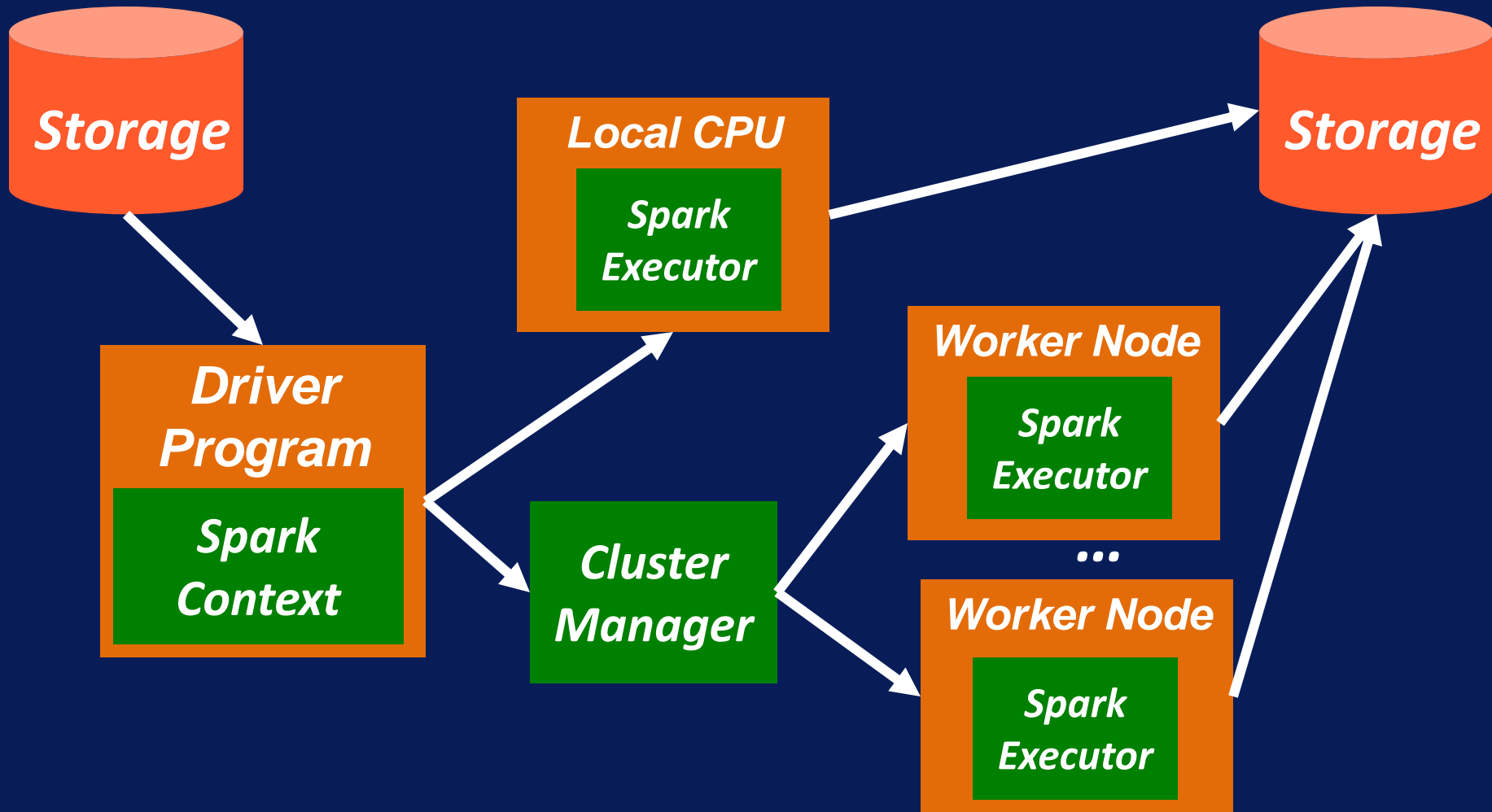


Spark Core: Programming In Spark



After this video you will be able to..

- Use two methods to create RDDs in Spark
- Explain what immutable means
- Interpret a Spark program as a pipeline of transformations and actions
- List the steps to create a Spark program



Creating RDDs

*Driver
Program*

```
In [1]: lines = sc.textFile("hdfs://user/cloudera/words.txt")
```

```
lines = sc.parallelize(["big", "data"])
```

```
numbers = sc.parallelize(range(10), 3)
```

Parallelize
range output
into 3 partitions

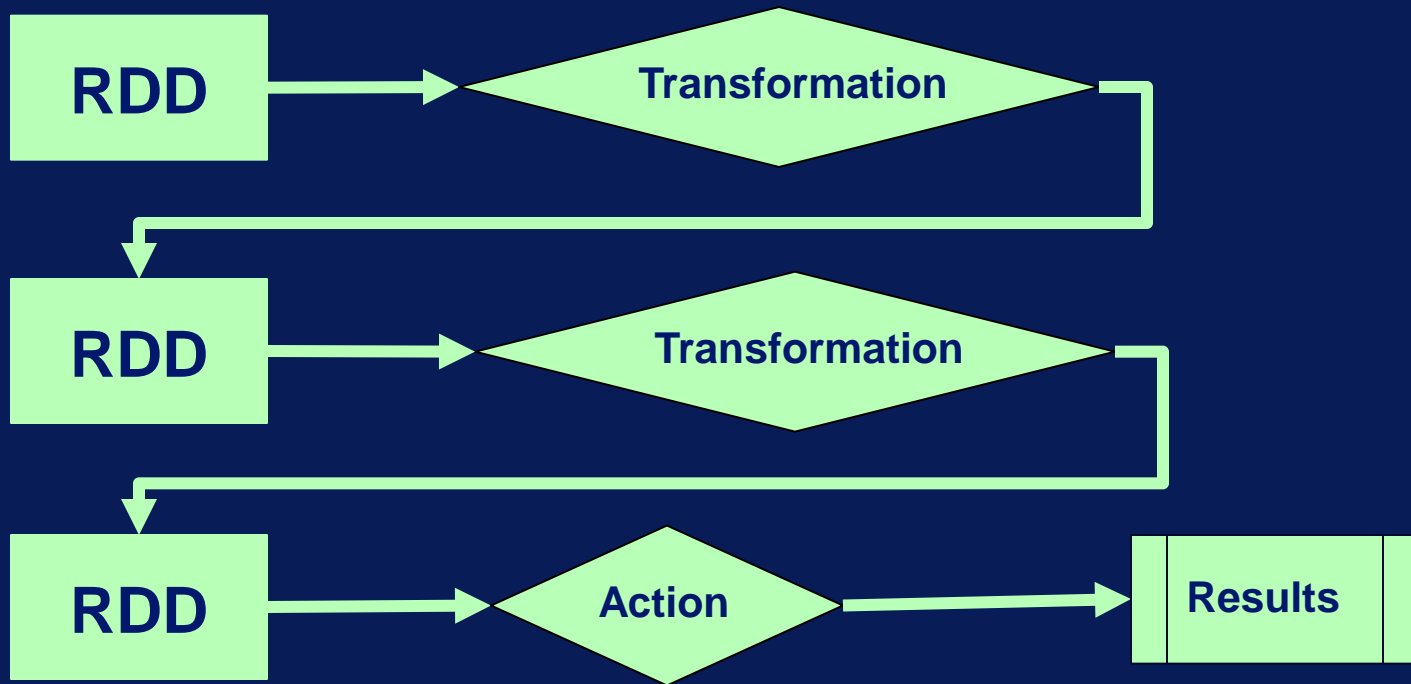
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

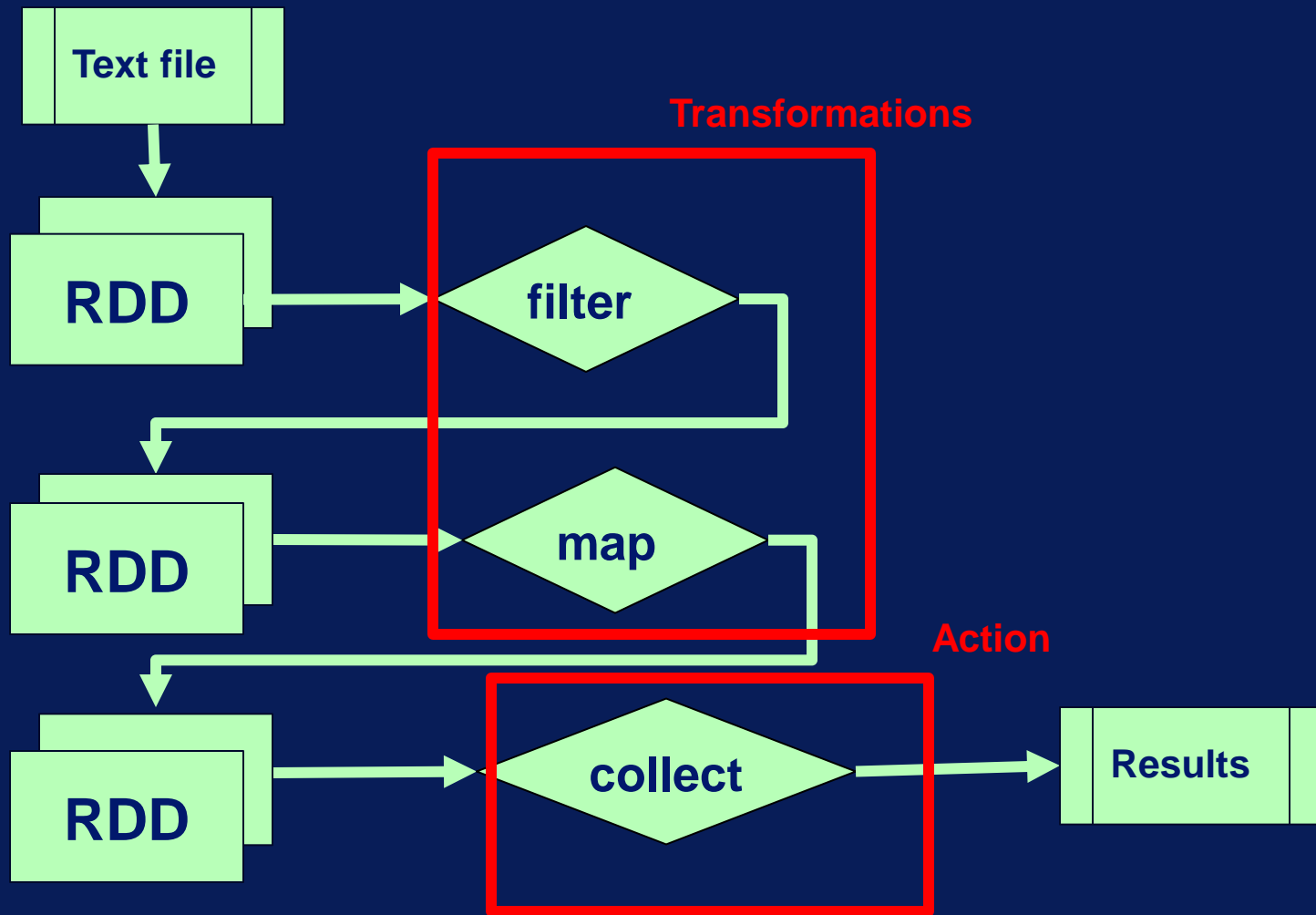
```
numbers.collect()
```

[0, 1, 2], [3, 4, 5], [6, 7, 8, 9]

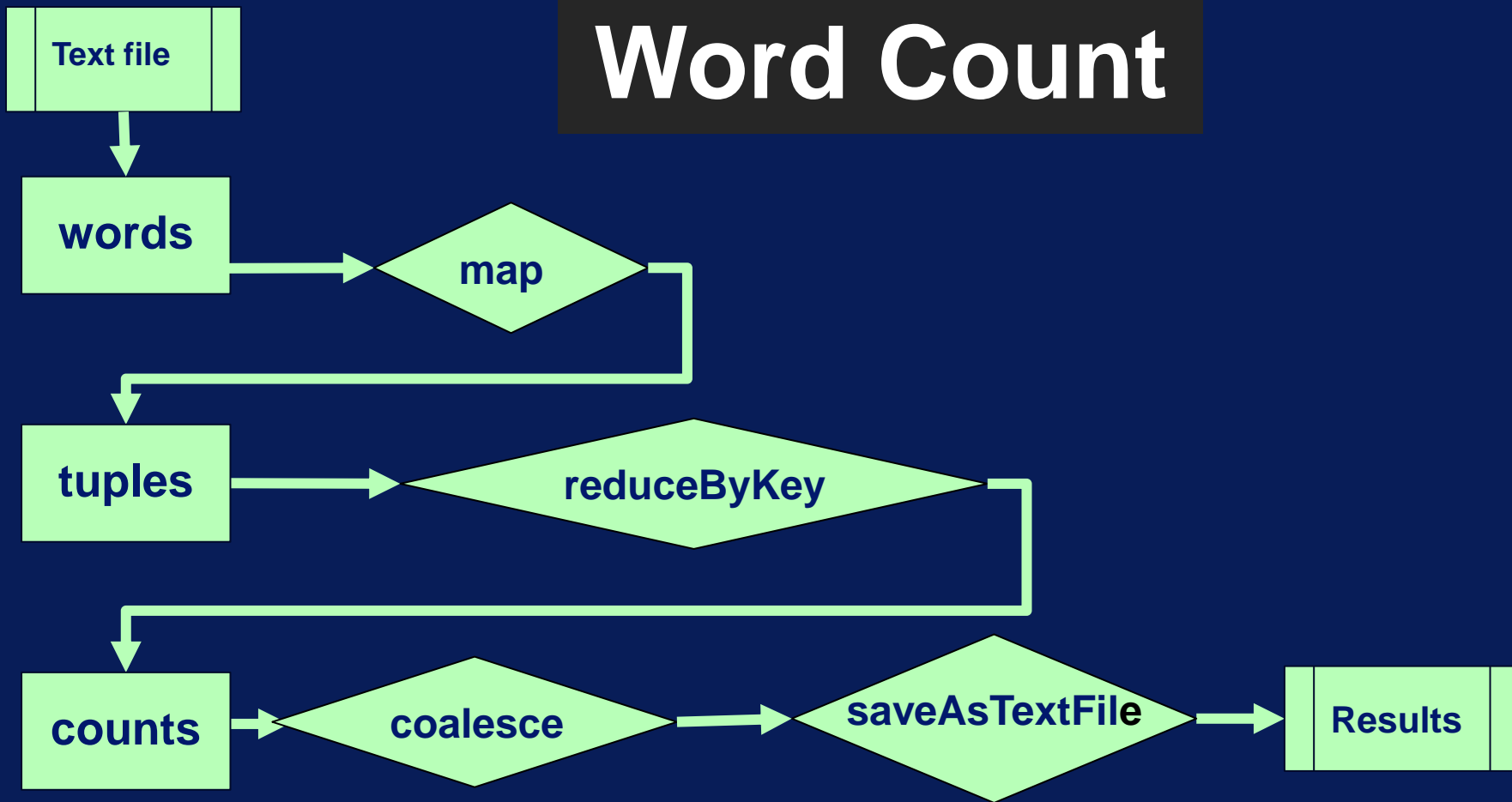
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

Processing RDDs





Word Count



Programming in Spark

Create RDDs



Apply transformations



Perform actions