

## Scalable Machine Learning with Apache Spark

### Day 1:

- Spark / ML overview
- Exploratory data analysis (EDA) and feature engineering with Spark
- SparkML: transformers, estimators, pipelines, and evaluators
- MLflow Tracking and Model Registry

### Day 2:

- Parallelizable hyperparameter tuning
- Databricks AutoML and Feature Store
- Integrating 3rd party packages (distributed XGBoost)
- Distributed inference of scikit-learn models with pandas UDFs
- Distributed training with pandas function API
- Pandas API on Spark for data manipulation