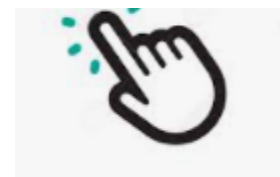


Spark Streaming Programming Guide

- [Overview](#)
- [A Quick Example](#)
- [Basic Concepts](#)



Overview

Spark Streaming is an extension of the core Spark API that enables scalable, high-throughput, fault-tolerant stream processing of live data streams.





Structured Streaming Programming Guide

- [Overview](#)
- [Quick Example](#)
- [Programming Model](#)



- Structured Streaming is a scalable and fault-tolerant stream processing engine built on the Spark SQL engine.
- You can express your streaming computation the same way you would express a batch computation on static data.
- The Spark SQL engine will take care of running it incrementally and continuously and updating the final result as streaming data continues to arrive.

Data stream

Unbounded Table



--	--	--

--	--	--

--	--	--

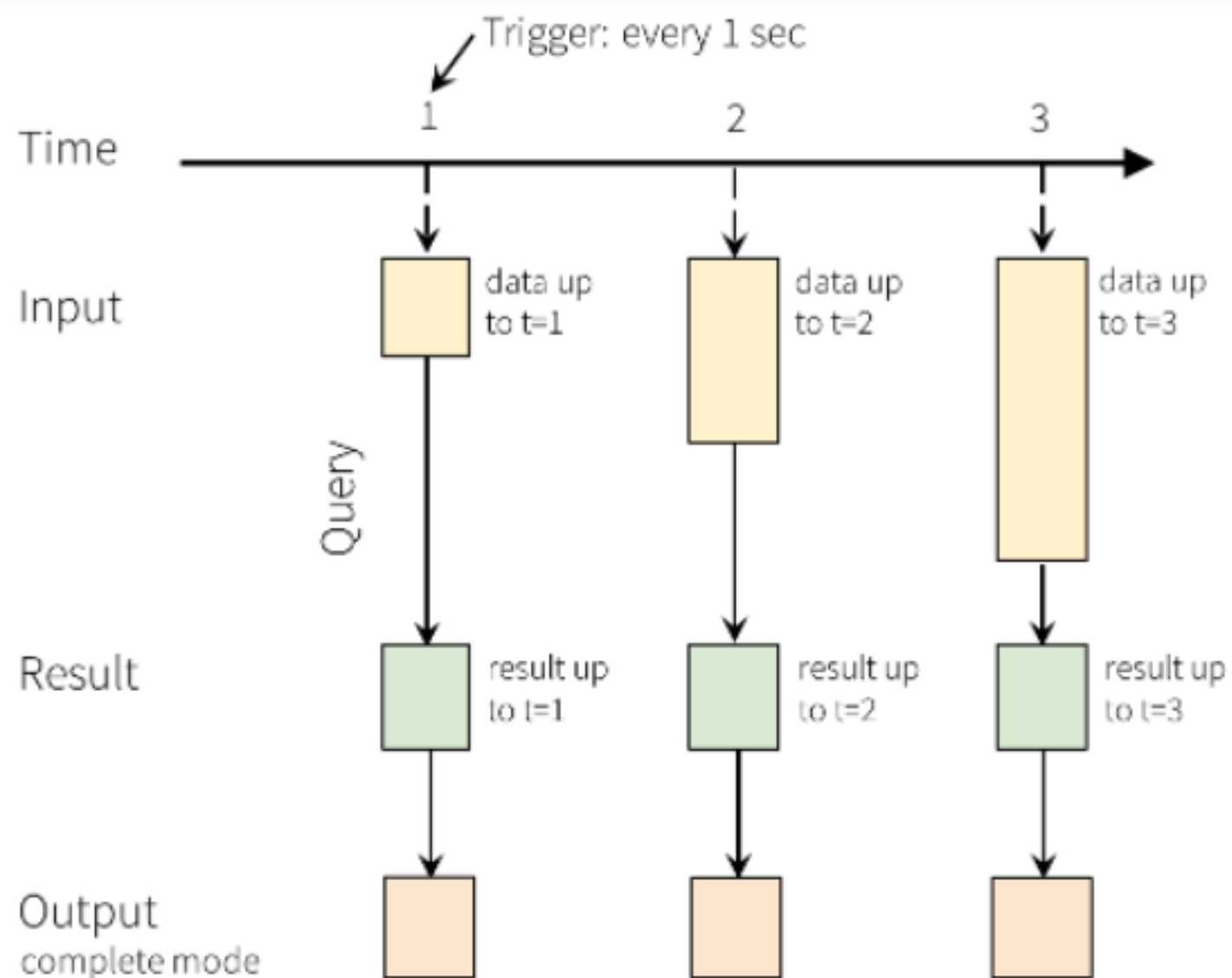
--	--	--

new data in the
data stream

=

new rows appended
to a unbounded table

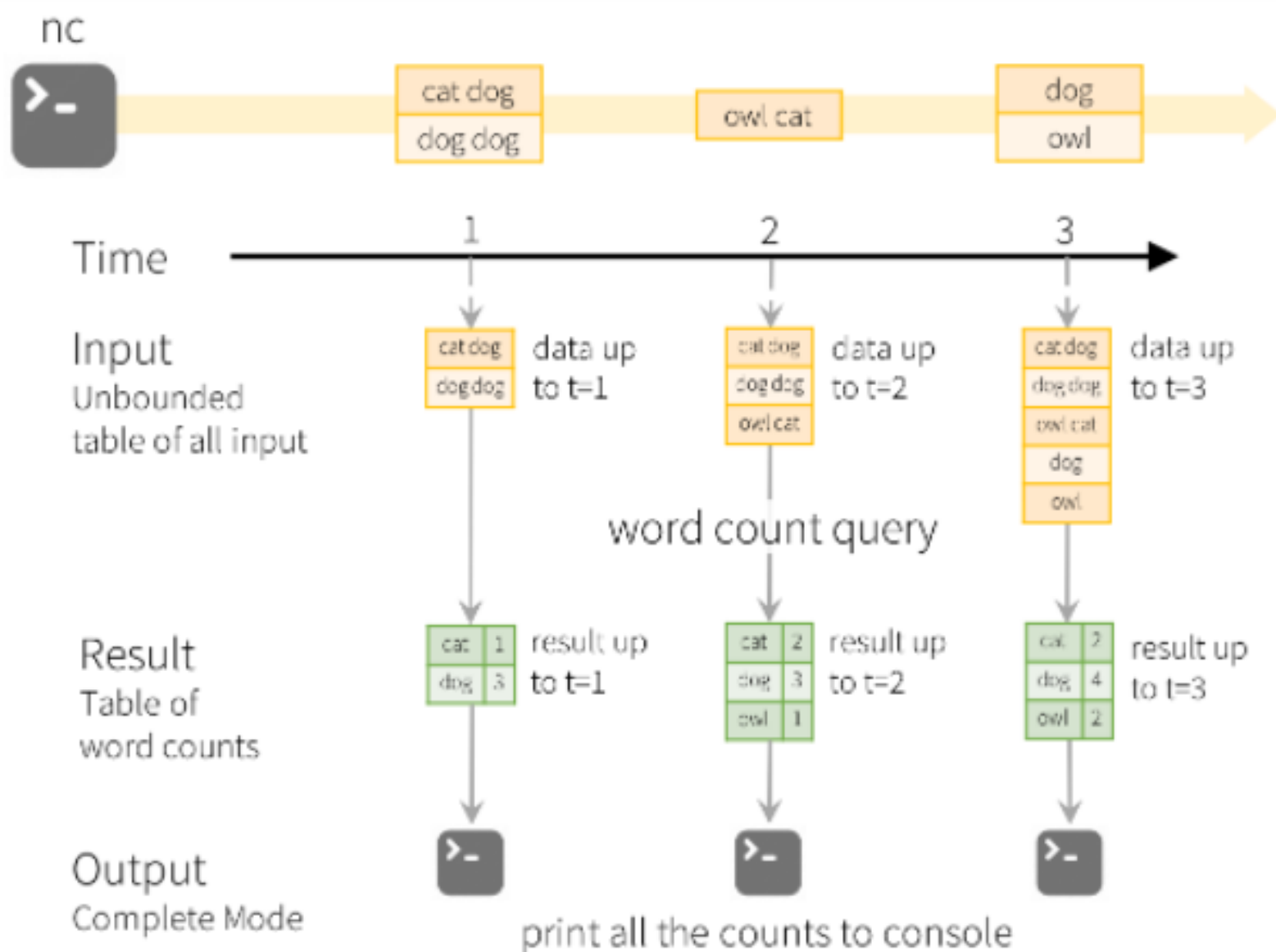
Data stream as an unbounded table



Programming Model for Structured Streaming

The "Output" is defined as what gets written out to the external storage. The output can be defined in a different mode:

- *Complete Mode* - The entire updated Result Table will be written to the external storage. It is up to the storage connector to decide how to handle writing of the entire table.
- *Append Mode* - Only the new rows appended in the Result Table since the last trigger will be written to the external storage. This is applicable only on the queries where existing rows in the Result Table are not expected to change.
- *Update Mode* - Only the rows that were updated in the Result Table since the last trigger will be written to the external storage (available since Spark 2.1.1). Note that this is different from the Complete Mode in that this mode only outputs the rows that have changed since the last trigger. If the query doesn't contain aggregations, it will be equivalent to Append mode.



Model of the Quick Example

