



PRESENTED BY



Building machine-learning apps with Spark

MLlib, ML Pipelines, and GraphX

strataconf.com

#StrataHadoop

Agenda

Overview	10 min	
Lab Environment Setup	10-15 min	IntelliJ/Scala IDE for Eclipse/Zeppelin
MLlib	90 min	Overview, Linear Regression, Random Forest, Clustering, Recommendations, FPG, Text Analytics
Break	10 min	
GraphX	50 min	Overview, Exploring Structures, Community-Affiliation, Algorithms, The AlphaGo Community, Wikipedia Page Rank
ML Pipelines	15 min	CrossValidation
Streaming MLlib	10 min	Streaming K-Means
Closing	10 min	