COMBINE BY KEY

-

NEED??

- Group By key is Expensive
  - Every Single key value pair will be shuffled
- Combine by key is one of optimization
- More generic than aggregate by key
- Input and output type can be different

INPUT

- Need 3 Inputs
- Create combiner
- Combiner Function
- Merge Function

AVERAGE SCORE EXAMPLE

```
// input
val in = Seq( "harry=78","raj=60" , "harry=75" , "harry=67" , "raj=89" , "raj=67" , "raj=72" ,"harry=72")
```

COMBINER SYNTAX

```
combineByKey(init_combiner, combiner_function, merger_function)
```

CREATE COMBINER

```
//combiner function
val init_comb = (score: Double) => (1, score)
```

COMBINER

- Combine values within a partition

```
//combine values within partition
val partitionCombiner = (part: (Int, Double), score: Double) => {
      (part._1 + 1, part._2 + score)
    }
```

MERGER FUNCTION

- Merge Output From Partition

```
// Partition Merger Function
val partMerger = (part1: (Int, Double), part2: (Int, Double)) => {
      (part1._1 + part2._1, part1._2 + part2._2)
   }
```