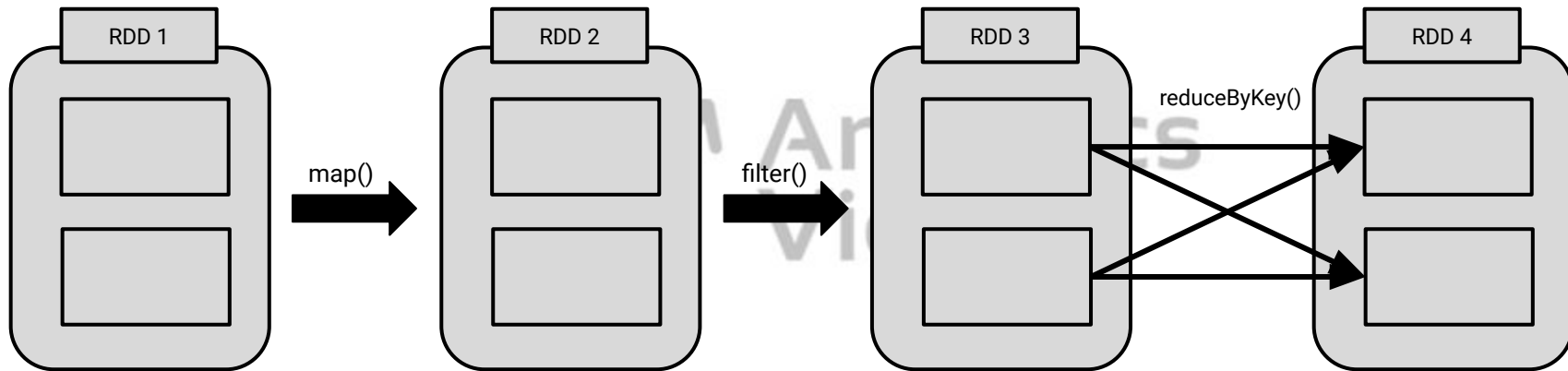
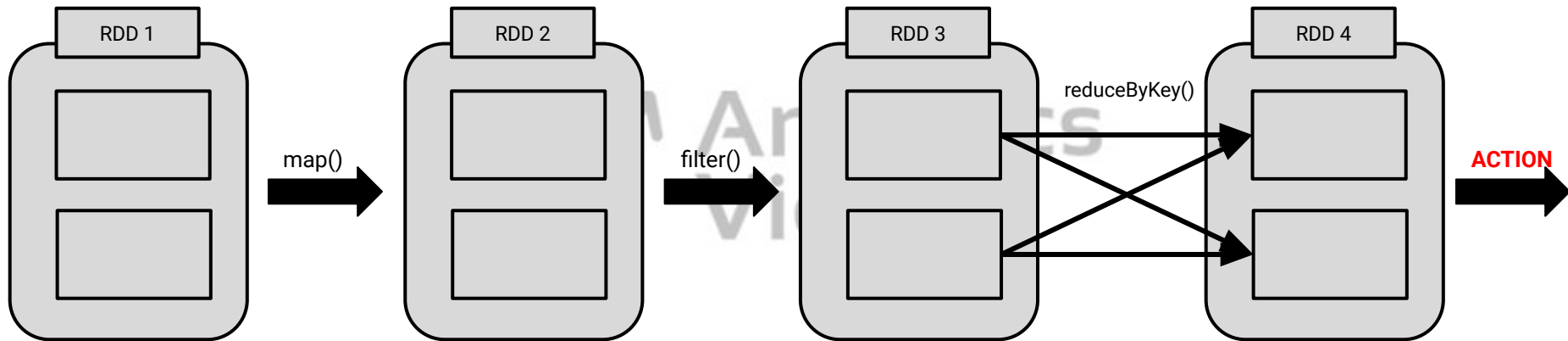


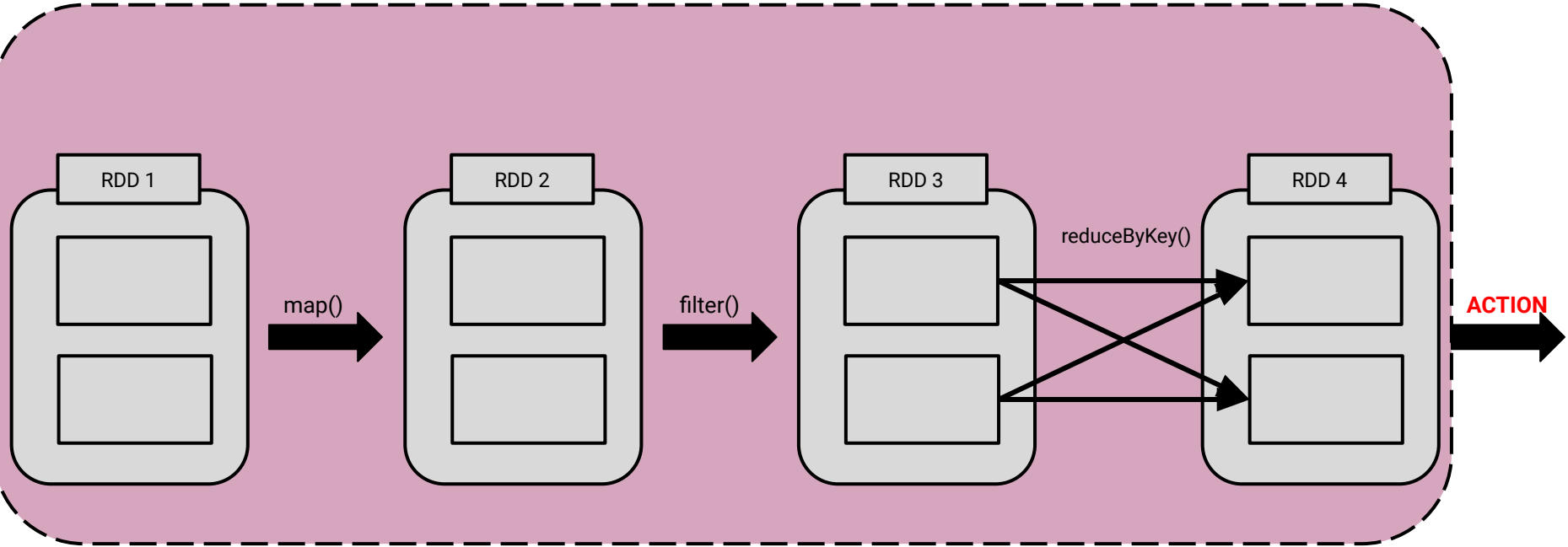
# Spark Execution

# Jobs, Stages, and Tasks

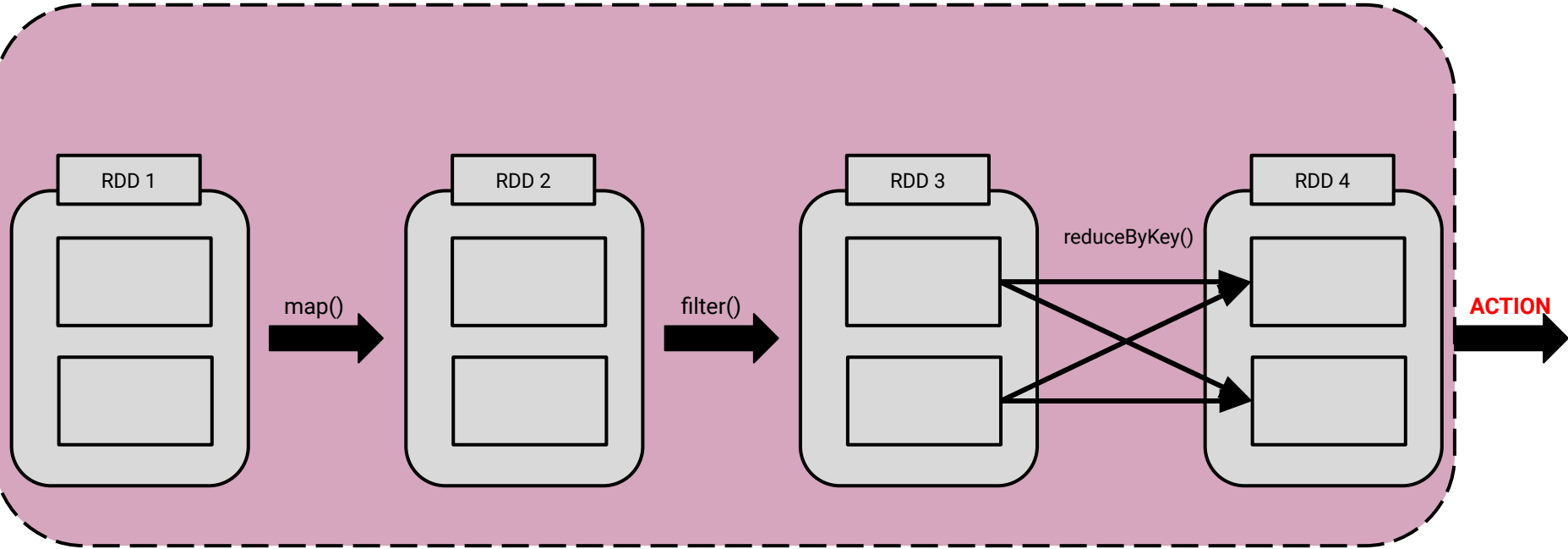




Job

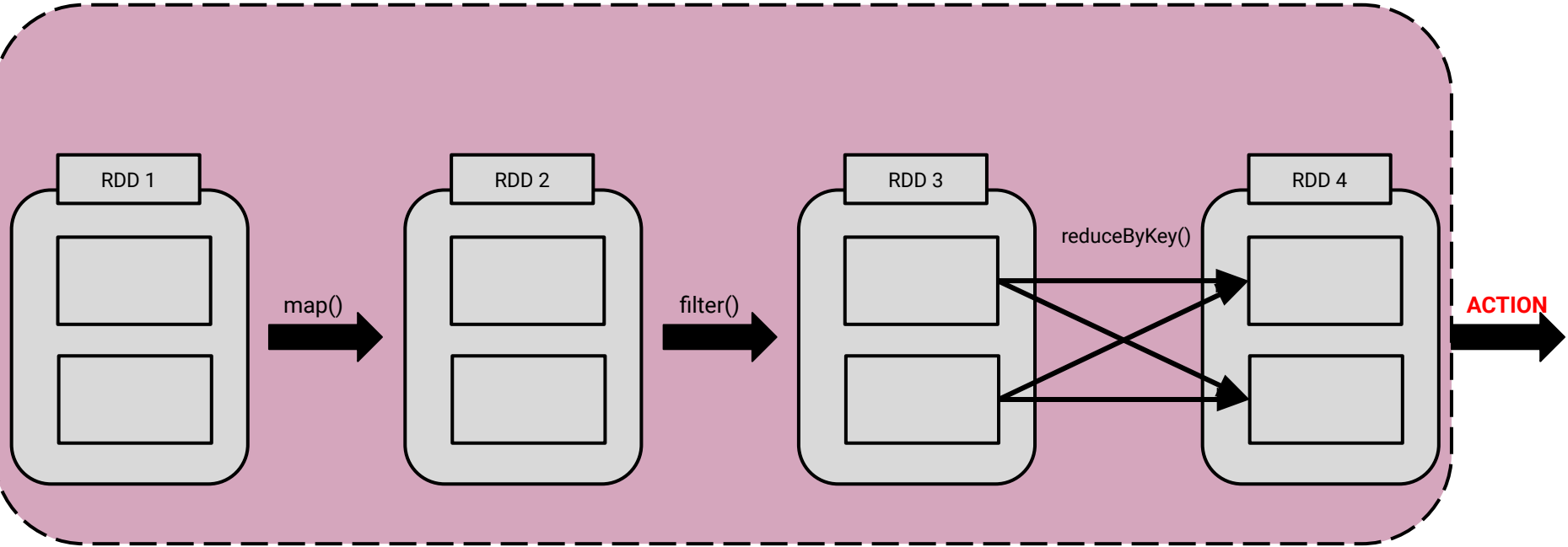


## Job



**Job** is a series of stages combined together.

Job



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

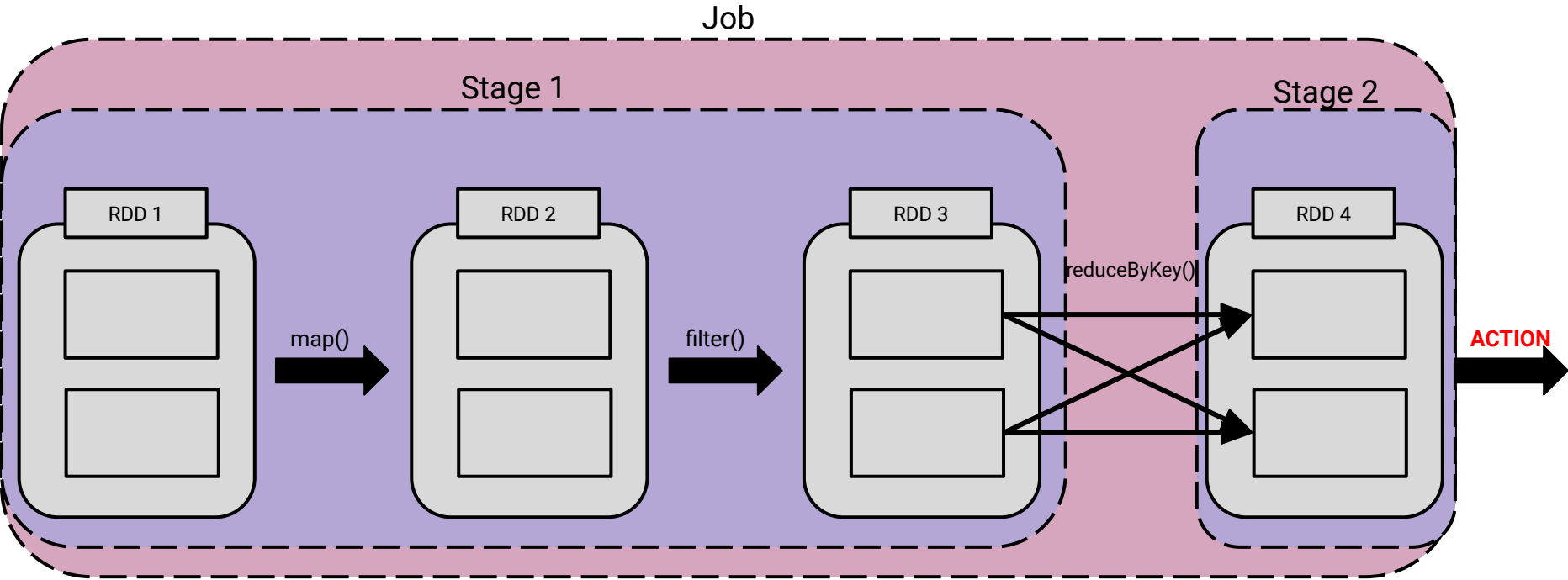
RDD 4

map()

filter()

reduceByKey()

ACTION





Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

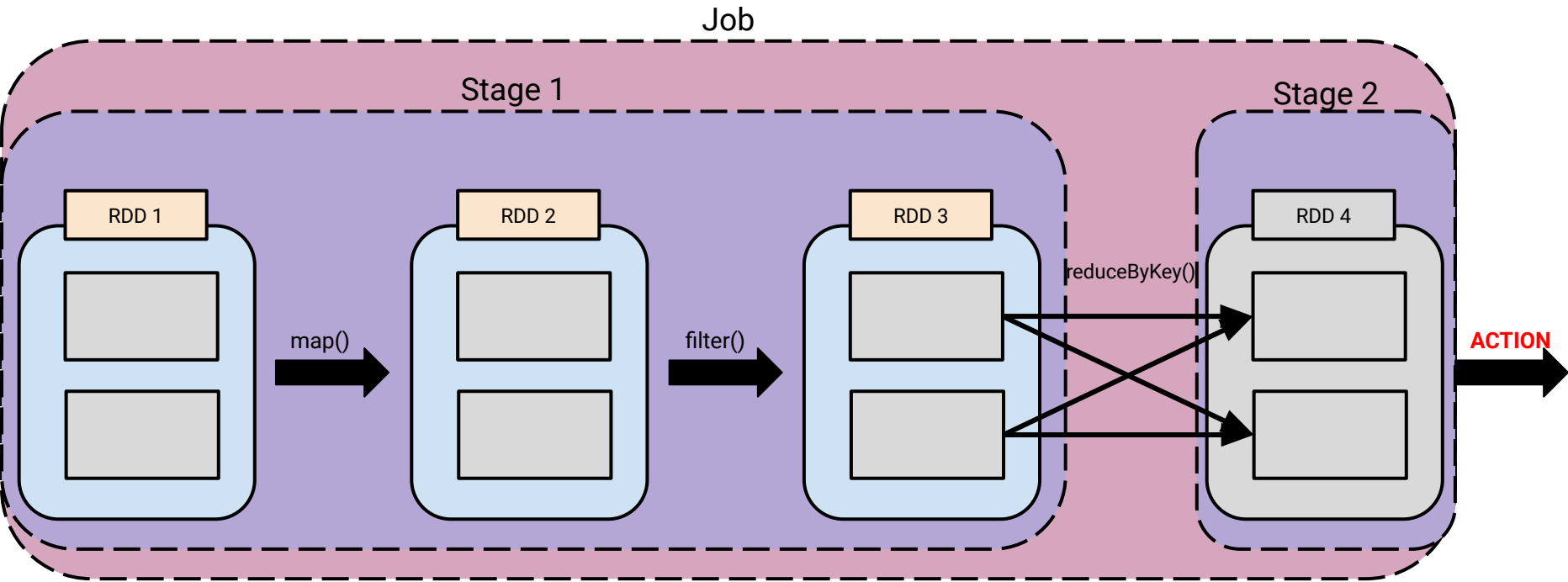
RDD 4

map()

filter()

reduceByKey()

ACTION



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

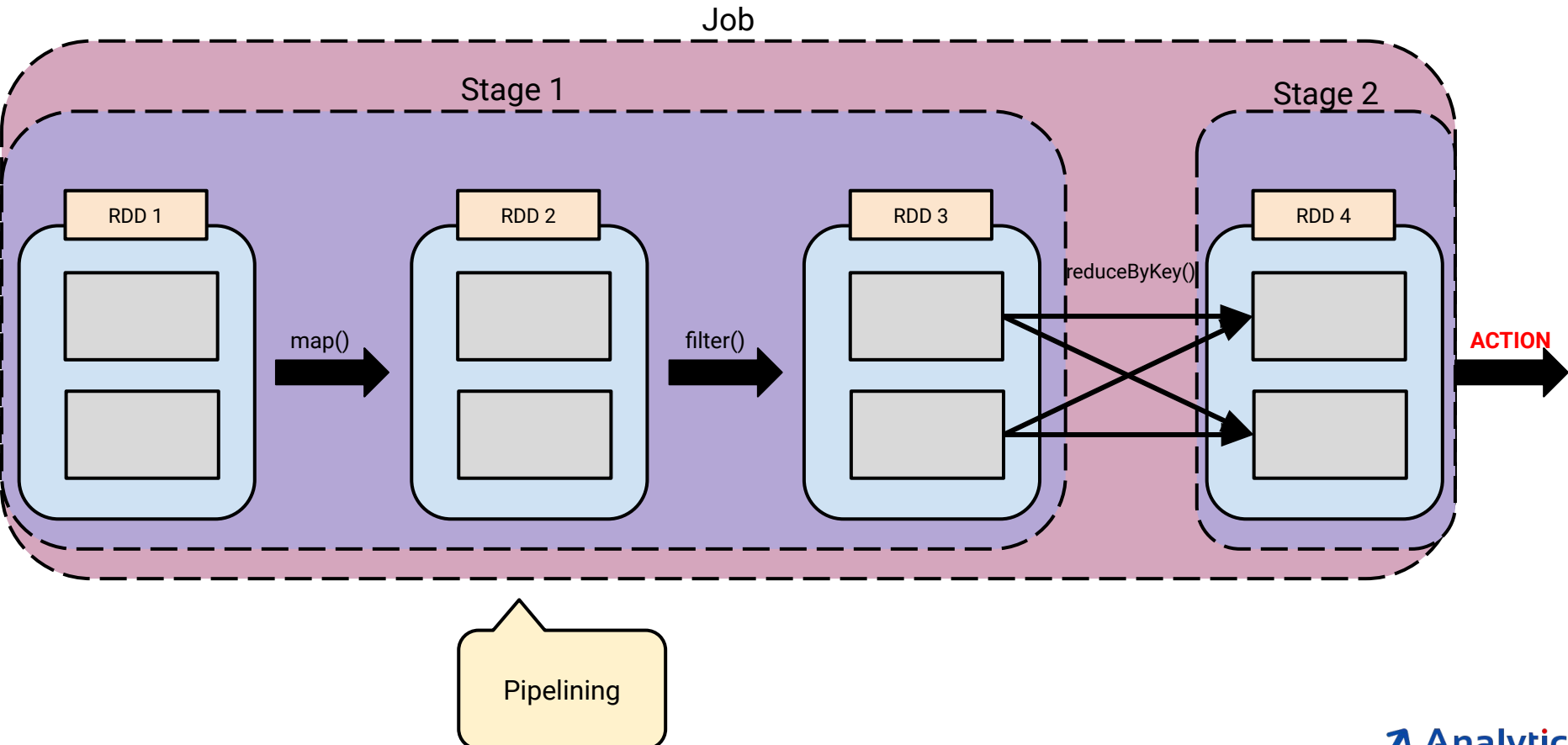
map()

filter()

reduceByKey()

ACTION

Pipelining



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

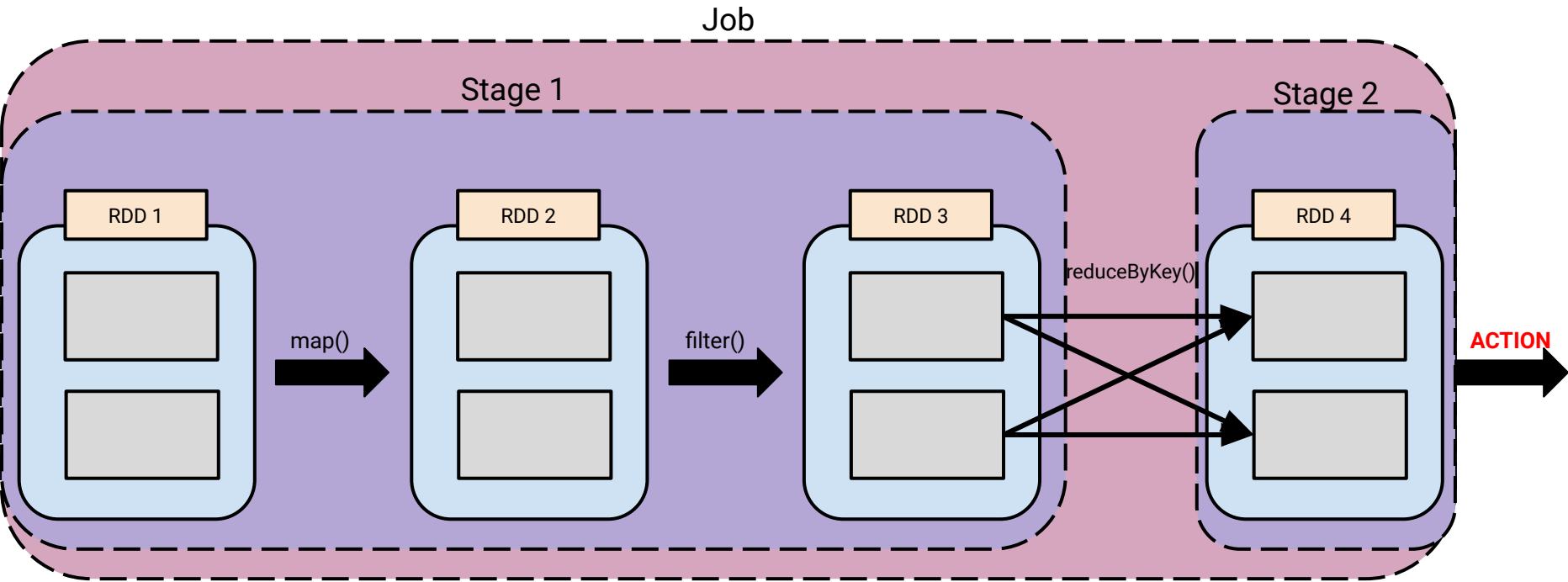
RDD 4

map()

filter()

reduceByKey()

ACTION



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

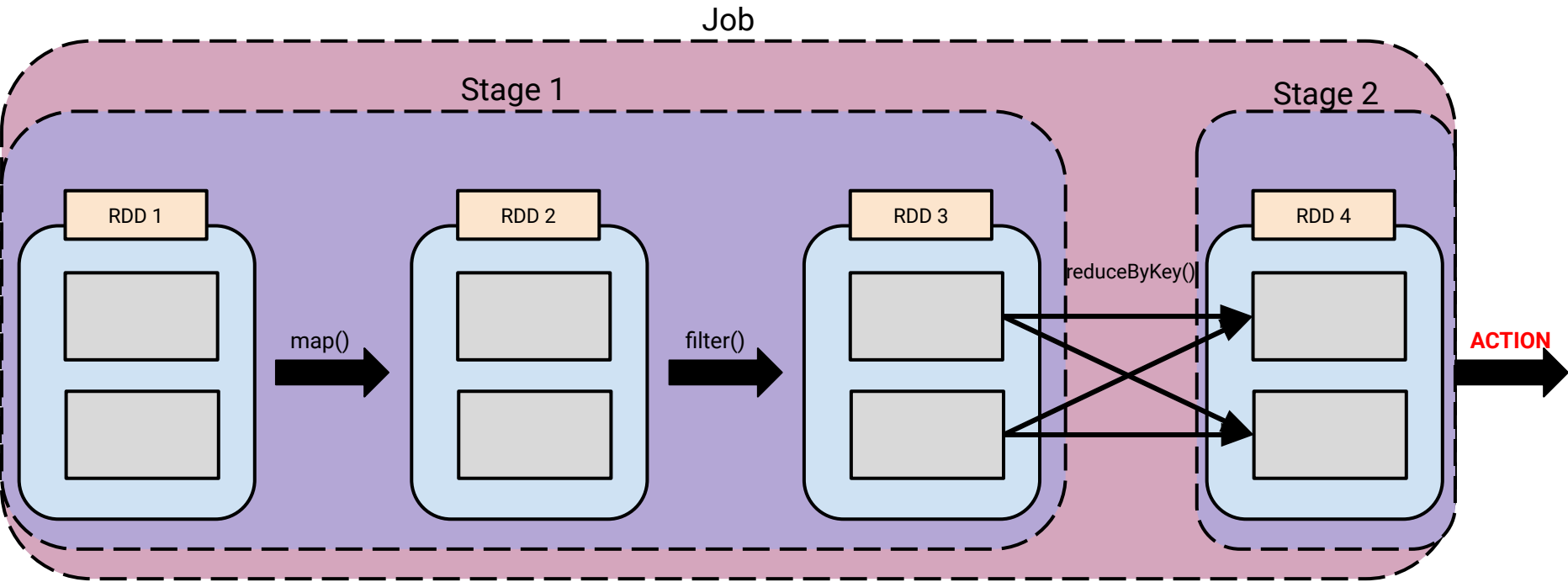
map()

filter()

reduceByKey()

ACTION

**Stage** is a group of tasks executed together.



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

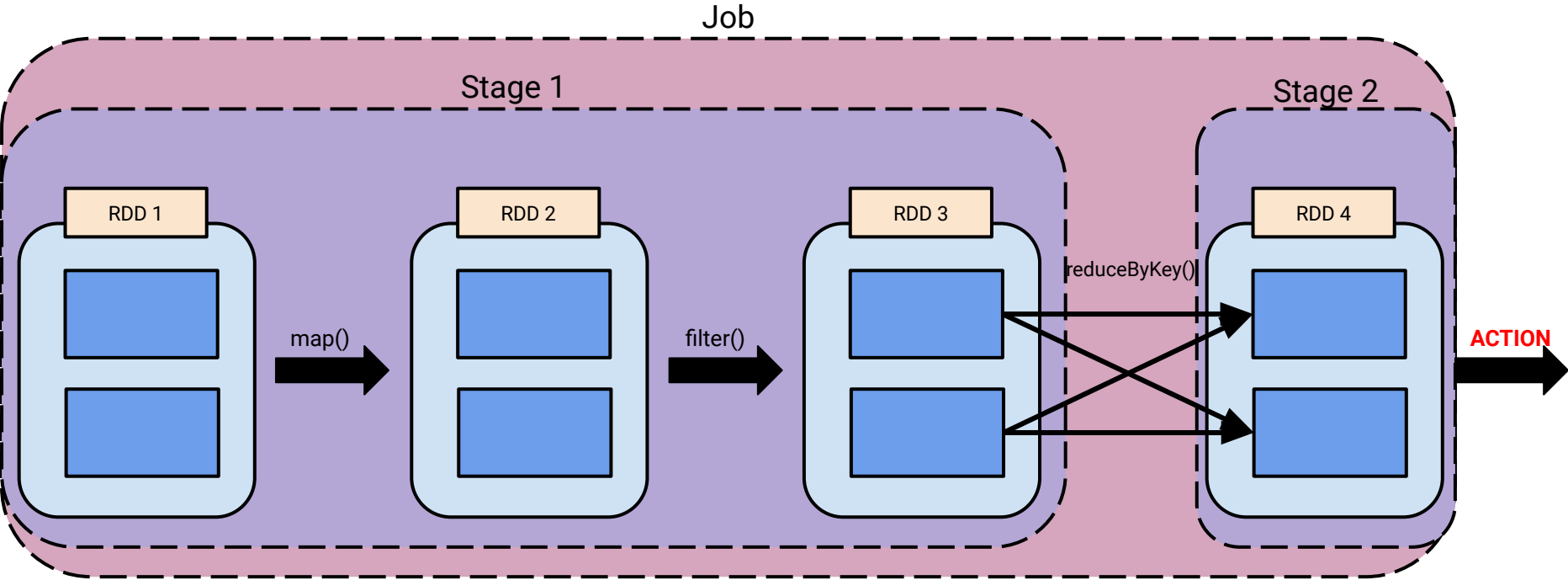
RDD 4

map()

filter()

reduceByKey()

ACTION



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

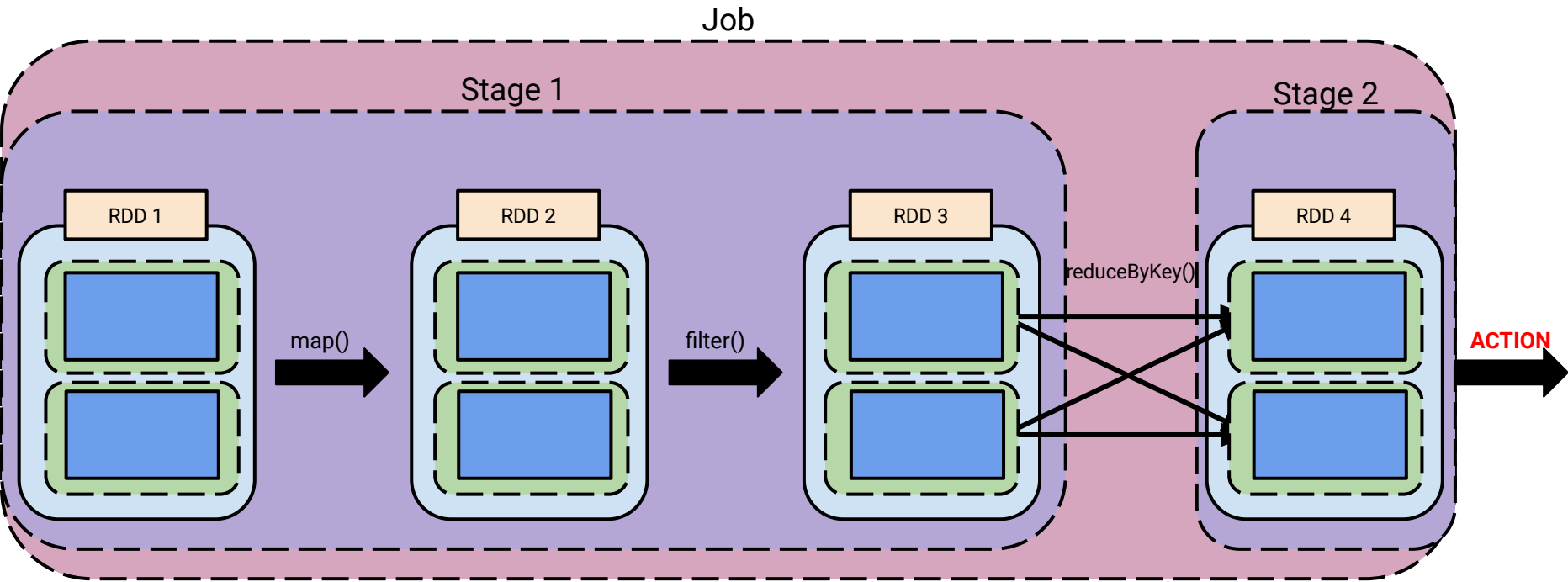
map()

filter()

reduceByKey()

ACTION

**Task** is a single unit of computation performed on a single partition of data



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

map()

filter()

reduceByKey()

ACTION

**Task** is a single unit of computation performed on a single partition of data

It is performed on a single core of Worker node.

Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

map()

filter()

reduceByKey()

ACTION



Job

Stage 1

Stage 2

RDD 1

RDD 2

RDD 3

RDD 4

map()

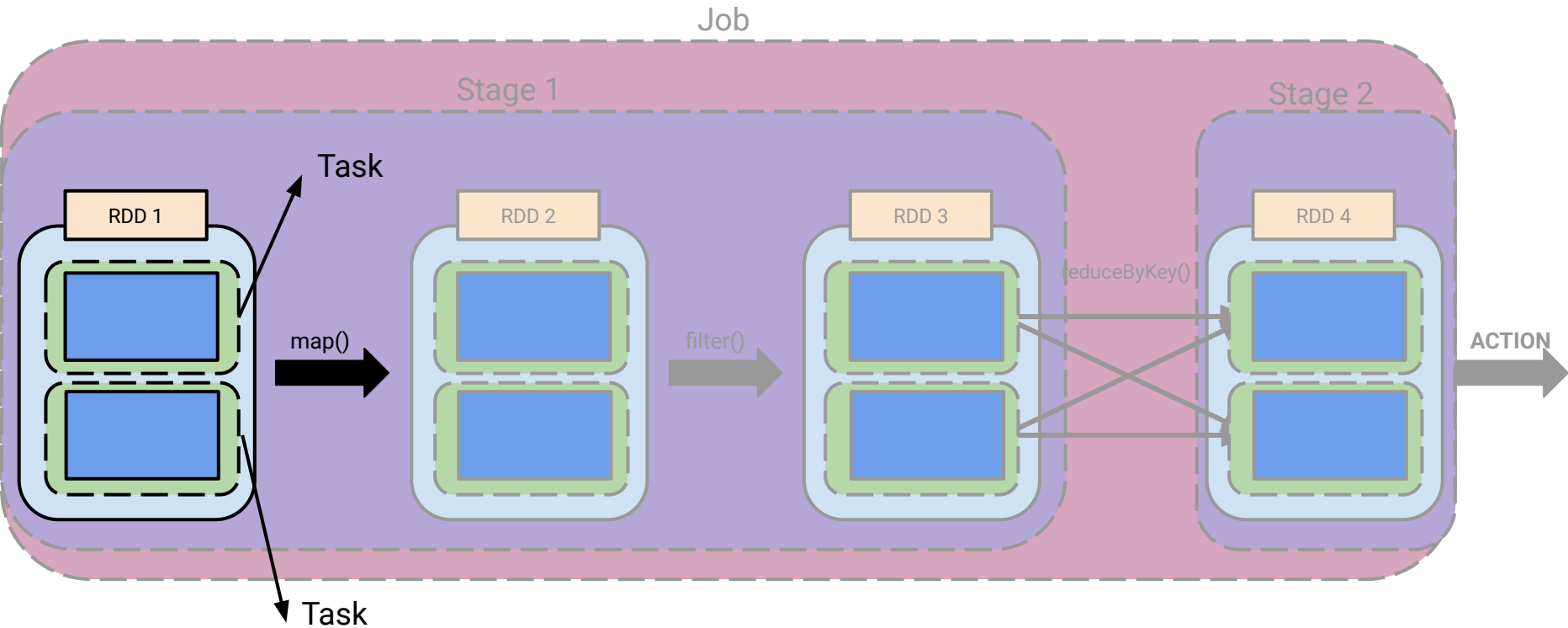
filter()

reduceByKey()

ACTION

Task

Task



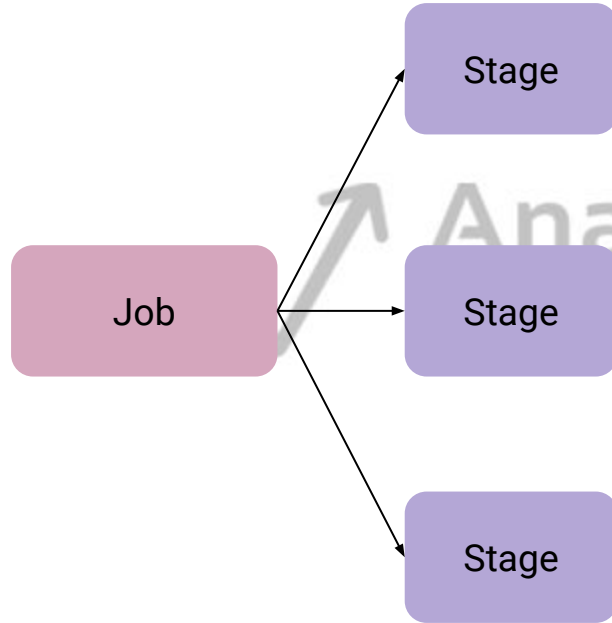
# Recap - Jobs, Stages, and Tasks



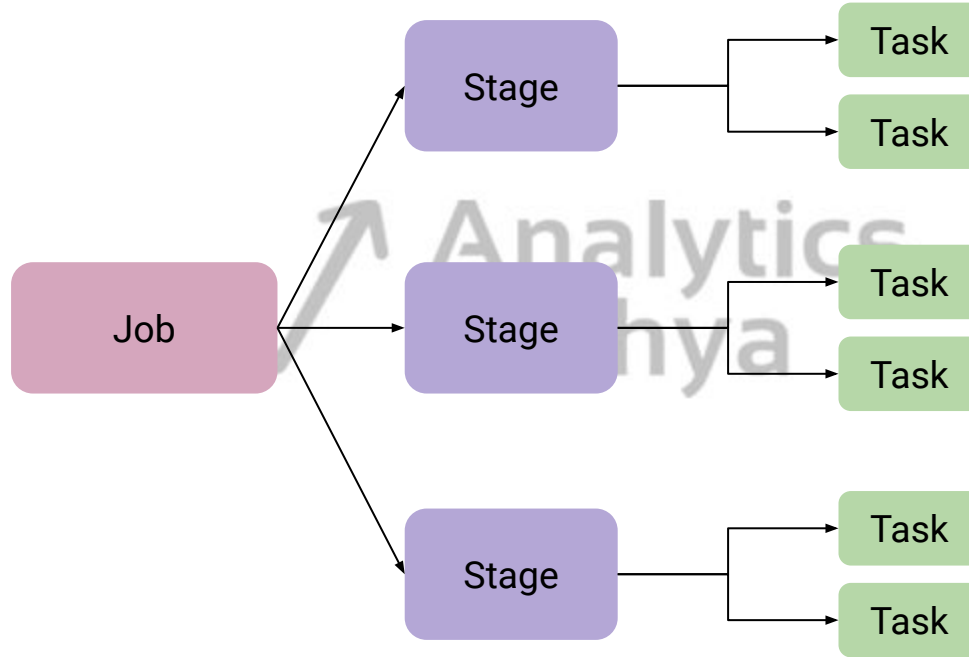
# Recap - Jobs, Stages, and Tasks



# Recap - Jobs, Stages, and Tasks



# Recap - Jobs, Stages, and Tasks





Thank You!!