- ____ 실시간 스트리밍
 - 1) 특징

Event-driven 아키텍처로 구현됨

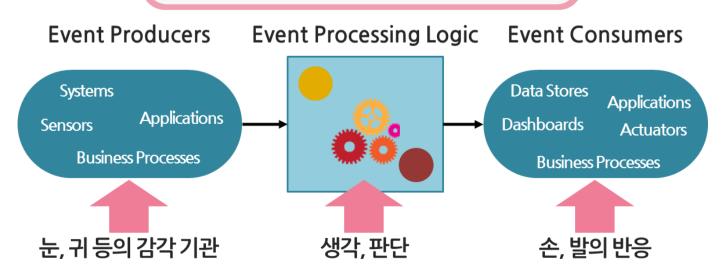
실시간이란 상대적인 개념으로 요건에 따라 실시간의 범위가 정의될 수 있음

낮은 수준의 지연 시간(Low Latency)

일정한 응답 속도의 보장과 예측 가능한 성공 제공

② 이벤트 기반 아키텍처

사람의 감각 기관 및 반응 구조와 유사함





1 실시간 스트리밍

	배치 처리	스트림 처리	
데이터 범위	데이터 세트의 모든 또는 대부분 데이터를 쿼리하거나 처리	롤링 타임 윈도우 내 데이터 또는 가장 최신 데이터 레코드의 데이터를 쿼리하거나 처리	
데이터 크기	대규모 데이터 배치	일부 레코드로 구성된 마이크로 배치 또는 개별 레코드	
성능	지연 시간이 몇 분에서 몇 시간	몇 초 또는 몇 밀리초의 지연 시간이 필요	
분석	복잡한 분석	간단한 응답 가능, 수집 및 롤링 지표	

배치 처리와 스트리밍 처리 비교

3 사례

주식 트레이딩

쇼핑몰 사용자 클릭 스트림 기반의 개인화 추천

실시간 분석, 추천, 광고

제조 공정 관리

지난 30분간 주문 상품 Top 10

실시간 금융 사기 방지

실시간 모니터링

보안 관제



스파크 스트리밍

RDD

• 스파크의 기본 데이터 셋 추상화 객체

Dstream(Discretized Stream)

- 불연속화 스트림
- 스파크 스트리밍의 데이터셋 추상화 객체



정의된 시간 간격 동안 이벤트 일괄 처리를 우선 수집하여 데이터를 마이크로 단위로 일괄로 처리

input data		batches of		batches of
stream	Spark Streaming	input data	Spark Engine	processed data



스파크 스트리밍

각 RDD는 일괄 처리 간격 동안 수집된 이벤트를 나타냄

RDD의 연속 집합은 DStream으로 수집됨

RDD@ time 1 RDD@ time 2 RDD@ time 3 RDD@ time 4

DStream -- lines from time 0 to 1 lines from time 1 to 2 lines from time 2 to 3 lines from time 3 to 4

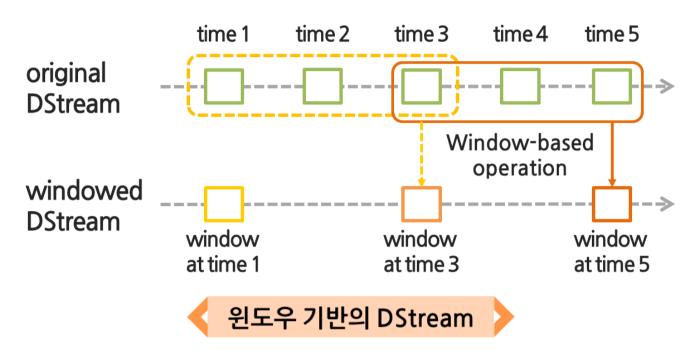
마이크로 배치 (Micro-batch)

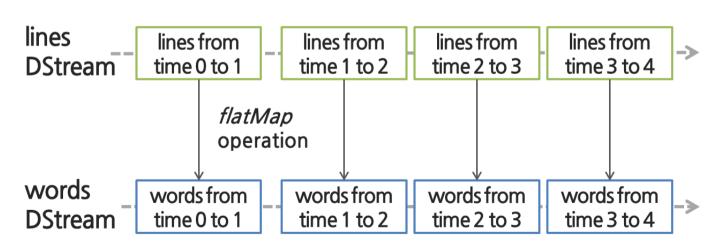
데이터 스트림을 개별 세그먼트로 나눈 후 각 세그먼트의 데이터를 스파크 엔진으로 처리

에이터를 초 단위의 타임 윈도우로 나눈 후 스파크 엔진을 실행함



스파크 스트리밍





마이크로 배치와 작은 배치의 연속