

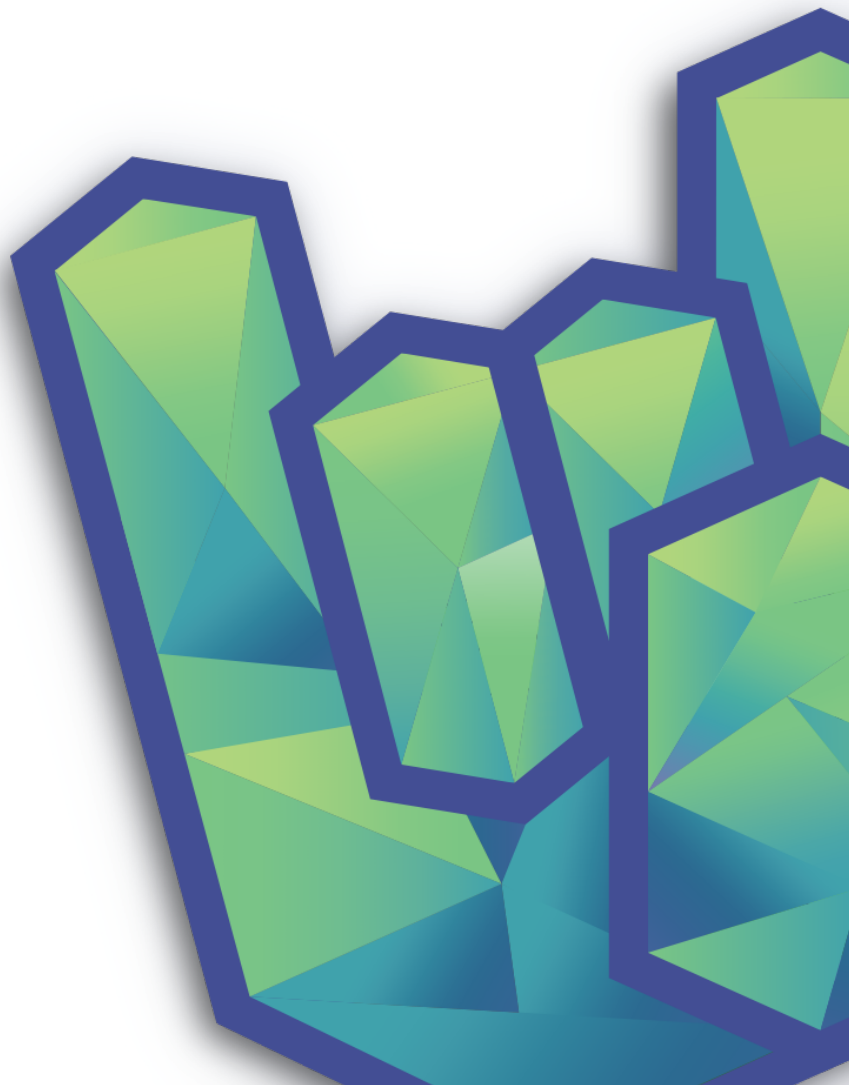
RDDs



Objective

Work with RDDs

- transformations
- partitioning
- exercises



Takeaways

Collection transformations: *map*, *flatMap*, *filter*

```
stocksRDD.map(_._price)
```

Actions: count, min, reduce

```
stocksRDD.min
```

← needs implicit ordering

Grouping

```
stocksRDD.groupBy(_._symbol)
```

← involves shuffling

Partitioning

```
stocksRDD.repartition(40)
```

← involves shuffling

```
repartitionedStocksRDD.coalesce(10)
```

← is NOT a full shuffle

Spark rocks

