

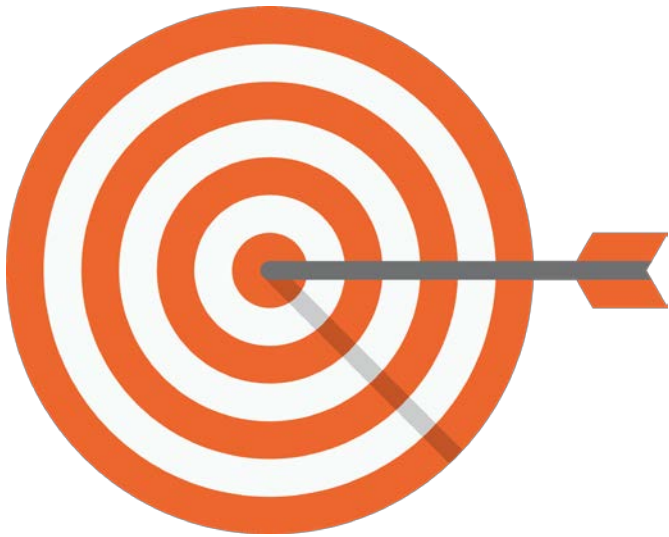
# Storm Deep Dive



Elton Stoneman

@EltonStoneman | [blog.sixeyed.com](http://blog.sixeyed.com)

# Module Goals



Storm Cluster Architecture

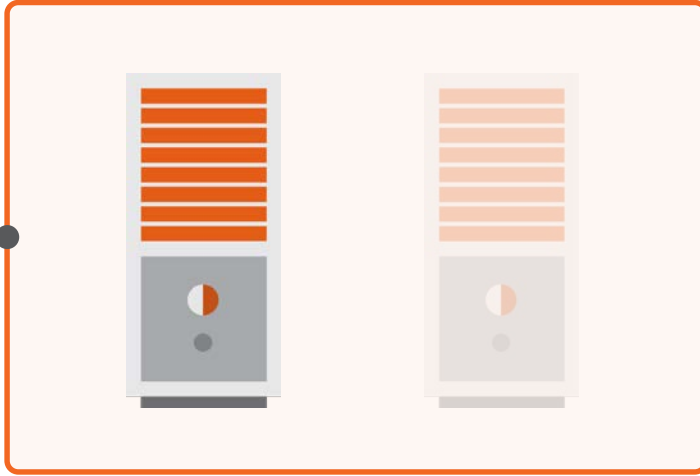
Runtime Compute Structure

Performance Tuning

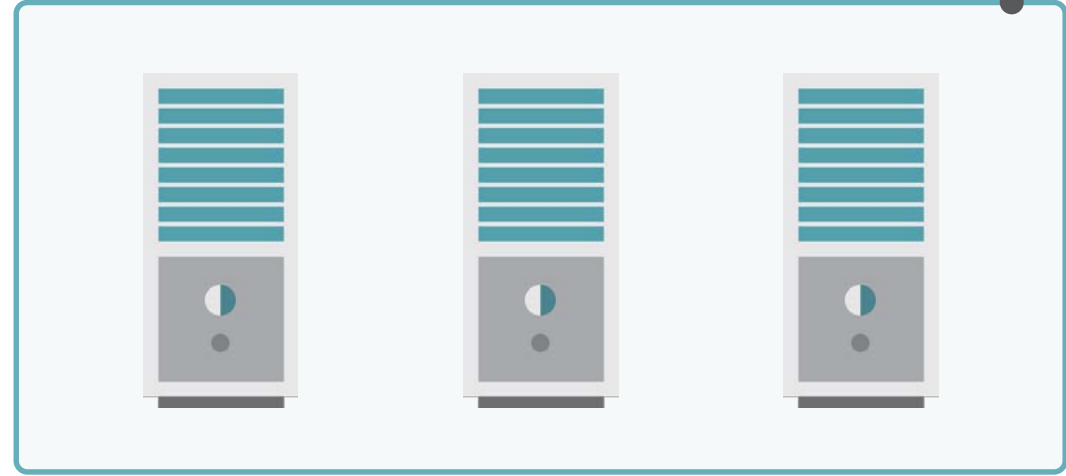
Guaranteed Processing

Monitoring & Testing

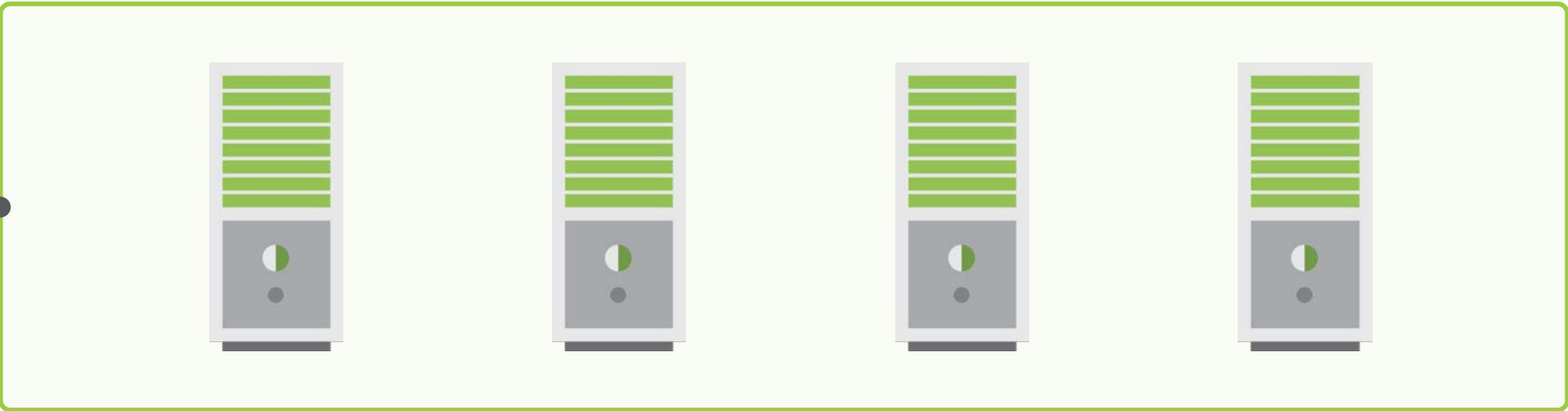
Nimbus  
Server(s)

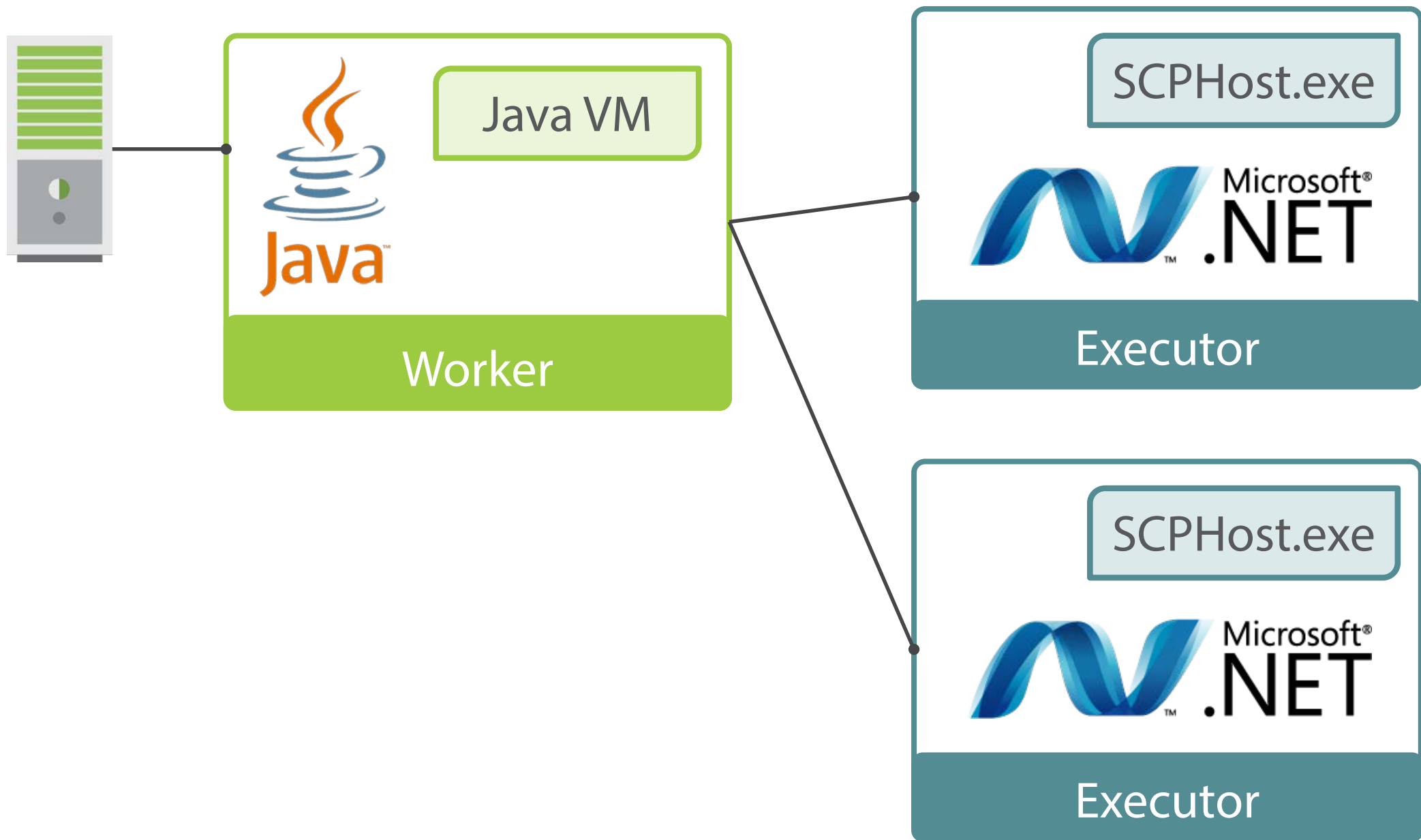


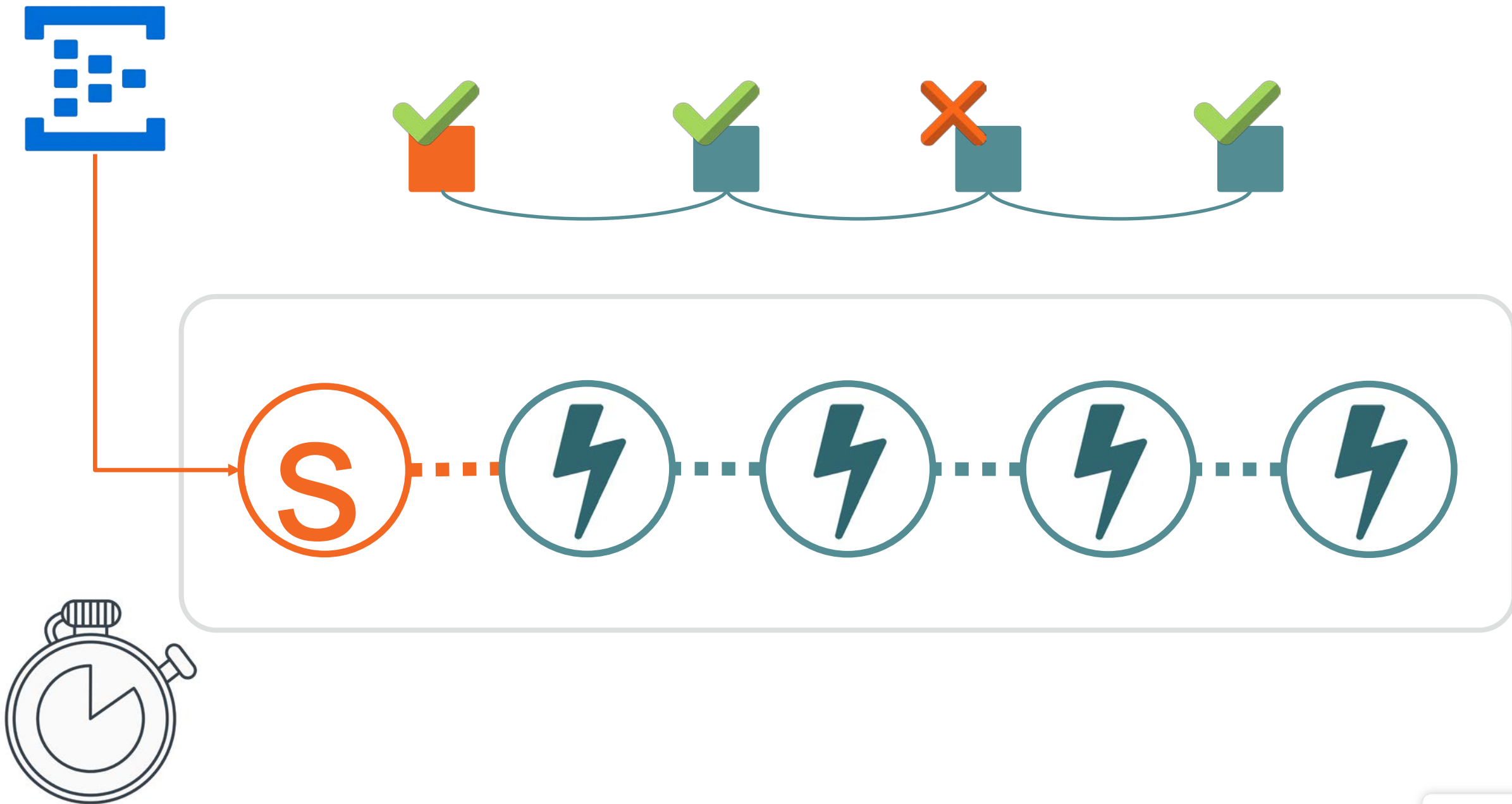
Zookeeper



Supervisor  
Nodes







## Spouts (All time)

Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error
EventHubSpout	16	16	8160	8160	0.000	0	0			

## Bolts (All time)

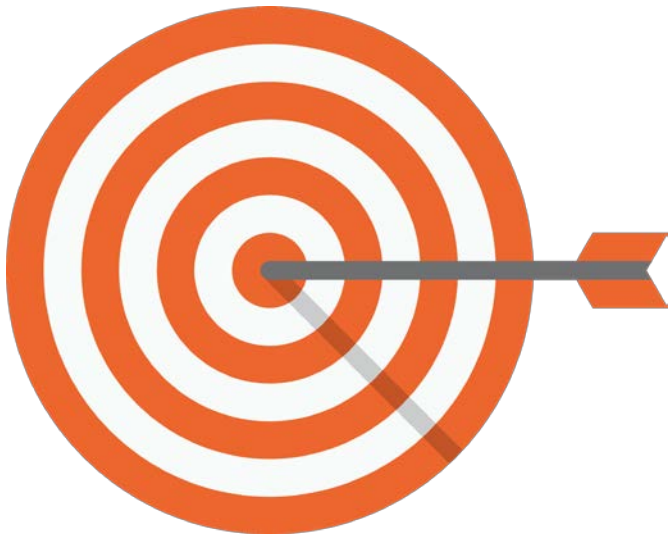
Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error
__acker	4	4	0	0	0.001	0.038	1060	0.019	1080	0			
__metricsbacktype.storm.metric.LoggingMetricsConsumer	1	1			0.000	0		0					
BatchSectorTimeBolt	64	64	20	0	0.000	0.000	200	0.000	0	0			
RaceResultBolt	32	32	0	0	0.000	0.000	0	0.000	0	0			
TimingEventBolt	256	256	880	780	0.342	13.705	880	20287.234	340	0			
UpdateSectorTimeBolt	128	128	0	0	0.000	0.000	0	0.000	0	0			





```
https://psod-hdi-storm.azurehdinsight.net/stormui/wn2/log?file=RaceTiming-1009-1607-11-1444403406-v Search
-09 15:11:24,019 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 70ms
-09 15:11:24,020 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 68ms
-09 15:11:24,021 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 63ms
-09 15:11:24,024 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 58ms
-09 15:11:24,024 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: 2fbe, raceDayTimestamp: 1270
-09 15:11:24,024 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 55ms
-09 15:11:24,025 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: 282d, raceDayTimestamp: 124
-09 15:11:24,029 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 70ms
-09 15:11:24,030 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 57ms
-09 15:11:24,035 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: 3330, raceDayTimestamp: 145
-09 15:11:24,036 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** UpdateSectorTimeBolt, did lookup - race ID: 8c43b7, took: 59ms
-09 15:11:24,038 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 62ms
-09 15:11:24,039 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 68ms
-09 15:11:24,040 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 204ms
-09 15:11:23,878 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 57ms
-09 15:11:23,946 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 67ms
-09 15:11:23,998 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 51ms
-09 15:11:23,900 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 71ms
-09 15:11:24,043 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 204ms
-09 15:11:24,045 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 81ms
-09 15:11:24,046 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 64ms
-09 15:11:23,904 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 83ms
-09 15:11:23,970 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 65ms
-09 15:11:24,037 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 66ms
-09 15:11:23,937 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 111ms
-09 15:11:23,991 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 54ms
-09 15:11:24,048 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: 2fbe, raceDayTimestamp: 1270
-09 15:11:24,003 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 94
-09 15:11:24,004 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 95
-09 15:11:24,004 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 96
-09 15:11:24,004 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 97
-09 15:11:24,004 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 98
-09 15:11:24,005 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** BatchSectorTimeBolt, added to batch - size: 99
-09 15:11:24,054 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: dcf5, raceDayTimestamp: 132
-09 15:11:24,041 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 227ms
-09 15:11:24,059 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 79ms
-09 15:11:24,060 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 57ms
-09 15:11:24,061 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 68ms
-09 15:11:24,062 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, PutTimestamp, took: 71ms
-09 15:11:24,063 [1] INFO Sixeyed_RaceTiming_EventProcessor_ComponentFactory [(null)] - ** TimingEventBolt, did lookup - timerId: 4495, raceDayTimestamp: 145
```

# Module Goals



Storm Cluster Architecture

Runtime Compute Structure

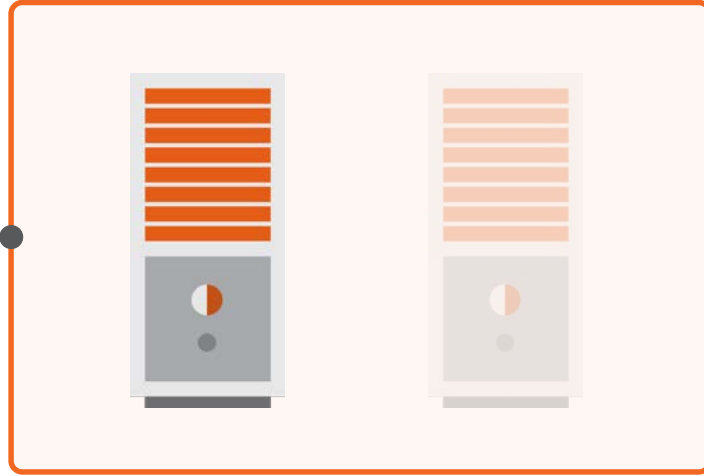
Performance Tuning

Guaranteed Processing

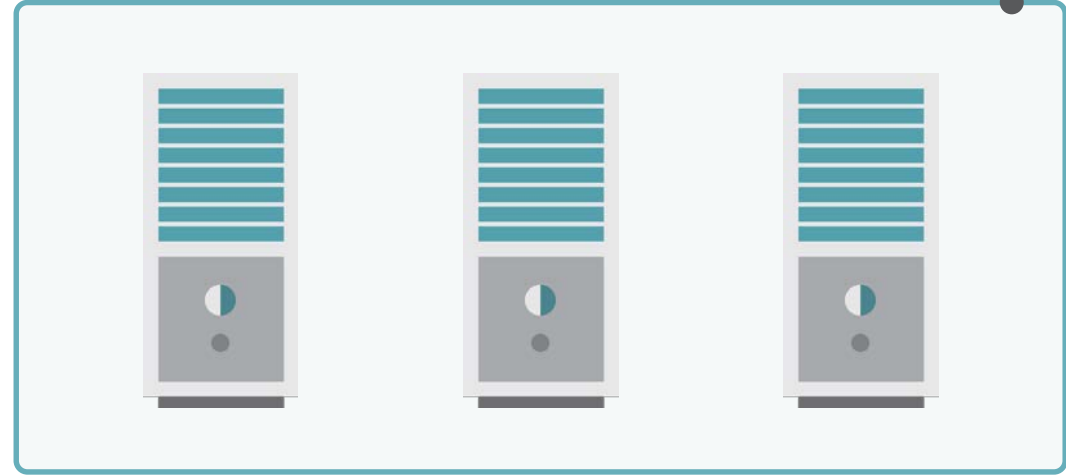
Monitoring & Testing



Nimbus  
Server(s)



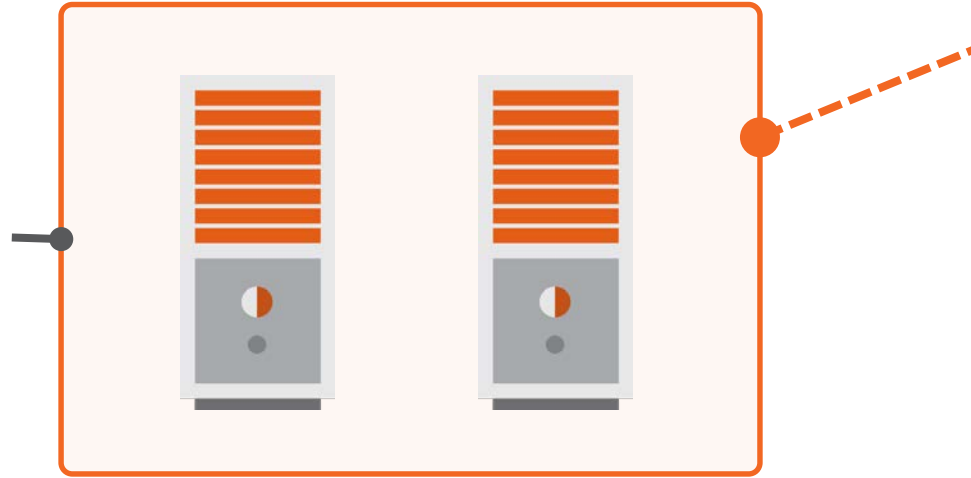
Zookeeper



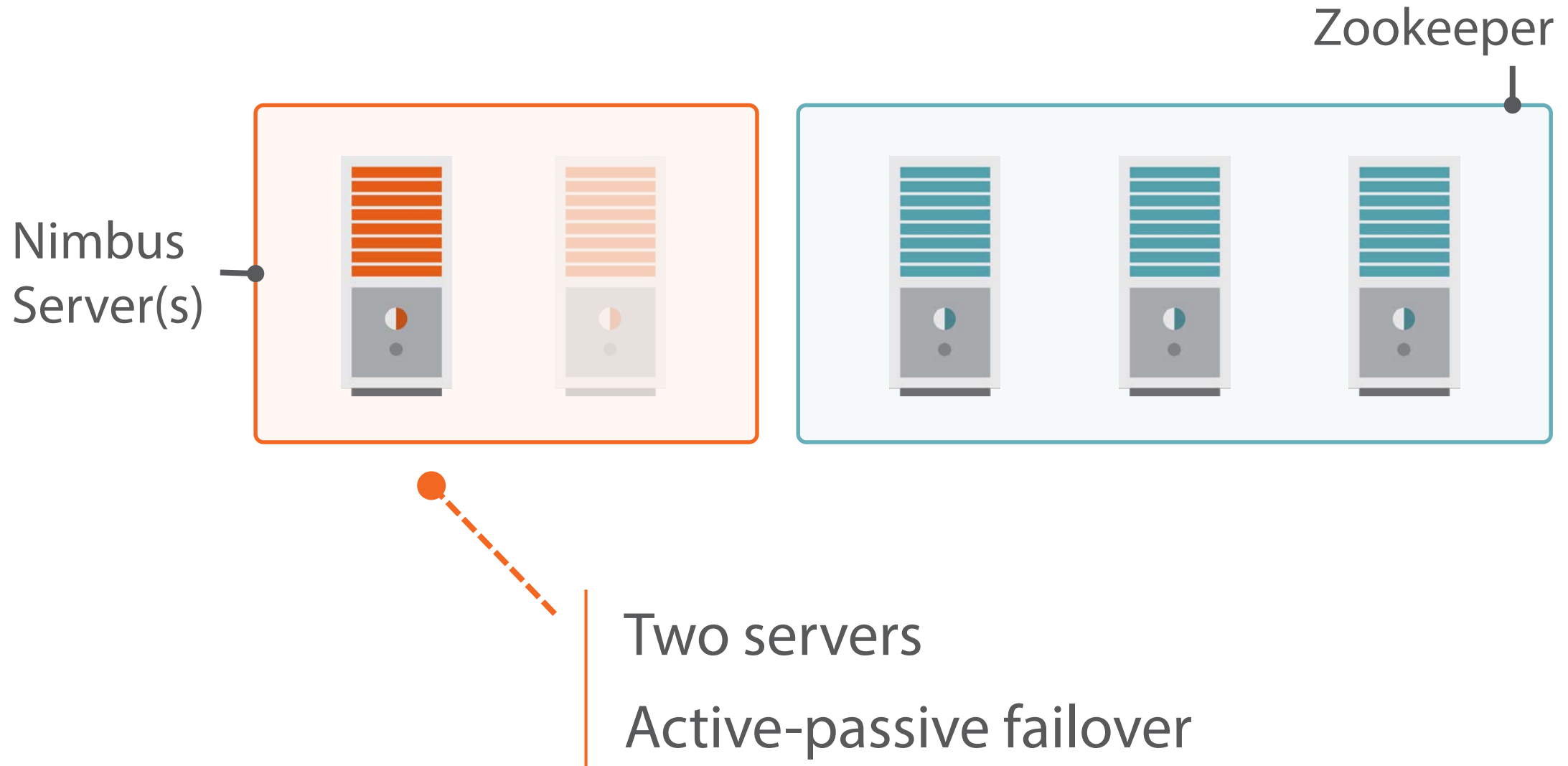
Supervisor  
Nodes



Nimbus  
Server(s)

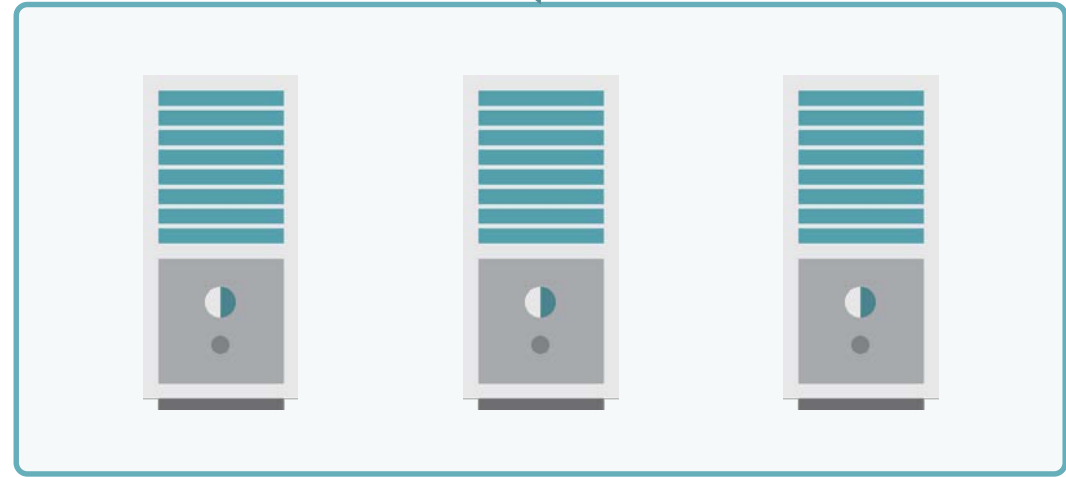


**Component** creation  
Component **monitoring**  
**Work** distribution

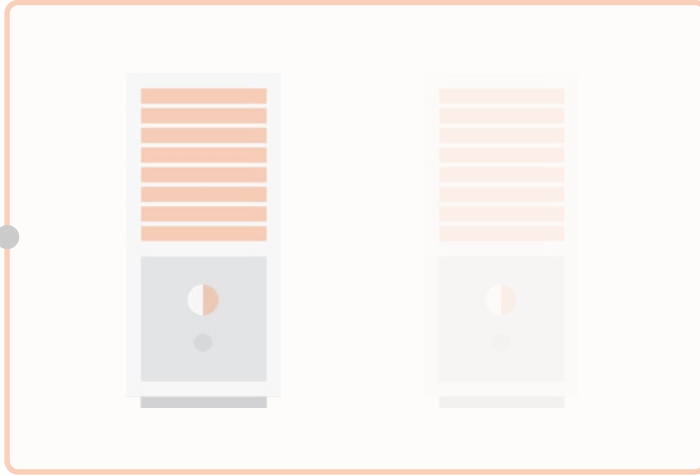




nimbus down



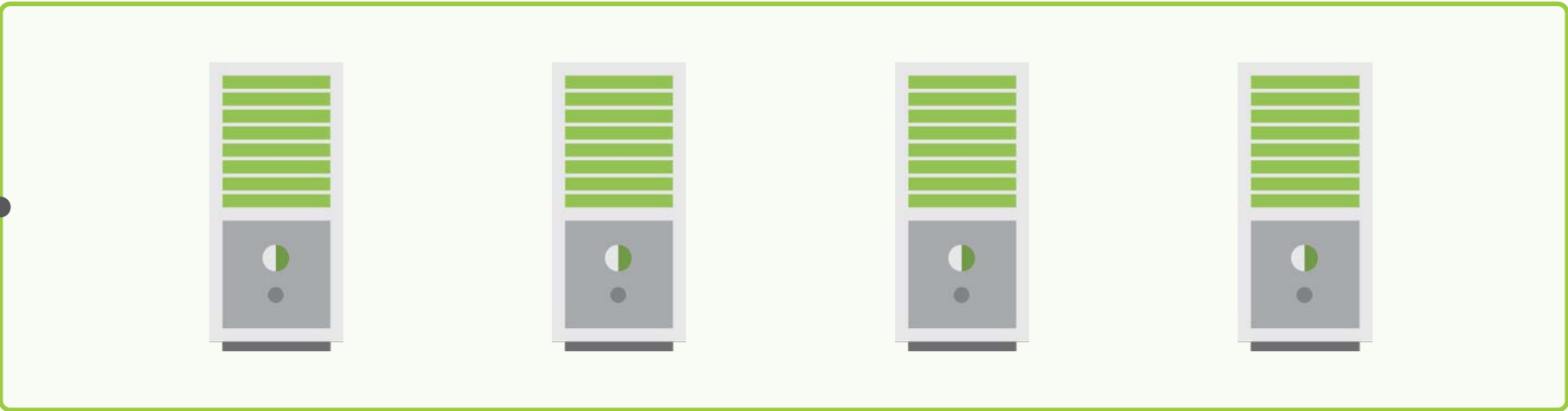
Nimbus  
Server(s)

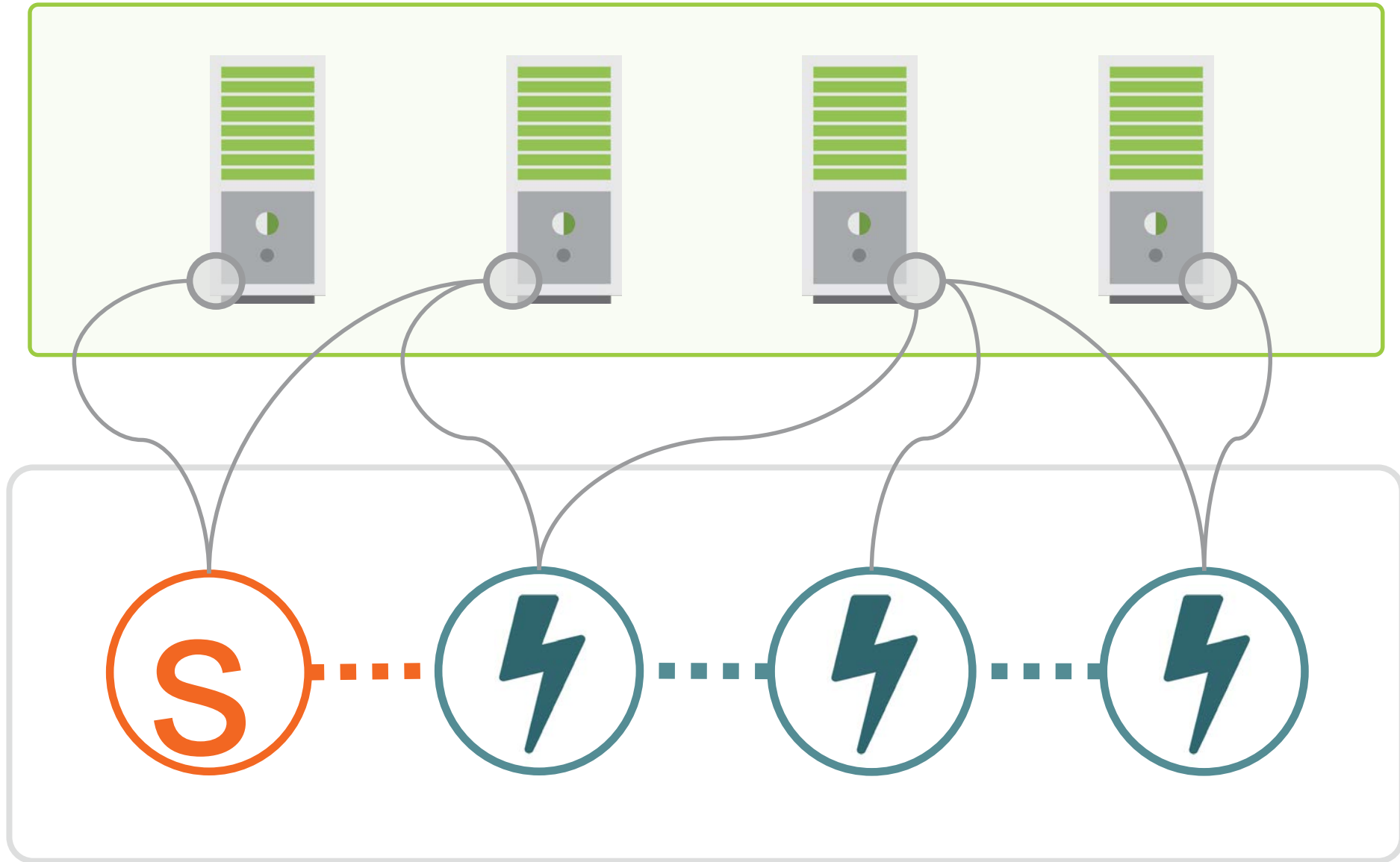


Zookeeper



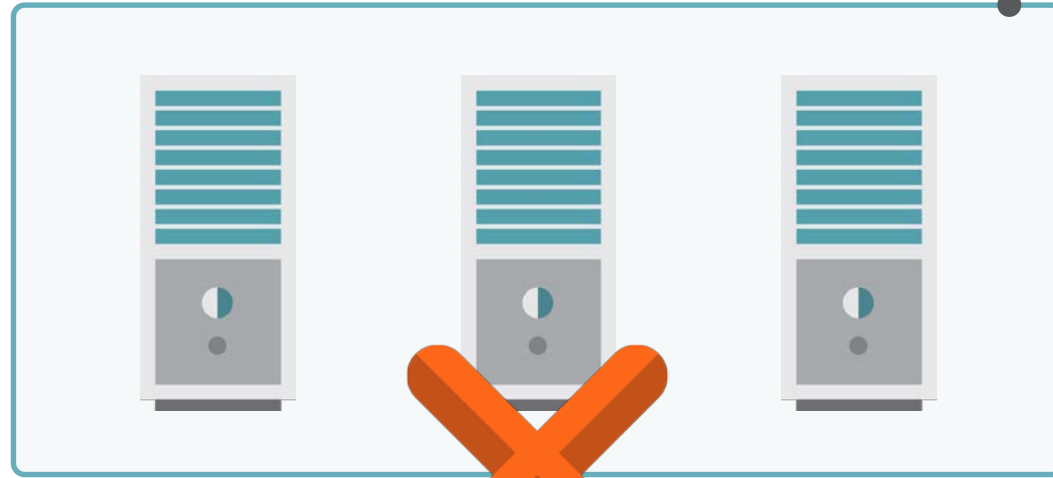
Supervisor  
Nodes







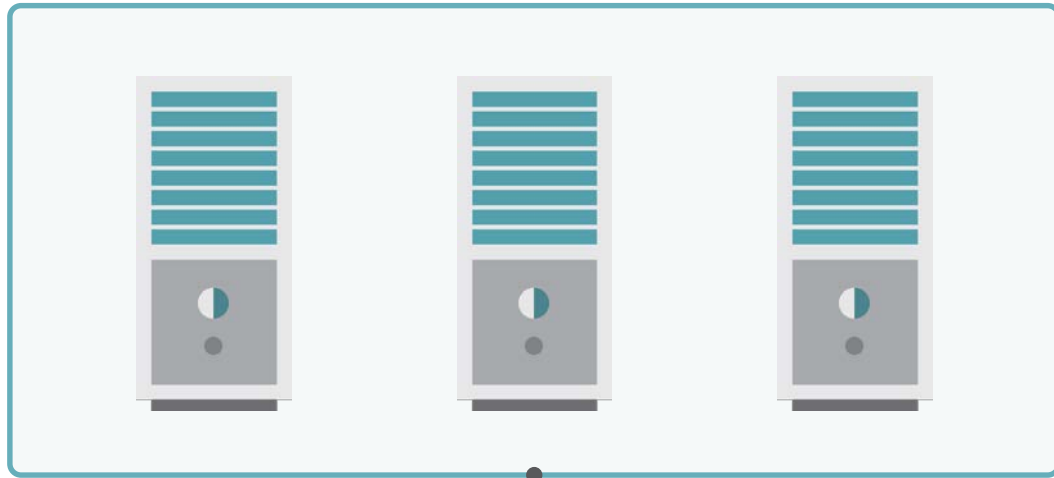
Zookeeper



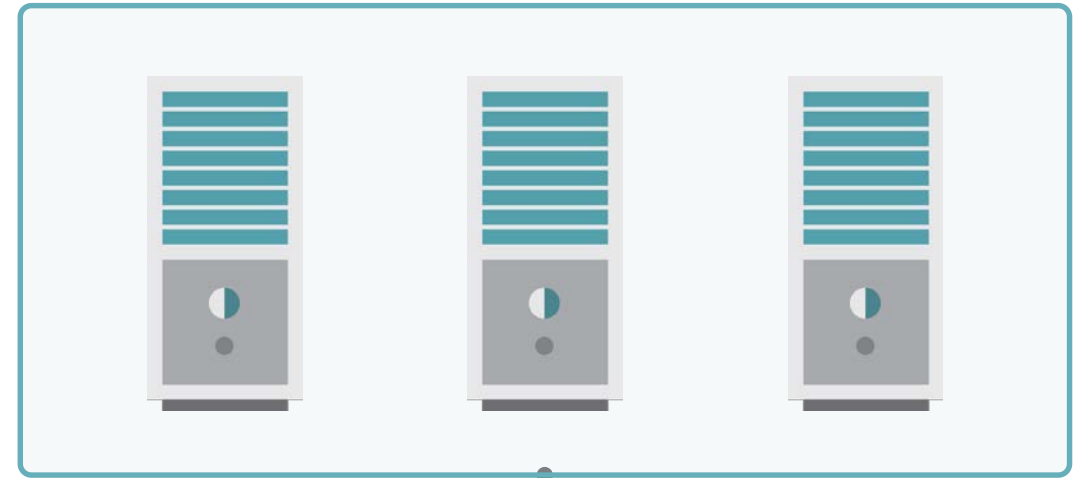
APACHE  
HBASE



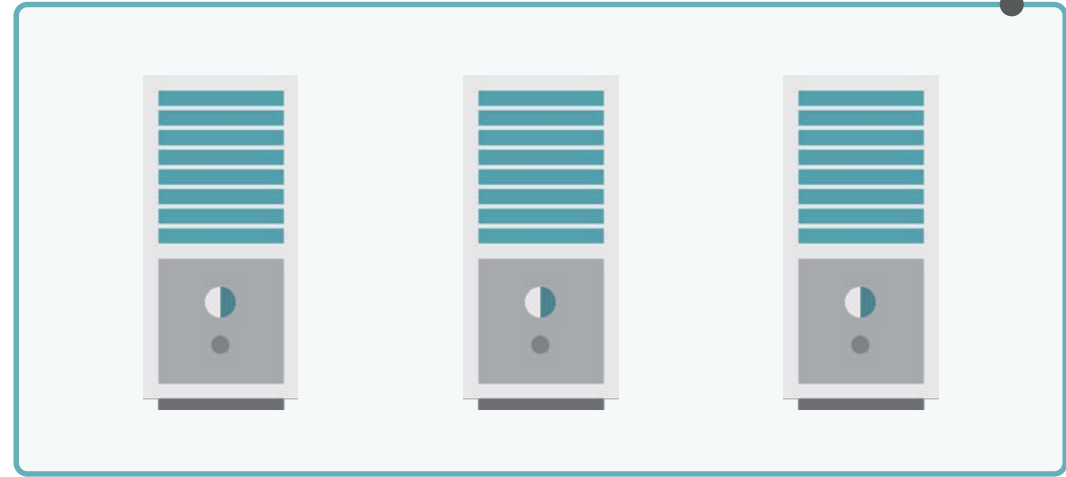




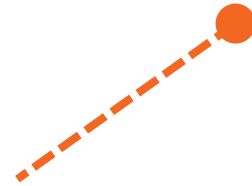
APACHE  
**HBASE**



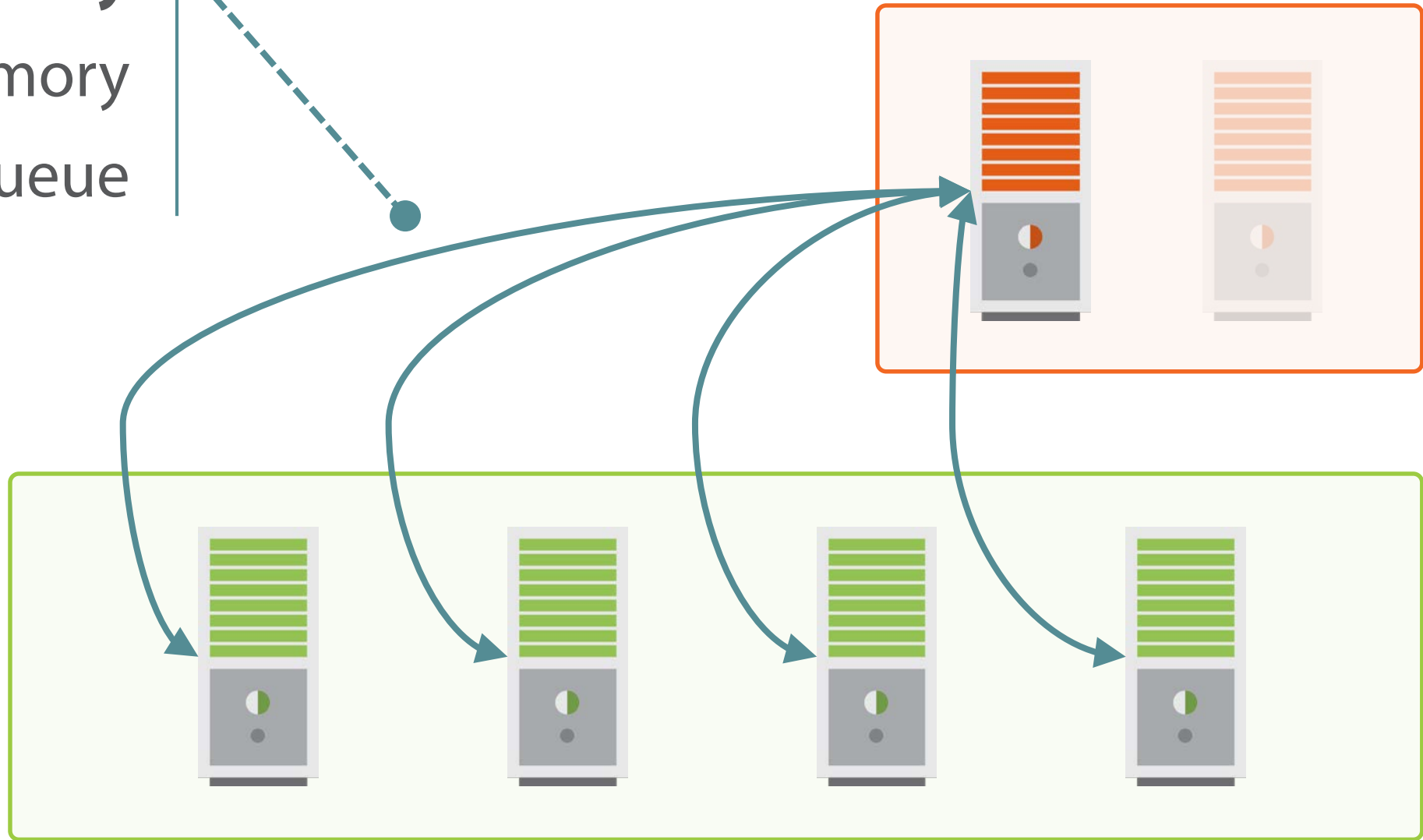
Zookeeper



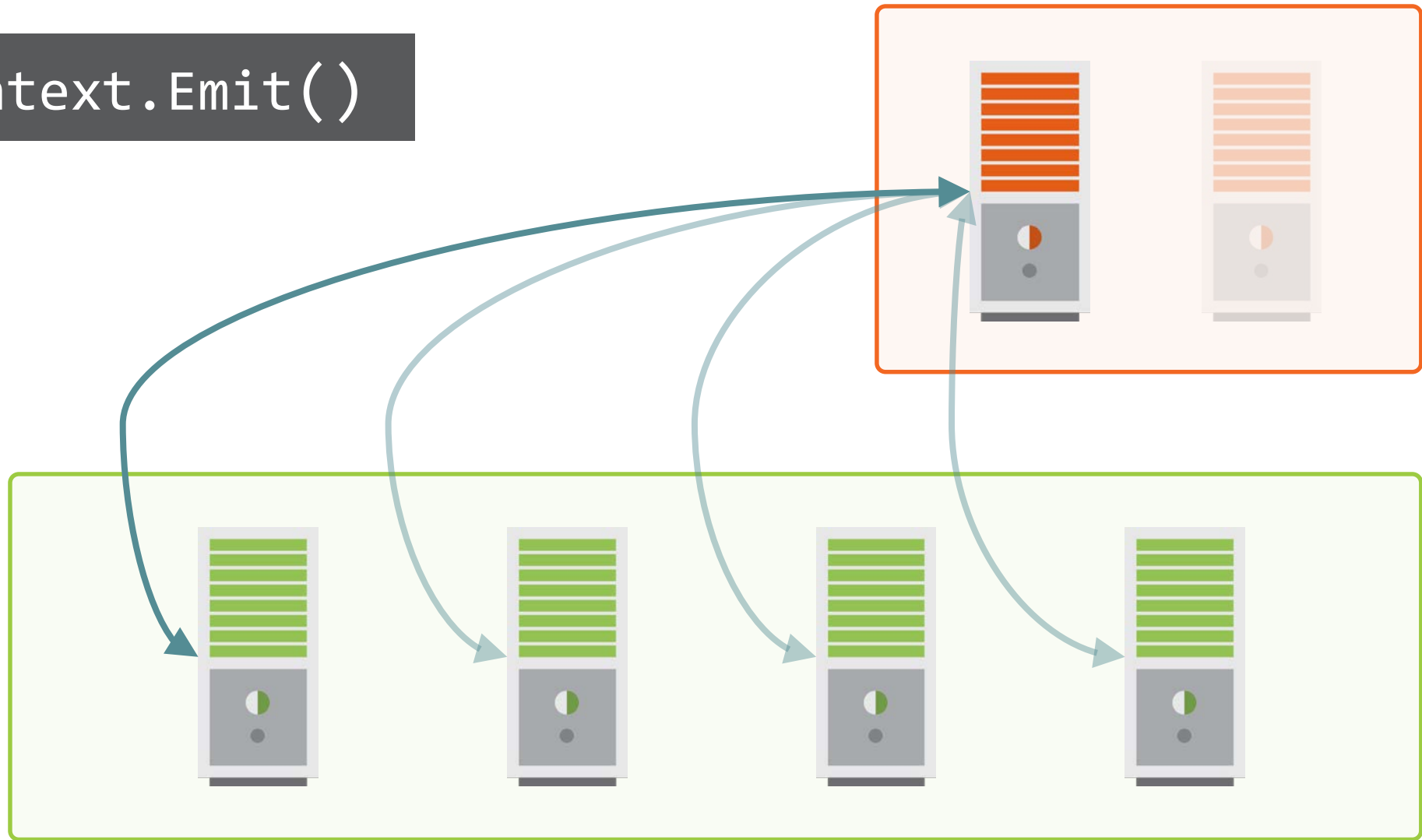
Configuration  
Persisted state  
Change notification



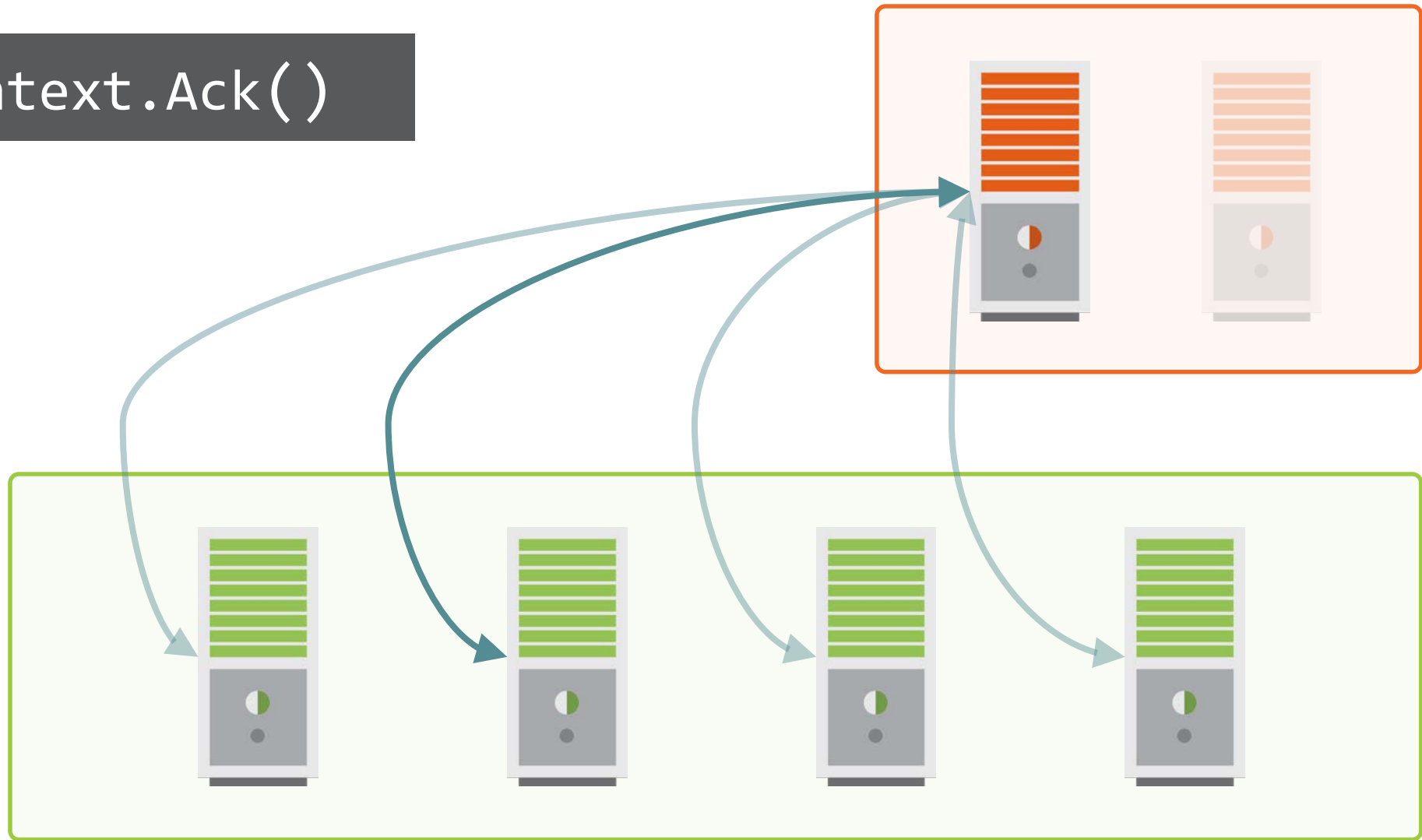
**Netty**  
In-memory  
Message queue



```
_context.Emit()
```

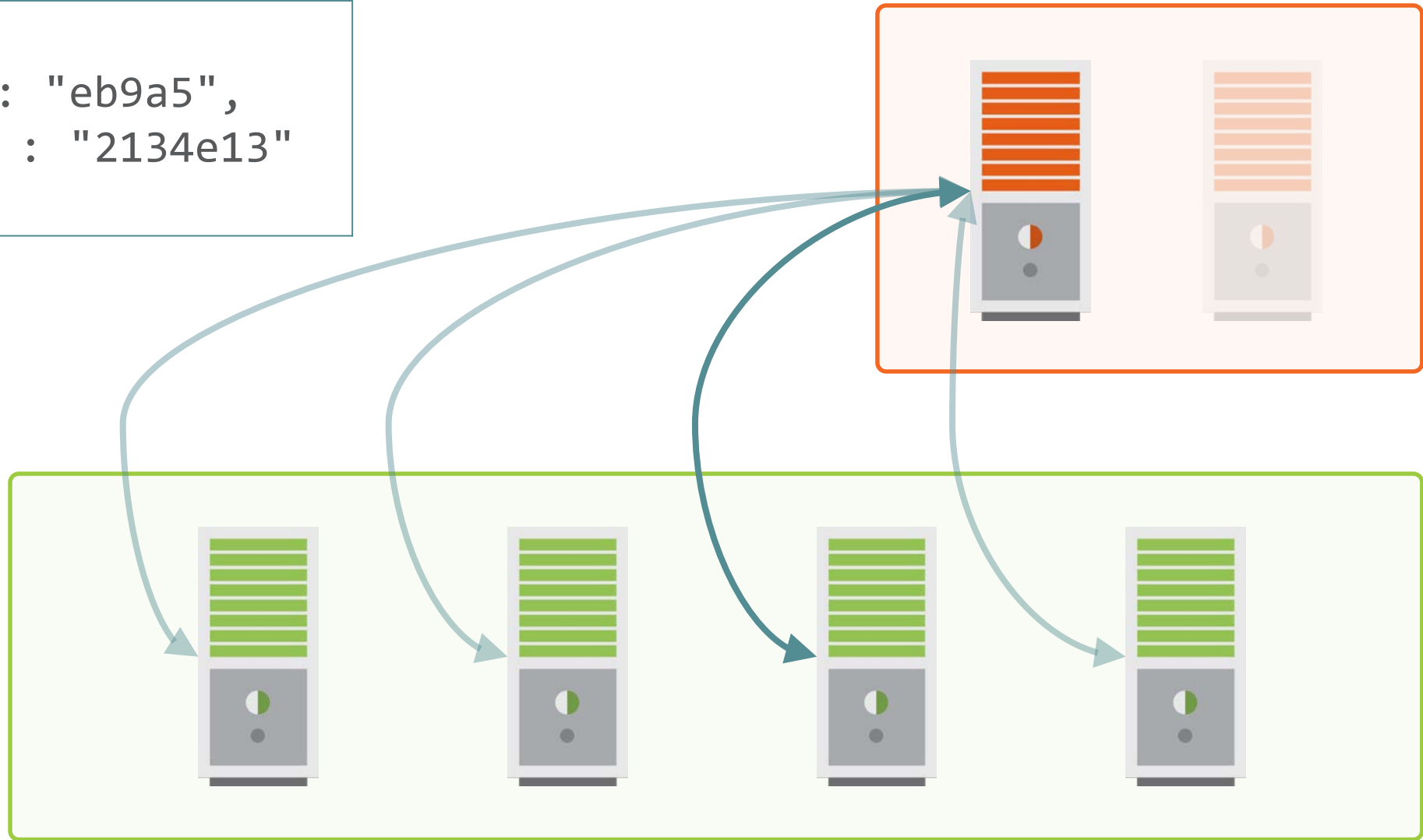


```
_context.Ack()
```



## Tuple

```
{  
  "raceId" : "eb9a5",  
  "racerId" : "2134e13"  
}
```



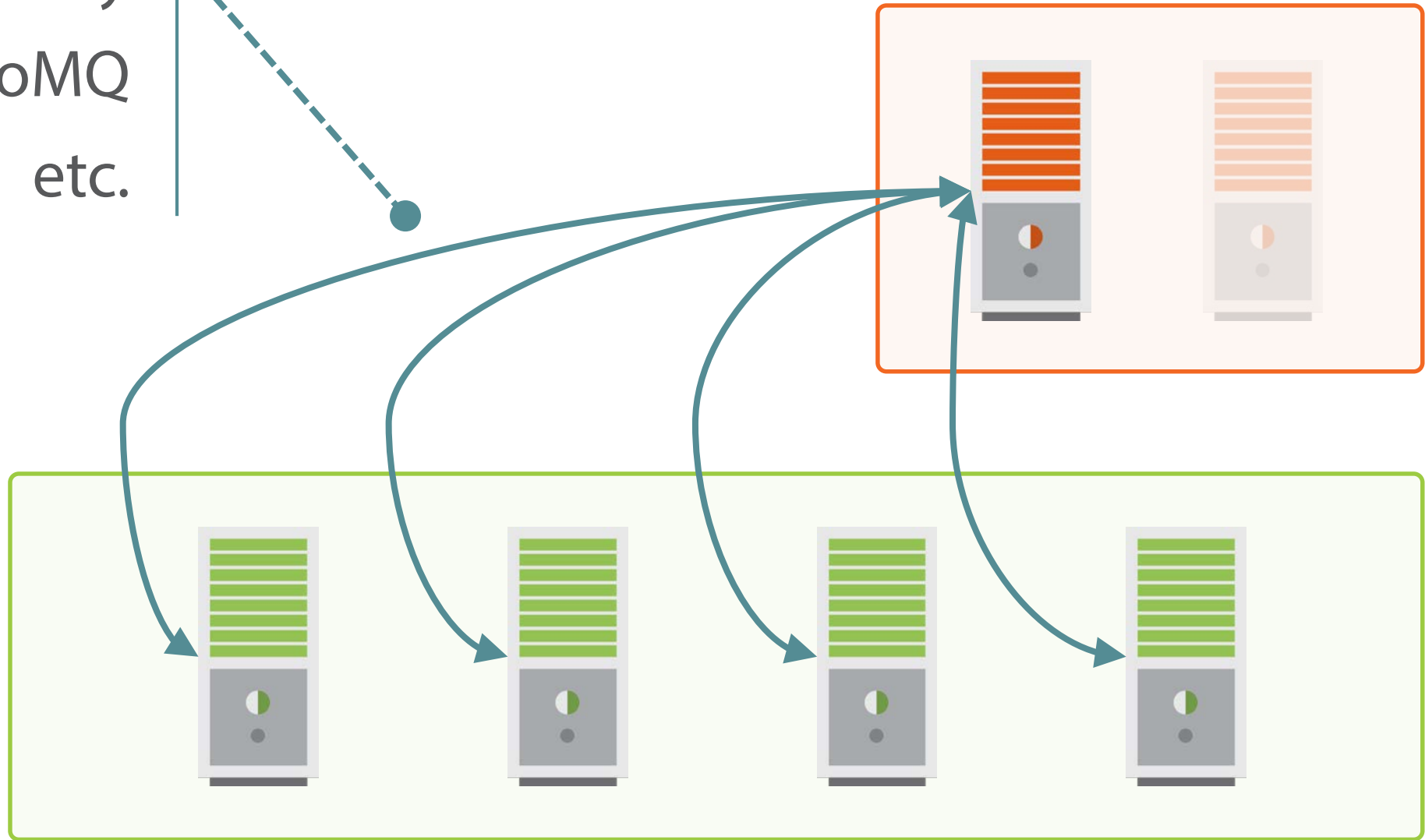


## Tuple

```
{  
  "raceId" : "eb9a5",  
  "racerId" : "2134e13"  
}
```

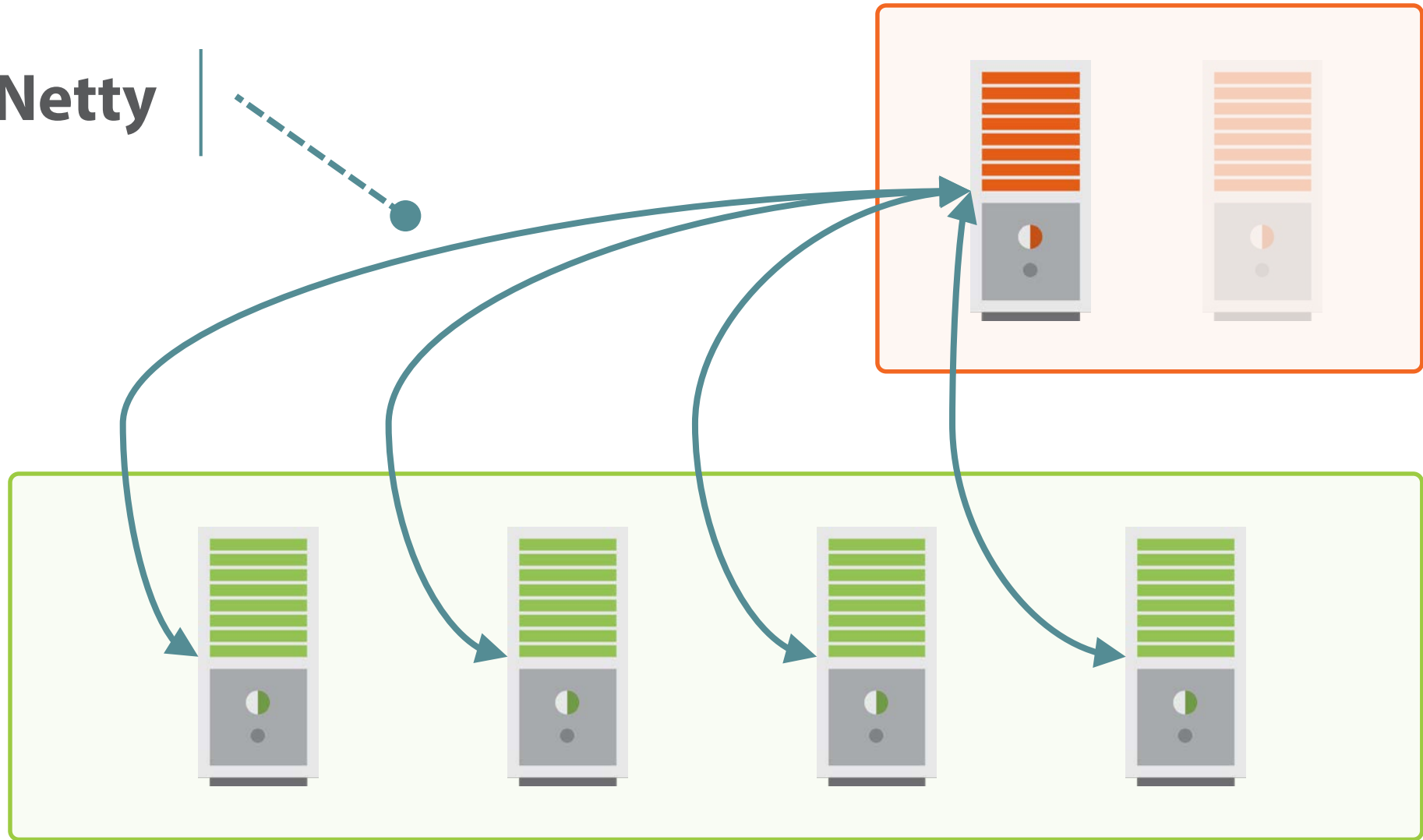


Netty  
ZeroMQ  
etc.





Usually **Netty**





# .NET Message Queue Fundamentals

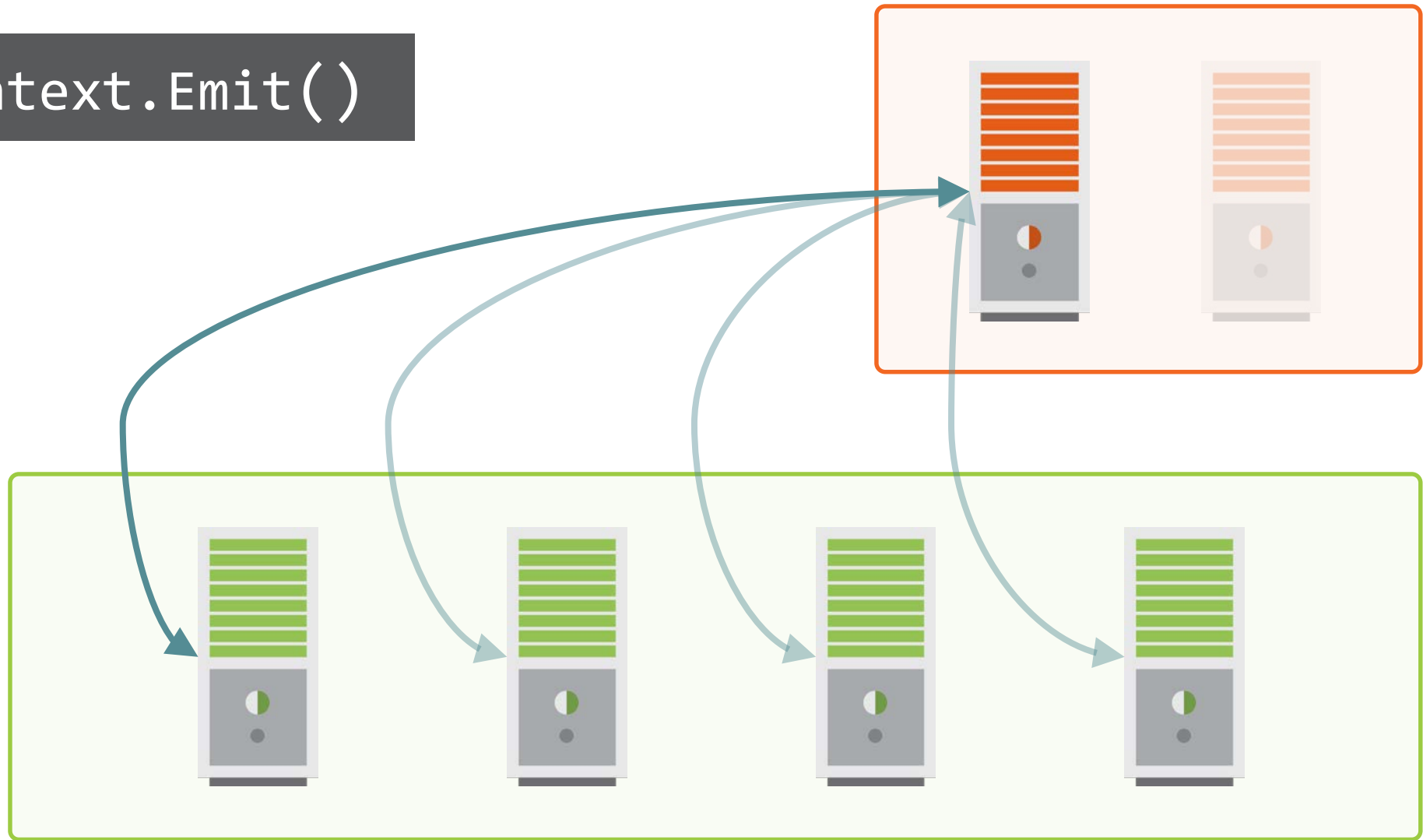
---

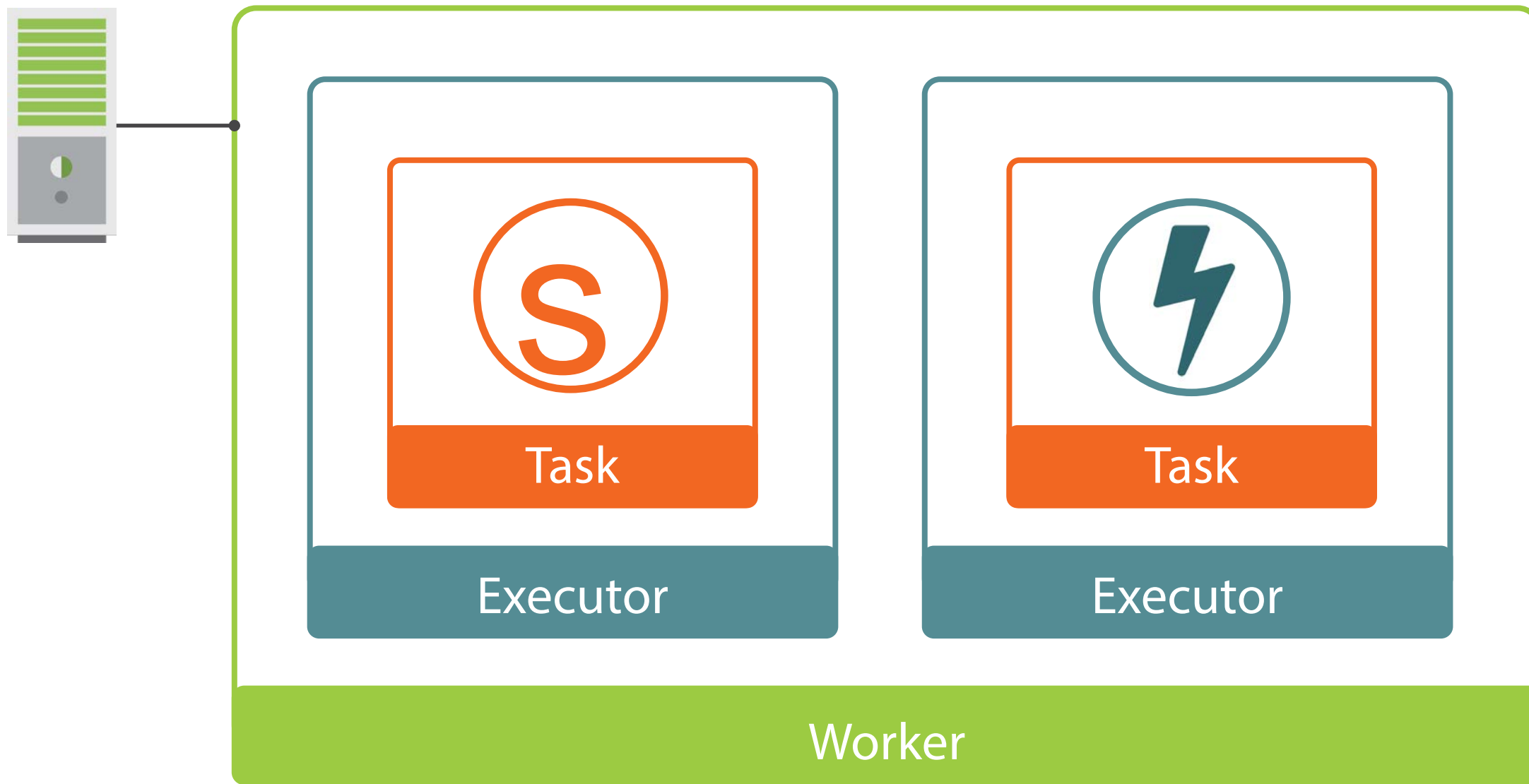


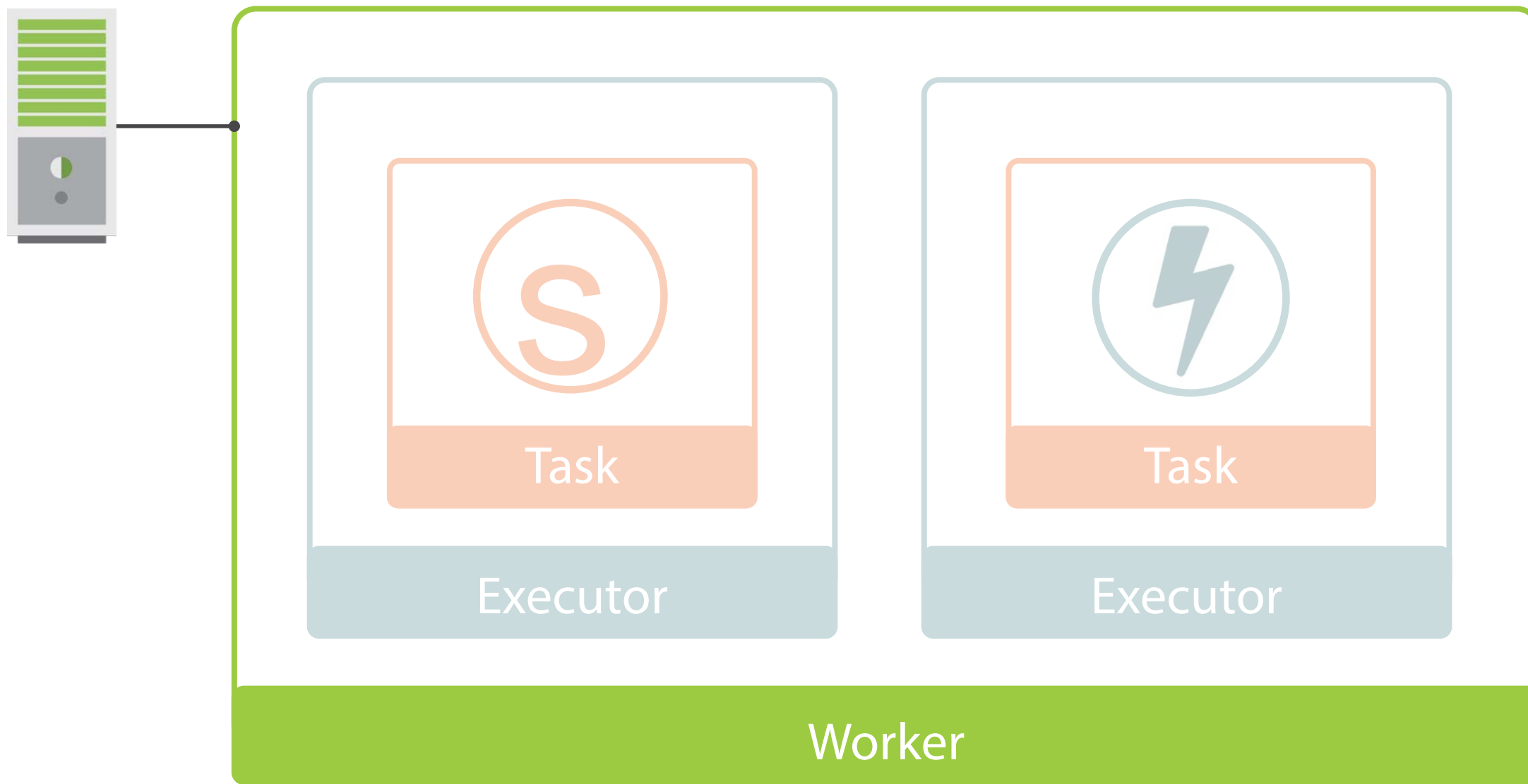
Elton Stoneman

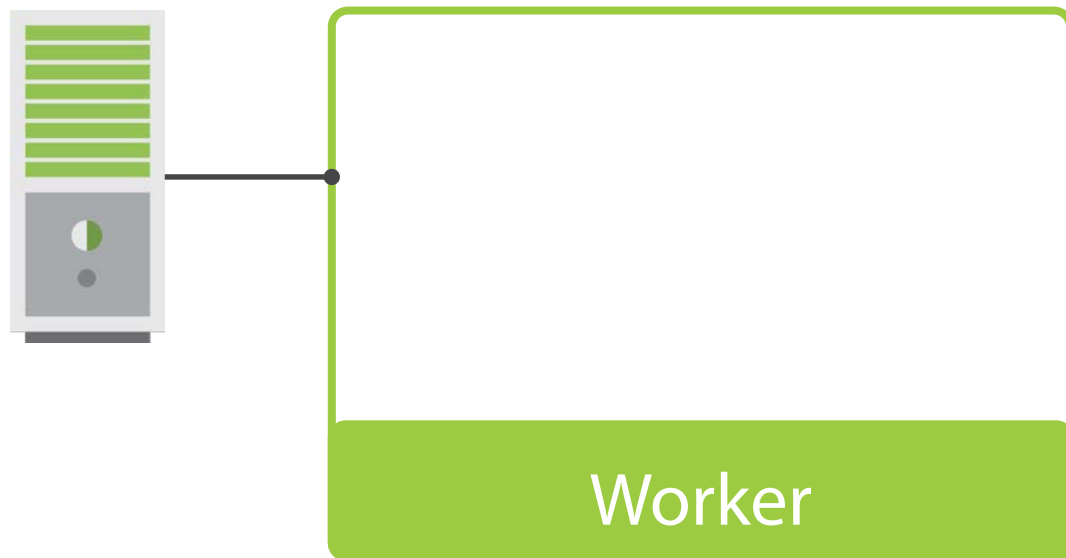
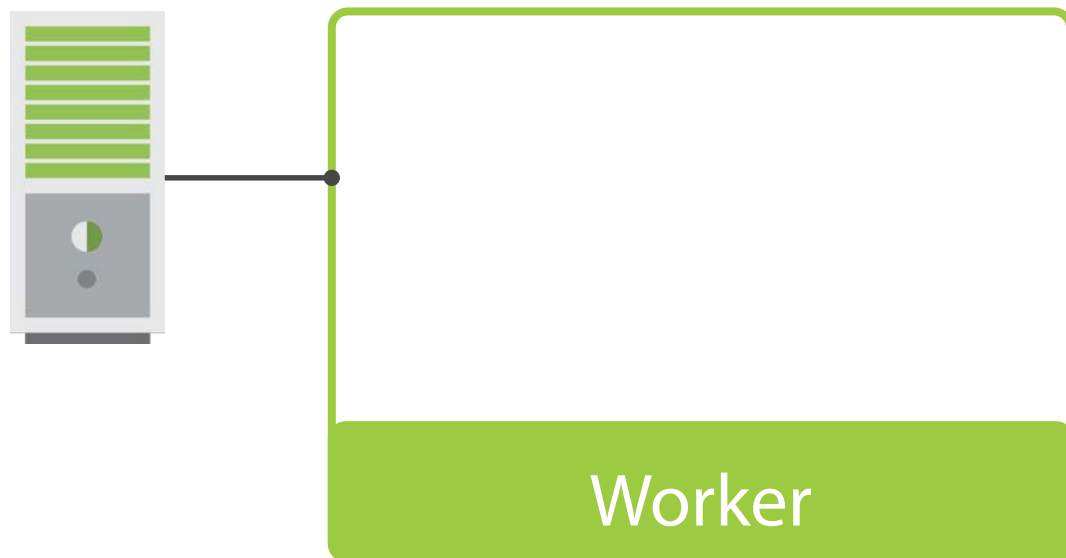
@EltonStoneman | [blog.sixeyed.com](http://blog.sixeyed.com)

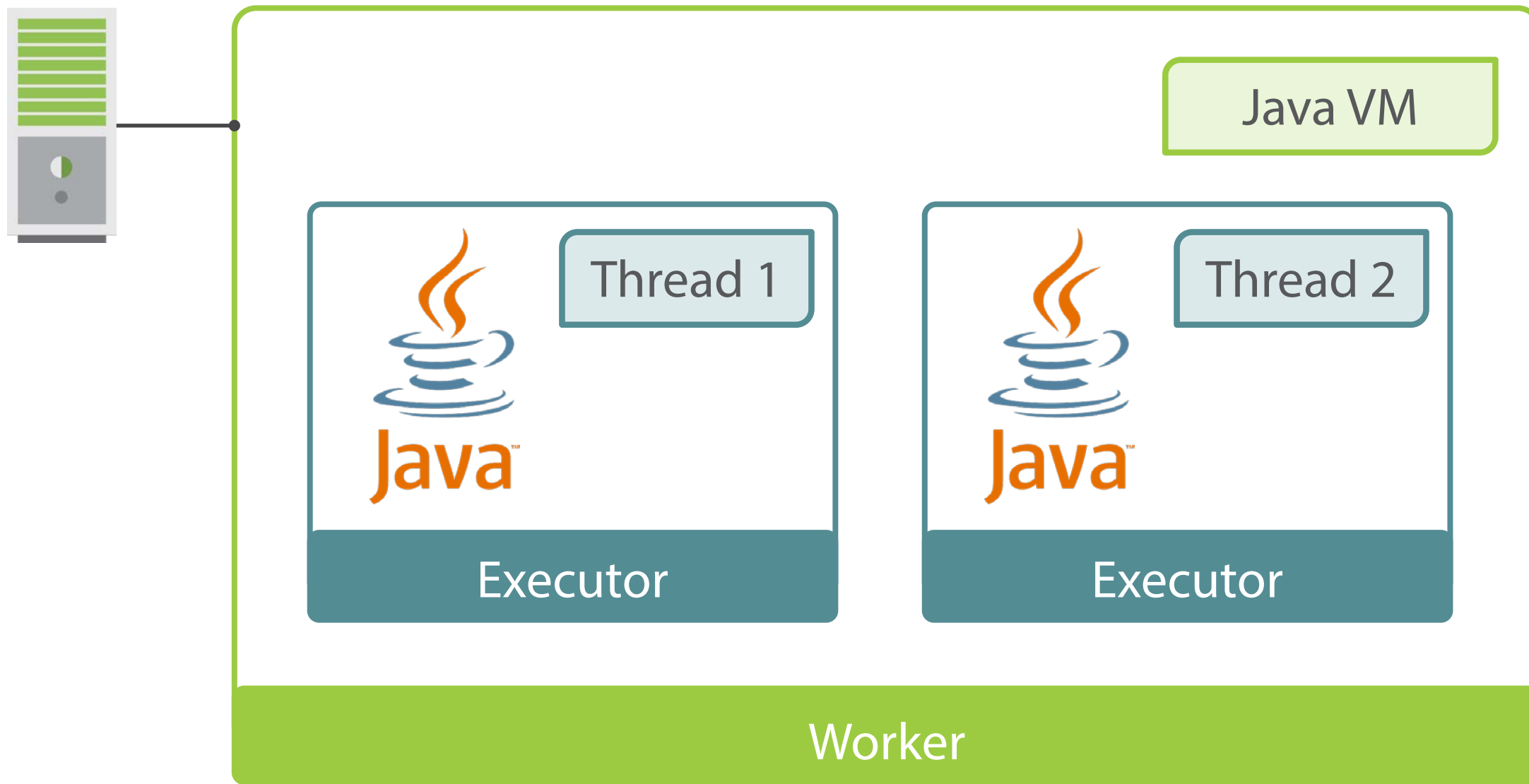
```
_context.Emit()
```

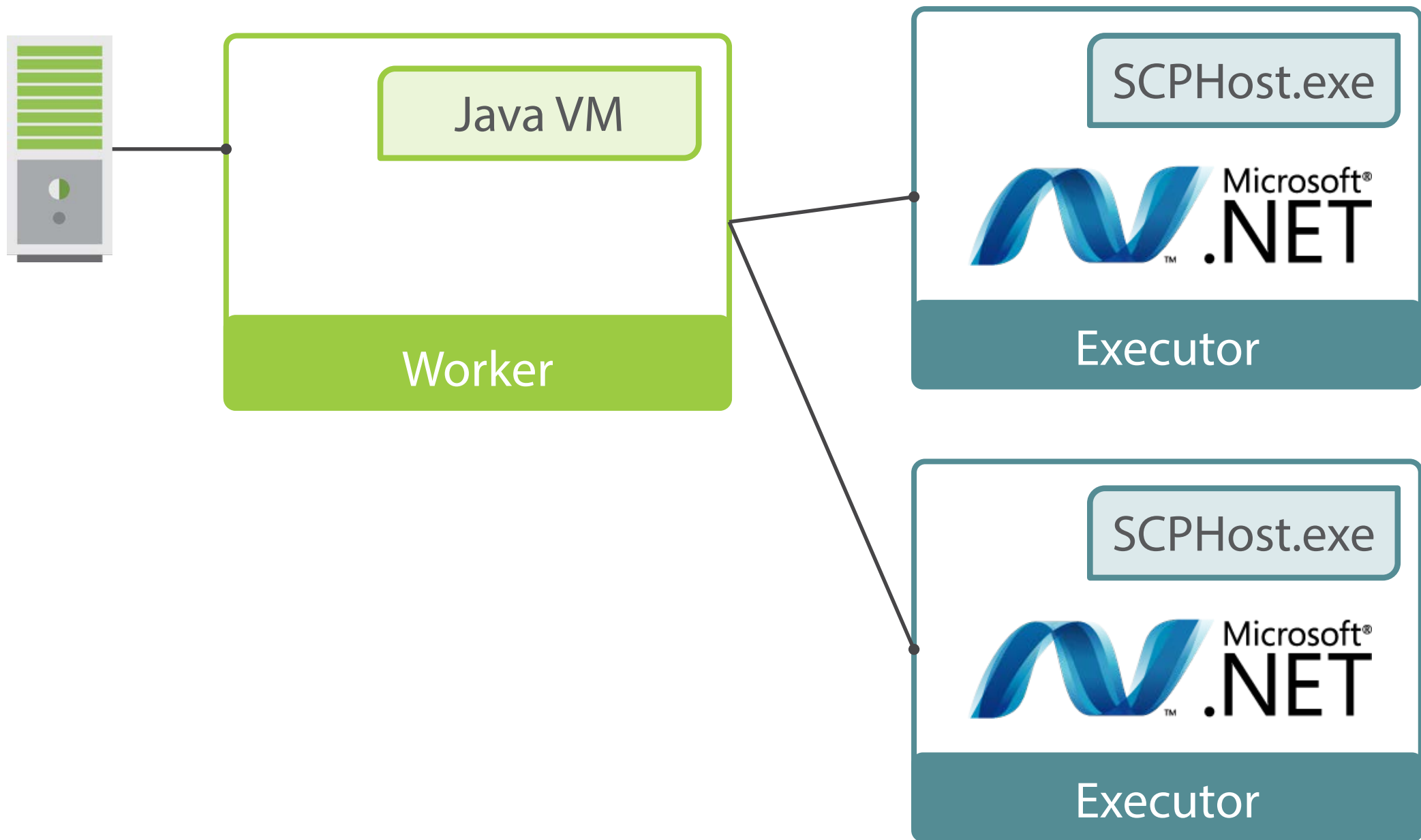
















Task



Task



Task

Executor



Task



Task

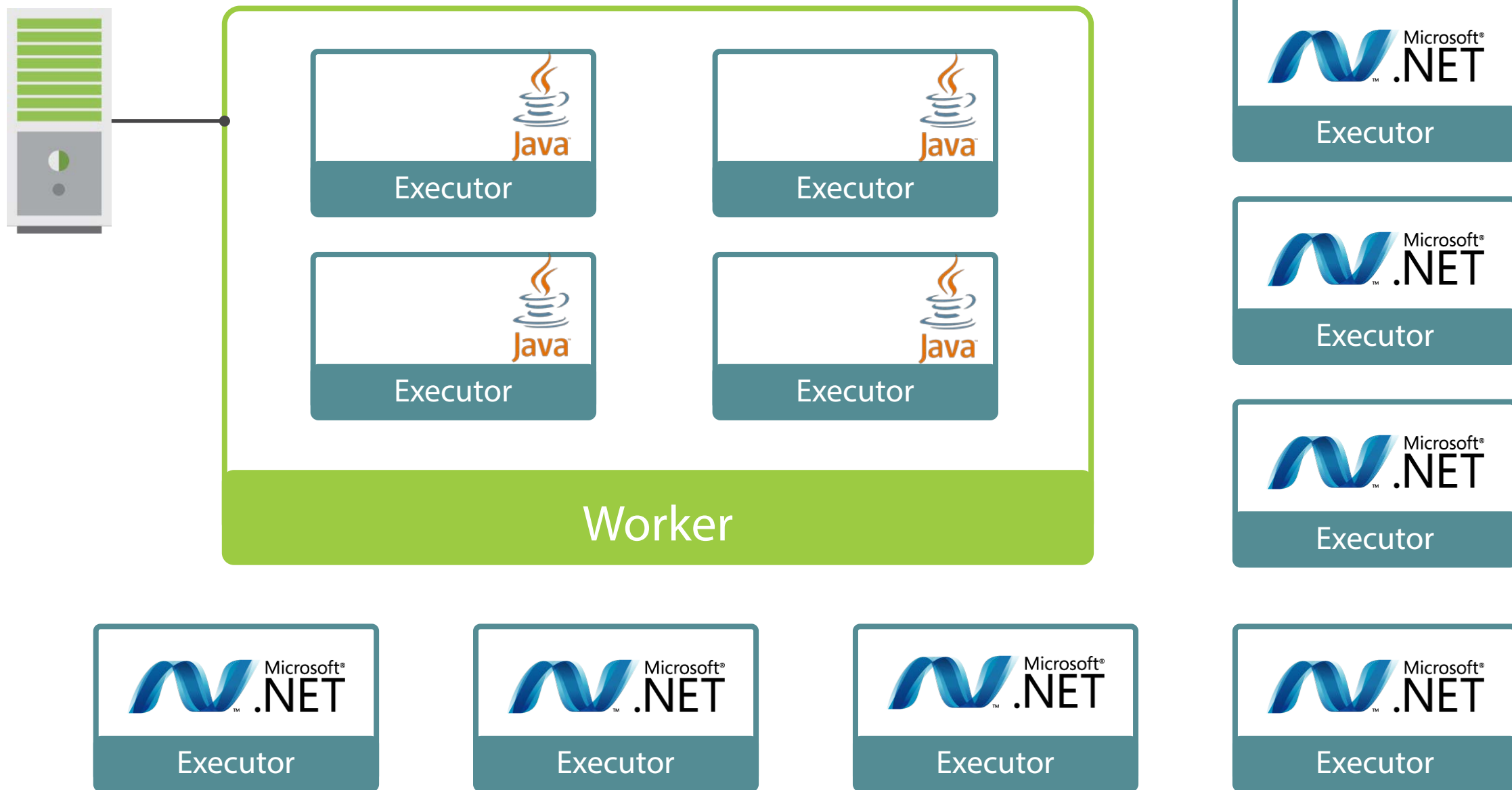
Executor

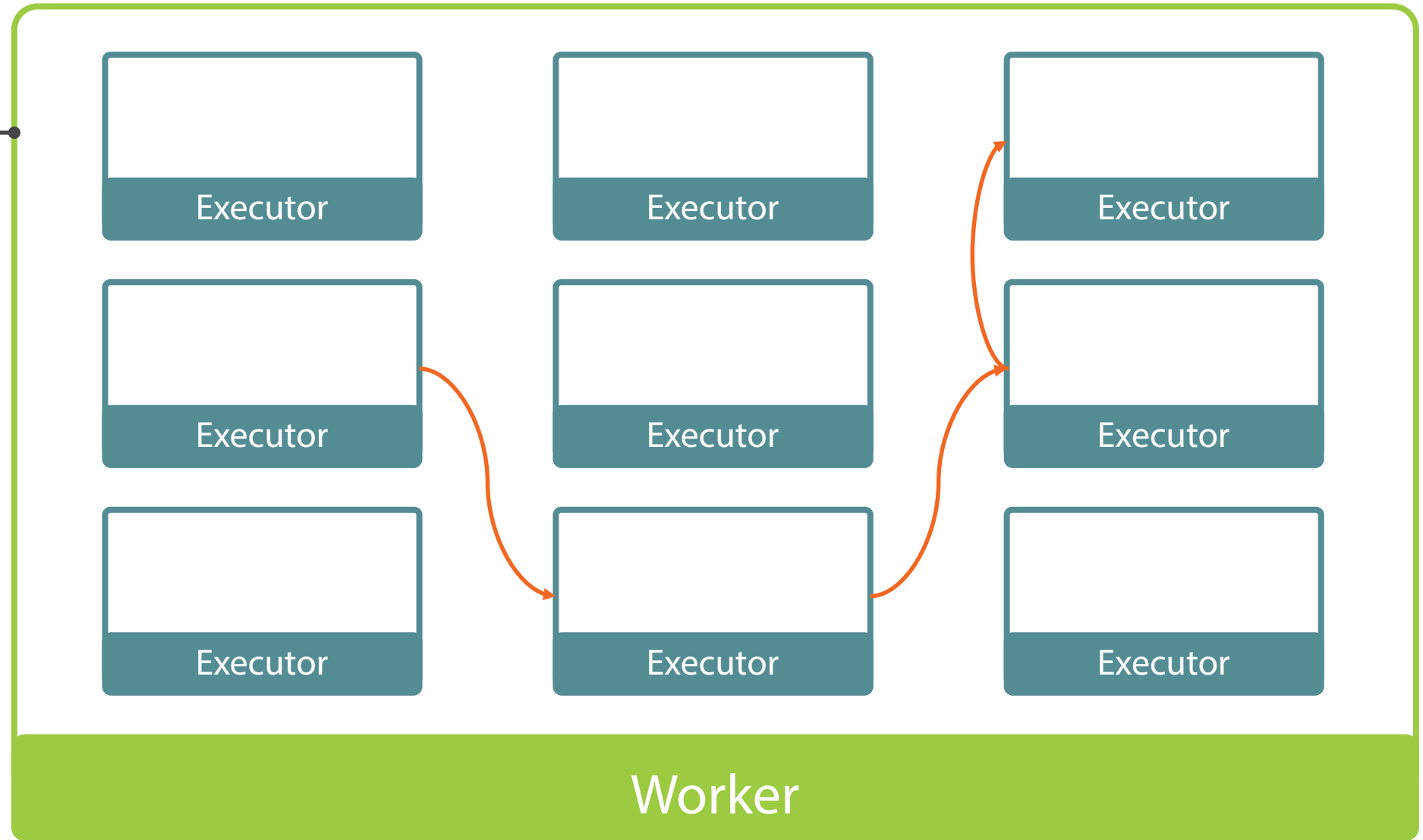


Task

Executor

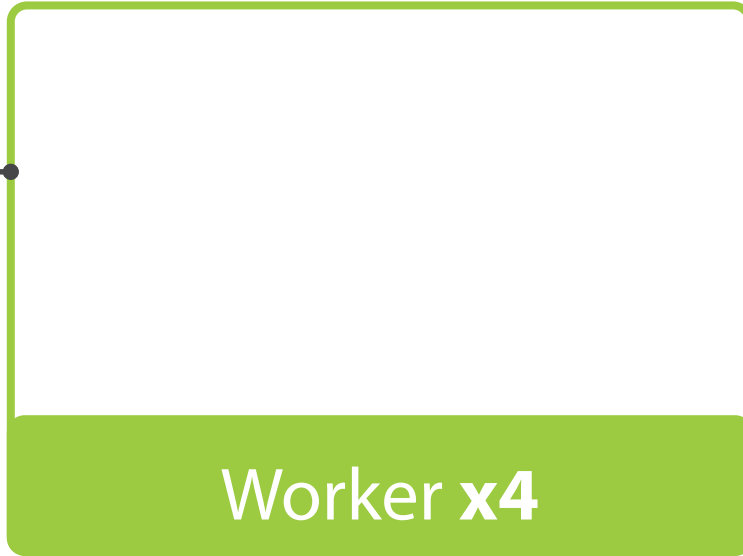








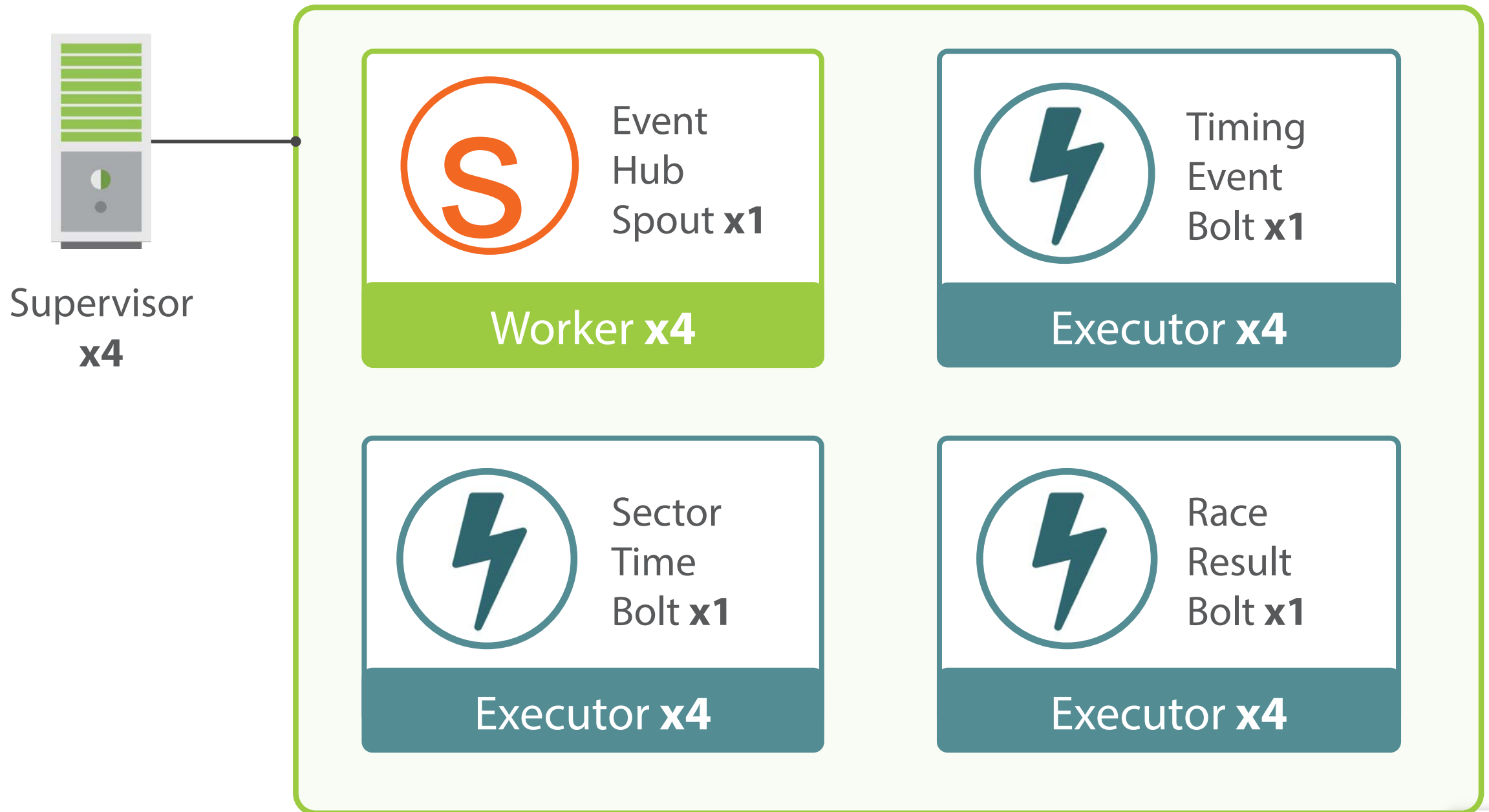
Supervisor  
**x4**

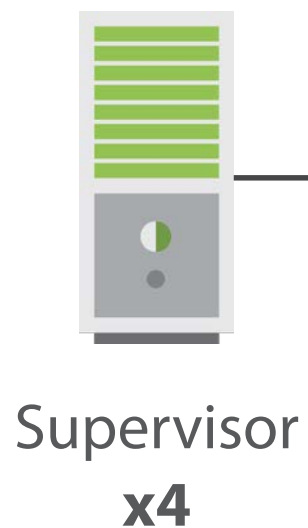


Worker **x4**









Java.exe  
Java.exe  
Java.exe  
Java.exe

Event  
Hub  
**Spout**

SCPHost.exe  
SCPHost.exe  
SCPHost.exe  
SCPHost.exe

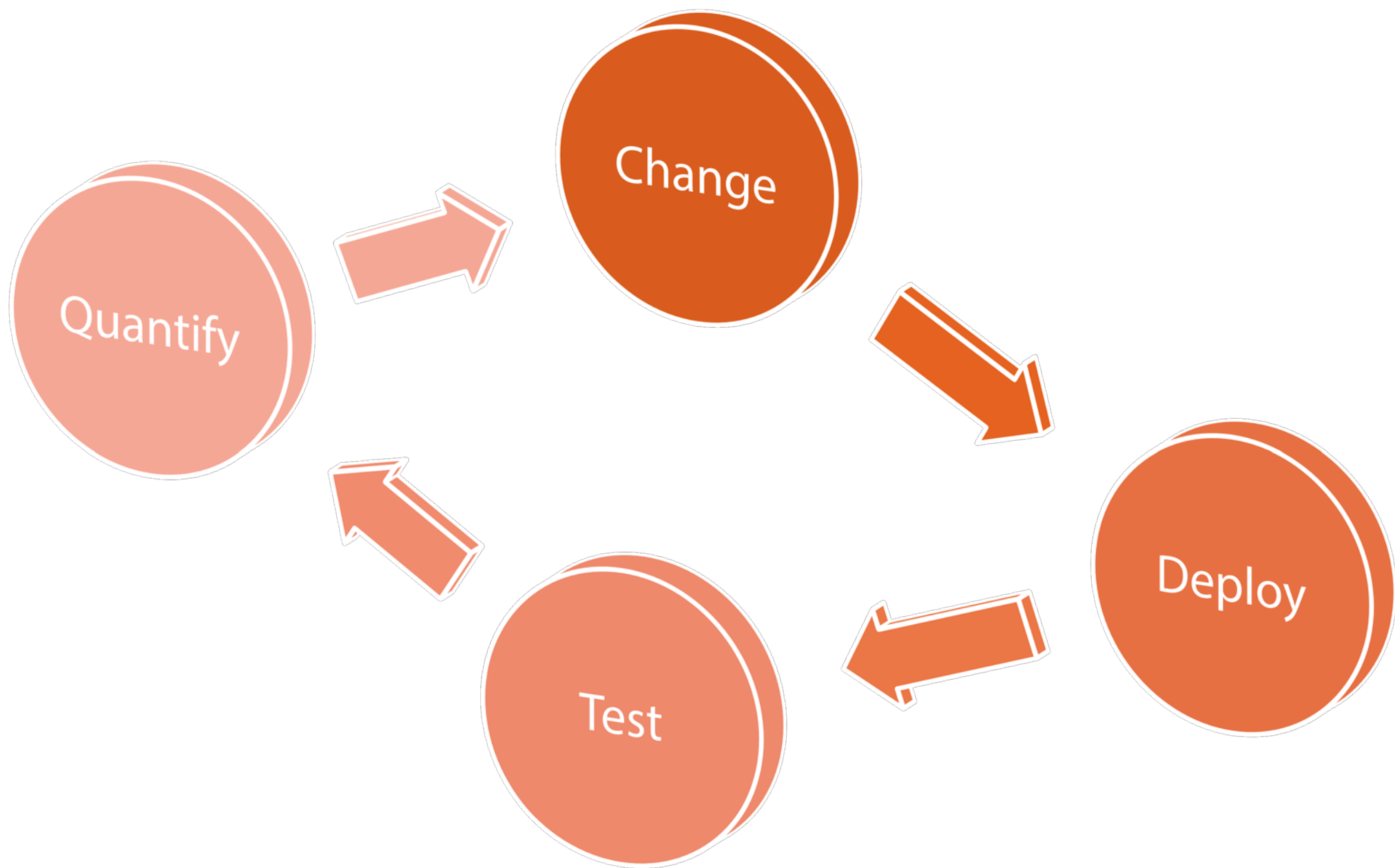
Timing  
Event  
**Bolt**

SCPHost.exe  
SCPHost.exe  
SCPHost.exe  
SCPHost.exe

Sector  
Time  
**Bolt**

SCPHost.exe  
SCPHost.exe  
SCPHost.exe  
SCPHost.exe

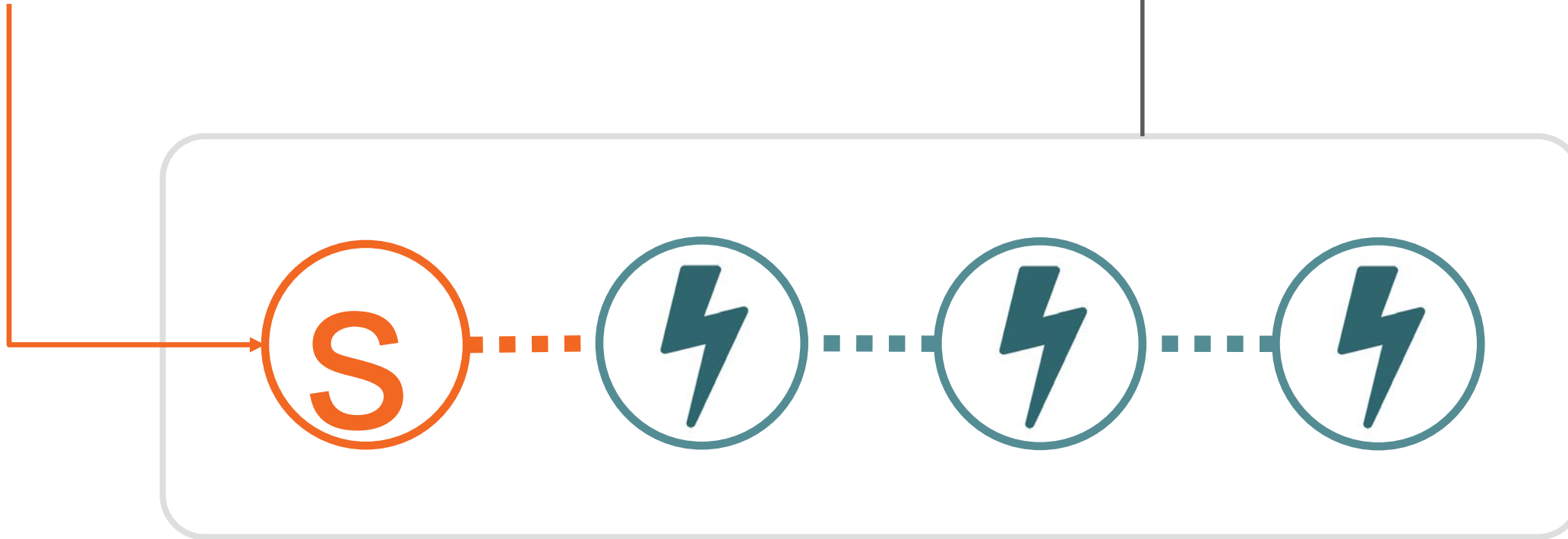
Race  
Result  
**Bolt**





**Topology v1**

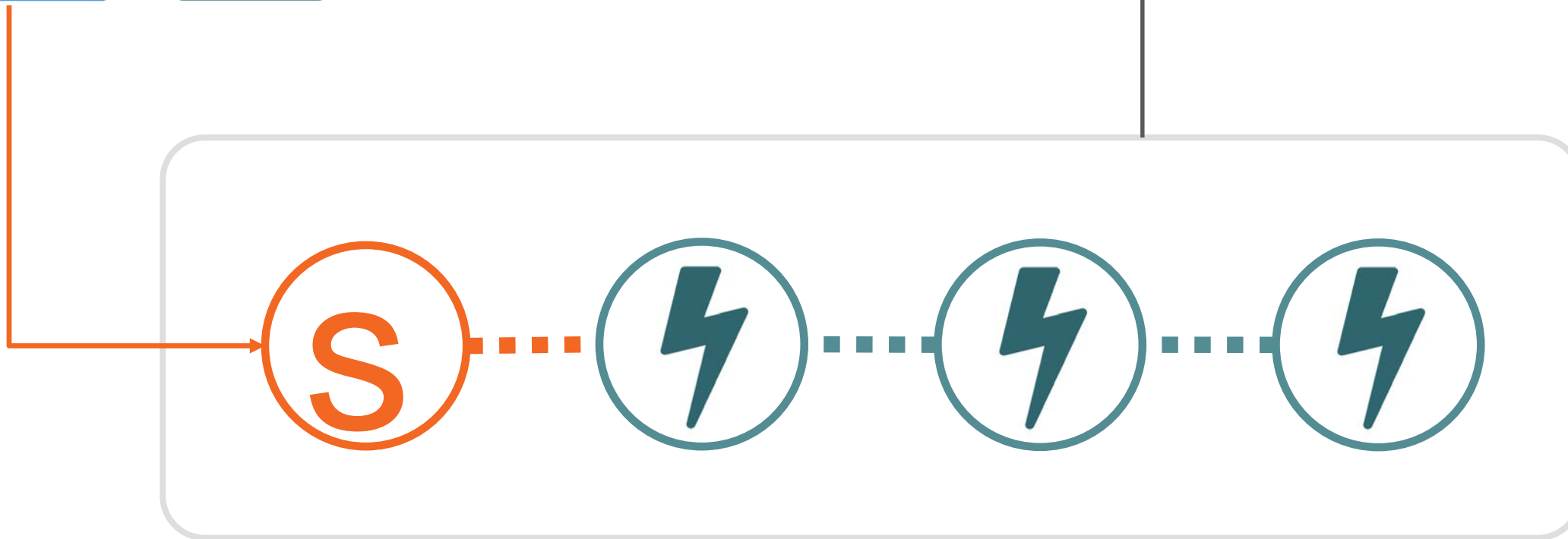
Last processed #400,000





**Topology v2**

Last processed #0

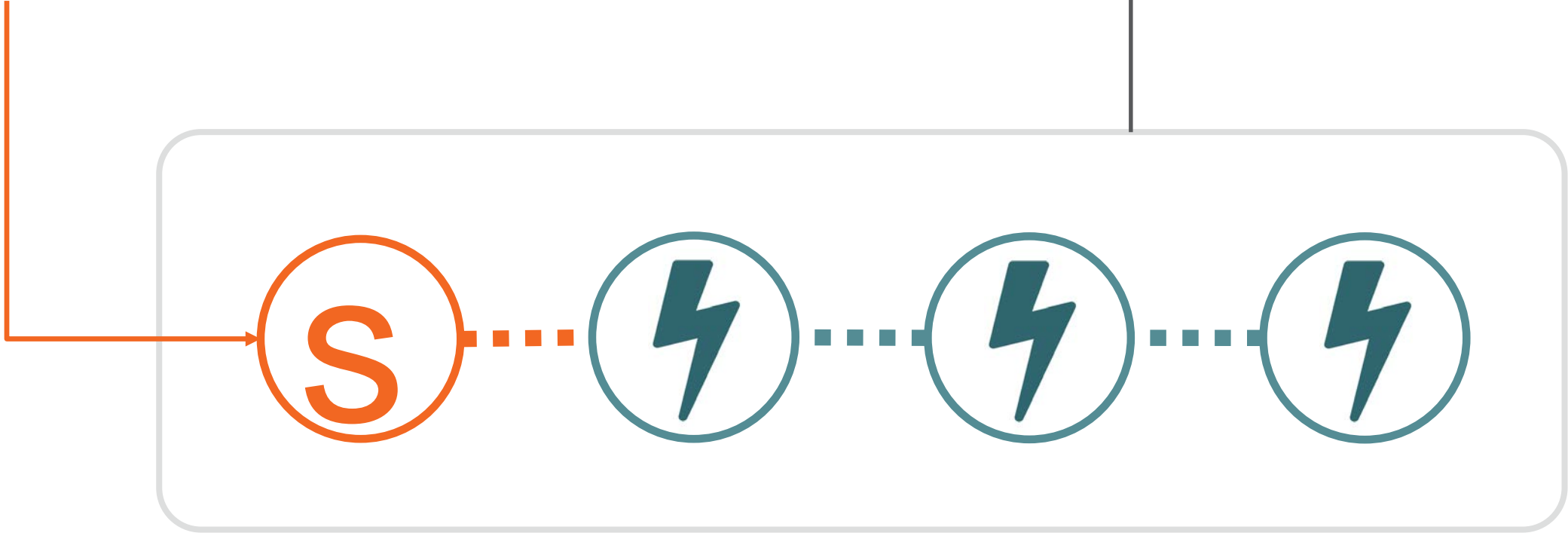




**Setup**  
Fill Event Hub

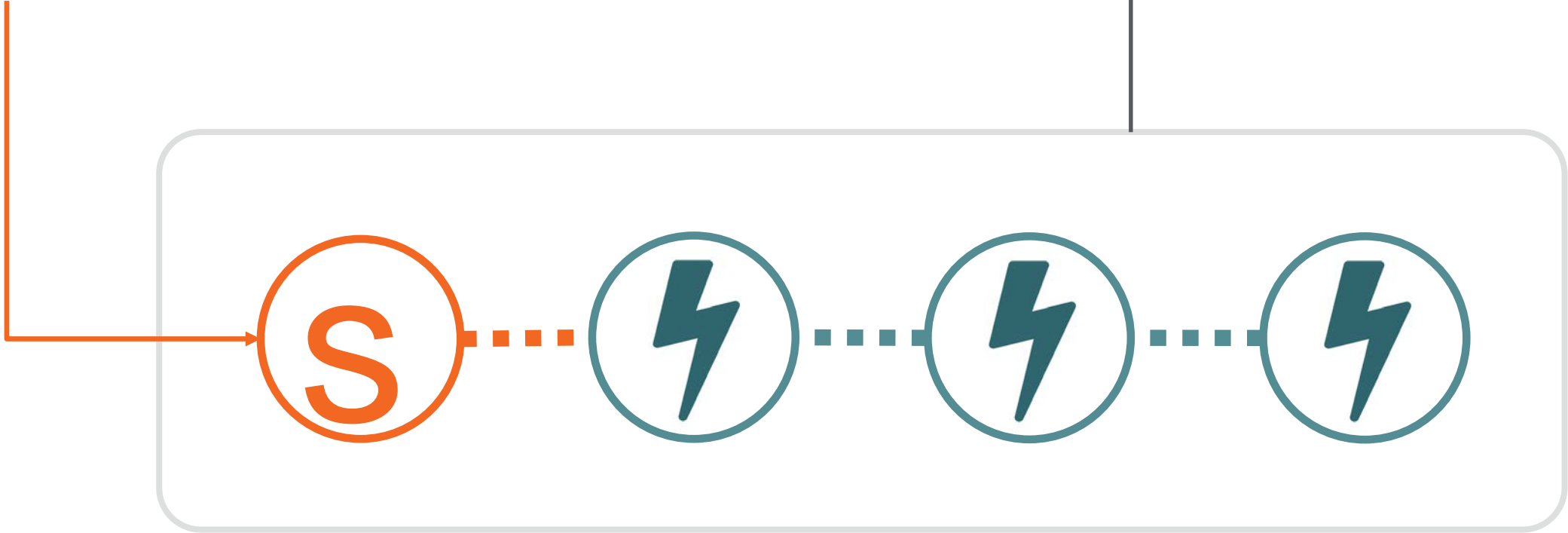


**Deploy**  
Topology v1

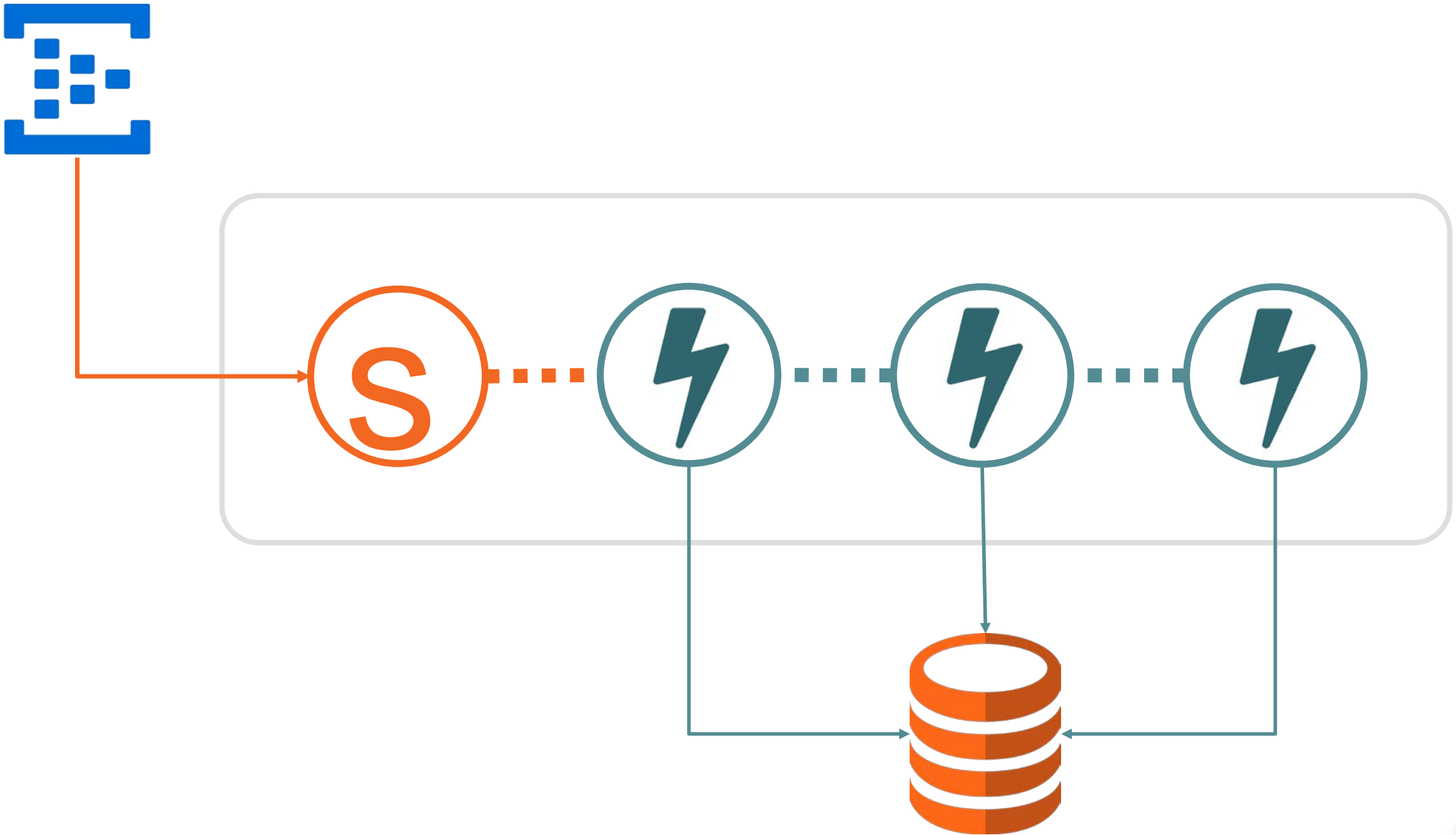


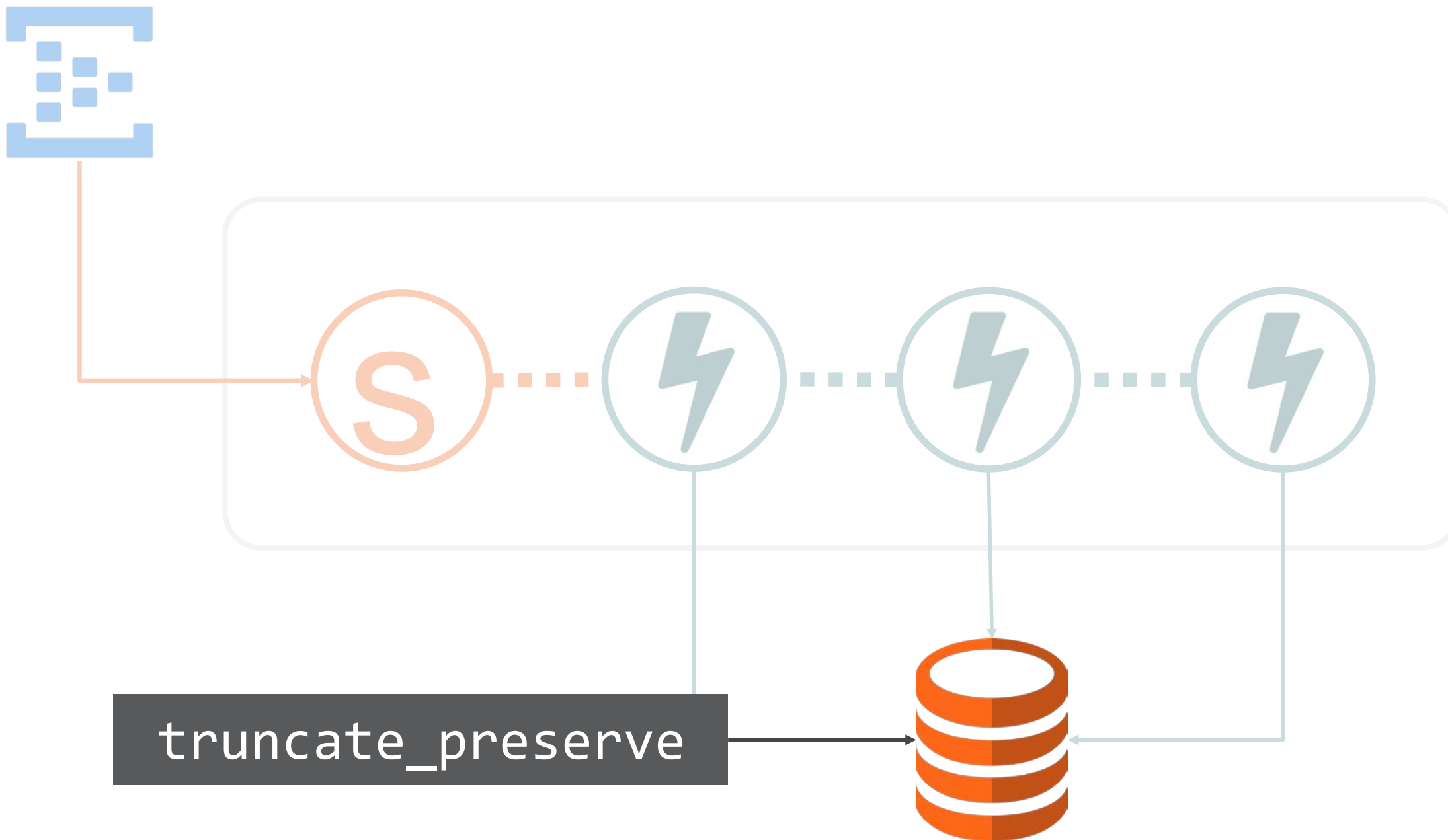


**Change**  
Topology v2





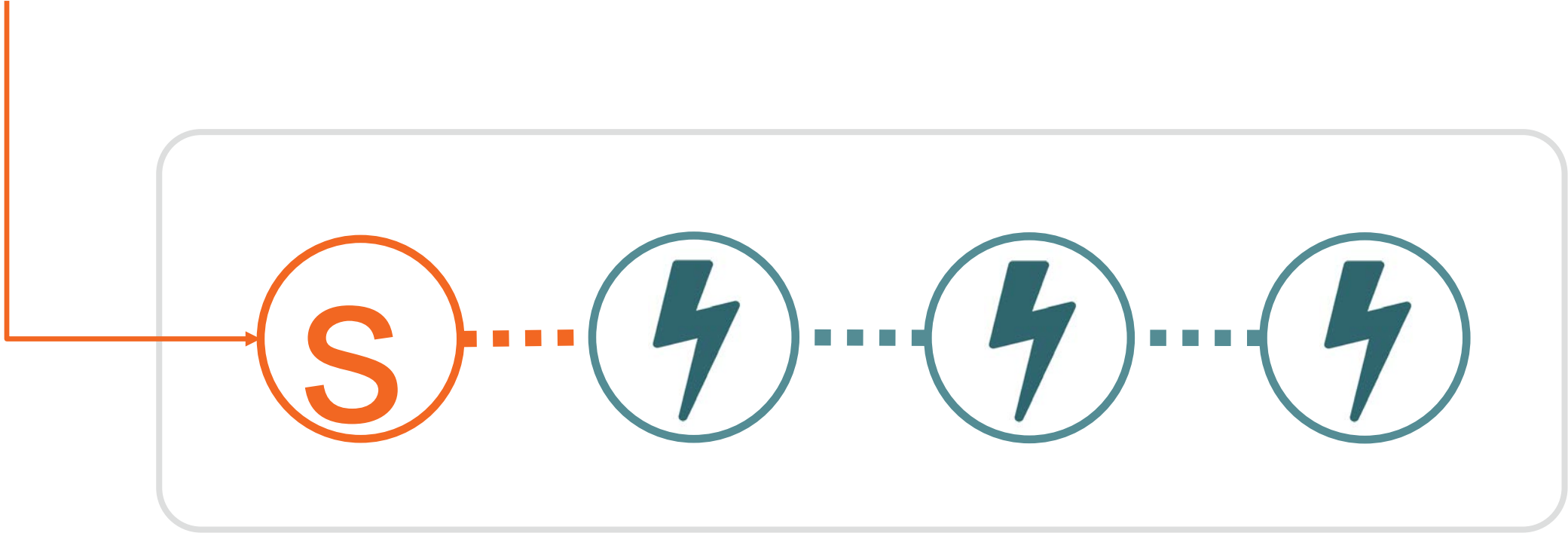


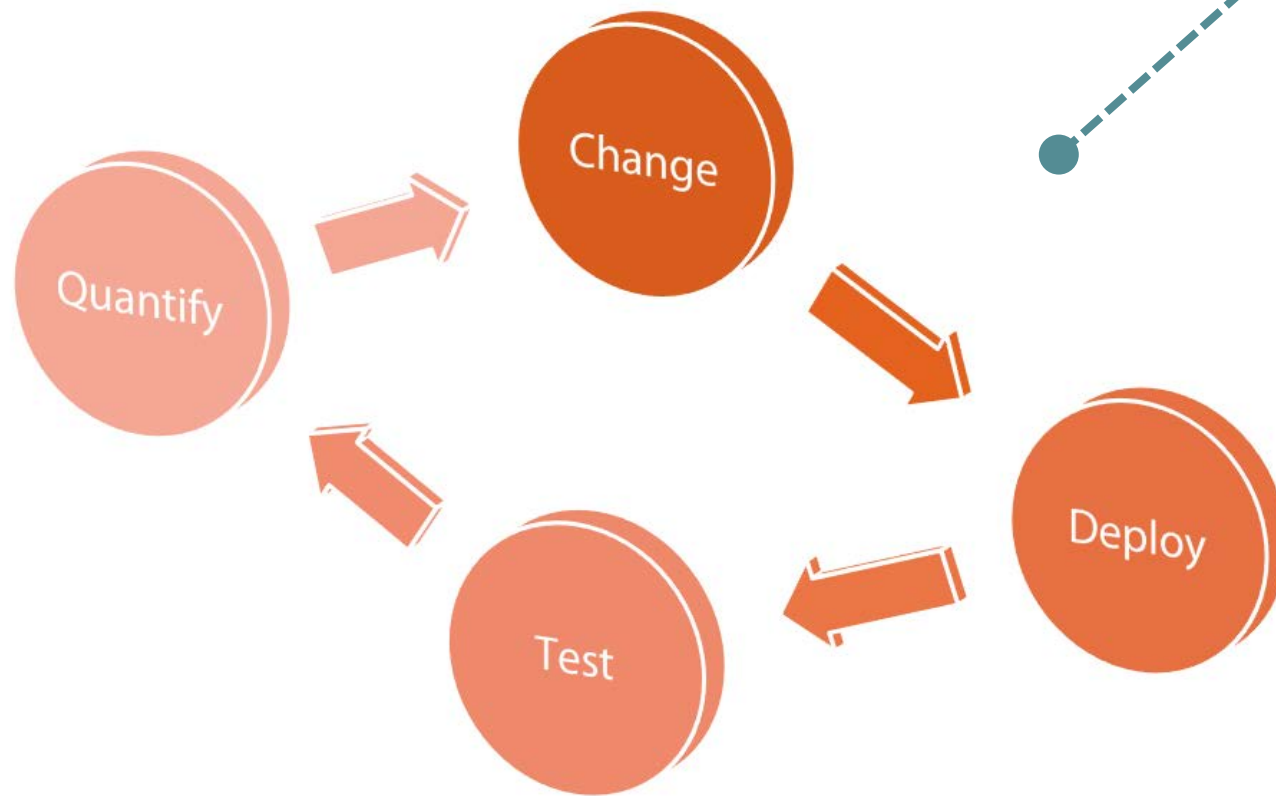




## Setup

Delete, re-create & fill Event Hub





**20min+** cycle

**>720** permutations

**70%** improvement?



Event Hub Spout  
**x16**



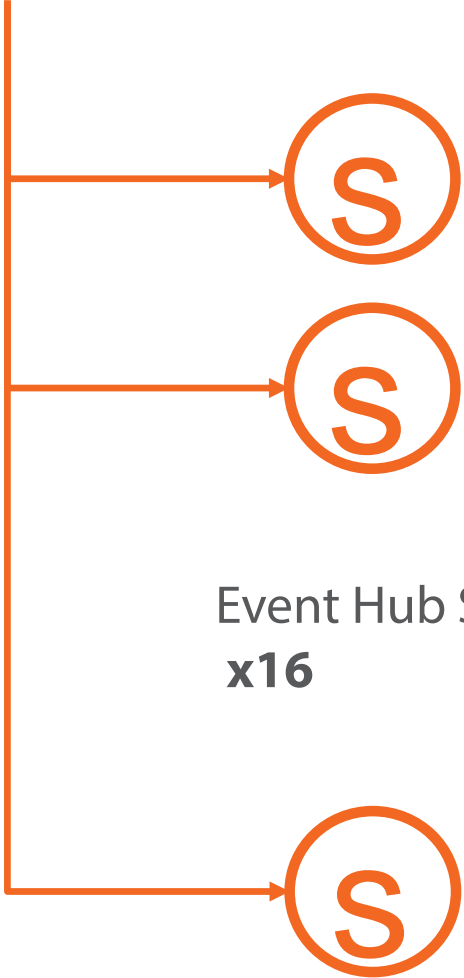
Timing Event Bolt  
**x32**



Sector Time Bolt  
**x16**



Race Result Bolt  
**x8**



Event Hub Spout  
**x16**



Timing Event Bolt  
**x32**



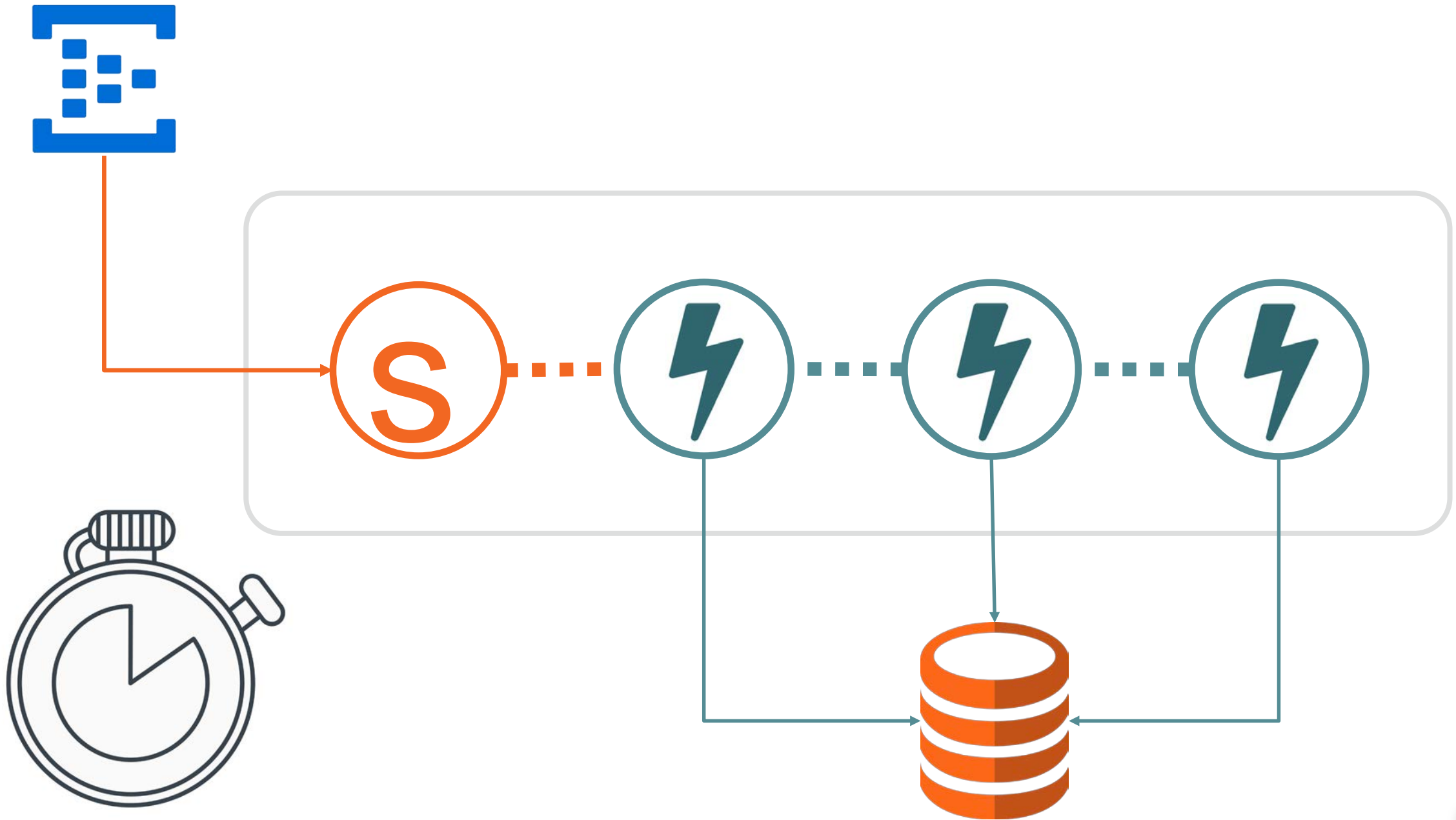
Batch Sector  
Time Bolt  
**x16**



Update Sector  
Time Bolt  
**x16**



Race Result Bolt  
**x8**



## Spouts (All time)

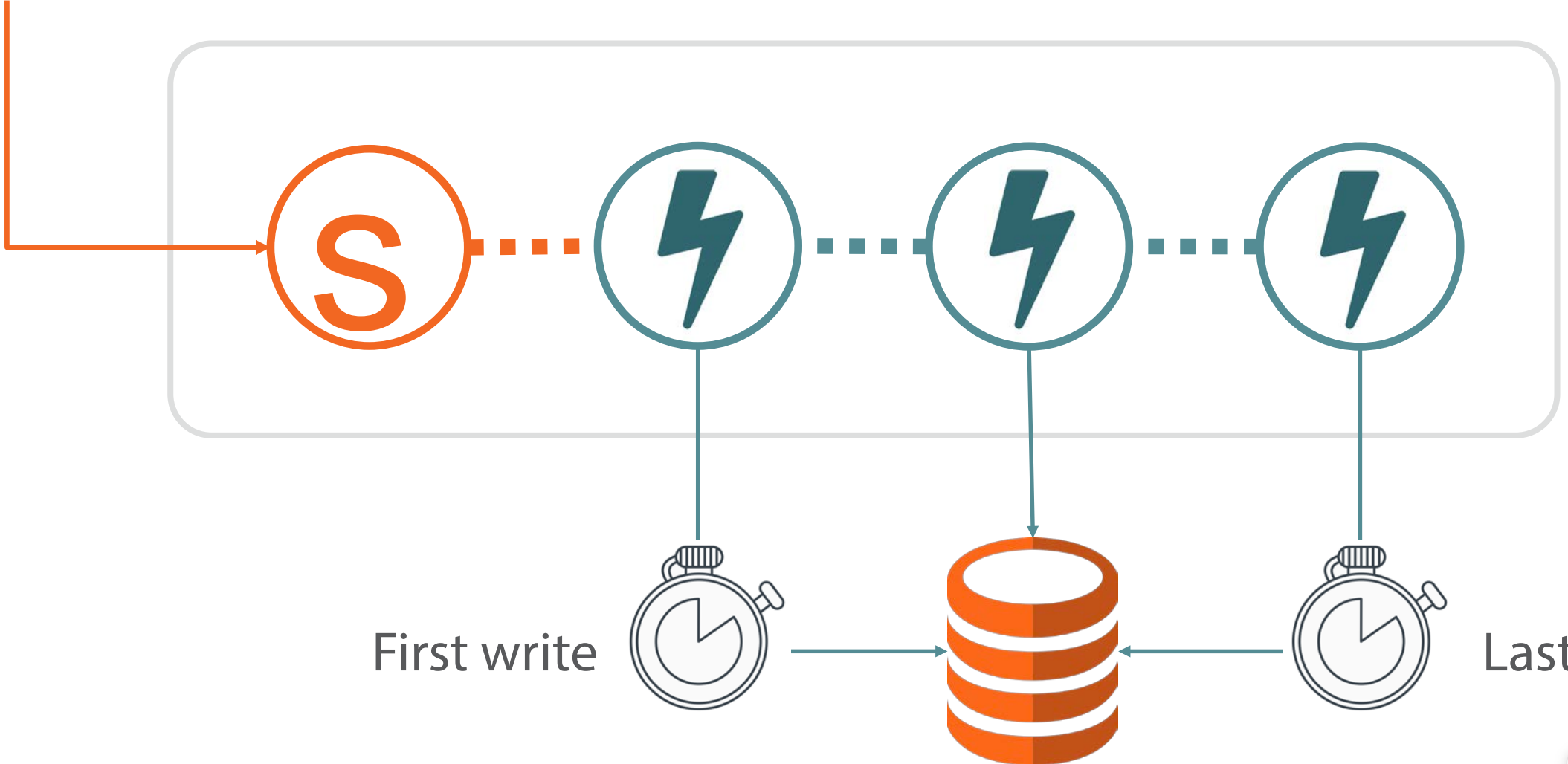
Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error
EventHubSpout	16	16	8160	8160	0.000	0	0			

## Bolts (All time)

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error
__acker	4	4	0	0	0.001	0.038	1060	0.019	1080	0			
__metricsbacktype.storm.metric.LoggingMetricsConsumer	1	1			0.000	0		0					
BatchSectorTimeBolt	64	64	20	0	0.000	0.000	200	0.000	0	0			
RaceResultBolt	32	32	0	0	0.000	0.000	0	0.000	0	0			
TimingEventBolt	256	256	880	780	0.342	13.705	880	20287.234	340	0			
UpdateSectorTimeBolt	128	128	0	0	0.000	0.000	0	0.000	0	0			

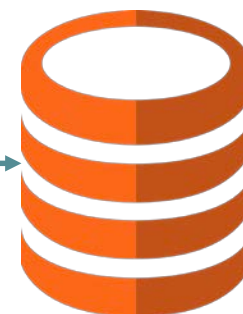








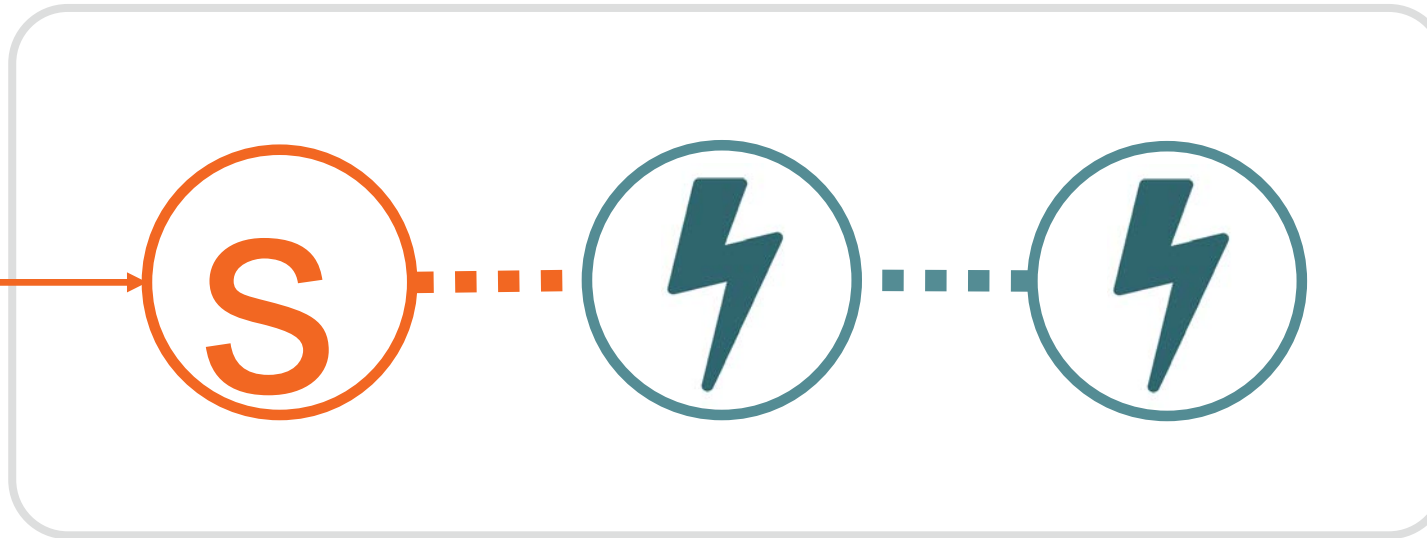
Baseline = **2.5K** msg/sec





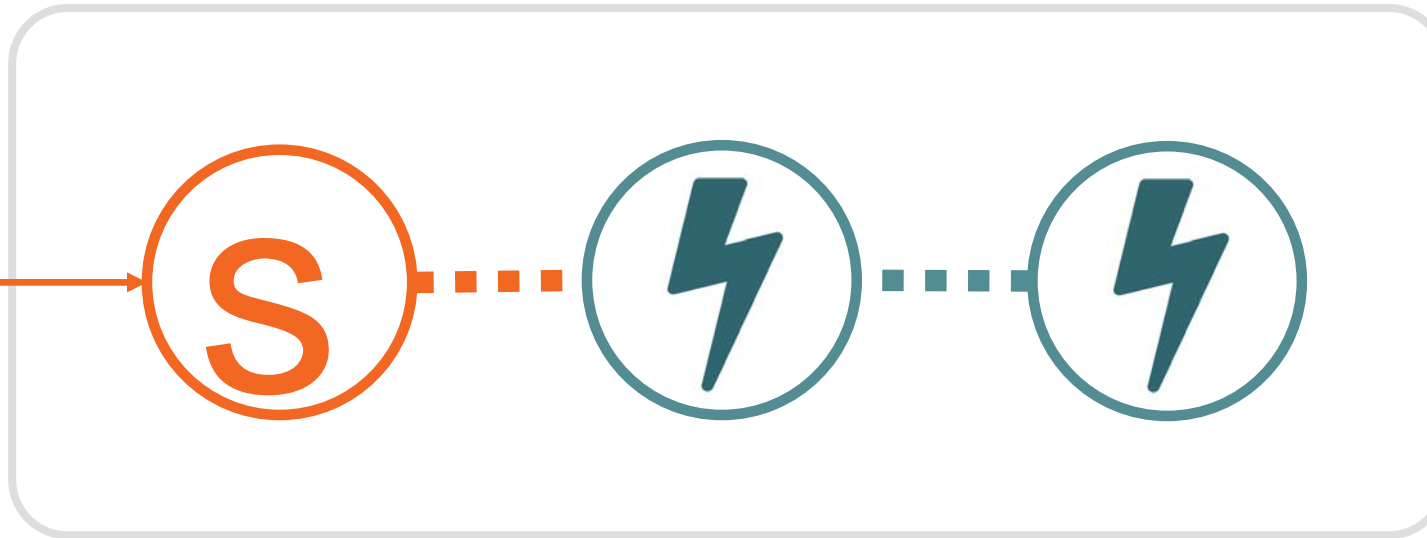
**32** partitions

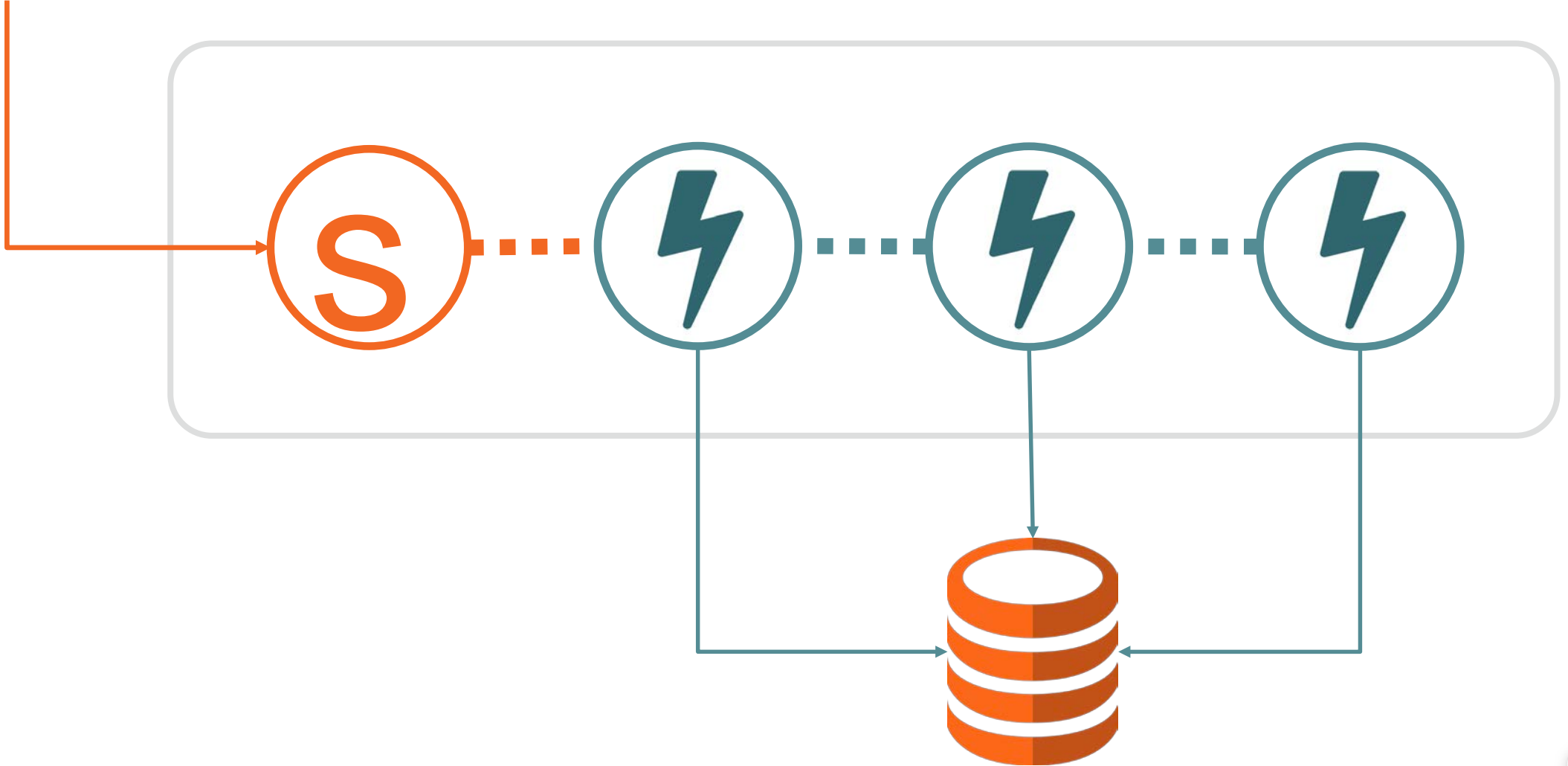
**20** Throughput Units





**Java topology**  
**8 Supervisor Nodes**

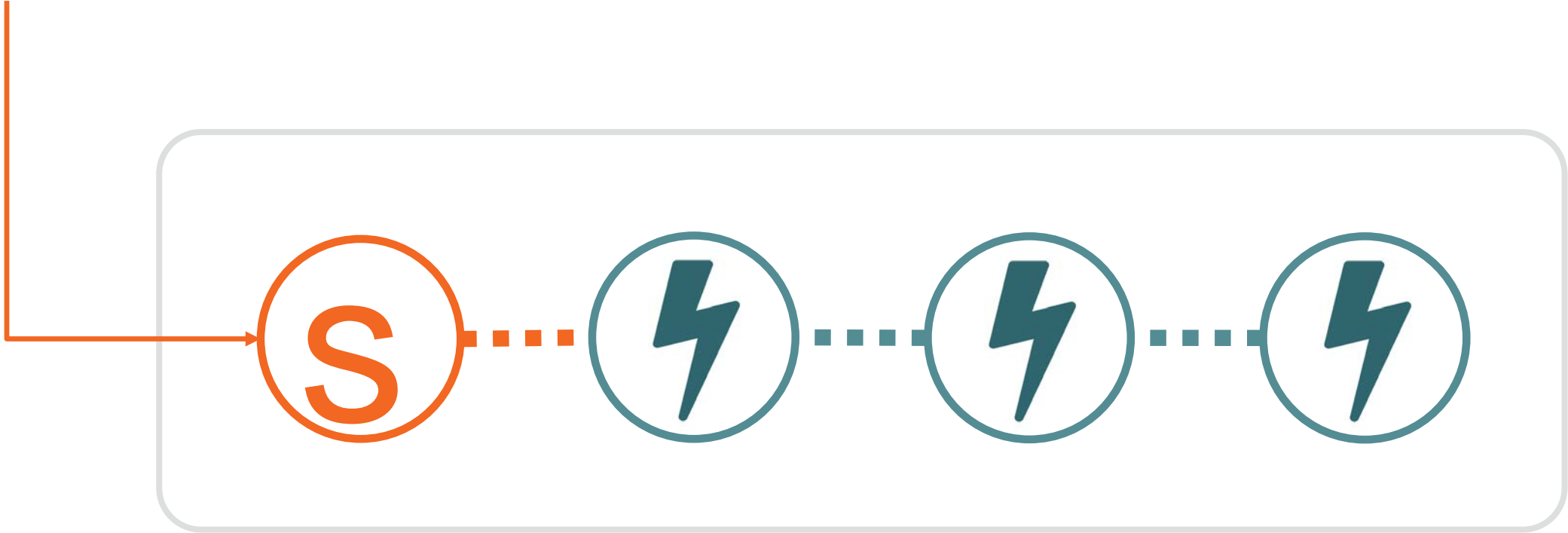






## Setup

Populate **400K** messages





Topology name  
Write timestamp

**races**

**au**

**RowKey**

**au:w**

**au:p**

a6545da436

1444641212016

RaceTiming-1012-1210

**timing-events**

**au**

**RowKey**

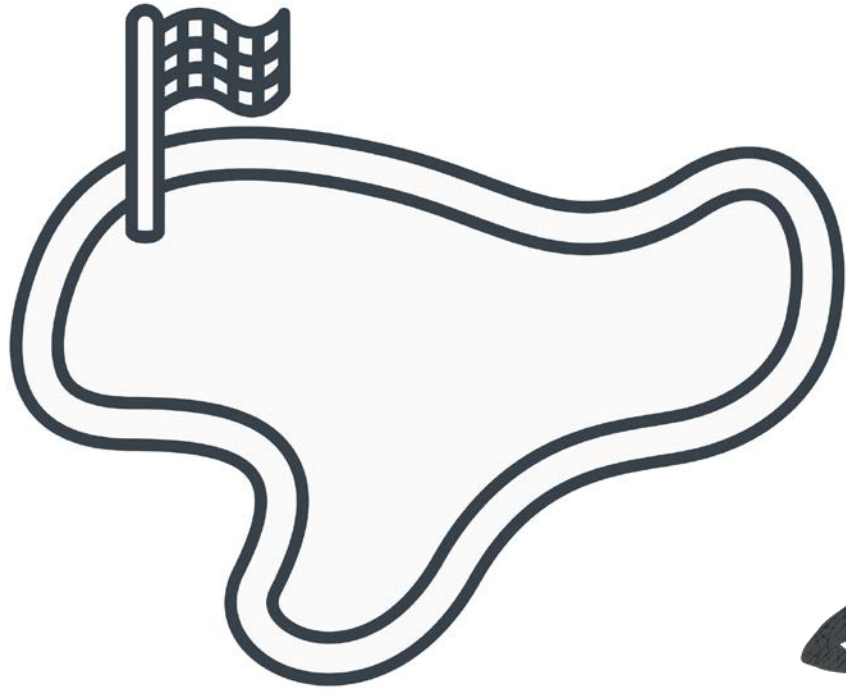
**au:w**

**au:p**

e4324|a31c4|a6545da436

1444641362204

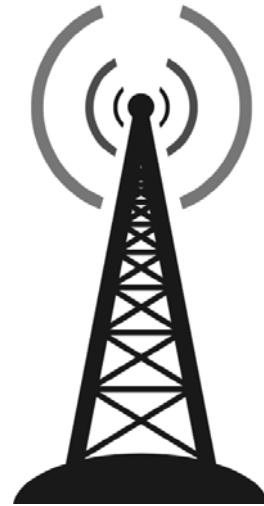
RaceTiming-1012-1210



x40



x100



x10

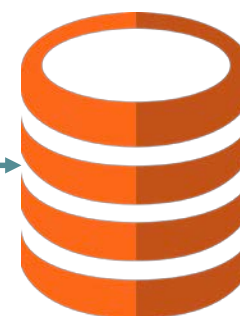
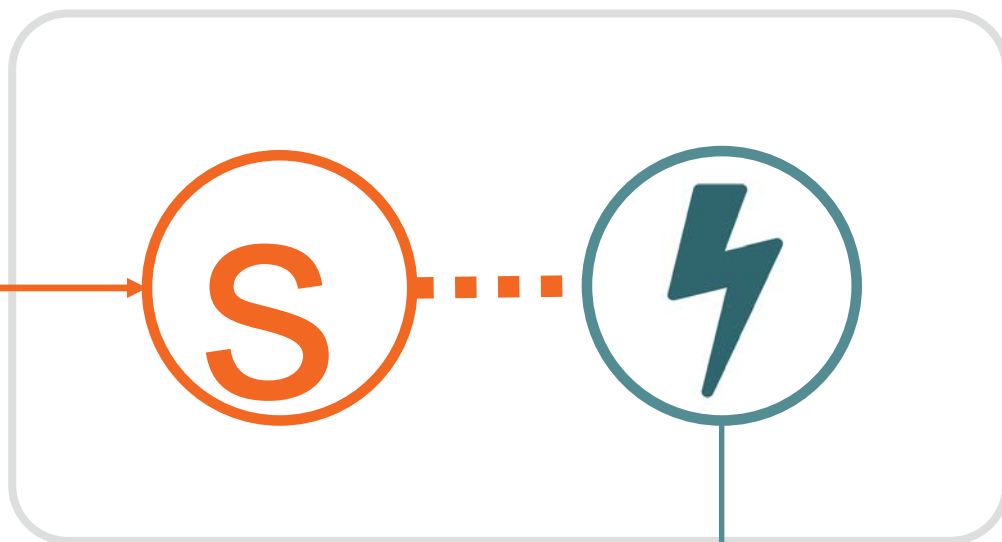


x10





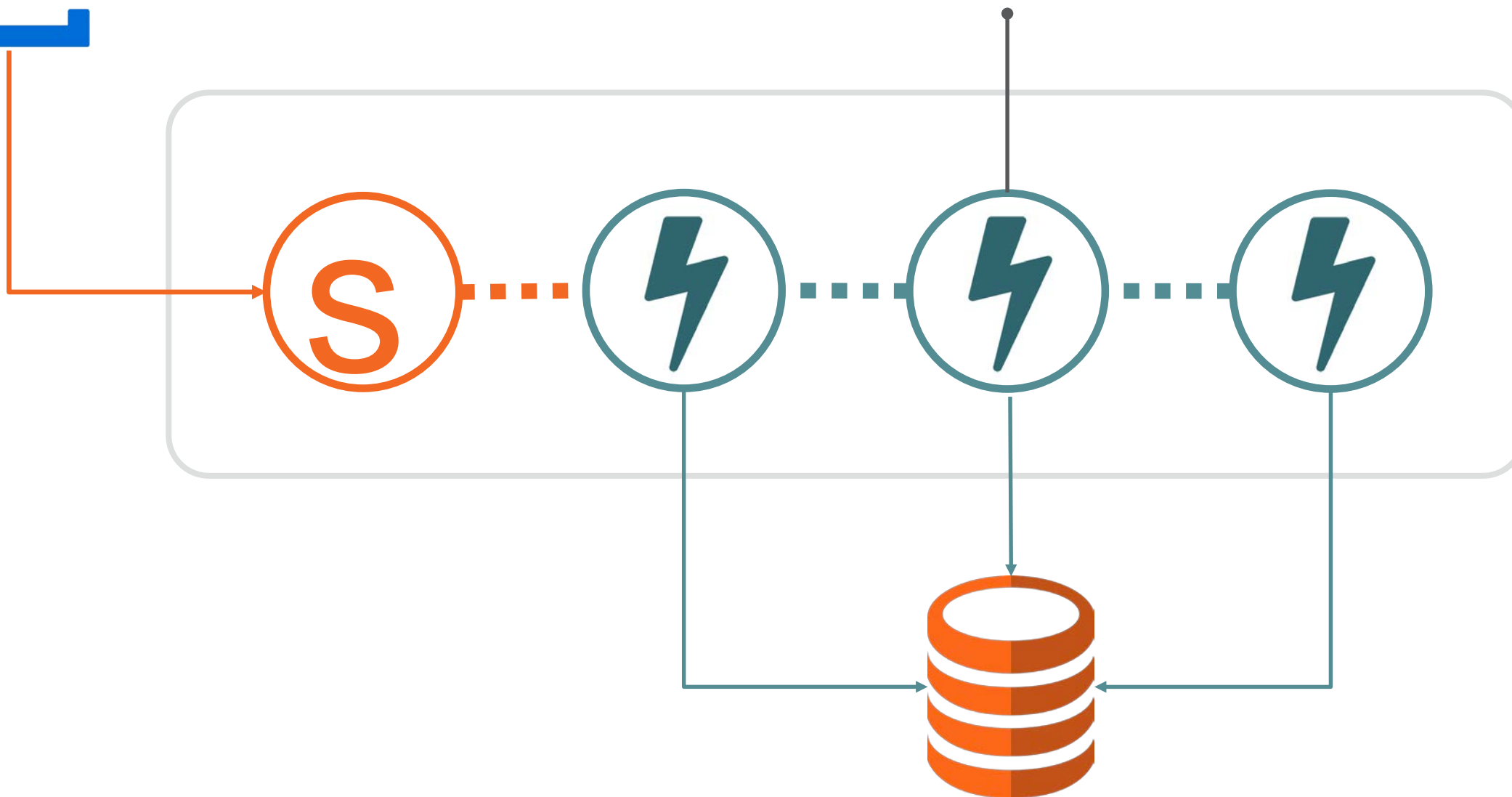
Baseline = **156** seconds  
= **2,560** msg/sec





## Sector Time Bolt

Buffer & update

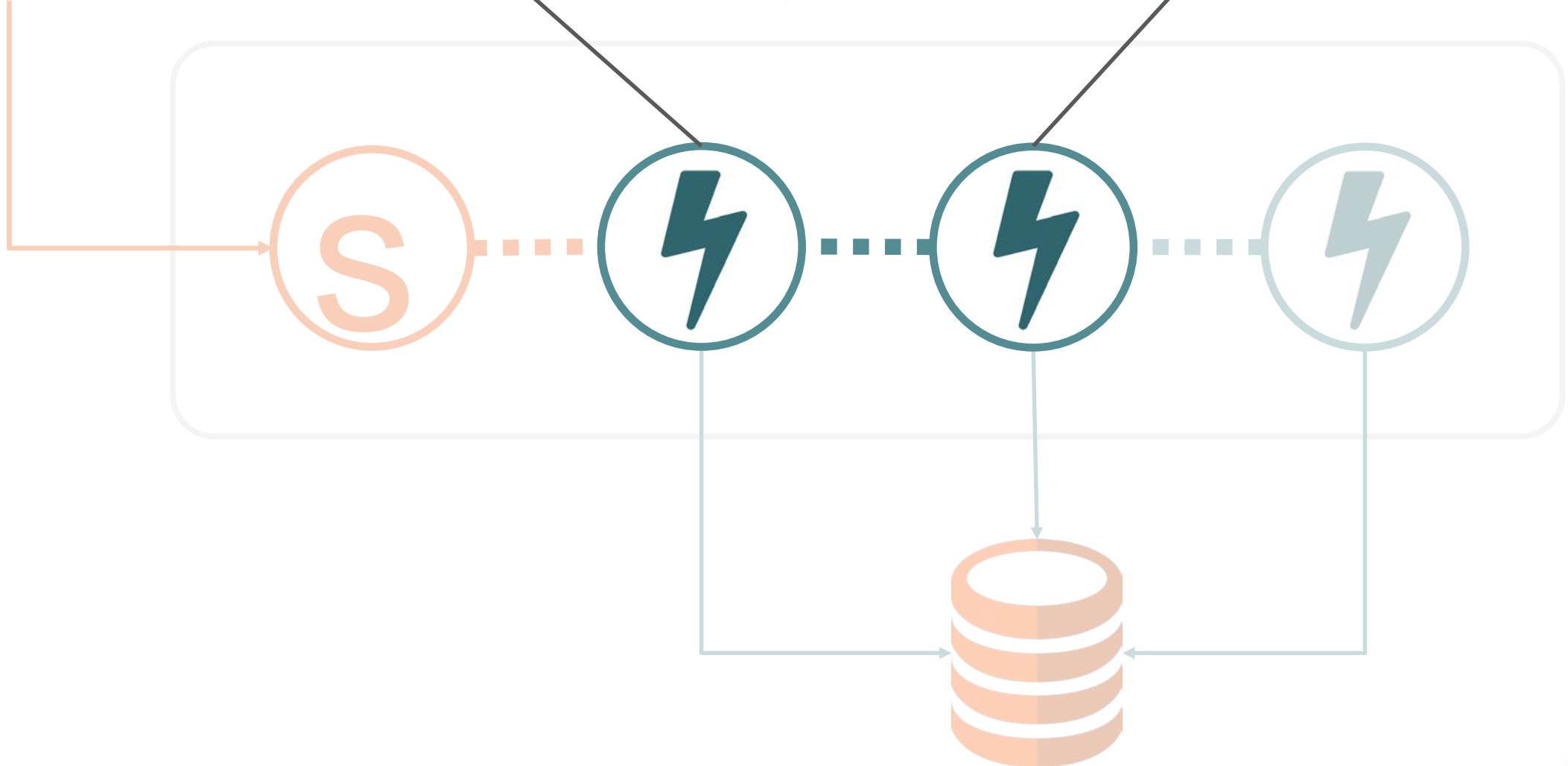




Timing Event **x256**



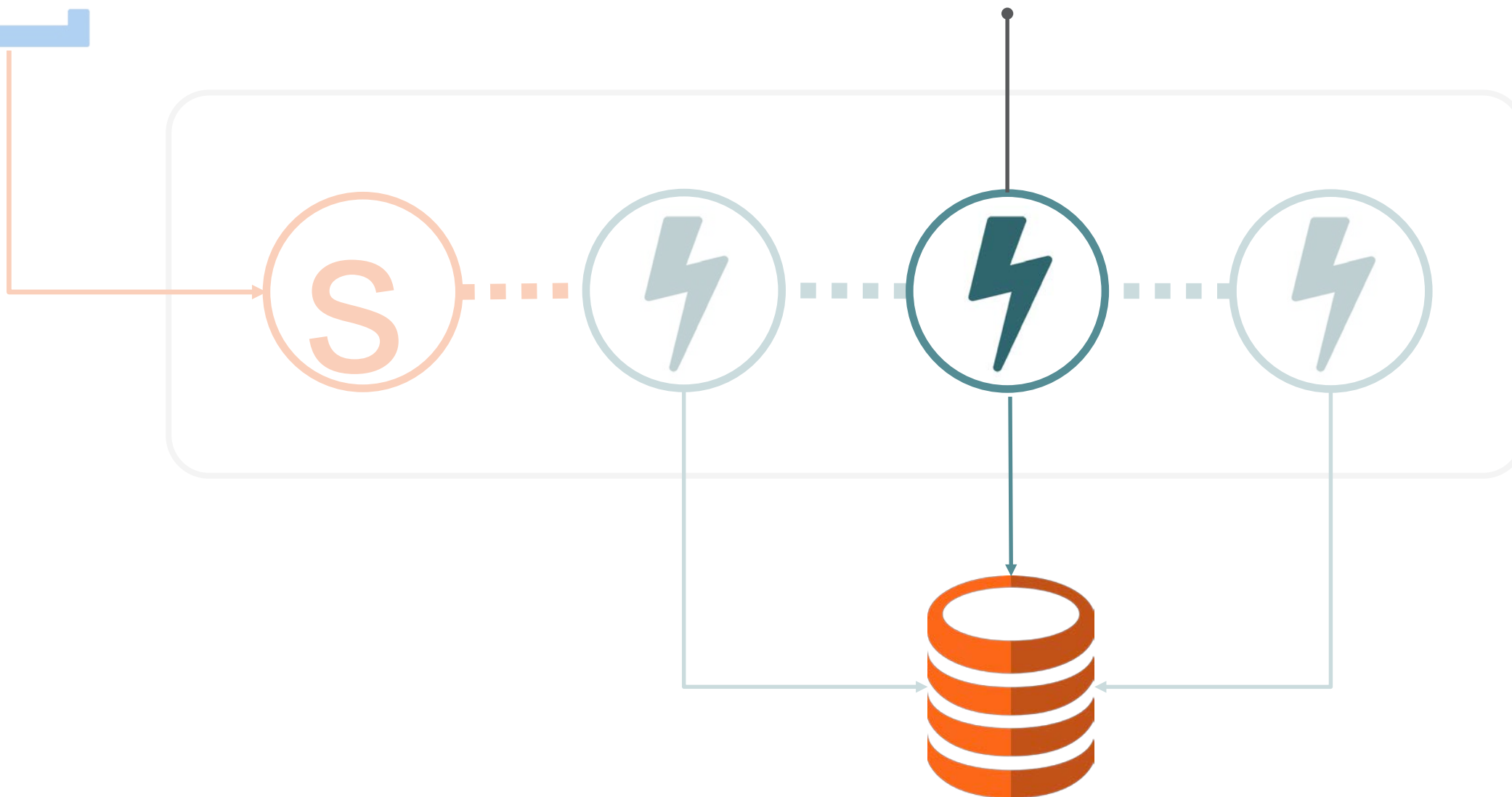
Sector Time **x40**





## Buffer flush

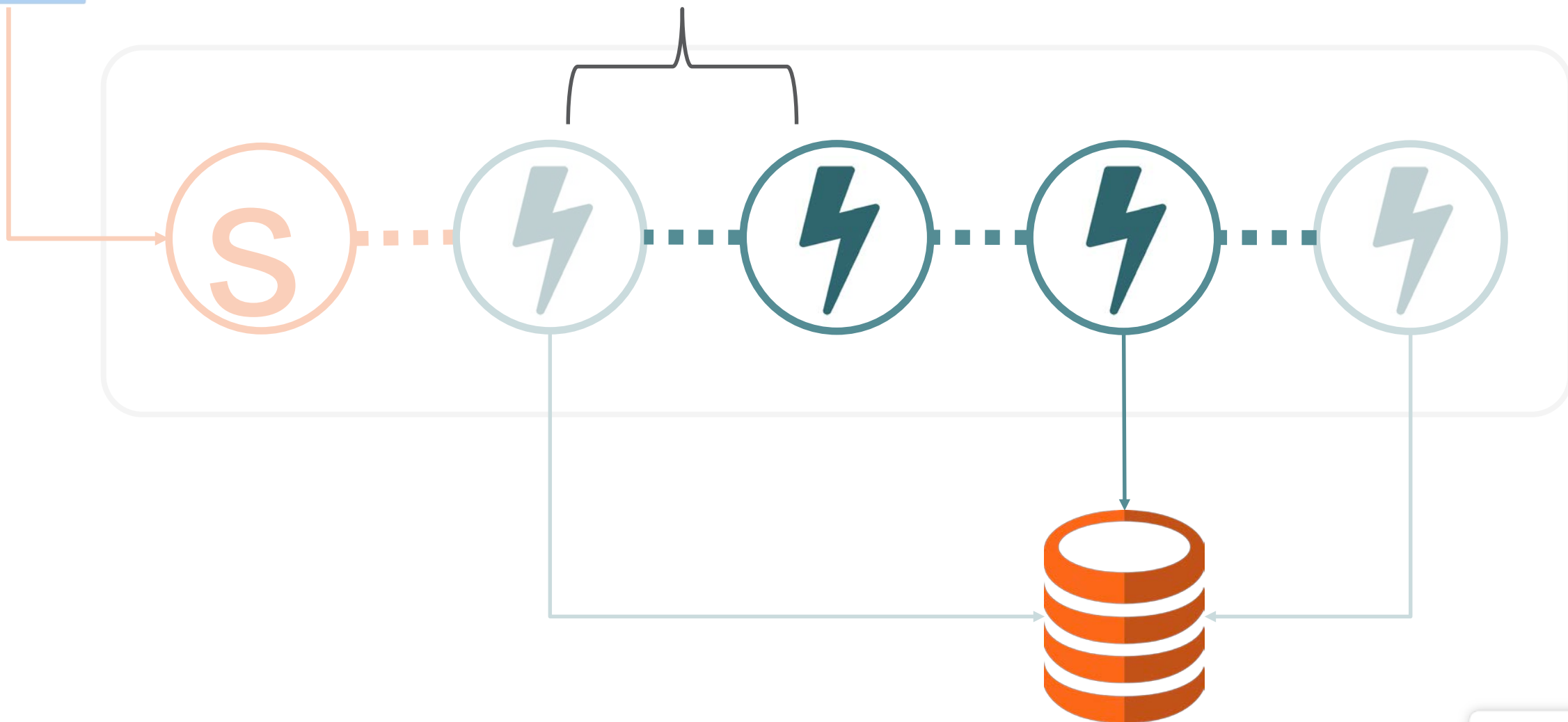
Sequential writes





## Field Grouping

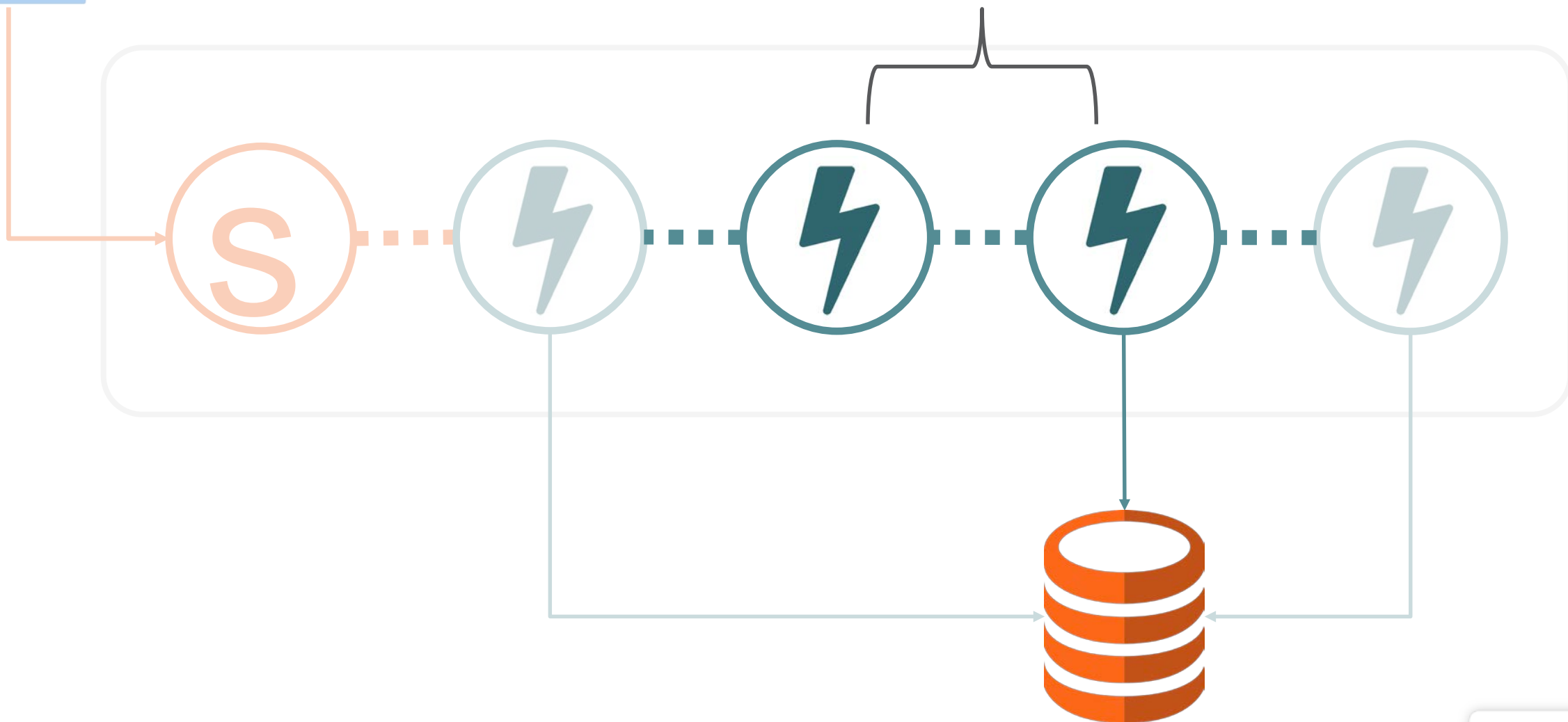
By Race ID + Timer ID



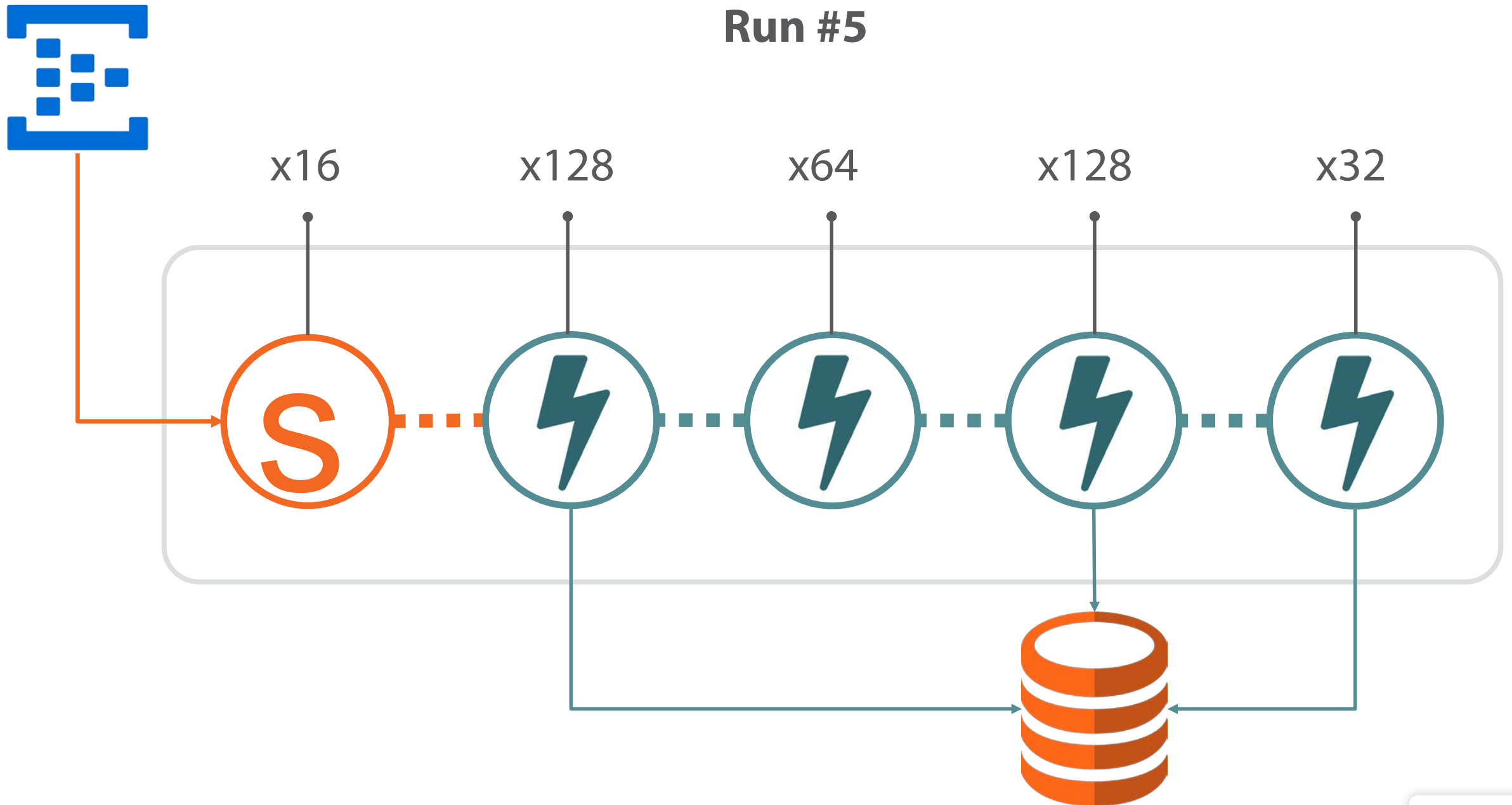


## Shuffle Grouping

Stateless writes



## Run #5

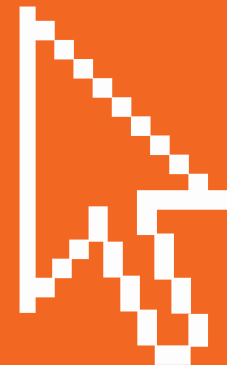


# Demo: Scaling Storm

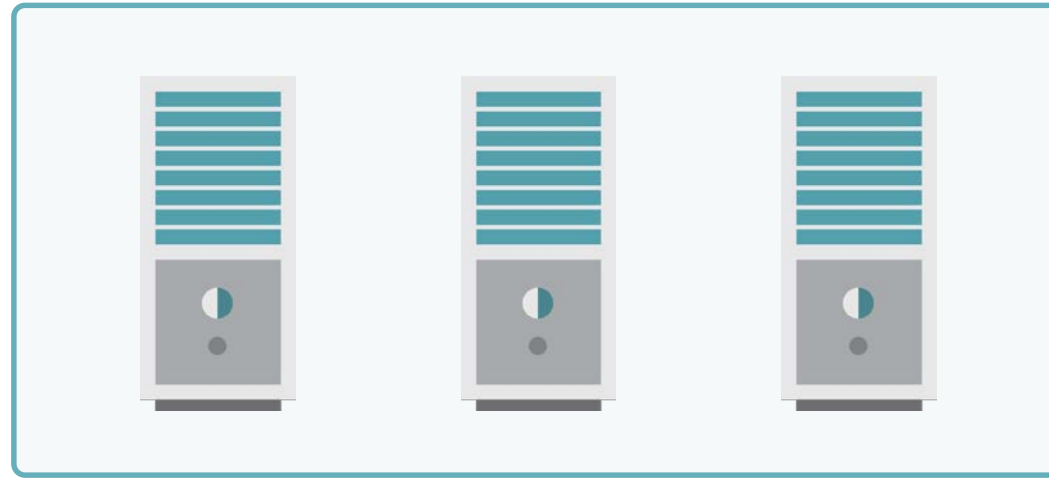
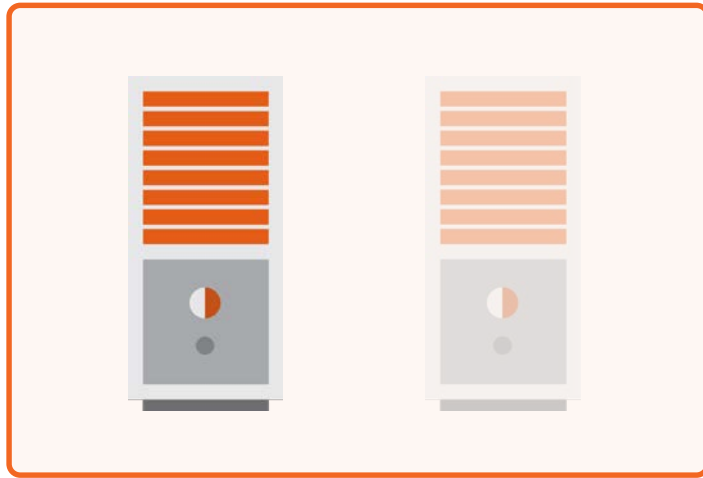
Adding Supervisor Nodes

Distributing More Workers

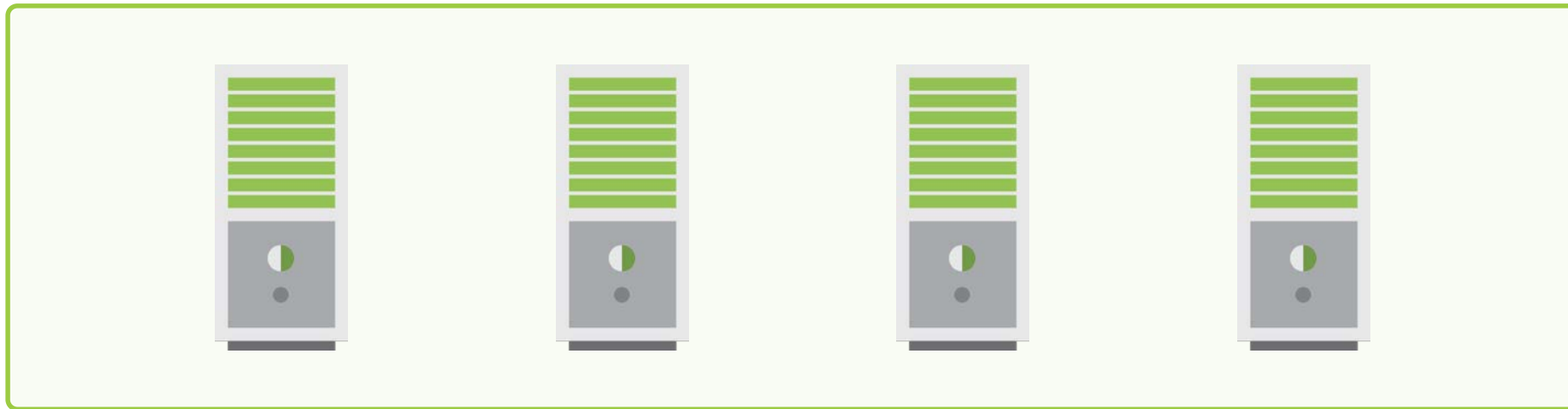
Performance Test







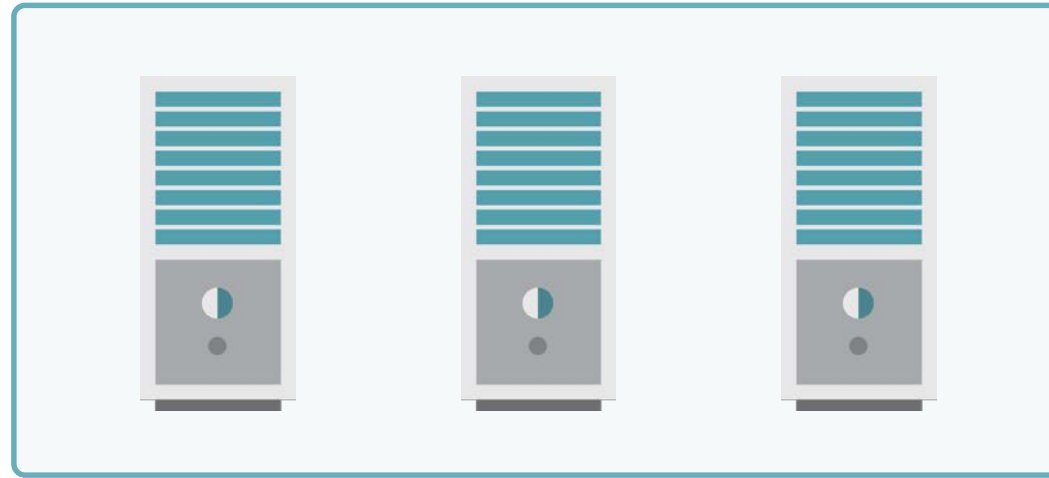
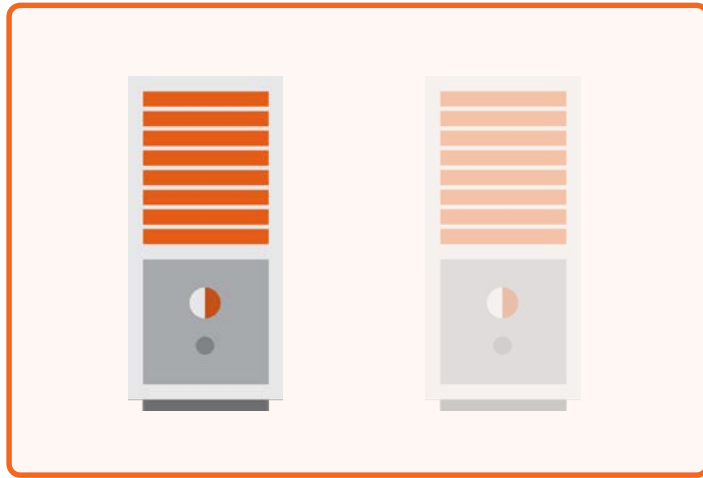
Management  
**x5**



Supervisor  
**x4**



Supervisor Nodes = **44%** of compute



Management  
**x5**



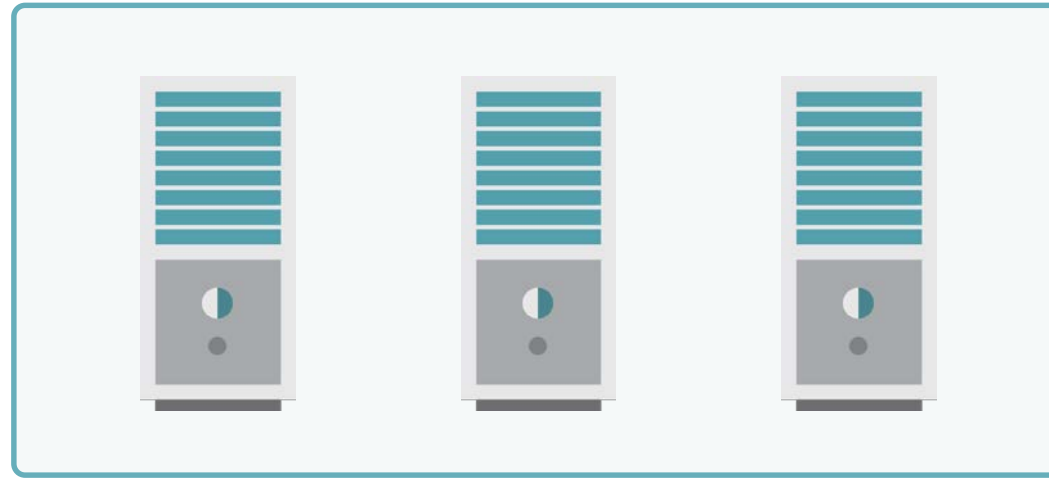
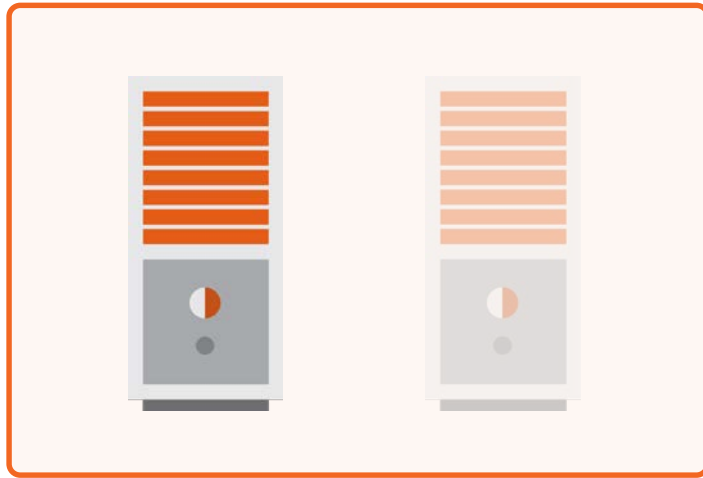
Supervisor  
**x8**



Supervisor Nodes = **62%** of compute



 Partitions **x32**



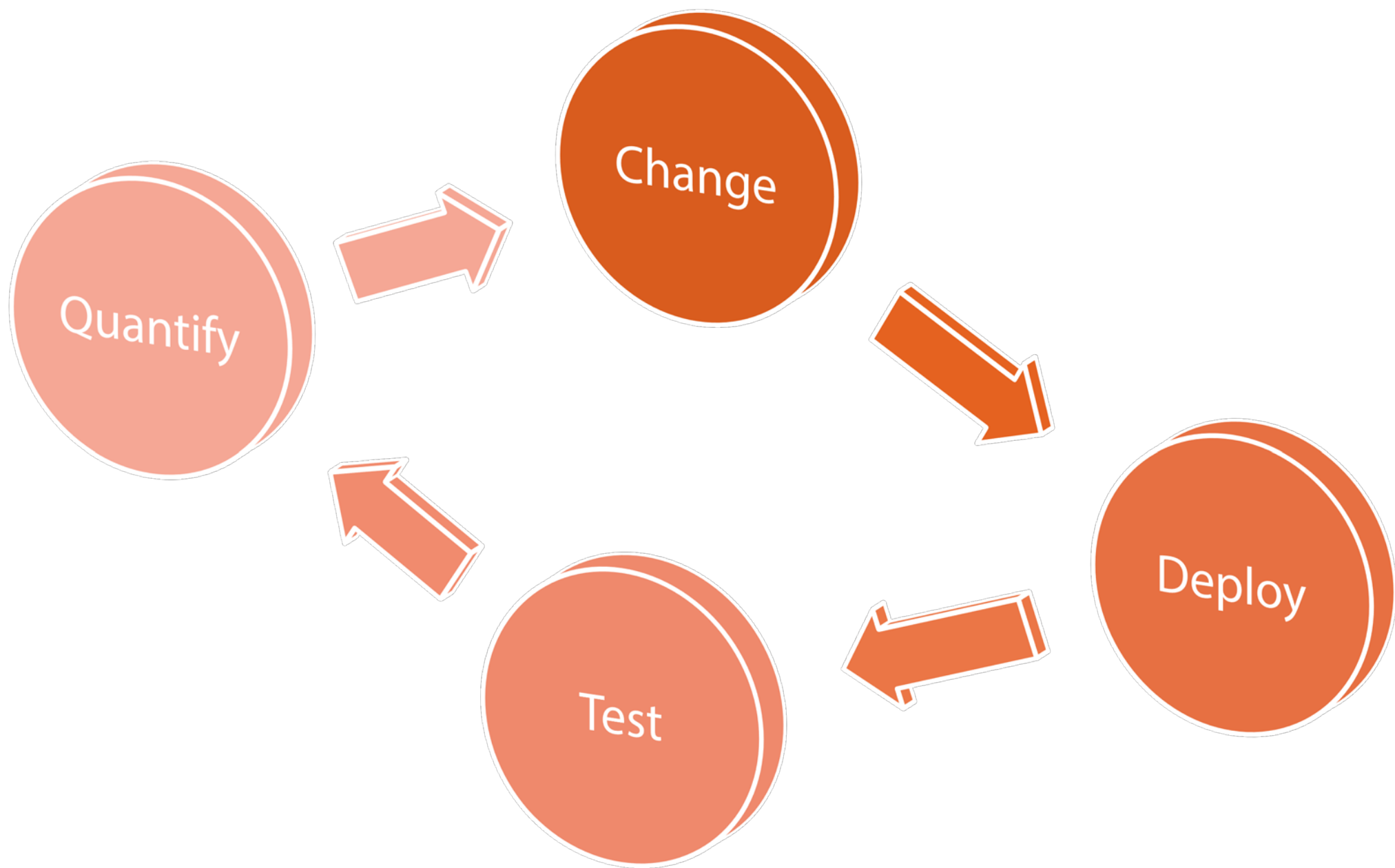
Management  
**x5**



Supervisor  
**x8**



Supervisor **x8** => Workers **x8**

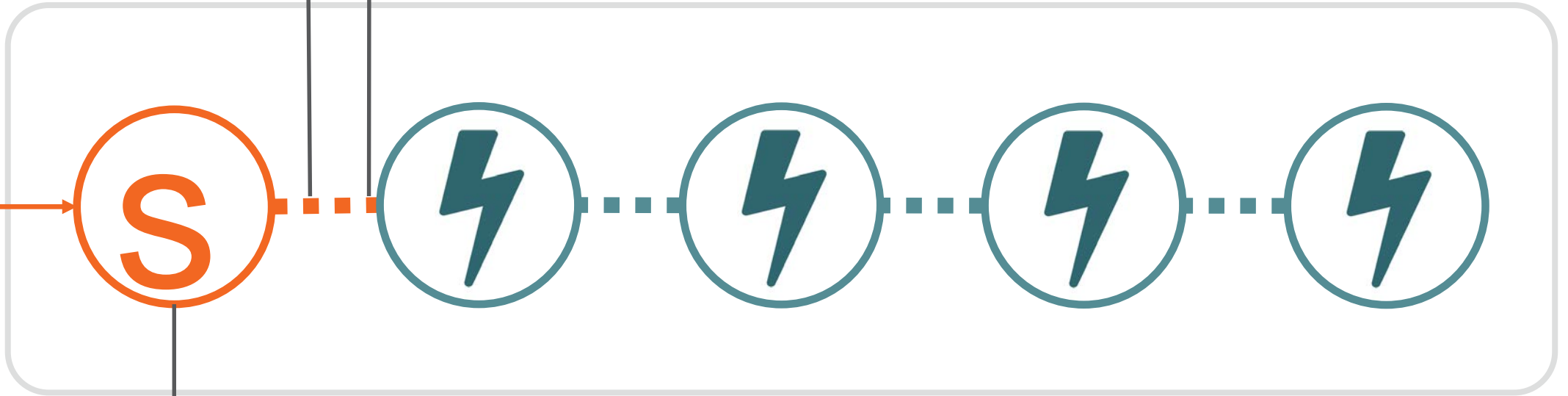




Sequence ID = 103



Sequence ID = 101

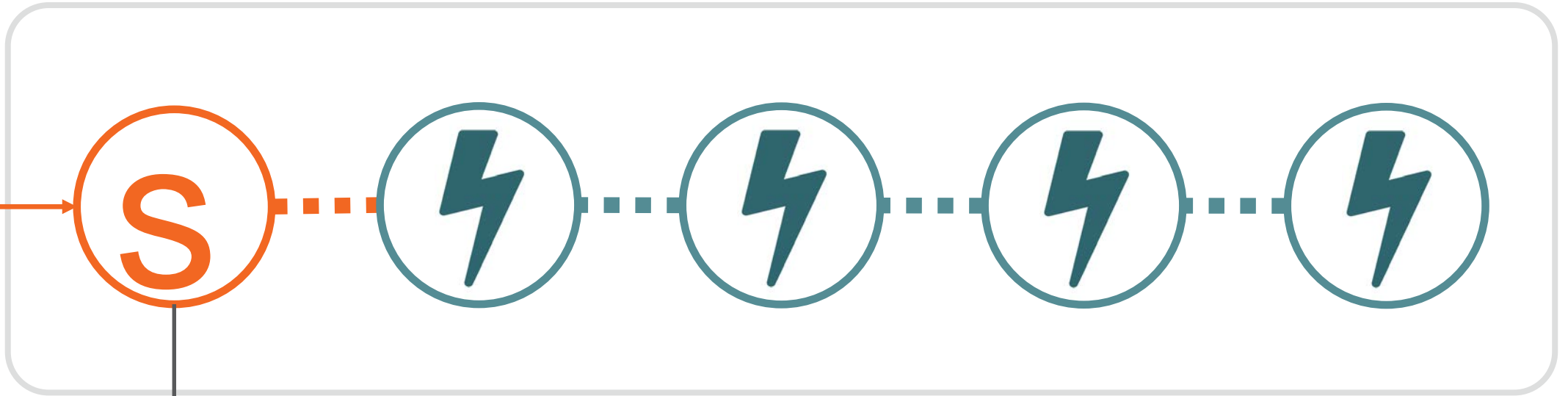


101: pending

103: pending



← ACK: 101

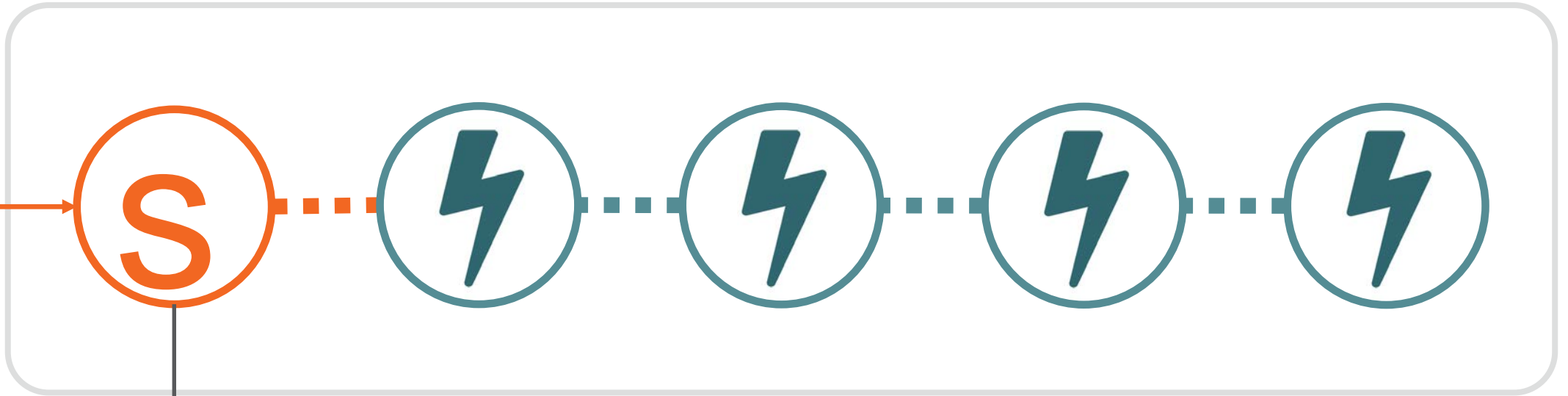


~~101: pending~~

103: pending



← FAIL: 103

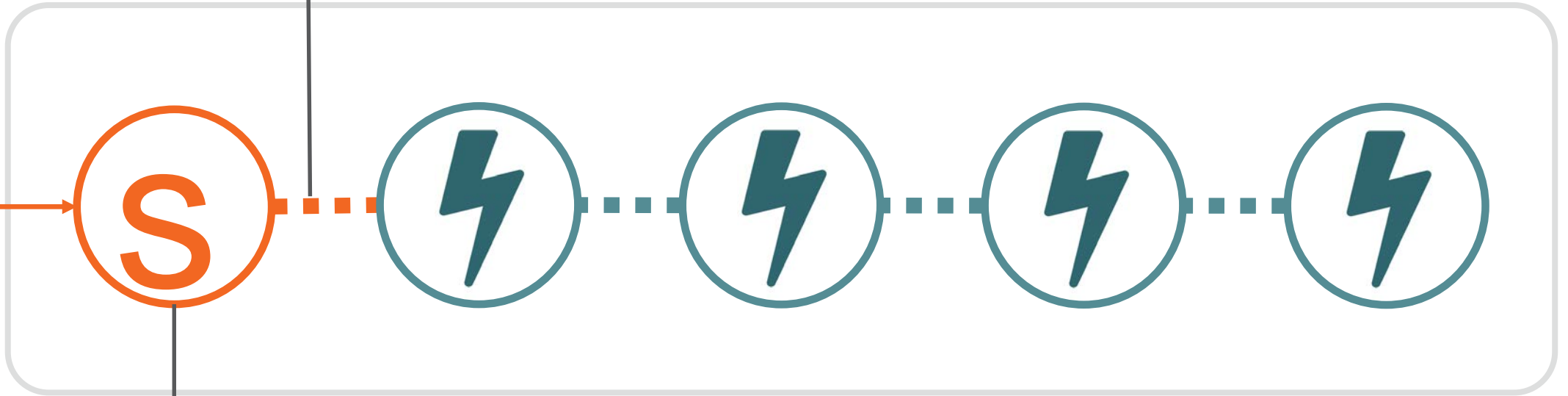


• 103: pending





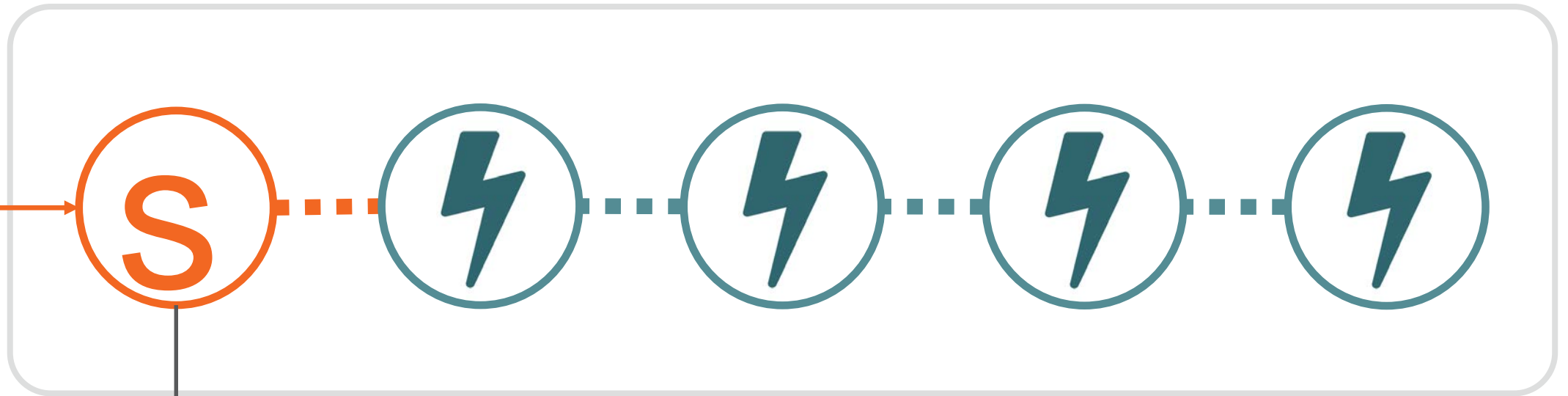
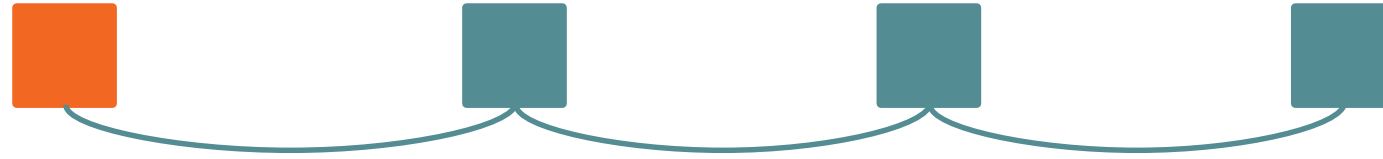
Sequence ID = 103



103: pending



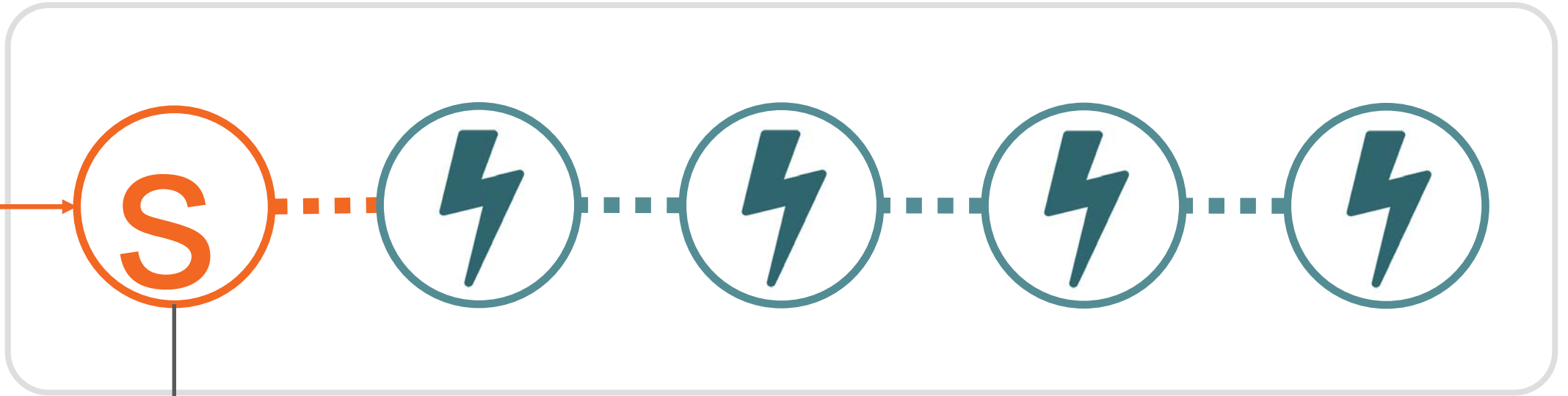
Sequence ID = 103



• 103: pending



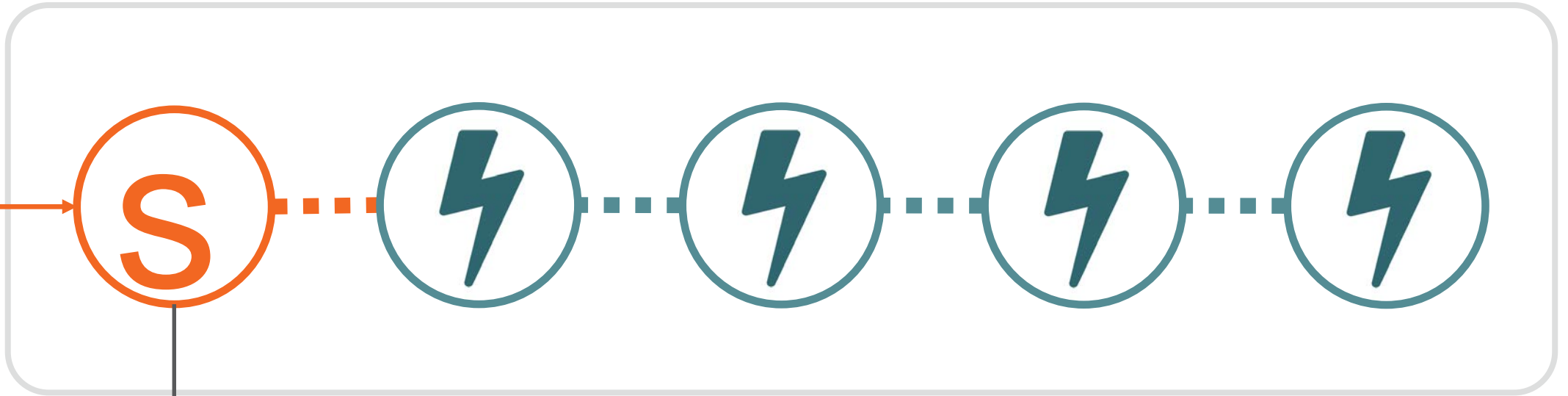
Sequence ID = 103



~~• 103: pending~~



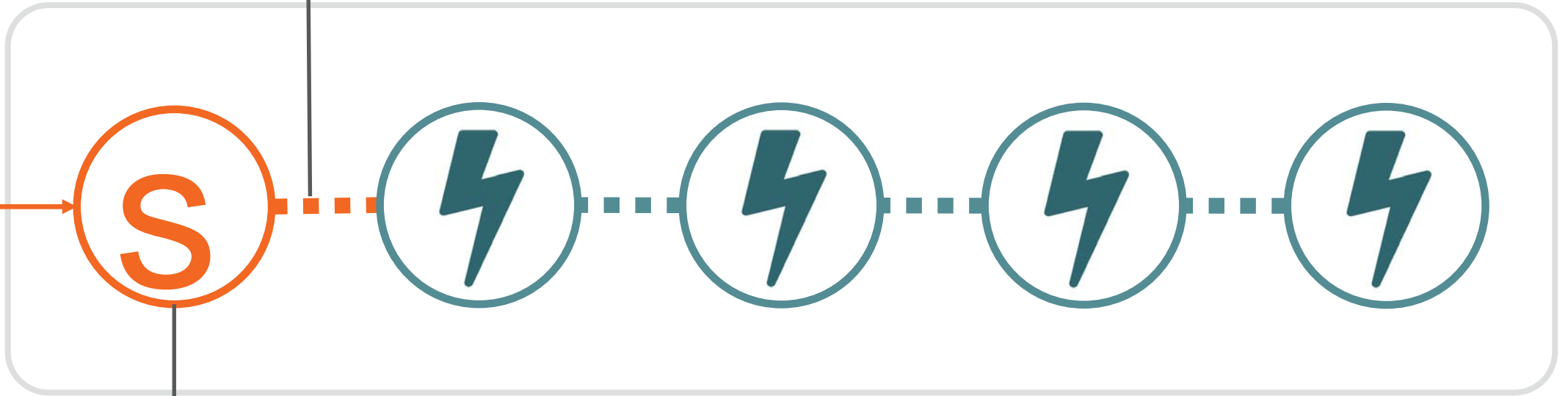
Sequence ID = 103



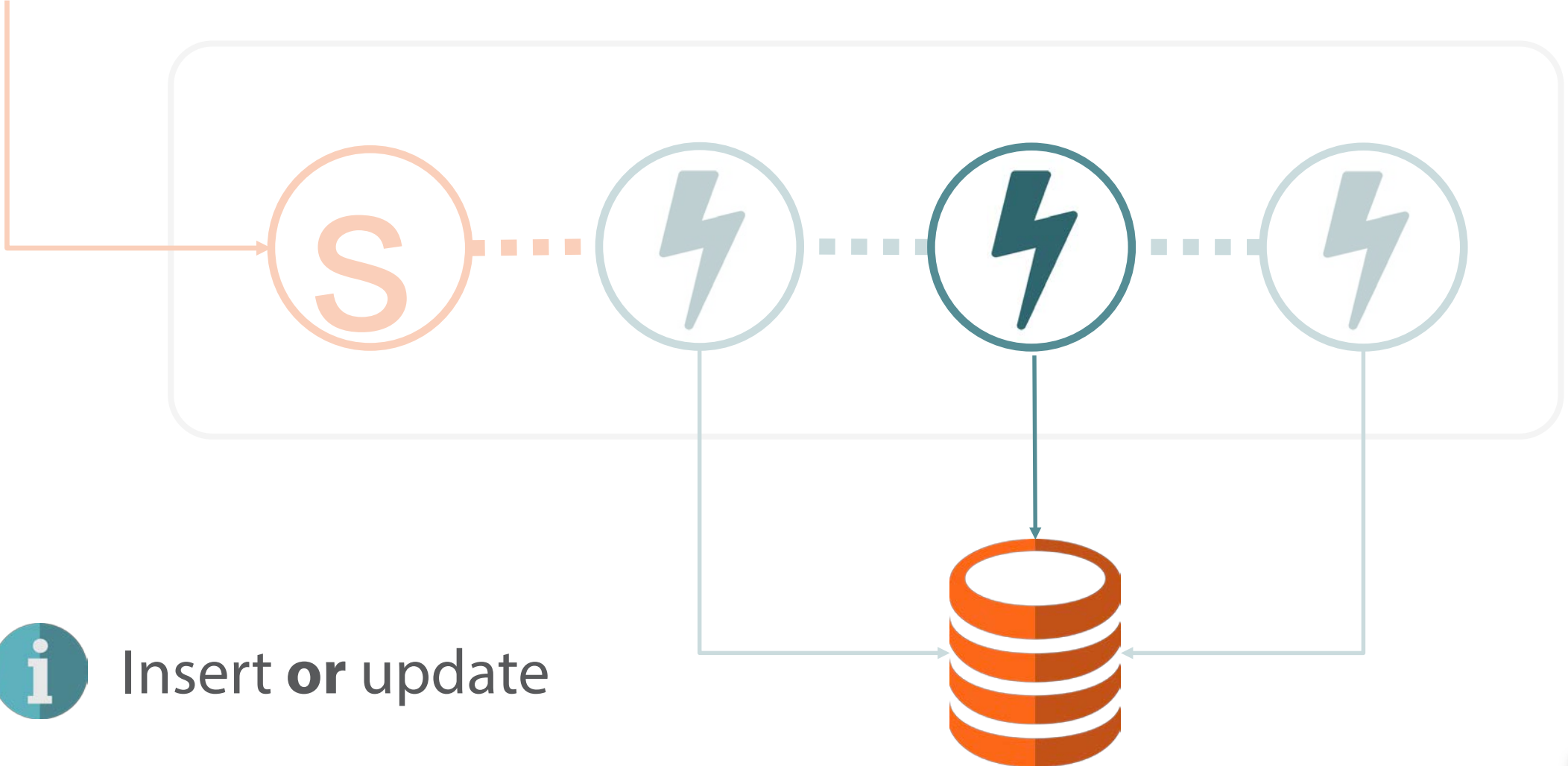
• 103: pending



Sequence ID = 103



103: pending



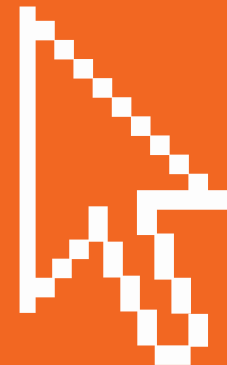
Insert **or** update

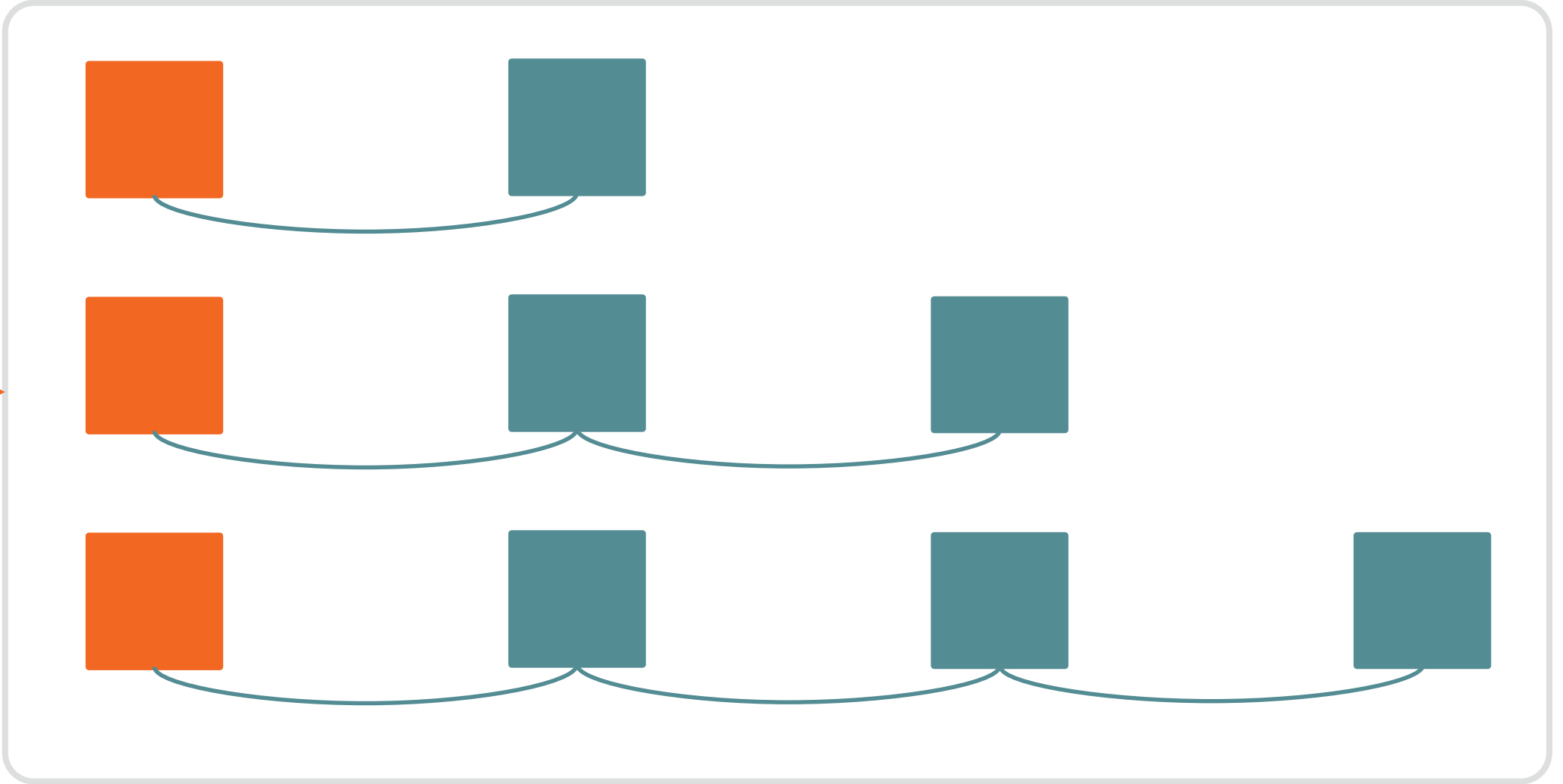
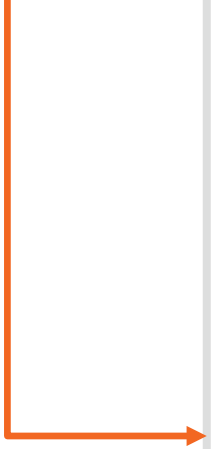
# Demo: Reliable Processing

Tuple Tree

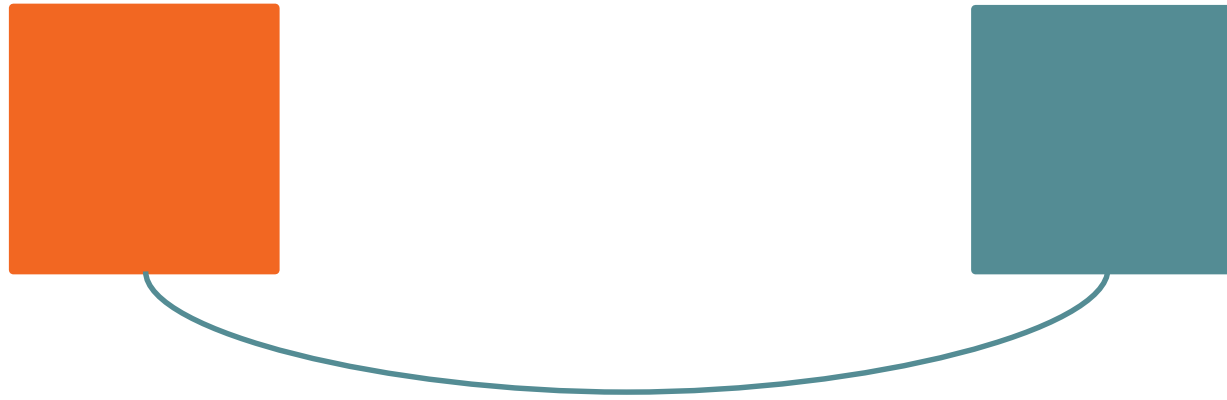
Ack and Fail

Performance Impact









## **Scenario 1 – Repeated data**

Two-tuple tree



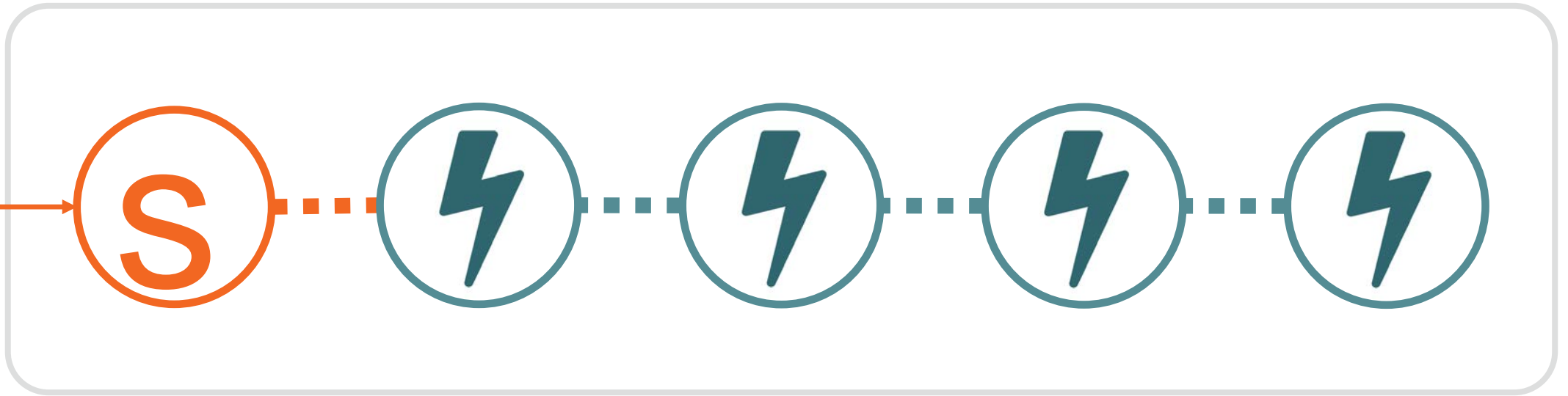
## Scenario 2 – Racer Continuing

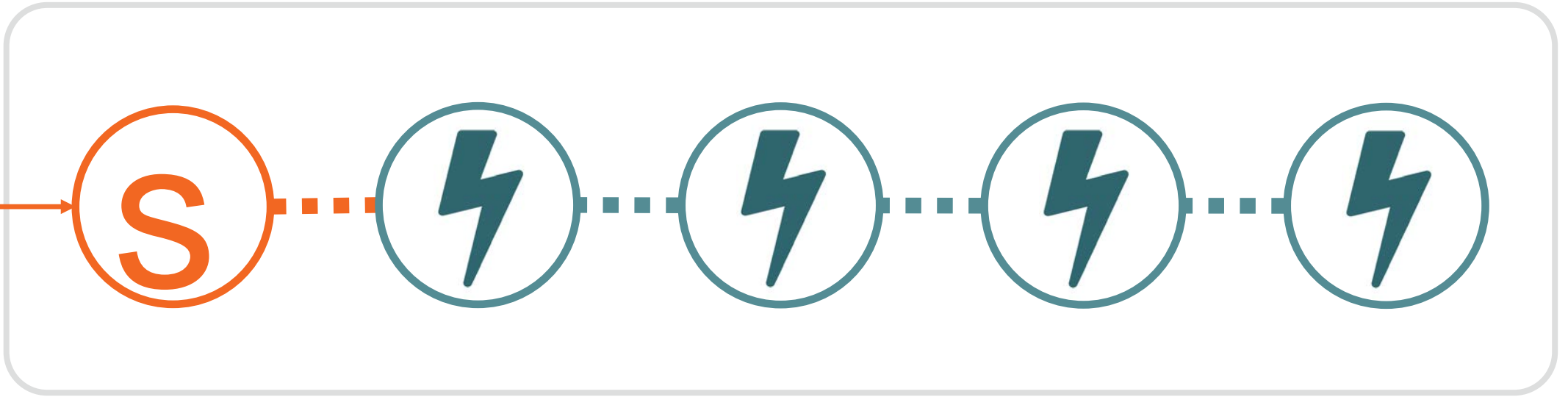
Three-tuple tree



## Scenario 3 – Racer Finished

Four-tuple tree





# Demo: Logging

Storm UI

Worker Log Files

Azure Table Storage

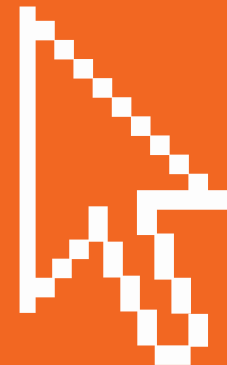


# Demo: Custom Logging

Log4net Config

Accessing the Logger

Writing Entries

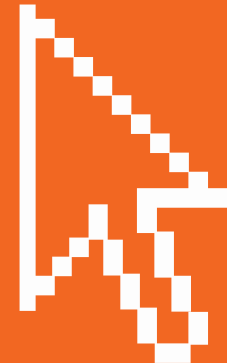


# Demo: Testing Storm

Component Unit Tests

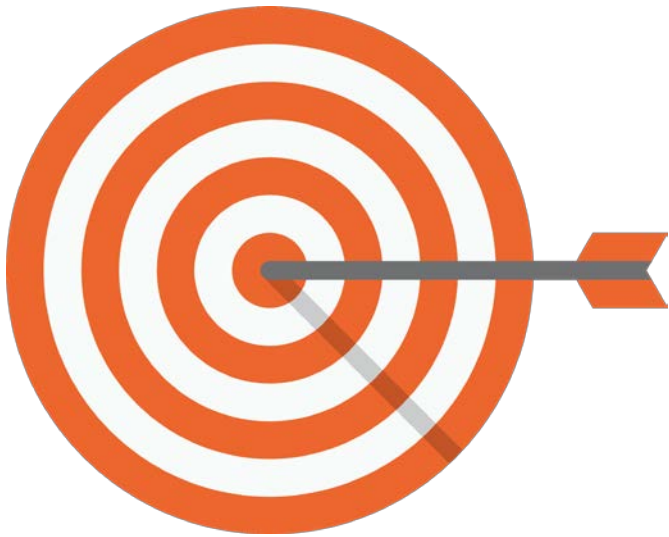
Topology Integration Tests

End-to-End Tests





# Module Goals



Storm Cluster Architecture



Runtime Compute Structure



Performance Tuning



Guaranteed Processing



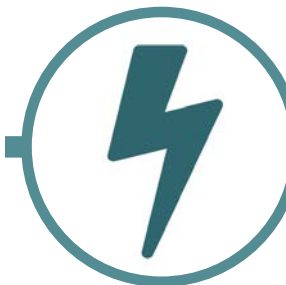
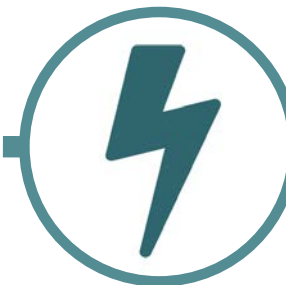
Monitoring & Testing

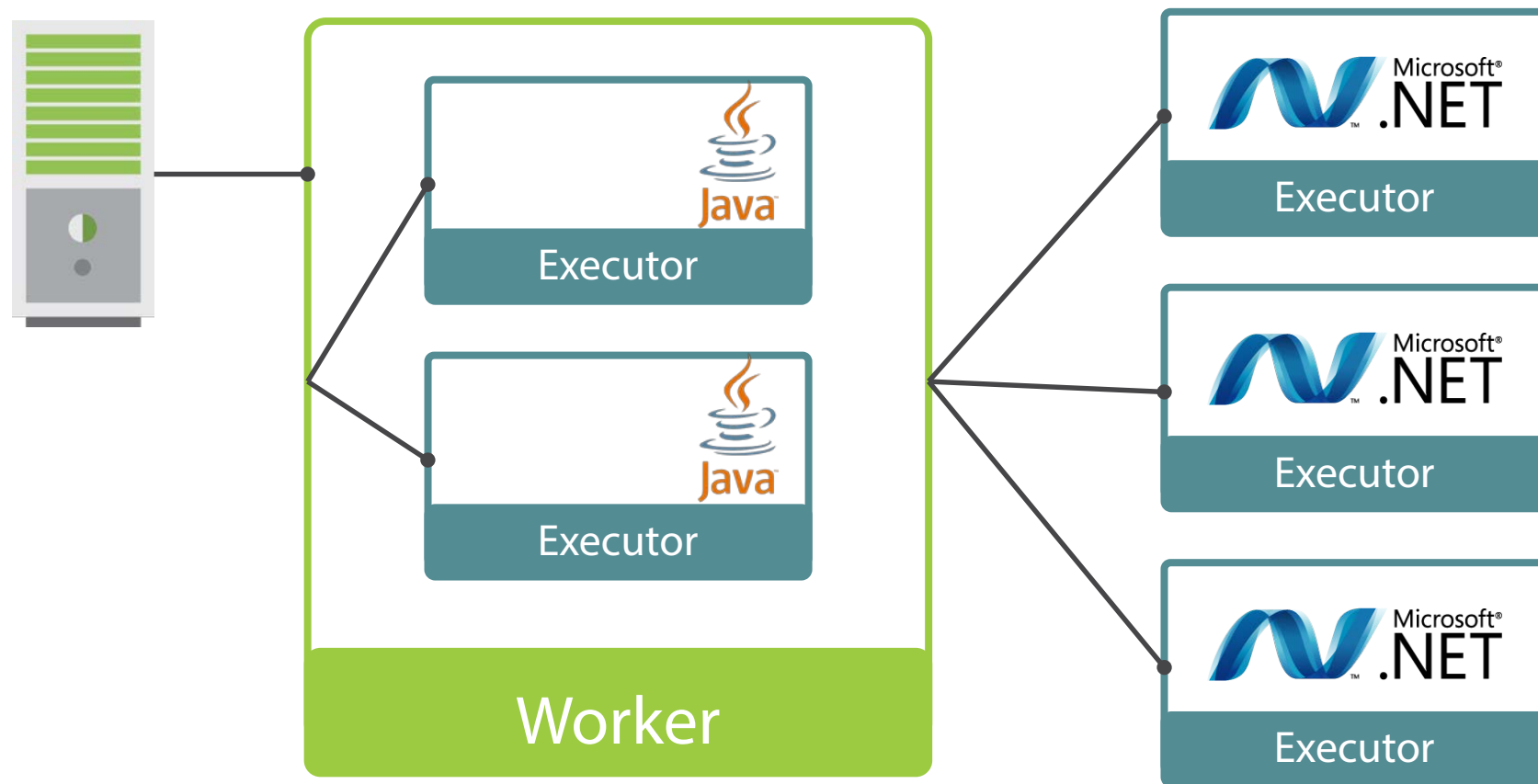




c:\RaceSimulator.exe

400K events





# workers

# executors

# Tuning Guidelines



**Workers:** 1 per supervisor

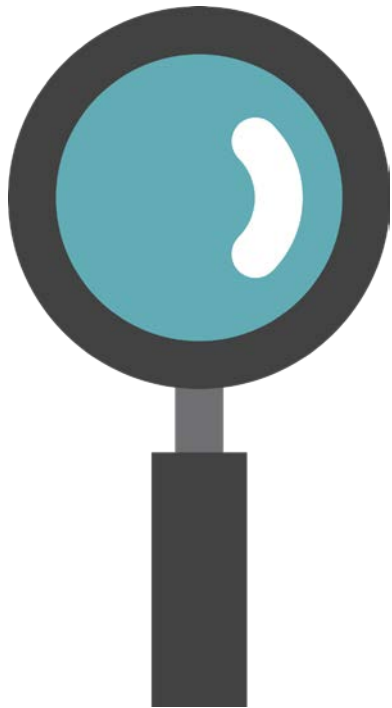
**Spouts:** 1 per partition

**CPU-bound Bolts:** 1 per core

**I/O-bound Bolts:** 1-10 per core

Or more...

# Coming Next



Apache Hive

Mapping & Querying HBase

Joining Across Stores

Writing Output to Azure