Solution: Resilient Distributed Datasets (RDDs)

- RDDs
  - Immutable, partitioned collection of records
  - Built through coarse grained transformations (map, join …)
  - Can be cached for efficient reuse