Troubleshooting & Operations

Konstantin Knauf, Solutions Architect
Nico Kruber, Solutions Architect
Seth Wiesman, Solutions Architect

Agenda

Morning

- 10:00am - 11:30am: Getting Started
  - Flink's Distributed Architecture "Recap"
  - Getting started with the hands-on exercises
- **Coffee Break**
- 11:45am - 1:15pm: Event-Time & Latency
  - Metrics & Monitoring
  - Watermarking
  - Latency

Agenda

Afternoon

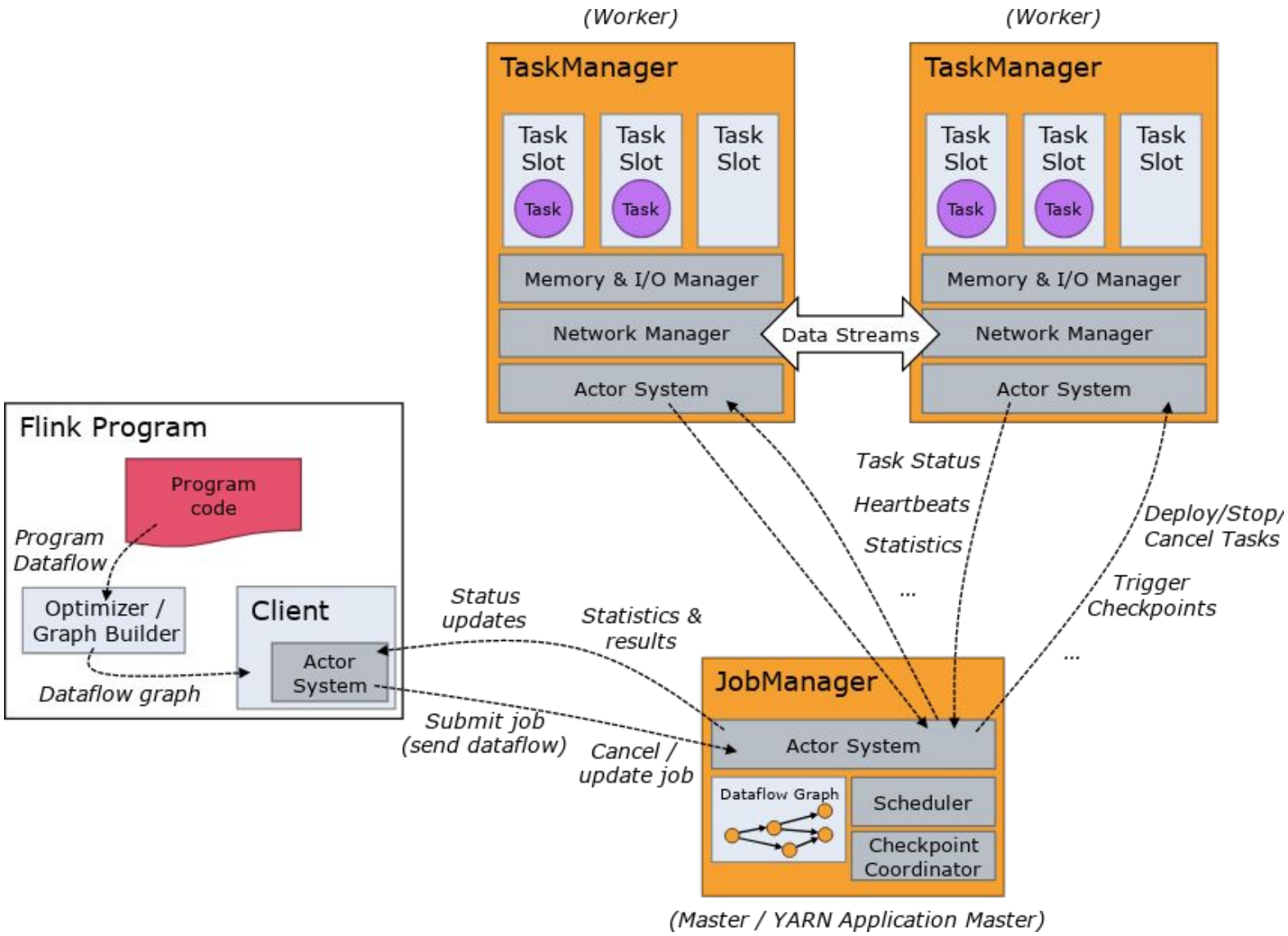
- 2:15pm - 3:45pm: Performance Tuning
  - Capacity Planning
  - Serialization
- **Coffee Break**
- 4:00pm - 5:30pm: Checkpointing & State Management
  - Statebackends & Checkpoint Tuning
  - State Migration

Flink's Distributed Architecture

Nico Kruber, Solutions Architect & Apache Flink committer

 |

Cluster Components



Job Definition

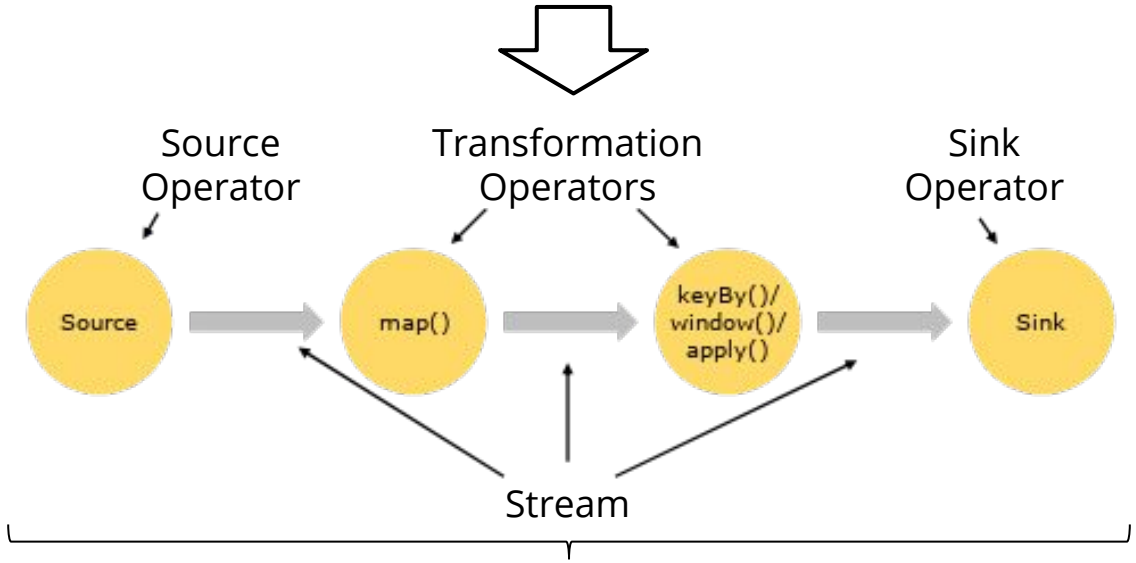
```
DataStream<String> lines = env.addSource(
  new FlinkKafkaConsumer<>(...));

DataStream<Event> events = lines.map((line) -> parse(line));

DataStream<Statistics> stats = events
  .keyBy("id")
  .timeWindow(Time.seconds(10))
  .apply(new MyWindowAggregationFunction());

stats.addSink(new BucketingSink<>(path));
```

Source

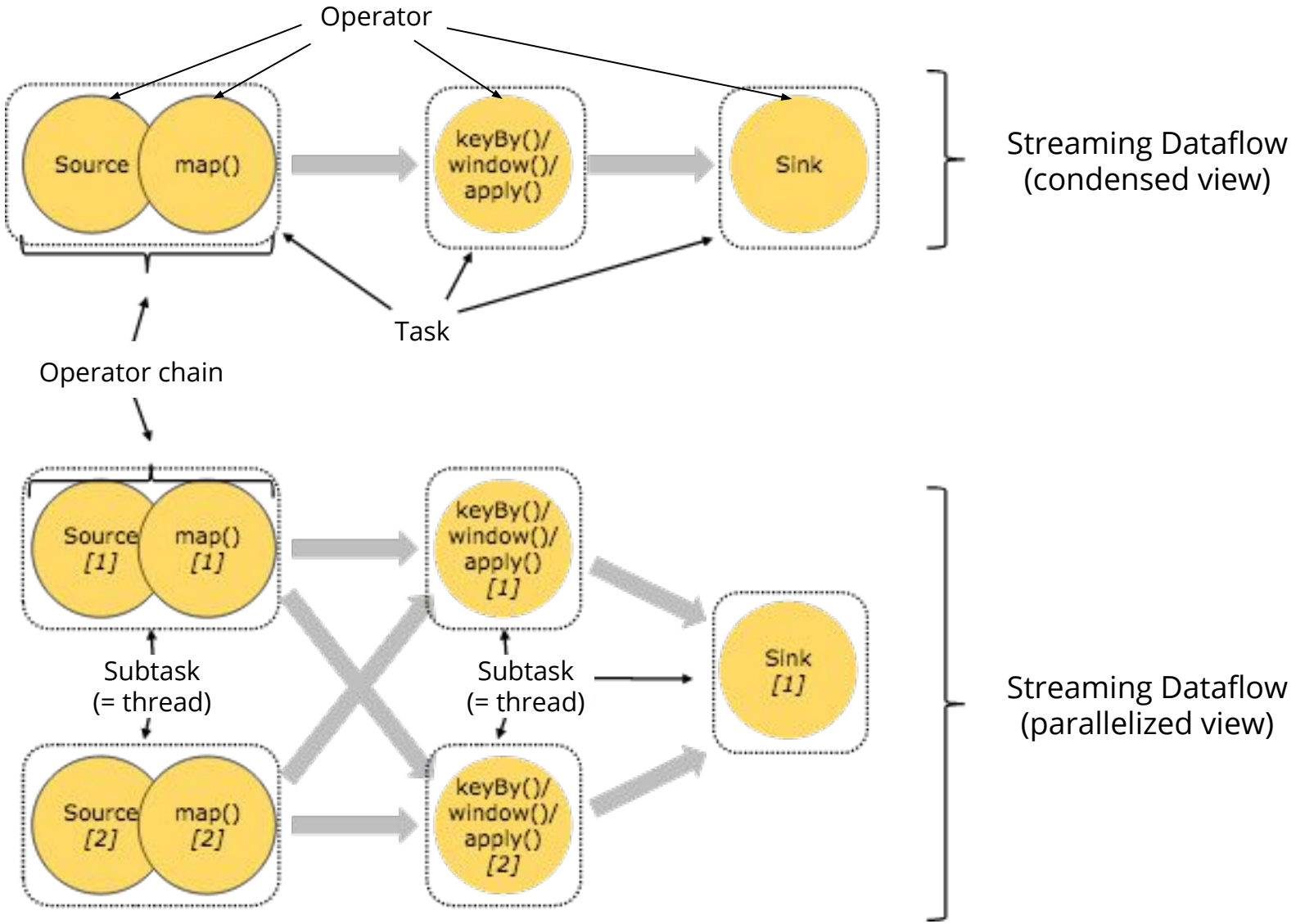
Transformation

Transformation

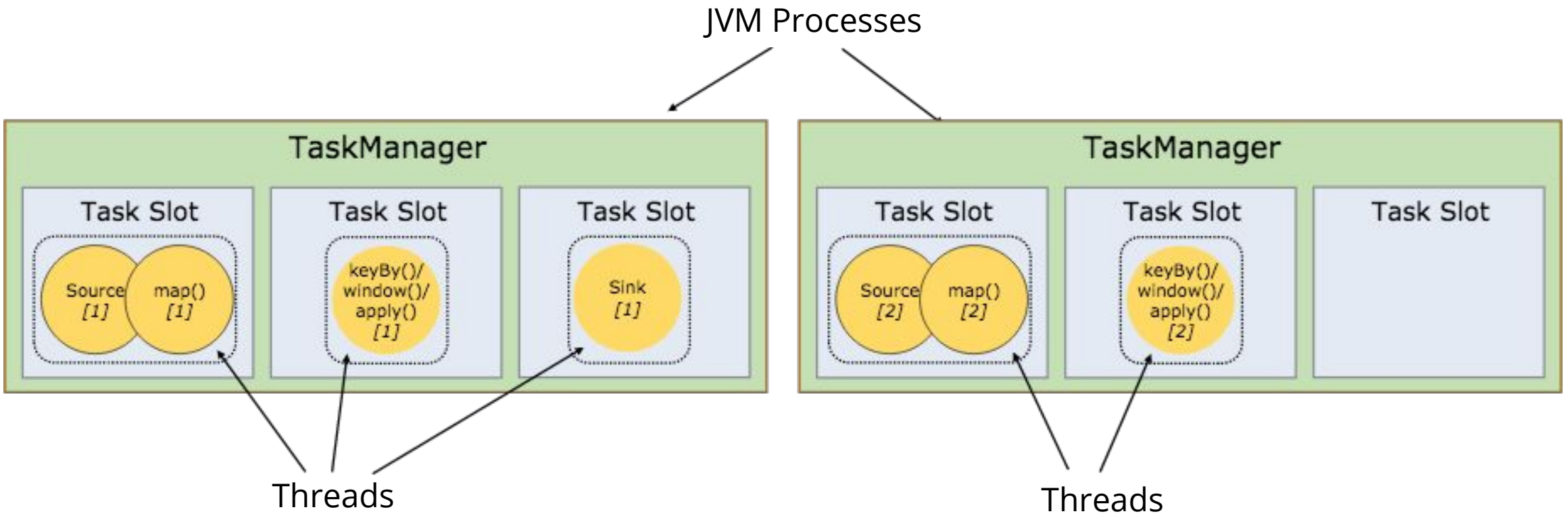
Sink



Job Components & Operator chaining

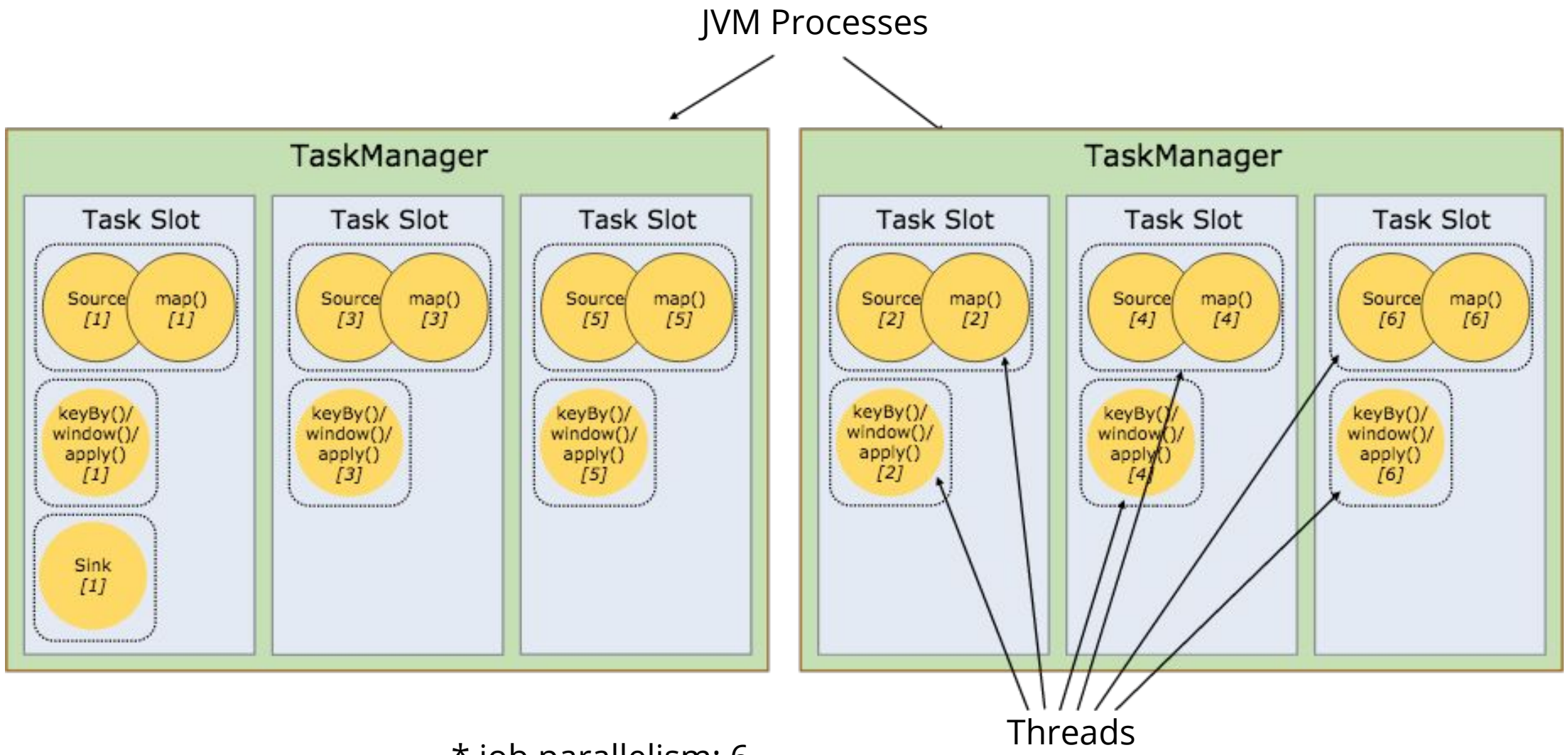


Task Deployment: Slots



* job parallelism: 2

Task Deployment: Slot Sharing



* job parallelism: 6

Fault Tolerance Guarantees

What happens if a worker goes down?

Flink supports different levels of guarantee for failure recovery:

Exactly once

- Each event affects the managed state exactly once.
- **Note**: This does *not* mean that events are processed exactly once!

At least once

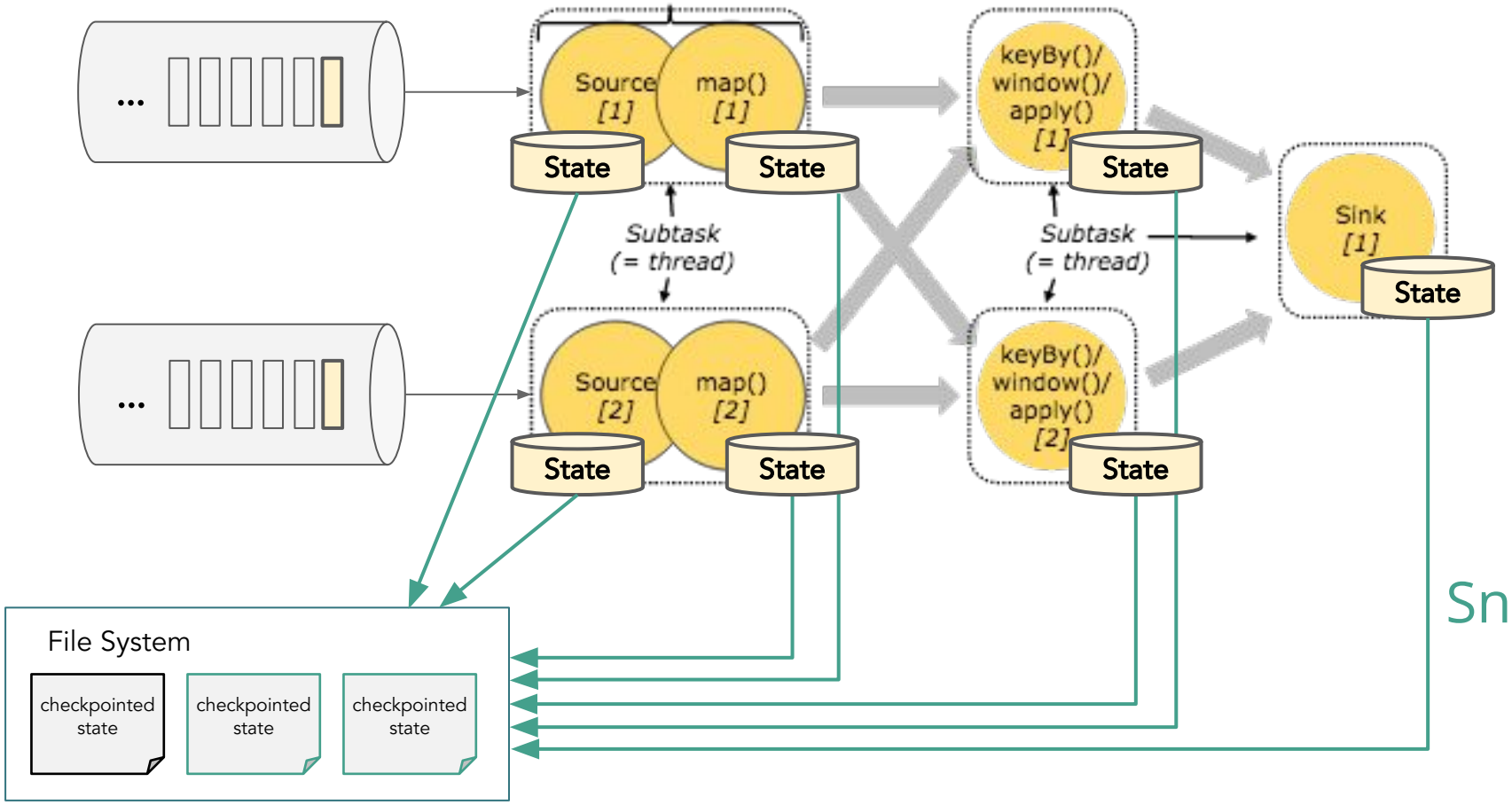
- Each event affects the declared state of a program at least once

Deactivated / None / At most once

- All state is lost in case of a failure

Job Lifecycle & Fault Tolerance

While running



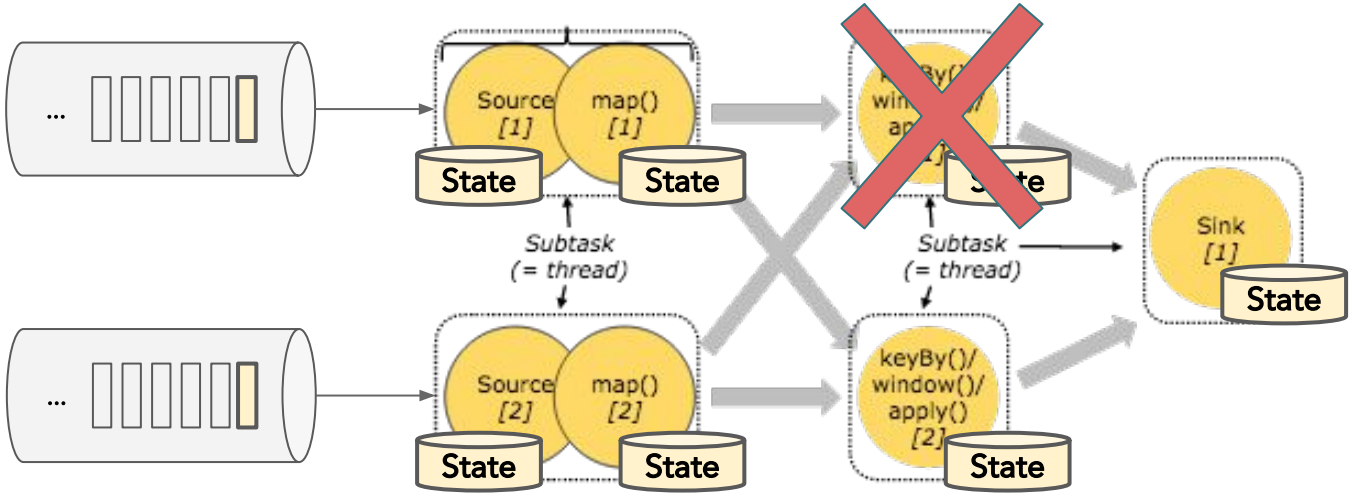
Snapshot*

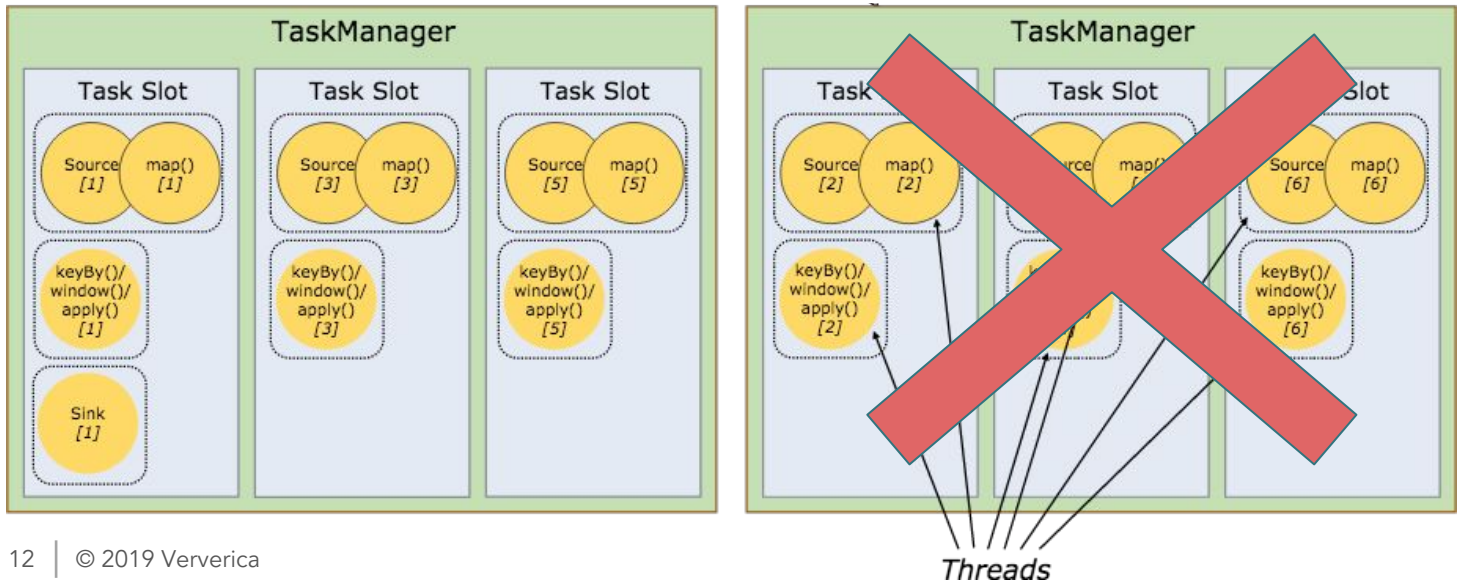
* Asynchronous Barrier Snapshotting
→ "checkpoint barriers" flow with data

Job Lifecycle & Fault Tolerance

On Failure

- TaskManager fails
  ⇒ cancel job on all TMs,
  ⇒ restart
  (may require new TMs/slots)

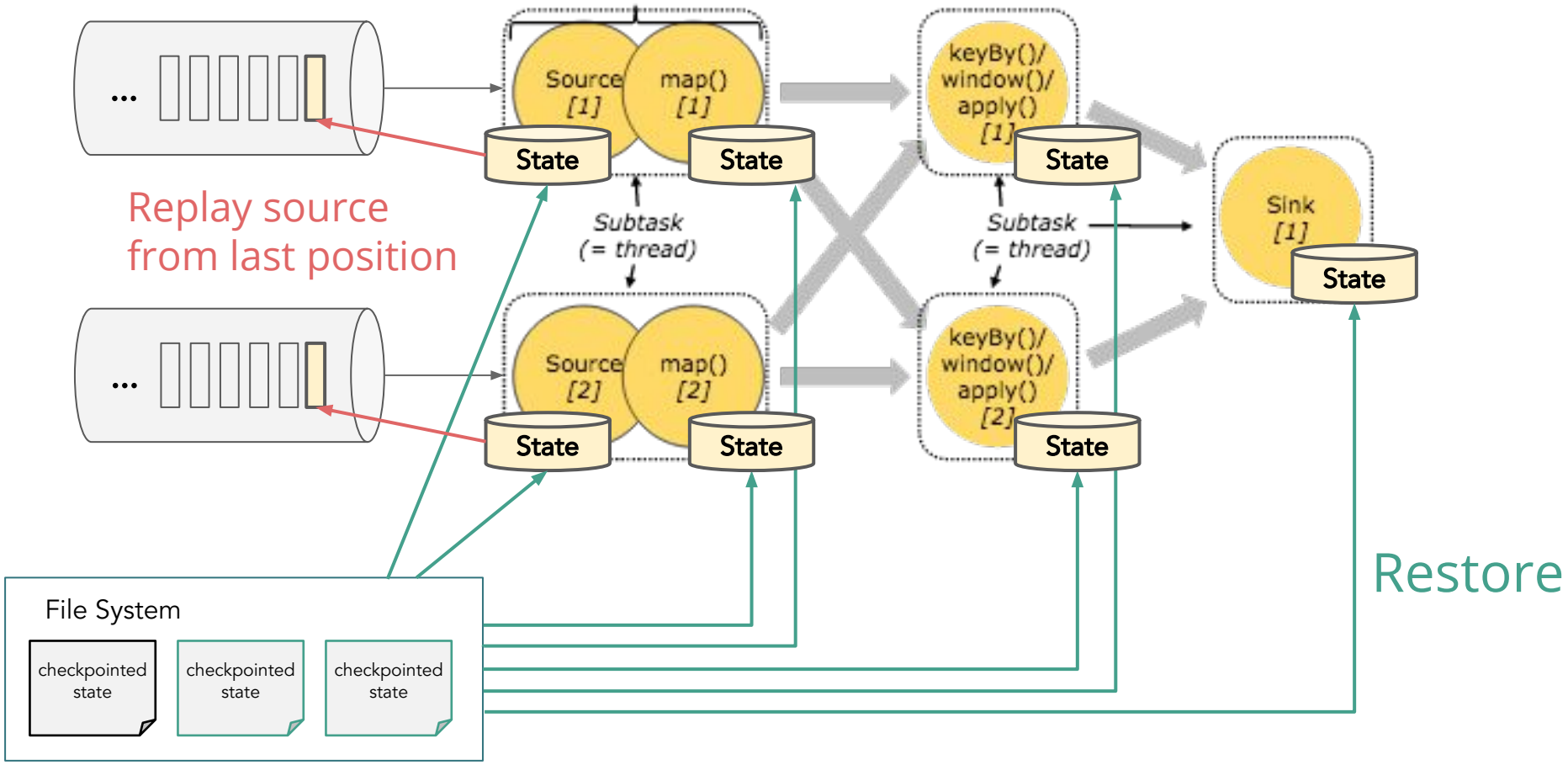




- Operator code fails
  ⇒ cancel job on all TMs,
  ⇒ restart
  (may require new TMs/slots)

Job Lifecycle & Fault Tolerance

After restart



 |

Exercises

Exercises

Getting Started

1. `Preparations`
   a. `git checkout`

      `https://github.com/ververica/flink-training-troubleshooting`

   b. **`git pull master origin`**

   c. Import project into IDE
2. Run the `TroubledStreamingJob` locally by executing `TroubledStreamingJobRunner`

   Note: The job is meant to restart periodically.
3. Run `TroubledStreamingJob` on Ververica Platform (see next slide)

Exercises

Getting Started

1. Login into https://training.flink-forward.org/ (Credentials provided)
2. Upload JAR
   a. `mvn package`
   b. `Upload via FLINK JARs-tab`
3. Create Deployment in Application Manager
   a. Upgrade Strategy: Stateless
   b. Initial State: RUNNING
   c. Restore Strategy: NONE
   d. Parallelism: 4
   e. Jobmanager/Taskmanager CPU: 1/1 (defaults)
   f. Jobmanager/Taskmanager Memory: 1G/2G (defaults)
   g. entryClass: com.ververica.training.TroubledStreamingJob

Getting Started Walk-Through

Exercises

Debugging Frequent Job Failures

Exercise 1
As we just saw `TroubledStreamingJob` is restarting frequently. Locate the problem and deploy a fix for it to Ververica Platform.

Note: Classes/Methods annotated with **`@DoNotTouchThis`** should not be modified during the training.

nico@ververica.com www.ververica.com @VervericaData