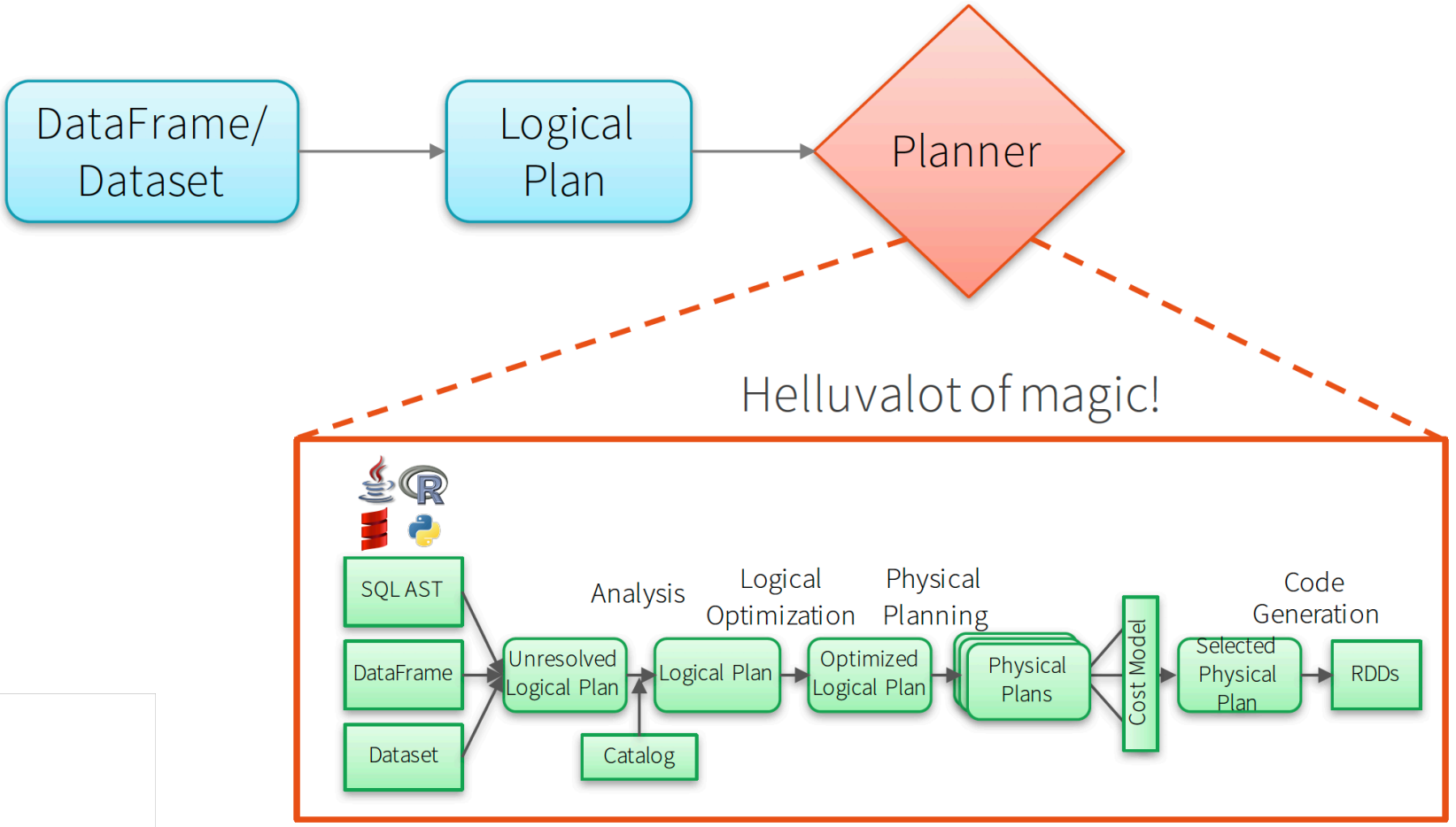
Structured Streaming

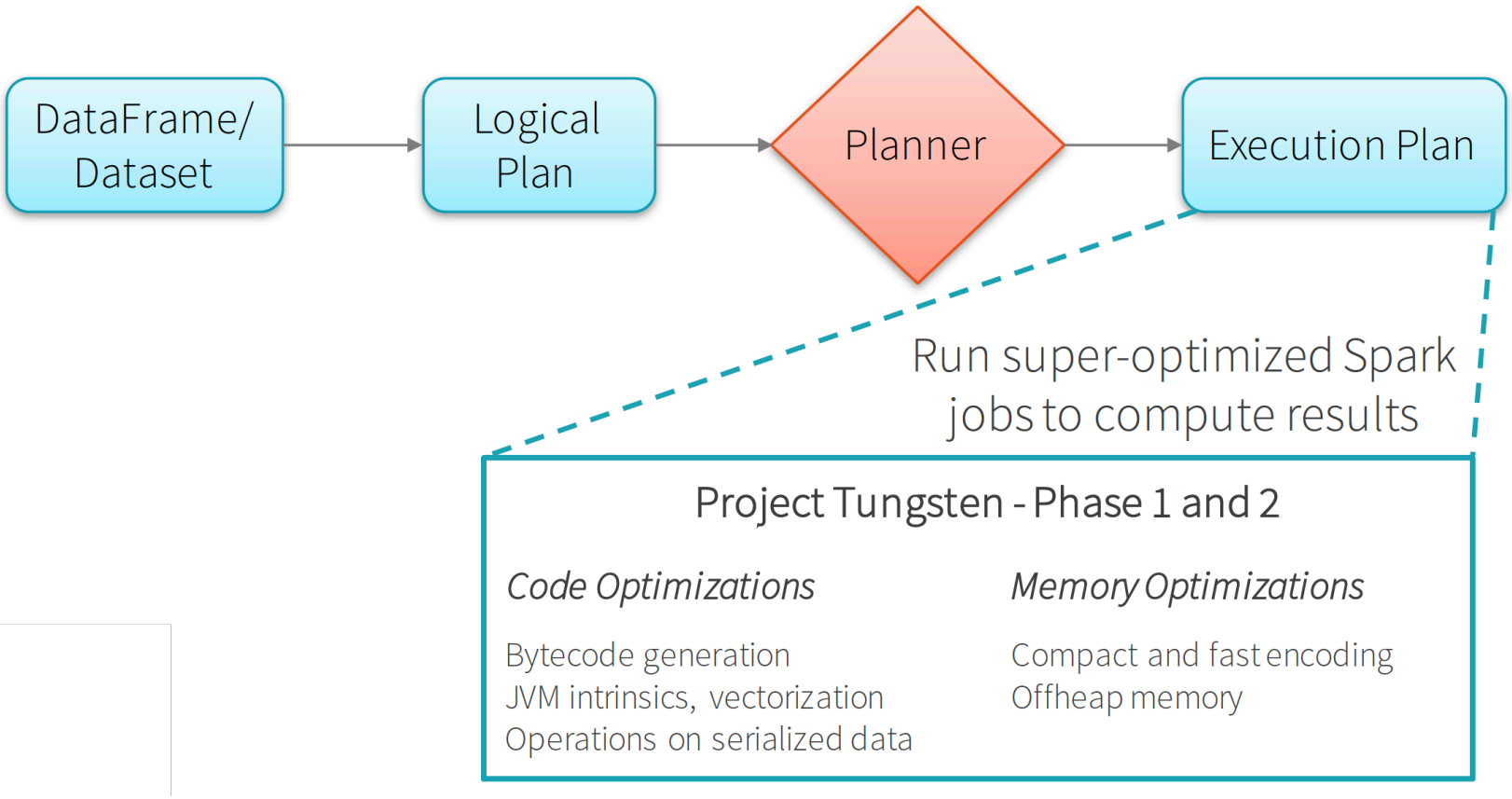
- High-level streaming API built on Datasets/DataFrames
  - Event time, windowing, sessions, sources & sinks
  - End-to-end exactly once semantics
- Unifies streaming, interactive and batch queries
  - Aggregate data in a stream, then serve using JDBC
  - Add, remove, change queries at runtime
  - Build and apply ML models

Internal execution

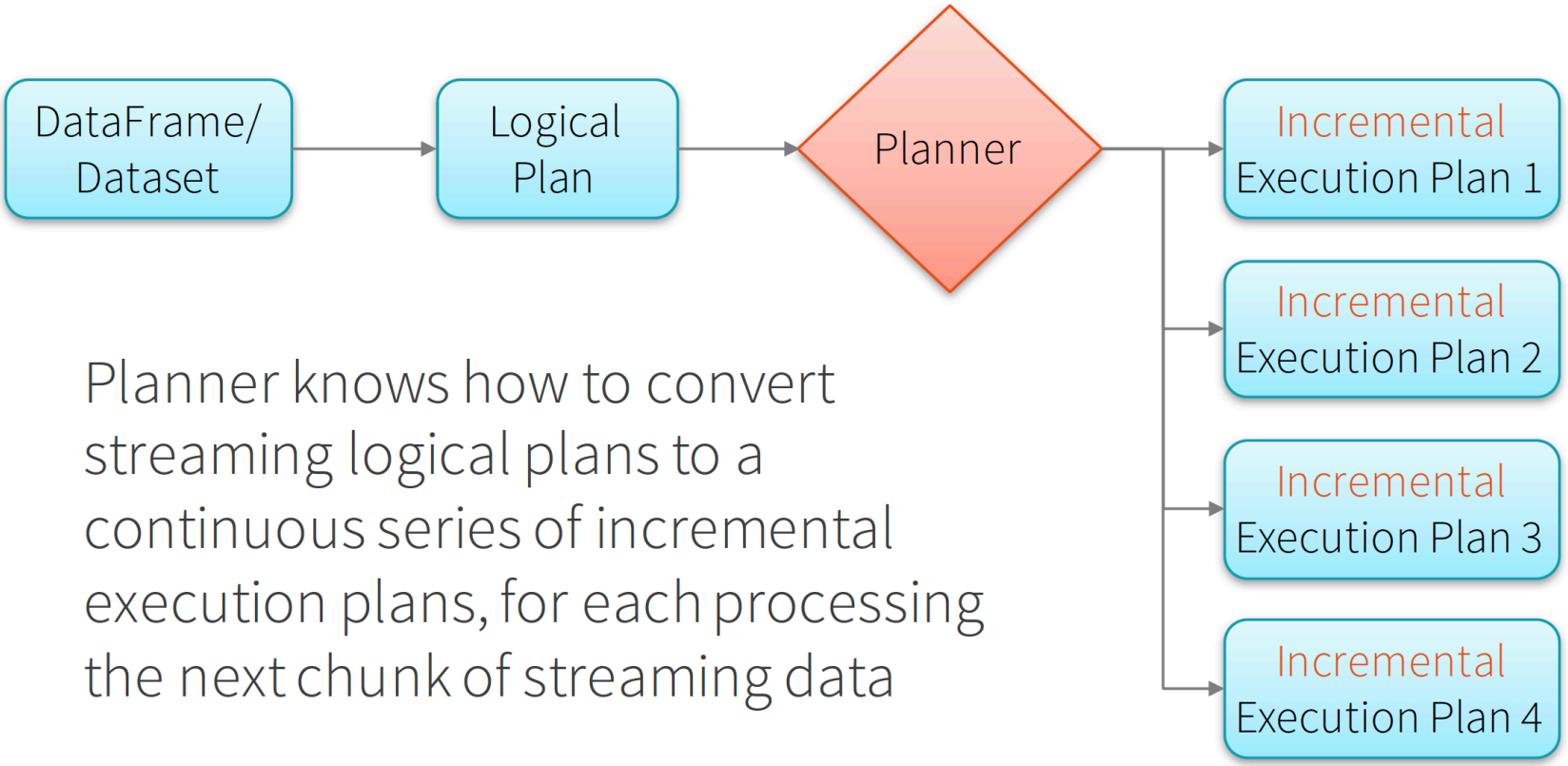
Batch Execution on Spark SQL



Batch Execution on Spark SQL

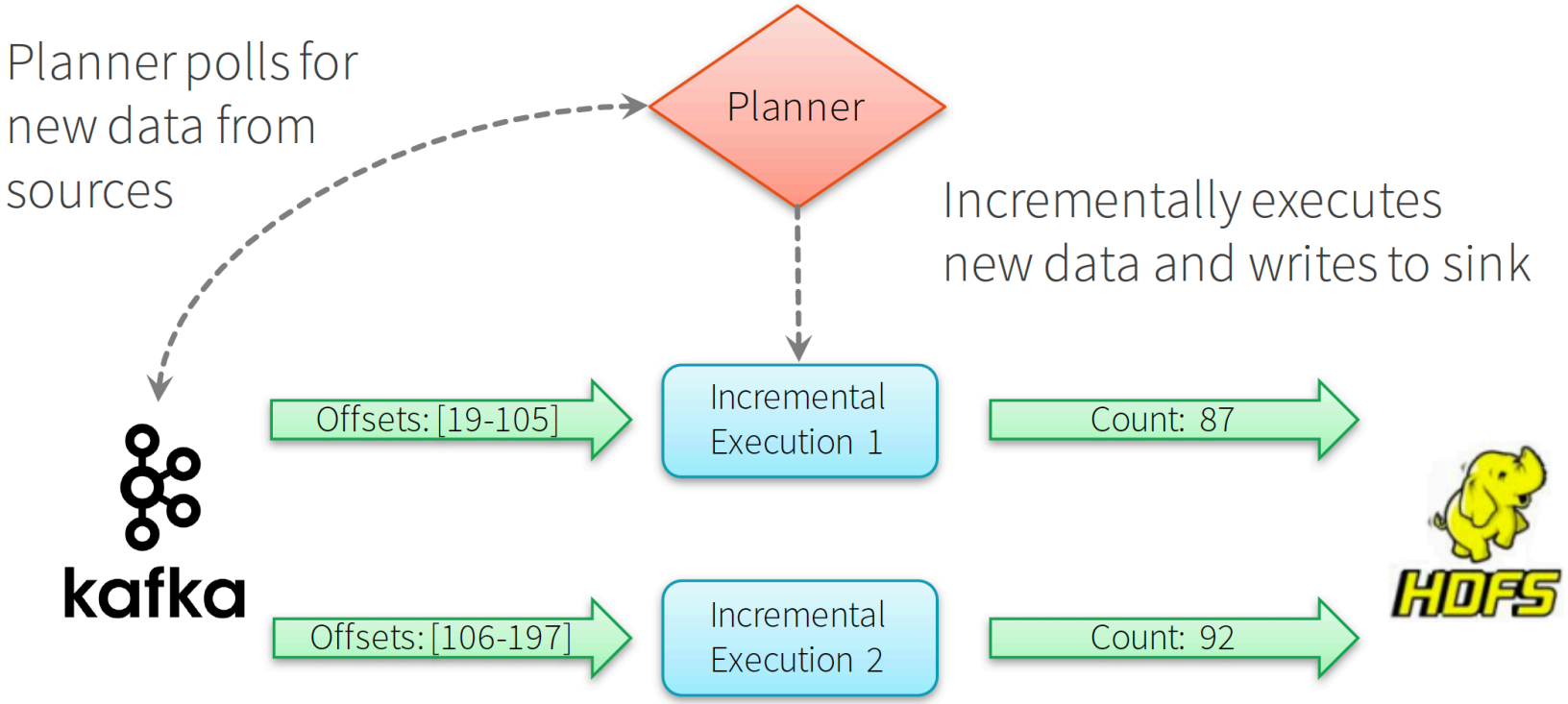


Continuous Incremental Execution



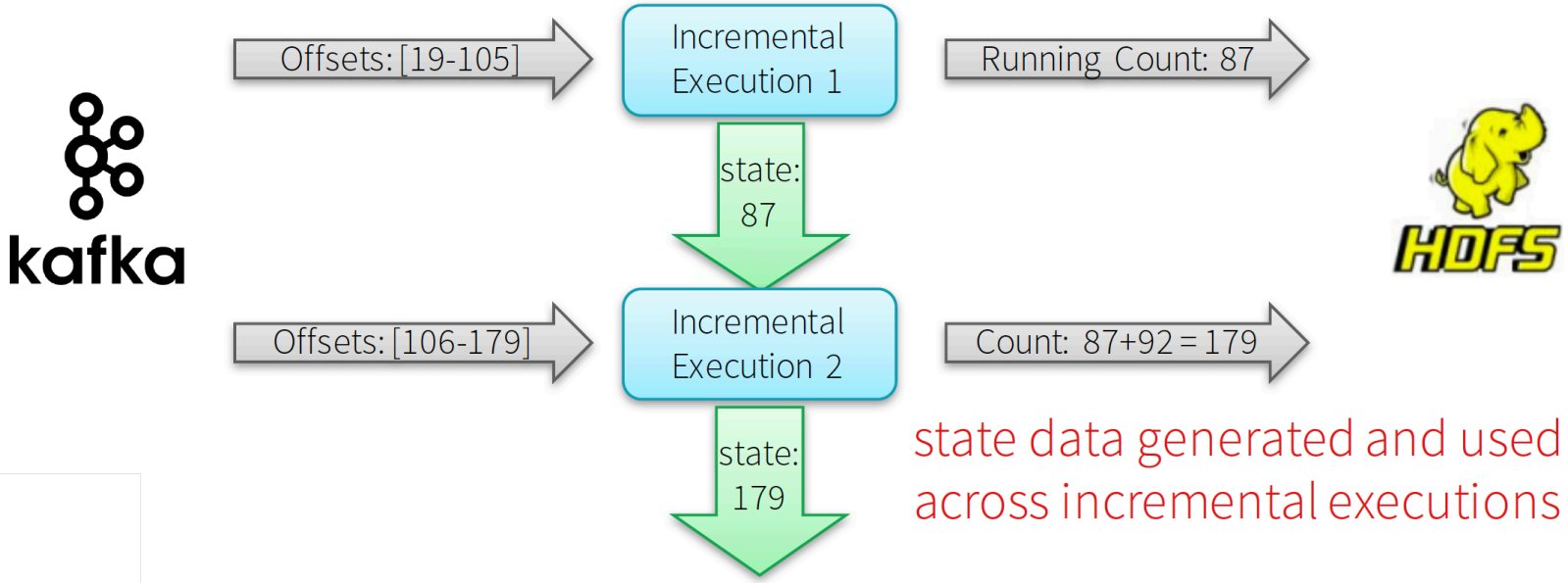
Planner knows how to convert streaming logical plans to a continuous series of incremental execution plans, for each processing the next chunk of streaming data

Continuous Incremental Execution



Continuous Aggregations

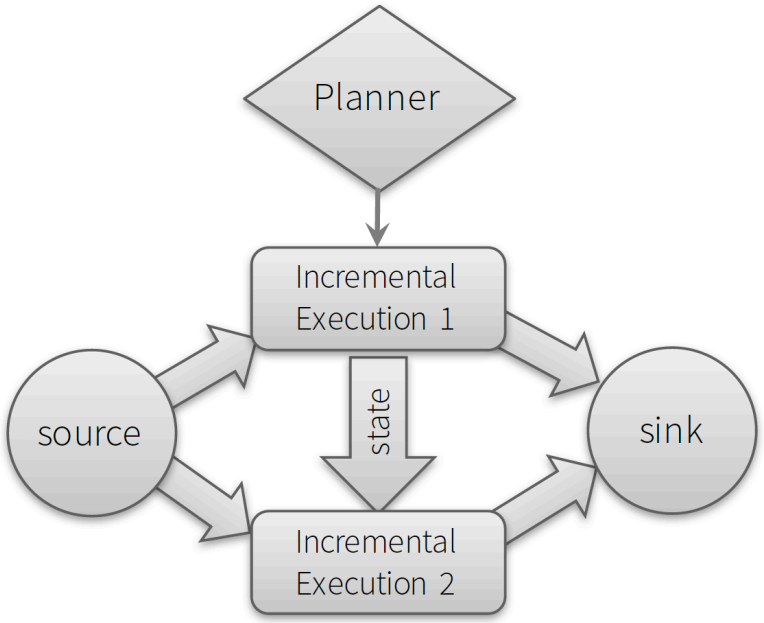
Maintain running aggregate as in-memory state backed by WAL in file system for fault-tolerance



state data generated and used across incremental executions

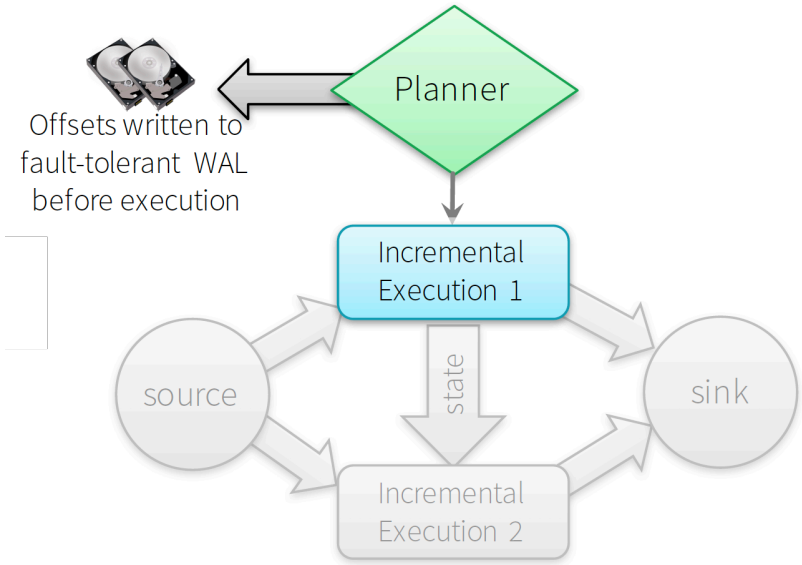
Fault-tolerance

- All data and metadata in the system needs to be recoverable / replayable



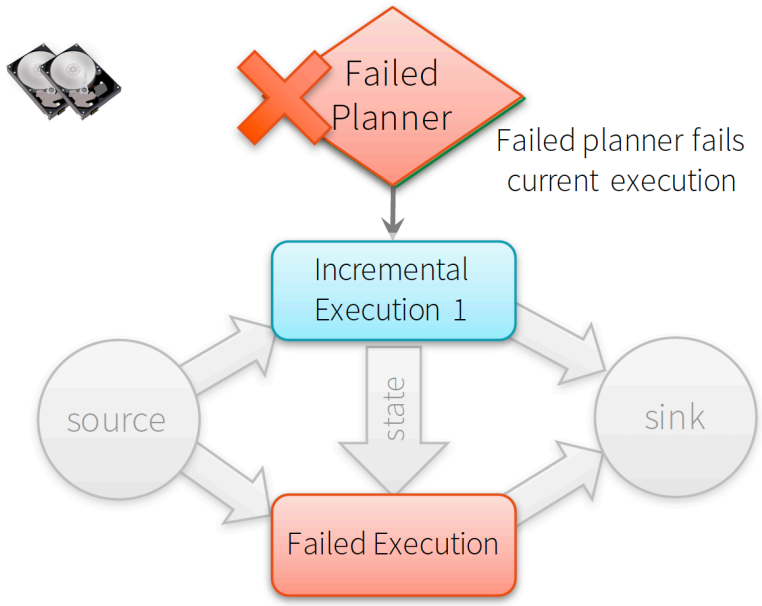
Fault-tolerant Planner

- Tracks offsets by writing the offset range of each execution to a write ahead log (WAL) in HDFS



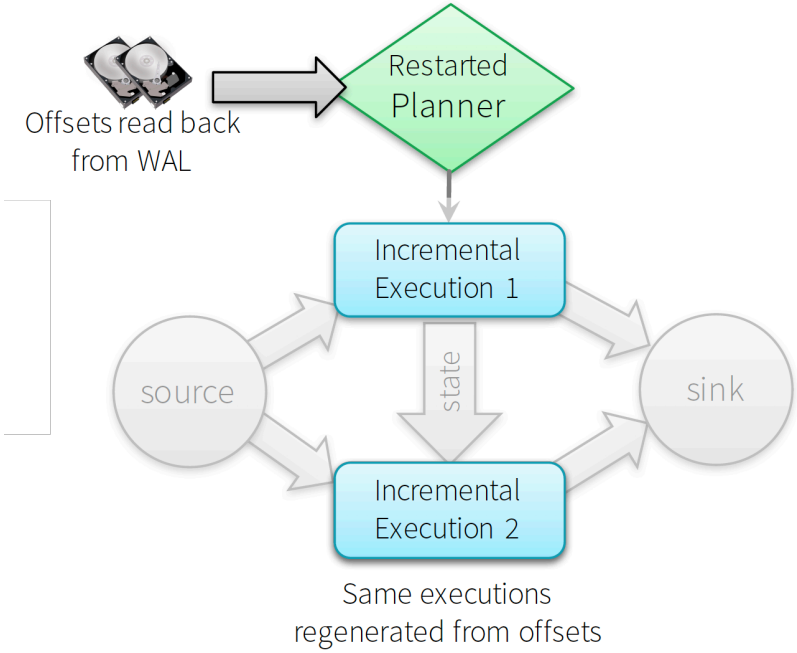
Fault-tolerant Planner

- Tracks offsets by writing the offset range of each execution to a write ahead log (WAL) in HDFS



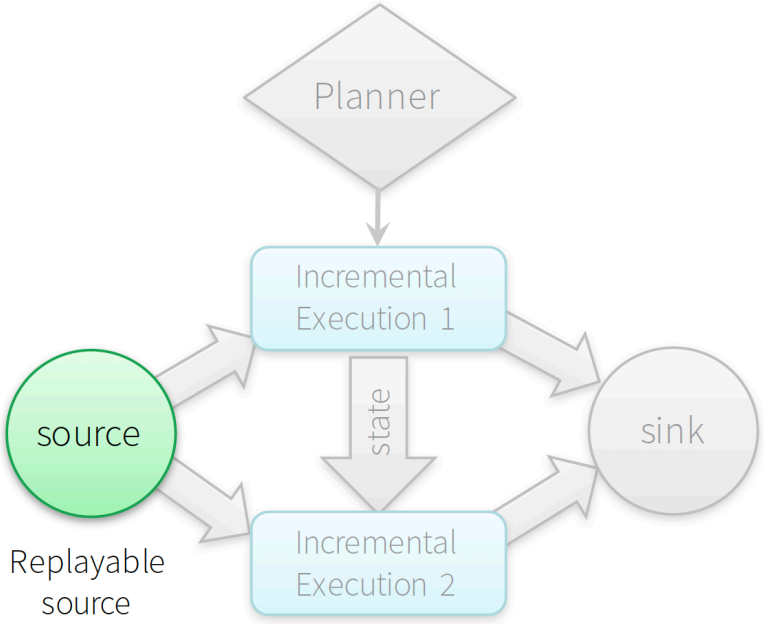
Fault-tolerant Planner

- Tracks offsets by writing the offset range of each execution to a write ahead log (WAL) in HDFS
- Reads log to recover from failures, and re-execute exact range of offsets



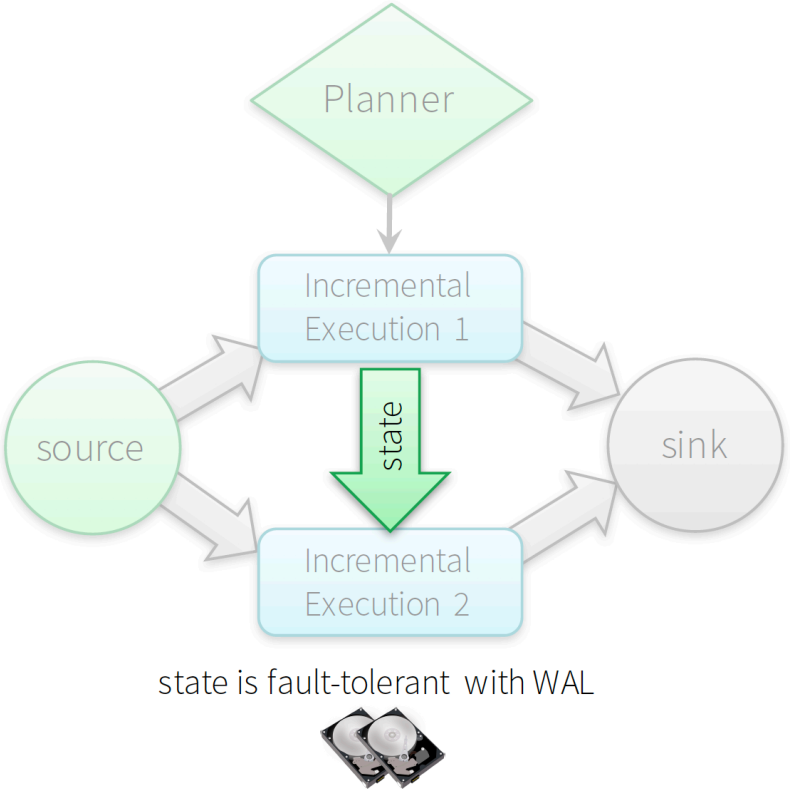
Fault-tolerant Sources

- Structured streaming sources are by design replayable (e.g. Kafka, Kinesis, files) and generate the exactly same data given offsets recovered by planner



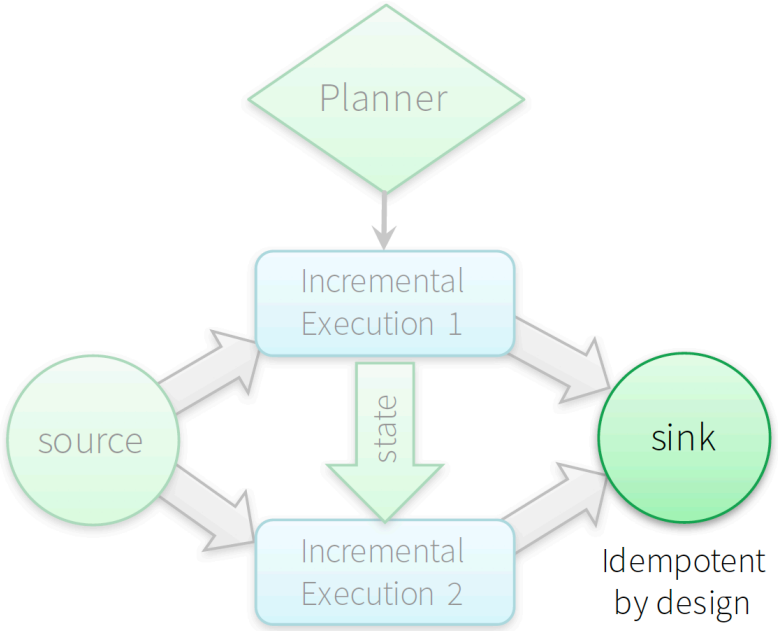
Fault-tolerant State

- Intermediate "state data" is a maintained in versioned, keyvalue maps in Spark workers, backed by HDFS
- Planner makes sure "correct version" of state used to reexecute after failure



Fault-tolerant Sink

- Sink are by design idempotent (deterministic), and handles re-executions to avoid double committing the output



Fault-tolerance

offset tracking in WAL

+

state management

+

fault-tolerant sources and sinks

=

end-to-end

exactly-once

guarantees

Structured streaming

Fast, fault-tolerant, exactly-once

stateful stream processing

without having to reason about streaming