Working with Streaming Joins

Janani Ravi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Join operations on streams

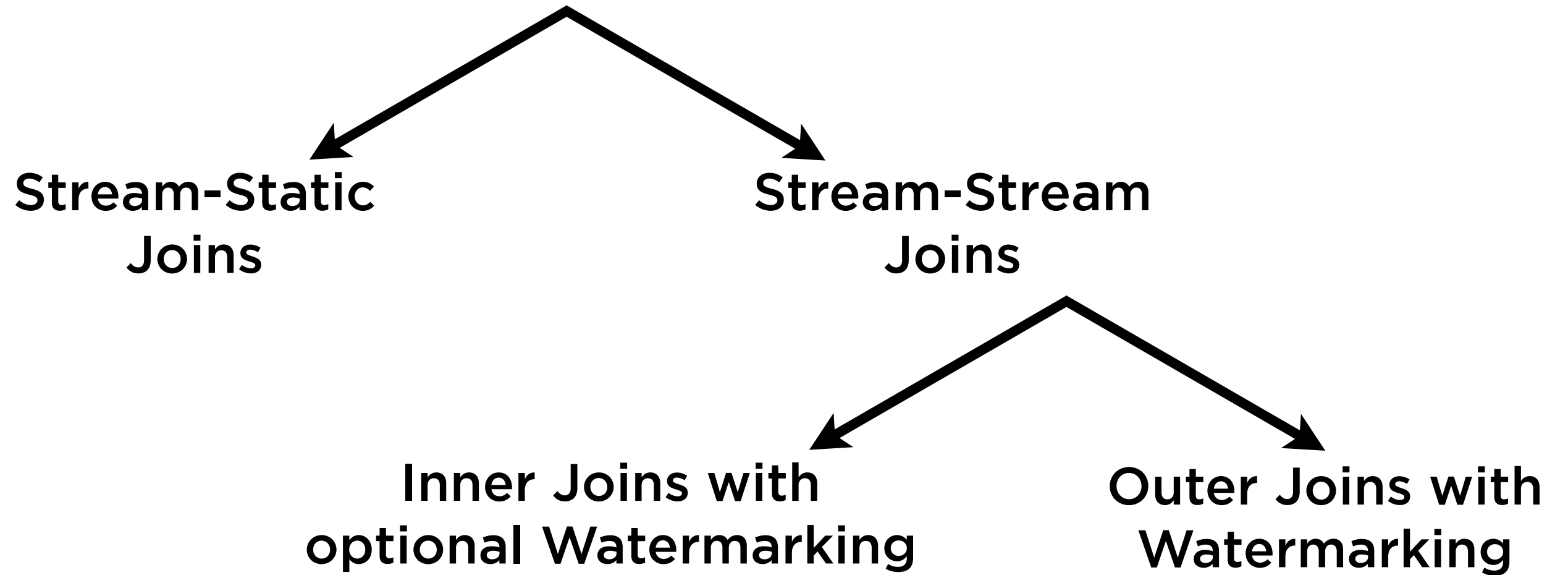**Stream-static joins**

Stream-stream joins
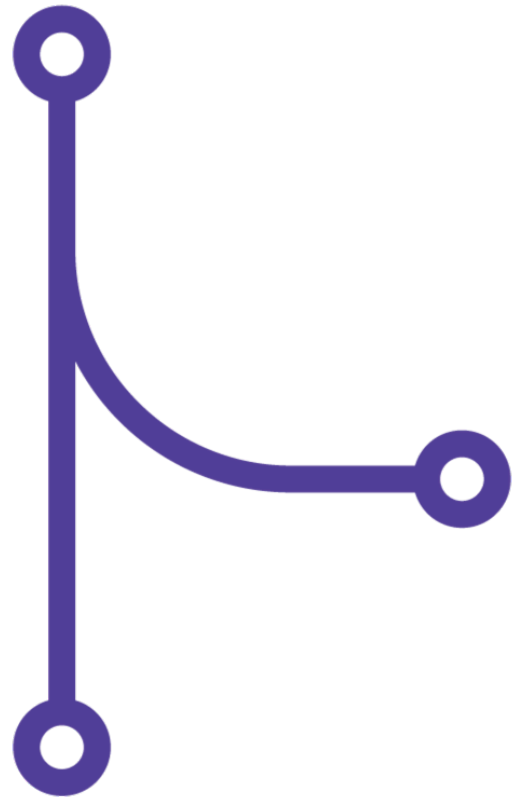
Streaming Joins

Joins in Structured Streaming

Stream-static joins

Stream-stream joins
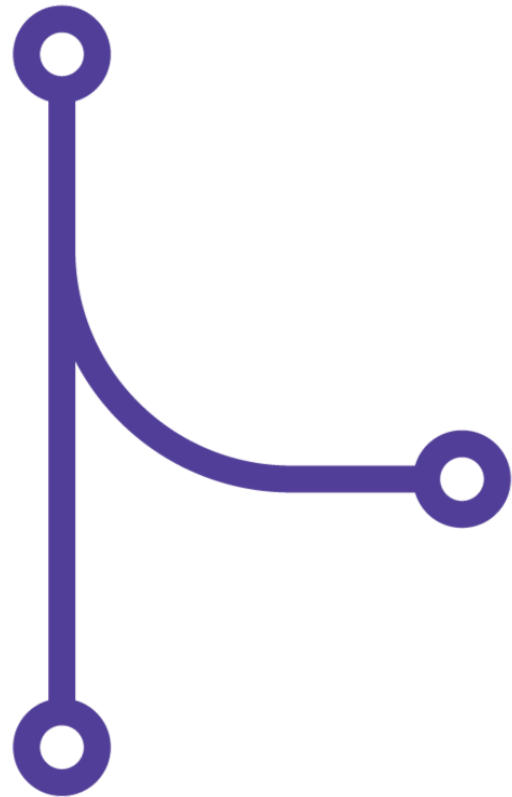
Joins in Structured Streaming

Stream-Static Joins

Stream-Stream Joins

Inner Joins with optional Watermarking

Outer Joins with Watermarking

Stream-Static Joins

Join static DataFrame with streaming DataFrame

Result of the streaming join is generated incrementally

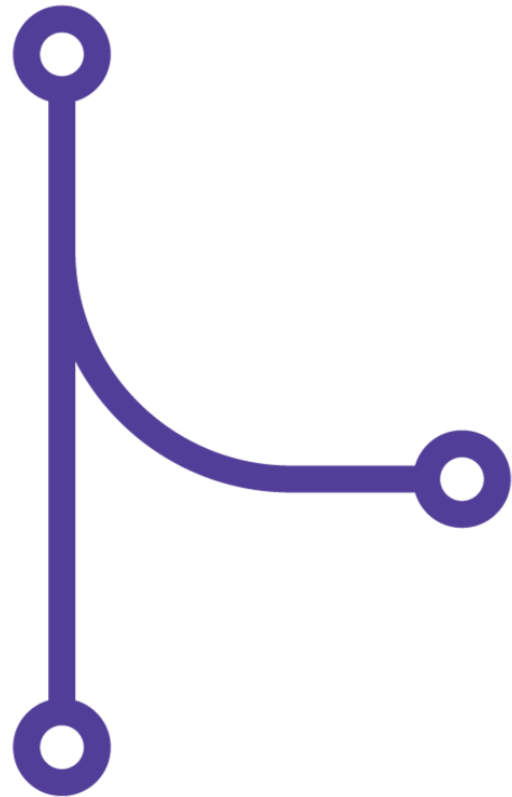Stream-static and static-stream joins are not stateful

Stream-Static Joins

First operand streaming, second operand static DataFrame

Supported types: Inner, Left outer

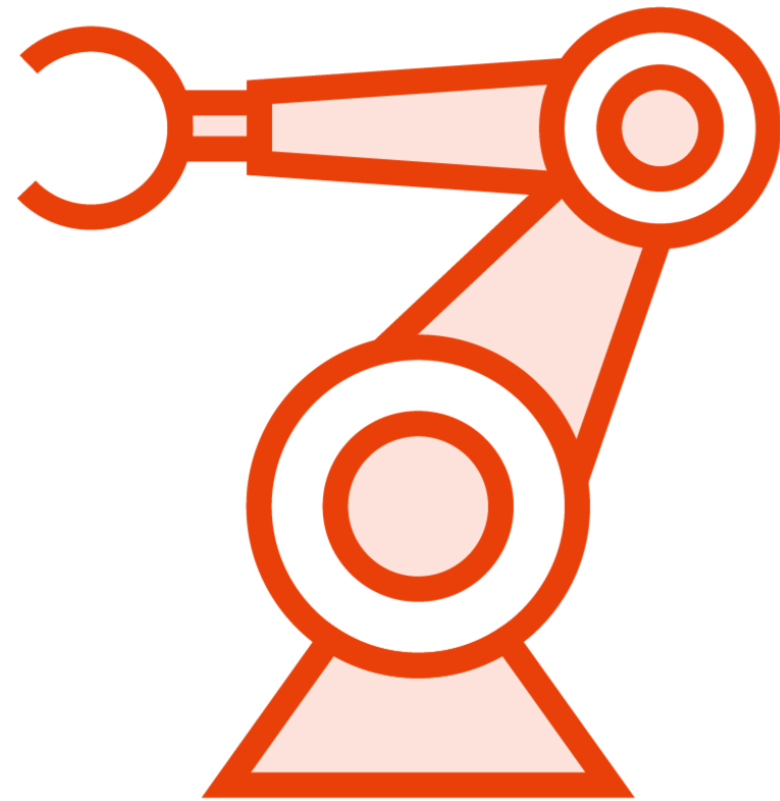**Unsupported types: Right and Full Outer**

Static-Stream Joins

First operand static, second operand streaming DataFrame

Supported types: Inner, Right outer

Unsupported types: Left and Full Outer

Stream-Stream Joins

Both operands are streaming DataFrames

Stream-stream joins are stateful

- Buffer past input as streaming state

- Allows automatic handling of late, out-of-order data

- Use watermarks to limit state

Stream-Stream Joins

Fully supported: Inner Joins

Conditionally supported: Left and Right Outer Joins

Unsupported: Full Outer Joins

Demo

Joining batch data with streaming data

Demo

Joining two streaming datasets

Summary

Join operations on streams

Stream-static joins

Stream-stream joins

Up Next:
Managing and Monitoring Streaming Queries