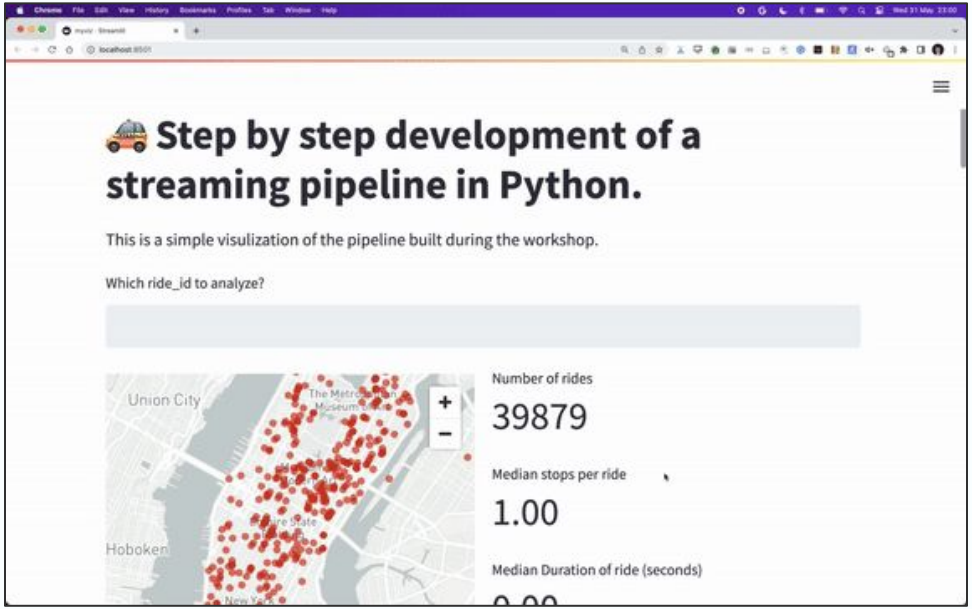

BEAM
SUMMIT

Workshop:
Step by step development of a streaming pipeline in python

Anthony Lazzaro
Strategic Cloud Engineer, Google Cloud

Israel Herraiz
Strategic Cloud Engineer, Google Cloud

A teaser!

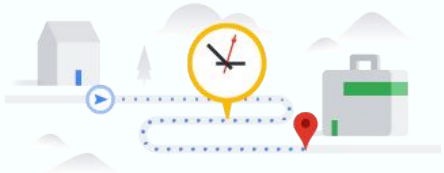


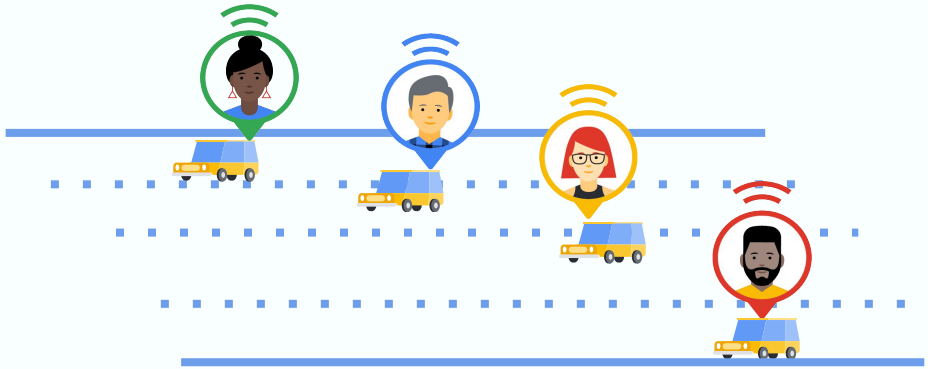
github.com/ant-laz/streamingworkshop

In this workshop we will use the NYC Taxi Tycoon Dataset

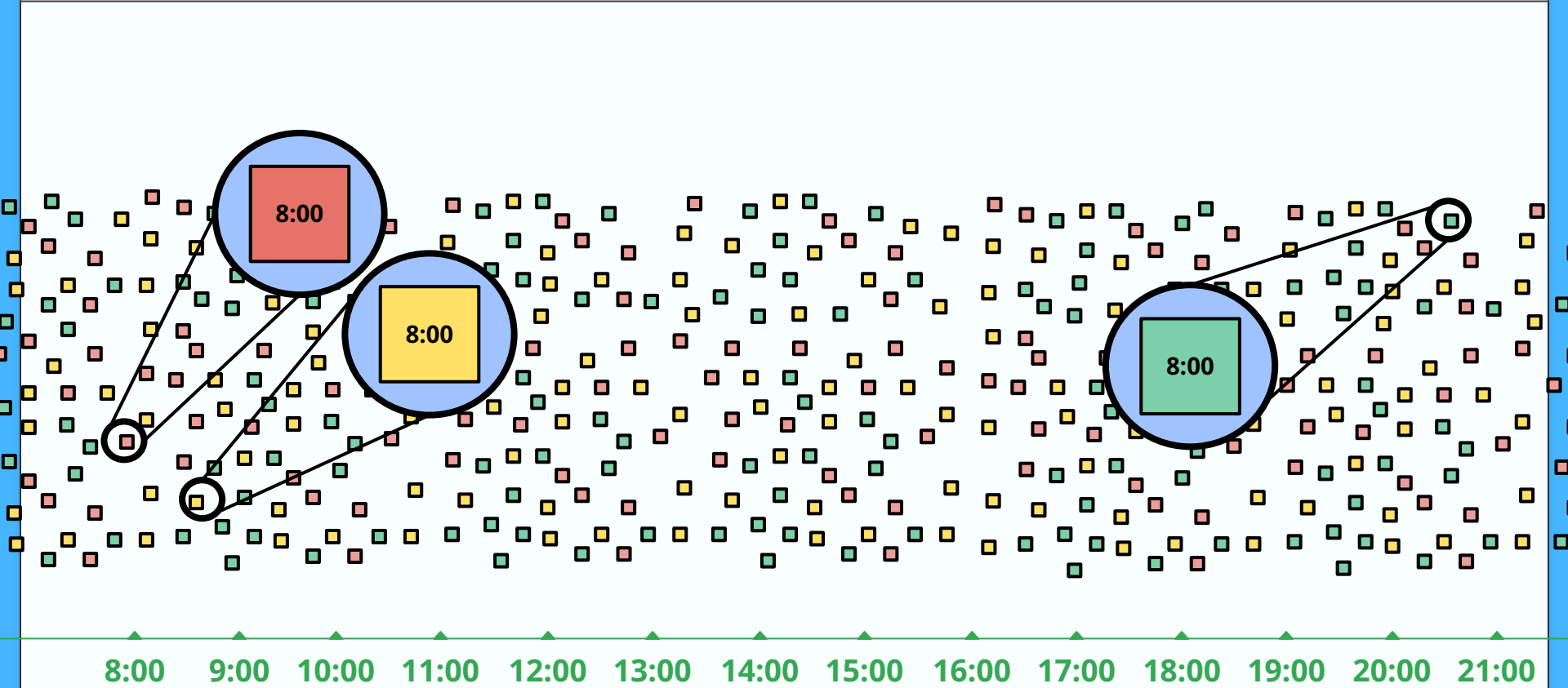
What do the events making up a ride look like in order ?

How long are my taxis drivers taking to complete rides ?

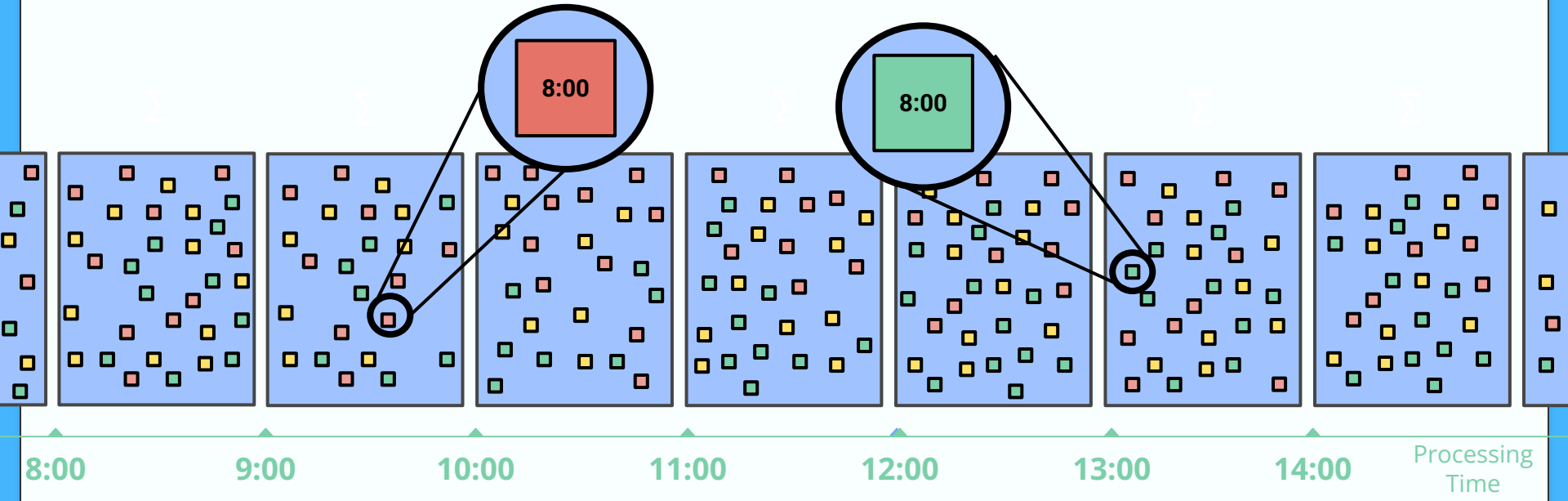




Arrival out of order – The problem with streaming



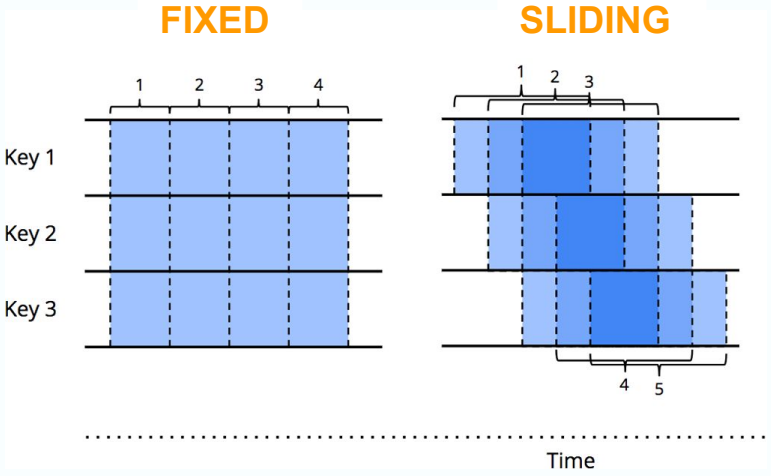
A problem that cannot be solved with micro-batching

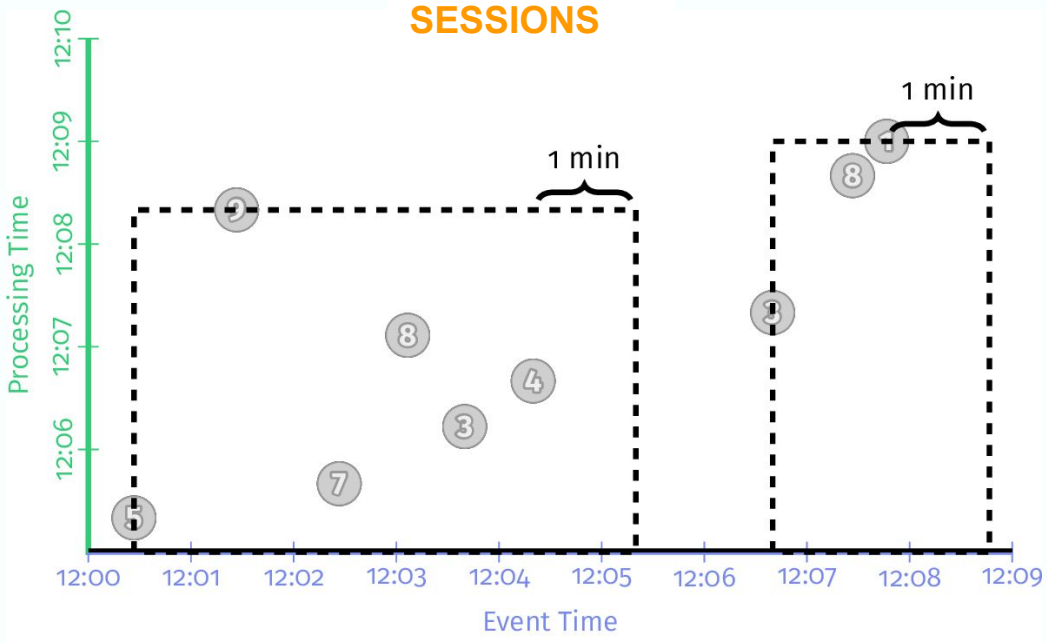


Event time vs processing time

Source: *Introduction to Apache Flink by Ellen Friedman, Kostas Tzoumas*

4. **WHERE** in event time to calculate ? → **Windowing**





source: "Streaming Systems" by Tyler Akidau, Slava Chernyak & Reuven Lax

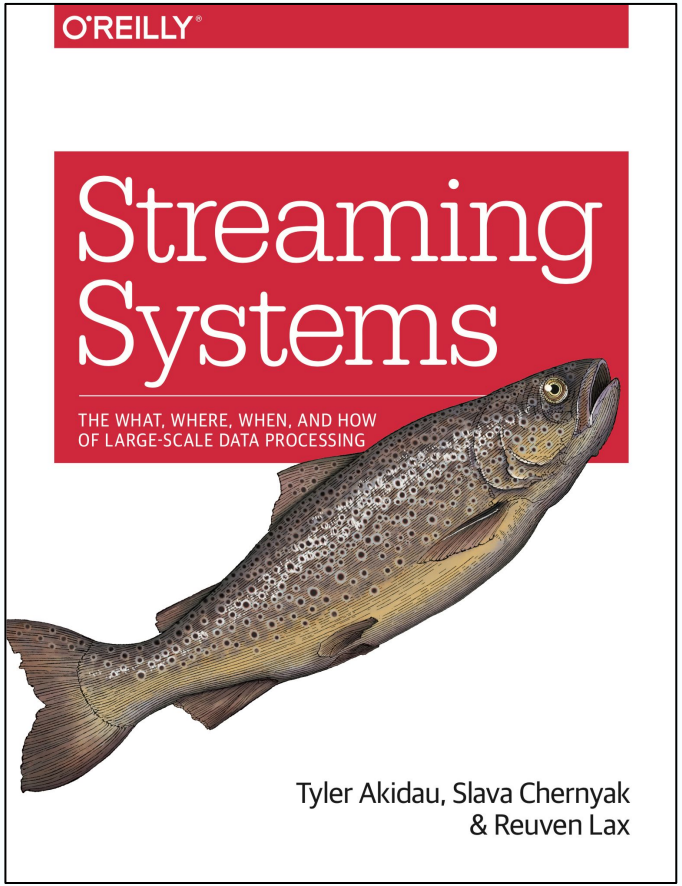
6 things you need to know to get starting with Beam & streaming !

1. **DATA** in Beam → **PCollection**
2. **WHAT** results are calculated ? → **PTransforms**
3. **TIME** in Beam → **Event vs Processing Time**
4. **WHERE** in event time to calculate ? → **Windowing**
5. **WHEN** in processing time emit results? → **Triggers + Watermarks**
6. **HOW** do refinements of results relate ? → **Accumulation mode**

source: "Streaming Systems" by Tyler Akidau, Slava Chernyak & Reuven Lax

The book on Streaming Systems

Tyler Akidau, Slava Chernyak, Reuven Lax

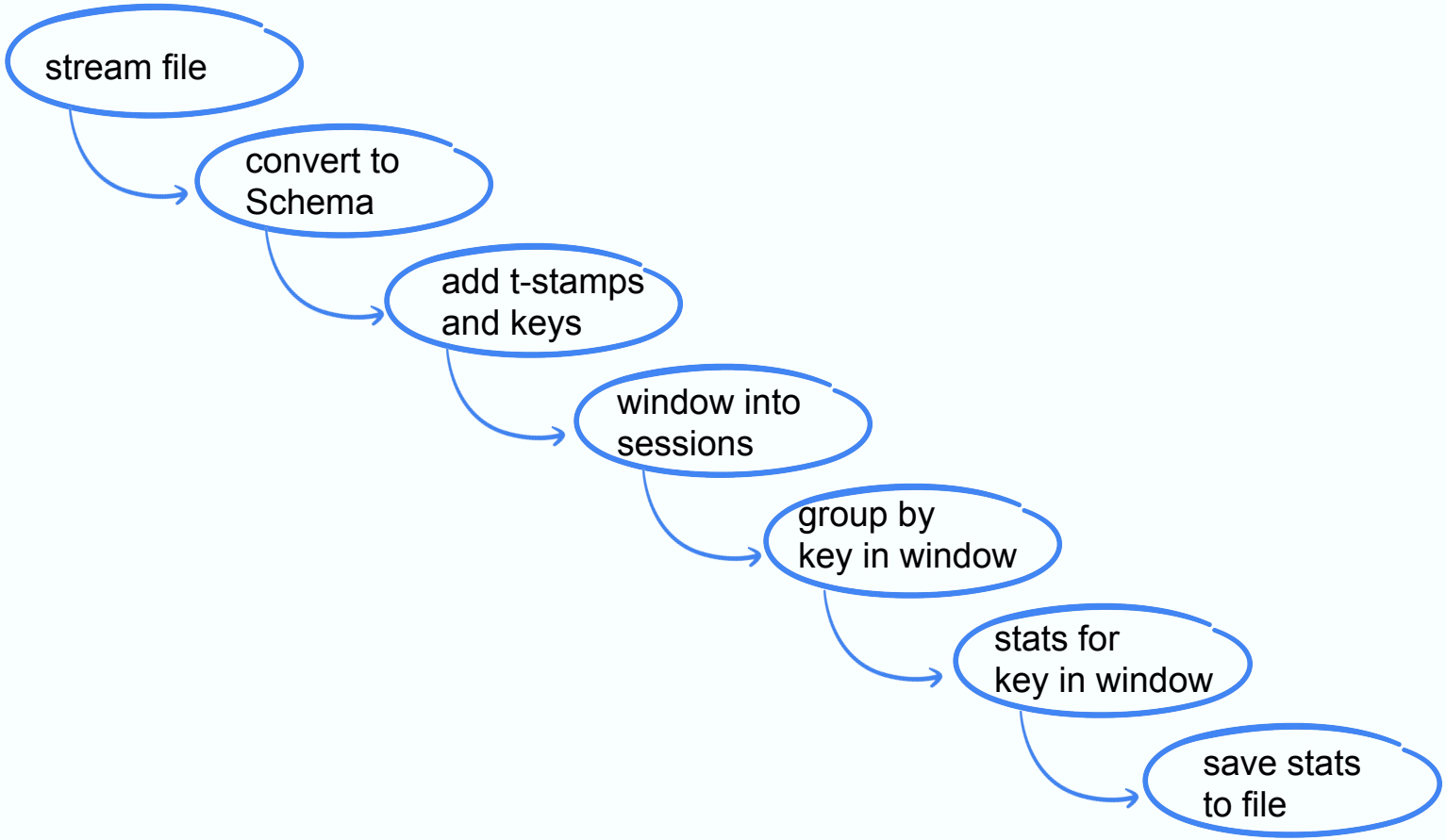


"If you care about the correctness of your streaming and batch processing jobs, this book is a must-read. It provides the most clear-thinking and logical discussion of the topic that I have seen, and its ideas are brilliantly explained."

—Martin Kleppmann
University of Cambridge

http://streamingsystems.net/

Overview of the pipeline we are going to build



Let's get coding !

1. Clone this repository →
2. Follow the README
3. Two branches
    a. main
    b. workshop14june2023
    c. challenge yourself !

https://github.com/ant-laz/streamingworkshop

Israel Herraiz & Anthony Lazzaro

Thank you