

2016
HADOOP
SUMMIT

SAN JOSE, CALIFORNIA

Session Agenda

#HS16SJ

Tuesday, June 28

7:30 - 7:30 Registration Open
 9:00 - 11:00 Opening Keynote
 11:00 - 11:30 Coffee Break in the Community Showcase

Business Tracks: ■ Modern Data Applications ■ Business Adoption
Technical Tracks: ■ Apache Committer Insights ■ Application Development ■ Cloud and Operations ■ Data Science, Analytics and Spark
■ Future of Apache Hadoop ■ Governance and Security ■ IoT and Streaming Sponsor

	Ballroom 220A	Ballroom 220B	Ballroom 220C	210A	210C	230A	230C	211	212	LL21A
11:30 - 12:10	How Macy's Creates Operational Insight on Hadoop. Macy's	What the #*\$* is a Business Catalog and Why You Need It! Hortonworks	From Zero to Data Flow in Hours with Apache NiFi Schlumberger	Open Source Ingredients for Interactive Data Analysis in Spark Microsoft	Cloudbreak - Internals Deep Dive Hortonworks	Big Data for Managers: From Hadoop to Streaming and Beyond SciSpike	Operationalizing YARN Based Hadoop Clusters in the Cloud - Lessons and Opportunities Qubole Inc	Model Management Demo in the Lambda Architecture Accenture	7 Enterprise Case Studies of IoT, Streaming Analytics, and Real Business Value TIBCO Software Inc	Apache Hadoop Crash Course (11:00-1:30)
12:20 - 1:00	Faster, Faster, Faster!: The True Story of a Mobile Analytics Data Mart on Hive Yahoo!, Inc.	Evolving HDFS to a Generalized Distributed Storage Subsystem Hortonworks	Streaming in the Wild with Apache Flink data Artisans	Near Real-time Outlier Detection and Interpretation AT&T, Sqrrl Data, Inc.	The "Connected Vehicle" (IoT and Streaming) - Supporting the Mission to Become Both an Automotive and Mobility Company Ford Motor Company	Apache Ambari - Simplified Cluster Operation and Troubleshooting Hortonworks	Building a Graph Database in Neo4j with Spark & Spark SQL to Gain New Insights from Log Data Hortonworks, TiVo	Knowledge from Noise: Geospatial Analytics at Progressive Insurance Progressive	There is a New Ranger in Town! End-to-End Security and Auditing in a Big-Data-as-a-Service Deployment BlueData	
1:00 - 2:10	Lunch in the Community Showcase Women in Big Data Lunch in Room LL21E									
2:10 - 2:50	A Multi-Colored YARN: Apps and First-Class Support for Services Hortonworks	In-Flux Limiting for a Multi-Tenant Logging Service Symantec Corporation	What is Data? And What Are You Doing? RFS Productions	Increasing Hadoop Resiliency & Performance with EMC Isilon EMC	Building a Smarter Home with Nifi and Spark Hortonworks	Governed Self Service Analytics at eBay eBay	Big Data in the Cloud, the Time has Come BlueData Inc., Dell Inc.	HDFS Analysis for Small Files Expedia	Preventative Maintenance of Robots in Automotive Industry Cisco	
3:00 - 3:40	The Industrial Internet: Big Data, Intelligent Machines, and Smarter Workforce GE Digital	HDFS: Optimization, Stabilization and Supportability Hortonworks	Successes, Challenges and Pitfalls Migrating a SAAS Business to Hadoop Marketo	On-Demand HDP Clusters Using Cloudbreak and Ambari Symantec Corporation	Investigating the Effects of Over Committing YARN Resources Yahoo!, Inc.	Analysis of Major Trends in Big Data Analytics Capital One Financial Corp.	Solving Performance Problems on Hadoop to Move Data Analytics Workloads into Production Actian Corporation	End-to-End Processing of 3.7 Million Telemetry Events per Second Using Lambda Architecture Hortonworks, Symantec Corporation	Deep Learning Using DL4J and Spark on HDP for Fun and Profit Skymind, Hortonworks	Apache Spark Crash Course (3:00-6:00)
3:40 - 4:10	Coffee Break in the Community Showcase									
4:10 - 4:50	Google Cloud Platform Empowers TensorFlow and Machine Learning Google Inc.	HDFS Tiered Storage Microsoft	The Life of an Internet of Things (IoT) Electron; its Journey to Become a Positive Influence for Something Greater Webtrends	Application of Active Learning for Fraud Labeling @ PayPal PayPal Inc.	Elephant In The Cloud: Hadoop on Cloud Challenges and Data Migration at Scale Flipkart Internet Pvt. Ltd.	How Hadoop and a Modern Data Platform Can Enable Transformation in Healthcare Blue Cross Blue Shield of Michigan	Use Cases from Batch to Streaming, MapReduce to Spark, Mainframe to Cloud: Today's ETL Does it All! Syncsort, Hortonworks	H2O: A Platform for Big Math h2o.ai	Top Three - Big Data Governance Issues and How Apache ATLAS resolves it for the Enterprise Hortonworks	
5:00 - 5:40	File Format Benchmark - Avro, JSON, ORC, and Parquet Hortonworks	IoT, Streaming Analytics and Machine Learning: Delivering Real-Time Intelligence With Apache NiFi SAS	SQL and Solr Search with Spark for Big Data Analytics in Your Browser Cloudera	Introducing Kafka Streams, the New Stream Processing Library of Apache Kafka Confluent	LEGO: Data Driven Growth Hacking Powered by Big Data Salesforce.com	Data Preparation and Munging for Data Science: A Field Guide Hortonworks	How to Build a Successful Data Lake Waterline Data	Bridging the Gap of Relational to Hadoop Using Sqoop @ Expedia Expedia	Large Scale Health Telemetry and Analytics with MQTT, Hadoop and Machine Learning DSLs Merck	
5:50 - 6:30	Netflix - Productionizing Spark on YARN for ETL at Petabyte Scale Netflix	SAP Hana and Hadoop: Better Together: Murphy Oil's Deployment Experience Hortonworks	Spark SQL versus Apache Drill: Different Tools with Different Rules MapR Technologies	Workload Automation + Hadoop? Oh Yeah! ...a Match Made in Heaven Malwarebytes	Scheduling Policies and Resource Types in YARN Hortonworks	HDP @ MD Anderson - Starting the Hadoop Journey at a Global Leader in Cancer Research MD Anderson Cancer Center	Instrument Your Instruments: Data-Driven Ops ósense	Processing and Retrieval of Geotagged Unmanned Aerial System Telemetry Hortonworks	Hive Metastore Security of Apache Ranger IBM	
6:30 - 7:30	Sponsor Reception in the Community Showcase									

Wednesday, June 29

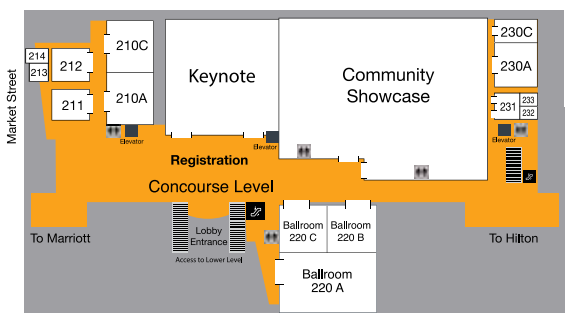
		Business Tracks:			Business Adoption			Technical Tracks:					
7:30 - 6:30		Registration Open			Modern Data Applications			Application Development			Cloud and Operations		
9:00 - 11:00		Opening Keynote			Apache Committer Insights			Governance and Security			Data Science, Analytics and Spark		
11:00 - 11:30		Coffee Break in the Community Showcase			Future of Apache Hadoop			IoT and Streaming			Sponsor		
		Ballroom 220A	Ballroom 220B	Ballroom 220C	210A	210C	230A	230C	211	212	LL21A		
11:30 - 12:10	Spark Uber Development Kit Uber	To Infinity and Beyond – Data-center Scale YARN Clusters through Federation Microsoft	Extending Hortonworks with Oracle’s Big Data Platform Oracle	Building Large-Scale Stream Infrastructures Across Multiple Data Centers with Apache Kafka Confluent	A New “Sparkitecture” for Modernizing Your Data Warehouse HPE	Querying the Internet of Things: Streaming SQL on Kafka/Samza and Storm/Trident Hortonworks	Accelerating Data Warehouse Migration to Hadoop Kyvos Insights, Inc., Impetus	Navigating the World of User Data Management and Data Discovery EMC	Hadoop Application Architectures - Fraud Detection Microsoft	Apache Nifi Crash Course (11:00-1:30)			
12:20 - 1:00	Machine Learning for Any Size of Data, Any Type of Data Google	Real-Time Hadoop: Keys for Success from Streams to Queries MapR Technologies	Debunking the Myths of HDFS Erasure Coding Performance LinkedIn, Intel	Using Hadoop for Cognitive Analytics IBM	Performance Comparison of Streaming Big Data Platforms Capital One, Yahoo!, Inc.	ACID Transactions in Hive Hortonworks	Quark: Simplify and Optimize SQL Queries Across Hadoop and RDBMS Qubole	Tracing Your Security Telemetry With Apache Metron Hortonworks	Beyond TCO: Architecting Hadoop for Adoption and Data Applications RBC				
1:00 - 2:10	Lunch in the Community Showcase												
2:10 - 2:50	The Next Generation of Data Processing & OSS Google	Enterprise-Grade Streaming Under 2ms on Hadoop Capital One	Apache HBase - State of the Union Hortonworks	Apache Kafka Security Netflix, Hortonworks	Blink - Improved Runtime for Flink and its Application in Alibaba Search Alibaba Inc	Yahoo’s Next-Generation User Profile Platform Yahoo!, Inc.	Simultaneous Localization and Mapping (SLAM) with Kafka and Spark Streaming IBM	Curb Your Insecurity - Tips for a Secure Cluster (with Spark too) Hortonworks	Lambda Architecture: How we Merged Batch and Real-Time Trulia				
3:00 - 3:40	Apache Hive 2.0 SQL Speed Scale Hortonworks	Scalable Realtime Analytics using Druid Hortonworks	Toward Better Multi-Tenancy Support from HDFS Hortonworks	Omid: A Transactional Framework for HBase Yahoo!, Inc.	Building and Managing Large Scale Data Pipelines with Complex Dependencies Using Apache Oozie Yahoo!, Inc.	Automated Systems for Loan Decisions Using AKKA and Spark Capital One	Future of Apache Hadoop – An Enterprise Architecture View PwC	“I’m Being Followed by Drones!” The Impact of IoT on the Future of Unmanned Aerial Systems Cognizant	Prescient Keeps Travelers Safe with Natural Language Processing and Geospatial Analytics Prescient	Internet of Things Crash Course (3:00-6:00)			
3:40 - 4:10	Coffee Break in the Community Showcase												
4:10 - 4:50	The Architectural Journey to our Modern Data Applications – DSC (Data Supply Chain) Ford Motor Company	The Ecosystem is Too Damn Big Datameer	The Intuit Analytics Cloud 101 Intuit	Fine-Grained Security for Spark and Hive Hortonworks	Hive Hbase Metastore - Improving Hive with a Big Data Metadata Storage Hortonworks	Statistical Analysis of Genomic Data with Hadoop Arizona State University	Designing and Implementing Your IoT Solution with Open Source MapR Technologies	Analyzing Telecom Fraud at Hadoop Scale Diyotta	Building a Data Analytics PaaS for Smart Cities EMC				
5:00 - 5:40	Reliable and Scalable Data Ingestion at Airbnb Airbnb	Lambda-less Stream Processing @ Scale in LinkedIn LinkedIn	Self-Service Analytics on Hadoop: Lessons Learned Comcast	Make Streaming Analytics Work For You: The Devil is in the Details Hortonworks, Neustar	Operating and Supporting Apache HBase - Best Practices and Improvements Hortonworks	Apache Eagle - Secure Hadoop in Real Time eBay Inc.	Real Time Visualizing Machine Learning with Spark GoPro	Filling the Data Lake Pentaho	The Stream is the Database - Revolutionizing Healthcare Data Architecture MapR Technologies, Liaison Technologies				
5:50 - 6:30	Managing a Large Multi-tenant Data Lake Comcast	Hybrid Data Platform – Cloud Environment Connected with On-premise Data Environment Impetus	High-Scale Entity Resolution in Hadoop eBay, Inc	Presto, What’s New in SQL-on-Hadoop and Beyond Teradata, Facebook	Scalable Optical Character Recognition with Apache NiFi and Tesseract Hortonworks	From Device to Data Center to Insights: Architectural Considerations for the Internet of Anything Hortonworks	Resource Aware Scheduling in Storm Yahoo! Inc.	“The Path to Wellness Through Big Data” PHEMI Systems, MARYCAIREMD	Disrupting Insurance with Advanced Analytics - The Next Generation Carrier Motorists Insurance, Saama Tech.				
6:30 - 9:30	10-Years of Hadoop Party in the Keynote and Community Showcase												

Thursday, June 30


7:30 - 4:00 Registration Open
 9:00 - 11:00 Opening Keynote
 11:00 - 11:30 Coffee Break in the Community Showcase

Business Tracks: Modern Data Applications (Blue), Business Adoption (Light Blue)
Technical Tracks: Apache Committer Insights (Orange), Application Development (Light Green), Cloud and Operations (Light Purple), Data Science, Analytics and Spark (Light Blue)
 Future of Apache Hadoop (Light Green), Governance and Security (Light Purple), IoT and Streaming (Yellow), Sponsor (White)

	Ballroom 220A	Ballroom 220B	Ballroom 220C	210A	210C	230A	230C	211	212	LL21A
11:30 - 12:10	War on Stealth Cyberattacks that Target Unknown Vulnerabilities Hortonworks	Cross-DC Fault-Tolerant View-FileSystem at Twitter Twitter	IoT, Big Data, Cloud – the Convergence of Marketing Terms? Dell Software	State of Security in Spark Hortonworks	The Future of Apache Storm Hortonworks	Yahoo’s Experience Running Pig on Tez at Scale Yahoo!, Inc.	Meeting Performance Goals in Multi-tenant Hadoop Clusters Duke University and Unravel Data Systems, YP	A Data Lake and a Data Lab to Optimize Operations and Safety Within a Nuclear Fleet EDF	Hadoop & Cloud Storage: Object Store Integration in Production Hortonworks	Data Science Crash Course (11:00-1:30)
12:20 - 1:00	(Big data) Squared: How YARN Timeline Service v.2 Unlocks 360-Degree Platform Insights at Scale Twitter Inc., Hortonworks	How We Re-Engineered Phoenix with a Cost-Based Optimizer Based on Calcite Hortonworks, Intel	Distributed Deep Learning on Hadoop Clusters Yahoo!, Inc	LLAP: Sub-Second Analytical Queries in Hive Hortonworks	HIPAA Compliance in the Public Cloud Amazon Web Services	Big Data Heterogeneous Mixture Learning on Spark NEC	Embeddable Data Transformation for Real-Time Streams Rocana	It’s Time: Launching Your Advanced Analytics Program for Success in a Mature Industry Like Oil and Gas ConocoPhillips, Hortonworks	Turning the Stream Processor into a Database: Building Online Applications on Streams data Artisans	
1:00 - 2:10	Lunch in the Community Showcase									
2:10 - 2:50	Combining Machine Learning Frameworks with Apache Spark Databricks	Hadoop in the Cloud – The What, Why and How from the Experts Microsoft	The Evolution of Big Data Pipelines at Intuit Intuit	The Columnar Era: Leveraging Parquet and Kudu for High-Performance Analytics Dremio	Managing Hadoop, HBase, and Storm Clusters at Yahoo Scale Yahoo!, Inc.	Integrating Apache Spark and NiFi for Data Lakes Think Big, a Teradata Company	Lego-Like Building Blocks of Storm and Spark-Streaming Pipelines for Rapid IOT and Streaming Analytics App Dev. Impetus Technologies	Debugging YARN Cluster in Production Hortonworks	Internet Of Things: What about Data Storage? Hortonworks	
3:00 - 3:40	Instilling Confidence and Trust - Big Data Security & Governance MasterCard	Cost and Resource Tracking for Hadoop Yahoo!, Inc.	Zero Downtime App Deployment Using Hadoop Hortonworks	Real-time, Streaming Adv. Analytics, Approx. & Recomm. using Apache Spark ML/GraphX, Kafka Stanford CoreNLP, and Twitter Algebird PipelineIO	Phoenix + HBase: An Enterprise Grade Data-Warehouse Appliance for Interactive Analytics? Hortonworks	Customer Journey - Sentiment Analysis for Fashion Retail Hortonworks, EXPRESS	Ingest and Stream Processing - What Will You Choose? Cloudera, StreamSets	Big Data Ready Enterprise Framework Wipro Technologies	Next Gen Big Data Analytics with Apache Apex DataTorrent	Apache Spark Crash Course (3:00-6:00)
3:40 - 4:10	Coffee Break on the Concourse									
4:10 - 4:50	Keep your Hadoop Cluster at its Best! Hortonworks	Building A Scalable Data Science Platform with R Microsoft	Modernizing Your Company’s Data Ecosystem SAS	Zeppelin + Livy: Bringing Multi Tenancy to Interactive Data Analysis Hortonworks	Effective Spark on Multi-Tenant Clusters Cloudera	GoodFit - An Efficient MRP (Multi-Resource Packing) Allocator for YARN Microsoft	Swimming Across the Data Lake - Lessons Learned and Keys to Success! Impetus	The DAP: Where Yarn, HBase, Kafka and Spark go to Production Cask	Fighting Fraud in Real Time by Processing 1M+ TPS Using Storm on Slider (YARN) Rocket Fuel, Inc	
5:00 - 7:00	BOF: Apache Spark, Apache Zeppelin & Data Science	BOF: Cloud & Operations	BOF: Streaming & Data Flow	BOF: Security & Governance	BOF: Apache Hive & Apache Pig	BOF: Apache Hadoop YARN	BOF: Apache Hadoop HDFS	BOF: Apache HBase		



W. San Carlos Street



HADOOP SUMMIT APP DOWNLOAD NOW!

Stay connected with the latest agenda, sponsors and show information.