Data Analysis with APACHE Spark™

3. Spark Structured Streaming

ADALTAS

What is Streaming Data

- Continuously generated data
- Coming from many data sources simultaneously
- Small in size (kB range)
- Examples: IoT, stock market, recommendation based on geo-location, electricity consumption...

Difference between Batch and Stream Processing

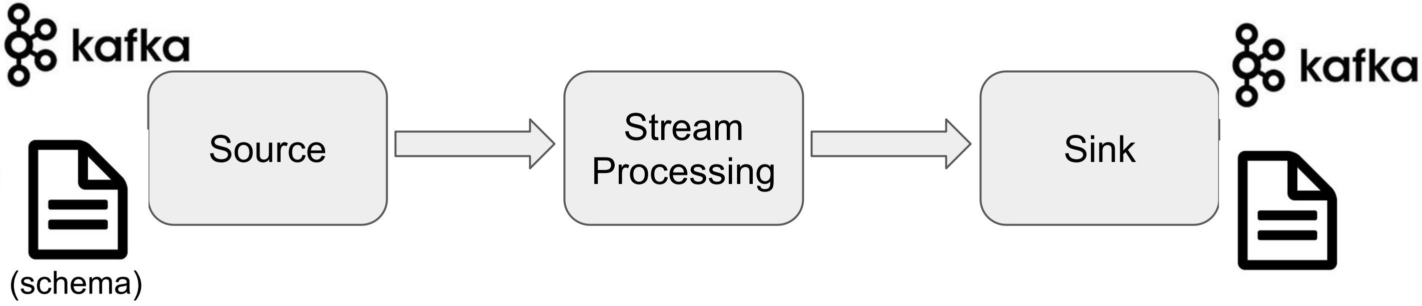
	Batch processing	Stream processing
Data scope	Queries or processing over all or most of the data in the dataset.	Queries or processing over data within a rolling time window, or on just the most recent data record.
Data size	Large batches of data.	Individual records or micro batches consisting of a few records.
Performance	Latencies in minutes to hours.	Requires latency in the order of seconds or milliseconds.
Analyses	Complex analytics.	Simple response functions, aggregates, and rolling metrics.

Streaming

- Unbounded data sets (in opposition to finite data sets);
- Unbounded data processing (in time);
- Spark Structured Streaming: fault-tolerant exactly-once processing

Streaming data sources and sinks

- Apache Kafka
- Amazon Kinesis
- Load files from S3 using Auto Loader
- Optimized Amazon S3 Source with Amazon SQS
- Azure Event Hubs
- Delta Lake tables
- Read and write streaming Avro data
- Write to arbitrary data sinks



Stream processing models

- **Traditional** vs **micro-batch model**

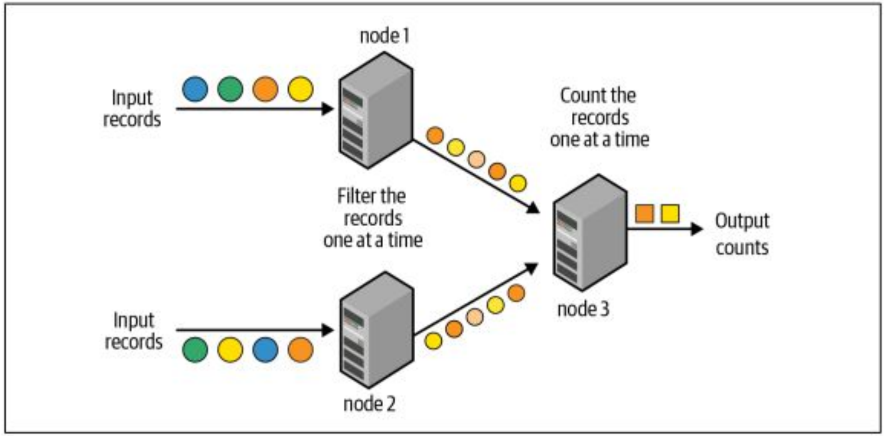


Figure 8-1. Traditional record-at-a-time processing model

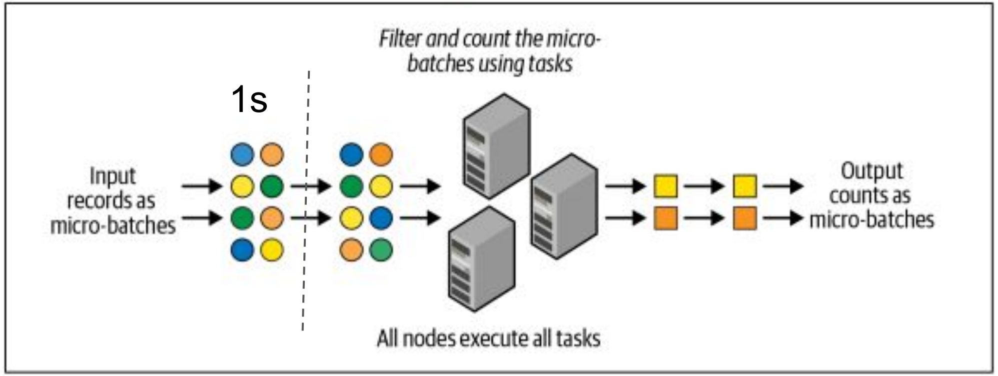


Figure 8-2. Structured Streaming uses a micro-batch processing model

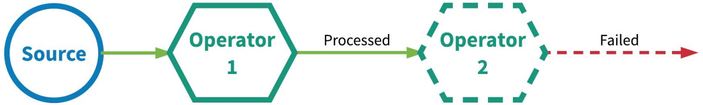
- redundancy
+ latency (ms)

+ redundancy
- latency (100 ms: exactly-once)

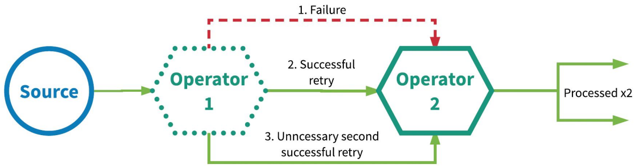
Different processing semantics

ADALTAS

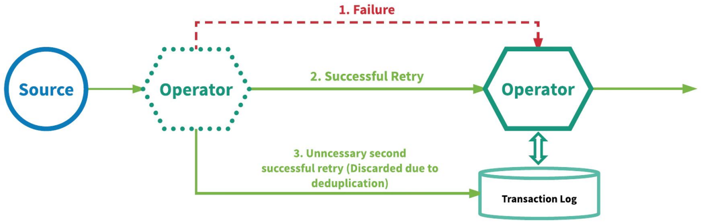
- At-most-once



- At-least-once

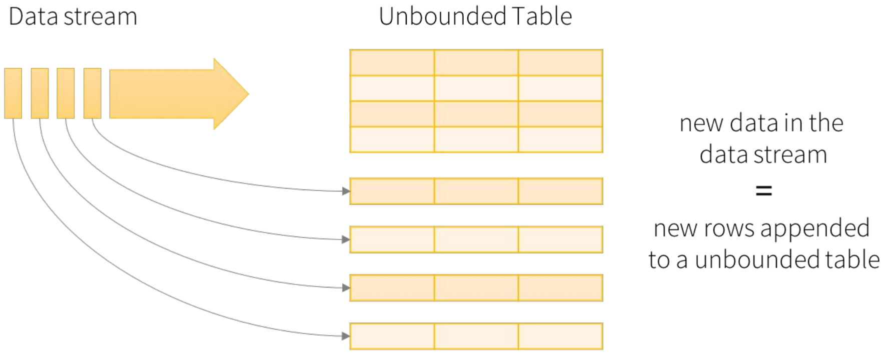


- Exactly-once

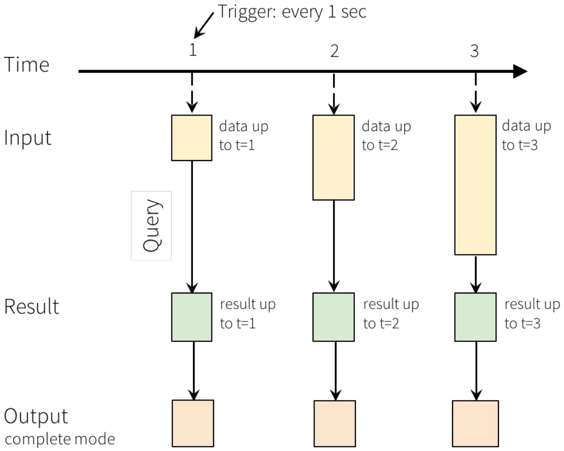


Programming model

- Streaming computation as standard batch-like query



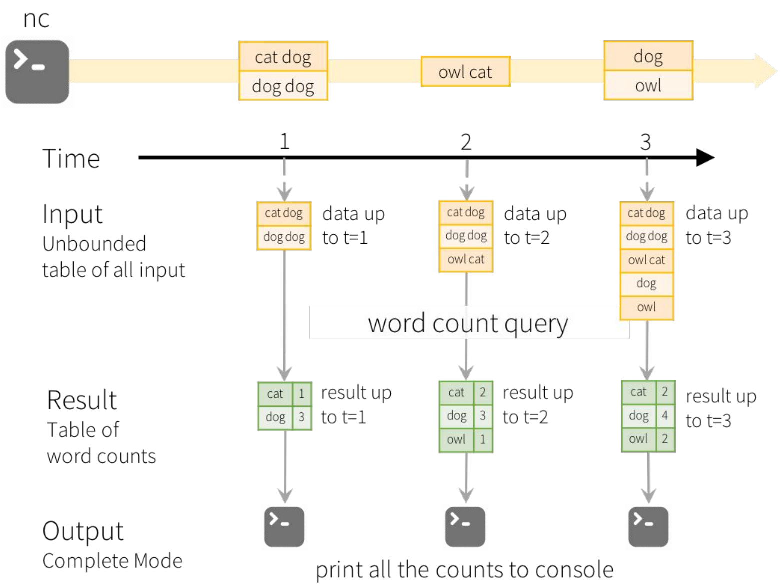
Data stream as an unbounded table



Programming Model for Structured Streaming

Programming model: example

- Word count

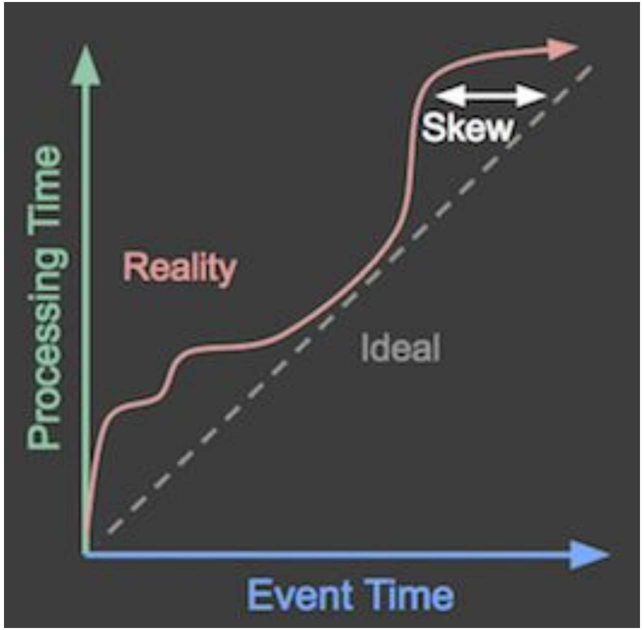


Output modes:

- complete - writes all the rows of a Result Table
- append - writes "new" rows only
- update - writes only the rows that were updated

Event-time vs Processing time

- event time - when the data was generated
- processing time - when Spark received the data



Operations on streaming DataFrames

- The syntax is very similar than what we saw with static DFs

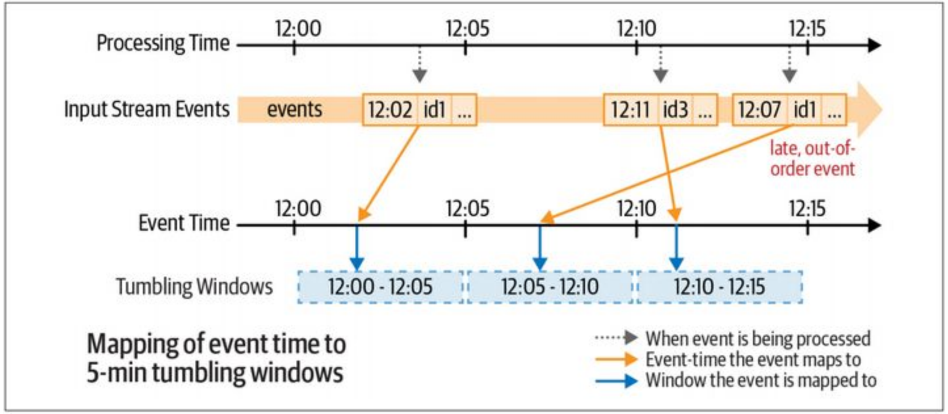
```
df = ...  # streaming DataFrame with IOT device data with schema { device: string, deviceType: string, signal: d
ouble, time: DateType }

# Select the devices which have signal more than 10
df.select("device").where("signal > 10")

# Running count of the number of updates for each device type
df.groupBy("deviceType").count()
```

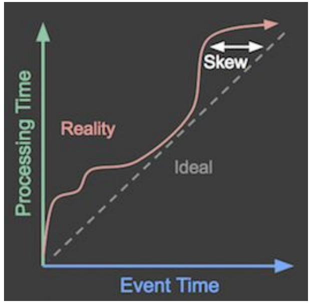
- SQL-like and RDD-like operations

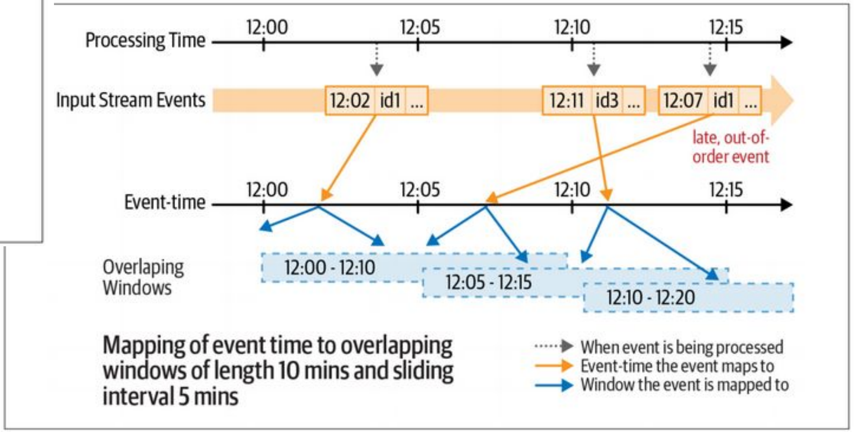
```
df.createOrReplaceTempView("updates")
spark.sql("select count(*) from updates")  # returns another streaming DF
```

- several operations are not supported: distinct, take(n), limit, chained aggregations, some types of joins...

Windows: tumbling, overlapping

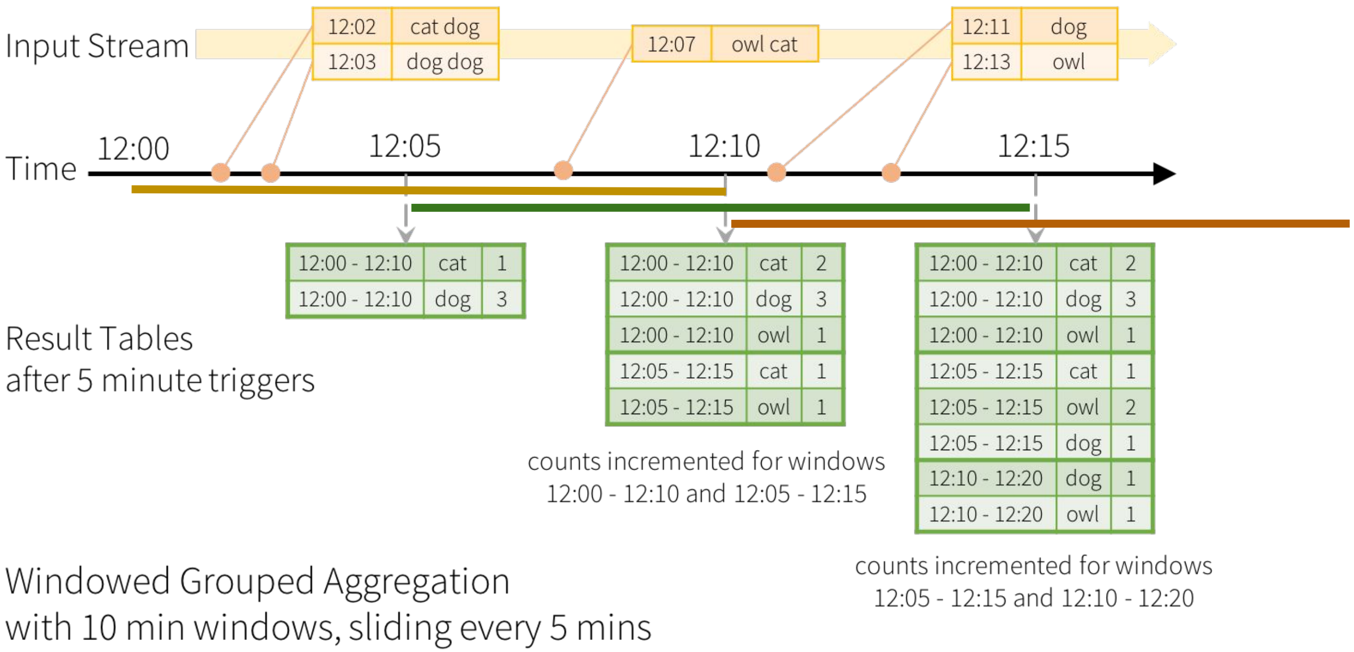






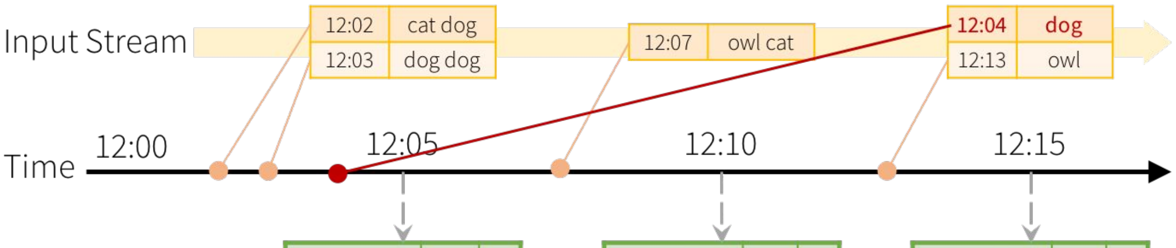
- Each event belongs to one or more windows

Aggregations over a sliding event-time window



Handling late data

late data that was generated
at 12:04 but arrived at 12:11



Result Tables
after 5 minute triggers

12:00 - 12:10	cat	1
12:00 - 12:10	dog	3

12:00 - 12:10	cat	2
12:00 - 12:10	dog	3
12:00 - 12:10	owl	1
12:05 - 12:15	cat	1
12:05 - 12:15	owl	1

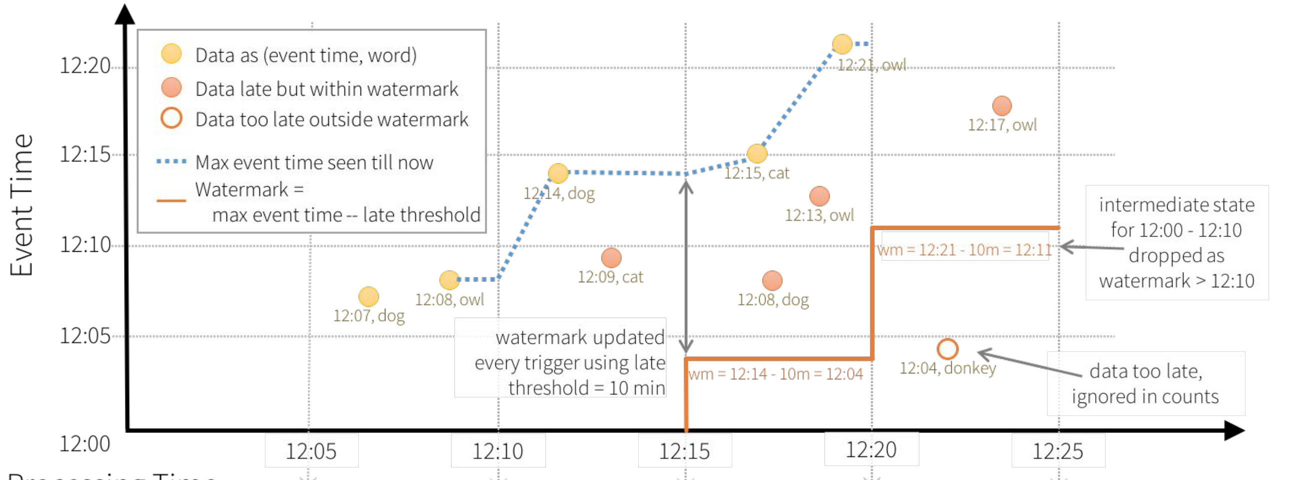
12:00 - 12:10	cat	2
12:00 - 12:10	**dog**	**4**
12:00 - 12:10	owl	1
12:05 - 12:15	cat	1
12:05 - 12:15	owl	2
12:10 - 12:20	owl	1

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

Late data handling in
Windowed Grouped Aggregation

How long will we wait?

Watermarking



Processing Time
with 5 min triggers

Result Tables after each trigger

12:10

12:00 - 12:10	owl	1
12:00 - 12:10	dog	1
12:05 - 12:15	owl	1
12:05 - 12:15	dog	1

12:15

12:00 - 12:10	owl	1
12:00 - 12:10	dog	1
12:00 - 12:10	cat	1
12:05 - 12:15	owl	1
12:05 - 12:15	dog	2
12:05 - 12:15	cat	1
12:10 - 12:20	dog	1

12:20

12:00 - 12:10	owl	1
12:00 - 12:10	dog	2
12:00 - 12:10	cat	1
12:05 - 12:15	owl	2
12:05 - 12:15	dog	3
12:05 - 12:15	cat	2
12:10 - 12:20	dog	1
12:10 - 12:20	cat	1
12:10 - 12:20	owl	1
...		

12:25

12:00 - 12:10	owl	1
12:00 - 12:10	dog	2
12:00 - 12:10	cat	1
12:05 - 12:15	owl	2
12:05 - 12:15	dog	3
12:05 - 12:15	cat	2
12:10 - 12:20	dog	1
12:10 - 12:20	cat	1
12:10 - 12:20	owl	2
...		

table *not* updated with too late data (12:04, donkey)

table updated with late data (12:17, owl)

purple rows are updated rows that are written to the sink as output

Watermarking in **Windowed Grouped Aggregation** with **Update Mode**

Vocabulary

- **Event-time** - a time when the event happens
- **Processing time** - a time when we receive the data of the event
- **Trigger** - how often we will collect the data
- **Window** - for how long will we collect the data
- **Watermark** - when the late data is discarded (how late is too late)