Help I need a stream processor learning to chose between Spark, Flink, Samza, and Storm

Andrew Psaltis

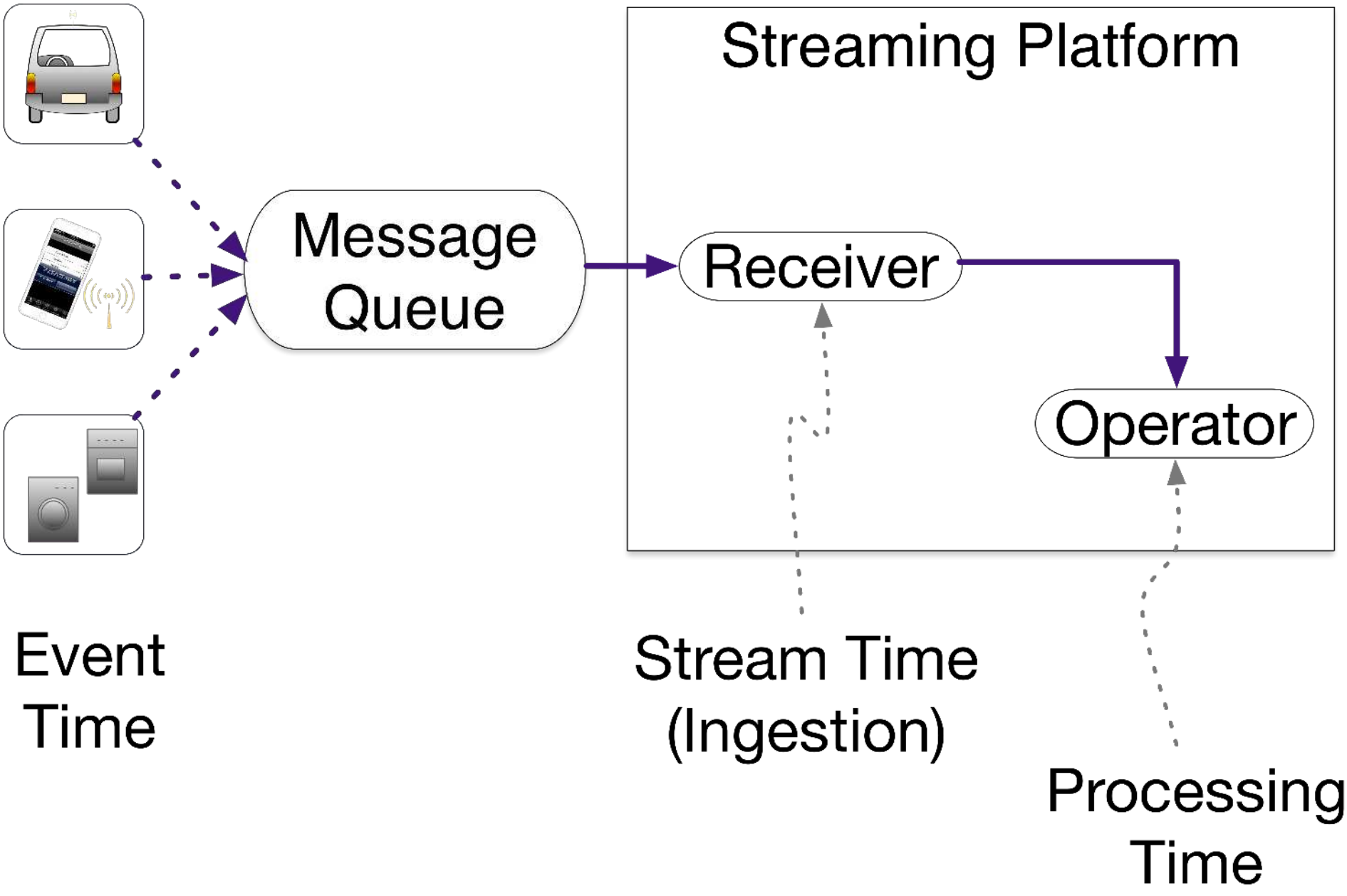
HDF/IoT/Cybersecurity Architect

apsaltis@hortonworks.com

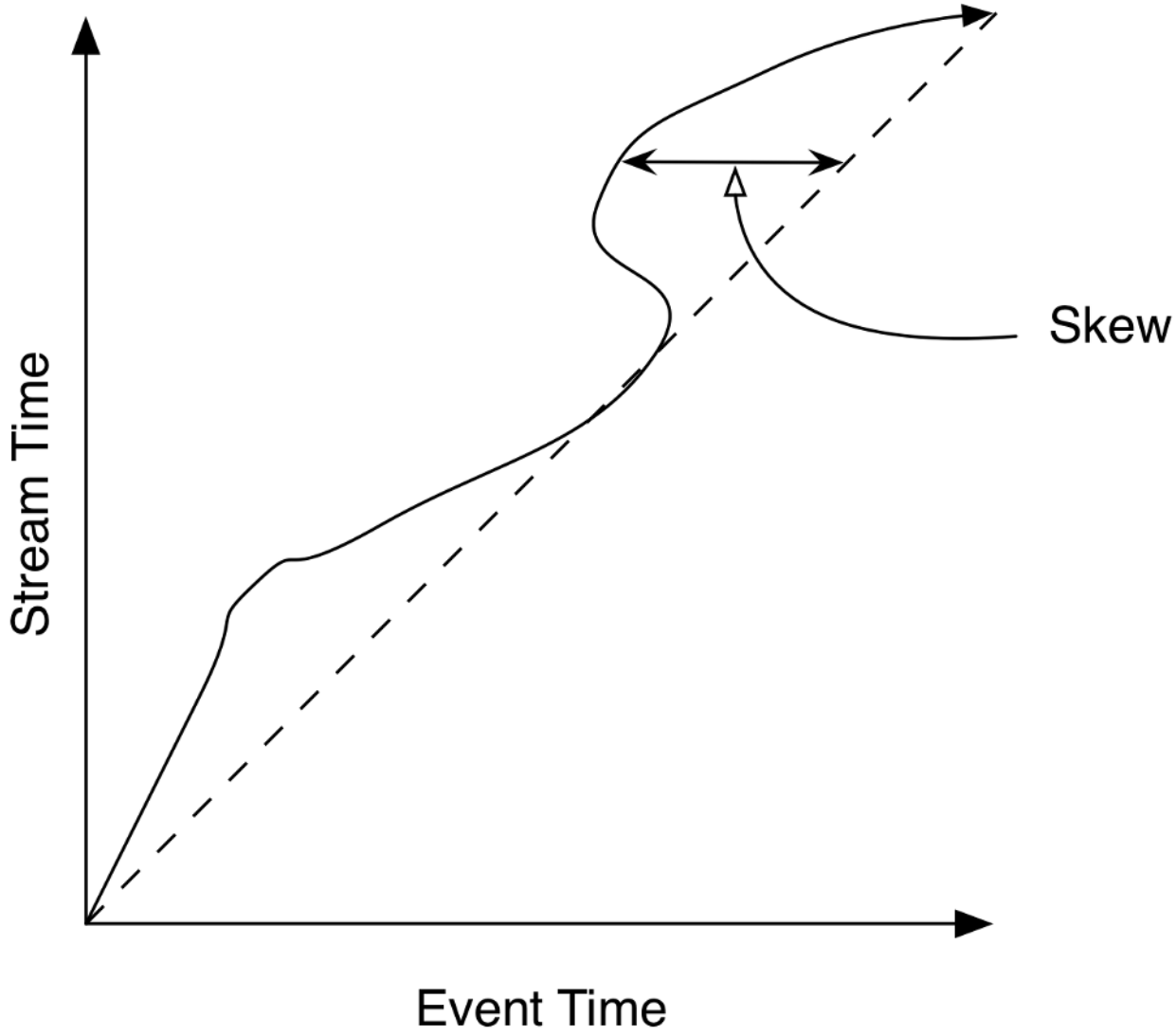
@itmdata

June 7, 2017

Thinking about time


Streaming Platform
Message Queue
Receiver
Operator
Event Time
Stream Time (Ingestion)
Processing Time

Time Skew

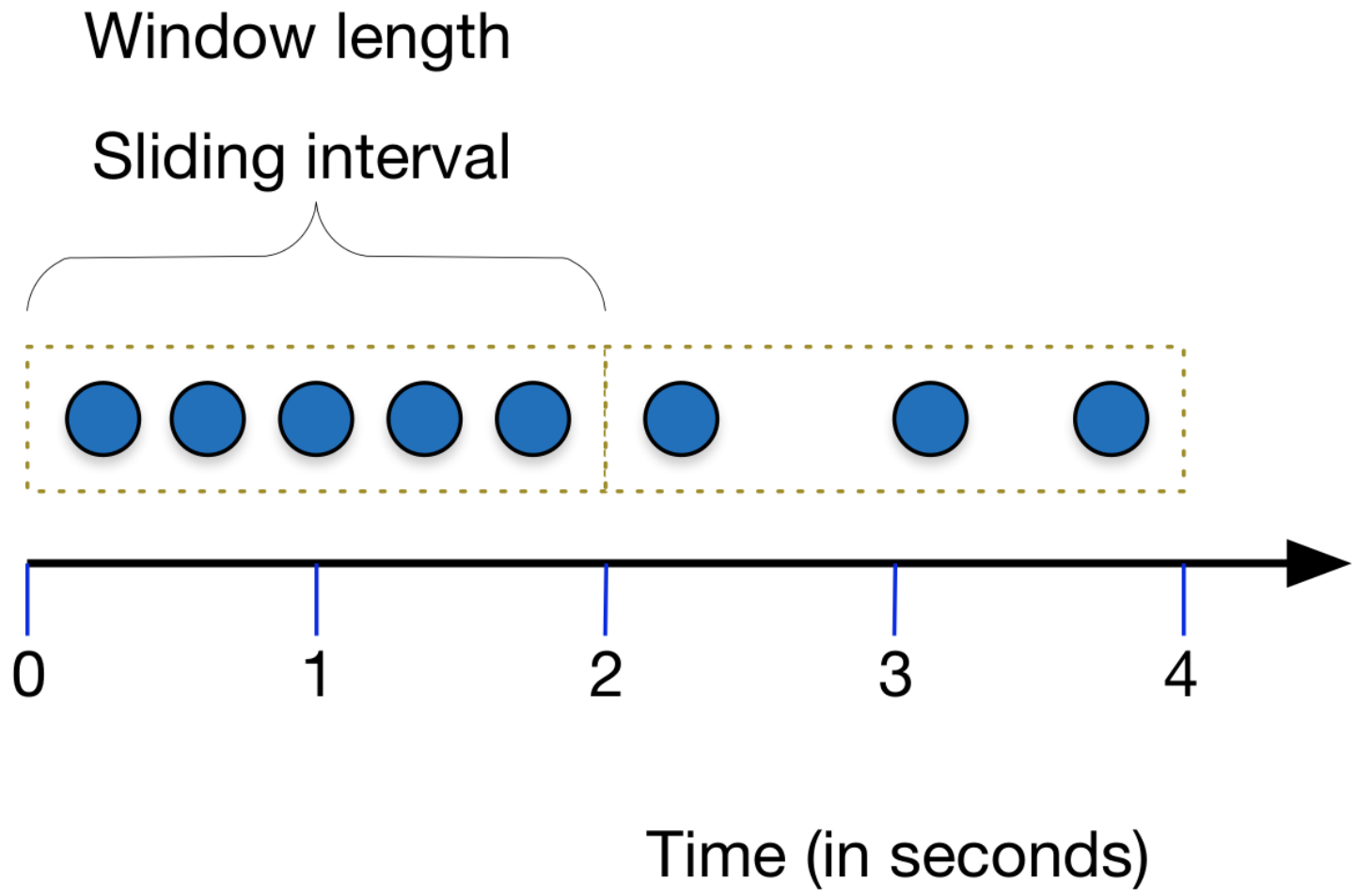


	Storm	Spark	Flink	Samza
Event	✔		✔	
Stream		✔	✔	✔
Processing	✔		✔	

Windows

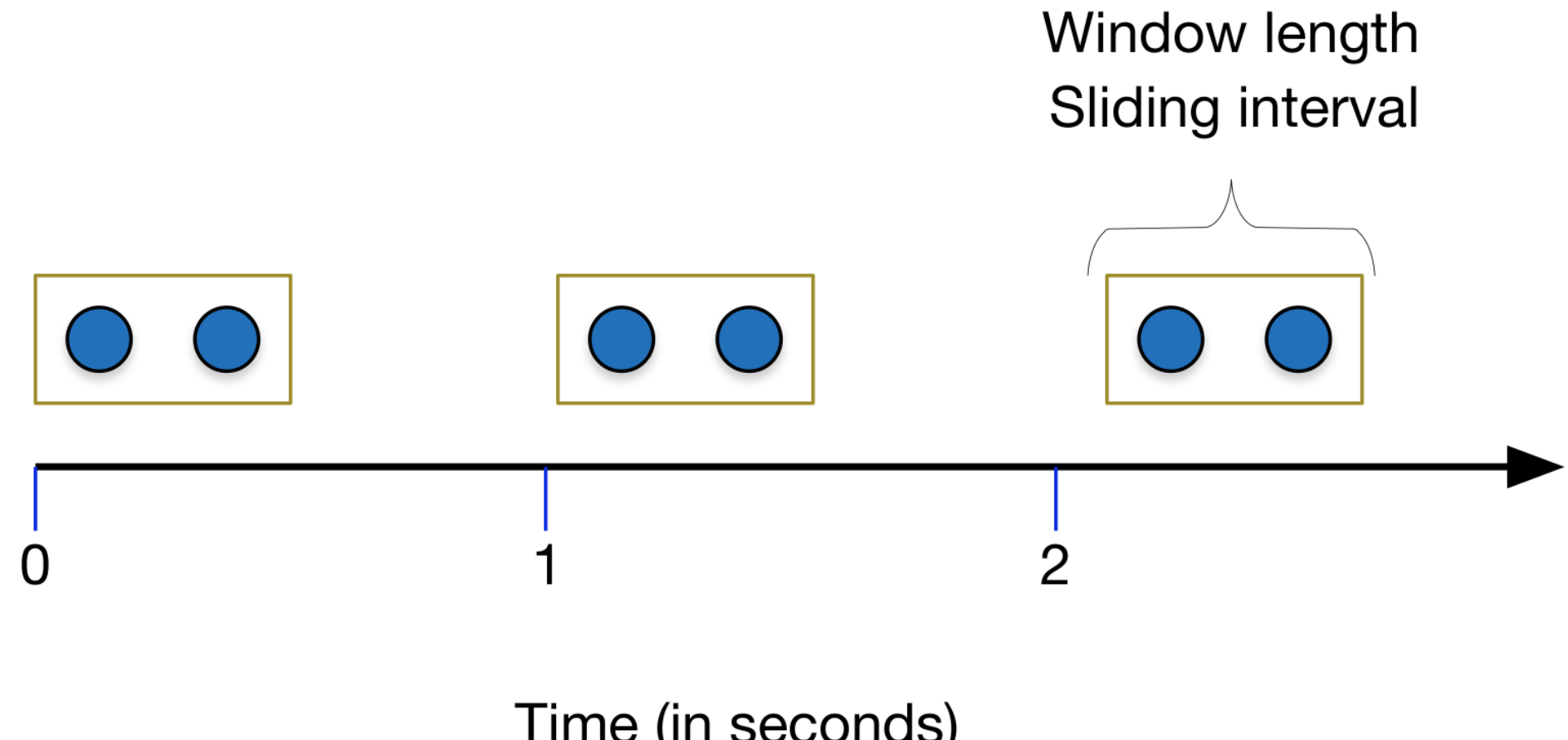
Tumbling Windows

Tumbling Time Windowing



Tumbling temporal window

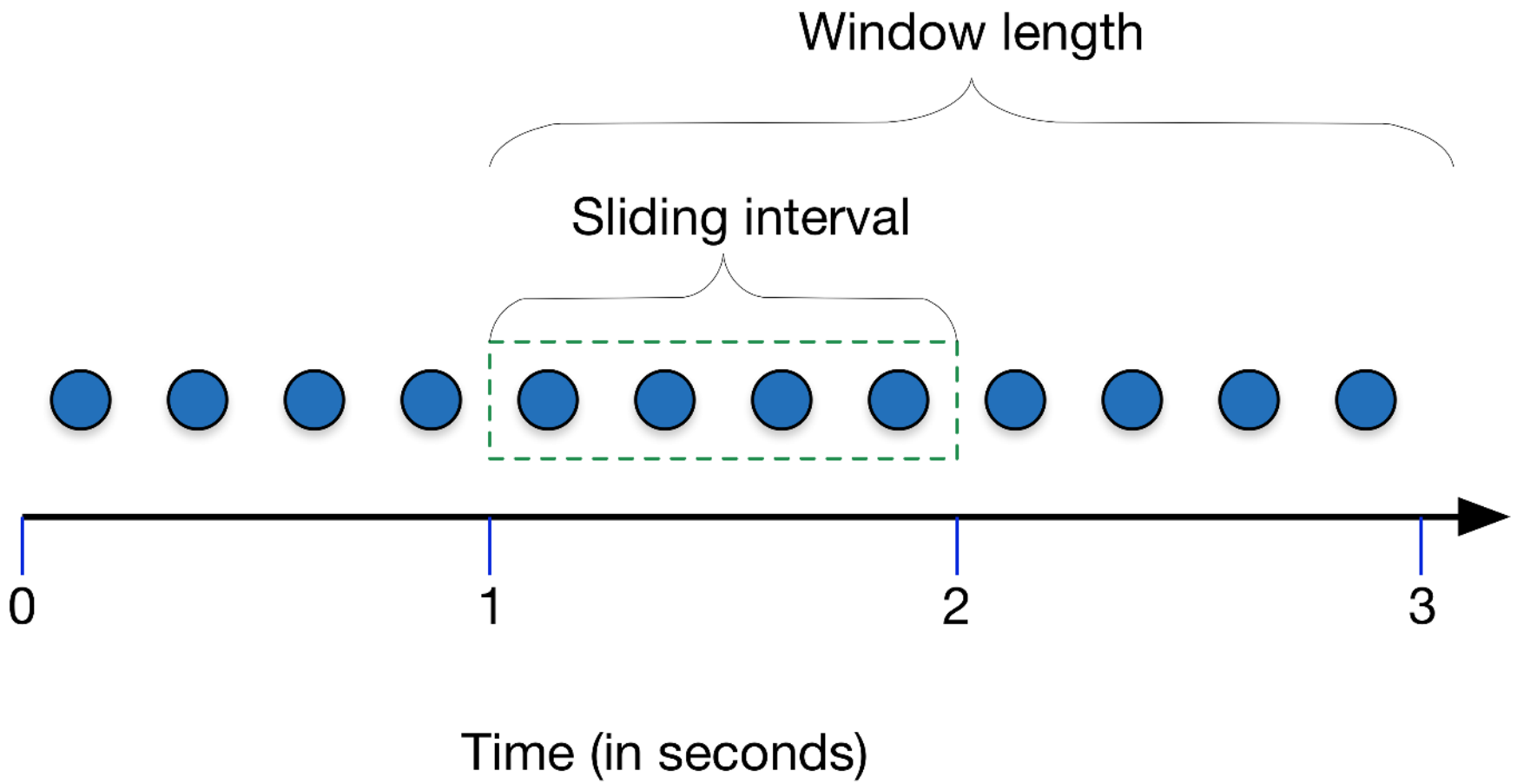
Tumbling Count Windowing



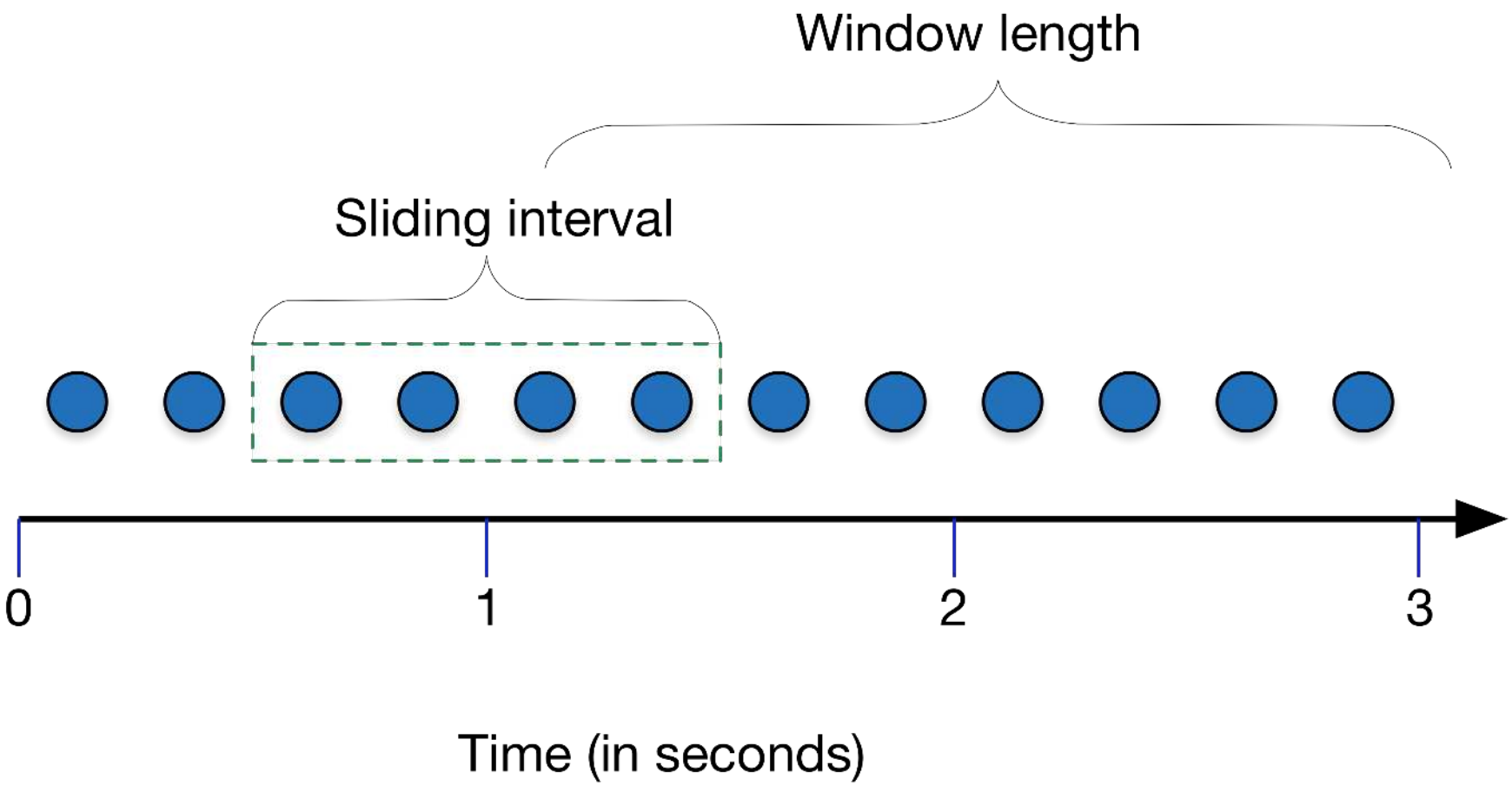
Tumbling count-window

Sliding Windows

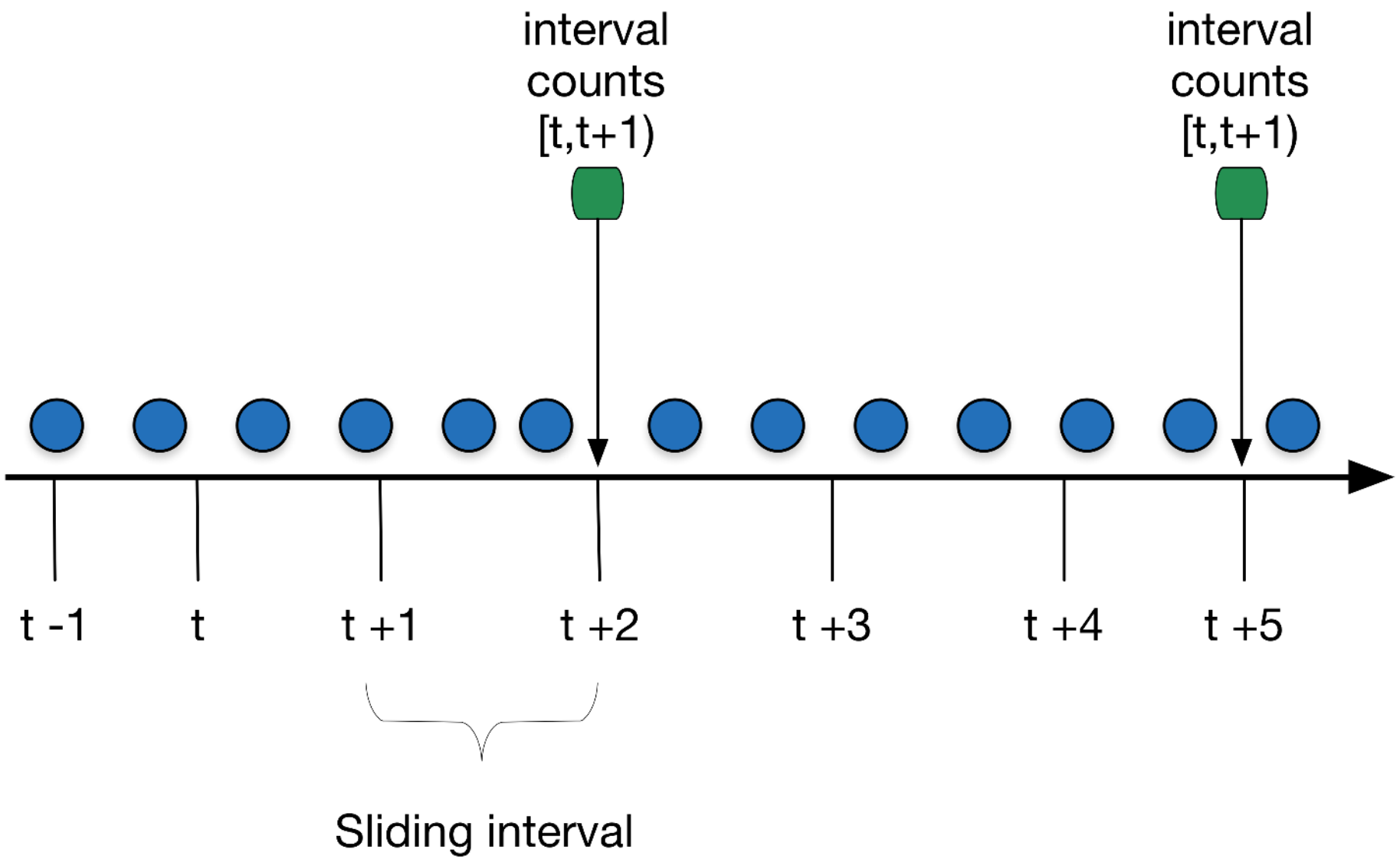
Sliding Time Window



Sliding Count Window



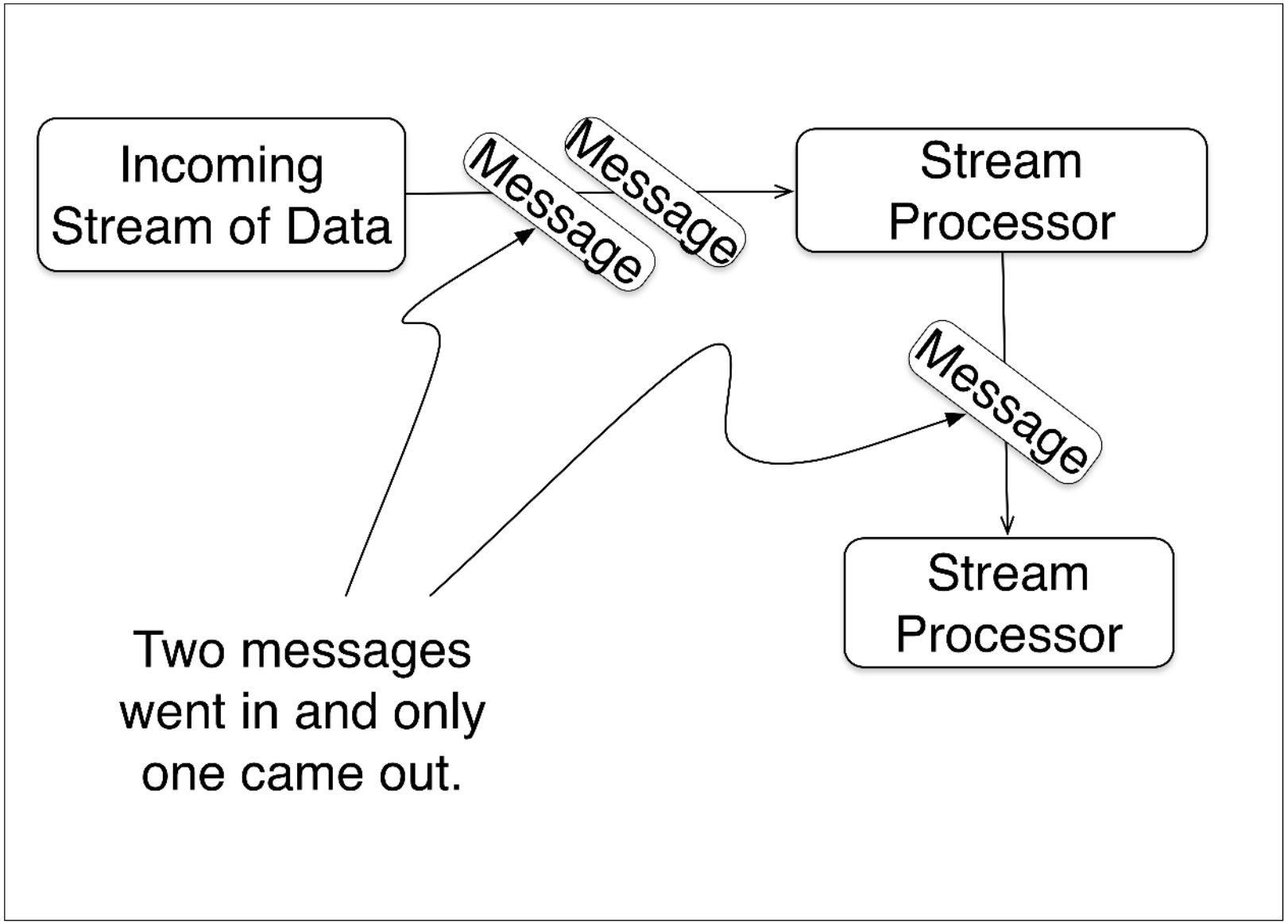
Handling Out of Order Data



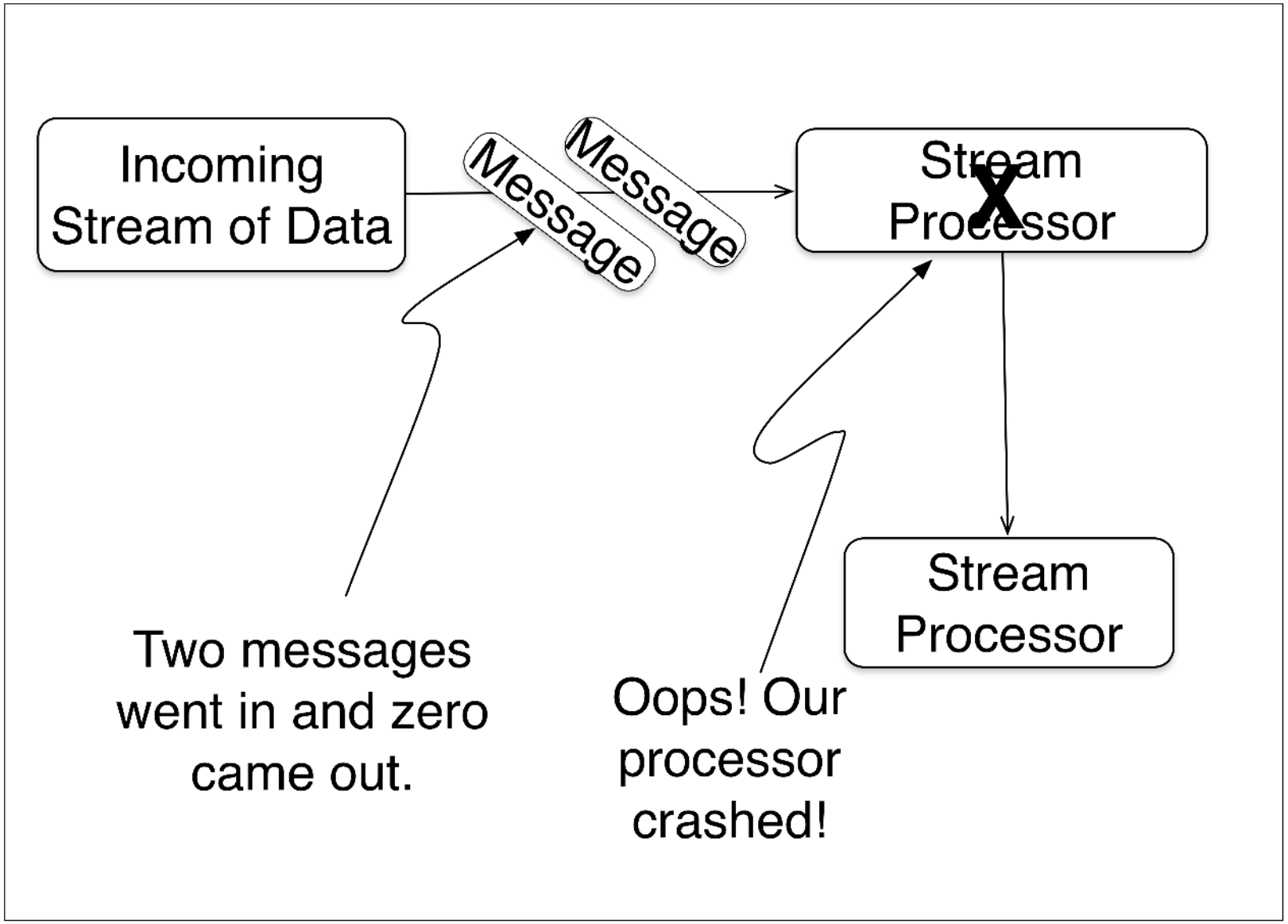
	Storm	Spark	Flink	Samza
Tumble Temporal	✔		✔	
Tumble Count	✔		✔	
Sliding Time	✔	✔	✔	✔^
Sliding Count	✔		✔	
Custom / Advanced			✔	
Out of Order Data	Discarded	✔	✔	

Processing Semantics

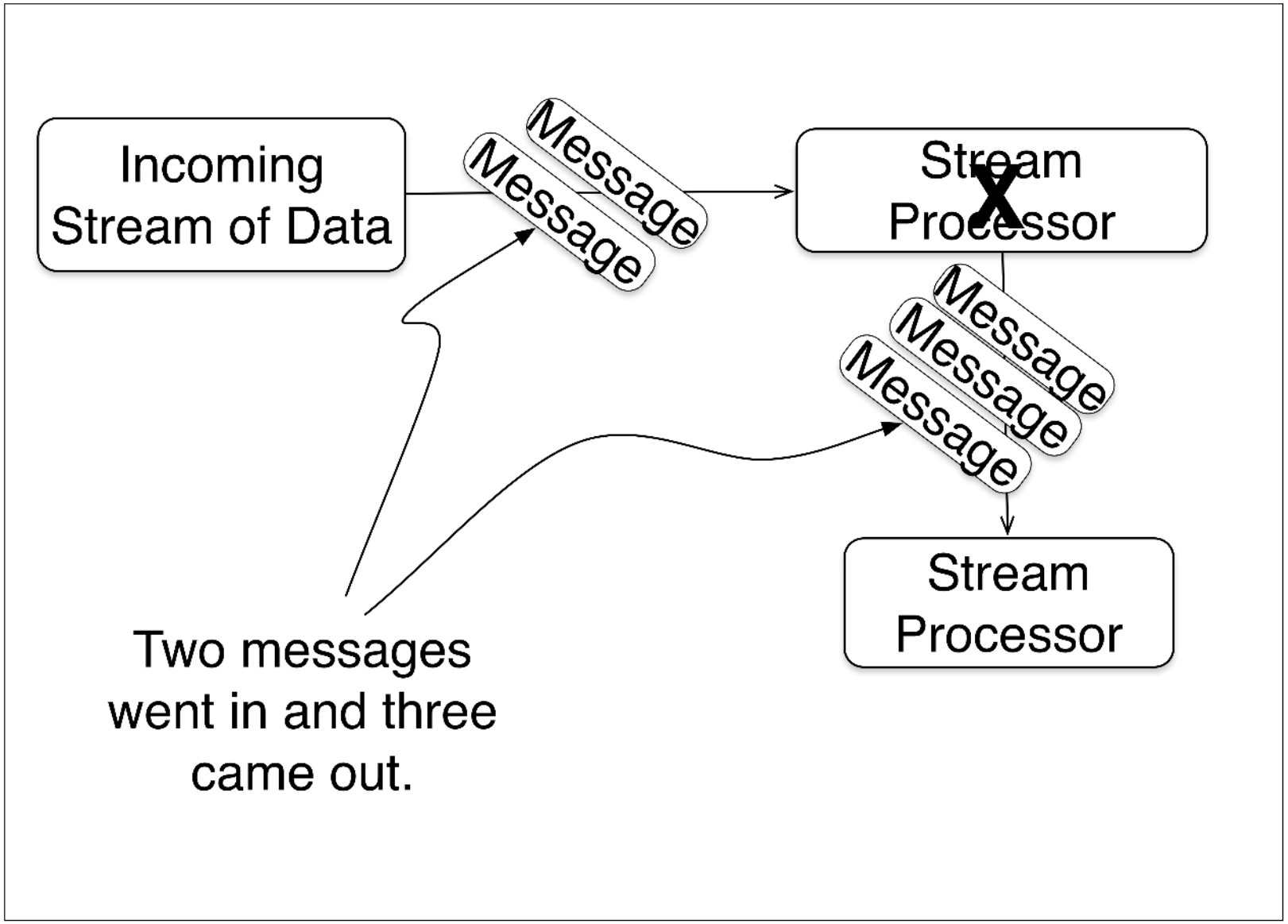
At-most-once


Incoming Stream of Data
Message
Message
Stream Processor
Message
Stream Processor
Two messages went in and only one came out.

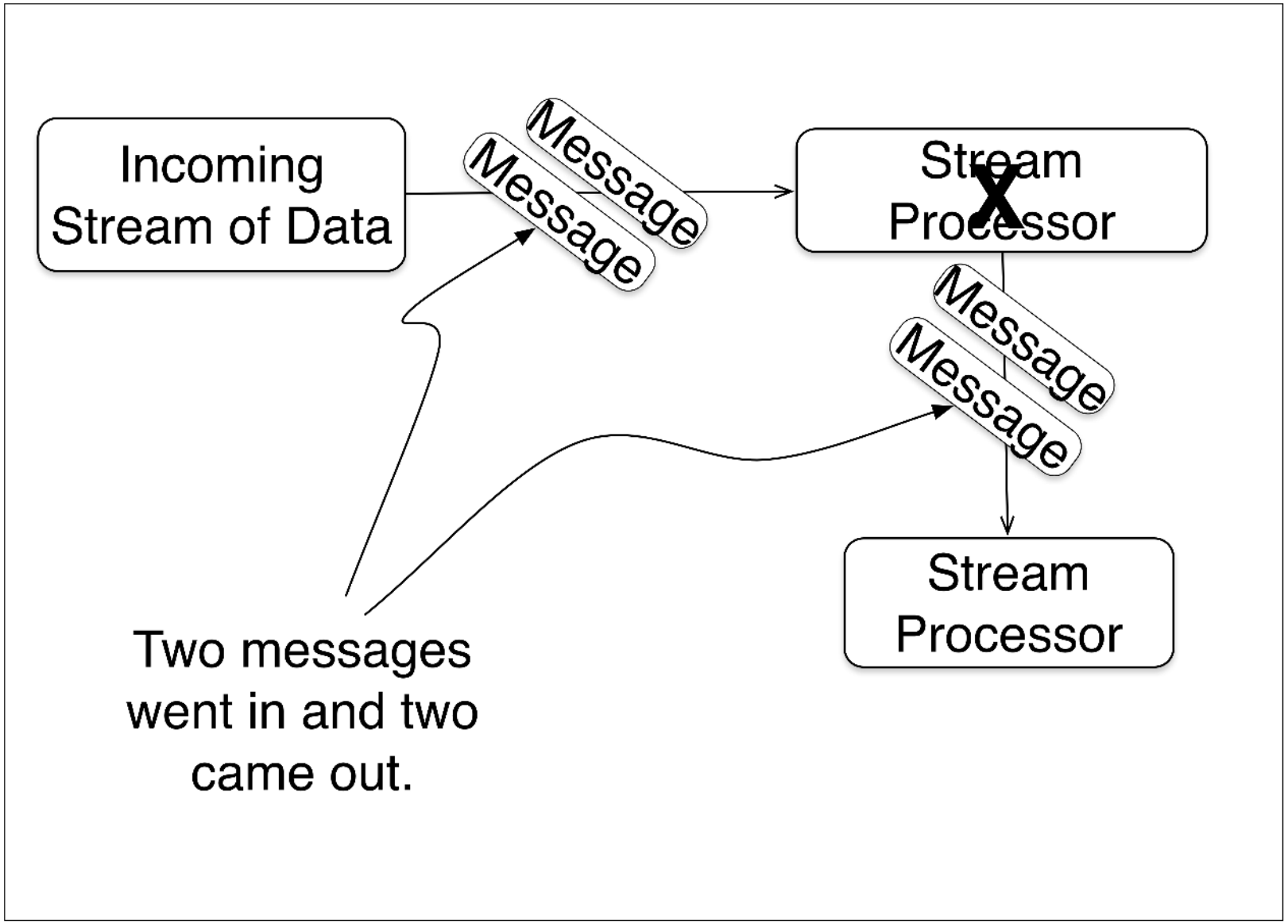
At-most-once


Incoming Stream of Data
Message
Message
Stream Processor
X
Stream Processor
Two messages went in and zero came out.
Oops! Our processor crashed!

At-least-once



Exactly-once

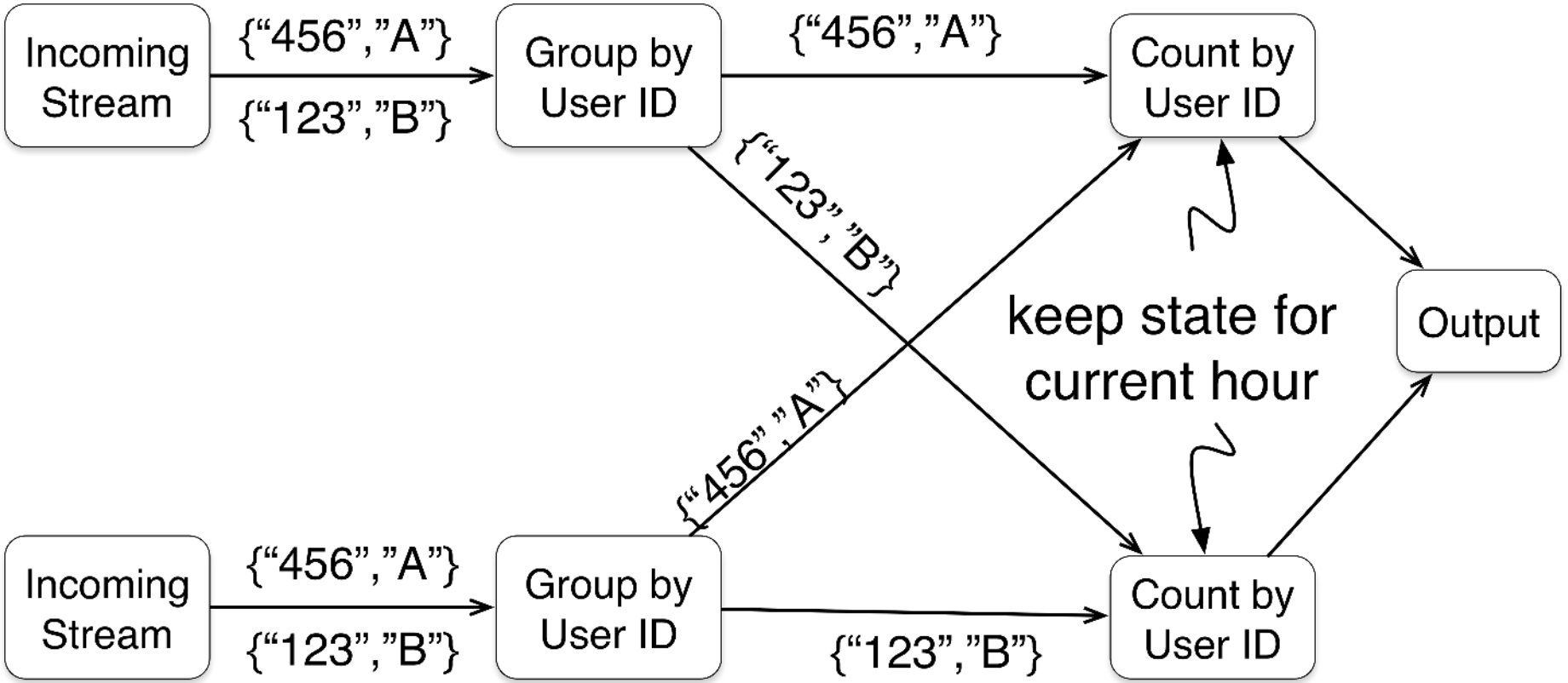


	Storm	Spark	Flink	Samza
At-most once	✔	✔	✔	
At-least once	✔ (Non–Tx)	✔	✔	✔
Exactly once	✔ (Trident)	✔	✔	
Best effort	✔			

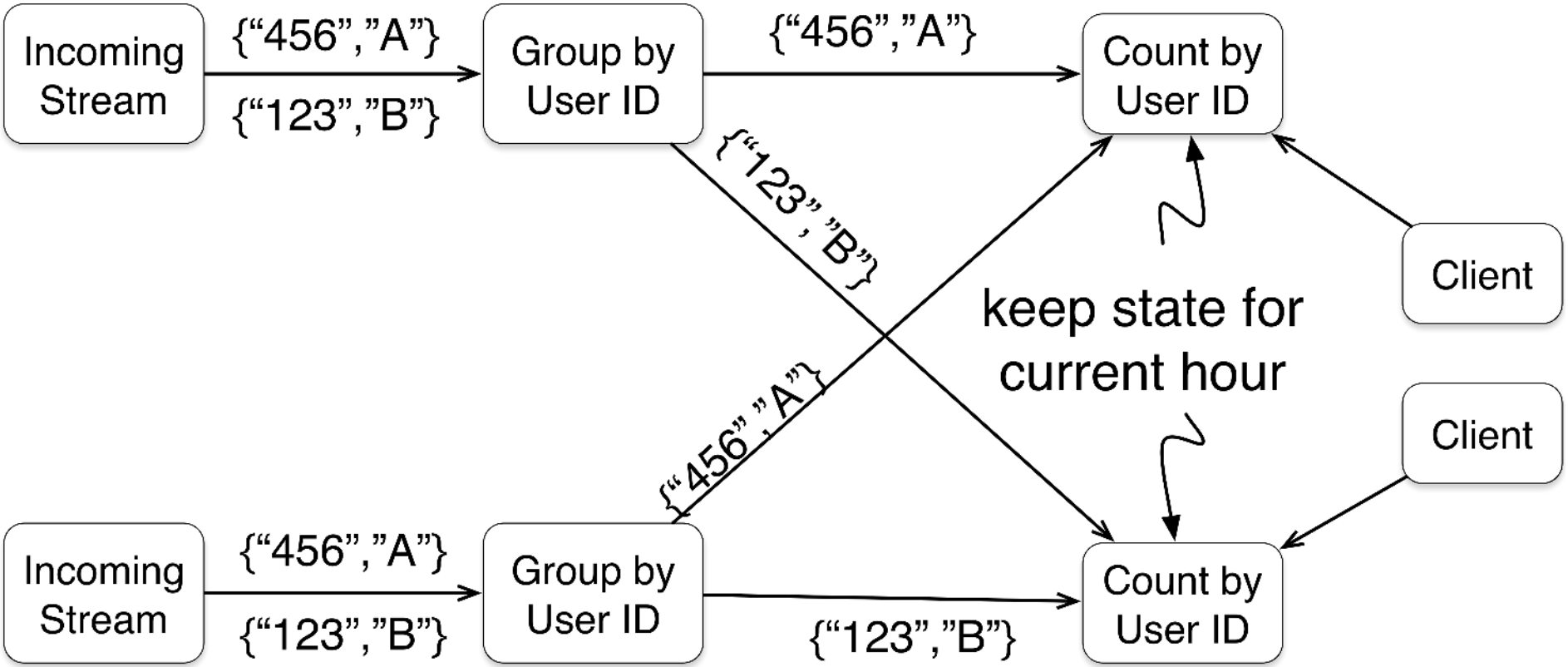
2 Types of State

Application (User-Defined) State

Keeping Simple State



Queryable Simple State

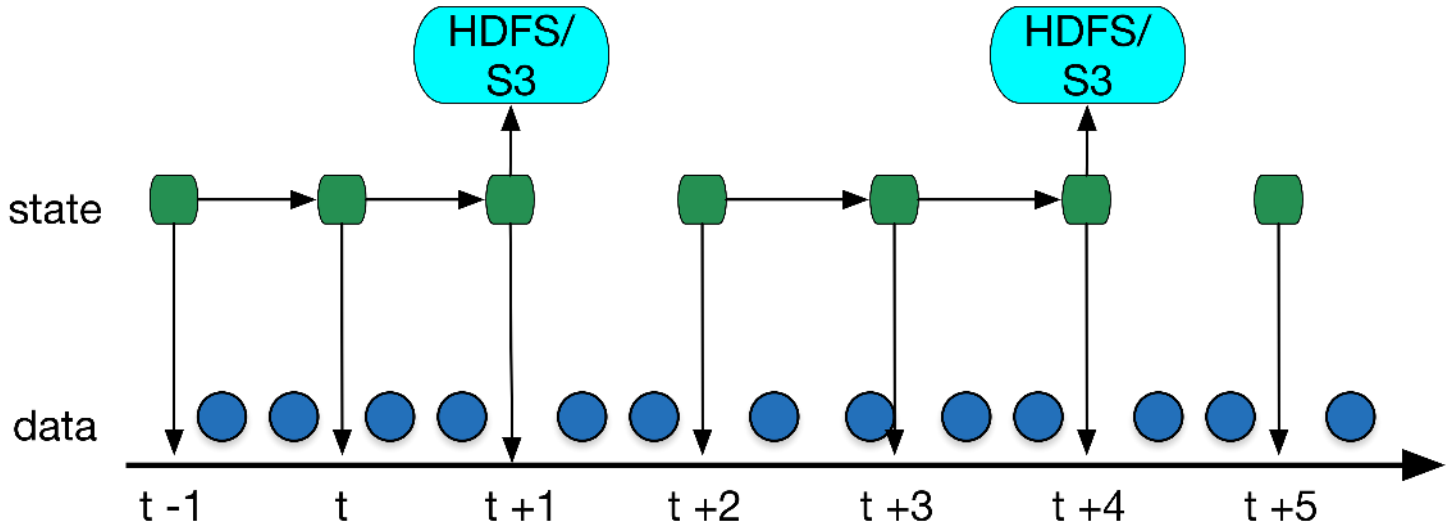


[Flink Queryable State (FLINK-3779)]

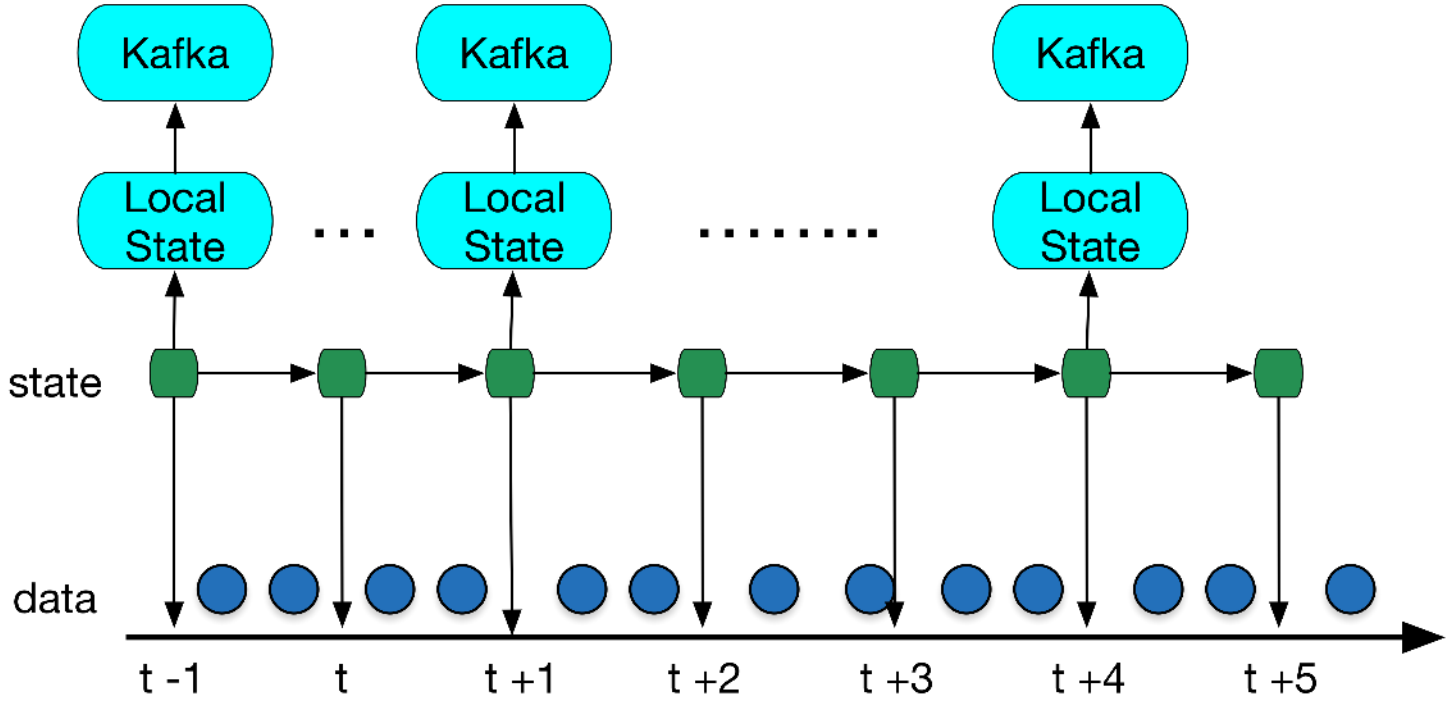

In-Memory
Replicated Queryable
Persistent Storage
Complexity and Features
Low
High

System State

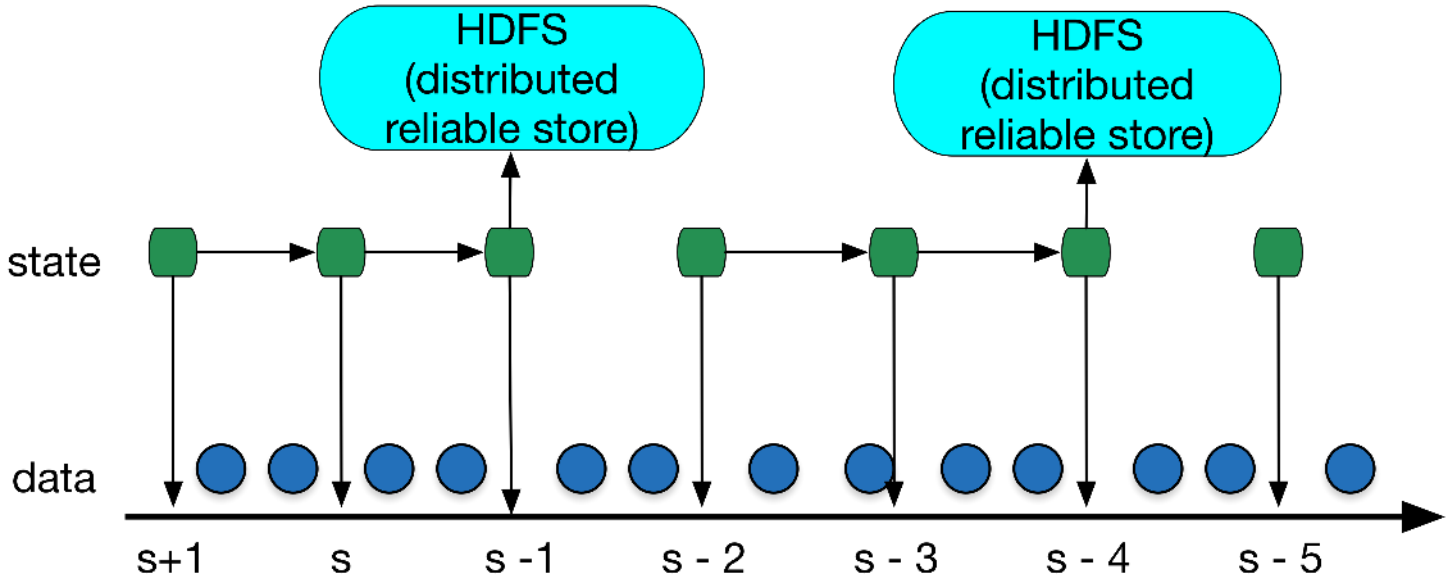
Spark Checkpointing



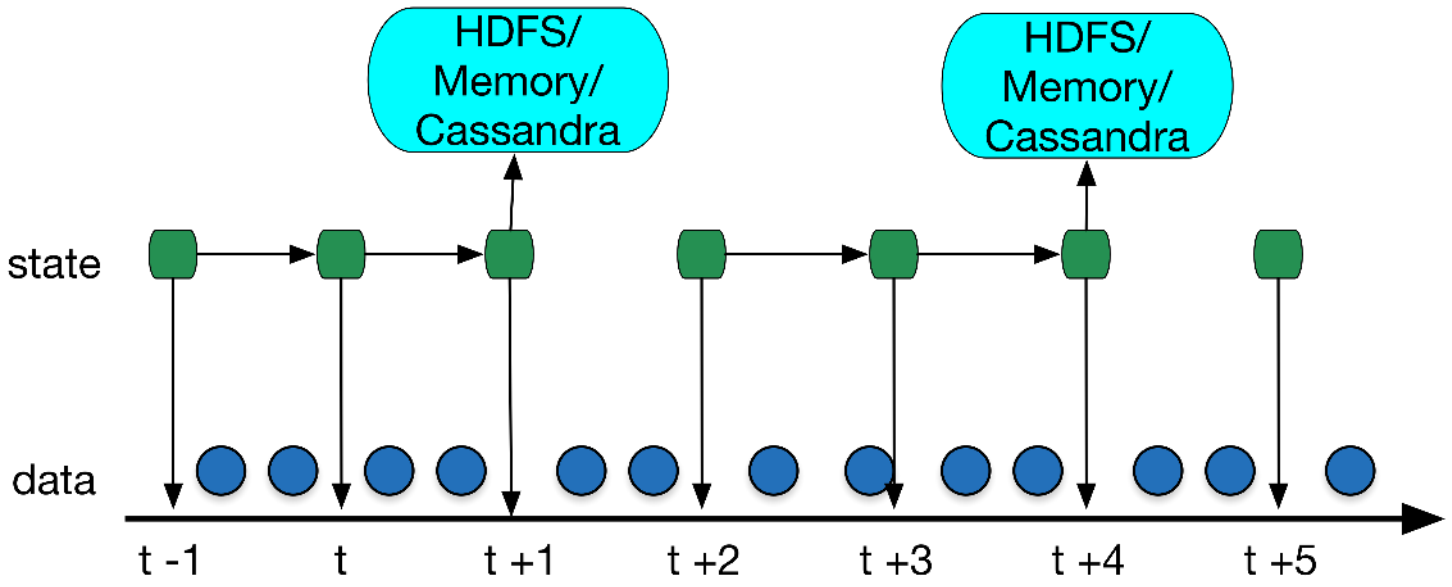
Samza Checkpointing



Flink Checkpointing



Storm (Trident) Checkpointing



	Storm (Trident)	Spark	Flink	Samza
At-least once				✔
Exactly once	✔ (We do the work)	✔	✔	

Thank You

Andrew Psaltis

HDF/IoT/Cybersecurity Architect

apsaltis@hortonworks.com

@itmdata

HORTONWORKS®
POWERING THE FUTURE OF DATA™