

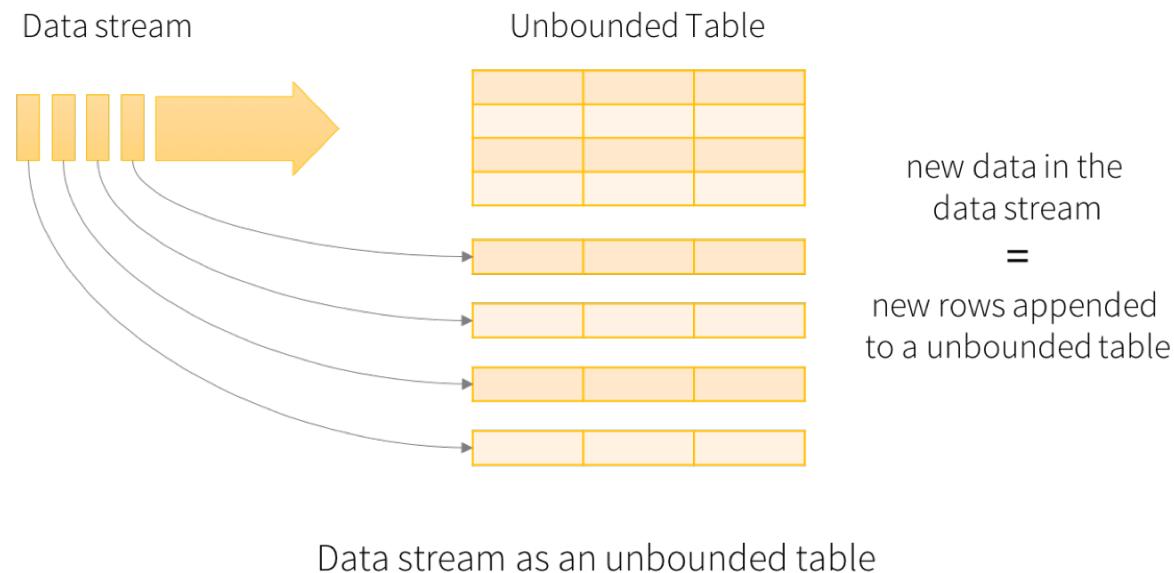


Whirlpools in the Stream

Misadventures in Structured Streaming

Jayesh Lalwani,
Lead Software Engineer, Quantum, Capital One

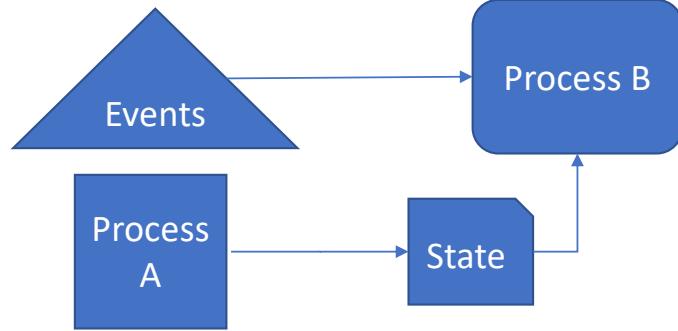
What is Structured Streaming



Dunderheaded Data stores!



Look Ma! No lookups!



- Spark allows you to join Streaming data with Batch Data
- BUT....
 - Spark doesn't support pushdown of join predicates. This means that the State is loaded into Spark memory for every microbatch
 - Caching a Batch dataframe “freezes” it

Solutions

- Cache dataframe & Restart the Streaming application
- Cache dataframe & Restart the Streaming query
- Cache the data outside of Spark
- Lookup batch data inside a map/flatmap operation
- Stream updates to Process B and do stream to stream join (2.3+ only)
 - works for limited use cases

Data store? We don't need no stinkin' data stores

- Outputs supported Batch Mode
 - JSON
 - CSV
 - Parquet
 - Orc
 - Text
 - JDBC
- Outputs supported Streaming Mode
 - JSON
 - CSV
 - Parquet
 - Text

WHERE IS JDBC?

My implementation of JDBC sink

- Features:
 - Can write streaming data to JDBC data stores
 - Uses the same code that is used by Batch JDBC sinks
 - Modes: Overwrite, Append
 - Supports *Atleast Once out of the box*
 - Supports *Exactly Once* with some configuration
- Fork: <https://github.com/GaalDornick/spark>
- Implementation:
<https://github.com/GaalDornick/spark/blob/master/sql/core/src/main/scala/org/apache/spark/sql/execution/streaming/JdbcSink.scala>
- Unit test:
<https://github.com/GaalDornick/spark/blob/master/sql/core/src/test/scala/org/apache/spark/sql/execution/streaming/JDBCSinkSuite.scala>

Clumsy footed Checkpoints



Missing the point with Checkpoints

- Checkpoint stores state
- Allows Streaming application to restart on failure
- Checkpoint should be stored on a location that is
 - Resilient to failure
 - Shared between executors and drivers (read-write many)
 - Immediately consistent
- S3 IS NOT IMMEDIATELY CONSISTENT

Solution

- NFS
- We use EFS on AWS
- On Kubernetes, you need a PV that supports RWX access mode
 - NFS
 - Ceph
 - GlusterFS

Checkpoints had a great fall

- <https://issues.apache.org/jira/browse/SPARK-21696>
- Leads to corrupt checkpoints.
- Solved by deleting the checkpoint once a day
- This means we lose data
- Fixed in 2.2.1

Anthropophagic Aggregations!



Aggregations? What Aggregations?

- Structured streaming allows you to aggregate a dataframe
 - But you cannot aggregate an aggregated data frame
- PARTITION BY clause is supported only for time windows

Solutions

- If possible, aggregate batch data before joining with stream
- Implement aggregations without Spark SQL using
groupBy..flatMapGroups

Jayesh Lalwani
Lead Software Engineer
Team Heartbeat
Jayesh.Lalwani@capitalone.com