DYNAMIC RESOURCE ALLOCATION, DO MORE WITH YOUR CLUSTER

Luc Bourlier
Lightbend

SPARK SUMMIT
EUROPE 2016

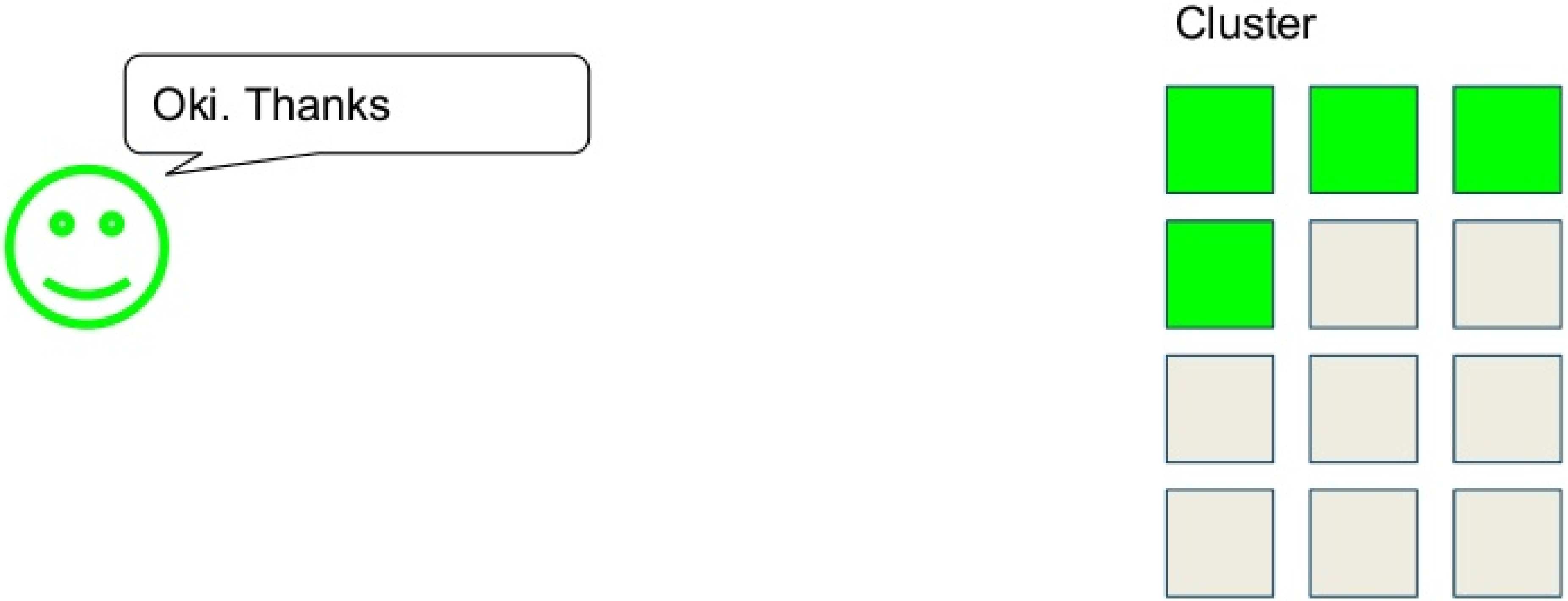
- Dynamic Resource Allocation
- ^^ in Spark
- External Shuffle Service
- Configuration
- Demo
- Spark Streaming

Dynamic Resource Allocation

Dynamic Resource Allocation

I'd like some resources for a job
Cluster

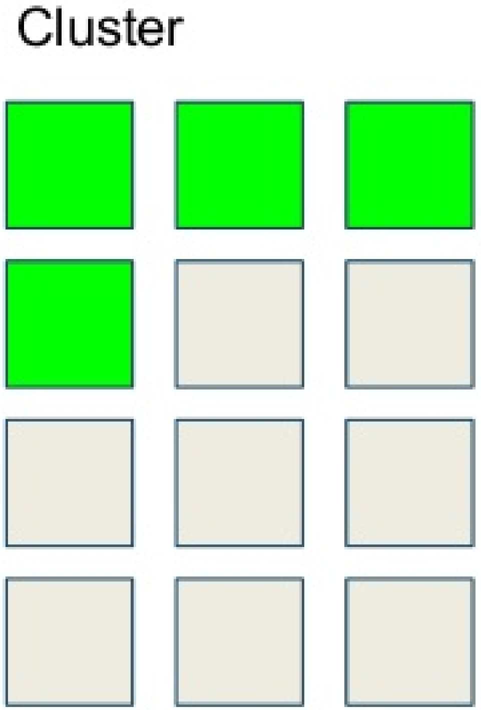
Dynamic Resource Allocation



Oki. Thanks

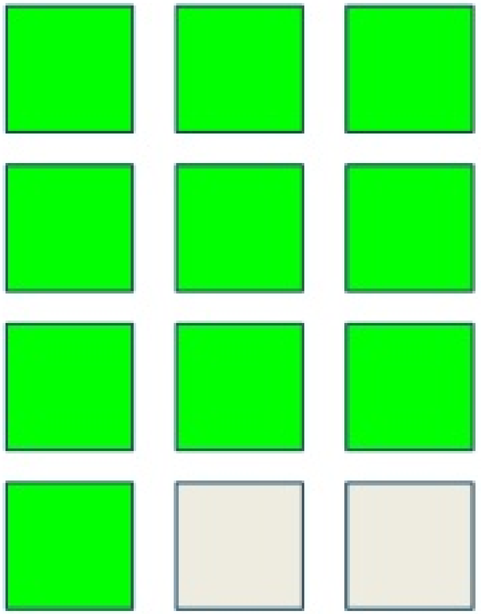
Cluster

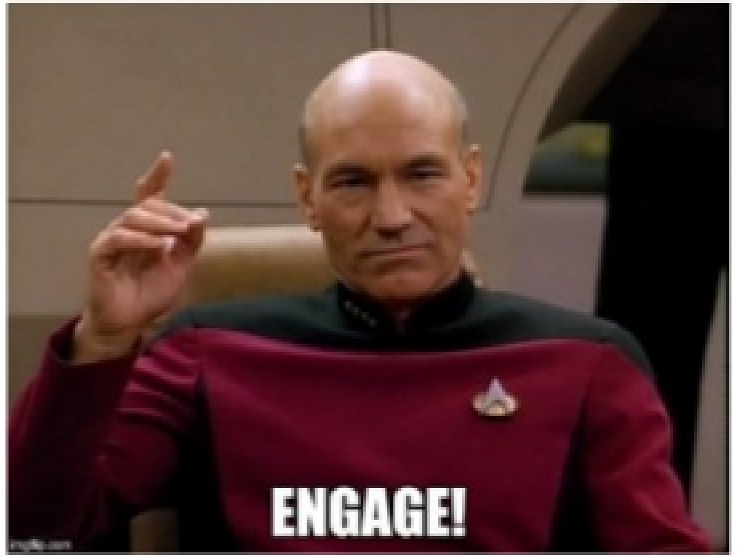
Dynamic Resource Allocation



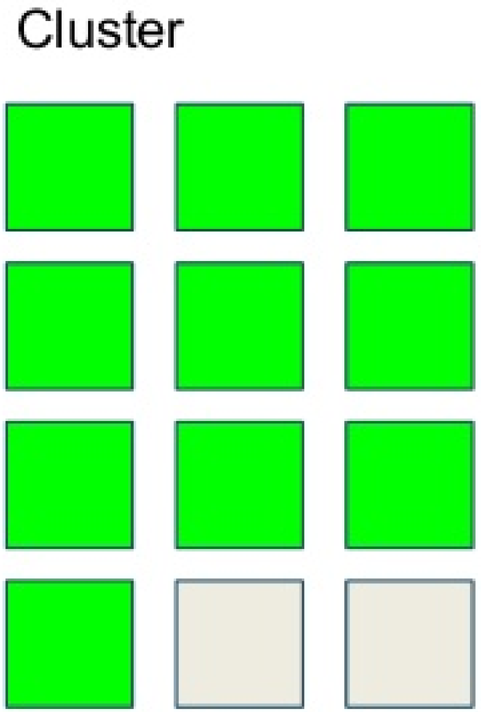
Dynamic Resource Allocation

Cluster



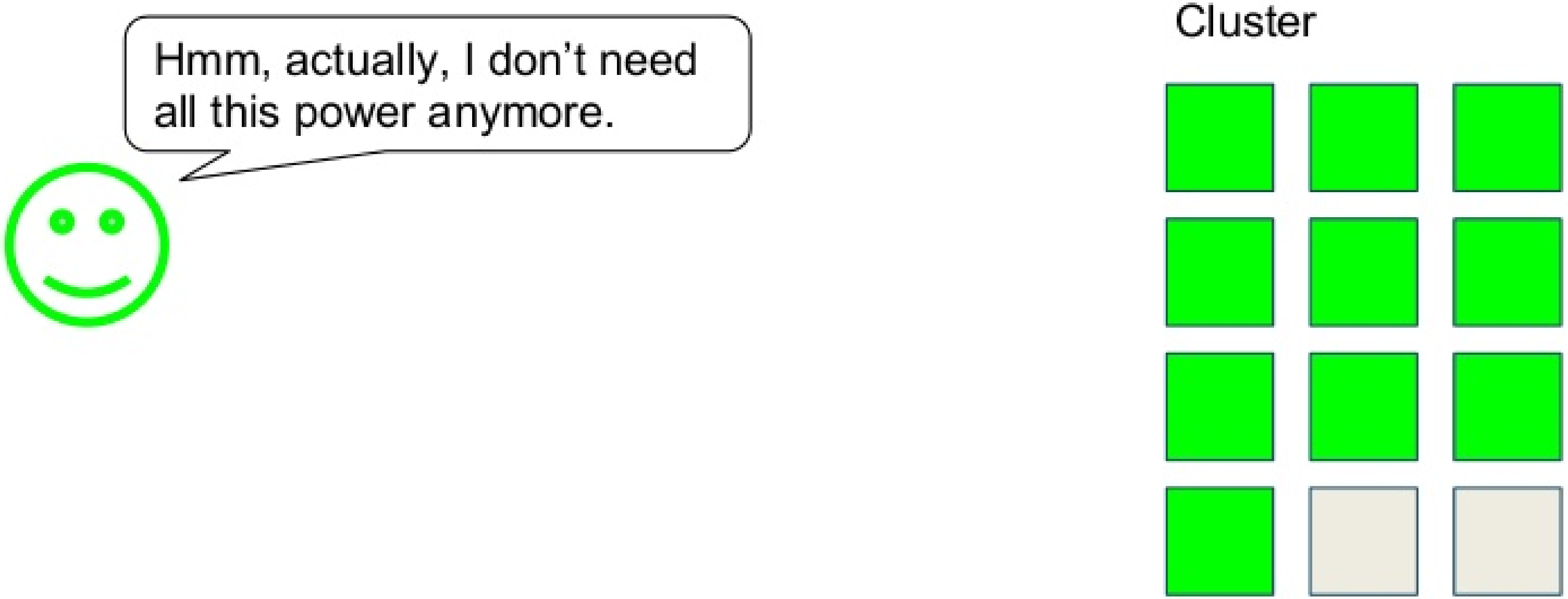


Dynamic Resource Allocation



Cluster

Dynamic Resource Allocation

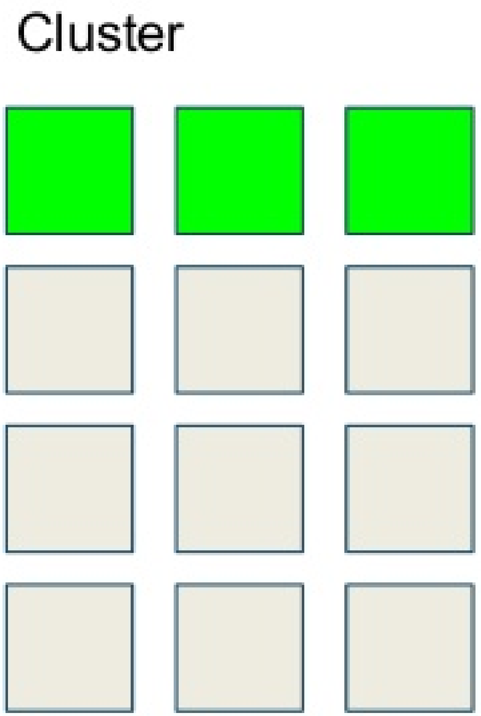


Hmm, actually, I don't need all this power anymore.

Cluster

Dynamic Resource Allocation

☺



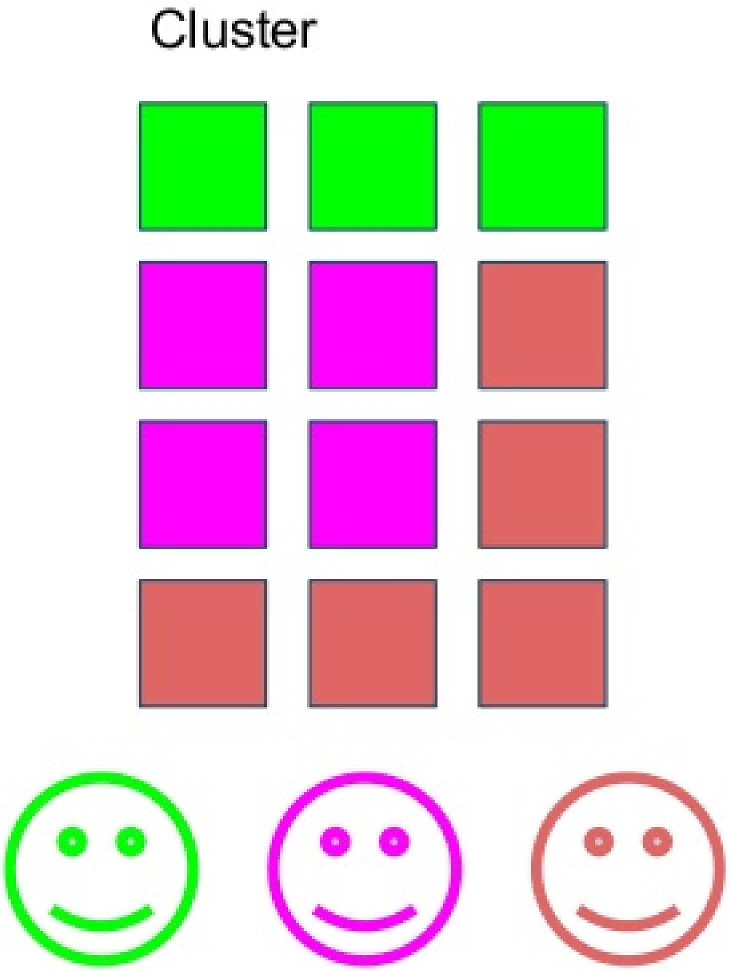
Dynamic Resource Allocation

Why?

- Shared cluster
- Optimization of resource usage

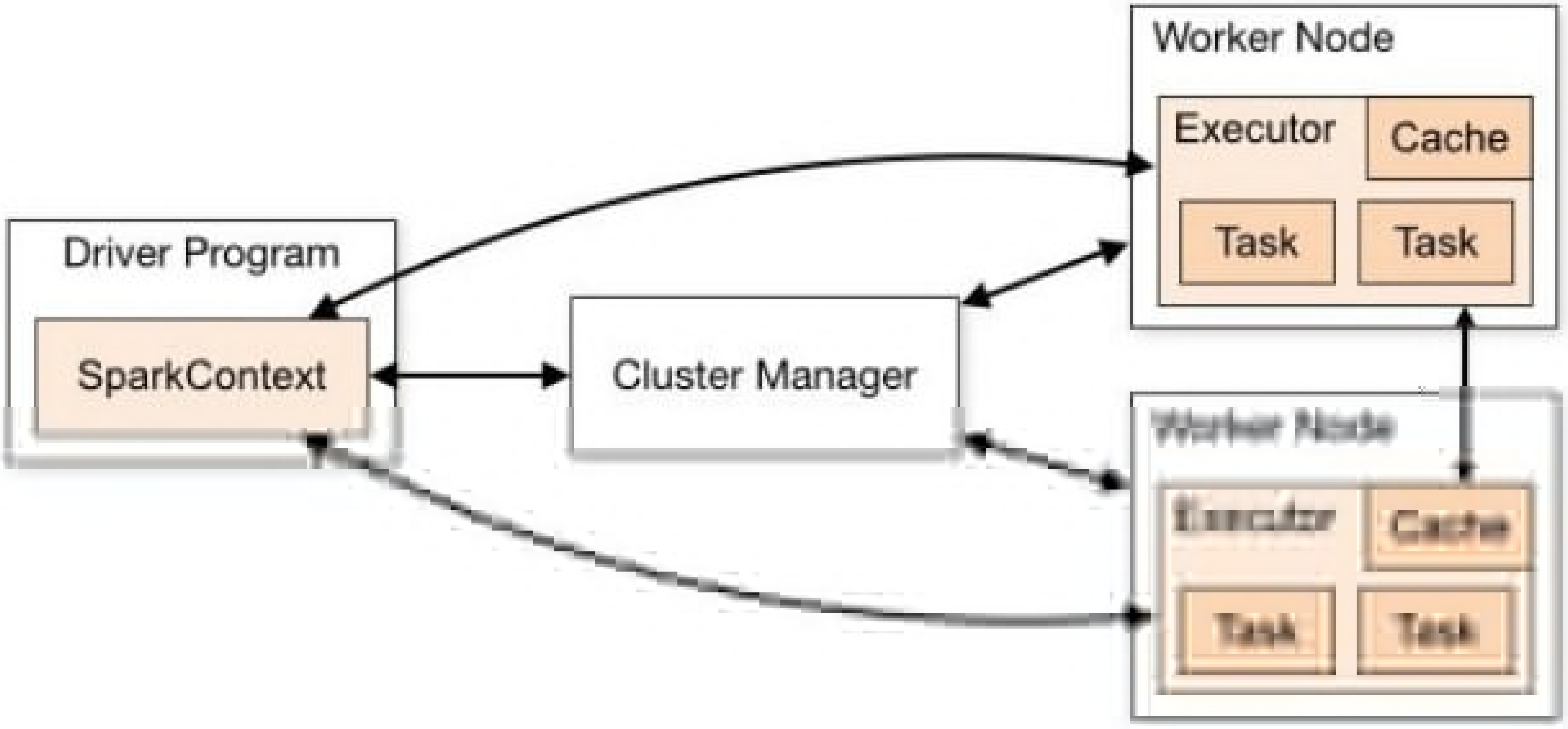
When?

- variable load job

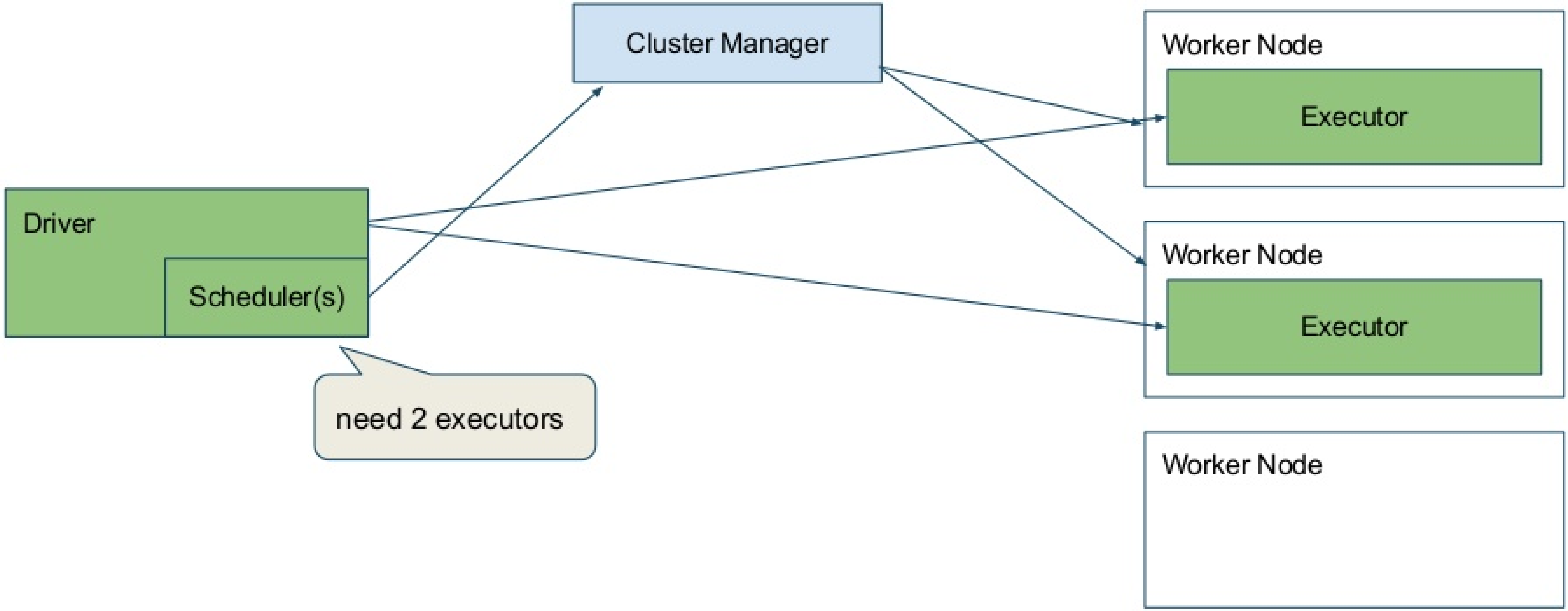


Spark Cluster Architecture

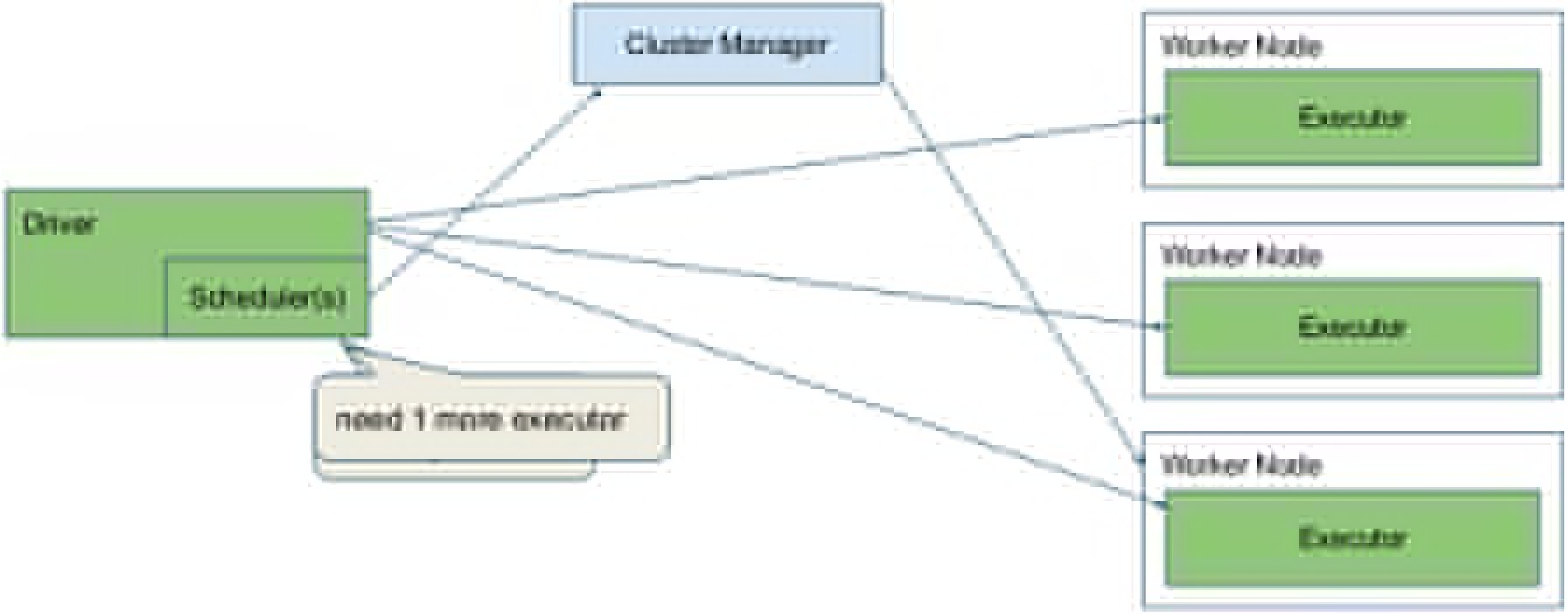
Spark Cluster Architecture



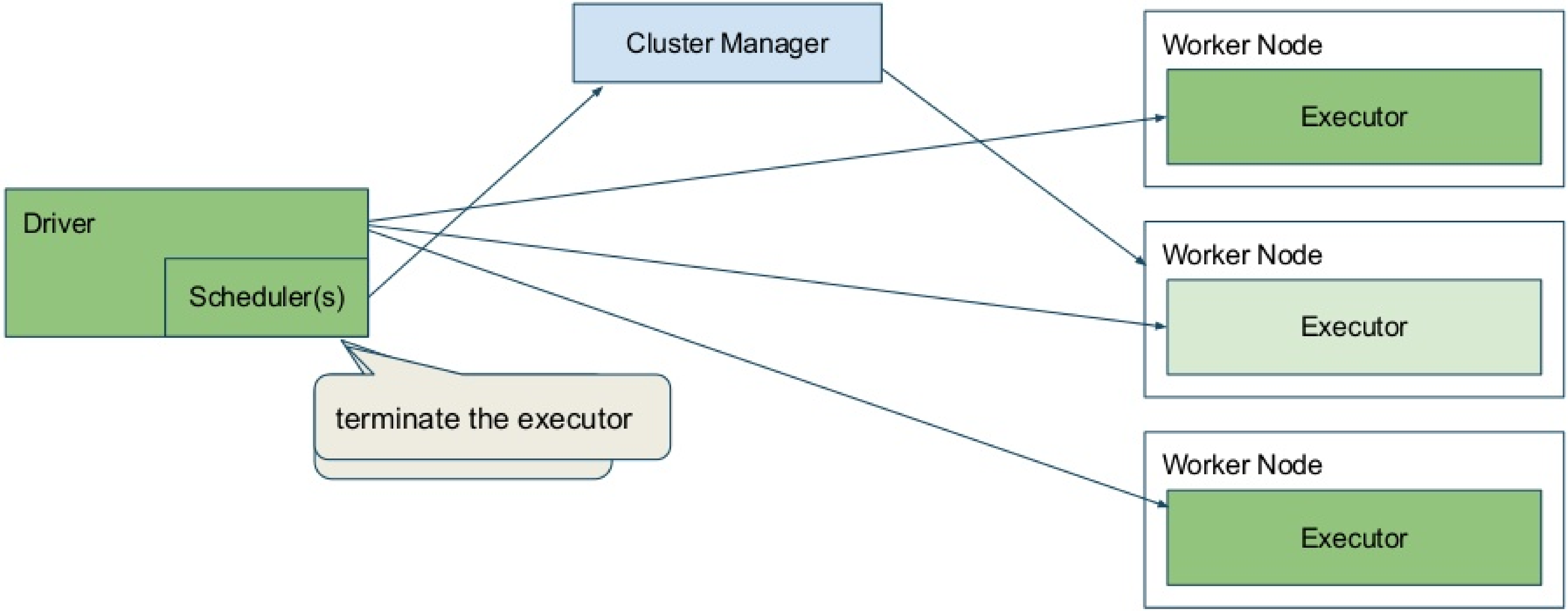
Spark Dynamic Allocation



Spark Dynamic Allocation



Spark Dynamic Allocation



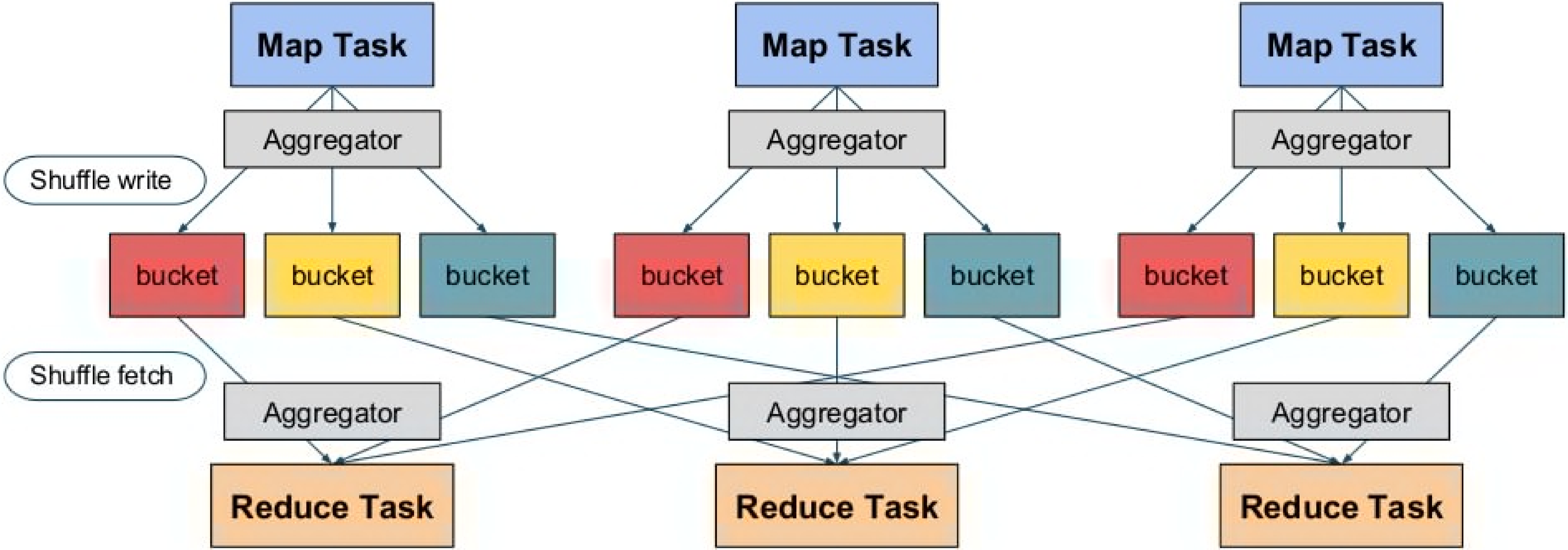
terminate the executor

External Shuffle Service

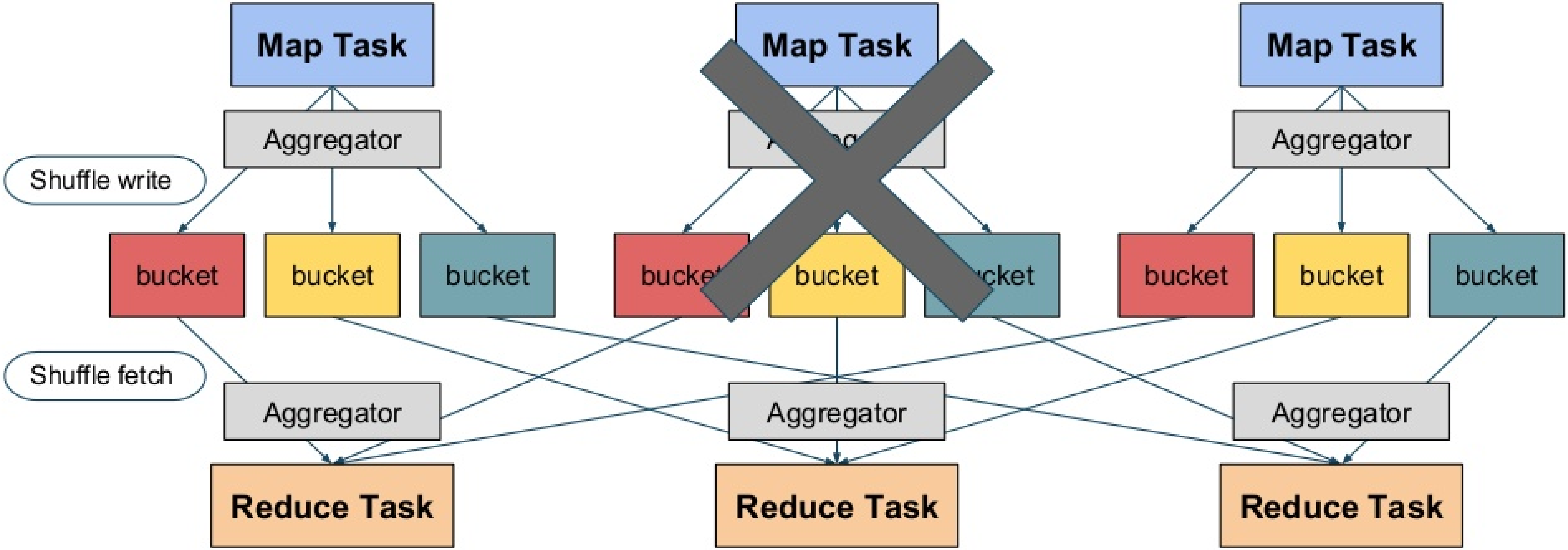
External Shuffle Service

- Did we lose any data?

External Shuffle Service



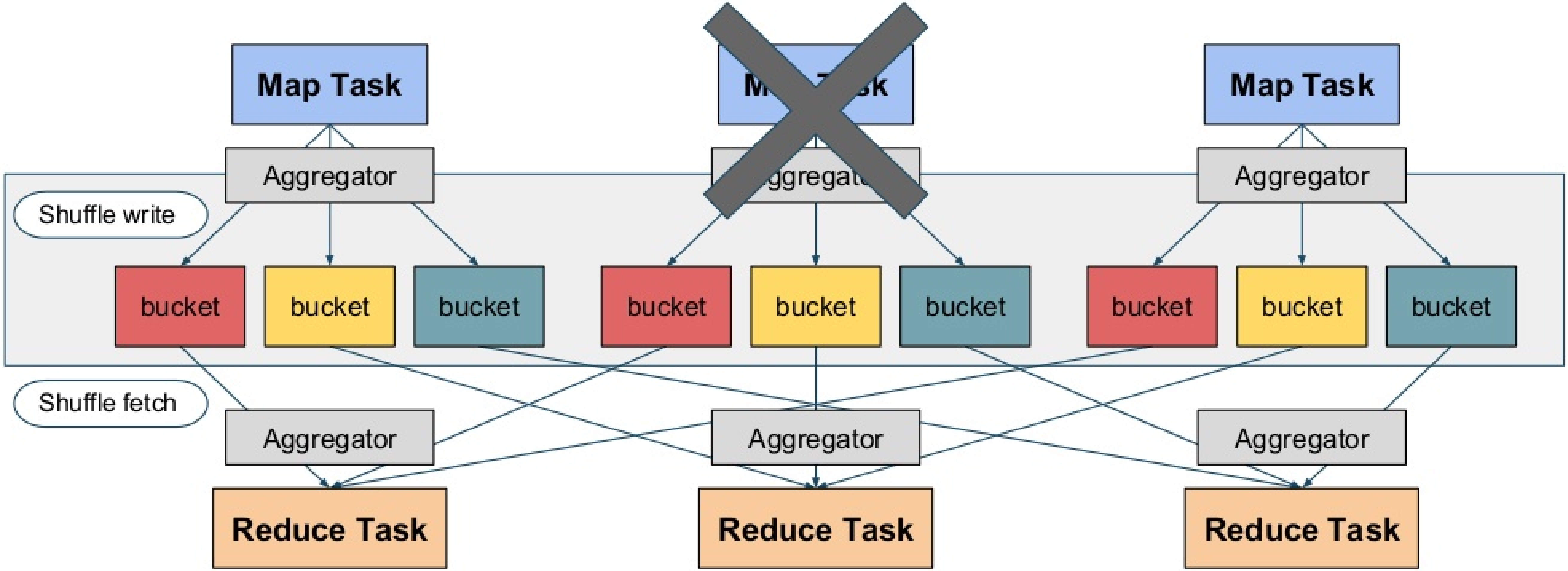
External Shuffle Service



SPARK SUMMIT
EUROPE 2016

Lightbend

External Shuffle Service



External Shuffle Service

- Extracted from executor
- Manage the local aggregated data for the shuffle operations
- Maintain the data until the application is done.

Configuration

Configuration

- Dynamic Allocation
  - `spark.dynamicAllocation.enabled`
  - `spark.dynamicAllocation.initialExecutors`
  - `spark.dynamicAllocation.maxExecutors`
  - `spark.dynamicAllocation.minExecutors`

Configuration

- Dynamic Allocation
  - `spark.dynamicAllocation.schedulerBacklogTimeout`
  - `spark.dynamicAllocation.executorIdleTimeout`

  - `spark.dynamicAllocation.sustainedSchedulerBacklogTimeout`

Configuration

- External Shuffle Service
    - `spark.shuffle.service.enabled`
    - `spark.shuffle.service.port`

Demo

Dynamic Allocation in Action

Configuration values?

Configuration Values?

It depends ….

No, seriously

Configuration Values

- `spark.dynamicAllocation.initialExecutors`
- `spark.dynamicAllocation.maxExecutors`
- `spark.dynamicAllocation.minExecutors`

Depends on workload and how many resources are potentially available to you.

Configuration Values

- `spark.dynamicAllocation.schedulerBacklogTimeout`

Too short, might trigger for short burst of tasks.

Too long, might be less effective.

- `spark.dynamicAllocation.sustainedSchedulerBacklogTimeout`

Executor start duration.

Default set to `schedulerBacklogTimeout`.

Configuration Values

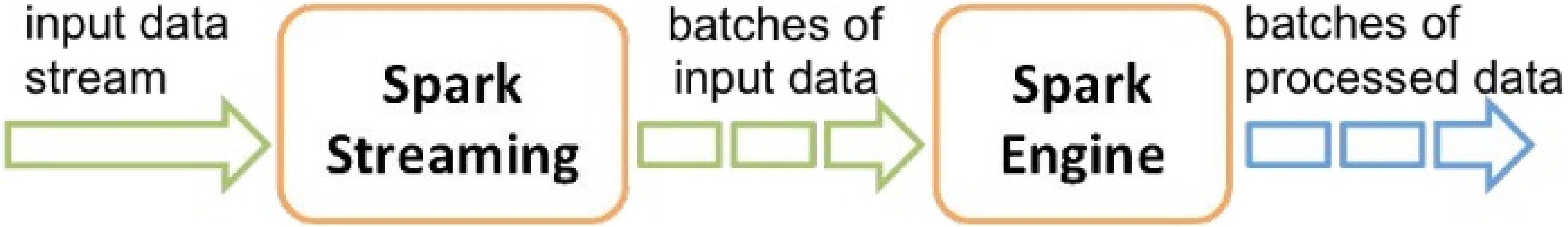
- `spark.dynamicAllocation.executorIdleTimeout`

Relative to the duration of the longer task.

No big drawback on being too long, except cost.

Spark Streaming

Spark Streaming



http://spark.apache.org/docs/latest/streaming-programming-guide.htm
l

Spark Streaming

- In most case, `schedulerBacklogTimeout` longer than batch interval.
- `executorIdleTimeout` a portion of batch interval.

- Should allow to manage processing delay.

- Not compatible with the dynamic rate estimator.

More Dynamic?

More Dynamic?

https://github.com/twosigma/Cook

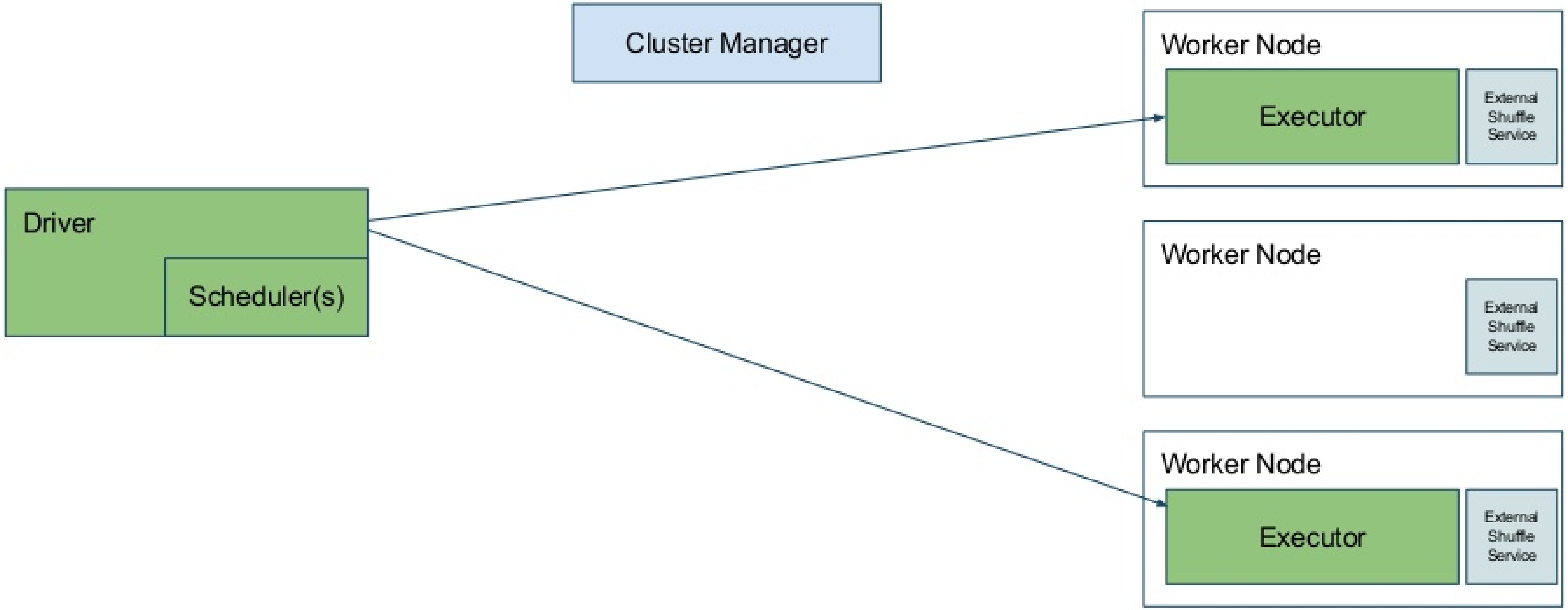
‘Fair’ job scheduler for Spark on top of Mesos

- Not a recommendation, just a suggestion.
- Some assembly required.

THANK YOU.

github.com/skyluc/tree/master/talks/sparksummit-eu-2016

External Shuffle Service


Cluster Manager
Driver
Scheduler(s)
Worker Node
Executor
External Shuffle Service
Worker Node
External Shuffle Service
Worker Node
Executor
External Shuffle Service