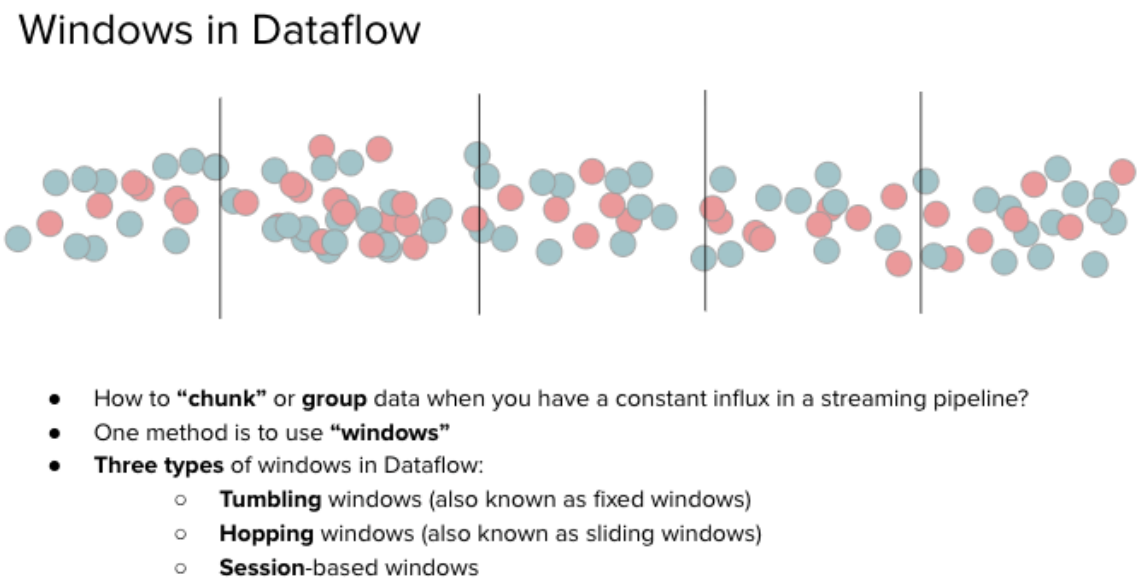
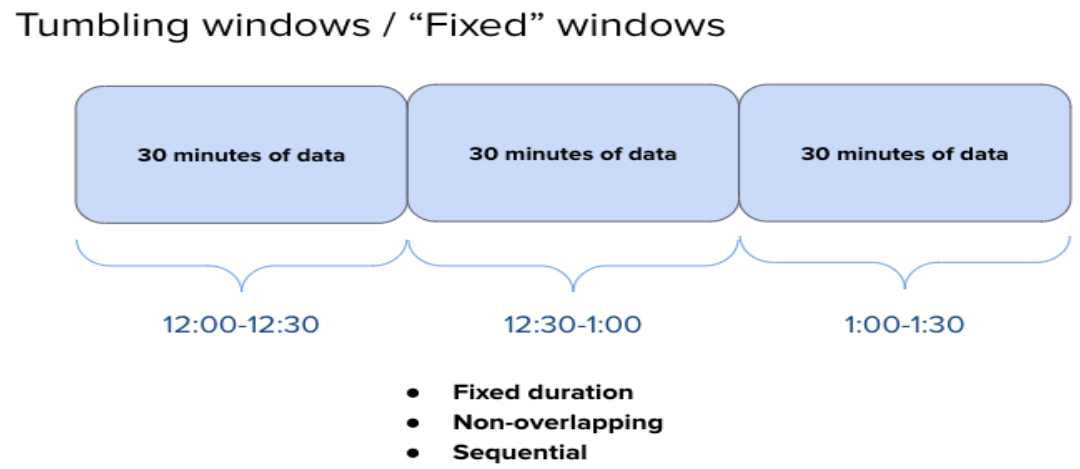
Dataflow: Windowing



Tumbling (Fixed) Windows

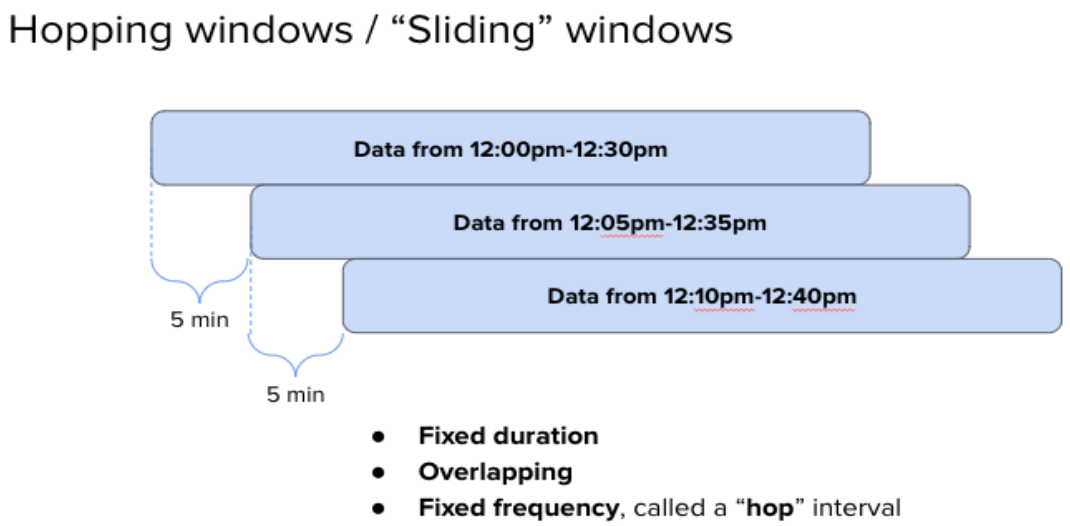


- **non-overlapping** intervals of time. Each window is the same size, and one window ends exactly where the next begins.
- For example, imagine we're working with a 30-minute tumbling window. The first window would start at 12:00 and end at 12:30. The next would go from 12:30 to 1:00, and so on. Any

data that arrives during one of those windows is grouped together and processed as a unit.

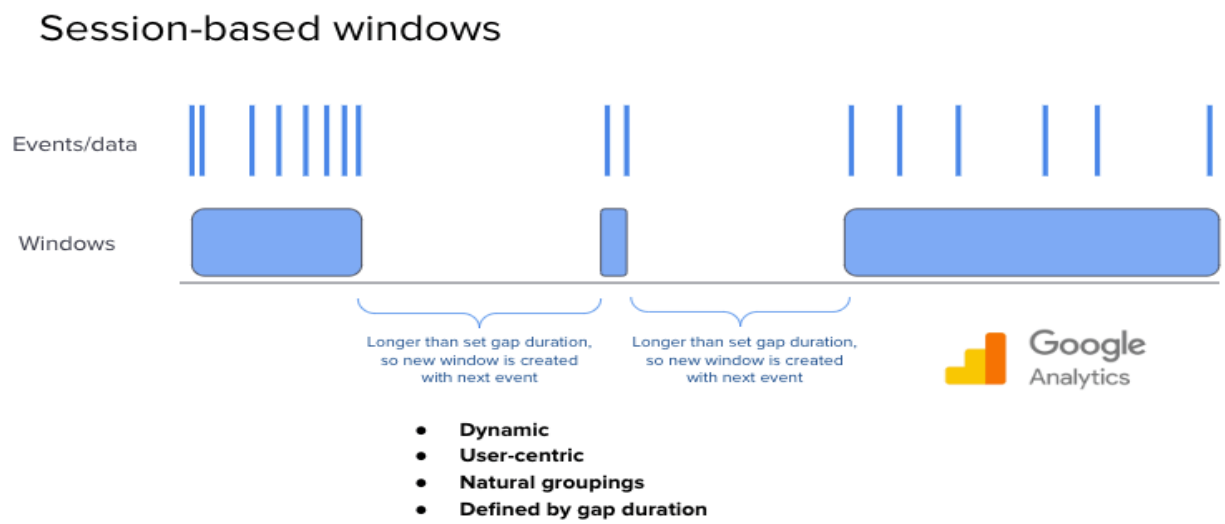
- This type of windowing is useful when you're calculating statistics like averages, counts, or totals over regular time intervals.
- It's also common in data visualization - think of a dashboard graph that shows activity by the hour.

Hopping (Sliding) Windows



- Hopping windows are similar to tumbling windows in that they have a fixed duration, but they also overlap. A new window starts at a regular interval (known as the "hop"), and it spans a fixed period of time.
- For example, let's say we still want 30-minute windows, but we want to create a new window every 5 minutes. The first window might go from 12:00 to 12:30, the next from 12:05 to 12:35, and so on.
- This setup is helpful when you want frequent updates over a longer span of data.
- Imagine calculating the average stock price every minute, but always using the last 10 or 20 minutes of data to do it. That's a great use case for hopping windows.
- In summary: hopping windows have a **fixed size, they overlap, and they're generated at regular intervals**, giving you a high-resolution view of recent trends.

Session-Based Windows



- Session-based windows are different. Instead of relying on fixed time intervals, session windows are built around gaps in the data.
- Let's say events are coming in from users, and sometimes there's a pause in activity. If **that pause - called a gap duration** - is longer than a certain threshold (say, five minutes), Dataflow considers the session to be over. The next event starts a new session.
- These windows are dynamic. They can be short or long, depending entirely on when the data arrives.
- For example, if a user is sending events every four minutes, those events will be grouped into the same session window. But if there's a six-minute break, the next event starts a new session.
- If you've used tools like **Google Analytics**, this may sound familiar.
- Google Analytics defines a session as ending after 30 minutes of inactivity. It's the same principle - **events are grouped** based on natural user behavior, not the clock.
- Session windows are great for user-centric data and behavior analysis, where the data doesn't follow a regular schedule.