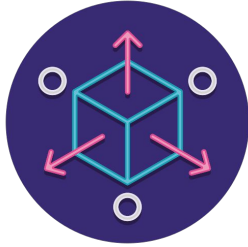


Structured Streaming



Structured Streaming



Scalable

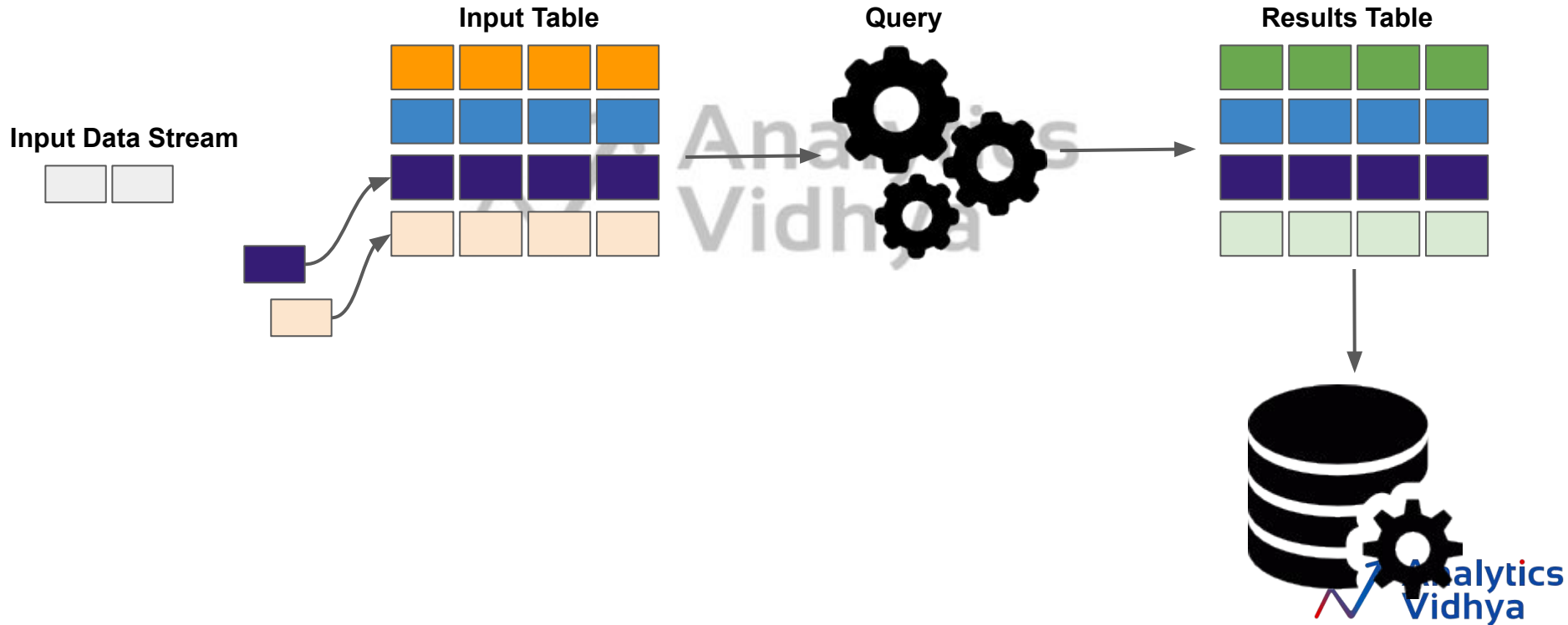


Fault Tolerant



Built on Spark SQL Engine

Structured Streaming



Structured Streaming

InputTable



Vidhya

Structured Streaming

Unbounded Table

--	--	--	--

--	--	--	--

--	--	--	--

Data stream as an Unbounded table

Structured Streaming

Unbounded Table

New data in the
data stream

=

--	--	--	--

--	--	--	--

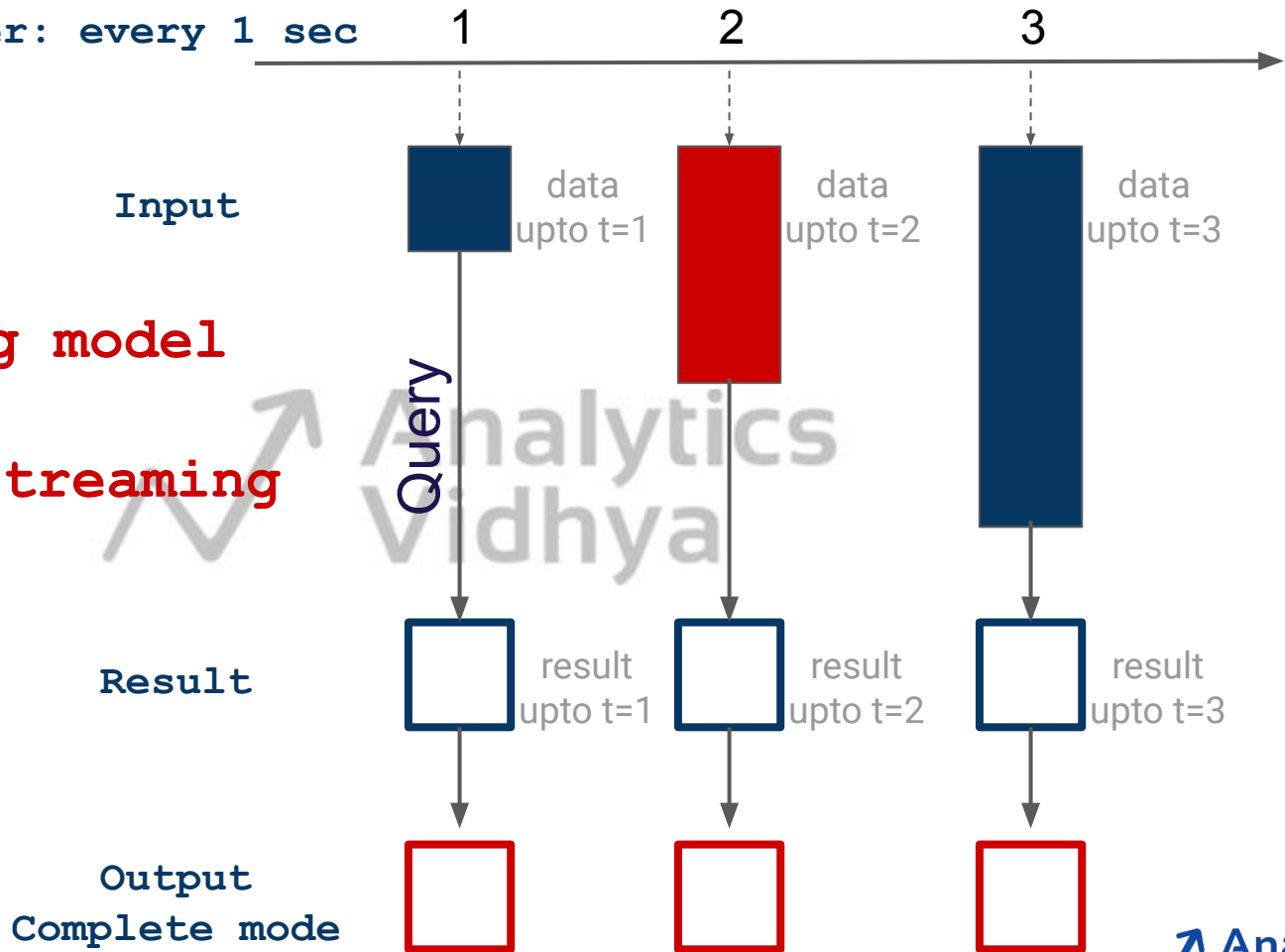
--	--	--	--

New rows
appended to the
unbounded table

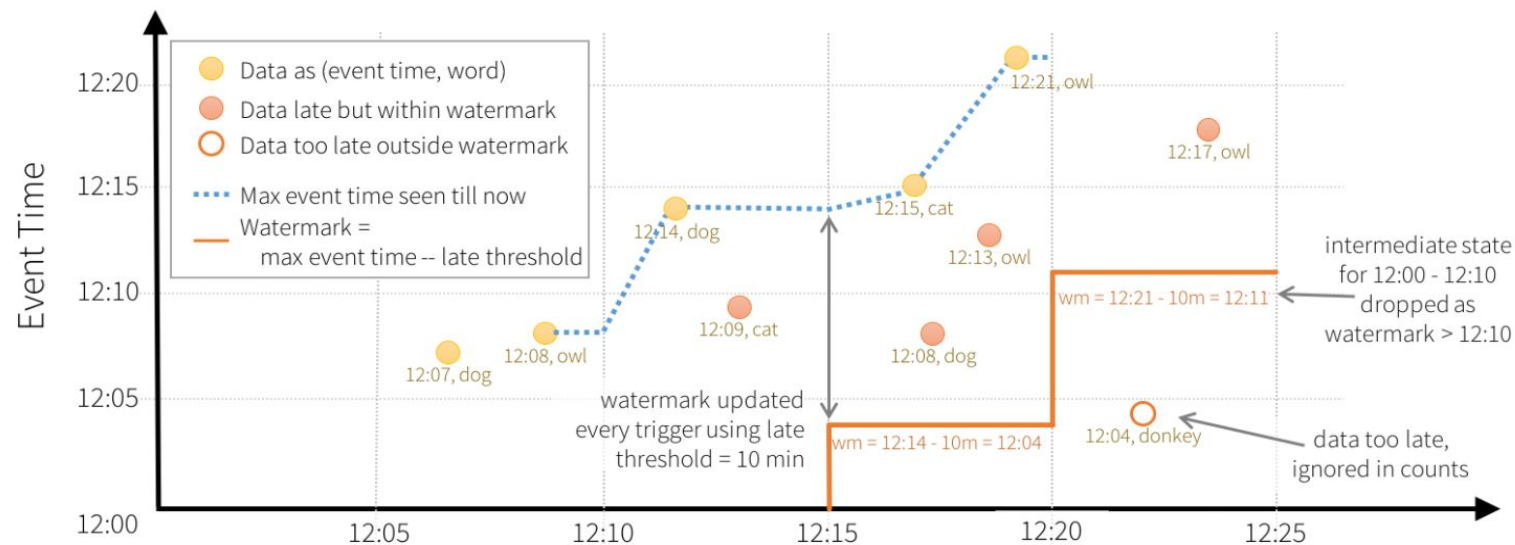
Data stream as an Unbounded table

Programming model of Structured Streaming

Trigger: every 1 sec



 Thank You
Analytics
Vidhya



Processing Time with 5 min triggers

Result Tables after each trigger

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

12:00 - 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

12:00 - 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
...		

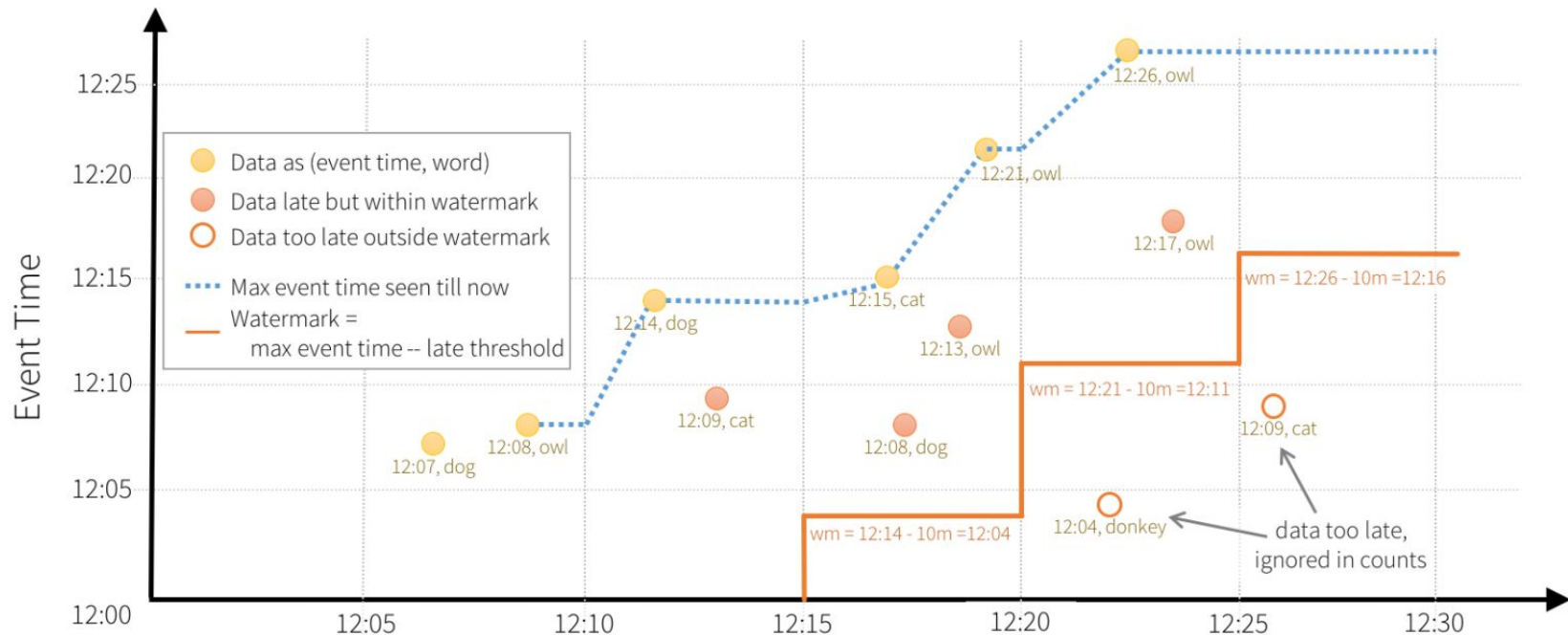
12:00 - 12:10	owl	1
12:00 - 12:10	dog	2
12:00 - 12:10	cat	1
12:05 - 12:15	owl	2
12:05 - 12:15	dog	3
12:05 - 12:15	cat	2
12:10 - 12:20	dog	1
12:10 - 12:20	cat	1
12:10 - 12:20	owl	2
...		

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



partial counts for window 12:00 - 12:10 maintained as internal state while waiting for late data, so not yet added to result table

final counts for 12:00 - 12:10 added to table when watermark > 12:10, late data counted, and intermediate state for window dropped

12:00 - 12:10	owl	1
12:00 - 12:10	cat	1
12:00 - 12:10	dog	2

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

Result Tables after each trigger

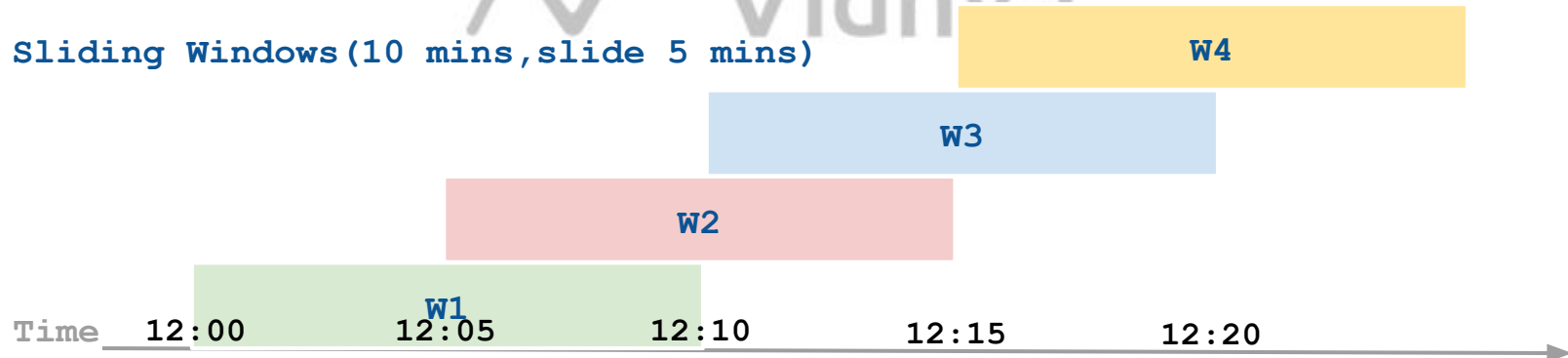
Watermarking in Windowed Grouped Aggregation with Append Mode

Types of Time Windows

Tumbling Windows (5 mins)



Sliding Windows (10 mins, slide 5 mins)



Types of Time Windows

Session Windows (gap duration 5 mins)

