



Introduction to Stream Processing with Apache Flink®

Kostas Kloudas

Vasia Kalavri

Jonas Traub

dataArtisans

Who are we?



- Kostas: software engineer @ data Artisans
- Vasia: PhD student @ KTH Stockholm
- Jonas: research associate @ TU Berlin

Overview



- What is Stream Processing?
- What is Apache Flink?
- Windowed computations over streams
- Handling time
- Handling node failures
- Handling planned downtime
- Handling code upgrades

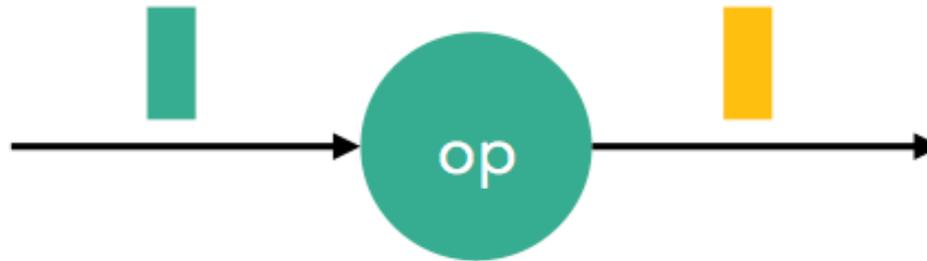
Demo instructions...



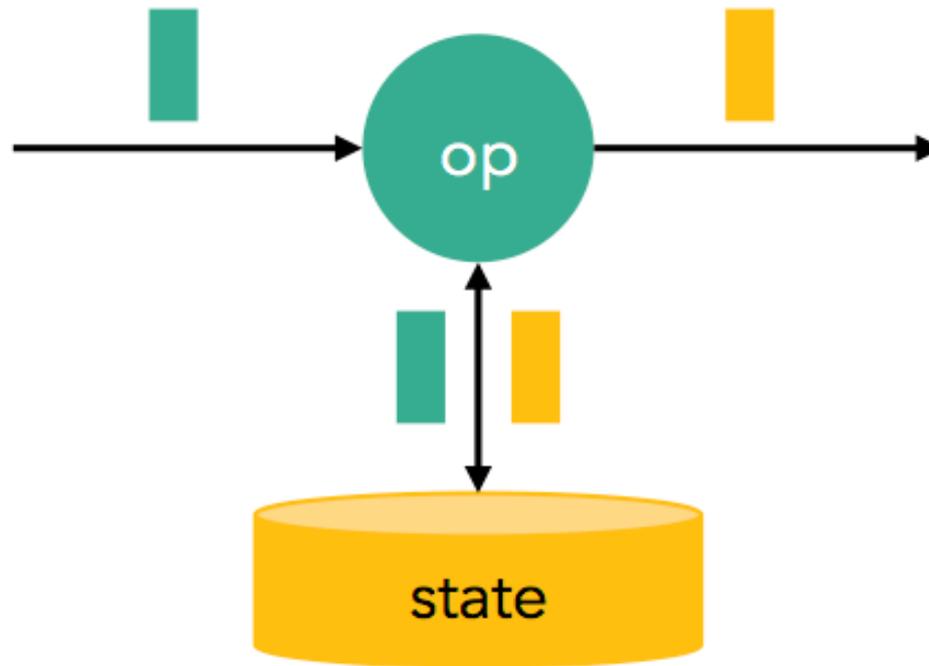
Robust Stream Processing with Apache Flink®: A Simple Walkthrough
<http://data-artisans.com/robust-stream-processing-flink-walkthrough/#more-1181>

Make sure you download: Apache Flink 1.0.3

Stateless stream processing



Stateful stream processing



Why should you care?



Data production is and has always been a continuous process.

Stream processing enables the obvious:
Continuous processing on data that is
continuously produced

What is Apache Flink?



A data processing engine

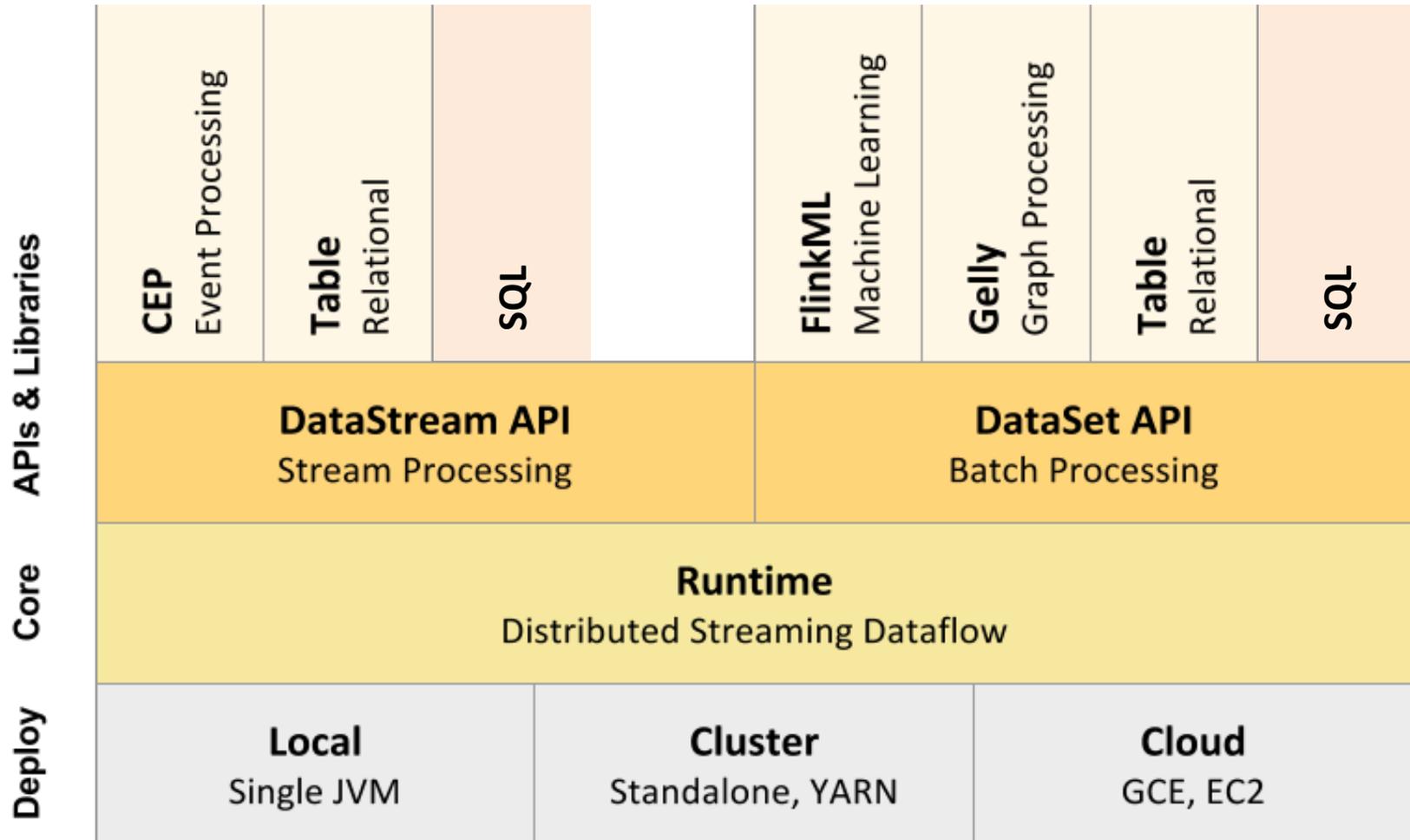


Apache Flink is an open source platform for distributed stream and batch processing



Apache Flink

The Apache Flink Ecosystem

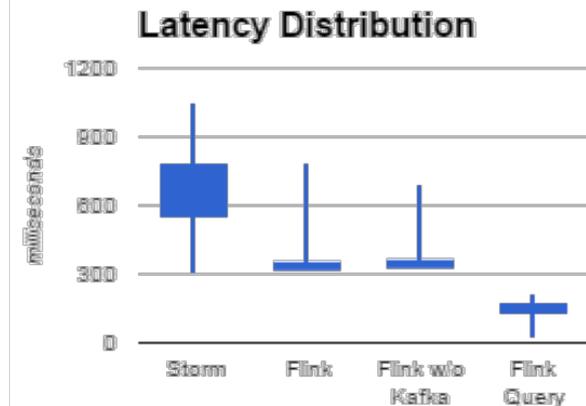
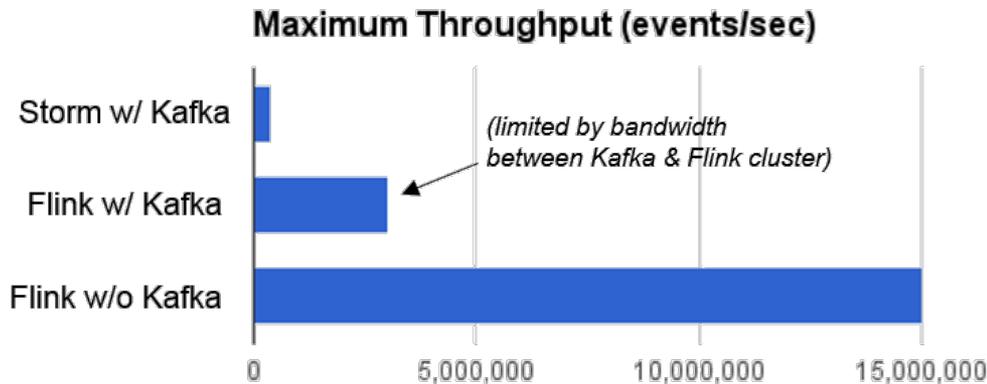


What does Flink provide?



■ High Throughput and Low Latency

- Yahoo! Benchmark : <https://yahooeng.tumblr.com/post/135321837876/benchmarking-streaming-computation-engines-at>
- Extended by Data Artisans: <http://data-artisans.com/extending-the-yahoo-streaming-benchmark/>



What does Flink provide?



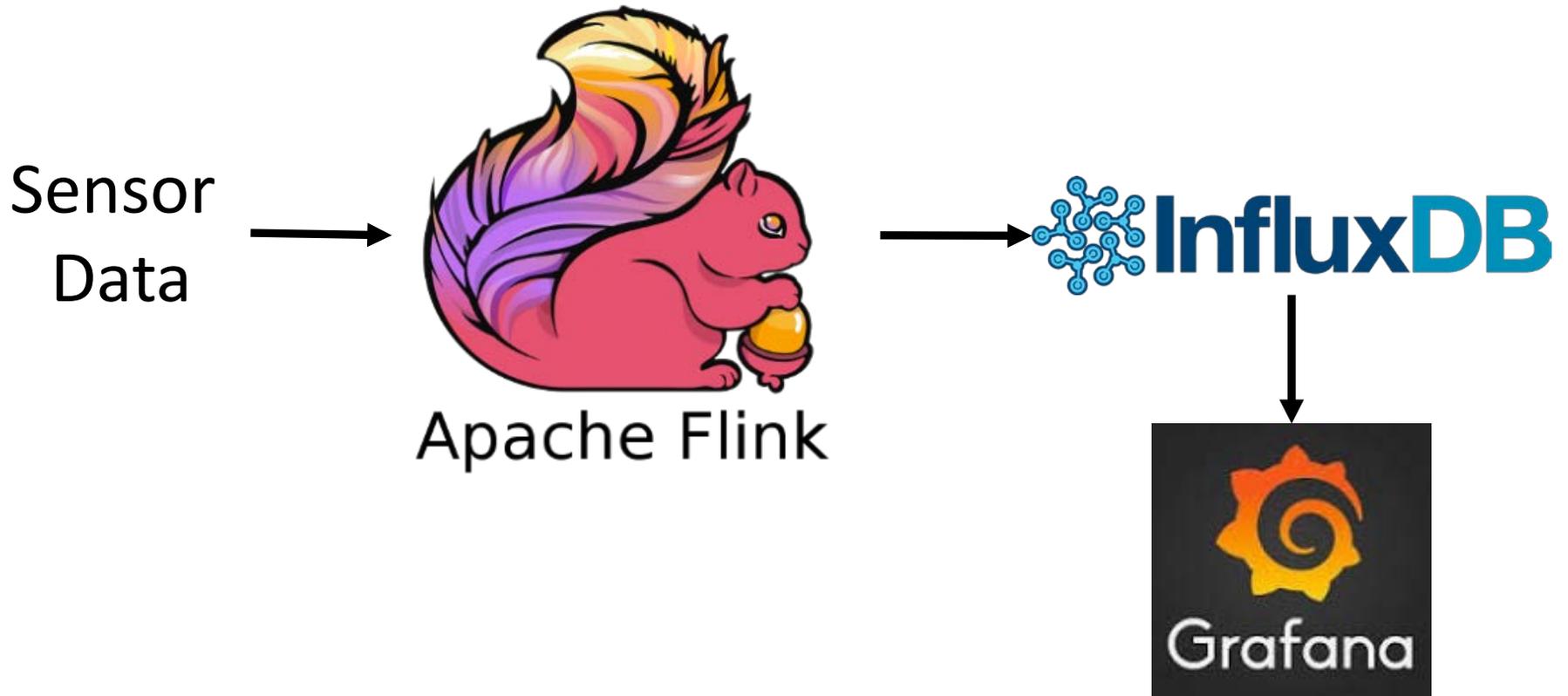
- High Throughput and Low Latency
- Event-time (out-of-order) processing
- Exactly-once semantics
- Flexible windowing
- Fault-Tolerance

Time for demo...



Robust Stream Processing with Apache Flink®: A Simple Walkthrough
<http://data-artisans.com/robust-stream-processing-flink-walkthrough/#more-1181>

Setup:



Windowed computations



Handling time



Handling time

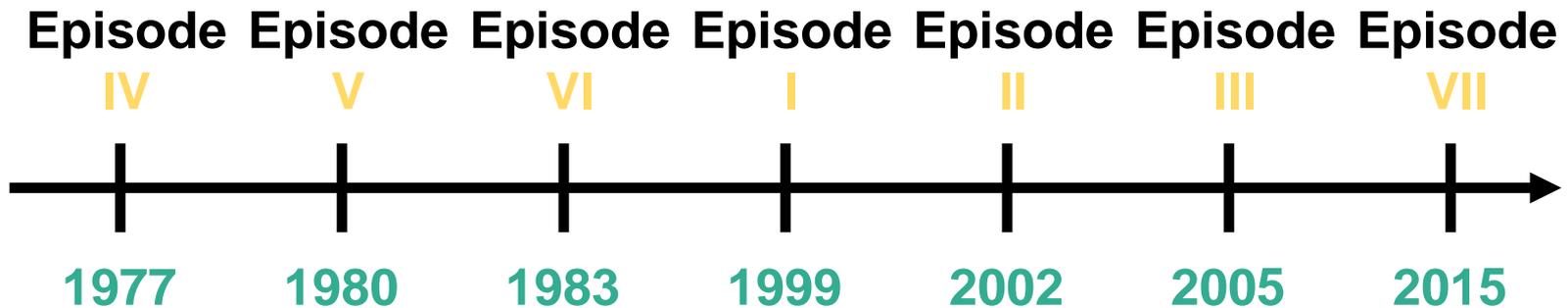


The system has to respect the same clock
as the data.

Event Time vs Processing Time



Event Time



Processing Time

Handling time: Watermarks

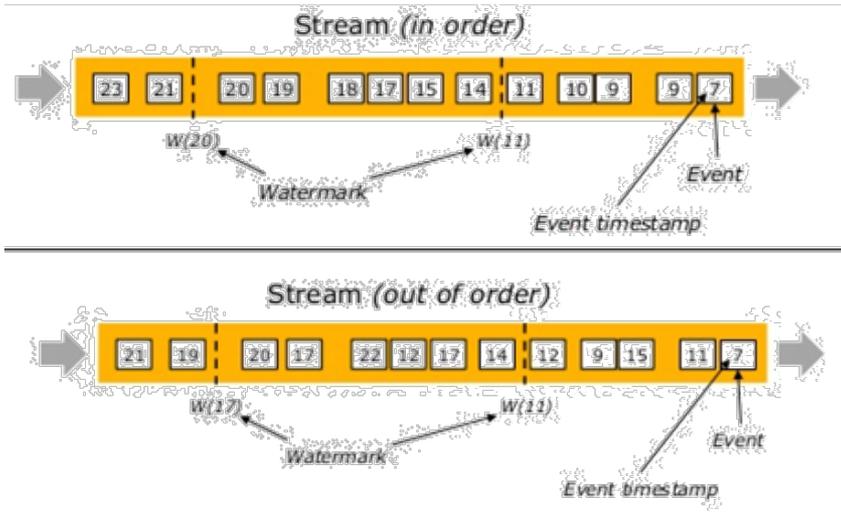


- Special events generated by the sources.
- A watermark for time T states that event time has progressed to T in that particular stream (or partition).
- No events with a timestamp smaller than T can arrive any more.

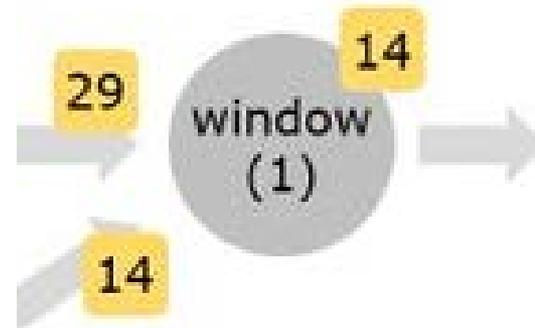
Handling time: Watermarks



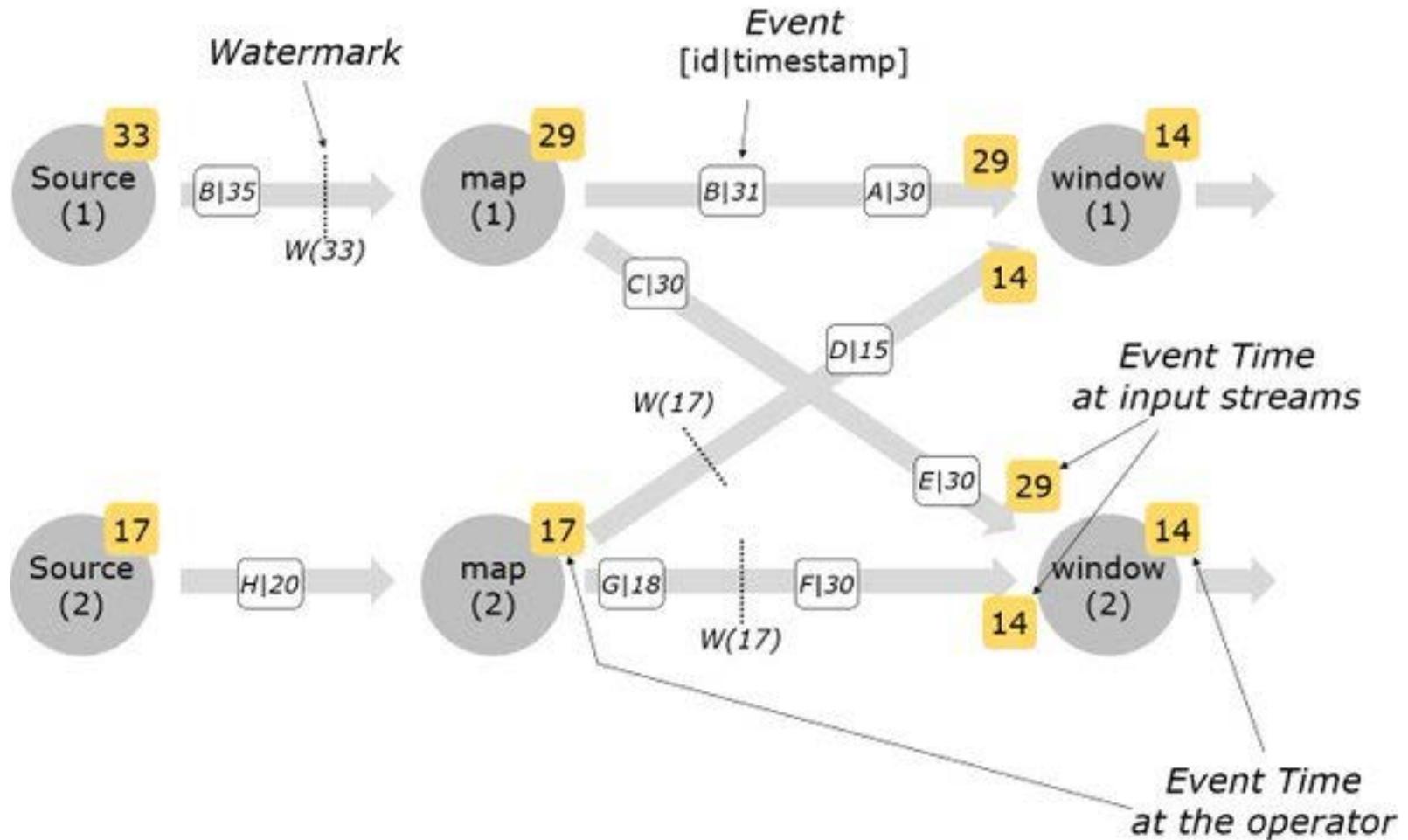
Sources emit elements and watermarks....



...operators always emit the lowest watermark



Handling time: Watermarks



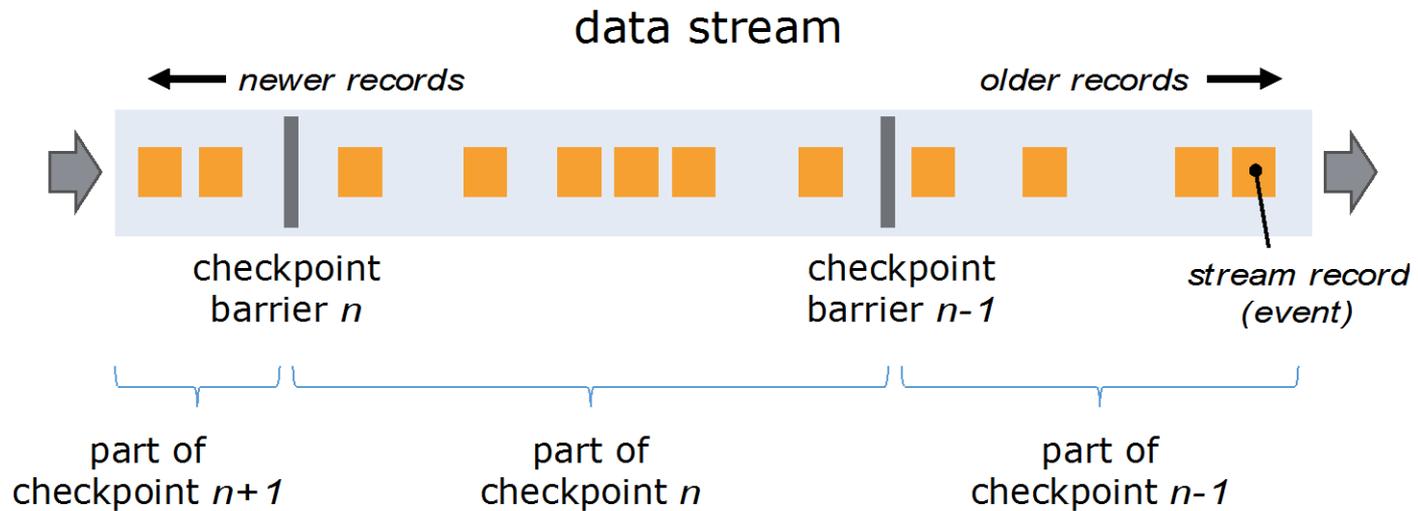
Handling node failures



Checkpoints



Sources emit elements and checkpoints...



Checkpoints



Handling planned downtime



Handling code upgrades



Is Apache Flink only that?



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



Apache Flink

Its lively community

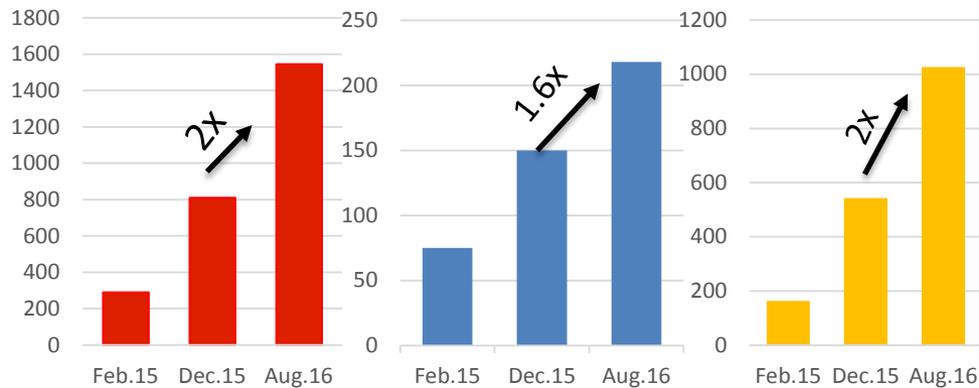


Apache Flink Community Growth

Stars on Github

Contributors

Forks on Github



■ You can join:

- **Follow:** @ApacheFlink, @dataArtisans
- **Read:** flink.apache.org/blog, data-artisans.com/blog
- **Subscribe:** (news | user | dev) @ flink.apache.org

Its Users



ERICSSON



[...https://flink.apache.org/poweredby.html](https://flink.apache.org/poweredby.html)

All of them will meet at...



<http://flink-forward.org/>

FlinkForward

REGISTRATION

PROGRAM ▾

NEWS

ABOUT ▾

12-14 SEP 2016

BERLIN



Flink Forward is the premier conference on Apache Flink®

REGISTER NOW

PALAIS >

All of them will meet at...



<http://flink-forward.org/>



Further Reading



■ Event-time processing:

- The Dataflow Model: <http://www.vldb.org/pvldb/vol8/p1792-Akidau.pdf>
- <http://data-artisans.com/how-apache-flink-enables-new-streaming-applications-part-1/>

■ Checkpointing and State:

- Distributed Snapshots: Determining Global States of Distributed Systems
<http://research.microsoft.com/en-us/um/people/lamport/pubs/chandy.pdf>
- Lightweight Asynchronous Snapshots for Distributed Dataflows
<https://arxiv.org/abs/1506.08603>
- Working with State in Flink: <https://ci.apache.org/projects/flink/flink-docs-master/dev/state.html>

■ Savepoints:

- <https://ci.apache.org/projects/flink/flink-docs-master/setup/savepoints.html>