

Quiz

- PTransforms such as Pardo does **NOT** mutate their input elements
- Pardo takes a **DoFn subclass** argument
- Pardo can **NOT** access random elements
- Pardo resembles **WHERE** SQL operation
- CoGroupByKey resembles **JOIN** SQL operation

Demos

- Deduplicate: group by key and take the first value
- Know how to switch from DirectRunner to DataflowRunner
- Should be able to finish with less than 12 workers
- ParDo supports side inputs (Takes_single.py)
- Flatten is similar to SQL UNION