livelessons
video instruction from technology experts

Lesson 1: Introduction to the Spark Environment

1.10 MapReduce with Spark: Programming with Key-Value Pairs

Key-Value Operations

```
pets = sc.parallelize([("cat", 1), ("dog", 1), ("cat", 2)])

pets.reduceByKey(lambda x, y: x + y) # => {(cat, 3), (dog, 1)}

pets.groupByKey() # => {(cat, [1, 2]), (dog, [1])}

pets.sortByKey() # => {(cat, 1), (cat, 2), (dog, 1)}
```

Functional Programming Primer

- Functions are `applied` to data (RDDs)
- RDDs are Immutable: `f(RDD) -> RDD2`
- Function application necessitates creation of new data

Review

- Client-Server execution model
- Spark leverages higher-order functions (`map()`, `filter()`, etc.)
- Transformations create new RDDs and are lazily evaluated
- Actions force materialization of RDD on driver