ANT310

Architecting for Real-Time Insights with Amazon Kinesis

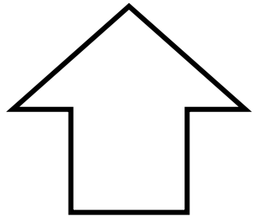
Hyobin An
Solutions Architect

AWS re:Invent

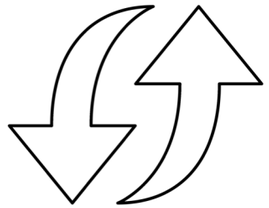
aws

What is streaming data?

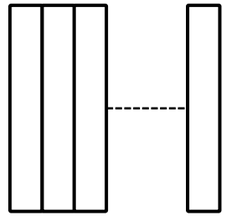
Typical characteristics



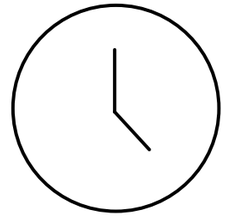
High volume



Continuous



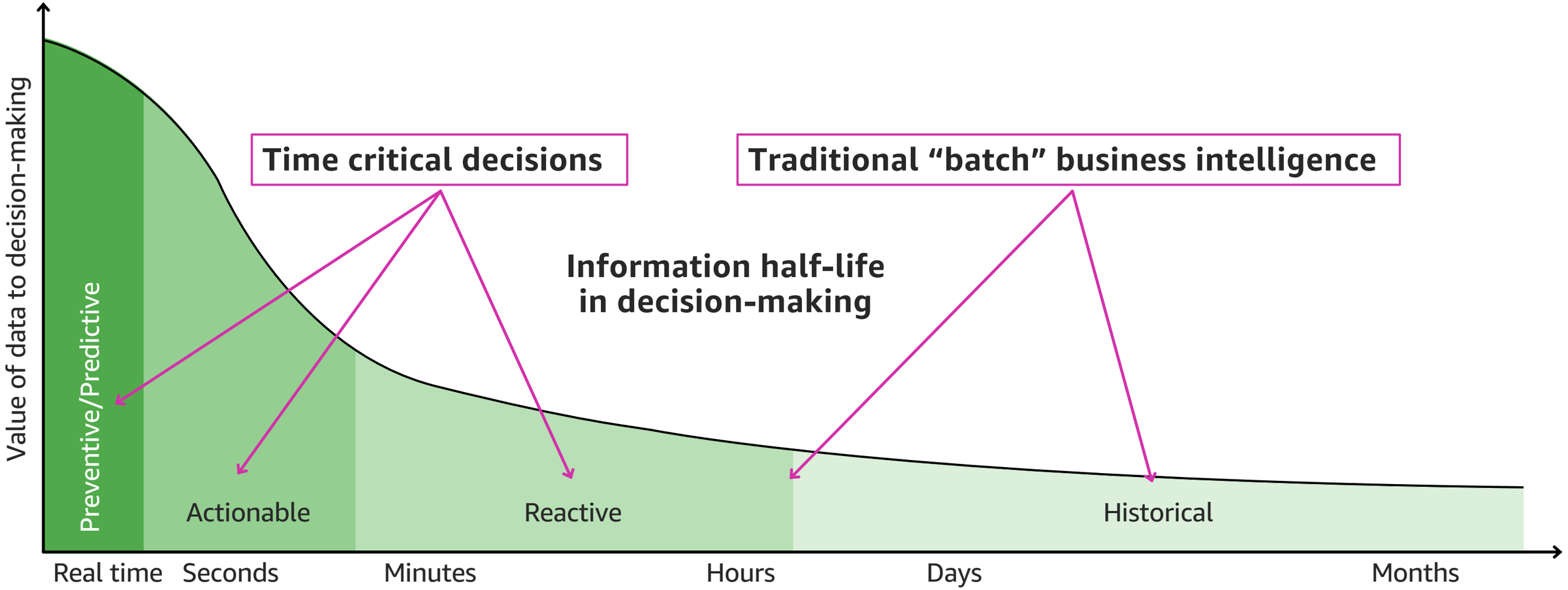
Ordered, incremental



Low-latency

Why streaming data?

Get actionable insights quickly



Source: Perishable insights, Mike Gualtieri, Forrester

Streaming with Amazon Kinesis

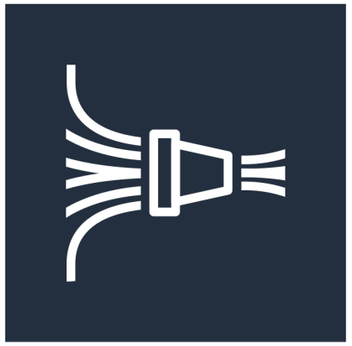
Easily collect, process, and analyze video and data streams in real time

Amazon Kinesis Video Streams

Capture, process, and store video streams

Amazon Kinesis Data Streams

Capture, process, and store data streams



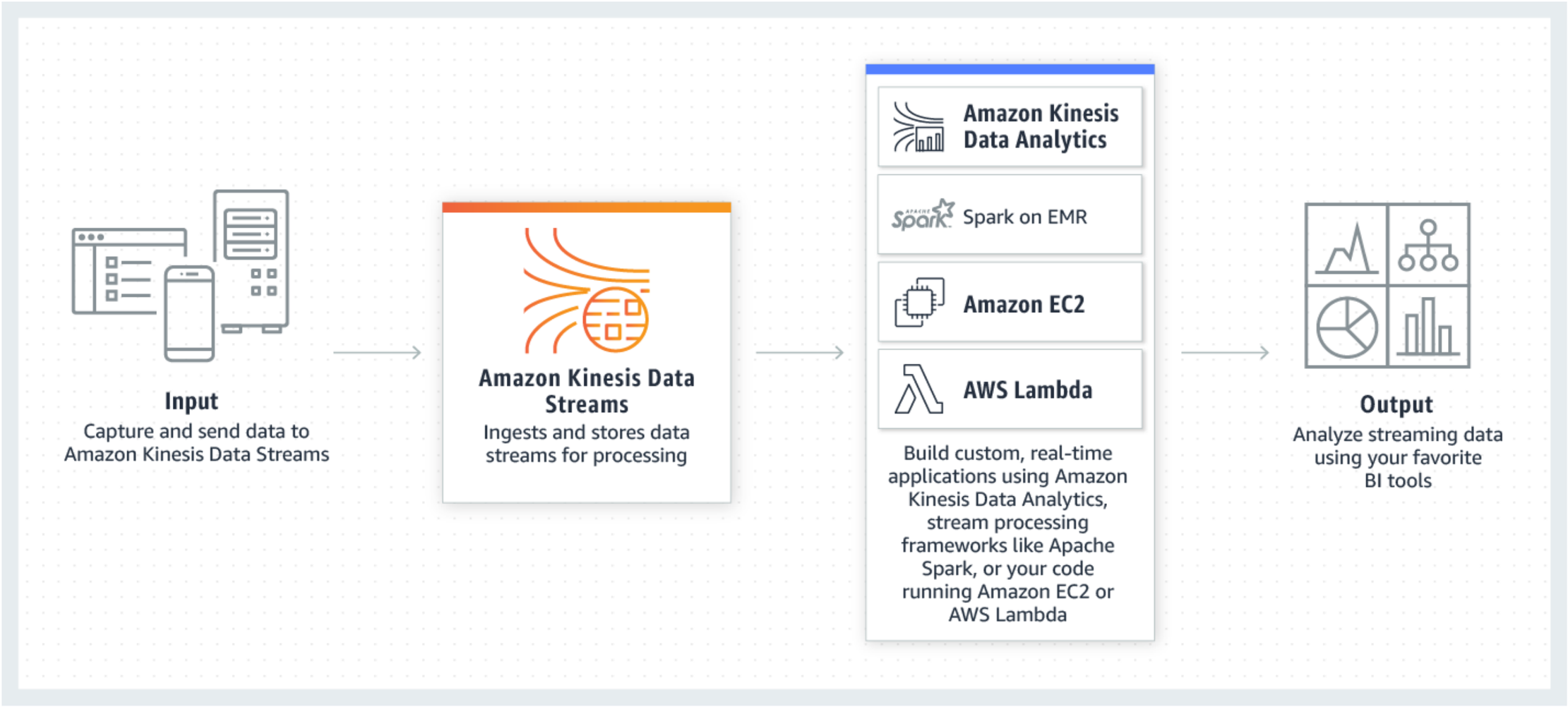
Amazon Kinesis Data Firehose

Load data streams into AWS data stores

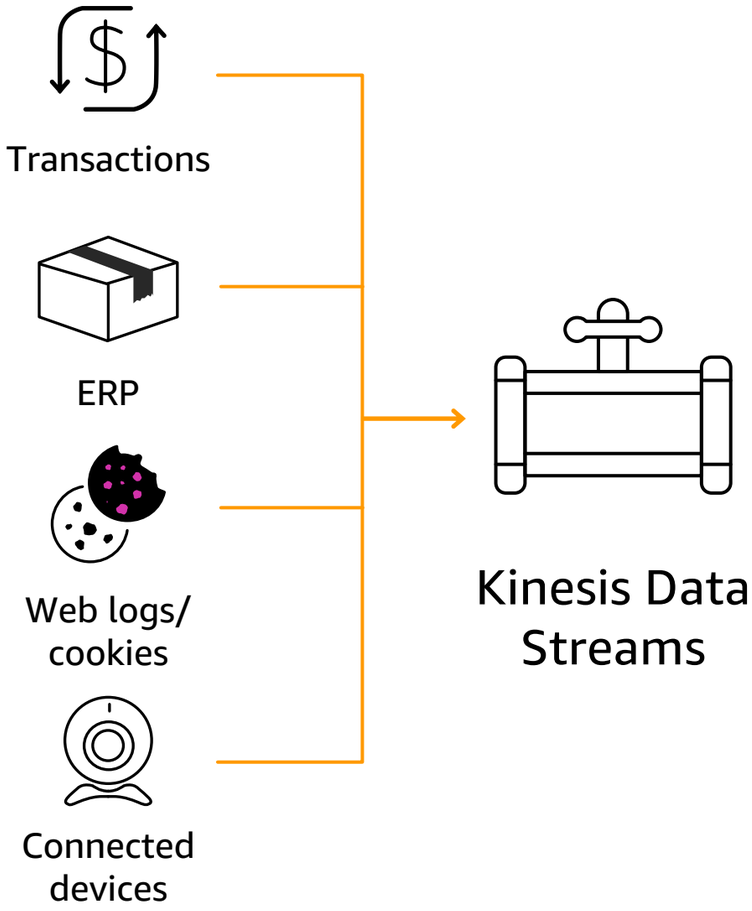
Amazon Kinesis Data Analytics

Analyze data streams in real time

Amazon Kinesis Data Streams overview



Data ingestion from a variety of sources



AWS SDKs

- Publish directly from application code via APIs
- AWS Mobile SDK
- Managed AWS sources: CloudWatch Logs, AWS IoT, Kinesis Data Analytics and more
- RDS Aurora via Lambda

Kinesis Agent

- Monitors log files and forwards lines as messages to Kinesis Data Streams

Kinesis Producer Library (KPL)

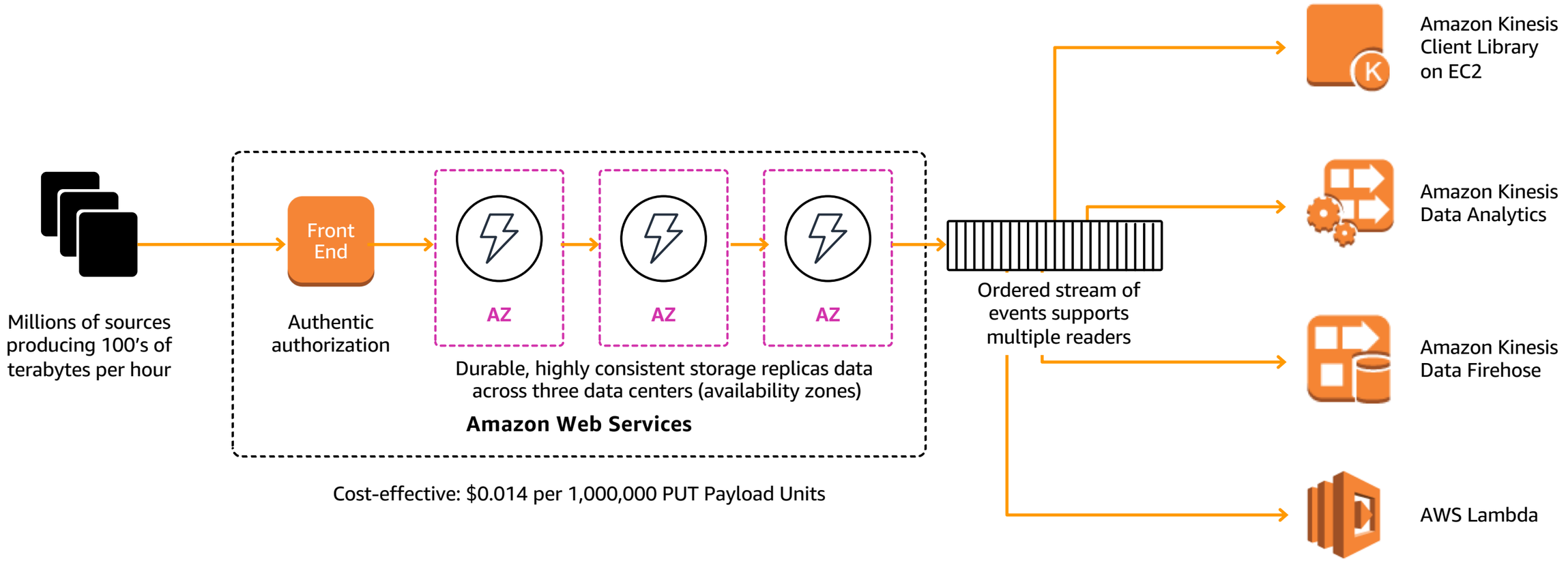
- Background process aggregates and batches messages

3rd party and open source

- Log4j appender
- Apache Kafka
- Flume, fluentd, and more …

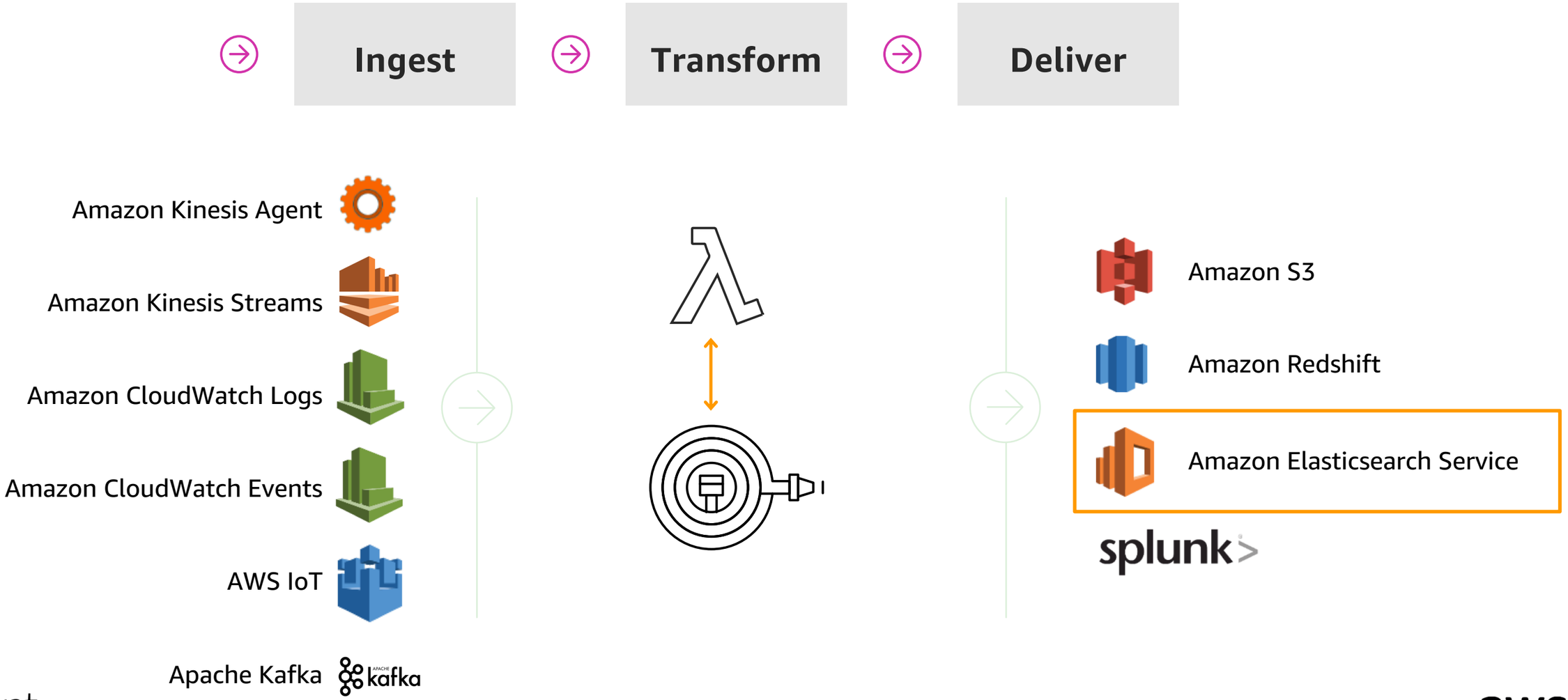
Data processing from a variety of consumers

Fully managed service for real-time processing of streaming data



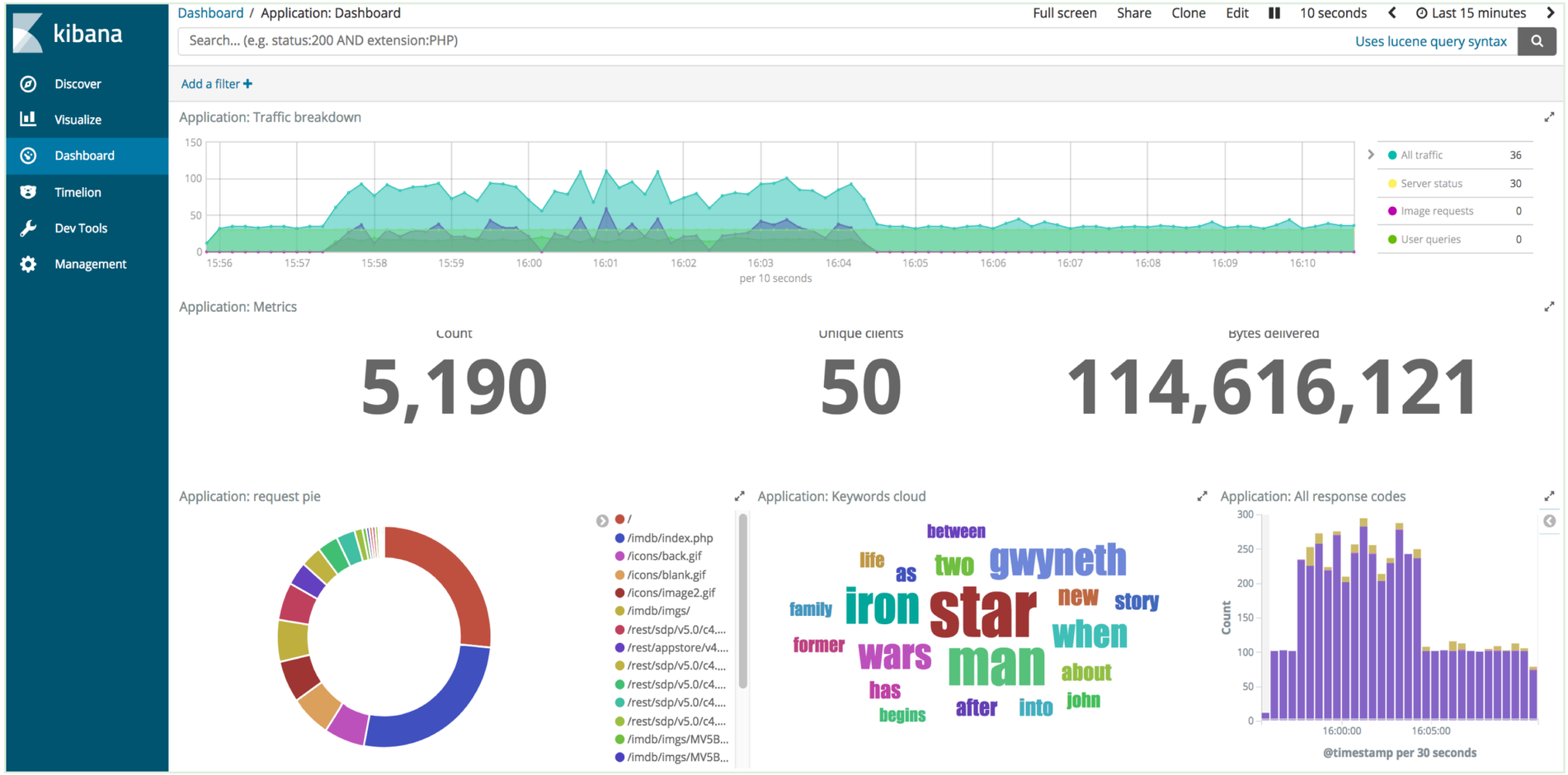
Cost-effective: $0.014 per 1,000,000 PUT Payload Units

Amazon Kinesis Data Firehose—How it works



Elasticsearch with Kibana

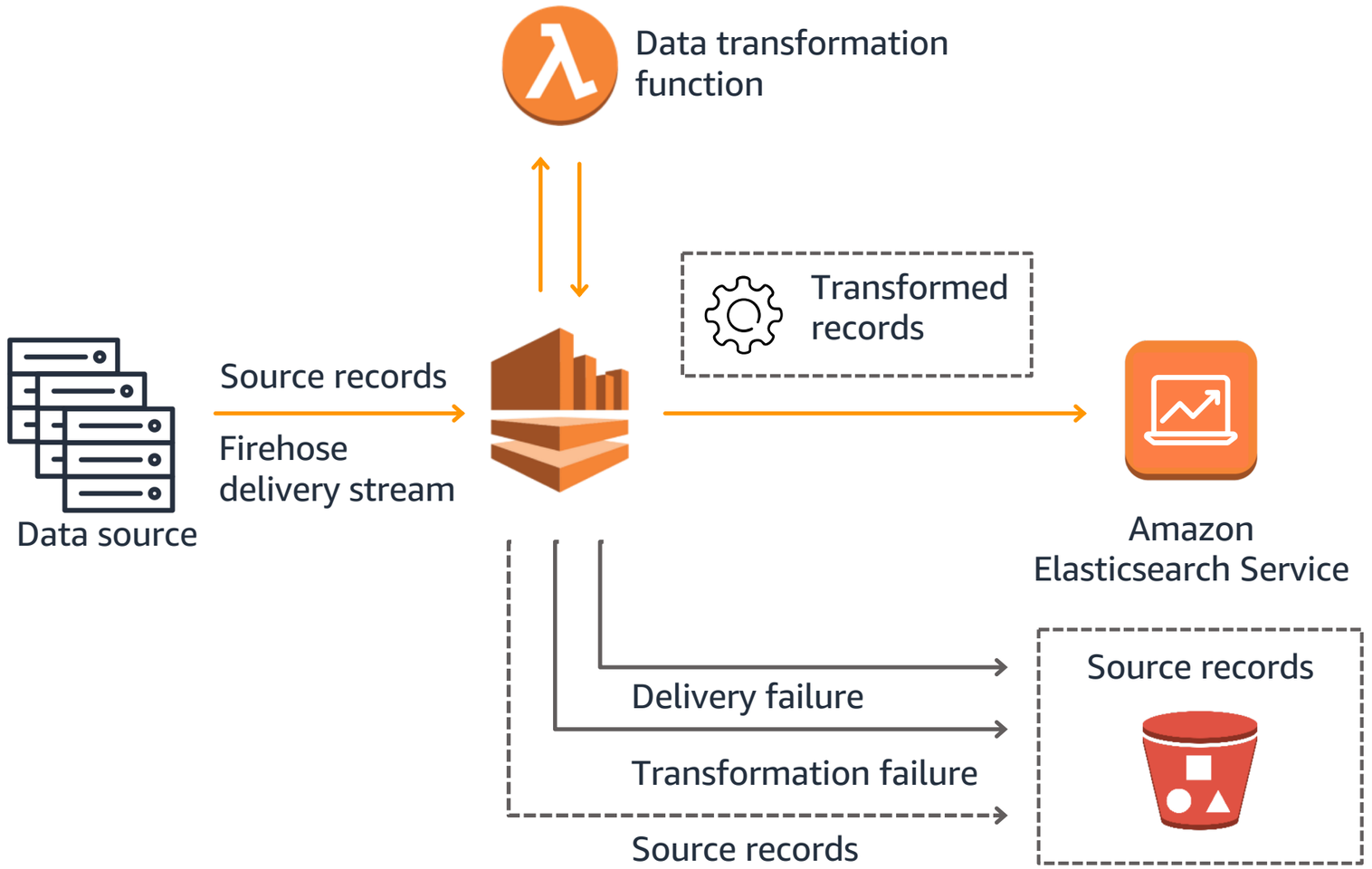
Derive insights from logs



Amazon Elasticsearch Service

Fully-managed service that is easy to deploy, manage, and scale

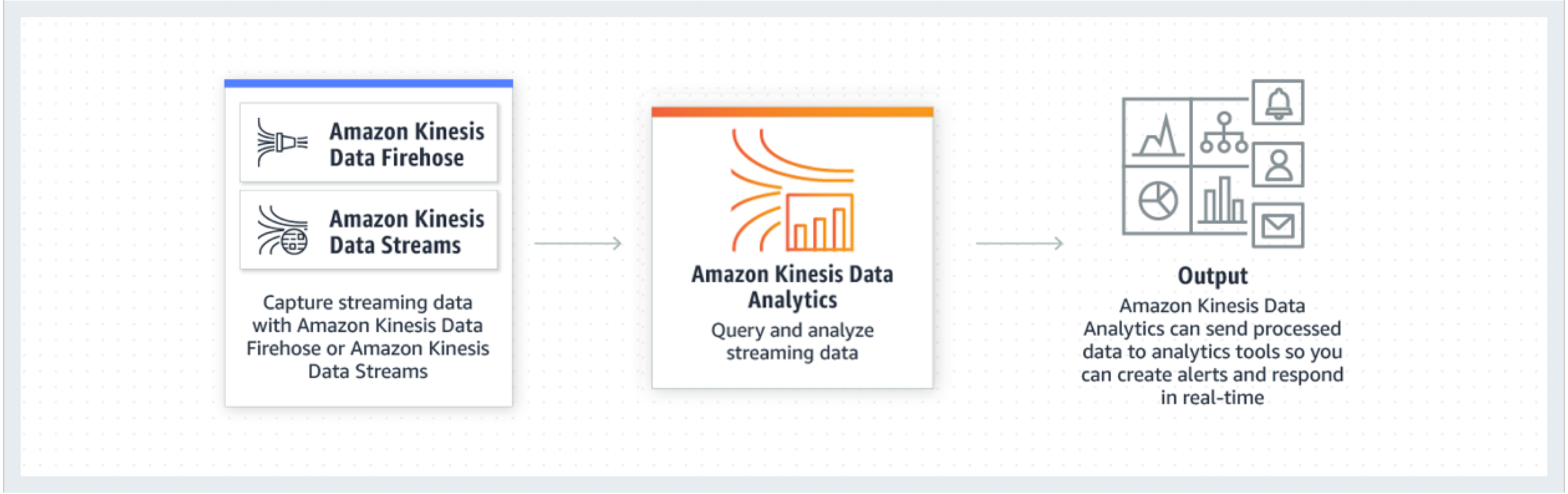
Amazon Kinesis Firehose delivery architecture



- For public access domains
- Easily transform data
- Serverless with built-in batching, index rollover, error handling

Amazon Kinesis Data Analytics

Continuously read, process, and deliver streaming data in real-time



SQL on streaming data?

Aggregations (count, sum, min, …) take granular real-time data and turn it into insights

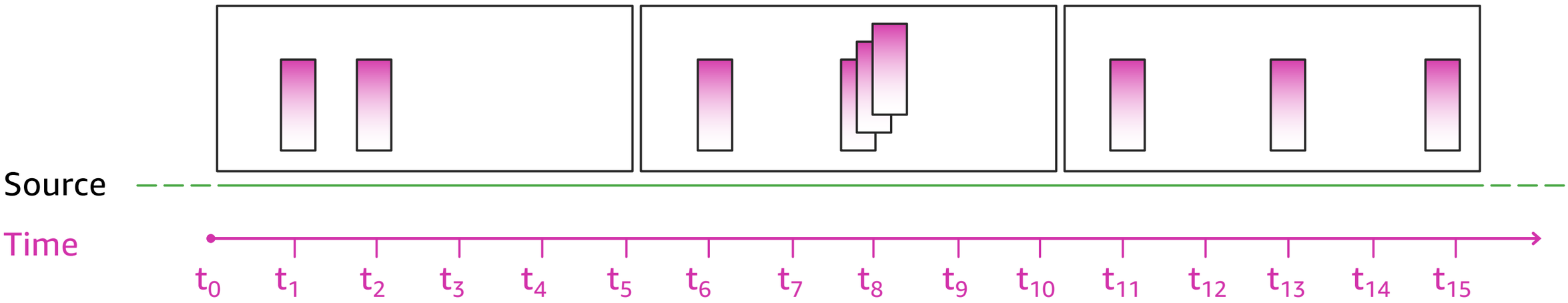
Data is continuously processed so you need to tell the application when you want results

Windows!

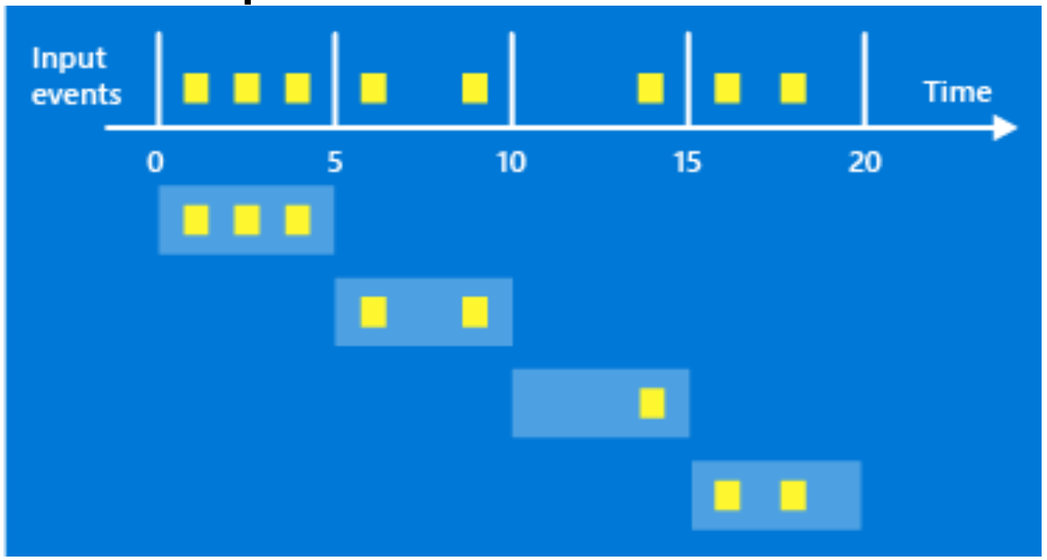
Window types

Sliding, tumbling, and stagger

Tumbling windows are fixed size and grouped keys do not overlap

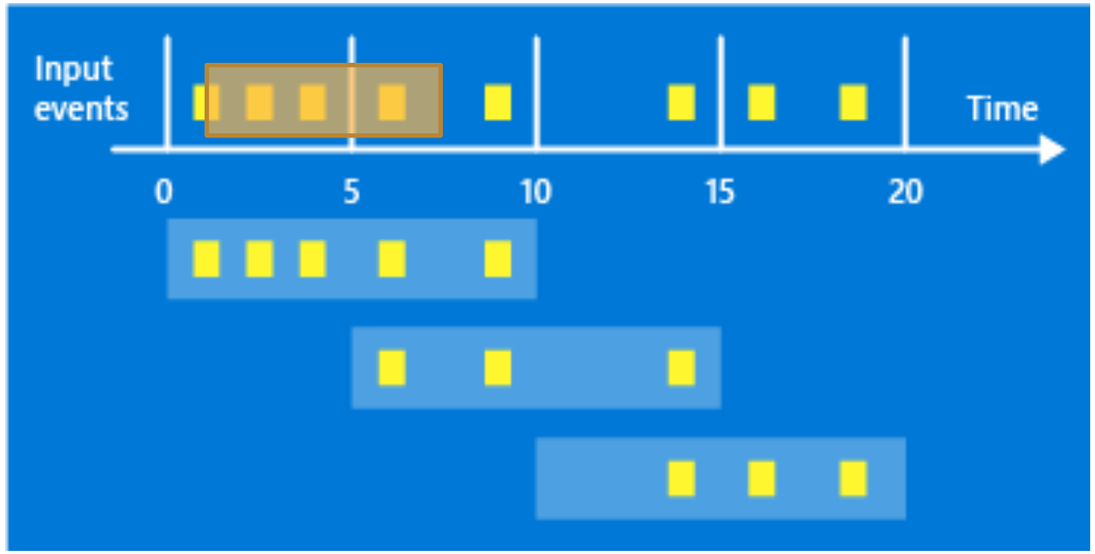


Compare 2 Windows



Tumbling Window

Aggregation by time interval



Sliding Window

Window that are constantly re-evaluated

- Output is created at the end of the window
- Window output is a single event based on the aggregation function used

Writing streaming SQL

Pump (continuous query) using stagger window

```
CREATE OR REPLACE PUMP calls_per_ip_pump AS
INSERT INTO calls_per_ip_stream
SELECT STREAM source_ip_address,
     COUNT(*)
FROM source_sql_stream_001
WINDOWED BY STAGGER(
     PARTITION BY source_ip_address
     RANGE INTERVAL '1' MINUTE);
```

Kinesis Customer Case

근 실시간으로
게임 이벤트
분석

NETFLIX

실시간으로
수십억 개의
네트워크 흐름
분석

SONOS

연결된
장치로부터 매주
10억 개의 이벤트

Zillow

근 실시간 주택
가치 평가
(Zestimates)

THOMSON REUTERS

10초 미만으로
갱신되는 라이브
클릭스트림 대시보드

IoT 예측 분석

HEARST corporation

250개 이상의
사이트에서 매일
100GB의
클릭스트림

AdRoll

50ms 이하
응답으로 연간
50억 회의 광고
노출

NORDSTROM

하루 1,000만 건의
온라인
스타일리스트 처리

lyft

100개 이상의
마이크로서비스간
통신