



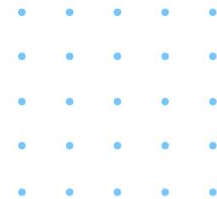
Xử lý dữ liệu Streaming

Streaming Processing



Event trong Streaming

Event in Streaming



Đăng nhập

Xem

Thêm vào giỏ
hàng

Mua

@9:00

@9:01

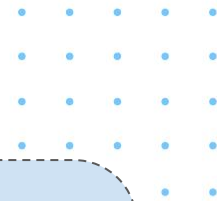
@9:10

@9:20



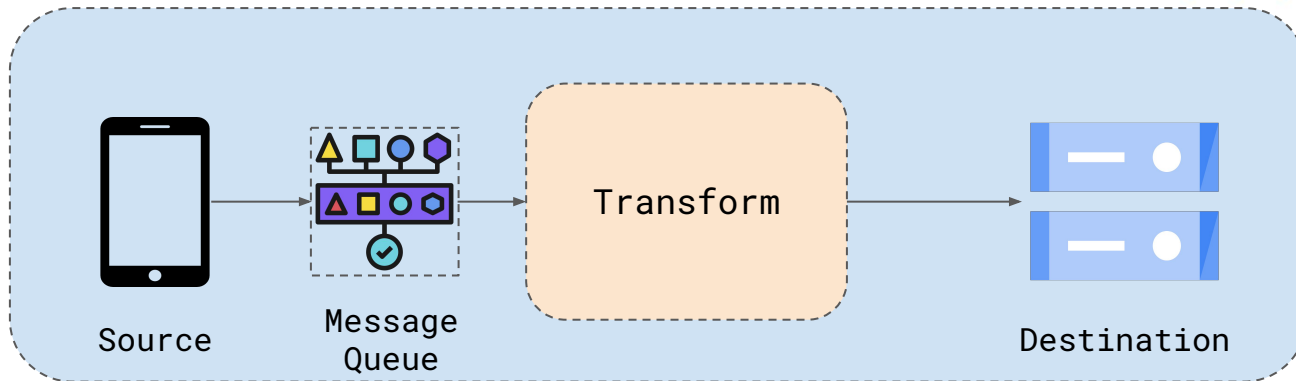
Khác biệt giữa Stream và Batch

Difference between Stream and Batch



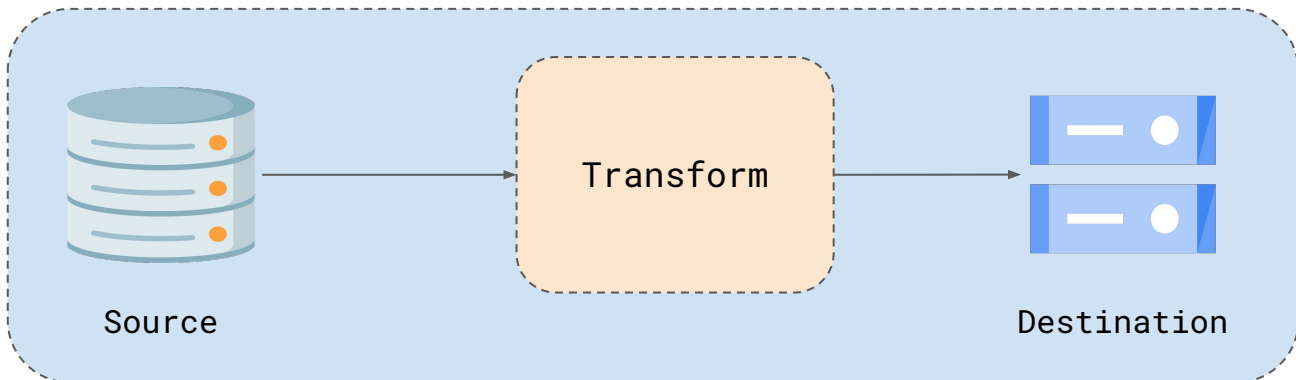
Streaming
Pipeline

Chạy liên tục



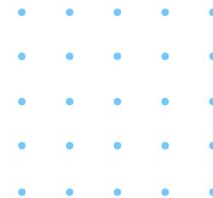
Batch
Pipeline

Chạy định kỳ



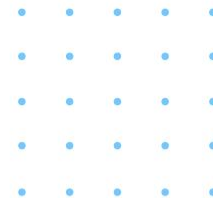
Streaming Pipeline

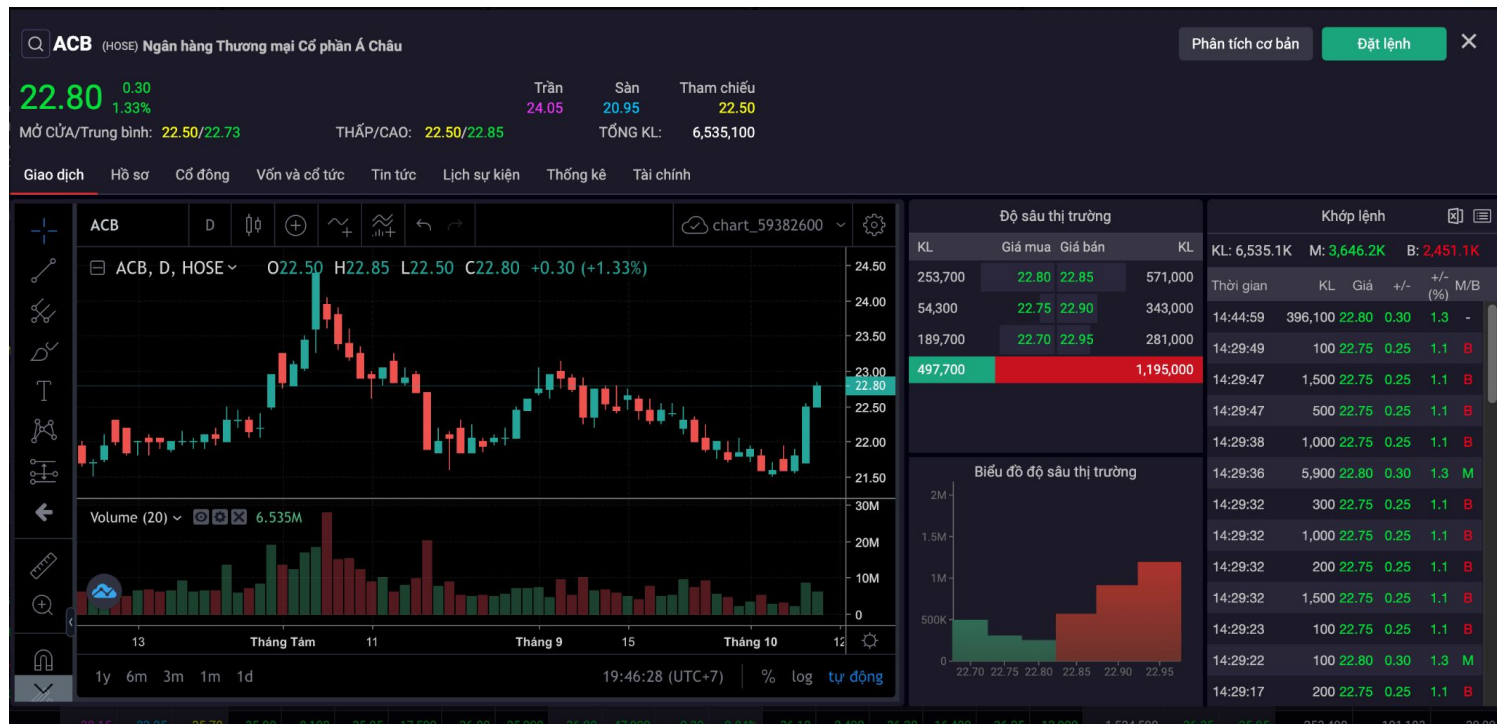
Streaming Pipeline



Streaming Pipeline

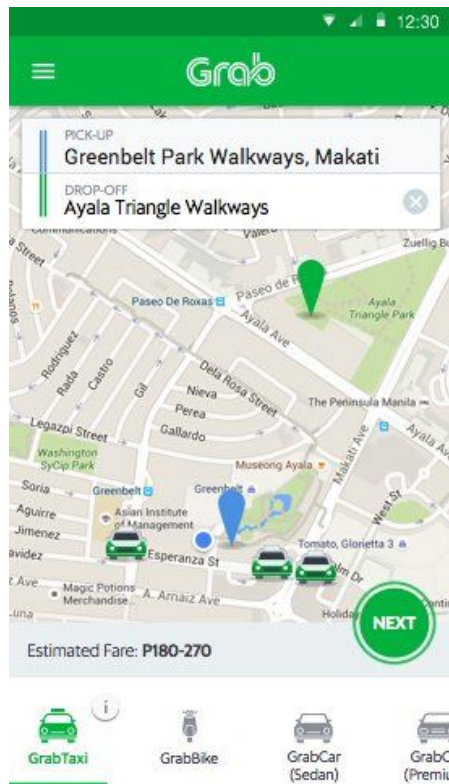
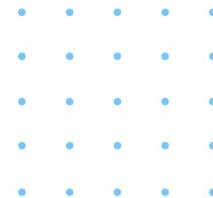
Streaming Pipeline





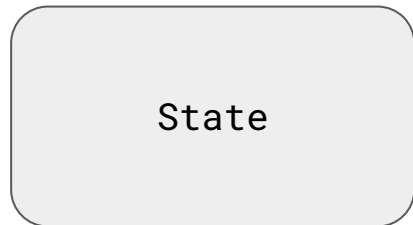
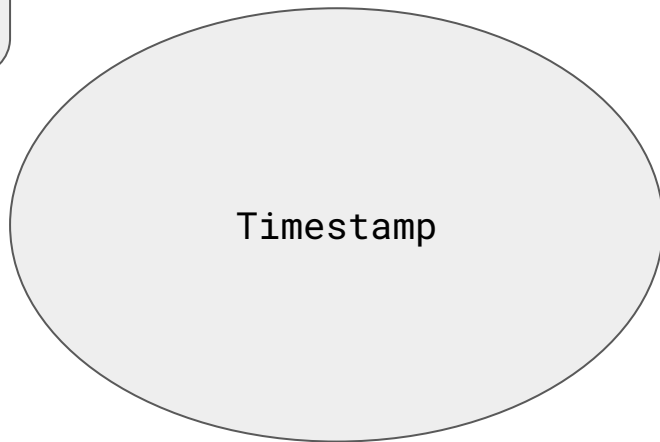
Ứng dụng

Uses Case



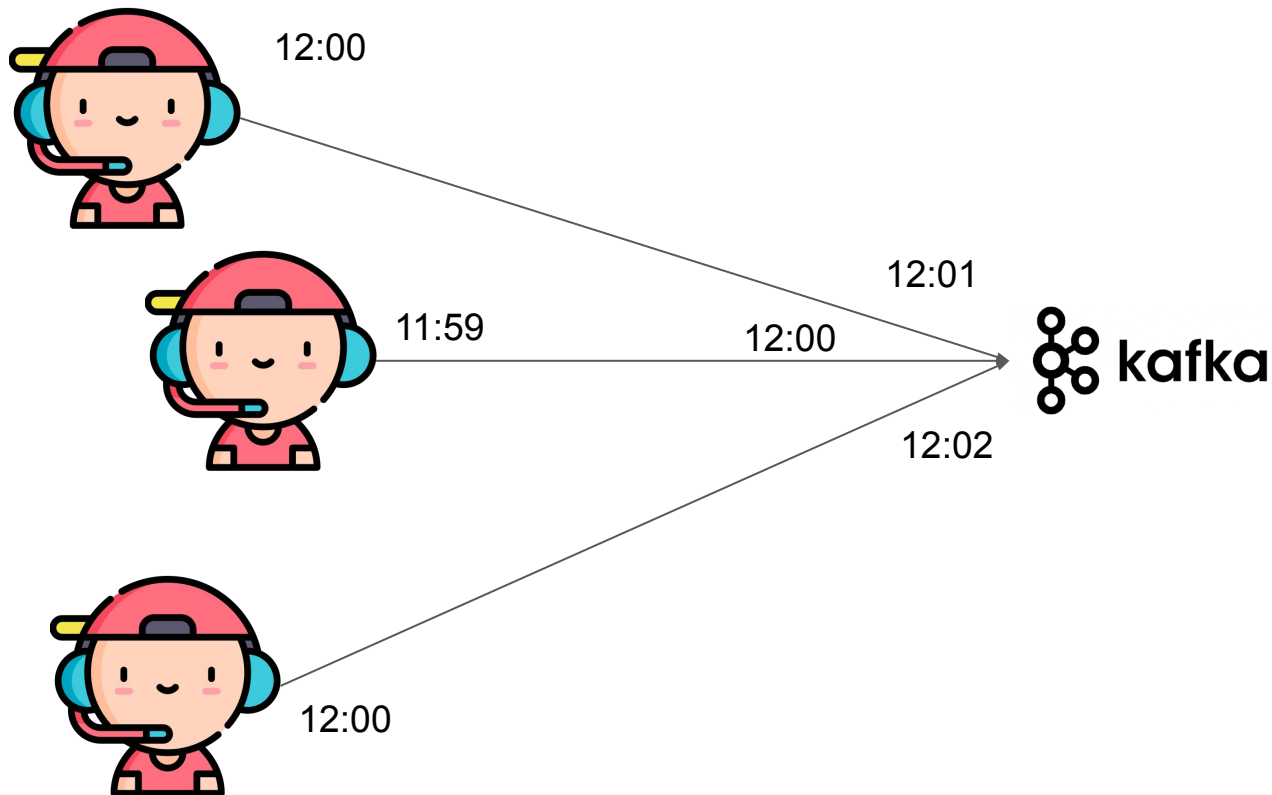
<https://www.grab.com/vn/en/>





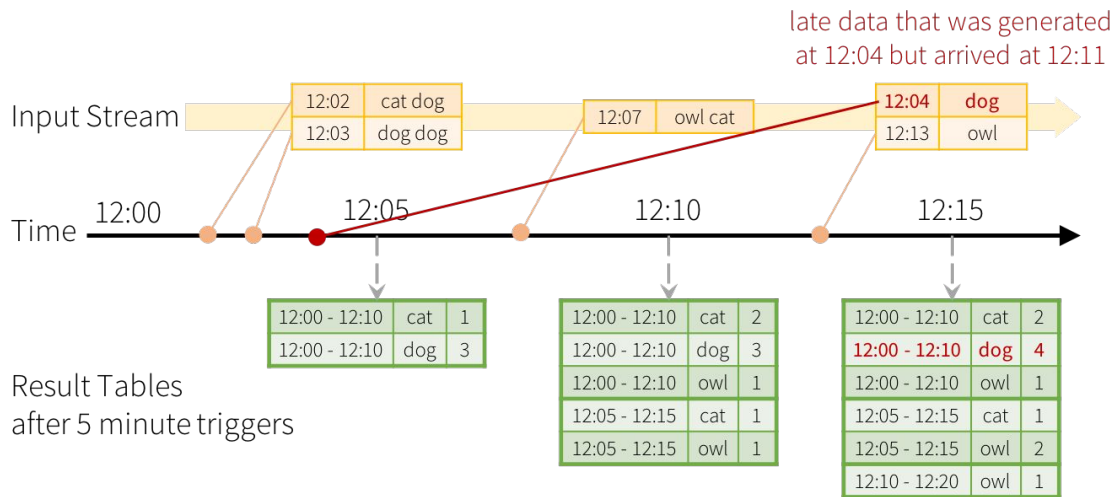
Event Time và Processing Time

Event Time and Processing Time



Late Event

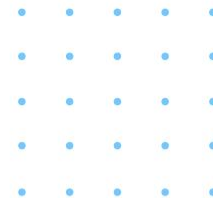
Event trễ



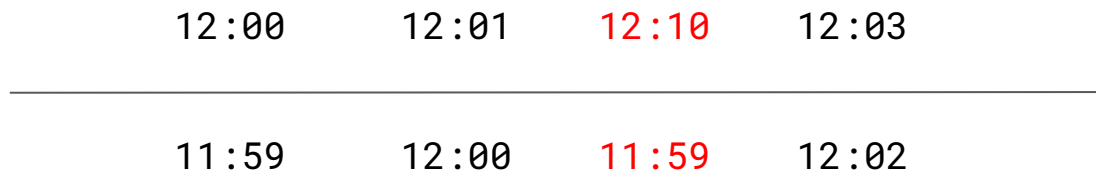
counts incremented only for window 12:00 - 12:10

Late data handling in
Windowed Grouped Aggregation

Trường hợp sự kiện xảy ra ở **12:04** nhưng **data lại đến vào lúc 12:11** nên chỉ có cập nhật đoạn thời gian **12:00 - 12:10**



Dùng 1 chuỗi time series có chứa các event time và processing time để tạo ra một mô hình phân loại event timestamp đến trễ.

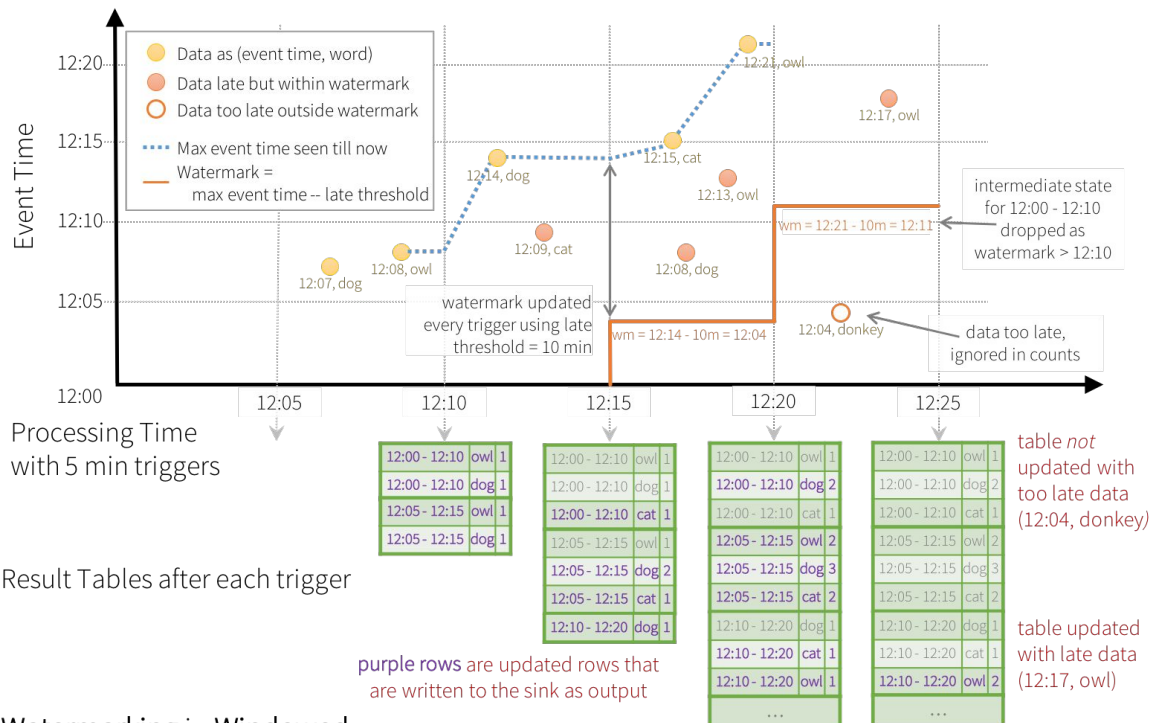


Ví dụ Spark $\text{Watermark} = \text{Latest Event Time} - \text{Latest Threshold}$



Watermark

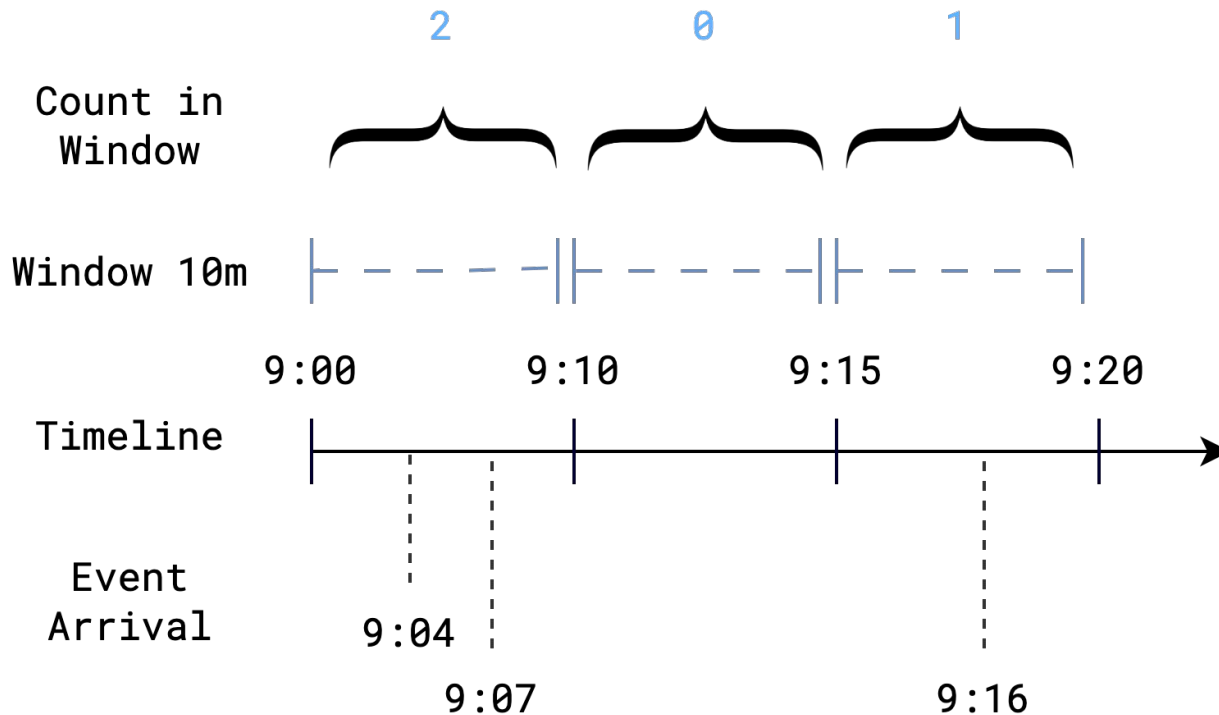
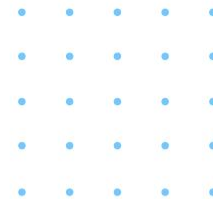
Watermark



Watermarking in Windowed
Grouped Aggregation with Update Mode

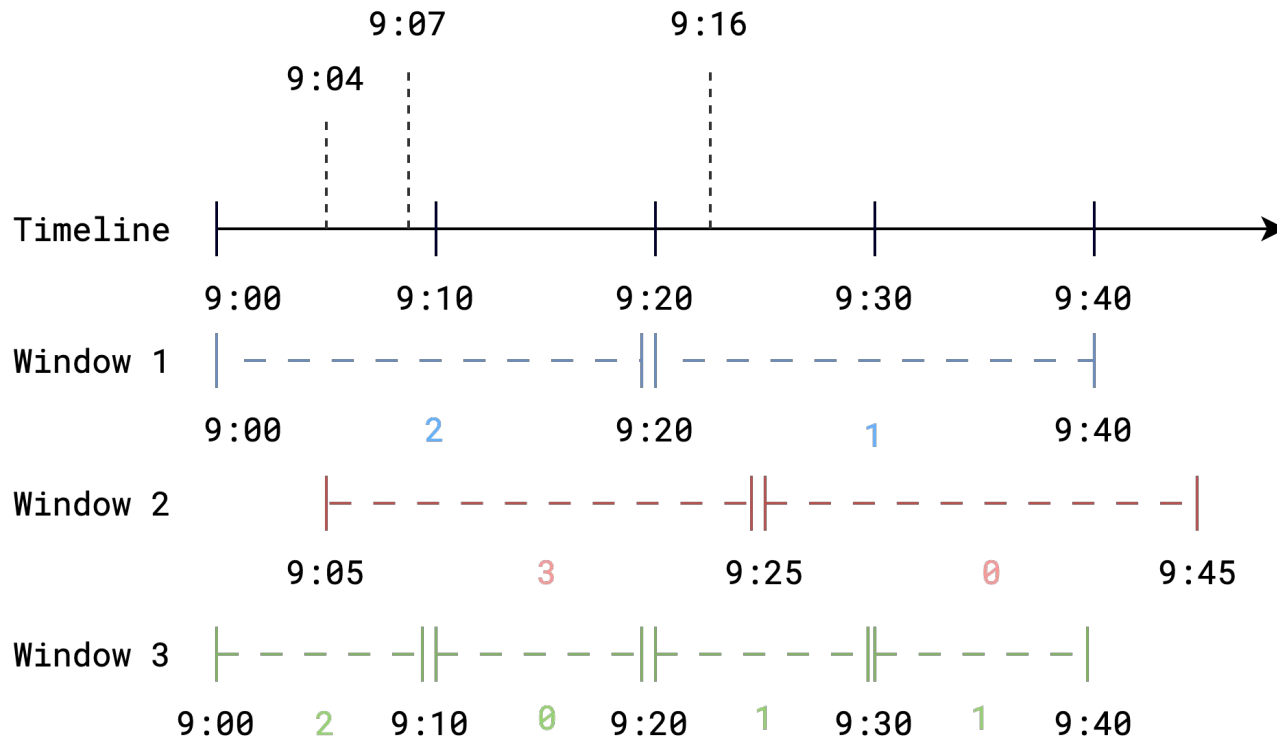
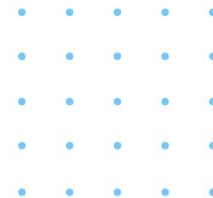
Fixed Time Window

Window cố định thời gian



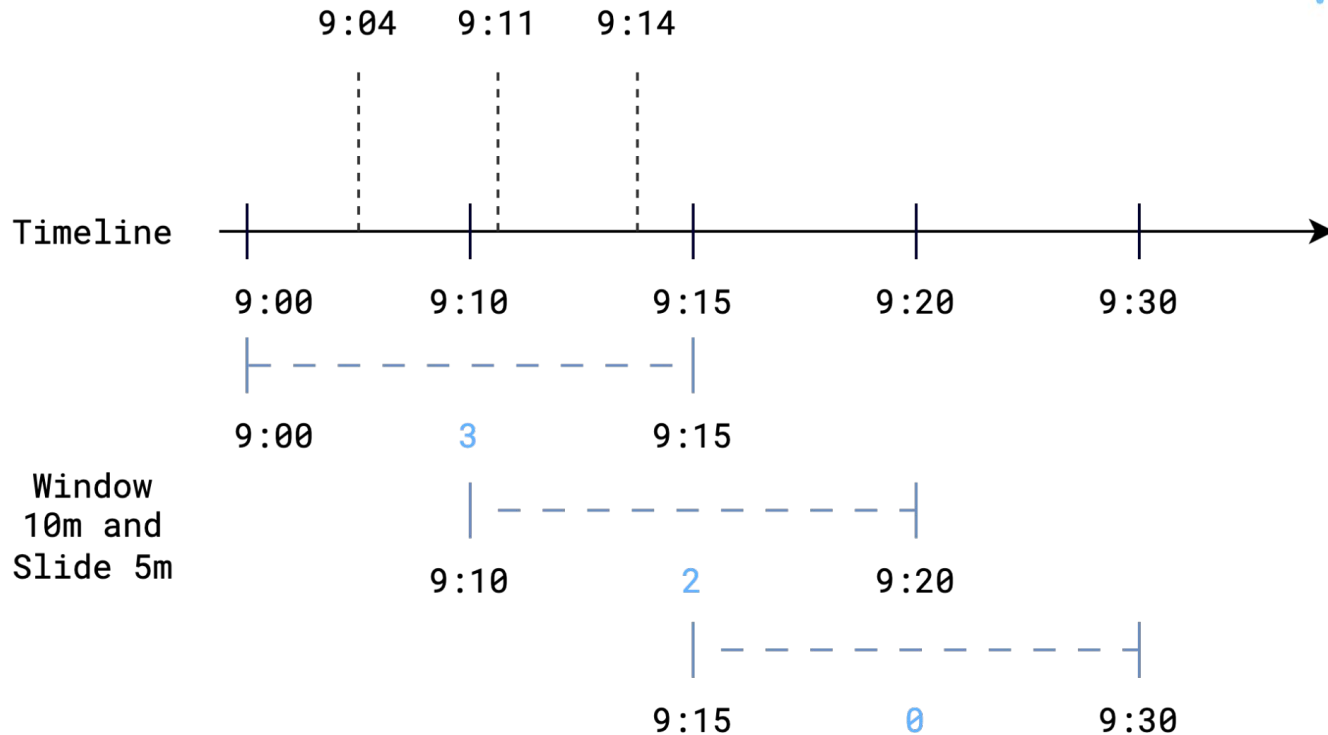
Fixed Time Window

Window cố định



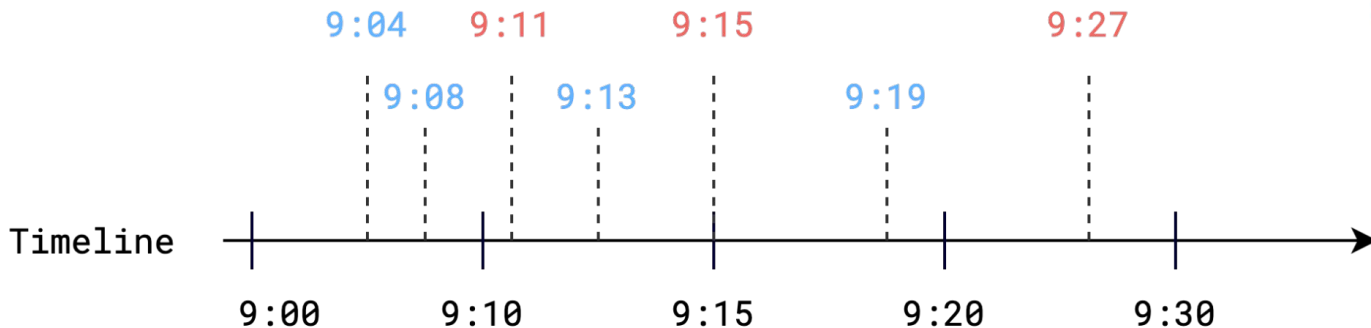
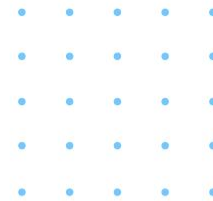
Sliding time Window

Window trượt



Session Window

Window phiên



Window
Session
5m Key, Blue

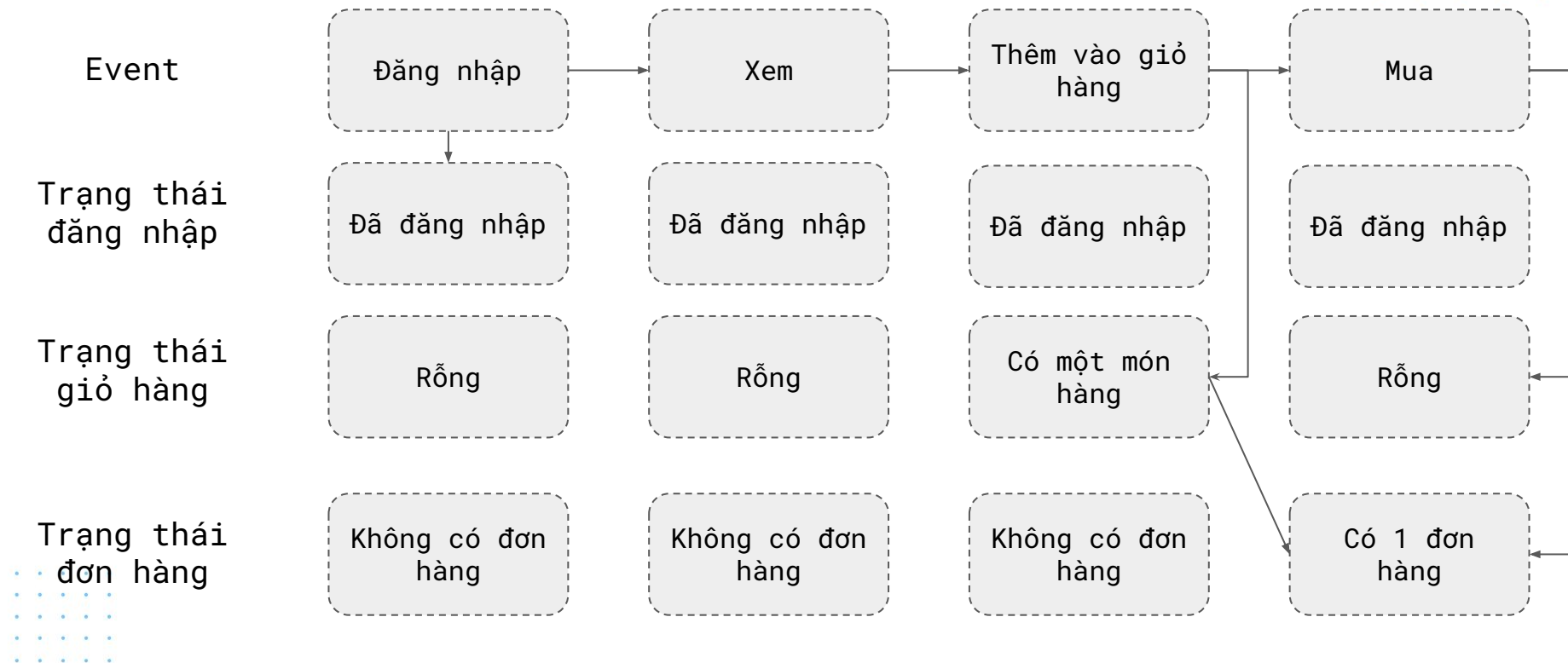


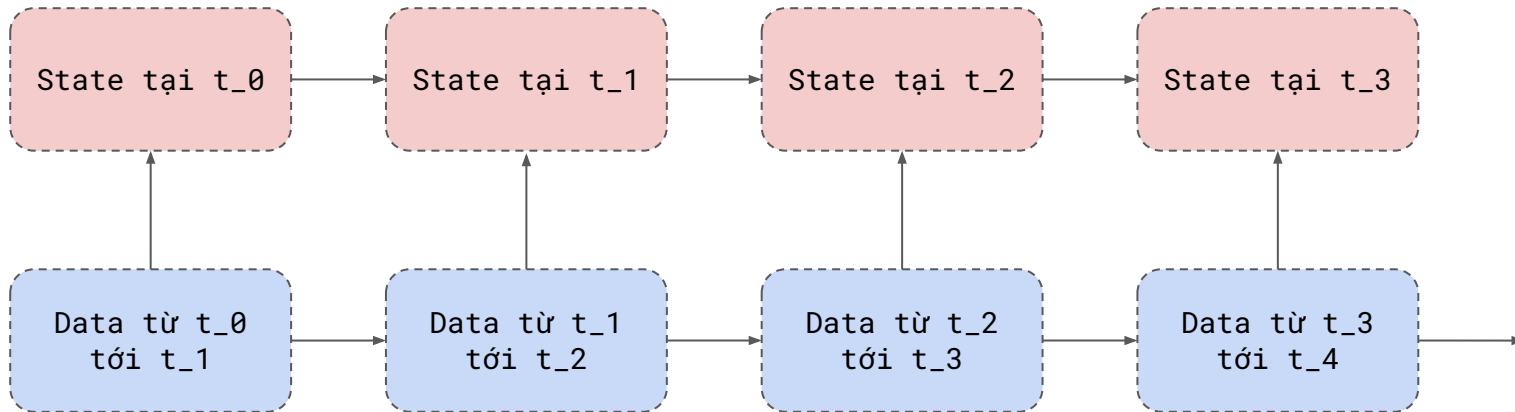
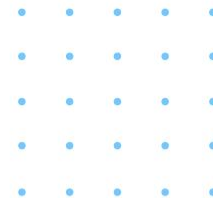
Window
Session
5m Key Red

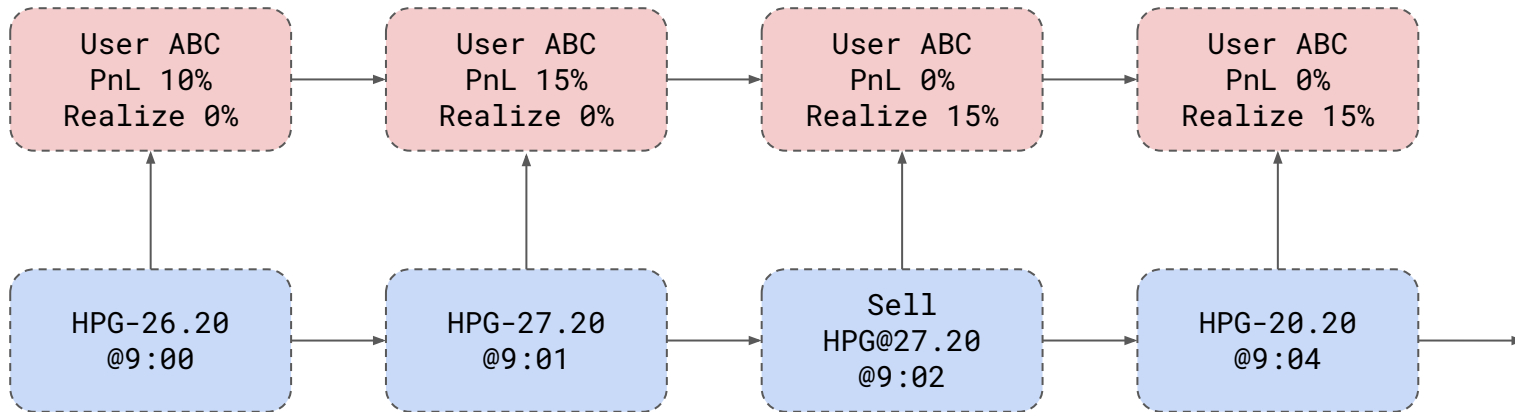
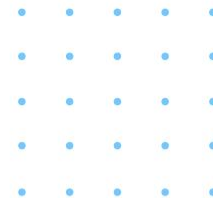


Event so với Trạng Thái

Event versus State

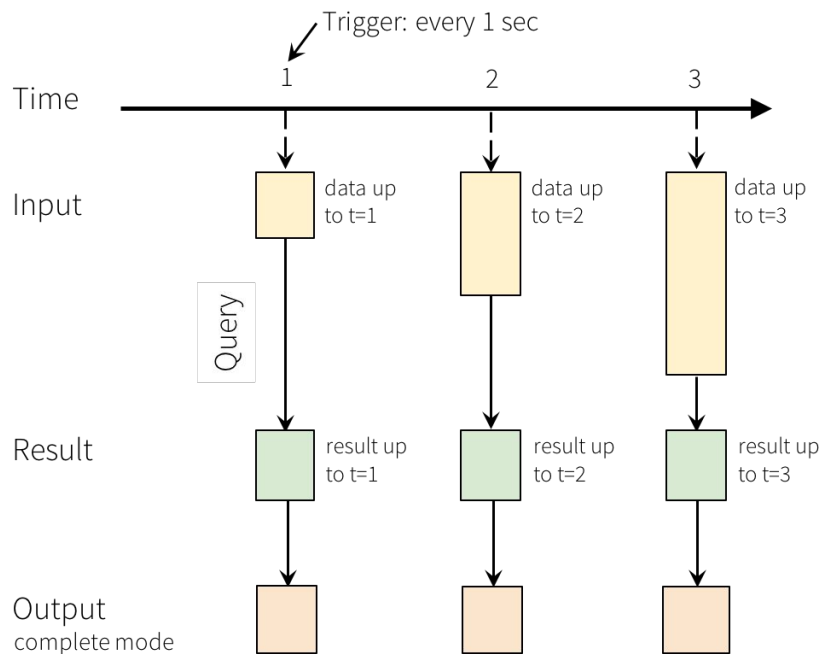






Trigger

Trigger



Programming Model for Structured Streaming

<https://spark.apache.org/docs/3.5.0/structured-streaming-programming-guide.html#triggers>

Các công cụ Streaming

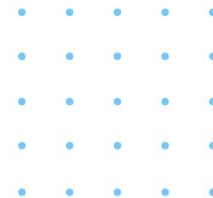
Streaming Tools





Spark Streaming Model

Spark Streaming Model



- **Dstream**
- **Structured Streaming**

