



# Spark Homework

Brian Cervantes Alvarez

March 5, 2025

## Part 1

---

```
# Part 1
rdd1 = sc.parallelize(range(25), 3)
print(rdd1.glom().collect())
```

## Part 2

---

```
rdd2 = sc.parallelize(range(50), 4)
print(rdd2.glom().collect())
rdd2_coalesced = rdd2.coalesce(2)
print(rdd2_coalesced.glom().collect())
```

## Part 3

---

```
rdd3 = sc.parallelize(range(26))
evenOdd = rdd3.map(lambda x: (x % 2 == 0, x)).partitionBy(2)
print(evenOdd.map(lambda pair: pair[1]).glom().collect())
```

## Part 4

---

```
strings = ["str" + str(i) for i in range(20)]
rdd4 = sc.parallelize(strings, 3)
print(rdd4.glom().collect())
rdd4_all_in_one = rdd4.coalesce(1).repartition(4)
print(rdd4_all_in_one.glom().collect())
```

## Part 5

---



```
rdd5 = sc.parallelize(range(100))
print(rdd5.repartition(20).glom().collect())

rdd5_kv = sc.parallelize(range(100)).map(lambda x: (x, x))
print(rdd5_kv.partitionBy(20).glom().collect())
```