Spark Core: Actions

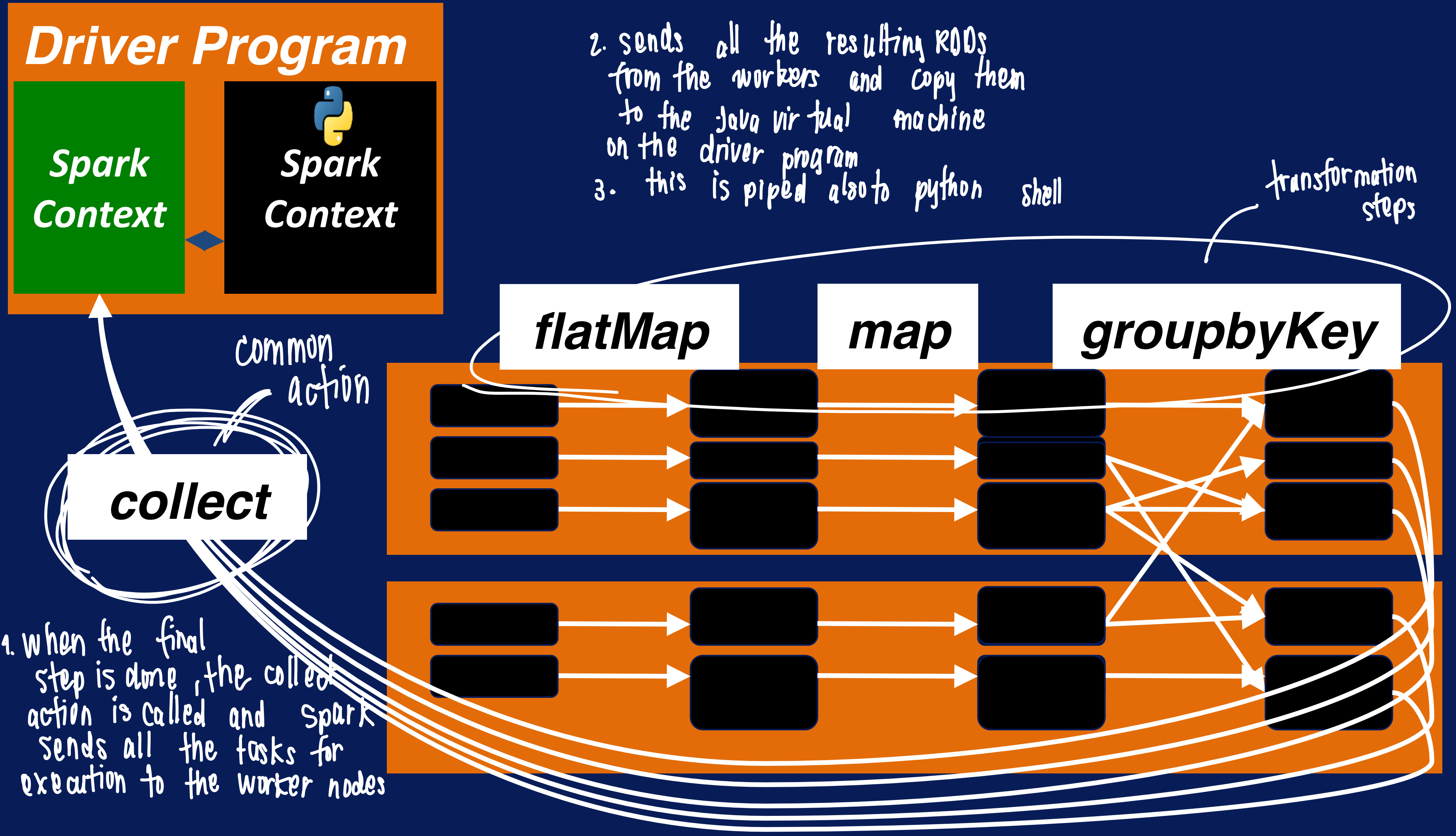
: RDD operations that trigger the evaluation of the transformation pipeline and return the final result to the driver program or save the results to a persistent storage

* "the last step in a spark pipeline"

transformations are evaluated after an action is performed

After this video you will be able to..

- Explain the steps of a Spark pipeline ending with a collect action
- List four common action operations in Spark



transformation steps

common action

1. when the final step is done, the collect action is called and Spark sends all the tasks for execution to the worker nodes

2. sends all the resulting RDDs from the workers and copy them to the Java vir tual machine on the driver program

3. this is piped also to python shell

Some Common Actions

Action	Usage
collect()	Copy all elements to the driver
take(n)	Copy first n elements
reduce(func) * most famous one	Aggregate elements with func (takes 2 elements, returns 1)
[redacted] save As Text File	Save to local file or HDFS

if results are too large to fit in the driver memory, then there's an opportunity to write them directly

no key, just a large area of some values

- running over and over to reduce werything to one single value

save the results to local disk or HDFs

* useful if output is large