



# Kafka Streams: Interactive Queries

Lee Dongjin | [dongjin@apache.org](mailto:dongjin@apache.org)



# Kafka Streams: 초간단 소개

- Kafka 0.10에서 추가된 Stream Processing Library
  - Framework (x) Library (o)
- Masterless
  - 전체 작업을 n조각 내서 각 thread에 할당된 부분만 처리함 (Consumer Group 기능 활용)
- Statestore
  - 중간 상태를 저장하는 key/value store (default: RocksDB)
  - Local Statestore의 변경 내역을 자동 생성되는 topic에 저장 (changelog topic)
  - Fast Lookup + Failover



## Kafka Streams: 초간단 예제 (wordcount)

```
final StreamsBuilder builder = new StreamsBuilder();
final KStream<String, String> words = builder.stream(wordsTopic);
final KTable<String, Long> counts = words
    .groupBy((key, value) -> value).count();
counts.toStream()
    .to(wordCountTopic, Produced.with(Serdes.String(), Serdes.Long()));
final KafkaStreams streams = new KafkaStreams(builder.build(), props);
streams.start();
```



# Kafka Streams: 문제

- 중간 상태를 보고 싶다면 어떻게 해야 할까?
  - "지금까지 집계된 단어별 개수를 확인하고 싶은데...?"
  - "지금까지 집계된 단어별 개수를 Remote API를 통해서 서비스하고 싶은데...?"
- 내가 원하는 key에 대한 값이 현재 작업중인 host에 없다면?
  - "전체 작업을 n조각 내서 처리한다" = "중간 상태도 n조각 나서 최대 n개 host에 분할될 수 있다."
  - "(k1, v1), (k2, v2)는 host A에, (k3, v3)은 host B에 있을 수 있다."
  - A가 v3을 필요로 한다면? 반대로 B가 v1, v2를 필요로 한다면?



# Kafka Streams: Interactive Query (1)

- 기능

- StateStore의 중간 상태를 읽을 수 있게 해 줌
- 내가 원하는 key를 가진 StateStore의 조각(partition)을 어느 host가 쥐고 있는지를 알려 줌.
  - "k3에 대한 값을 저장하고 있는 StateStore는 host B가 쥐고 있습니다."



## Kafka Streams: Interactive Query (2)

```
// 1. StateStore를 생성하는 연산을 호출할 때 Materialized 옵션을 사용해서
// QueriableStore 이름을 지정한다. (i.e., "'wordcount-store'라는 이름을 붙이겠다.")
final KTable<String, Long> counts = words
    .groupBy((key, value) -> value).count(Materialized.as("wordcount-store"));

// 2. KafkaStreams 객체에 주어질 옵션값에 StreamsConfig.APPLICATION_SERVER_CONFIG를
// 추가한다. (i.e., '이 process는 7070번 포트에서 Rest API 요청을 받고 있습니다.')
properties.put(StreamsConfig.APPLICATION_SERVER_CONFIG, "localhost:7070");
```



## Kafka Streams: Interactive Query (3)

```
// "어느 instance가 'wordcount-store' 'marvin'라는 key에 대한 값을 주고 있는가?"
final StreamsMetadata metadata =
    streams.metadataForKey("wordcount-store", "marvin", new StringSerializer());
HostInfo hostInfo = metadata.hostInfo();

// local에 위치한 KeyValue Store 내용을 질의한다.
final ReadOnlyKeyValueStore<String, Long> myCount =
    streams.store("wordcount-store",
        QueryableStoreTypes.<String, Long>keyValueStore());
```



## Kafka Streams: Interactive Query (4)

- Q1. 원격 질의를 하는/받는 code는 어떻게 구현해야 하나요?
  - 님 편한 걸로.
- Q2. KeyValue Store만 있나요?
  - 상태를 누적시키는 방식에 따라 변종이 있습니다.
  - Timestamp가 딸려나오는 것과 아닌 것.

Class	StateStore
KTable	(Timestamped)KeyValueStore
SessionWindowedStream	(Timestamped)SessionWindowStore(*)
TimeWindowedStream	(Timestamped)WindowStore





# 정리

- Kafka Streams는 상태 누적 연산을 지원한다.
  - KTable, SessionWindowedStream, TimeWindowedStream
  - 그리고 그 내용을 Interactive Query 기능을 사용해서 열어볼 수 있다.
- Interactive Query가 해 주는 것:
  - 현재 process에서 잡고 있는 StateStore 의 Read-only View
  - 현재 process에서 잡고 있는 StateStore 조각(partition)에 포함되지 않은 값이 저장되어 있는 위치.
- 직접 해 줘야 하는 것:
  - Remote API 구현



# 질문?

- Kafka 한국 사용자 모임 (KafkaKru)
  - <https://www.facebook.com/groups/kafka.kru/>