

Robust Stream Processing with Apache Flink

Jamie Grier
@jamiegrier
jamie@data-artisans.com



dataArtisans



Who am I?



- Director of Applications Engineering at data Artisans
- Previously working on streaming computation at Twitter, Gnip and Boulder Imaging
- Involved in various kinds of stream processing for about a decade
- High-speed video, social media streaming, general frameworks for stream processing



Overview

- What is Apache Flink?
- What is **Stateful** Stream Processing?
- Windowed computation over streams
- **Robust** Time Handling (Event Time vs Processing Time)
- **Robust** Failure Handling
- **Robust** Planned Downtime Handling
- **Robust** Reprocessing

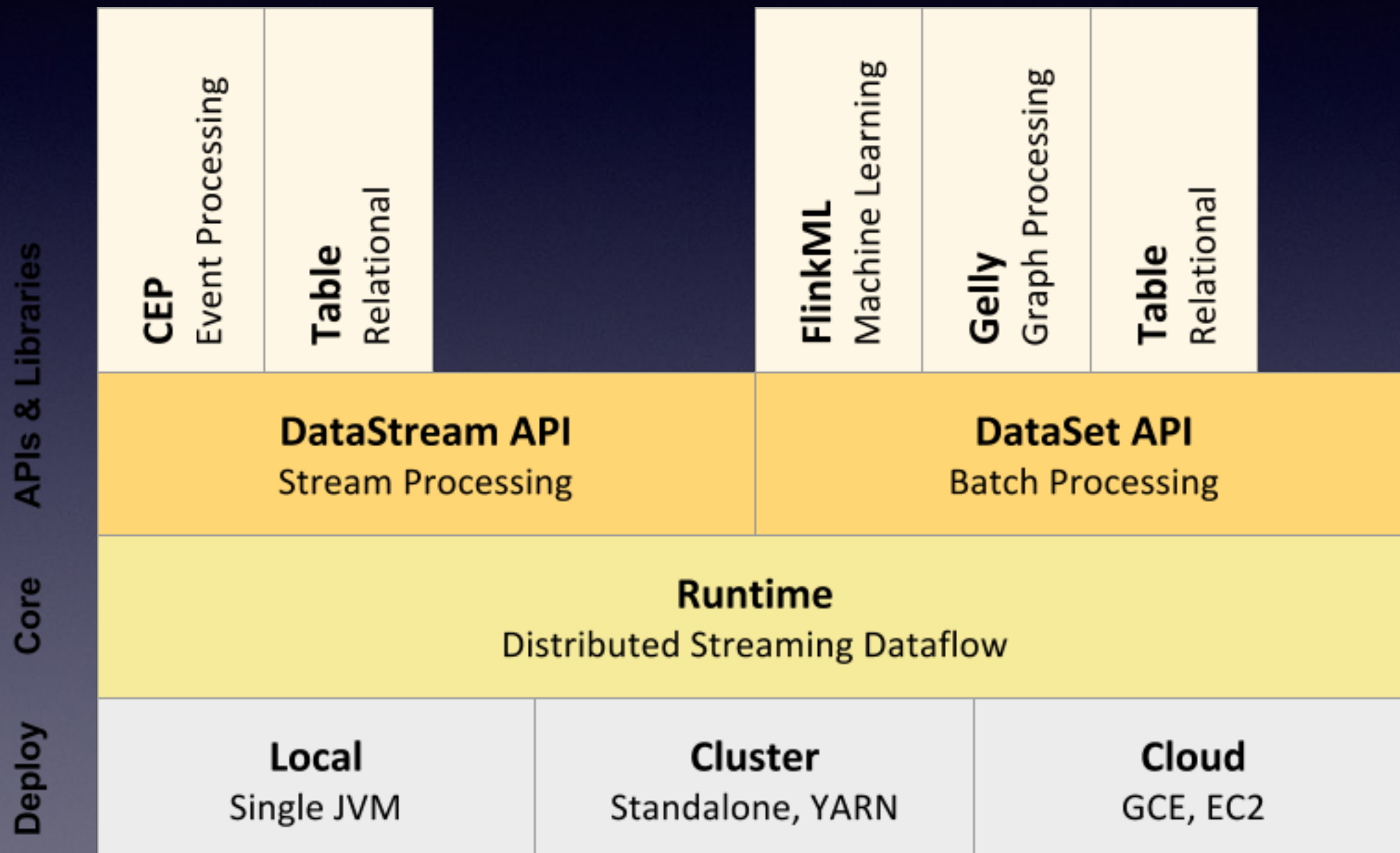


What is Apache Flink?

Apache Flink is an open source platform for distributed stream and batch data processing.

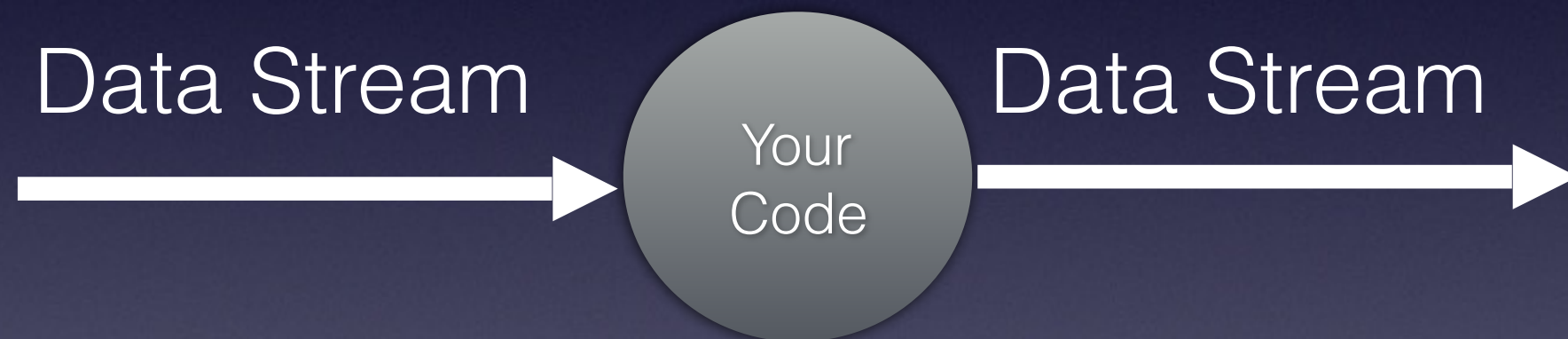


What is Apache Flink?



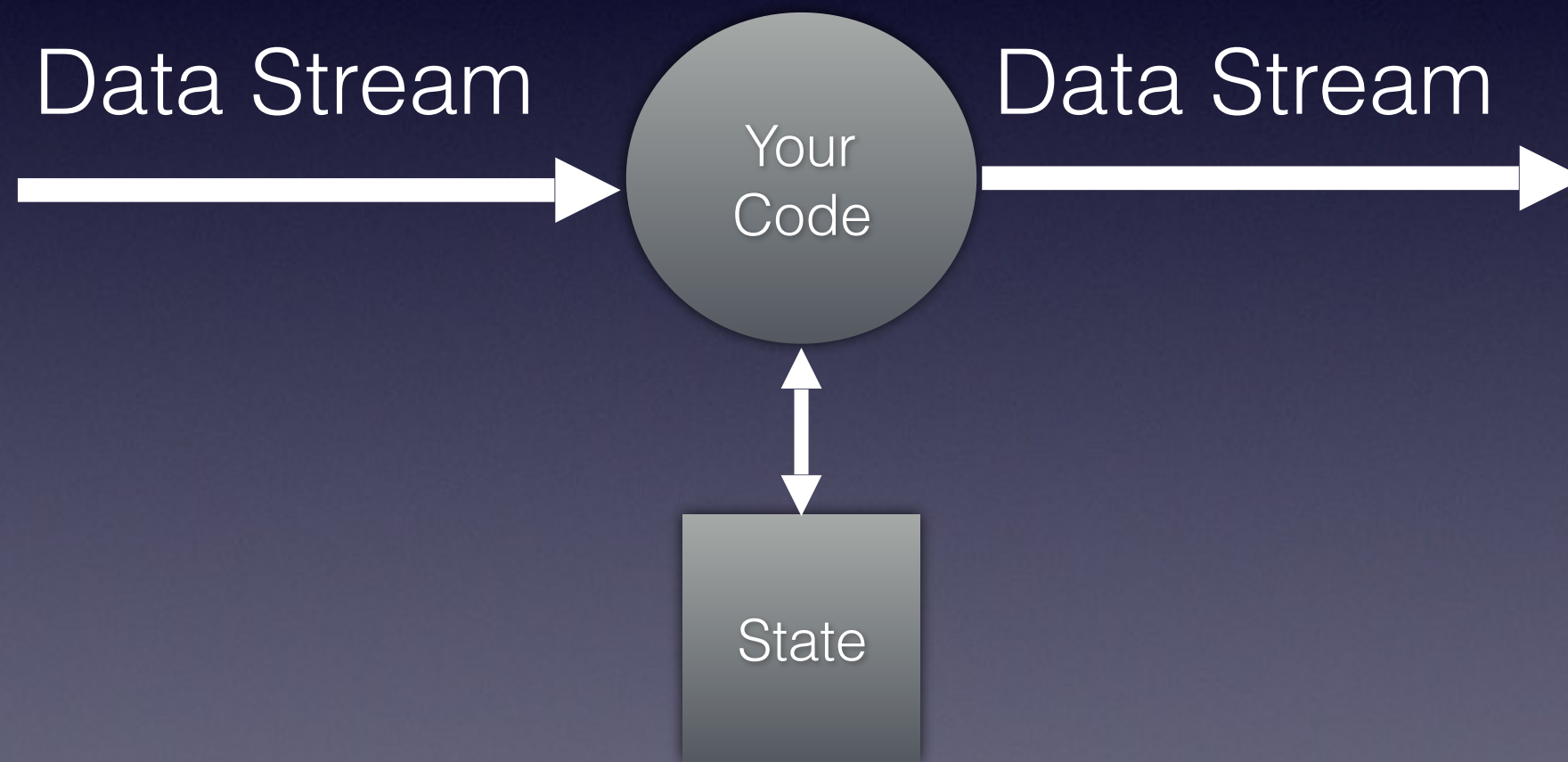


Stream Processing



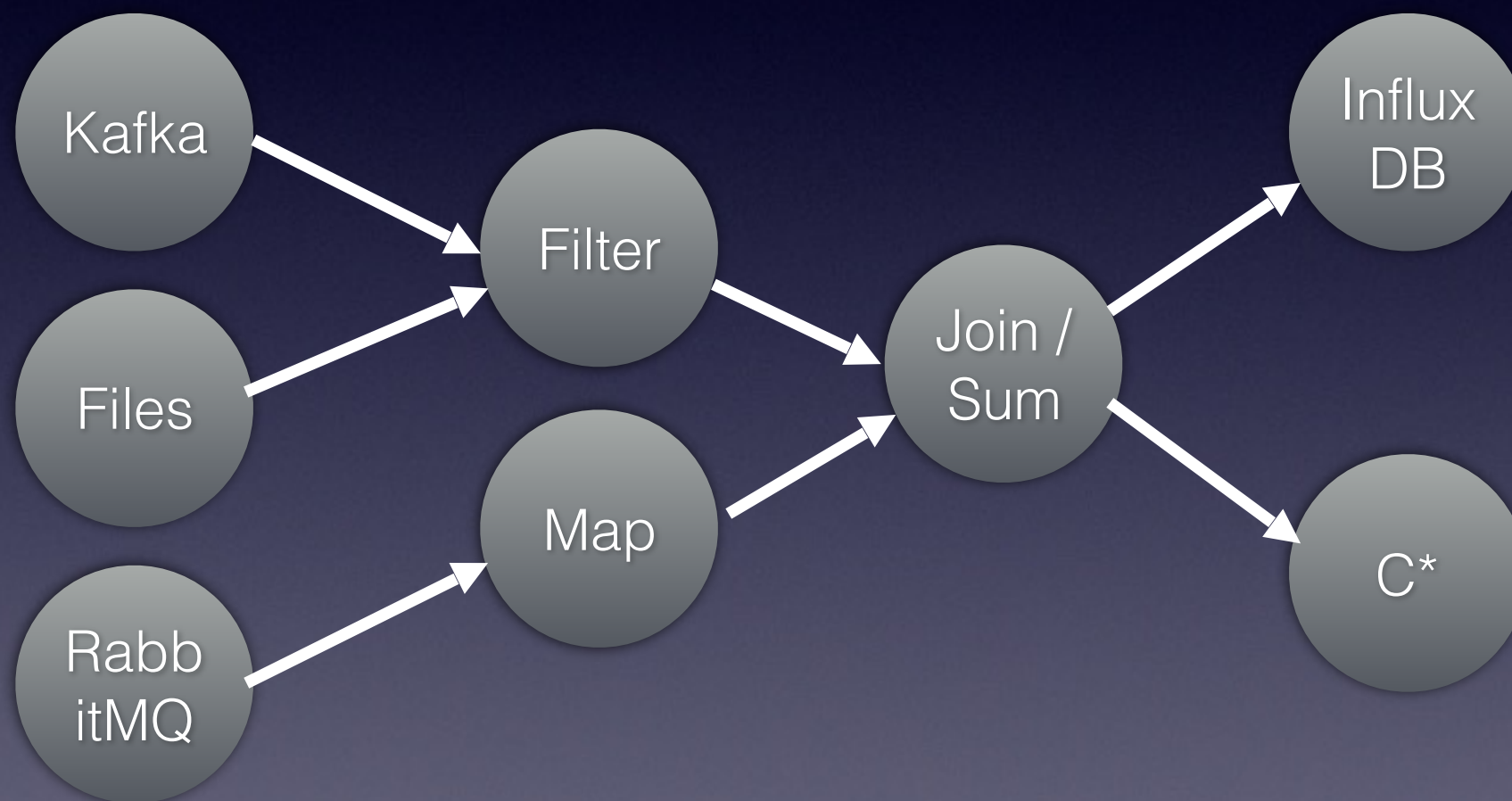


Stateful Stream Processing



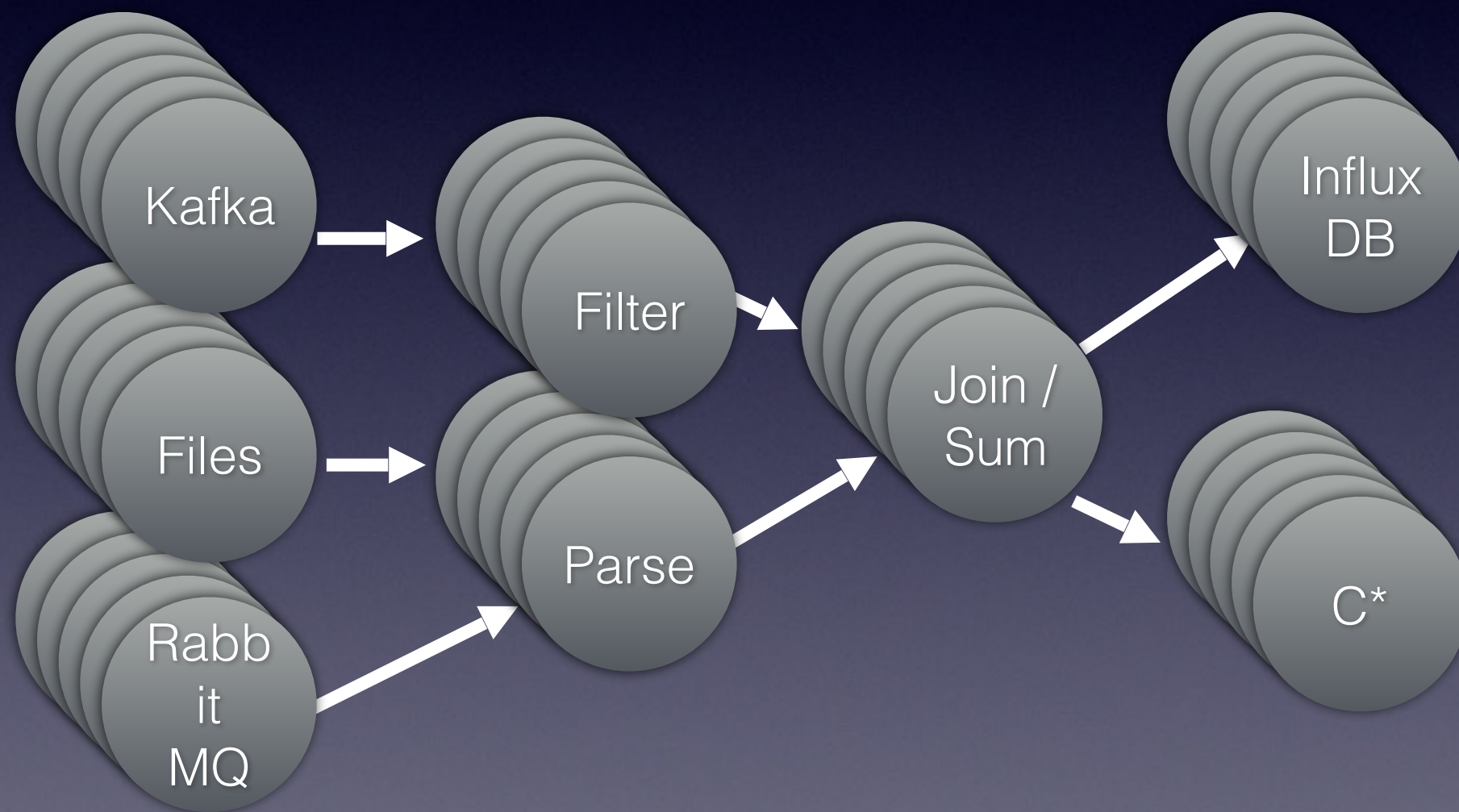


More Complex Example





Distributed and Parallel Deployment





Robust Stream Processing with Apache Flink



Code Example!



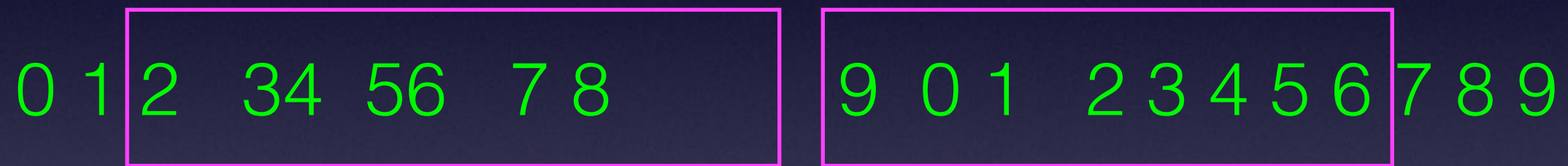
Windowing



Processing Time vs Event Time



Windowing in Processing Time



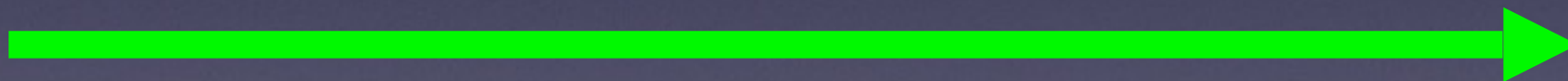
Event Time



Processing Time



Windowing in Event Time



Event Time



Processing Time = Errors!



Event Time = Accuracy



Failure Handling



Downtime Handling



Data Reprocessing



dataArtisans

We're Hiring!

<http://data-artisans.com/careers>

12-14 SEP 2016

BERLIN

CALL FOR SUBMISSIONS

PALAIS

Flink Forward 2016, Berlin

Submission deadline: June 30, 2016

Early bird deadline: July 15, 2016

www.flink-forward.org



Questions?



Thanks!