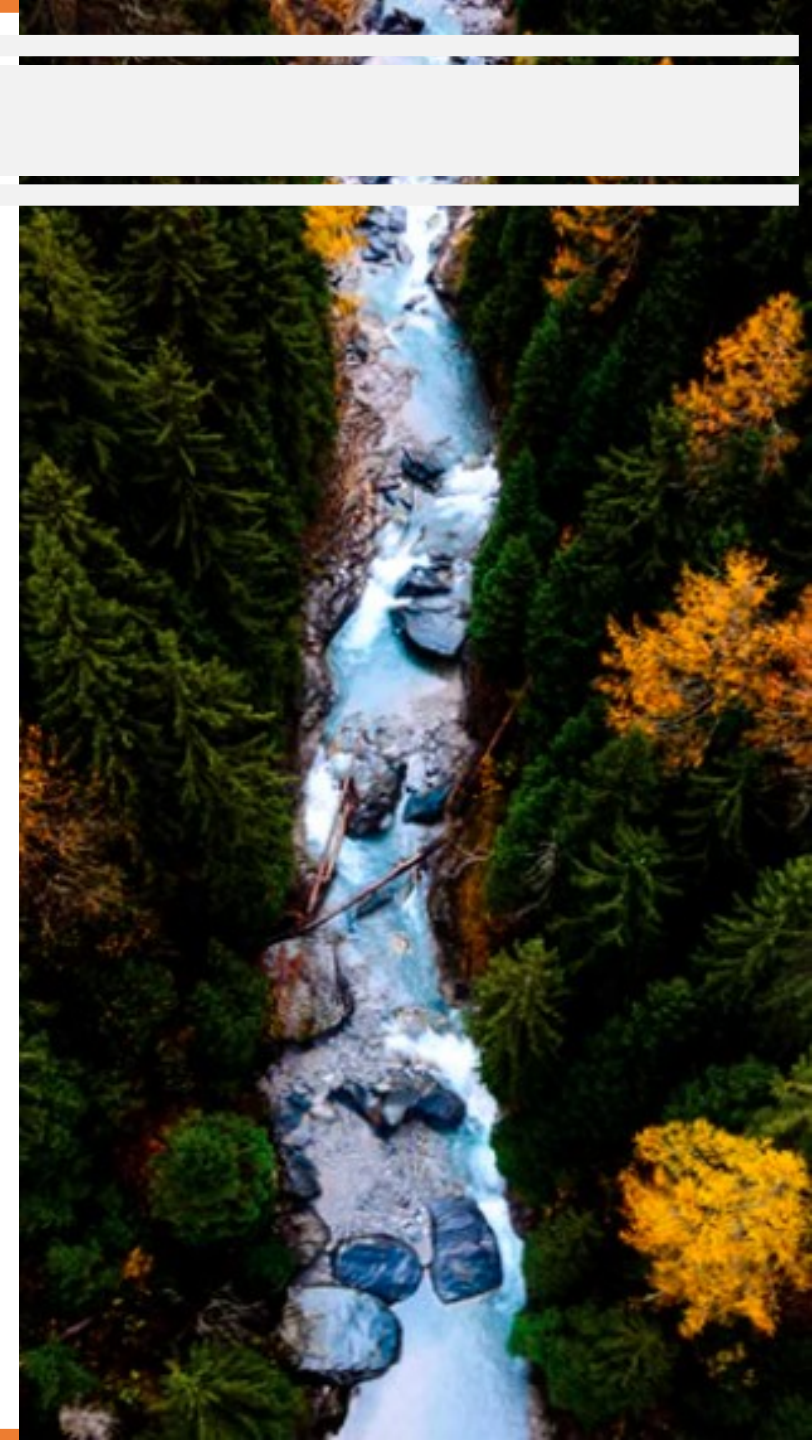




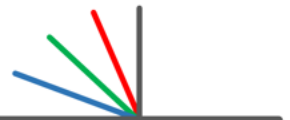
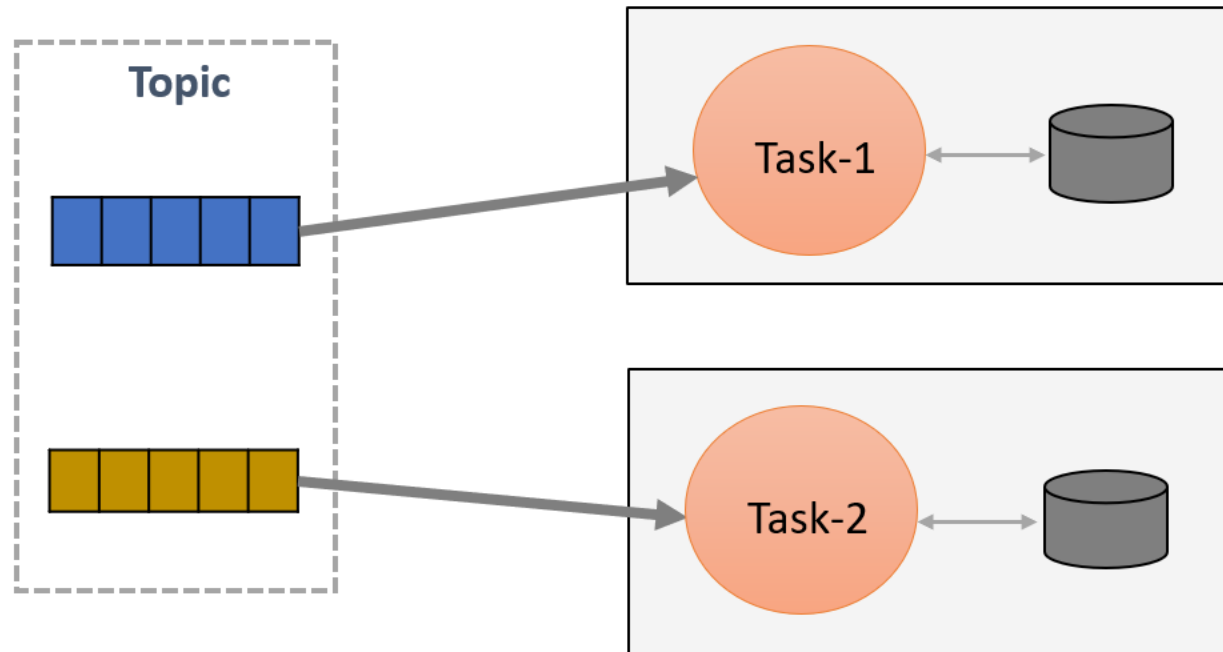
Streaming Aggregates – Core Concept

Aggregation?

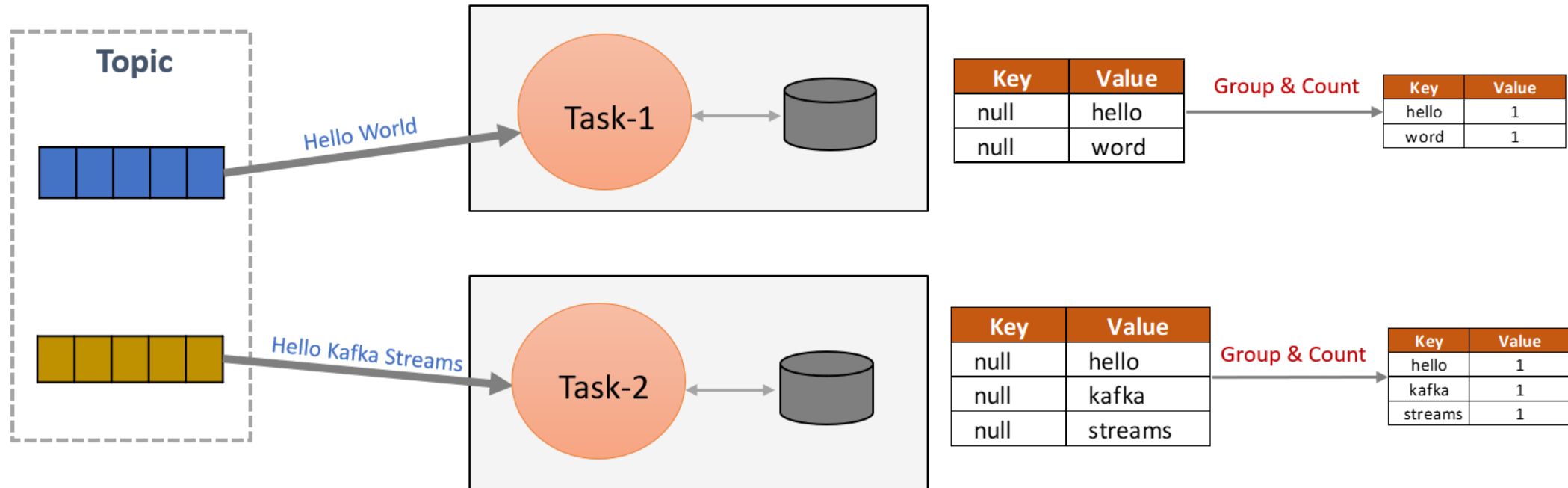
1. Group By a Key
 - Key is the most critical element for aggregation
 - Data must be partitioned on the Key
2. Apply Aggregation formula



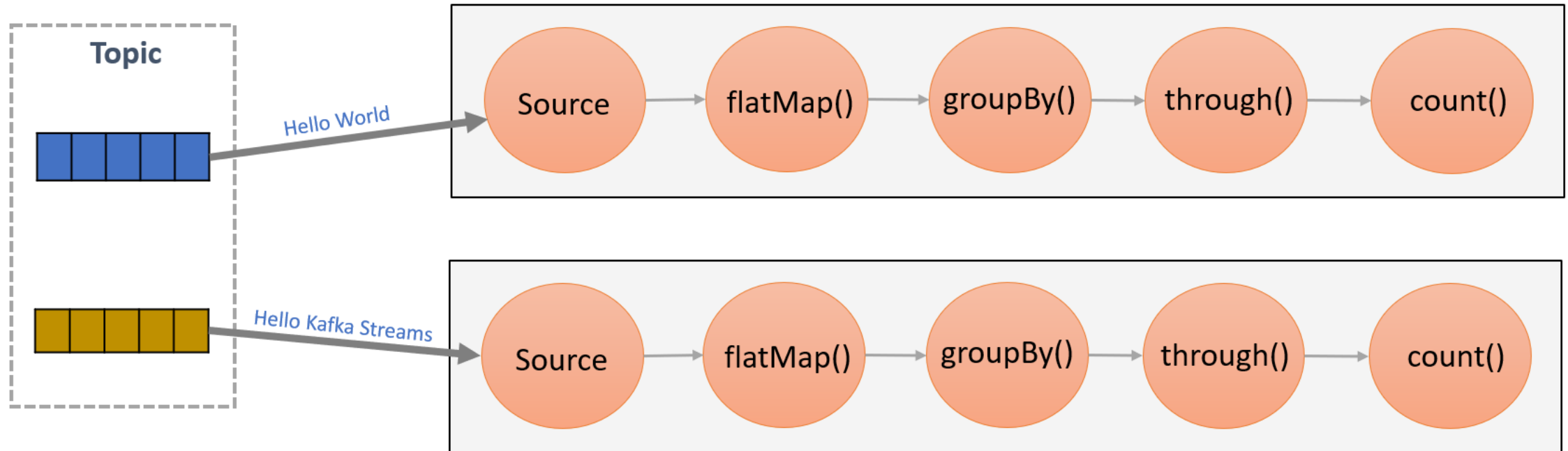
Real-time Aggregation



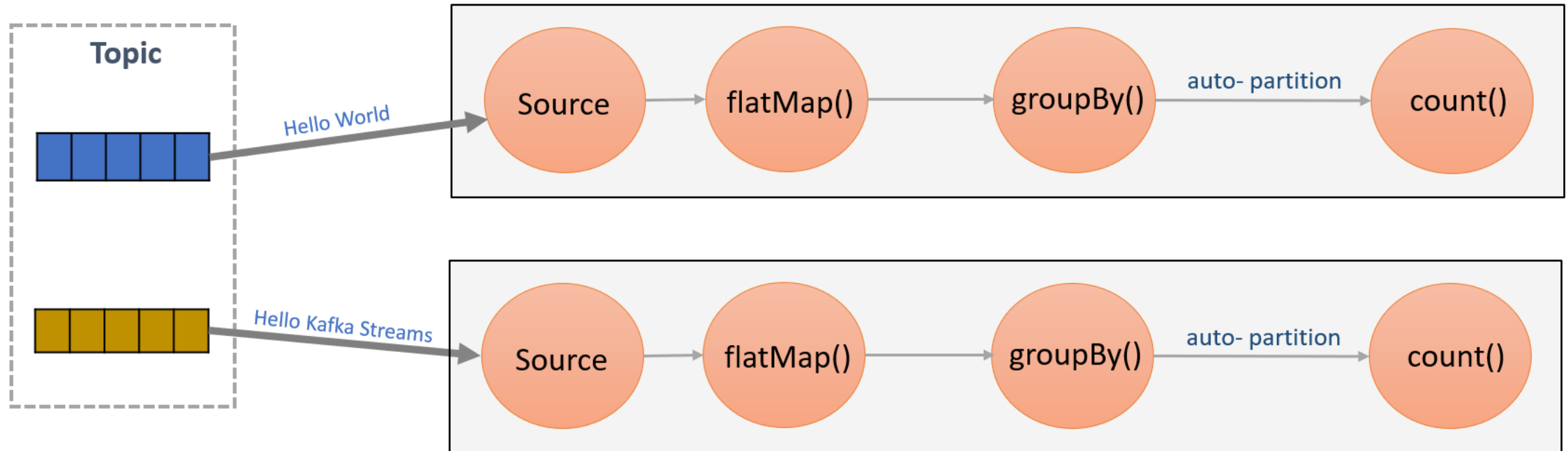
Real-time Aggregation



Real-time Aggregation



Real-time Aggregation



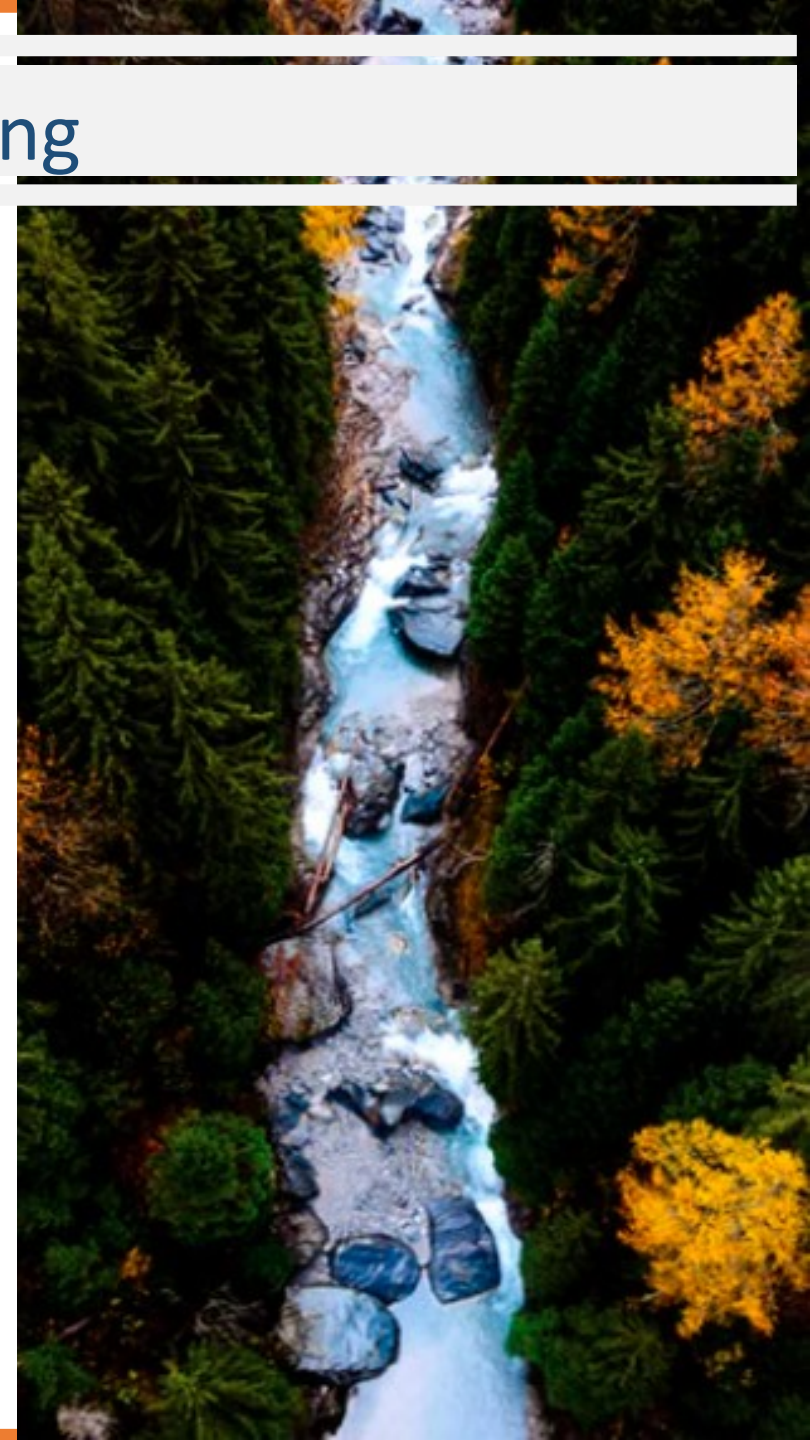
Key changing Vs. Key preserving

Streams API

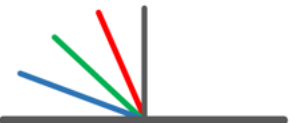
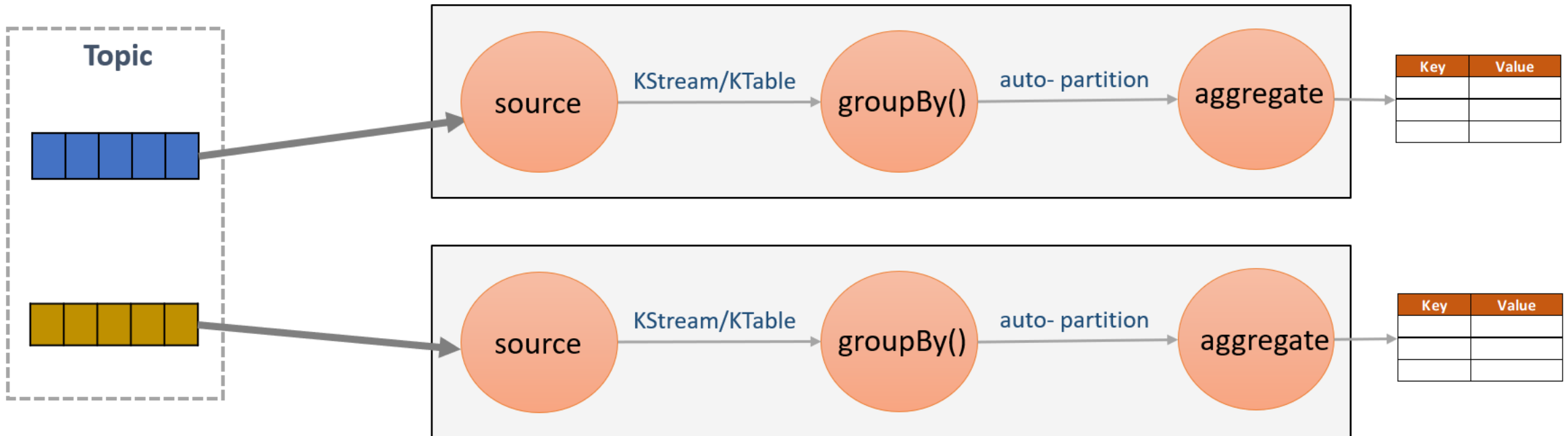
1. Key preserving API — `mapValues()`, `flatMapValues()`, `groupByKey()`
2. Key changing API — `map()`, `flatMap()`, `groupBy()`

Auto repartitioning

1. Use a Key changing API
2. Apply aggregation or join



Real-time Aggregation



Aggregation?

1. Group By a Key

- Key is the most critical element for aggregation
- Data must be partitioned on the Key

2. Apply Aggregation formula

- `count()`
- `reduce()`
- `aggregate()`

