Serialization:

Kryo Me a River

Objective

Use the fast Kryo serializer

- performance improvement
- memory efficiency

Kryo Serialization

Fast and memory efficient serializer

- can't serialize everything

- needs explicit class registration

```scala
val sparkConf = new SparkConf()
sparkConf
  .set("spark.serializer", "org.apache.spark.serializer.KryoSerializer")
  .set("spark.kryo.registrationRequired", "true")
  .registerKryoClasses(
    Array(
      classOf[Person],
      classOf[Array[Person]],
    )
)
```

← set Kryo as serializer

← tip: mandatory registration

← register classes
(tip: register arrays as well)

```scala
val spark = SparkSession.builder()
  .appName("Kryo serialization")
  .config(sparkConf)
  .master("local[*]")
  .getOrCreate()
```

← pass config to SparkSession

With caching & shuffles: 10-20% perf boost, 20-30% less memory

Spark rocks