-summit.org/2017/speakers/tom-auer/) (Cornell University) (30 minutes) ENTERPRISE TBA (30 minutes) STREAMING Scalable Monitoring Using Apache Spark and Friends (https://spark-summit.org/2017/events/scalable-monitoring-using-apache-spark-and-friends/) Utkarsh Bhatnagar (https://spark-summit.org/2017/speakers/utkarsh-bhatnagar/) (Tinder) (30 minutes) RESEARCH Accelerating SparkML Workloads on the Intel Xeon+FPGA Platform (https://spark-summit.org/2017/events/accelerating-sparkml-workloads-on-the-intel-xeon-fpga-platform/) Zhongyue Nah (https://spark-summit.org/2017/speakers/zhongyue-nah/) (Intel) (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)

6:10 PM Attendee Reception

DAY 3 • WEDNESDAY, JUNE 7 • ENTERPRISE DAY

8:00 AM Registration

9:00 AM Keynotes (<https://spark-summit.org/2017/events/keynotes/>)

10:30 AM Break

ROOM 1 ROOM 2 ROOM 3 ROOM 4 ROOM 5 ROOM 6 ROOM 7 ROOM 8 ROOM 9

11:00 AM

DEVELOPER

Dr. Elephant for Monitoring and Tuning Apache Spark Jobs on Hadoop (<https://spark-summit.org/2017/events/dr-elephant-for-monitoring-and-tuning-apache-spark-jobs-on-hadoop/>)

Carl Steinbach (<https://spark-summit.org/2017/speakers/carl-steinbach/>) (LinkedIn)

Simon King (<https://spark-summit.org/2017/speakers/simon-king/>) (Pepperdata)
(30 minutes)

MACHINE LEARNING

Embracing a Taxonomy of Types to Simplify Machine Learning (<https://spark-summit.org/2017/events/embracing-a-taxonomy-of-types-to-simplify-machine-learning/>)

Leah McGuire (<https://spark-summit.org/2017/speakers/leah-mcguire/>) (Salesforce.com)
(30 minutes)

SPARK ECOSYSTEM

Apache Spark on Kubernetes (<https://spark-summit.org/2017/events/apache-spark-on-kubernetes/>)

Anirudh Ramanathan (<https://spark-summit.org/2017/speakers/anirudh-ramanathan/>) (Google)

Tim Chen (<https://spark-summit.org/2017/speakers/tim-chen/>) (Hyperpilot)
(30 minutes)

SPARK EXPERIENCE AND USE CASES

Spinach: Providing Ad-Hoc Query Support on Top of Spark SQL (<https://spark-summit.org/2017/events/spinach-providing-ad-hoc-query-support-on-top-of-spark-sql/>)

Daoyuan Wang (<https://spark-summit.org/2017/speakers/daoyuan-wang/>) (Intel)

Yuanjian Li (<https://spark-summit.org/2017/speakers/yuanjian-li/>) (Baidu)
(30 minutes)

ENTERPRISE

Archiving, E-Discovery, and Supervision with Spark and Hadoop (<https://spark-summit.org/2017/events/archiving-e-discovery-and-supervision-with-spark-and-hadoop/>)

Jordan Volz (<https://spark-summit.org/2017/speakers/jordan-volz/>) (Cloudera)
(30 minutes)

DATA SCIENCE

Yelp Ad Targeting at Scale with Apache Spark (<https://spark-summit.org/2017/events/yelp-ad-targeting-at-scale-with-apache-spark/>)

Inaz Alaei-Novin (<https://spark-summit.org/2017/speakers/inaz-alaei-novin/>) (Yelp)

Joe Malicki (<https://spark-summit.org/2017/speakers/joe-malicki/>) (Yelp)
(30 minutes)

RESEARCH

Debugging Big Data Analytics in Apache Spark with BigDebug (<https://spark-summit.org/2017/events/debugging-big-data-analytics-in-apache-spark-with-bigdebug/>)

Matteo Interlandi (<https://spark-summit.org/2017/speakers/matteo-interlandi/>) (UCLA)

Muhammad Ali Gulzar (<https://spark-summit.org/2017/speakers/muhammad-ali-gulzar/>) (UCLA)
(30 minutes)

SPONSORED SESSIONS

TBA

(30 minutes)

TECHNICAL DEEP DIVES

Deep Dive Into Apache Spark Multi-User Performance (<https://spark-summit.org/2017/events/deep-dive-into-apache-spark-multi-user-performance/>)

Mikhail Genkin (<https://spark-summit.org/2017/speakers/mikhail-genkin/>) (IBM)

Peter Lankford (<https://spark-summit.org/2017/speakers/peter-lankford/>) (STAC)
(30 minutes)

11:40 AM	DEVELOPER Productive Use of the Apache Spark Prompt (https://spark-summit.org/2017/events/productive-use-of-the-apache-spark-prompt/) Sam Penrose (https://spark-summit.org/2017/speakers/sam-penrose/) (Mozilla) (30 minutes) MACHINE LEARNING Identify Disease-Associated Genetic Variants Via 3D Genomics Structure and Regulatory Landscapes Using Deep Learning Frameworks (https://spark-summit.org/2017/events/identify-disease-associated-genetic-variants-via-3d-genomics-structure-and-regulatory-landscapes-using-deep-learning-frameworks/) Yi-Hsiang Hsu (https://spark-summit.org/2017/speakers/yi-hsiang-hsu/) () (30 minutes) SPARK ECOSYSTEM Homologous Apache Spark Clusters Using Nomad (https://spark-summit.org/2017/events/homologous-apache-spark-clusters-using-nomad/) Alex Dadgar (https://spark-summit.org/2017/speakers/alex-dadgar/) (Hashicorp) (30 minutes) SPARK EXPERIENCE AND USE CASES Social Media, Spark, Machine Learning, and Data Visualization to Find Patterns and Insight (https://spark-summit.org/2017/events/social-media-spark-machine-learning-and-data-visualization-to-find-patterns-and-insight/) Erik Schlegel (https://spark-summit.org/2017/speakers/erik-schlegel/) (Microsoft) (30 minutes) ENTERPRISE Next Generation Workshop Car Diagnostics at BMW Powered by Apache Spark (https://spark-summit.org/2017/events/next-generation-workshop-car-diagnostics-at-bmw-powered-by-apache-spark/) Bernhard Schlegel (https://spark-summit.org/2017/speakers/bernhard-schlegel/) (BMW) (30 minutes) DATA SCIENCE Data Wrangling with PySpark for Data Scientists Who Know Pandas (https://spark-summit.org/2017/events/data-wrangling-with-pyspark-for-data-scientists-who-know-pandas/) Andrew Ray (https://spark-summit.org/2017/speakers/andrew-ray/) (Silicon Valley Data Science) (30 minutes) RESEARCH Building Genomic Data Processing and Machine Learning Workflows Using Apache Spark (https://spark-summit.org/2017/events/building-genomic-data-processing-and-machine-learning-workflows-using-apache-spark/) Anupama Joshi (https://spark-summit.org/2017/speakers/anupama-joshi/) (Epinomics) Matt Negulescu (https://spark-summit.org/2017/speakers/matt-negulescu/) (Epinomics) (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES Deep Dive Into Apache Spark Multi-User Performance (continues) (https://spark-summit.org/2017/events/deep-dive-into-apache-spark-multi-user-performance-continues/) (30 minutes)
12:20 PM	DEVELOPER TBA (30 minutes) MACHINE LEARNING Large-Scale Ads CTR Prediction with Spark and Deep Learning: Lessons Learned (https://spark-summit.org/2017/events/large-scale-ads-ctr-prediction-with-spark-and-deep-learning-lessons-learned/) Yanbo Liang (https://spark-summit.org/2017/speakers/yanbo-liang/) (Hortonworks) (30 minutes) SPARK ECOSYSTEM Interoperating a Zoo of Data Processing Platforms Using Rheem (https://spark-summit.org/2017/events/interoperating-a-zoo-of-data-processing-platforms-using-rheem/) Yasser Idris (https://spark-summit.org/2017/speakers/yasser-idris/) (Qatar Computing Research Institute) (30 minutes) SPARK EXPERIENCE AND USE CASES Spark, GraphX, and Blockchains: Building a Behavioral Analytics Platform for Forensics, Fraud, and Finance (https://spark-summit.org/2017/events/spark-graphx-and-blockchains-building-a-behavioral-analytics-platform-for-forensics-fraud-and-finance/) Bryan Cheng (https://spark-summit.org/2017/speakers/bryan-cheng/) (BlockCypher) Karen Hsu (https://spark-summit.org/2017/speakers/karen-hsu/) (BlockCypher) (30 minutes) ENTERPRISE Big Data at Audi: Root Cause Analysis in an Automotive Paint Shop Using MLlib (https://spark-summit.org/2017/events/big-data-at-audi-root-cause-analysis-in-an-automotive-paint-shop-using-mllib/) Christian Raimann (https://spark-summit.org/2017/speakers/christian-raimann/) (Audi Business Innovation GmbH) Christoph Kreibich (https://spark-summit.org/2017/speakers/christoph-kreibich/) (Audi) (30 minutes) DATA SCIENCE Smart Scalable Feature Reduction With Random Forests (https://spark-summit.org/2017/events/smart-scalable-feature-reduction-with-random-forests/) Erik Erlandson (https://spark-summit.org/2017/speakers/erik-erlandson/) (Red Hat) (30 minutes) RESEARCH Neuro-Symbolic AI for Sentiment Analysis (https://spark-summit.org/2017/events/neuro-symbolic-ai-for-sentiment-analysis/) Michael Malak (https://spark-summit.org/2017/speakers/michael-malak/) (Oracle) (30 minutes) SPONSORED SESSIONS Women in Big Data Lunch (https://spark-summit.org/2017/events/women-in-big-data-lunch/) (30 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)
12:50 PM	Lunch

	ROOM 1	ROOM 2	ROOM 3	ROOM 4	ROOM 5	ROOM 6	ROOM 7	ROOM 8	ROOM 9
2:00 PM	DEVELOPER Improving Apache Spark with S3 (https://spark-summit.org/2017/events/improving-apache-spark-with-s3/) Ryan Blue (https://spark-summit.org/2017/speakers/ryan-blue/) (Netflix) (30 minutes) MACHINE LEARNING Building Competing Models Using Apache Spark DataFrames (https://spark-summit.org/2017/events/building-competing-models-using-apache-spark-dataframes/) Abdulla Al-Qawasmeh (https://spark-summit.org/2017/speakers/abdulla-al-qawasmeh/) (Credit Karma) (30 minutes) SPARK ECOSYSTEM Cassandra and SparkSQL: You Don't Need Functional Programming for Fun (https://spark-summit.org/2017/events/cassandra-and-sparksql-you-dont-need-functional-programming-for-fun/) Russell Spitzer (https://spark-summit.org/2017/speakers/russell-spitzer/) (DataStax) (30 minutes) SPARK EXPERIENCE AND USE CASES Tuning Apache Spark for Large-Scale Workloads (https://spark-summit.org/2017/events/tuning-apache-spark-for-large-scale-workloads/) Gaoxiang Liu (https://spark-summit.org/2017/speakers/gaoxiang-liu/) (Facebook) Sital Kedia (https://spark-summit.org/2017/speakers/sital-kedia/) (Facebook) (30 minutes) ENTERPRISE Changing the Way Viacom Looks at Video Performance (https://spark-summit.org/2017/events/changing-the-way-viacom-looks-at-video-performance/) Mark Cohen (https://spark-summit.org/2017/speakers/mark-cohen/) (Viacom) Michael Rosencrantz (https://spark-summit.org/2017/speakers/michael-rosencrantz/) (Viacom, Inc.) (30 minutes) DATA SCIENCE Fully-Reproducible ML Deployment with Spark, Pachyderm, and MLeap (https://spark-summit.org/2017/events/fully-reproducible-ml-deployment-with-spark-pachyderm-and-mleap/) Daniel Whitenack (https://spark-summit.org/2017/speakers/daniel-whitenack/) (Pachyderm) Hollin Wilkins (https://spark-summit.org/2017/speakers/hollin-wilkins/) (Combust, Inc.) (30 minutes) DATA SCIENCE Natural Language Processing with CNTK and Apache Spark (https://spark-summit.org/2017/events/natural-language-processing-with-cntk-and-apache-spark/) Ali Zaidi (https://spark-summit.org/2017/speakers/ali-zaidi/) (Microsoft) (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES Sparklyr: Recap, Updates, and Use Cases (https://spark-summit.org/2017/events/sparklyr-recap-updates-and-use-cases/) Javier Luraschi (https://spark-summit.org/2017/speakers/javier-luraschi/) (RStudio) (30 minutes)								

2:40 PM	DEVELOPER Demystifying DataFrame and Dataset (https://spark-summit.org/2017/events/demystifying-dataframe-and-dataset/) Dr. Kazuaki Ishizaki (https://spark-summit.org/2017/speakers/dr-kazuaki-ishizaki/) (IBM) (30 minutes) MACHINE LEARNING Real-Time Image Recognition with Apache Spark (https://spark-summit.org/2017/events/real-time-image-recognition-with-apache-spark/) Nikita Shamgunov (https://spark-summit.org/2017/speakers/nikita-shamgunov/) (MemSQL) (30 minutes) SPARK ECOSYSTEM Applying SparkSQL to Big Spatio-Temporal Data Using GeoMesa (https://spark-summit.org/2017/events/applying-sparksql-to-big-spatio-temporal-data-using-geomesa/) Anthony Fox (https://spark-summit.org/2017/speakers/anthony-fox/) (CCRI) (30 minutes) SPARK EXPERIENCE AND USE CASES Performance Optimization of Recommendation Training Pipeline at Netflix (https://spark-summit.org/2017/events/performance-optimization-of-recommendation-training-pipeline-at-netflix/) Hua Jiang (https://spark-summit.org/2017/speakers/hua-jiang/) (Netflix) (30 minutes) ENTERPRISE From Data to Actions and Insights at Conviva (https://spark-summit.org/2017/events/from-data-to-actions-and-insights-at-conviva/) Rui Zhang (https://spark-summit.org/2017/speakers/rui-zhang/) (Conviva, Inc.) Yan Li (https://spark-summit.org/2017/speakers/yan-li/) (Conviva) (30 minutes) DATA SCIENCE Large-Scaled Insurance Analytics Using Tweedie Models in Apache Spark (https://spark-summit.org/2017/events/large-scaled-insurance-analytics-using-tweedie-models-in-apache-spark/) Yanwei Zhang (https://spark-summit.org/2017/speakers/yanwei-zhang/) (Uber) (30 minutes) DATA SCIENCE ADMM-Based Scalable Machine Learning on Apache Spark (https://spark-summit.org/2017/events/admm-based-scalable-machine-learning-on-apache-spark/) Mohak Shah (https://spark-summit.org/2017/speakers/mohak-shah/) (Robert Bosch LLC) Sauptik Dhar (https://spark-summit.org/2017/speakers/sauptik-dhar/) (Robert Bosch LLC) (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES Sparklyr: Recap, Updates, and Use Cases-Continues (https://spark-summit.org/2017/events/sparklyr-recap-updates-and-use-cases-continues/) (30 minutes)
	DEVELOPER TBA (30 minutes) MACHINE LEARNING No More Cumbersomeness: Automatic Predictive Modeling on Apache Spark (https://spark-summit.org/2017/events/no-more-cumbersomeness-automatic-predictive-modeling-on-apache-spark/) Masato Asahara (https://spark-summit.org/2017/speakers/masato-asahara/) (NEC) Ryohei Fujimaki (https://spark-summit.org/2017/speakers/ryohei-fujimaki/) (NEC) (30 minutes) SPARK ECOSYSTEM Just-in-Time Analytics and the Need for Autonomous Database Administration (https://spark-summit.org/2017/events/just-in-time-analytics-and-the-need-for-autonomous-database-administration/) Kristian Alexander (https://spark-summit.org/2017/speakers/kristian-alexander/) (Algebraix Data Corp.) Wes Holler (https://spark-summit.org/2017/speakers/wes-holler/) (Algebraix) (30 minutes) SPARK EXPERIENCE AND USE CASES Machine Learning as a Service: Apache Spark MLlib Enrichment and Web-Based Codeless Modeling (https://spark-summit.org/2017/events/machine-learning-as-a-service-apache-spark-mllib-enrichment-and-web-based-codeless-modeling/) Zhengyi Le (https://spark-summit.org/2017/speakers/zhengyi-le/) (Suning R&D) (30 minutes) ENTERPRISE Leveraging Apache Spark to Disrupt Airline Pricing Distribution (https://spark-summit.org/2017/events/leveraging-apache-spark-to-disrupt-airline-pricing-distribution/) Anton Diego (https://spark-summit.org/2017/speakers/anton-diego/) (EveryMundo) Daniel Santana (https://spark-summit.org/2017/speakers/daniel-santana/) (EveryMundo LLC) (30 minutes) DATA SCIENCE Write Graph Algorithms Like a Boss (https://spark-summit.org/2017/events/write-graph-algorithms-like-a-boss/) Andrew Ray (https://spark-summit.org/2017/speakers/andrew-ray/) (Silicon Valley Data Science) (30 minutes) DATA SCIENCE TBA (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)

3:50 PM Break

ROOM 1 ROOM 2 ROOM 3 ROOM 4 ROOM 5 ROOM 6 ROOM 7 ROOM 8 ROOM 9

4:20 PM

DEVELOPER

Cost-Based Optimizer in Apache Spark 2.2 (<https://spark-summit.org/2017/events/cost-based-optimizer-in-apache-spark-22/>)
(30 minutes)

MACHINE LEARNING

Deep Learning in Security—Are we ready? (<https://spark-summit.org/2017/events/deep-learning-in-security-are-we-ready/>)
Dr. Jisheng Wang (<https://spark-summit.org/2017/speakers/dr-jisheng-wang/>) (Niara)
(30 minutes)

SPARK ECOSYSTEM

Getting Ready to Use Redis with Apache Spark (<https://spark-summit.org/2017/events/getting-ready-to-use-redis-with-apache-spark/>)
Tague Griffith (<https://spark-summit.org/2017/speakers/tague-griffith/>) (Redis Labs)
(30 minutes)

SPARK EXPERIENCE AND USE CASES

Why You Should Care about Data Layout in the Filesystem (<https://spark-summit.org/2017/events/why-you-should-care-about-data-layout-in-the-filesystem/>)
Vida Ha (<https://spark-summit.org/2017/speakers/vida-ha/>) (Databricks)
(30 minutes)

ENTERPRISE

Leveraging Spark in Ecommerce platform (<https://spark-summit.org/2017/events/leveraging-spark-in-ecommerce-platform/>)
Shafaq Abdullah (<https://spark-summit.org/2017/speakers/shafaq-abdullah/>) (Honest Company)
(30 minutes)

DATA SCIENCE

Using AI for Providing Insights and Recommendations on Activity Data (<https://spark-summit.org/2017/events/using-ai-for-providing-insights-and-recommendations-on-activity-data/>)
Alexis Roos (<https://spark-summit.org/2017/speakers/alexis-roos/>) (Salesforce)
Sammy Namhari (<https://spark-summit.org/2017/speakers/sammy-namhari/>) (Salesforce)
(30 minutes)

DATA SCIENCE

Apache SparkR Under the Hood: How to Debug your SparkR Applications (<https://spark-summit.org/2017/events/apache-sparkr-under-the-hood-how-to-debug-your-sparkr-applications/>)
Hossein Falaki (<https://spark-summit.org/2017/speakers/hossein-falaki/>) (Databricks)
(30 minutes)

SPONSORED SESSIONS

TBA
(30 minutes)

TECHNICAL DEEP DIVES

Real-Time Machine Learning with Redis, Apache Spark, Tensor Flow, and more (<https://spark-summit.org/2017/events/real-time-machine-learning-with-redis-apache-spark-tensor-flow-and-more/>)
Cihan Biyikoglu (<https://spark-summit.org/2017/speakers/cihan-biyikoglu/>) (Redis Labs)
(30 minutes)

3:00 PM	DEVELOPER Continuous Application with FAIR Scheduler (https://spark-summit.org/2017/events/continuous-application-with-fair-scheduler/) Robert Xue (https://spark-summit.org/2017/speakers/robert-xue/) (Groupon) (30 minutes) MACHINE LEARNING Add Deep Learning to Big Data Analytics on Apache Spark Using BigDL (https://spark-summit.org/2017/events/add-deep-learning-to-big-data-analytics-on-apache-spark-using-bigdl/) Xianyan Jia (https://spark-summit.org/2017/speakers/xianyan-jia/) (Intel) Yuhao Yang (https://spark-summit.org/2017/speakers/yuhao-yang/) (Intel) (30 minutes) SPARK ECOSYSTEM From R Script to Production Using rsparkling (https://spark-summit.org/2017/events/from-r-script-to-production-using-rsparkling/) Navdeep Gill (https://spark-summit.org/2017/speakers/navdeep-gill/) (H2O.ai) (30 minutes) SPARK EXPERIENCE AND USE CASES RubiOne: Apache Spark as the Backbone of a Retail Analytics Development Environment (https://spark-summit.org/2017/events/rubione-apache-spark-as-the-backbone-of-a-retail-analytics-development-environment/) Adrian Petrescu (https://spark-summit.org/2017/speakers/adrian-petrescu/) (Rubikloud) (30 minutes) ENTERPRISE Stream All Things—Patterns of Modern Data Integration (https://spark-summit.org/2017/events/stream-all-things-patterns-of-modern-data-integration/) Gwen Shapira (https://spark-summit.org/2017/speakers/gwen-shapira/) (Confluent) (30 minutes) DATA SCIENCE NLP with MLlib: Global Empire-Building for Fun and Profit (https://spark-summit.org/2017/events/nlp-with-mllib-global-empire-building-for-fun-and-profit/) Michelle Casbon (https://spark-summit.org/2017/speakers/michelle-casbon/) (Qordoba) (30 minutes) DATA SCIENCE TBA (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES Real-Time Machine Learning with Redis, Apache Spark, Tensor Flow, and more (continued) (https://spark-summit.org/2017/events/real-time-machine-learning-with-redis-apache-spark-tensor-flow-and-more-continued/) (30 minutes)
5:40 PM	DEVELOPER SparkOscope: Enabling Apache Spark Optimization through Cross Stack Monitoring (https://spark-summit.org/2017/events/sparkoscope-enabling-apache-spark-optimization-through-cross-stack-monitoring/) Yiannis Gkoufas (https://spark-summit.org/2017/speakers/yiannis-gkoufas/) (IBM) (30 minutes) MACHINE LEARNING Deep Learning with Apache Spark and GPUs (https://spark-summit.org/2017/events/deep-learning-with-apache-spark-and-gpus/) Pierce Spitler (https://spark-summit.org/2017/speakers/pierce-spitler/) (Bitfusion) Tim Gasper (https://spark-summit.org/2017/speakers/tim-gasper/) (Bitfusion) (30 minutes) SPARK ECOSYSTEM Distributed End-to-End Drug Similarity Analytics and Visualization Workflow (https://spark-summit.org/2017/events/distributed-end-to-end-drug-similarity-analytics-and-visualization-workflow/) Anahita Bhiwandiwalla (https://spark-summit.org/2017/speakers/anahita-bhiwandiwalla/) (Intel) Dina Suehiro (https://spark-summit.org/2017/speakers/dina-suehiro/) (Intel Corporation) (30 minutes) SPARK EXPERIENCE AND USE CASES The Smart Data Warehouse: Goal-Based Data Production (https://spark-summit.org/2017/events/the-smart-data-warehouse-goal-based-data-production/) Sim Simeonov (https://spark-summit.org/2017/speakers/sim-simeonov/) (Swoop) (30 minutes) ENTERPRISE TBA (30 minutes) DATA SCIENCE Very Large Data Files, Object Stores, and Deep Learning—Lessons Learned While Looking for Signs of Extra-Terrestrial Life (https://spark-summit.org/2017/events/very-large-data-files-object-stores-and-deep-learning-lessons-learned-while-looking-for-signs-of-extra-terrestrial-life/) Gil Vernik (https://spark-summit.org/2017/speakers/gil-vernik/) (IBM Corporation) Graham Mackintosh (https://spark-summit.org/2017/speakers/graham-mackintosh/) (IBM) (30 minutes) DATA SCIENCE TBA (30 minutes) SPONSORED SESSIONS TBA (30 minutes) TECHNICAL DEEP DIVES TBA (30 minutes)
8:00 PM	Attendee Party

Home (<https://spark-summit.org/2017/>)
Schedule (<https://spark-summit.org/2017/schedule/>)
Training (<https://spark-summit.org/2017/apache-spark-training/>)
Venue (<https://spark-summit.org/2017/venue/>)
Hackathon (<http://go.spark-summit.org/2017-hackathon>)
Sponsors (<https://spark-summit.org/2017/sponsors/>)

Event Policy (<https://spark-summit.org/event-policy/>)
About Spark Summit (<https://spark-summit.org/about/>)
Terms of Use (<https://spark-summit.org/terms-of-use/>)

Spark Summit Home (/)
Past Events
Subscribe to the Newsletter



Organized by (<https://databricks.com/>)

If you have questions, or would like information on sponsoring a Spark Summit, please contact organizers@spark-summit.org (<mailto:organizers@spark-summit.org>).

Apache, Apache Spark, Spark, and the Spark logo are trademarks of the Apache Software Foundation (<http://www.apache.org>). The Apache Software Foundation has no affiliation with and does not endorse the materials provided at this event.