Optimizing Apache Spark on Databricks

Day 1

- Review of Spark architecture and Spark UI
- Skew
- Spill
- Shuffle
- Storage
- Serialization

Day 2:

- Ingestion basics
- Predicate push downs
- Disk partitioning
- Z-ordering
- Bucketing
- Optimization with Adaptive Query Execution (AQE)
- Designing and configuring clusters for high performance