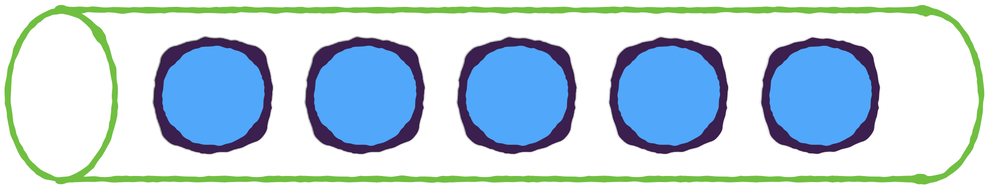
Data Representation

Unbounded stream
of entities



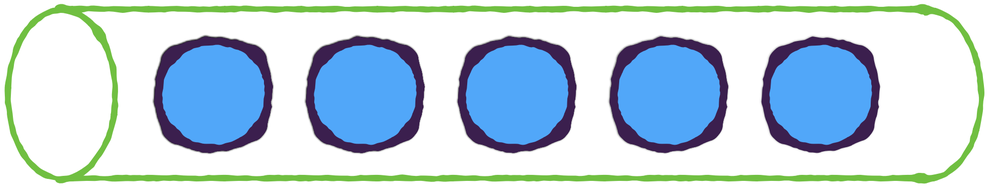
DataStream

Bounded collection
of entities

DataSet

Data Representation

DataStream



DataSet

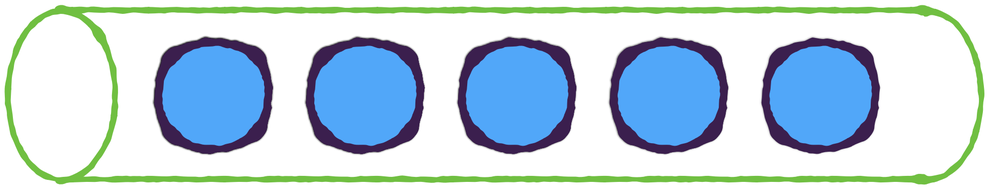
Immutable collections

Can be transformed but

not mutated

Data Representation

DataStream

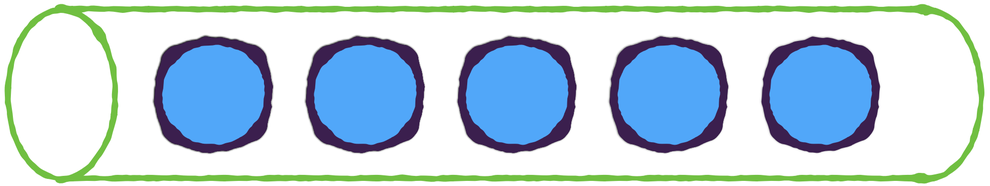


DataSet

Contain duplicates

Data Representation

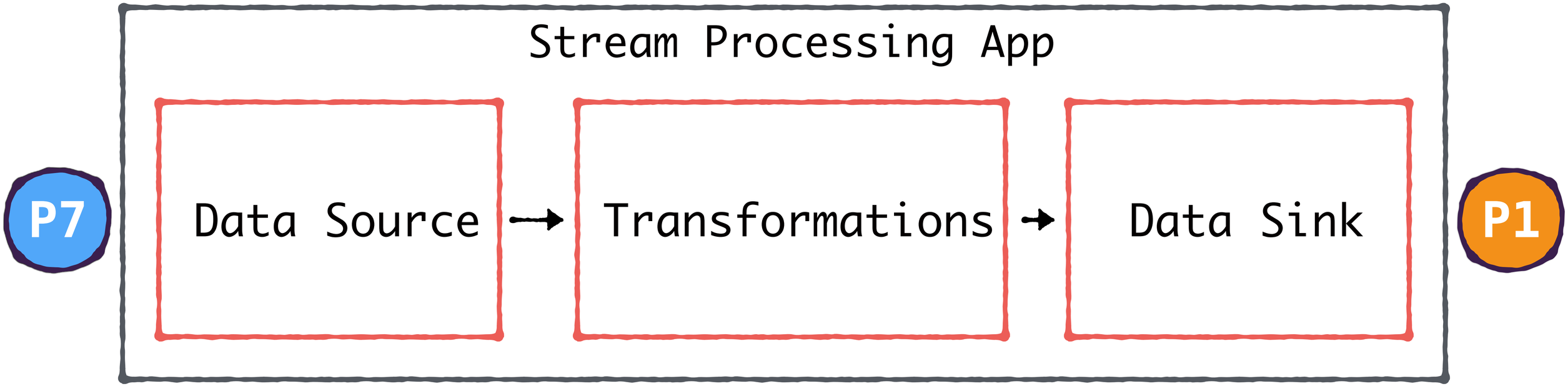
DataStream



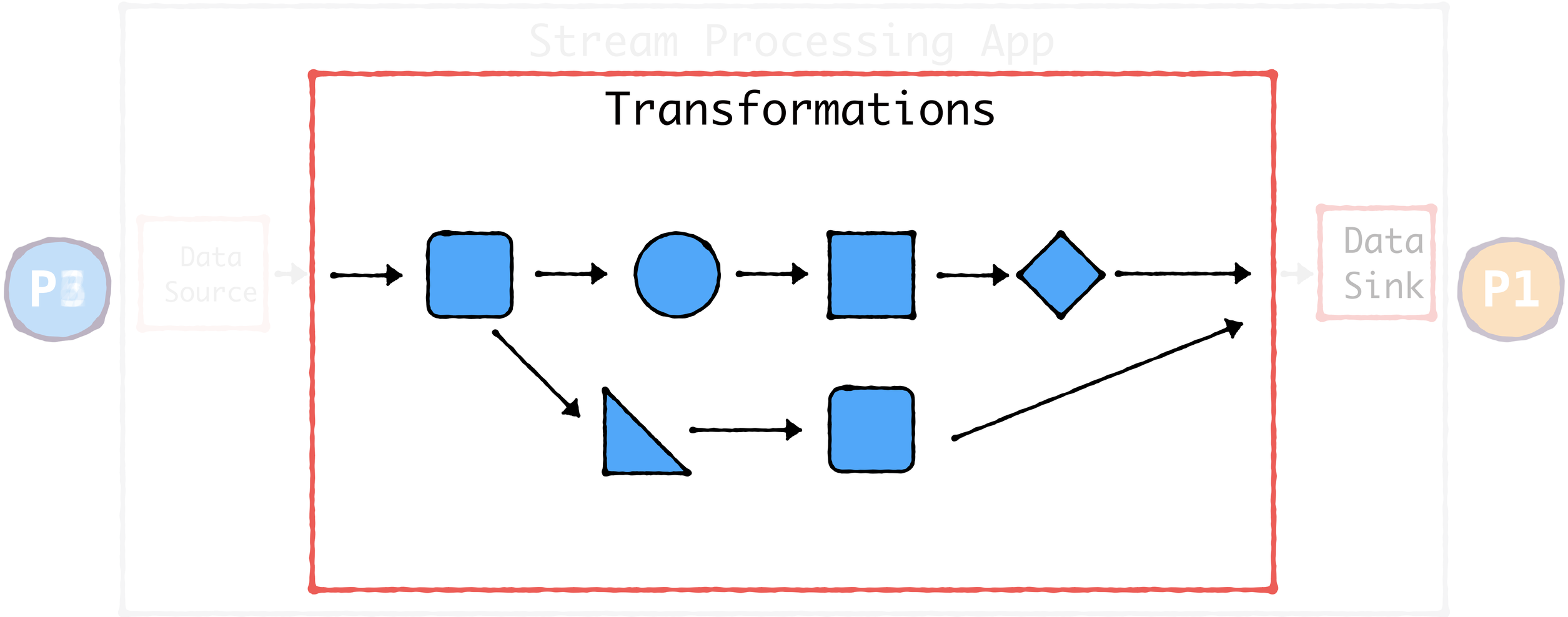
DataSet

Entities of one of the supported data types

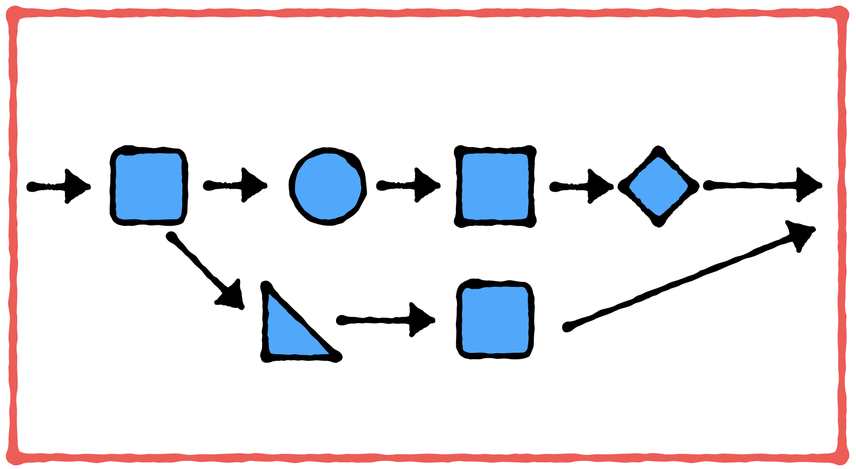
Flink Programming Model



Flink Programming Model



Flink Transformations



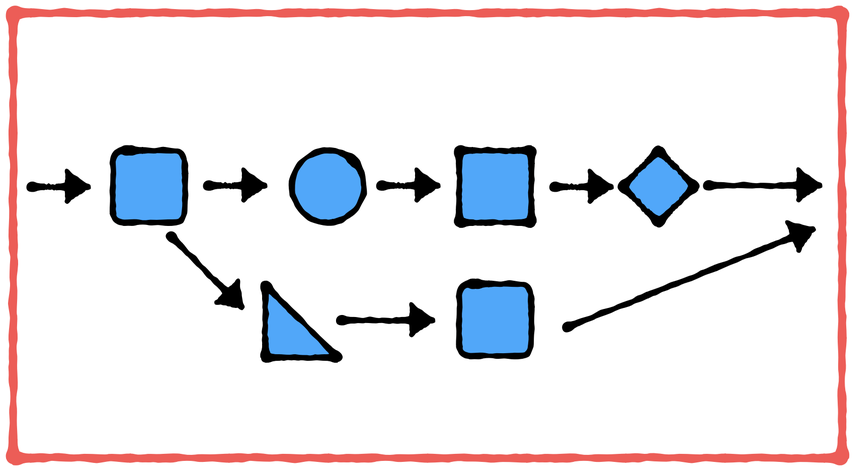
filter

map

flatMap

Act on one entity at a time

Flink Transformations

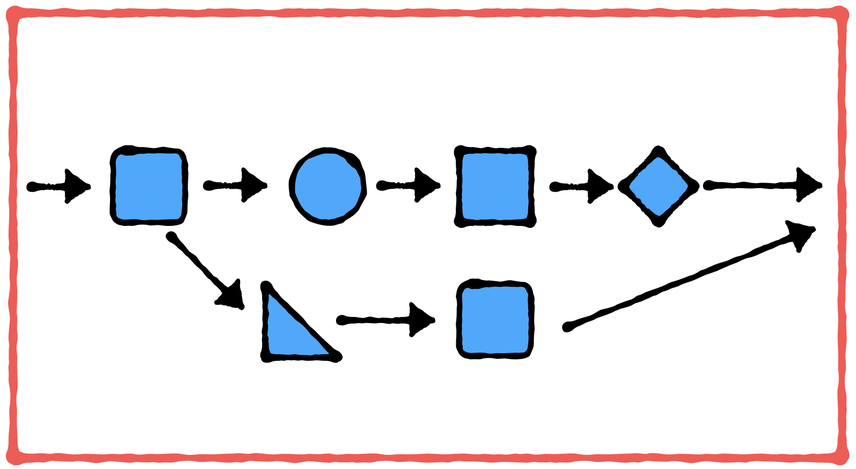


reduce

sum

Combine/aggregate
multiple entities

Flink Transformations

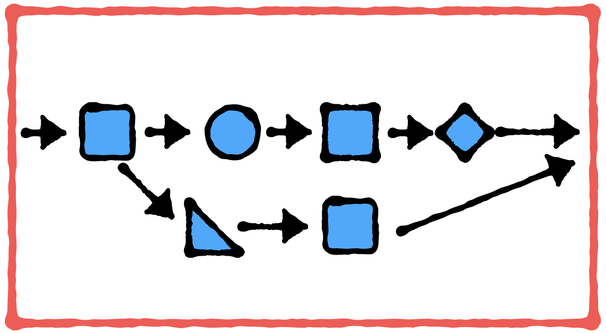


keyBy

groupBy

Group data based
on a key(s)

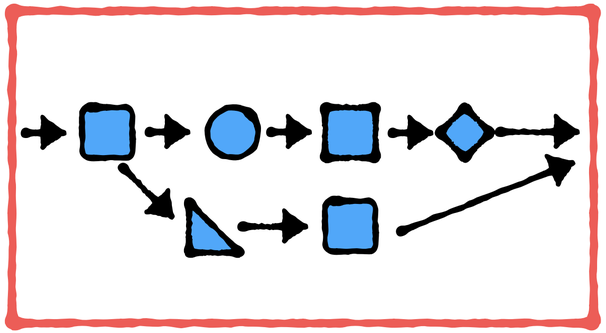
Lazy Evaluation



User expresses a chain of operations

Operations are added to a plan

Lazy Evaluation

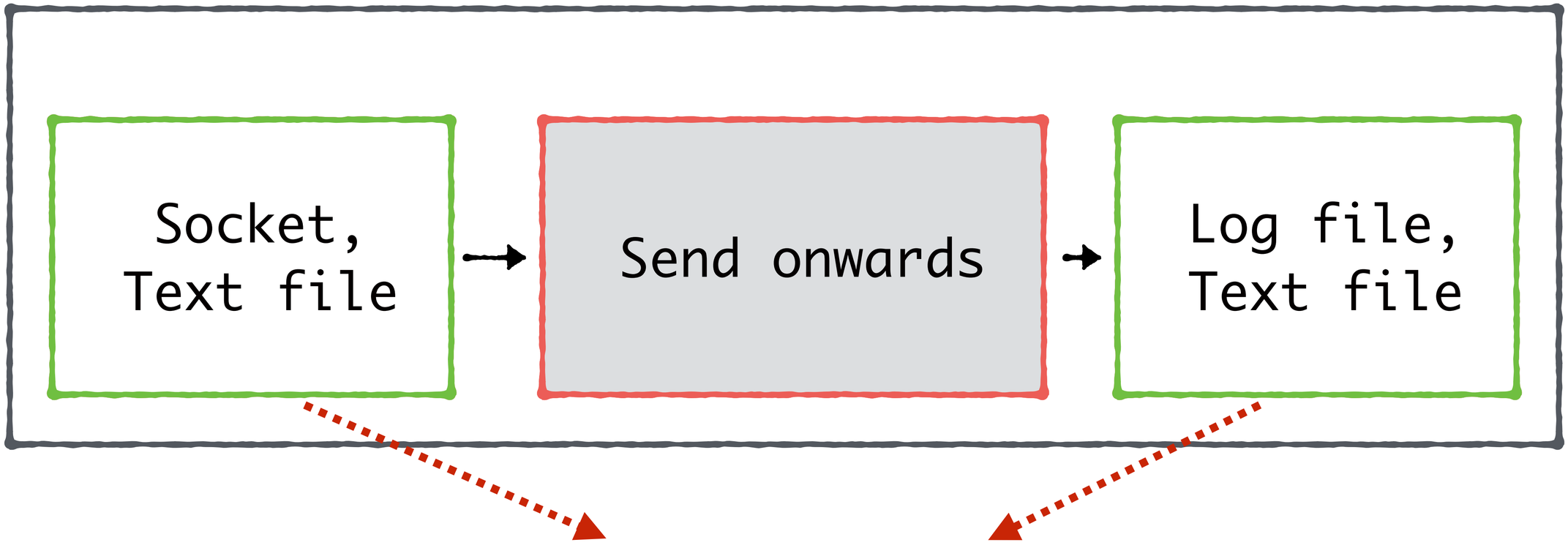


User expresses a chain of operations

Operations are added to a plan

Executed when **execute()** is called

Hello World



Built-in sources, sinks

Flink Program

Set up execution environment

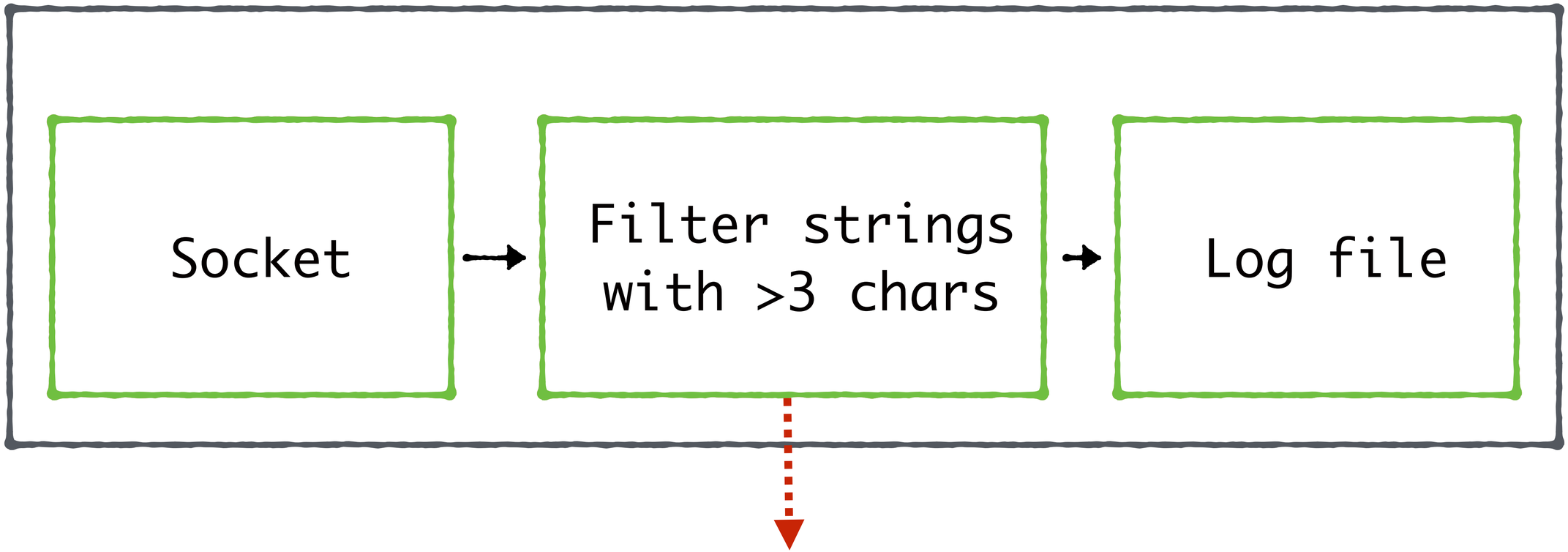
Express data transformations

Trigger execution

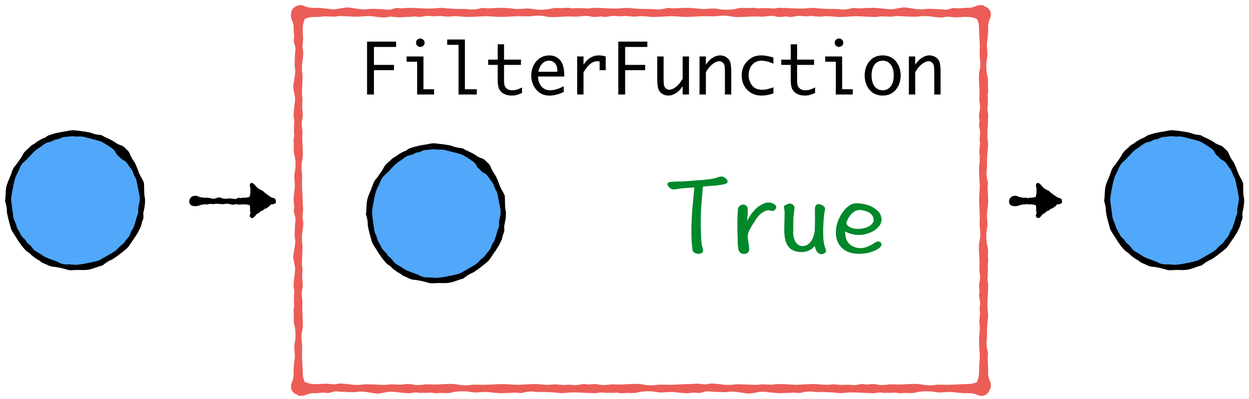
Load/create data from source(s)

Write data to sink(s)

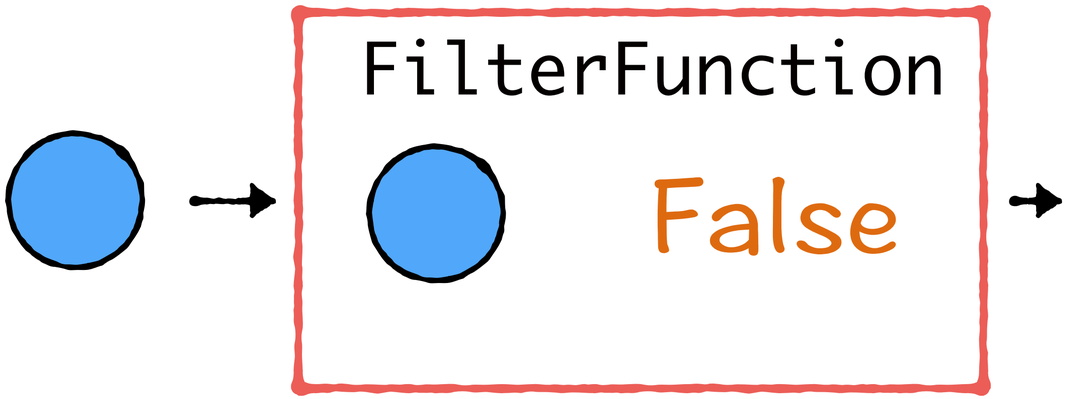
Filter


Socket
Filter strings with >3 chars
Log file
Implement a FilterFunction

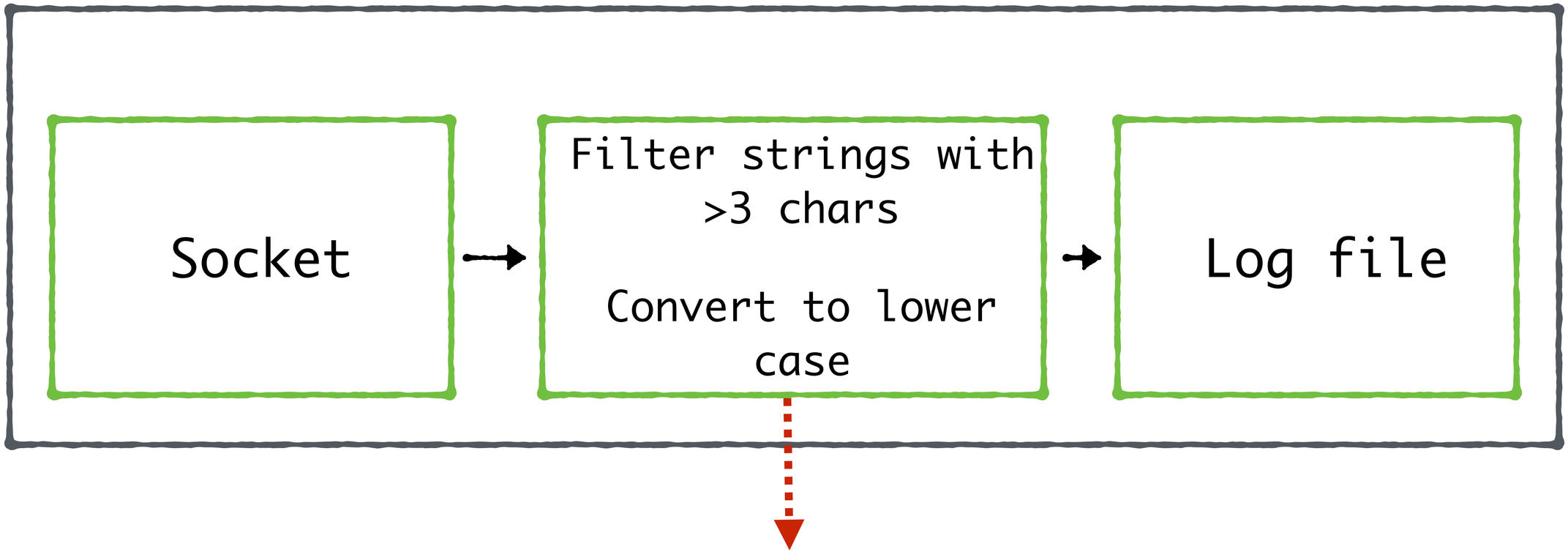
FilterFunction



FilterFunction



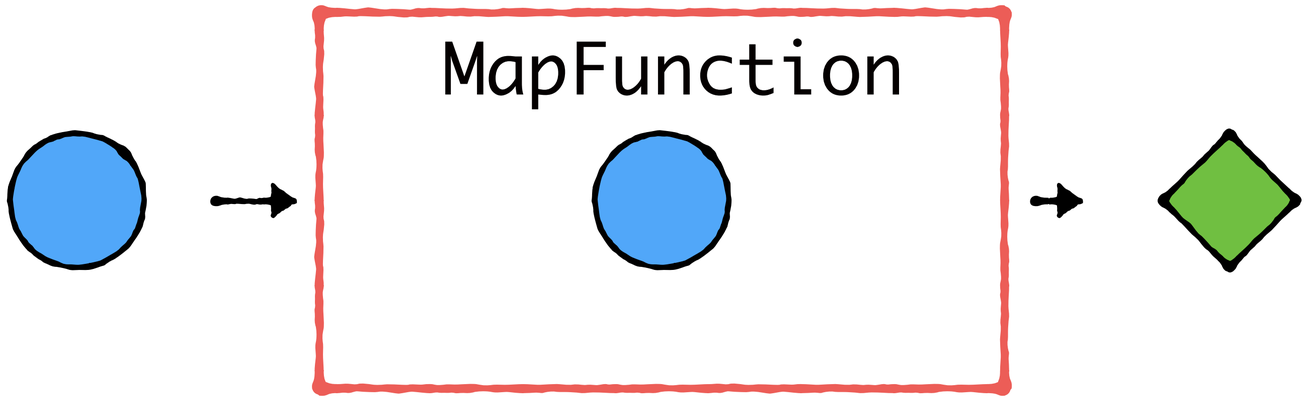
Map



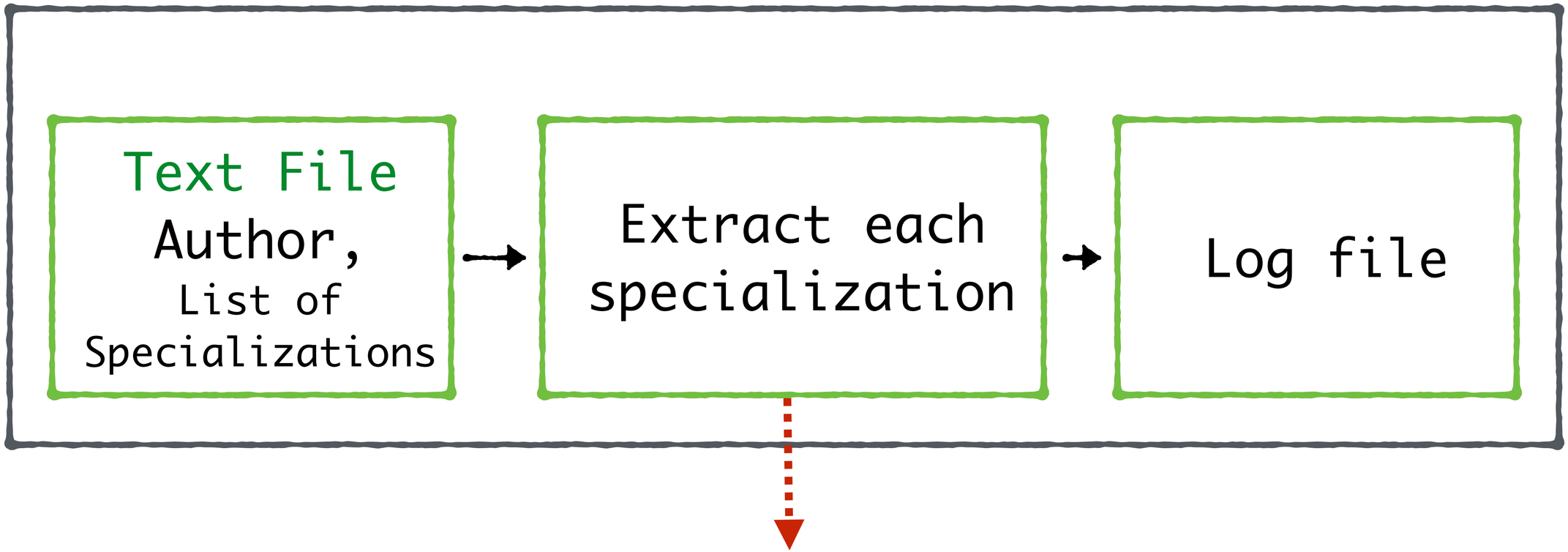
Chain 2 operations

Implement a MapFunction

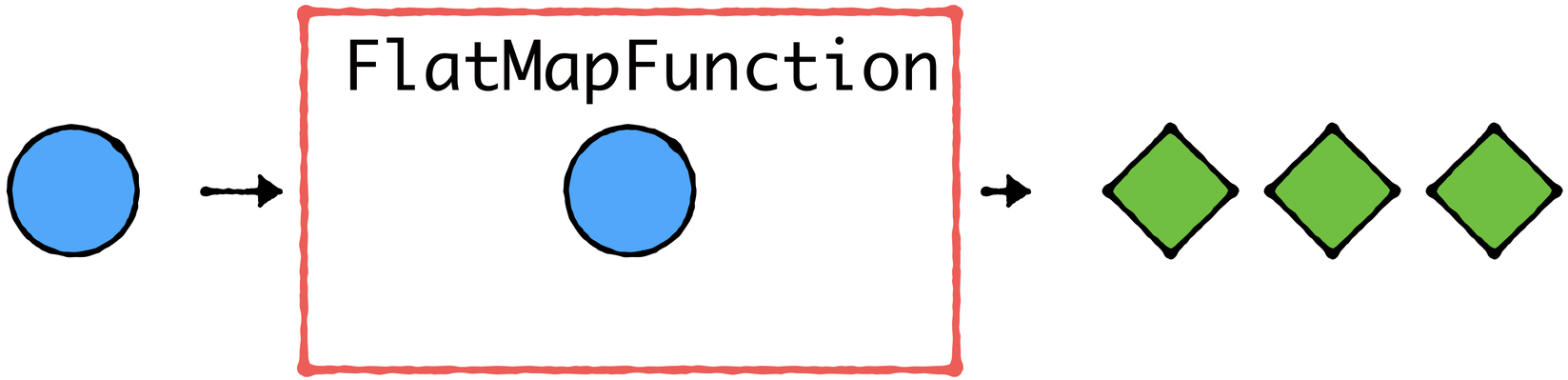
MapFunction



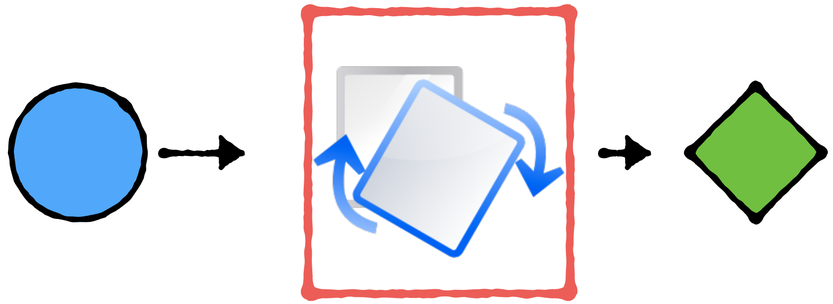
flatMap


Text File
Author,
List of
Specializations
Extract each
specialization
Log file
Implement a FlatMapFunction

FlatMapFunction

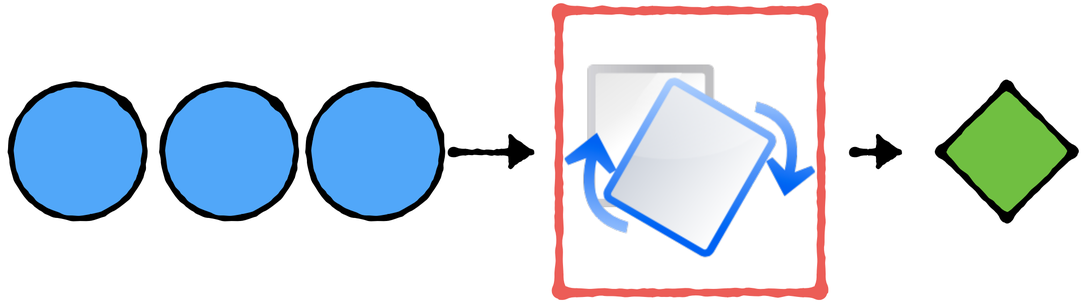


Transformation applied on single stream entity

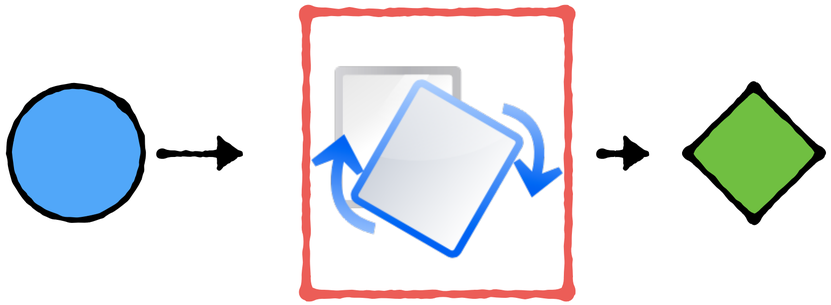


Stateless Transformation

Transformation applied across multiple stream entities



Stateful Transformation



Stateless Transformation

filter

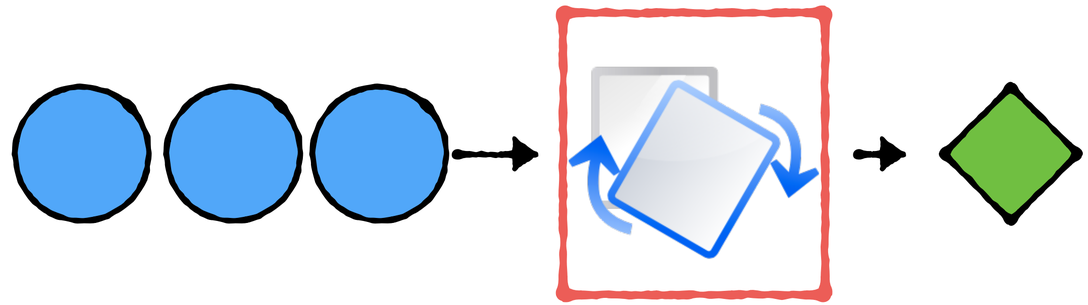
map

flatMap

reduce

sum

keyby



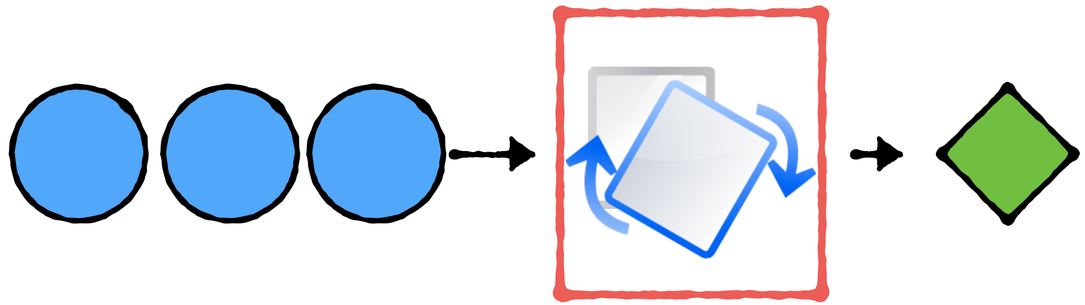
Stateful Transformation

Accumulate data

entire stream

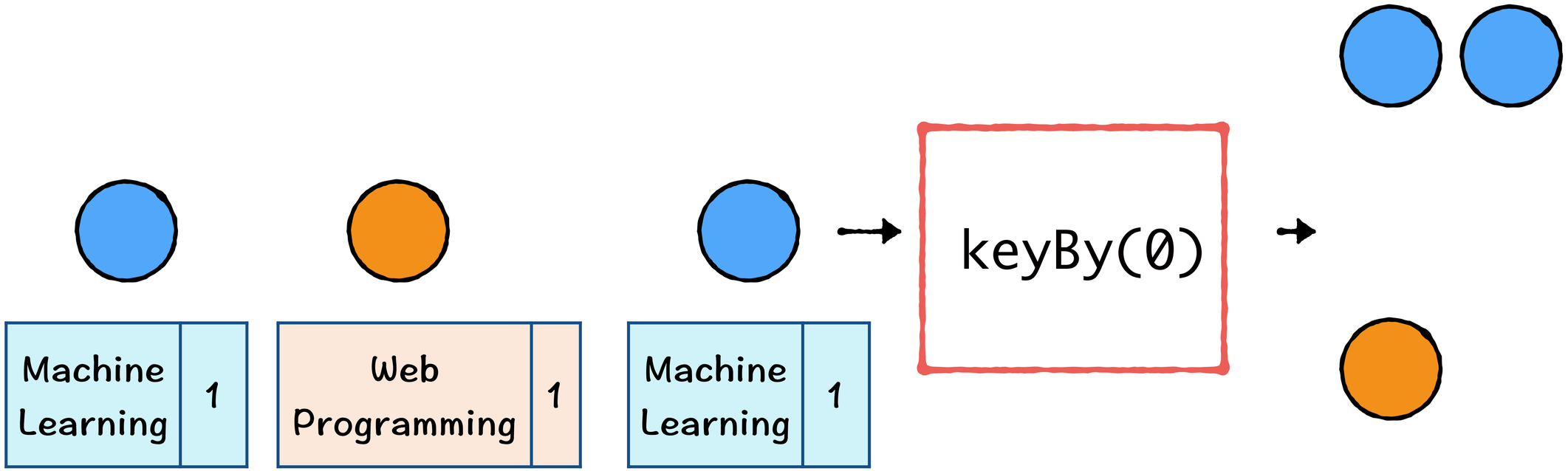
window

per key, per operator



Stateful Transformation

Keyed Streams



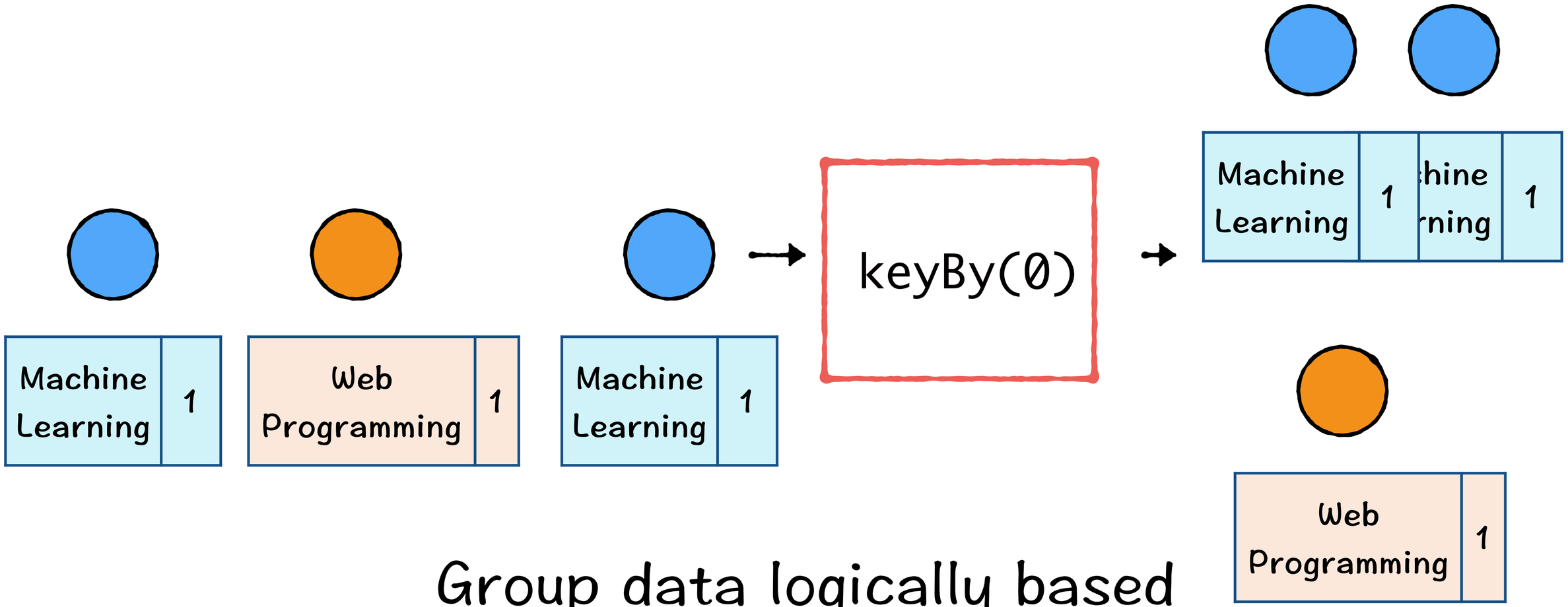
Group data logically based on a key

Keyed Streams

Group data logically based
on a key

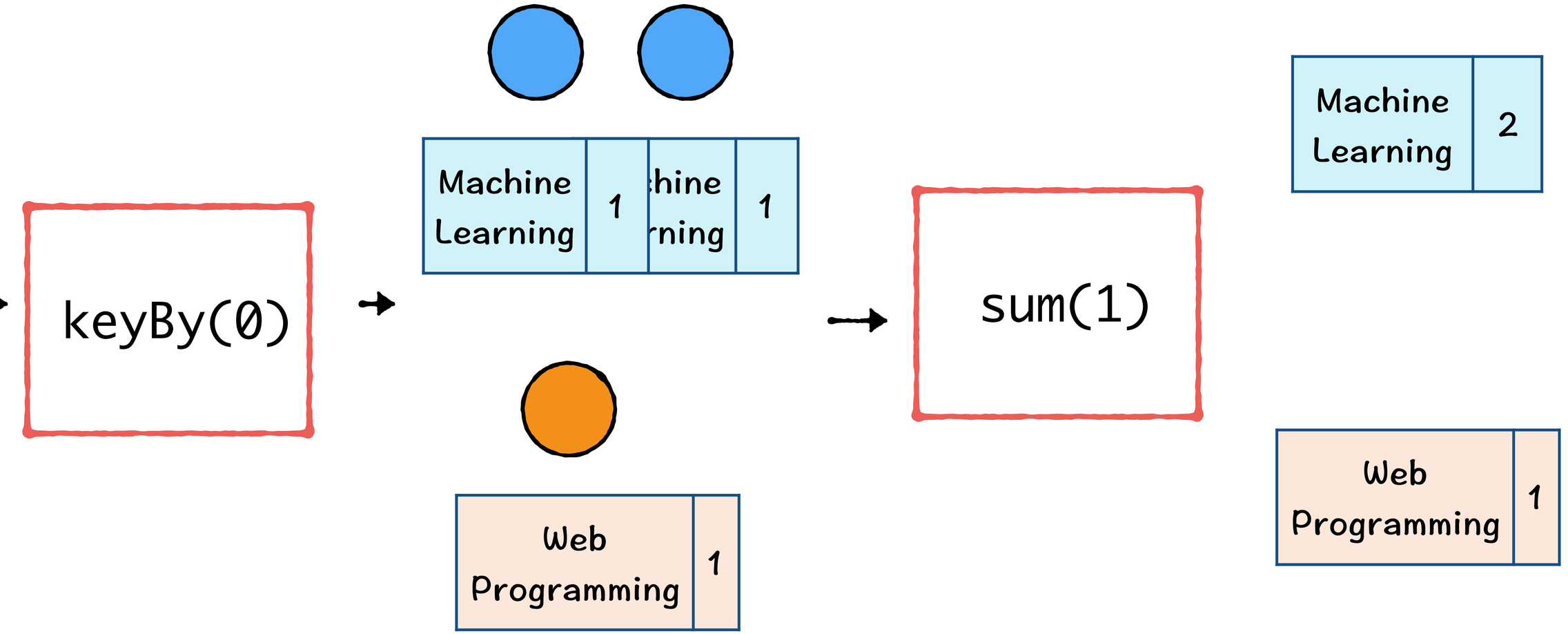
Keys in Flink are virtual,
not physical

Transformations on Keyed Streams



Group data logically based on a key

Transformations on Keyed Streams



Operation is applied on each group

Number Aggregations

Built-in aggregations for keyed streams

sum

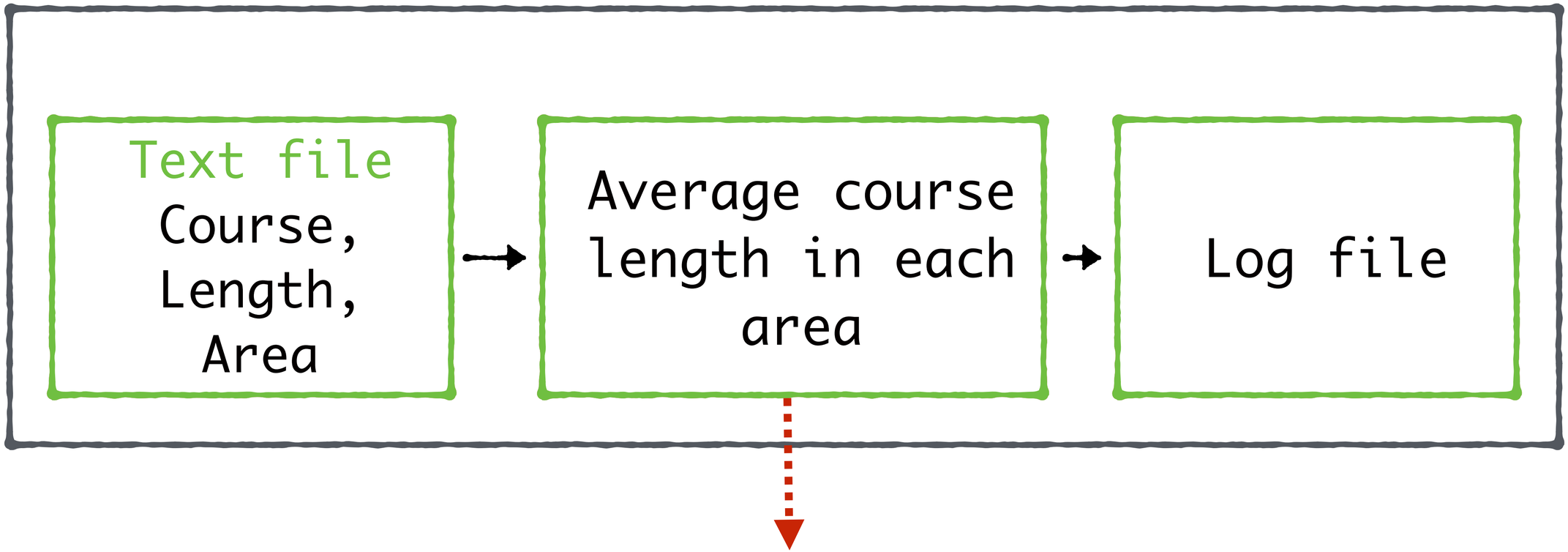
min

max

minby

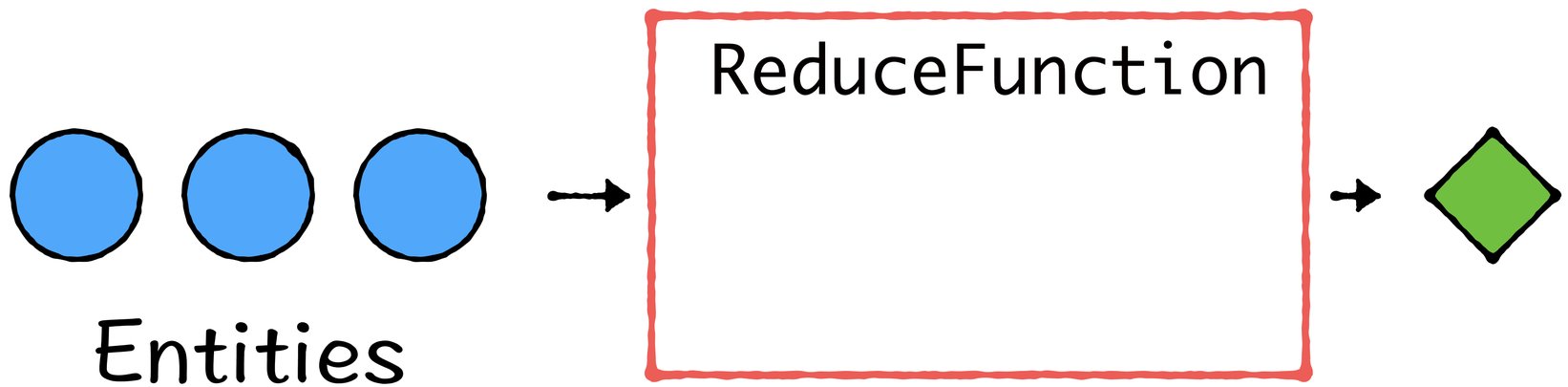
maxby

Reduce


Text file
Course,
Length,
Area
Average course length in each area
Log file
Key stream by area
Implement a ReduceFunction

ReduceFunction

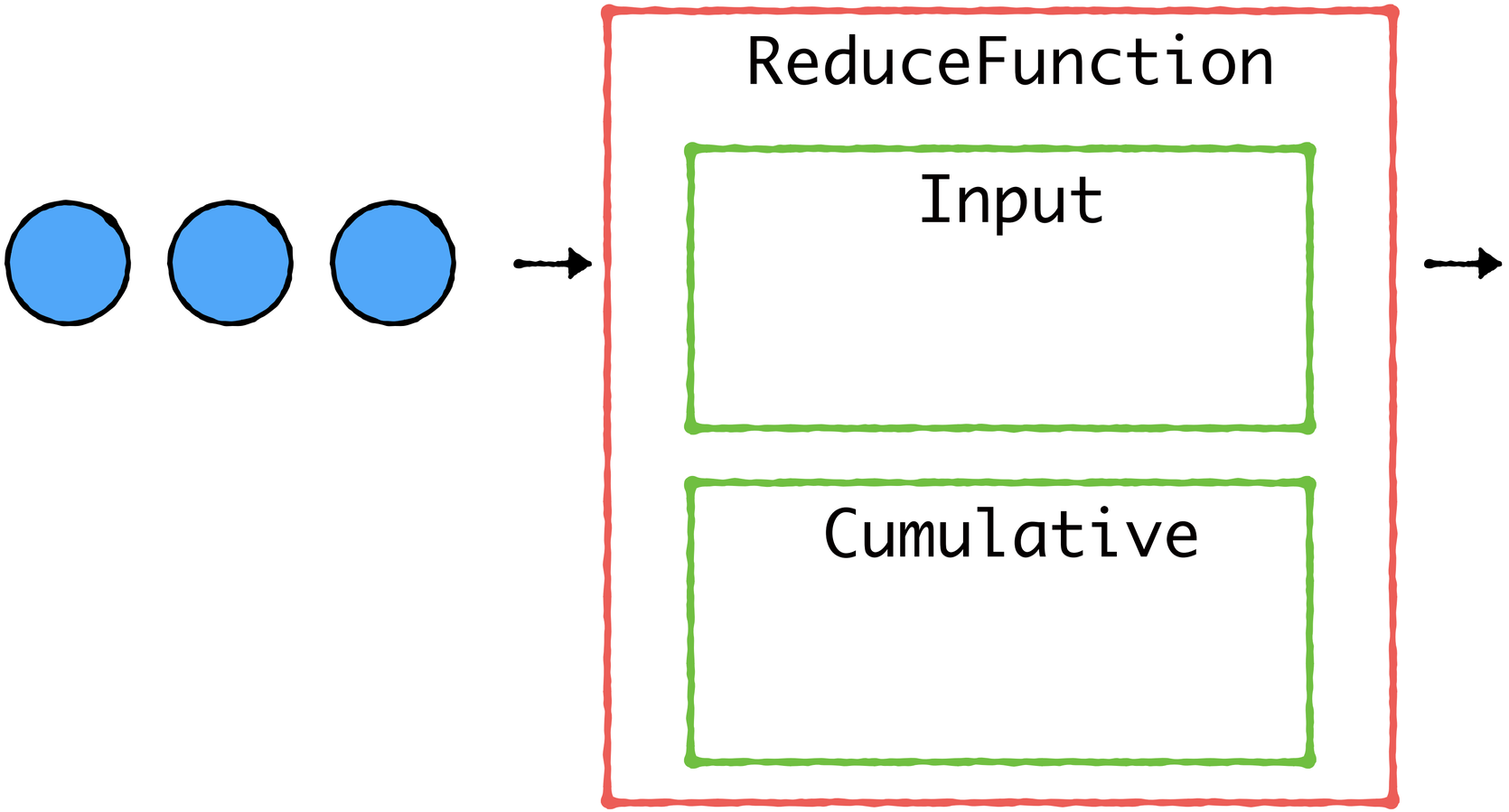
Applies on a KeyedStream



Entities belonging to one group

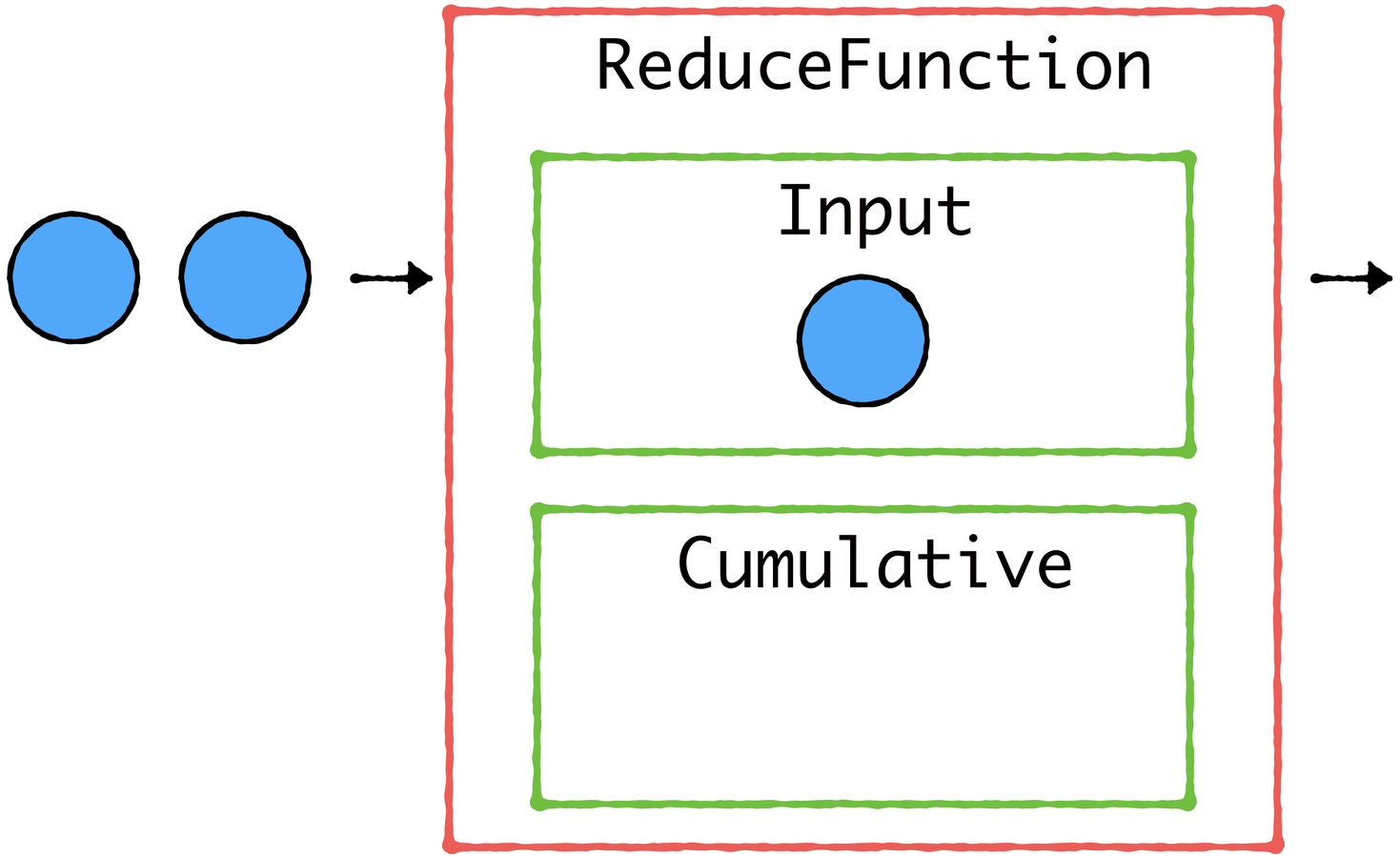
ReduceFunction

Reduce function combines 2 values at a time



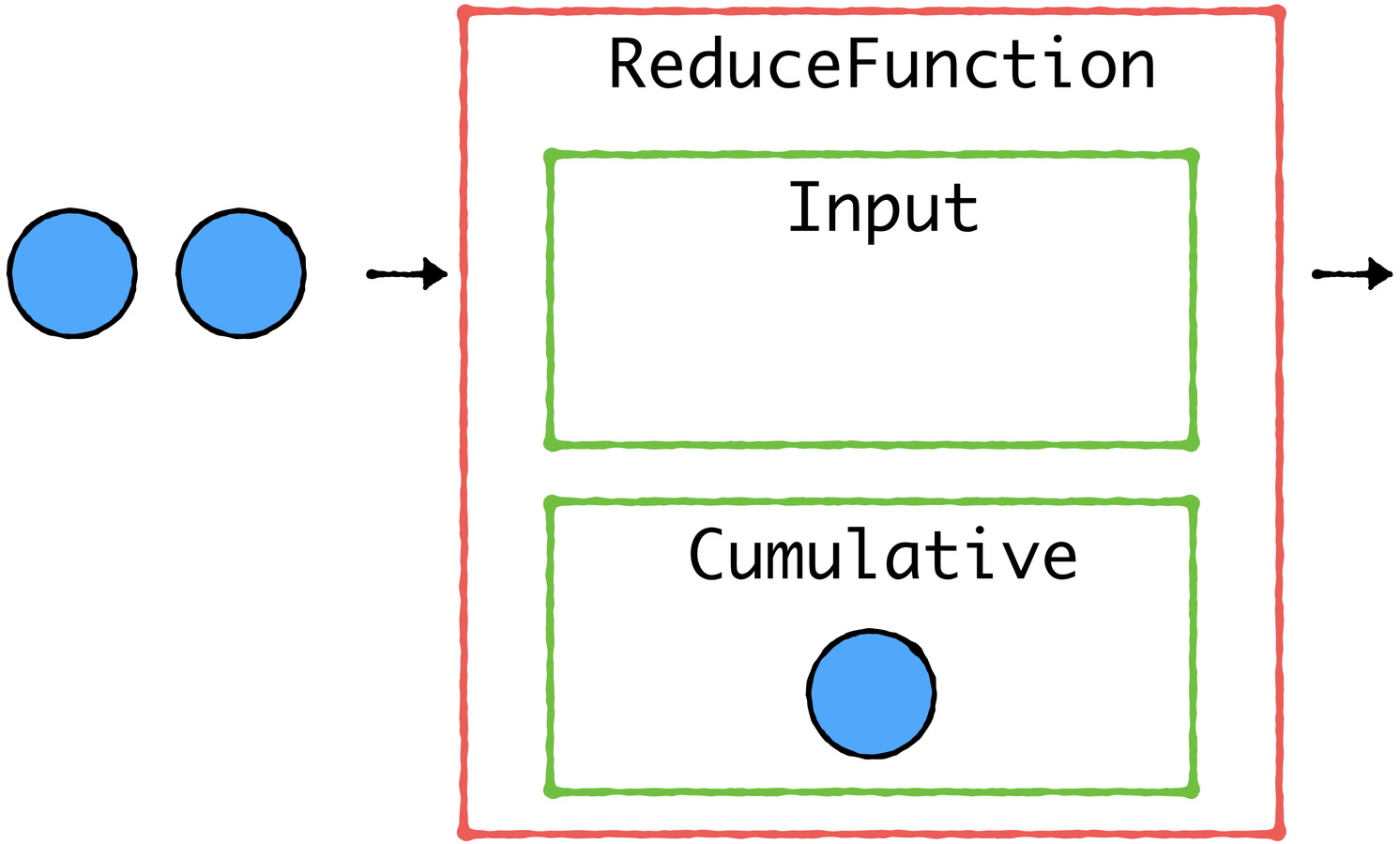
ReduceFunction

Reduce function combines 2 values at a time



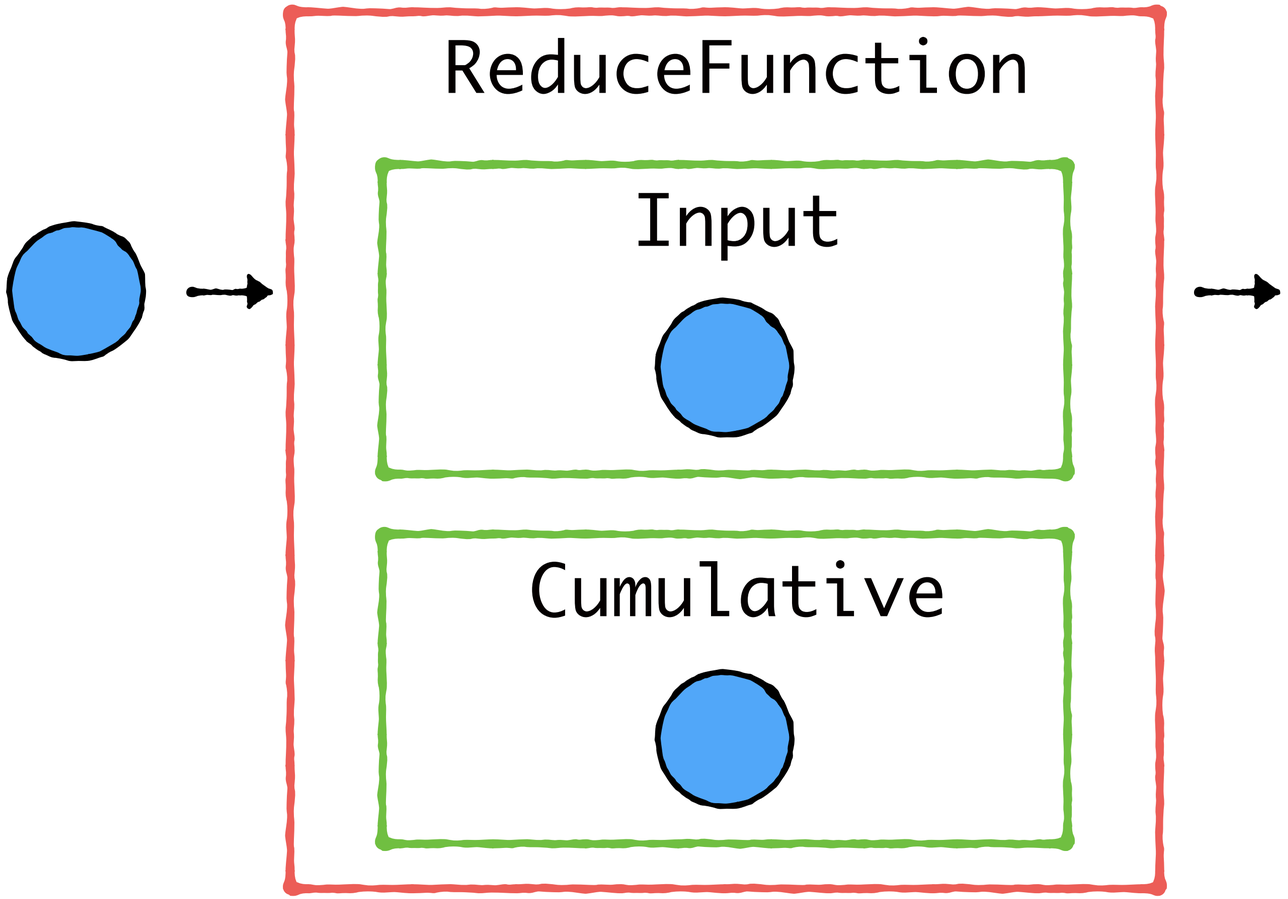
ReduceFunction

Reduce function combines 2 values at a time



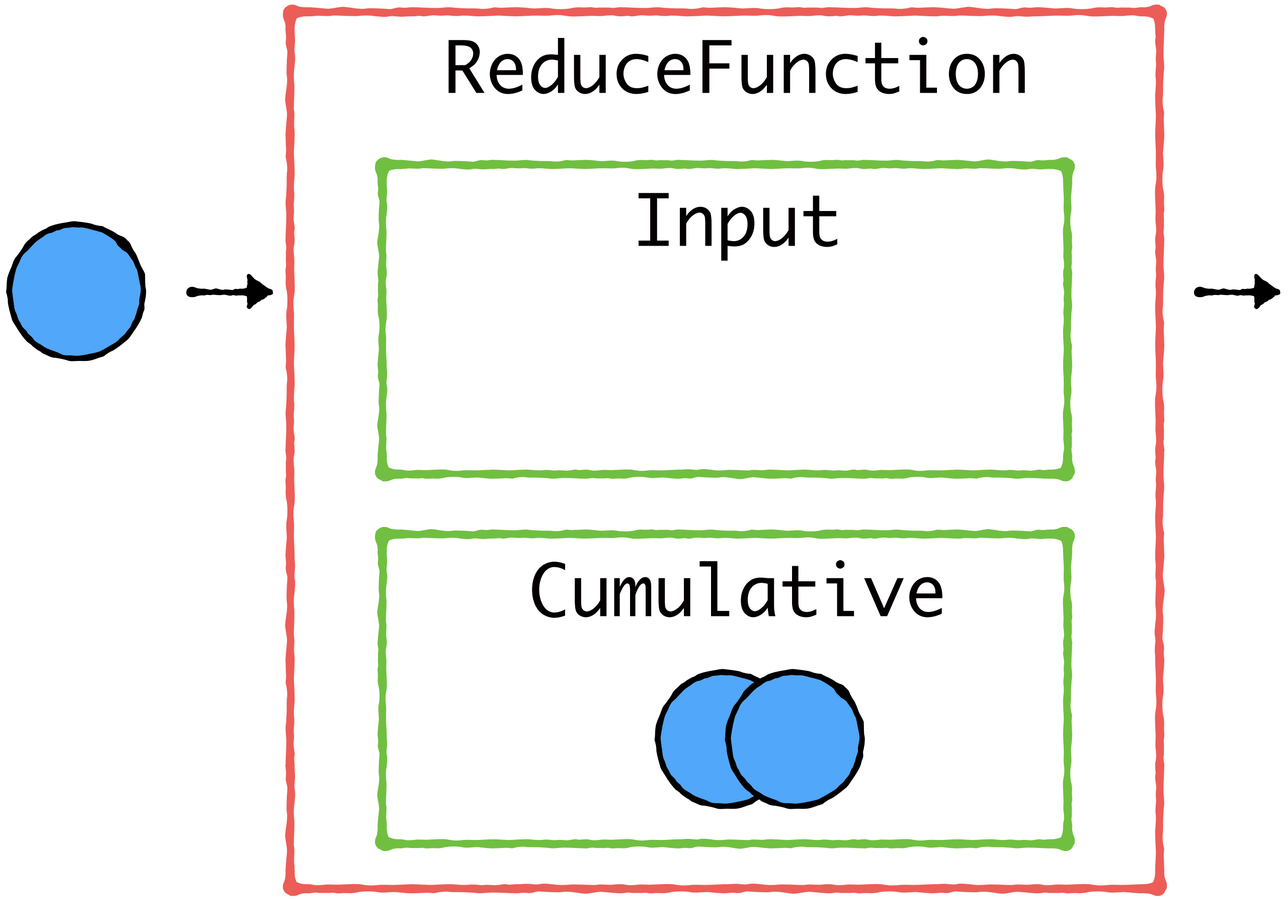
ReduceFunction

Reduce function combines 2 values at a time



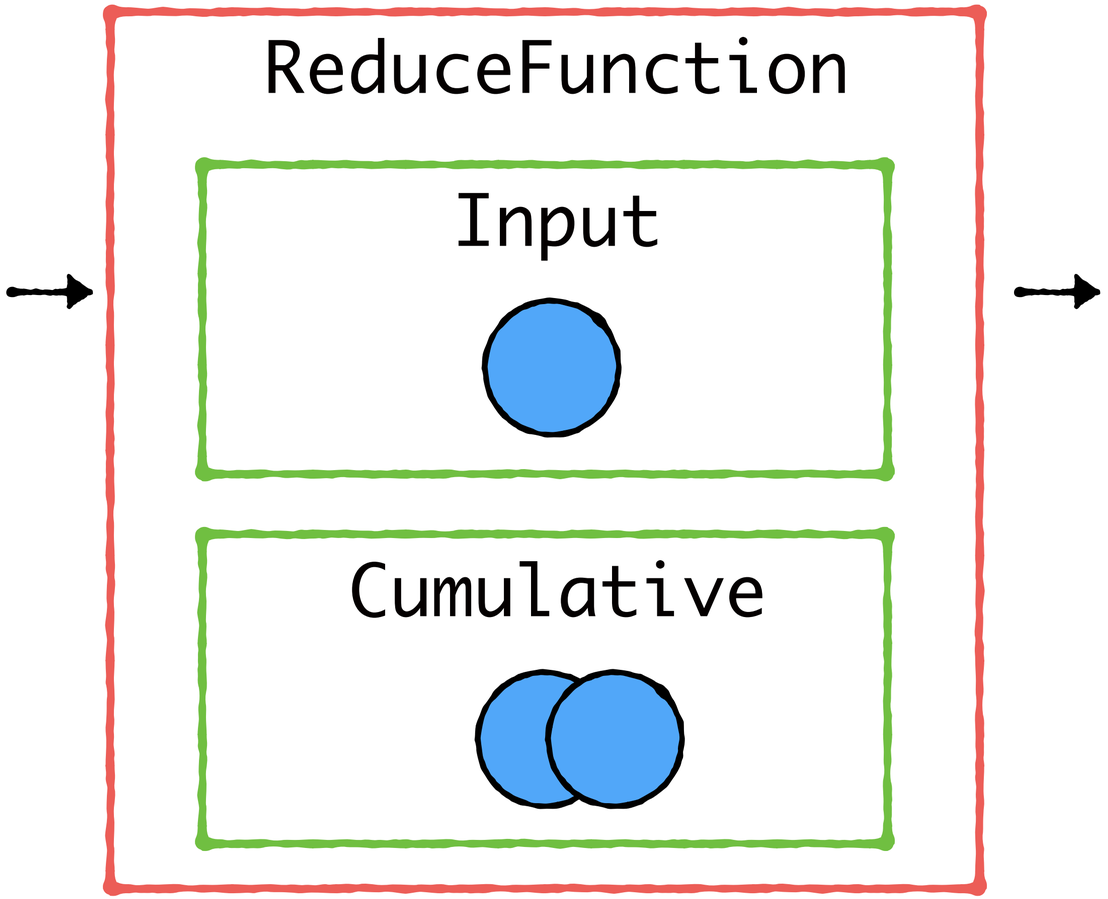
ReduceFunction

Reduce function combines 2 values at a time



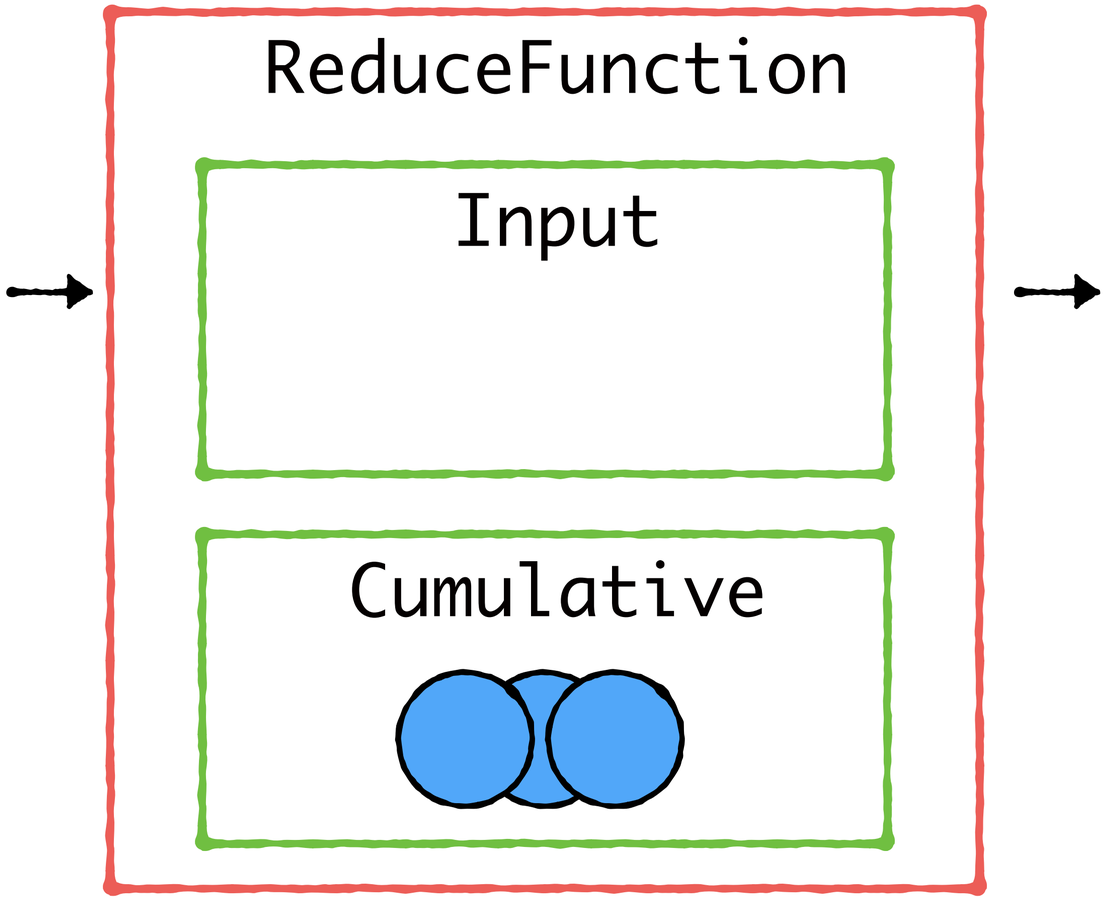
ReduceFunction

Reduce function combines 2 values at a time



ReduceFunction

Reduce function combines 2 values at a time



ReduceFunction

Reduce function combines 2 values at a time