Checkpointing

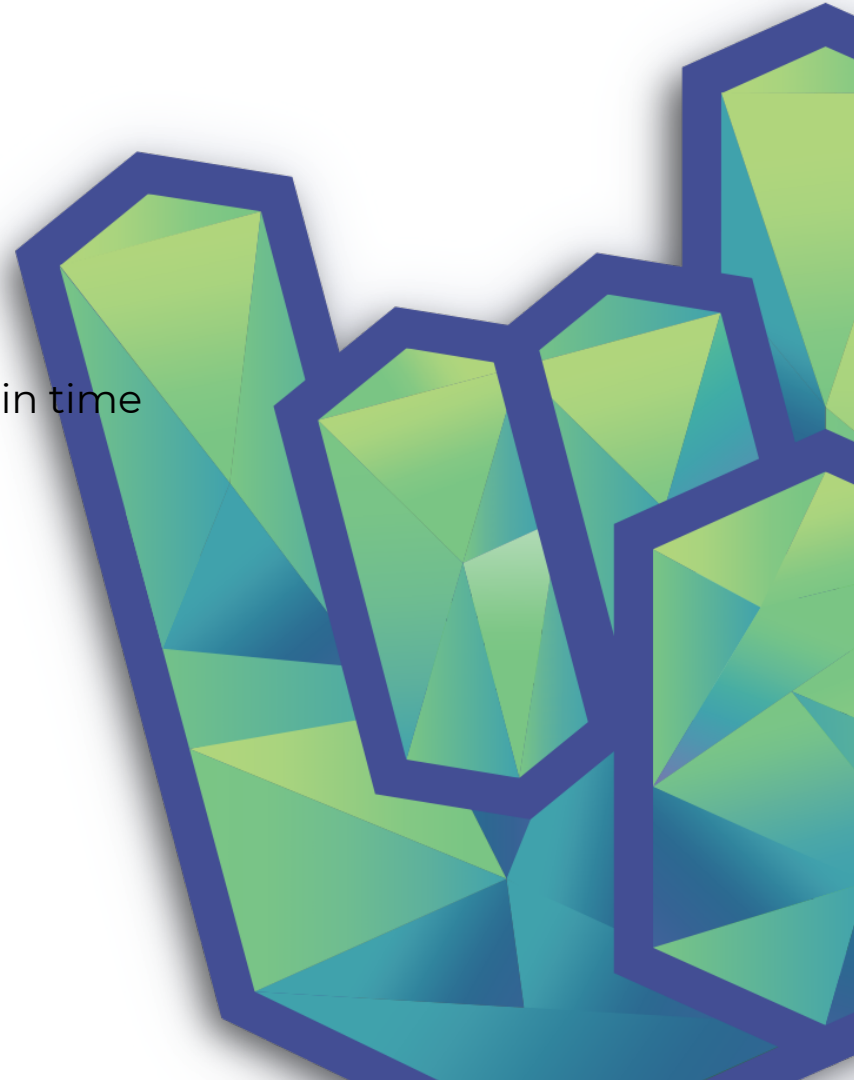
Checkpoints

Checkpoint = the entire state at an exact point in time

Distributed system, inherently unreliable

Need to deal with failures

- killed processes
- failed/unreachable machines



Checkpoints - Naive

Taking a checkpoint

- pause the application and data ingestion
- wait for in-flight data to be processed
- copy the state to the checkpoint backend
- resume data ingestion

Restoring a checkpoint after failure

- restart the entire application
- copy the checkpoint data to all the stateful tasks
- resume data ingestion

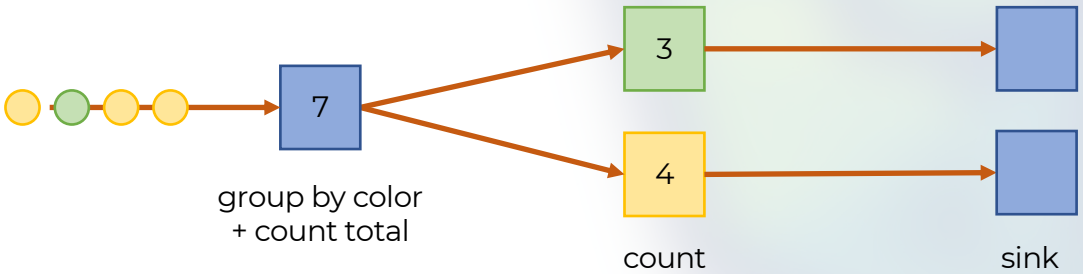
In modern data infrastructures, this approach is unacceptable

- prolonged "hiccups" in data updates (read: increased latency)
- the incoming data might be too much to handle after resuming

Checkpoints - Flink

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



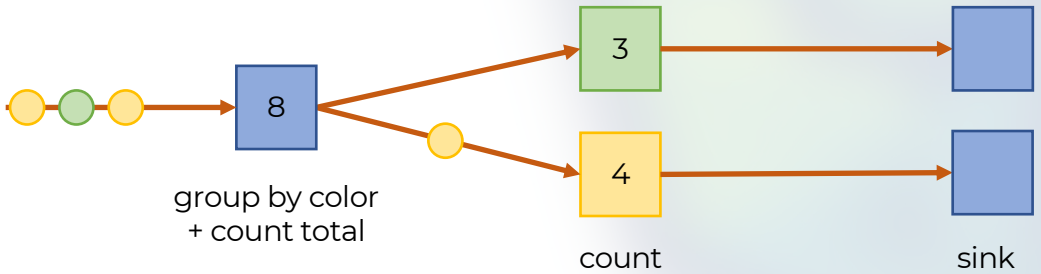
Checkpoints - Flink

Steps

- the stream is running

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



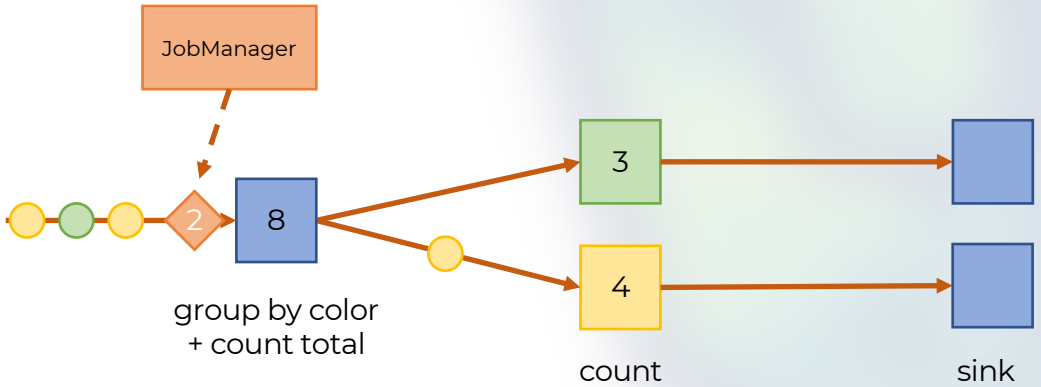
Checkpoints - Flink

Steps

- the stream is running
- the JobManager adds a checkpoint barrier

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



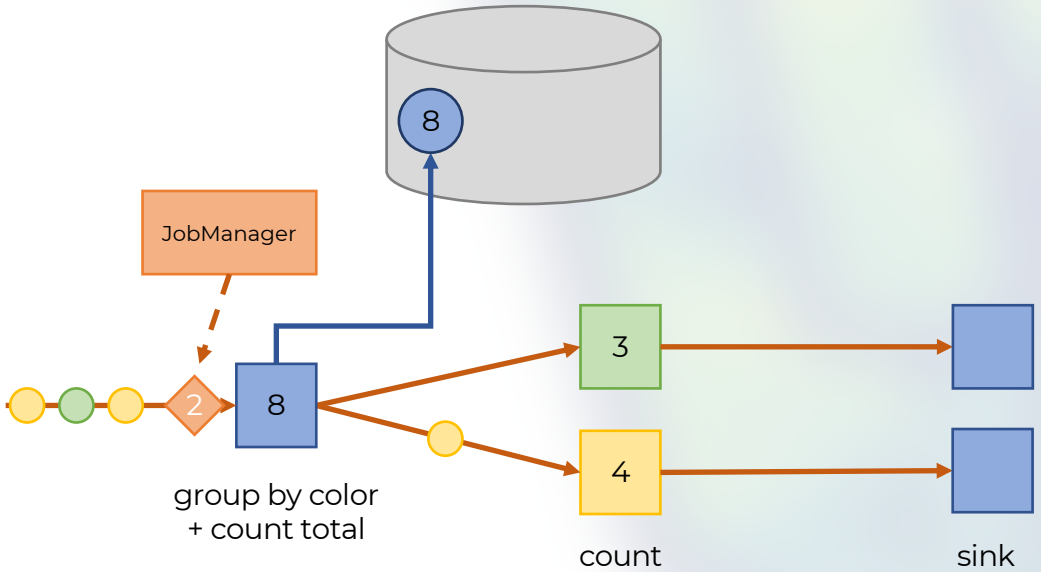
Checkpoints - Flink

Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



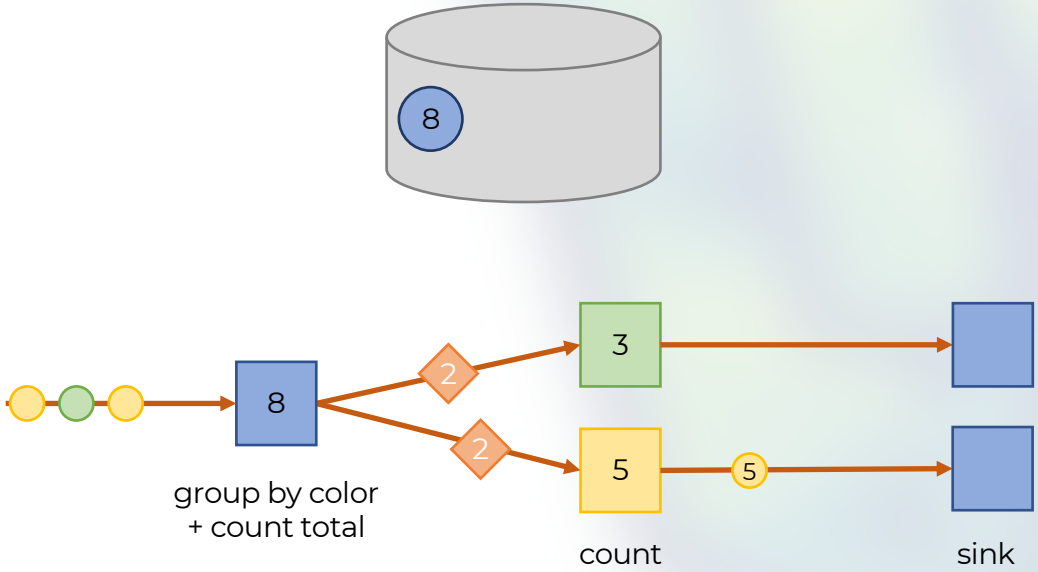
Checkpoints - Flink

Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



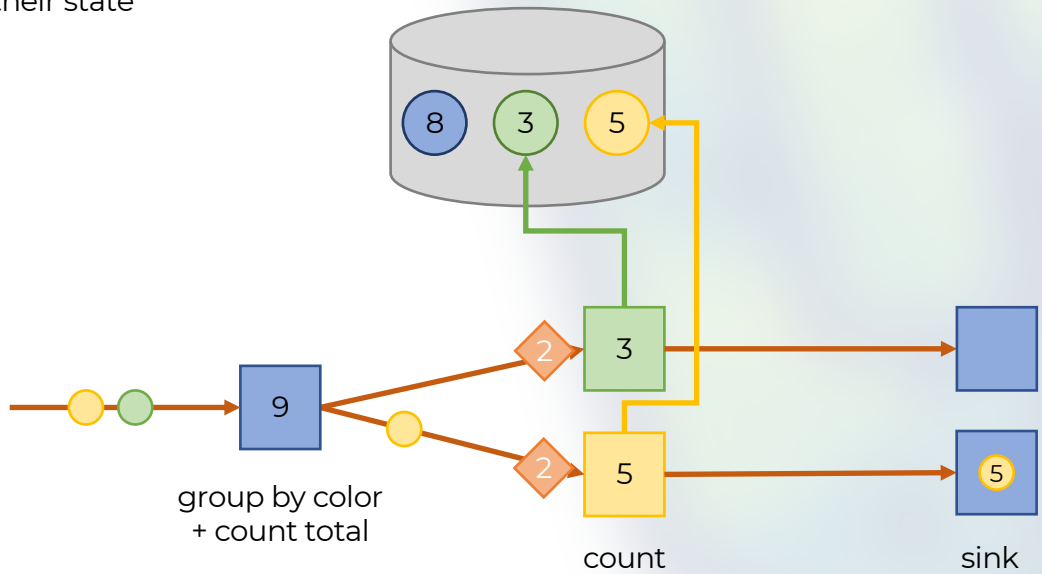
Checkpoints - Flink

Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier
- barrier arrives at tasks, they save their state

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



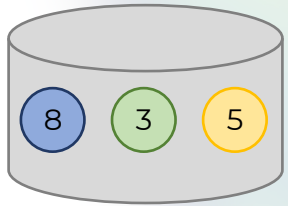
Checkpoints - Flink

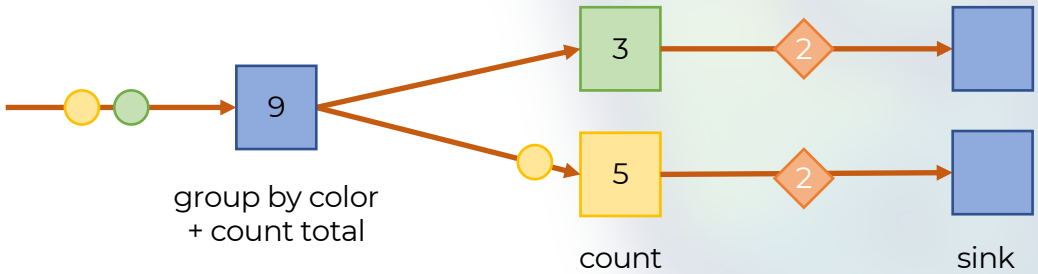
Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier
- barrier arrives at tasks, they save their state
- if new elements arrive while saving, buffer
- barrier moves forward, elements continue to flow

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored





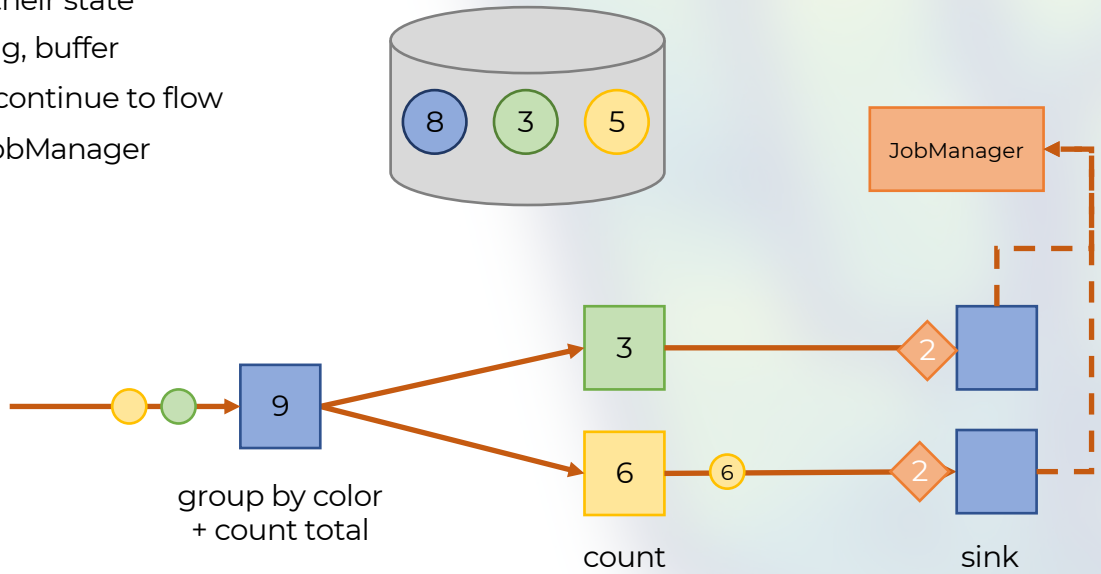
Checkpoints - Flink

Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier
- barrier arrives at tasks, they save their state
- if new elements arrive while saving, buffer
- barrier moves forward, elements continue to flow
- the sinks ack the checkpoint to JobManager

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored



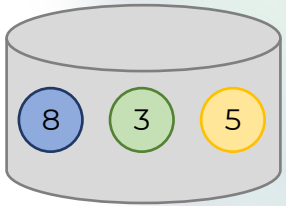
Checkpoints - Flink

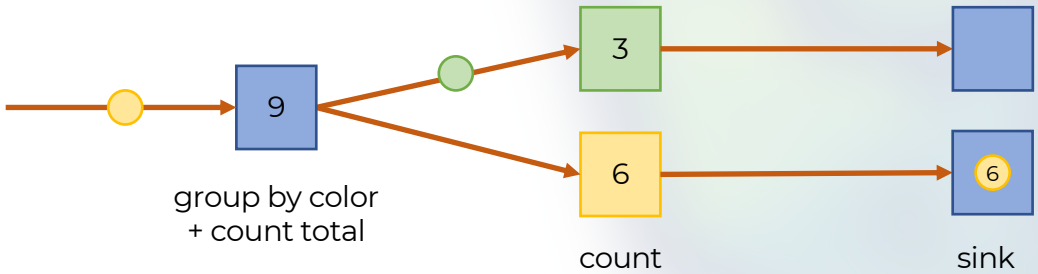
Steps

- the stream is running
- the JobManager adds a checkpoint barrier
- task saves its state and forwards the barrier
- barrier arrives at tasks, they save their state
- if new elements arrive while saving, buffer
- barrier moves forward, elements continue to flow
- the sinks ack the checkpoint to JobManager
- checkpoint complete, move on

A rolling checkpoint algorithm

- checkpoint barrier emitted in line with the data
- as the barrier arrives at tasks, they store their state
- incoming elements are buffered until state is stored





Flink rocks