Accumulators

What are Accumulators?

- Variables that are only added to through an associative and commutative operation, and can therefore be efficiently supported in parallel.
- Keep track of the progress of certain operations happening in parallel.
  - Examples: sums, using MapReduce
- Currently only numeric accumulators are supported natively by spark, but programmers can add support for new types.

Accumulator Use

- Created from initial value `v` by calling `SparkContext.accumulator(v)`.
- Tasks running on cluster add to it using add method or `+=` operator

```
>>> accum = sc.accumulator(0)
Accumulator<id=0, value=0>

>>> sc.parallelize([1, 2, 3, 4]).foreach(lambda x: accum.add(x))
...
10/09/29 18:41:08 INFO SparkContext: Tasks finished in 0.317106 s

scala> accum.value
10
```