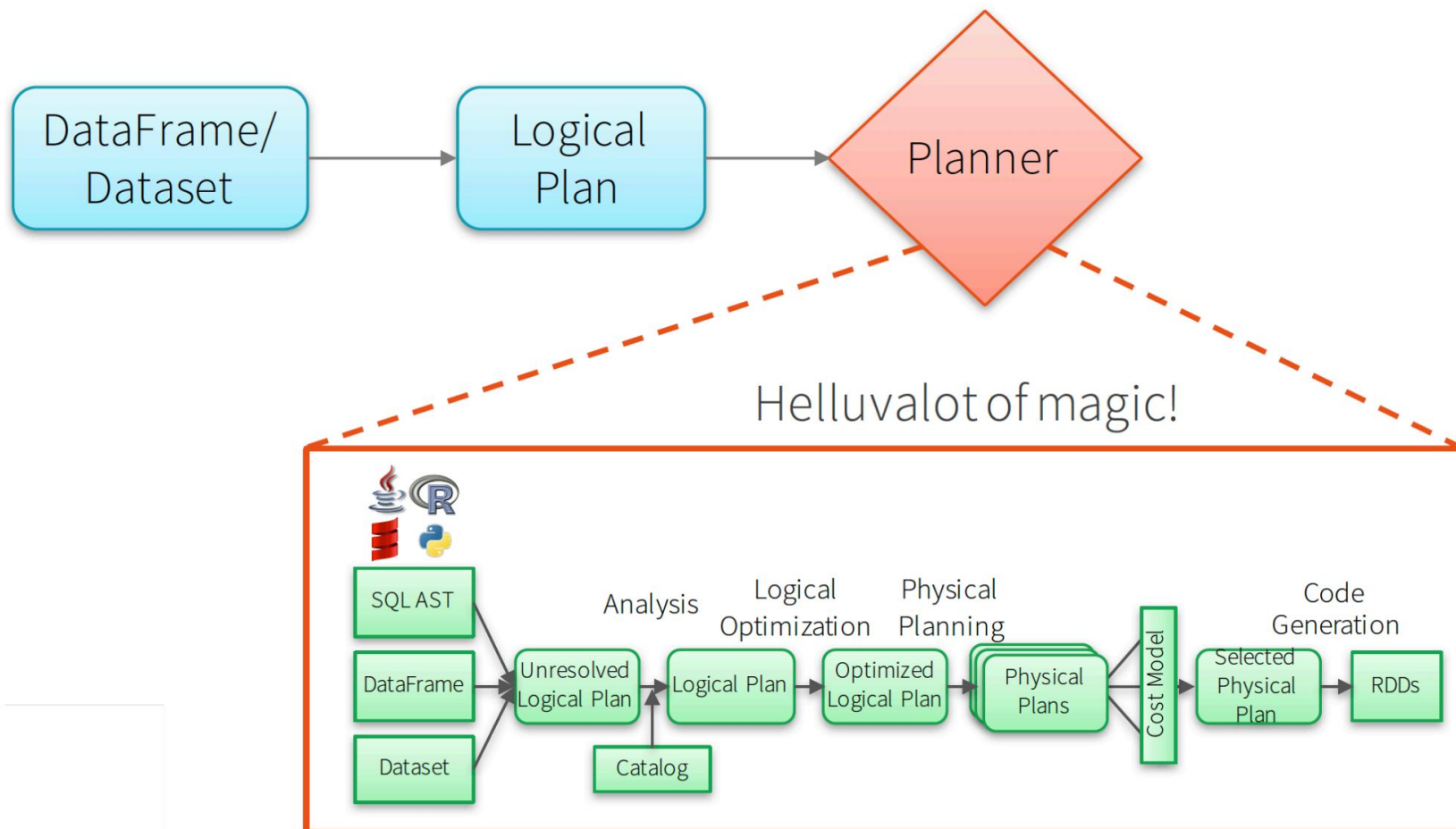


Phát trực tuyến có cấu trúc

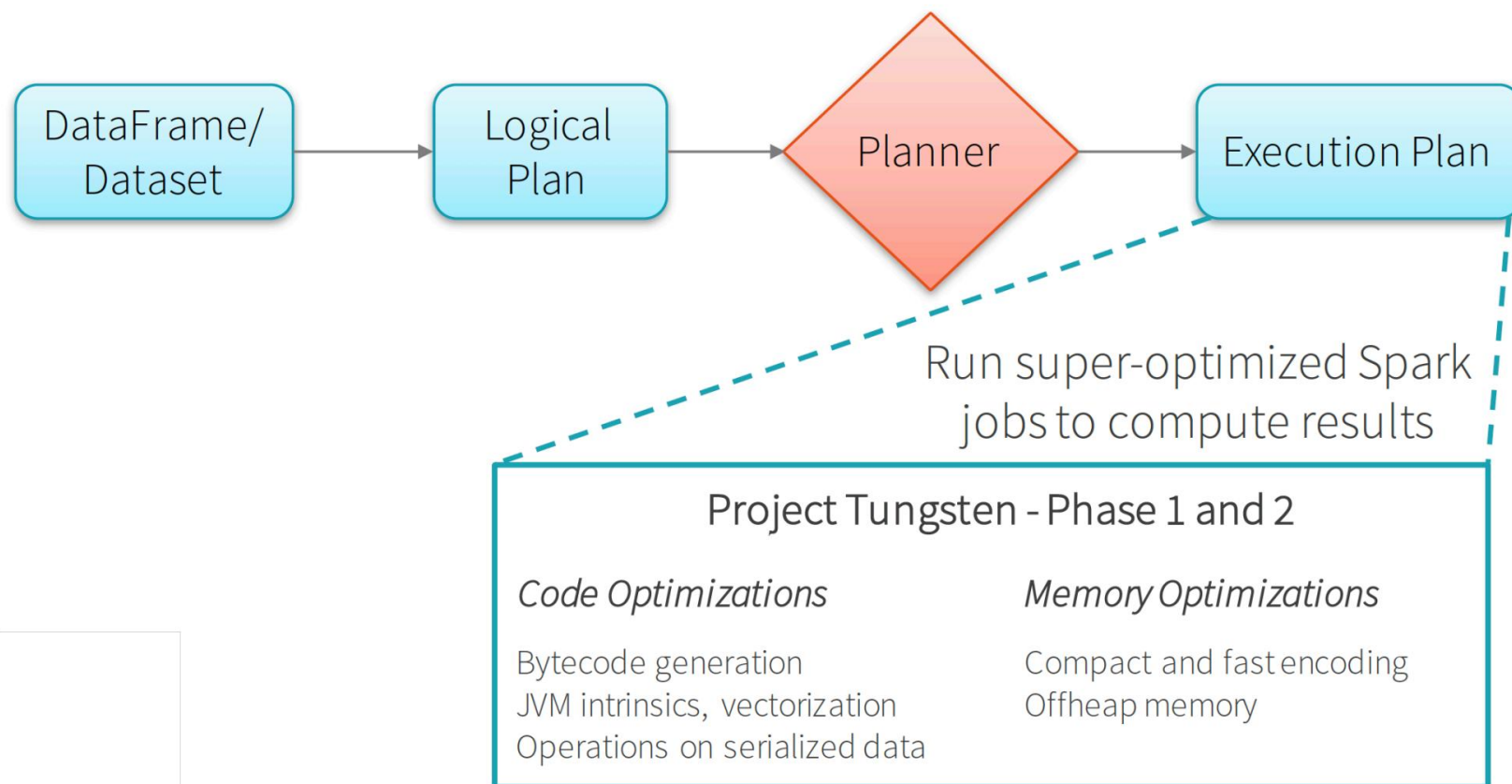
- API phát trực tuyến cấp cao đư ợc xây dựng trên Bộ dữ liệu/Khung dữ liệu
 - Thời gian sự kiện, cửa sổ, phiên, nguồn và bộ thu
 - **Ngữ nghĩa đầu cuối chính xác một lần**
- Hợp nhất các truy vấn phát trực tuyến, tư ơ ng tác và hàng loạt
 - Tổng hợp dữ liệu trong một luồng, sau đó phục vụ bằng JDBC
 - Thêm, xóa, thay đổi truy vấn khi chạy
 - Xây dựng và áp dụng các mô hình ML

Thực hiện nội bộ

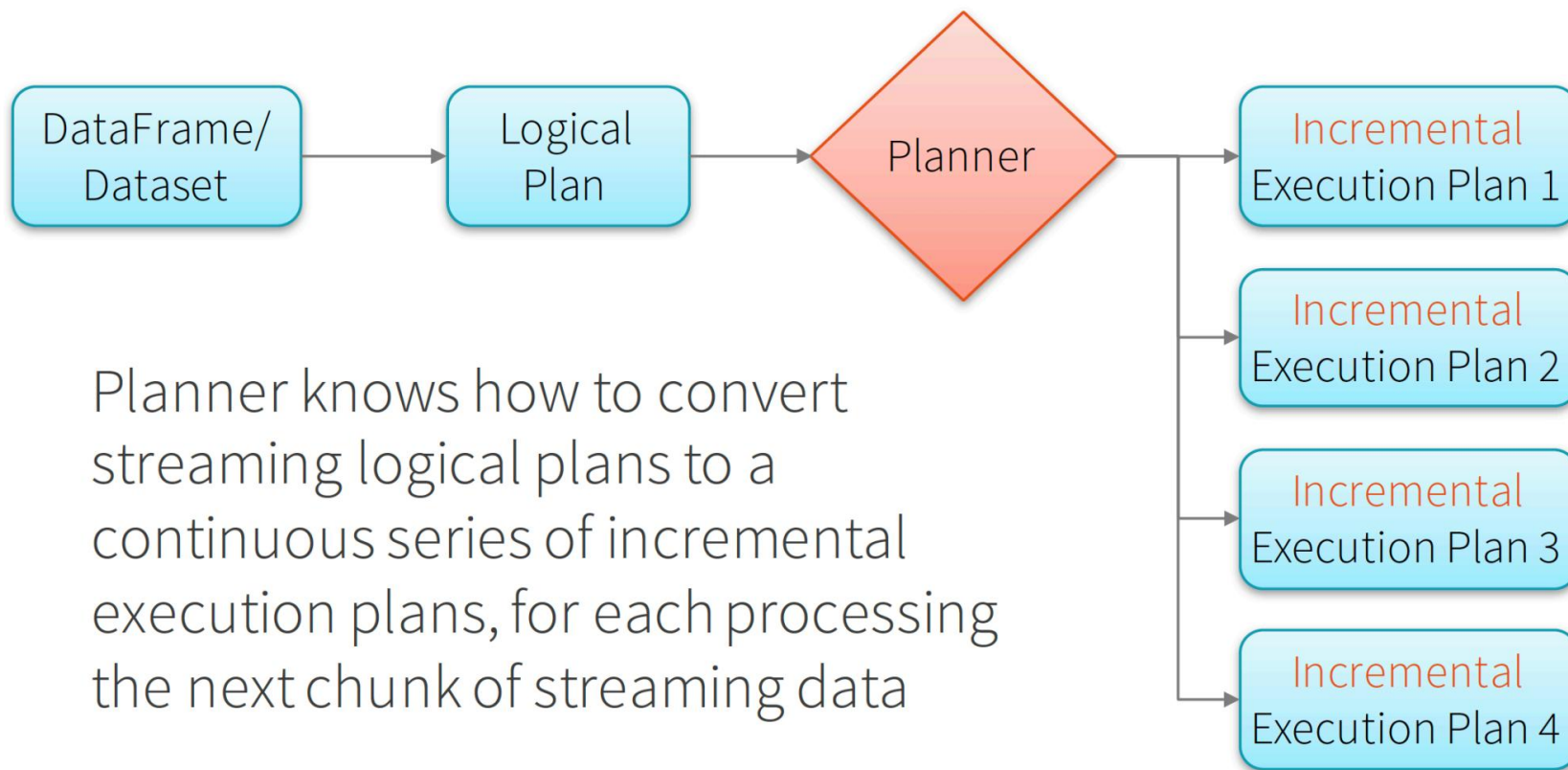
Thực hiện hàng loạt trên Spark SQL



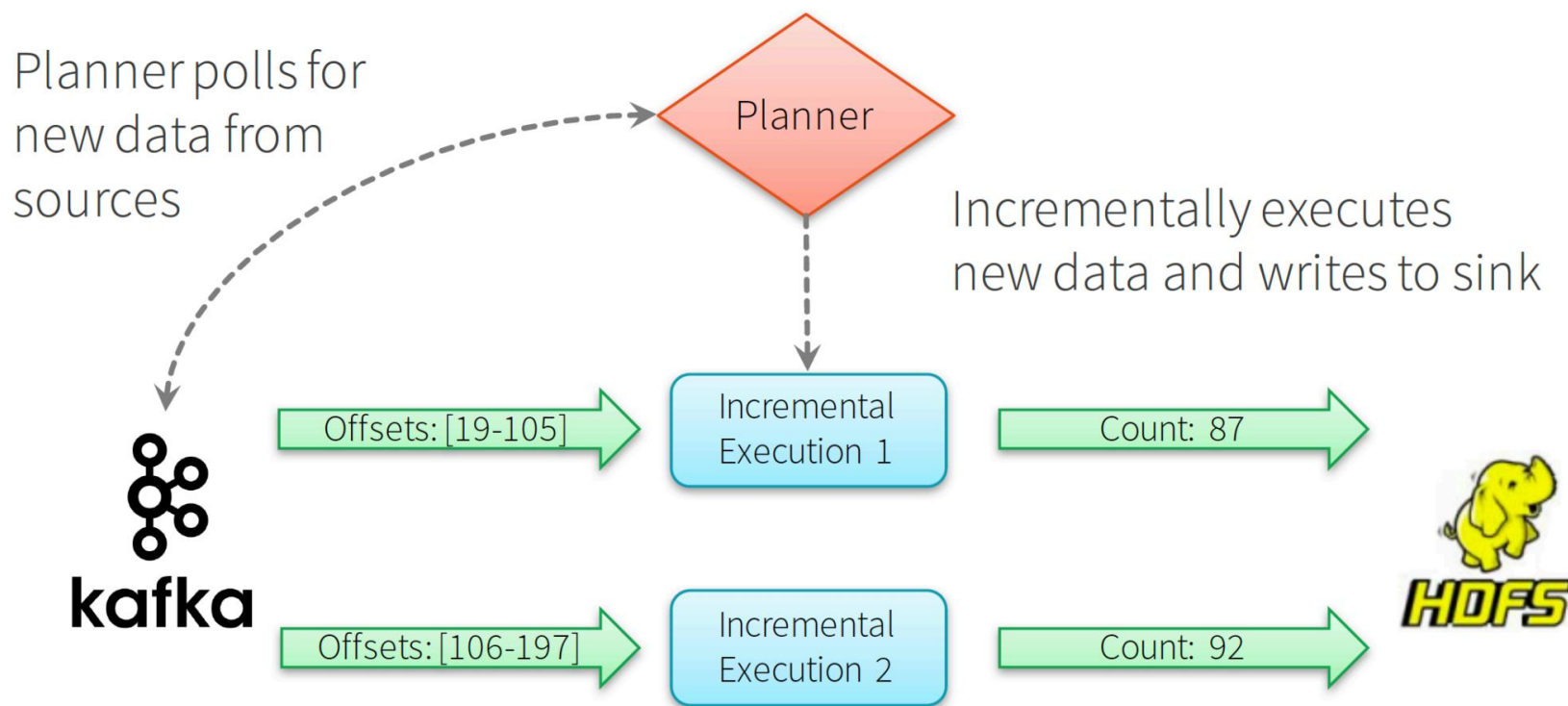
Thực hiện hàng loạt trên Spark SQL



Thực hiện gia tăng liên tục

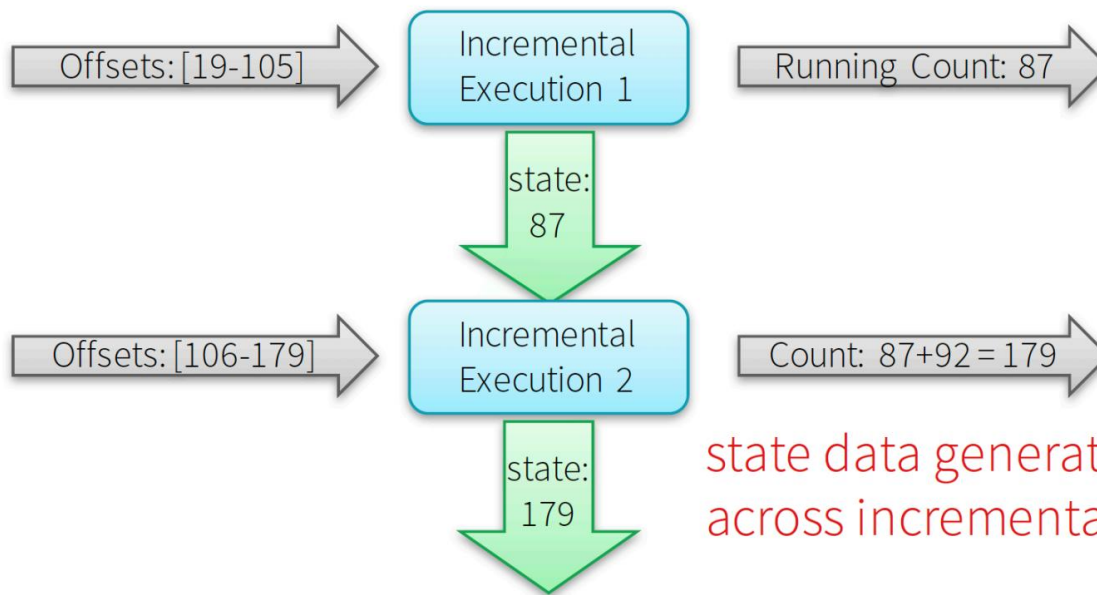


Thực hiện gia tăng liên tục



Tổng hợp liên tục

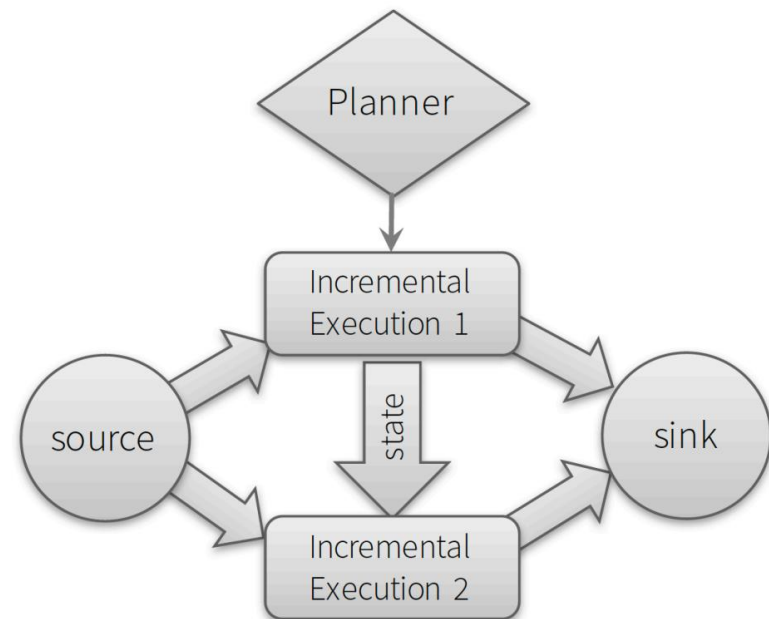
Maintain running aggregate as **in-memory state** backed by **WAL in file system** for fault-tolerance



state data generated and used across incremental executions

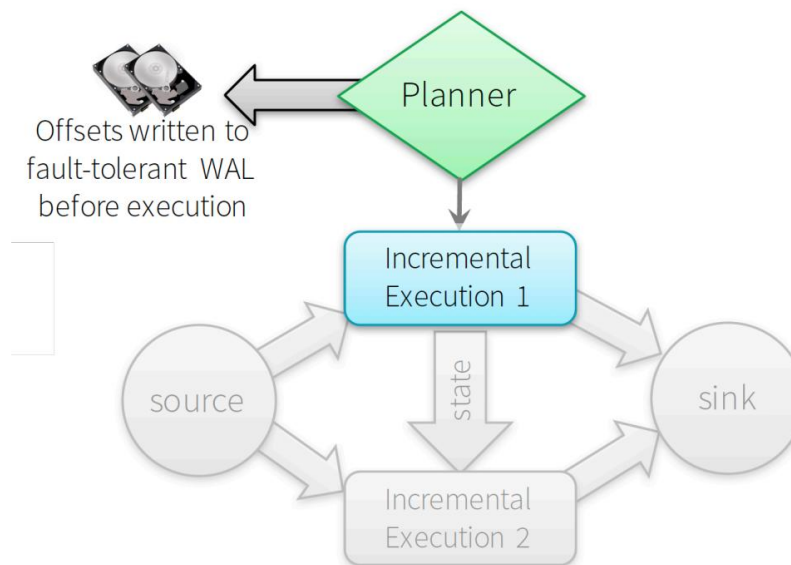
Khả năng chịu lỗi

- Tất cả dữ liệu và siêu dữ liệu trong hệ thống cần có thể phục hồi/phát lại được



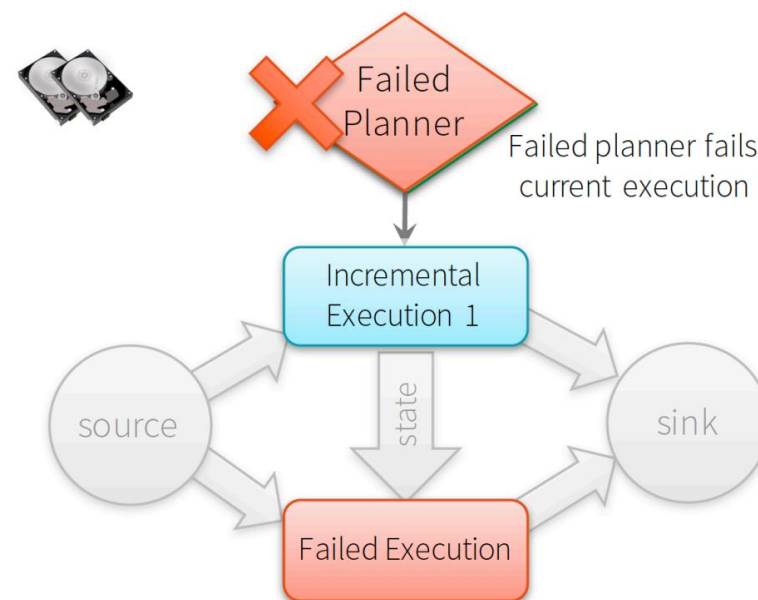
Trình lập kế hoạch chịu lỗi

- Theo dõi các bù trừ bằng cách ghi phạm vi bù trừ của mỗi lần thực hiện vào nhật ký ghi trước (WAL) trong HDFS



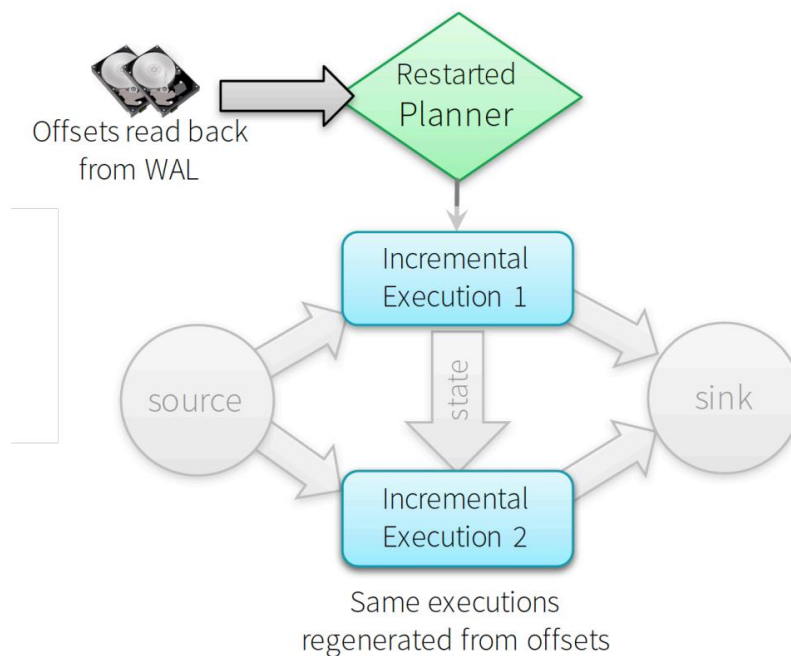
Trình lập kế hoạch chịu lỗi

- Theo dõi các bù trừ bằng cách ghi phạm vi bù trừ của mỗi lần thực hiện vào nhật ký ghi trước (WAL) trong HDFS



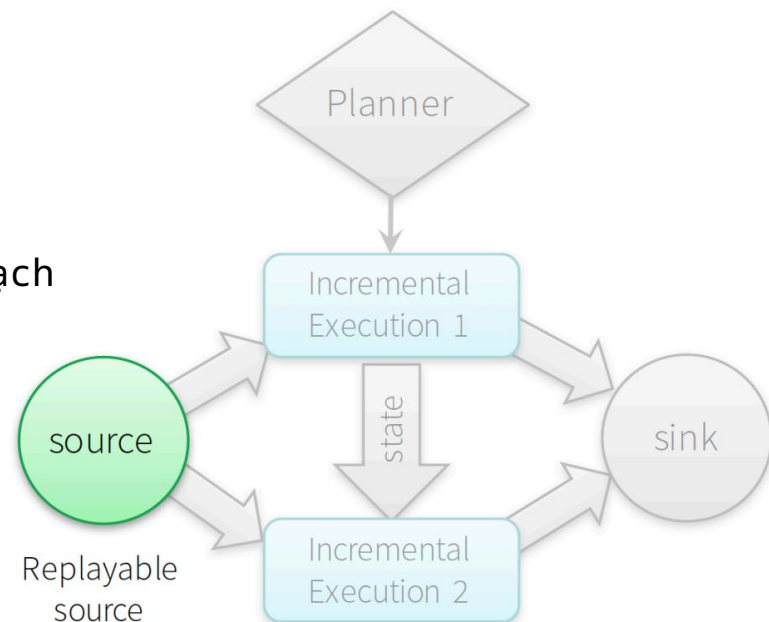
Trình lập kế hoạch chịu lỗi

- Theo dõi các bù trừ bằng cách ghi phạm vi bù trừ của mỗi lần thực hiện vào nhật ký ghi trước (WAL) trong HDFS
- Đọc nhật ký để phục hồi sau lỗi và thực hiện lại phạm vi bù trừ chính xác



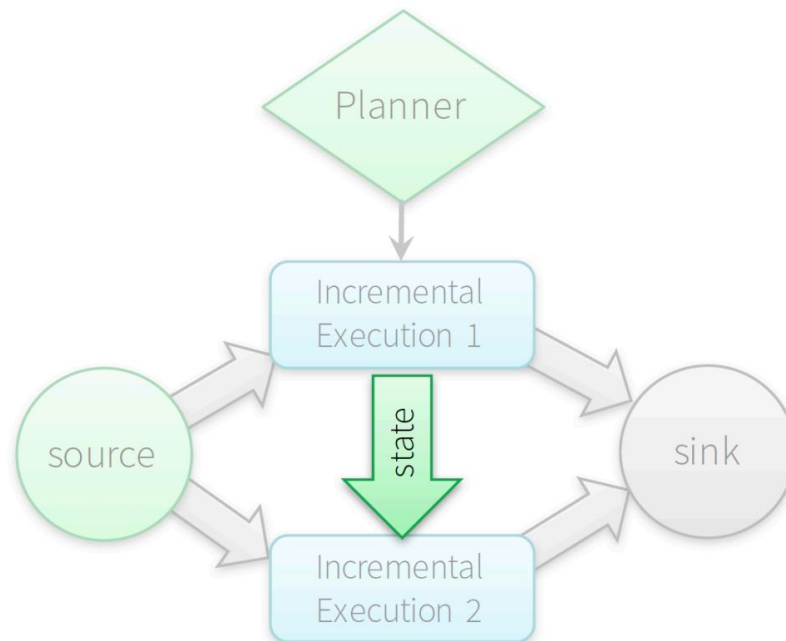
Nguồn chịu lỗi

- Các nguồn phát trực tuyến có cấu trúc được thiết kế có thể phát lại (ví dụ Kafka, Kinesis, tệp) và tạo ra dữ liệu giống hệt nhau khi có các độ lệch được khôi phục bởi trình lập kế hoạch



Trạng thái chịu lỗi

- "Dữ liệu trạng thái" trung gian được duy trì trong các bản đồ giá trị khóa có phiên bản trong Spark worker, được hỗ trợ bởi HDFS
- Người lập kế hoạch đảm bảo "phiên bản đúng" của trạng thái được sử dụng để thực thi lại sau khi lỗi

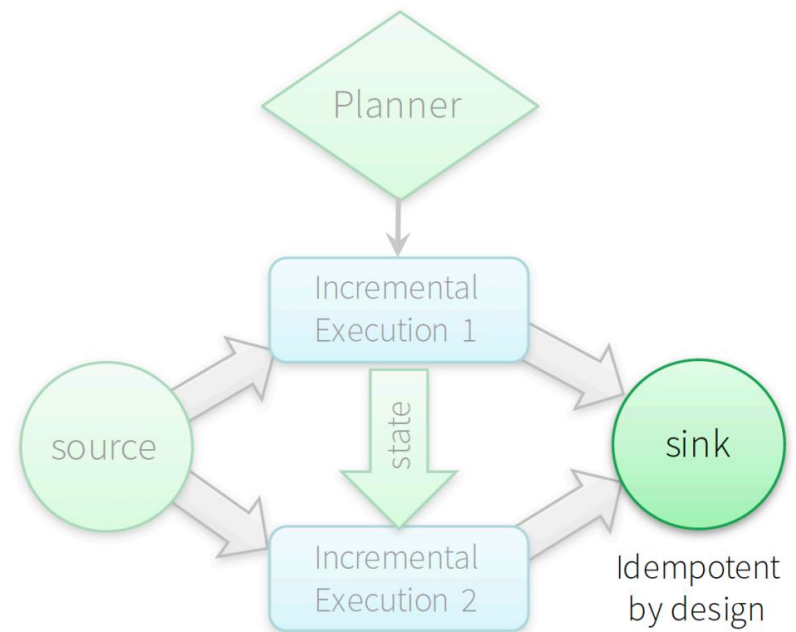


state is fault-tolerant with WAL



Bồn rửa chịu lỗi

- Sink được thiết kế theo nguyên lý idempotent (xác định) và xử lý việc thực thi lại để tránh cam kết đầu ra hai lần



Khả năng chịu lỗi

theo dõi bù trừ trong WAL

+

quản lý nhà nước

+

nguồn và bồn chứa có khả năng chịu lỗi

=

từ đầu đến cuối

đảm bảo chính

xác một lần

Phát trực tuyến có cấu trúc

Xử lý luồng trạng thái nhanh, có
khả năng chịu lỗi, chính
xác một lần mà không cần phải lý giải về luồng