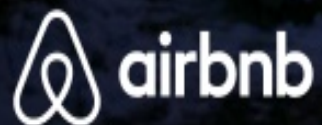


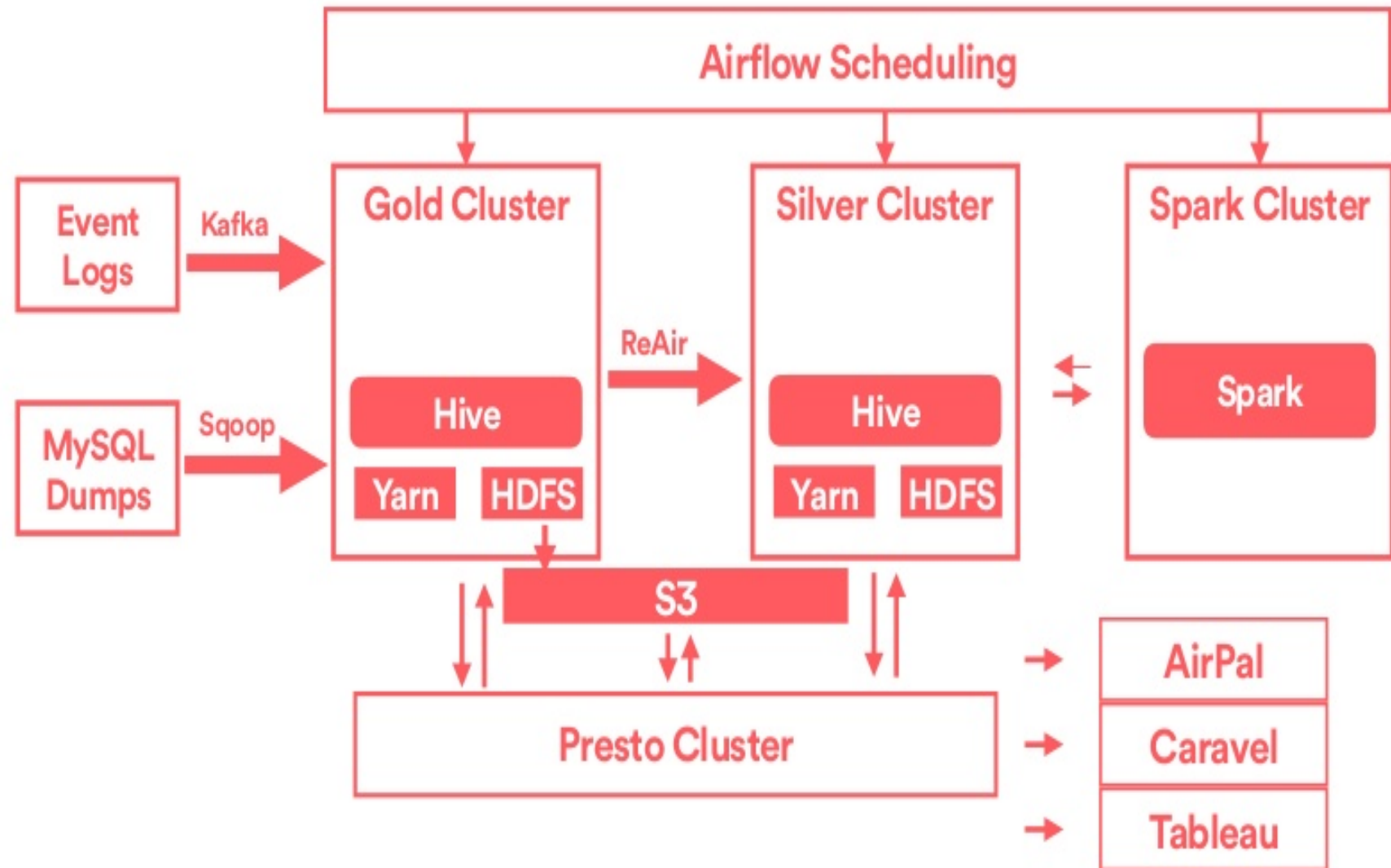
LIYIN TANG & JINGWEI LU

AirStream

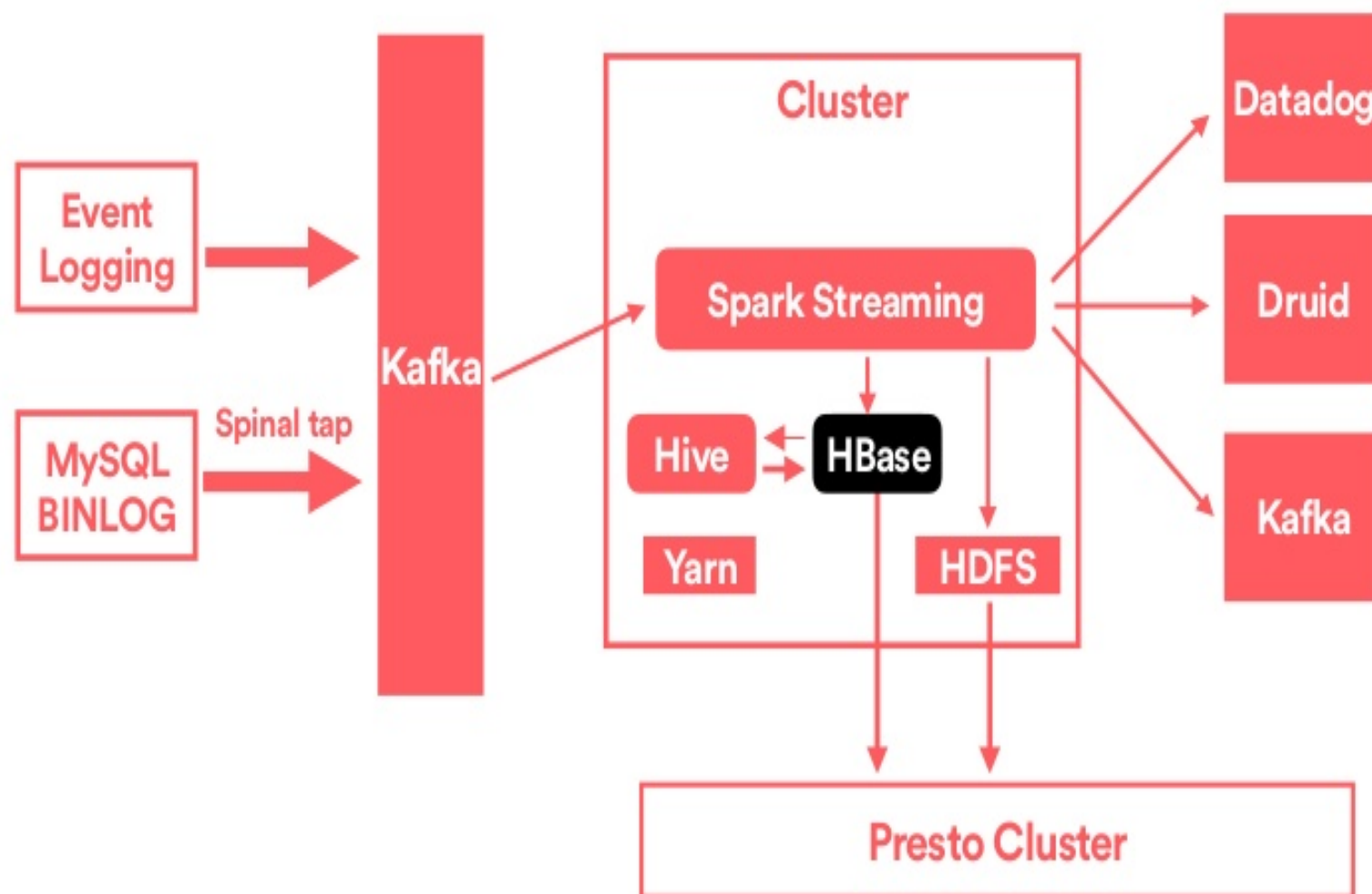


Data Infrastructure at Airbnb

Batch Infrastructure



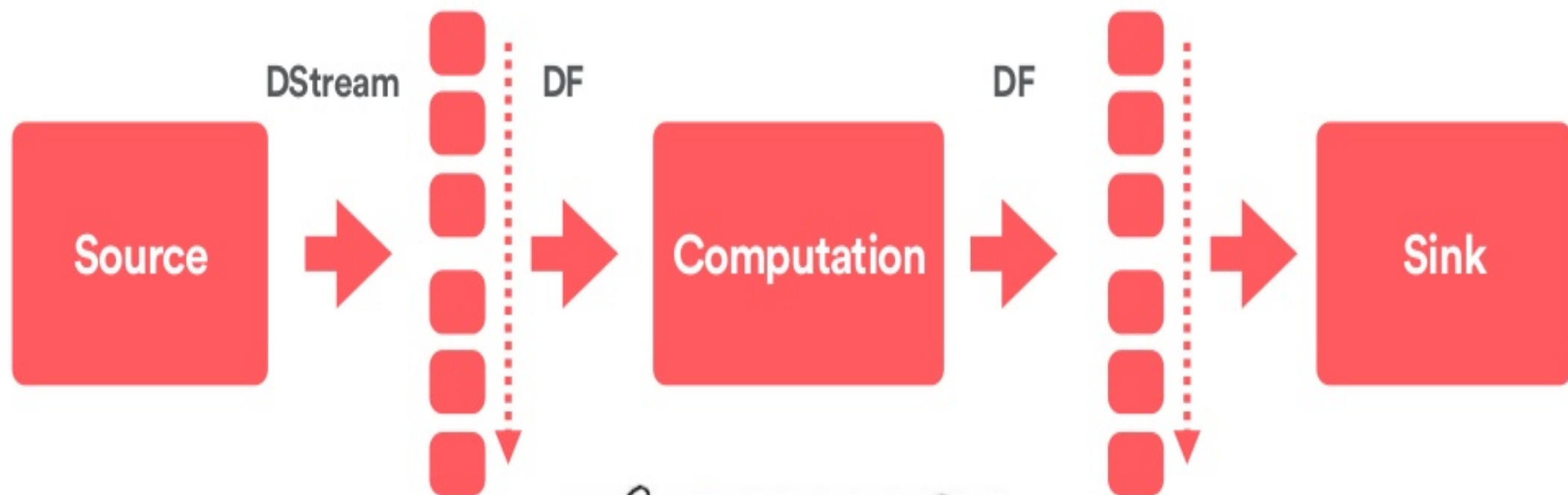
Streaming at Airbnb



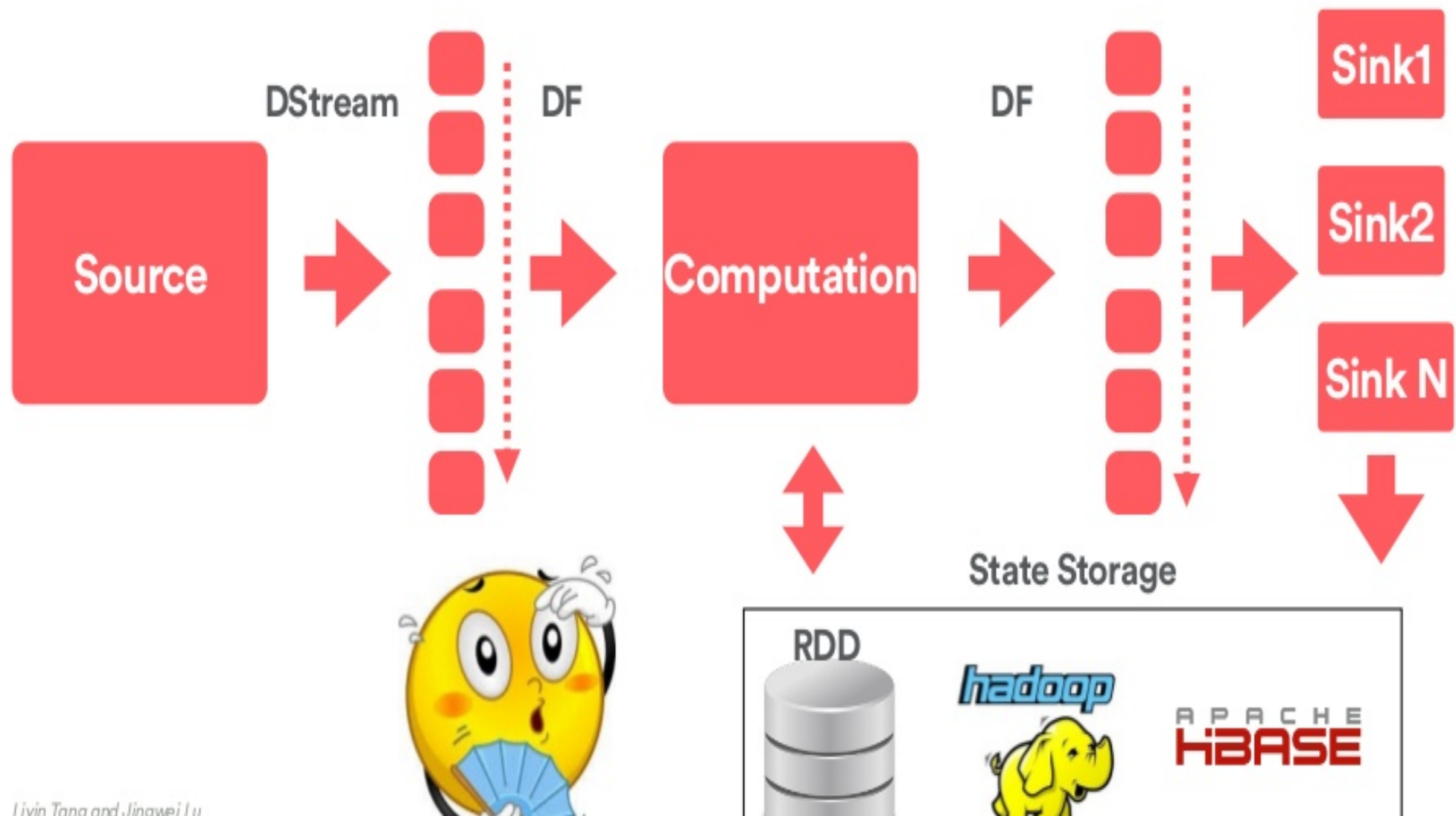
Growing Pain

The background of the slide is a solid red color. Overlaid on this background is a dense, repeating pattern of stylized, hand-drawn teardrop or leaf-like shapes. These shapes are drawn with a darker red, slightly irregular line, giving them a sketchy, organic appearance. They are scattered across the entire red field, creating a textured effect.

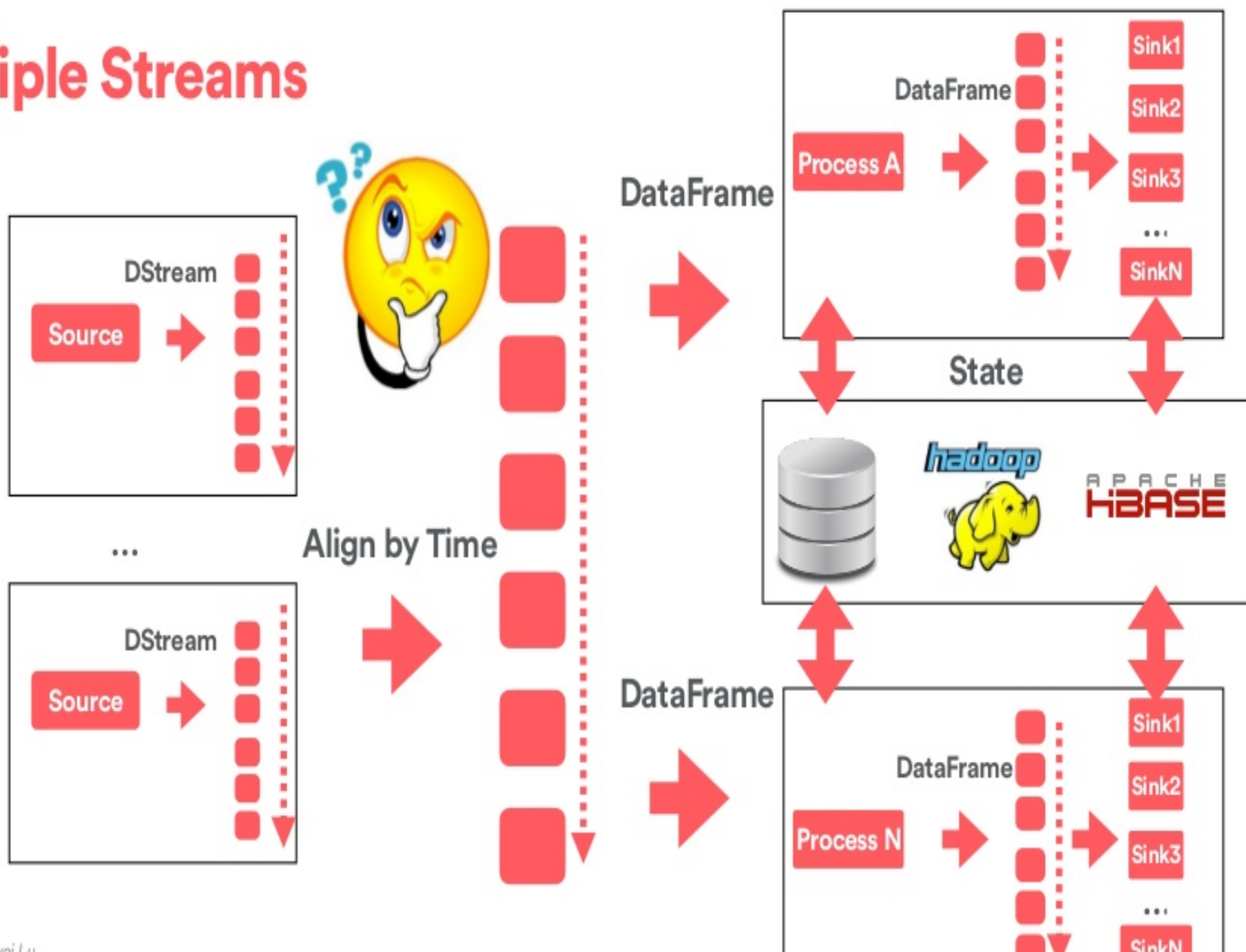
Stateless



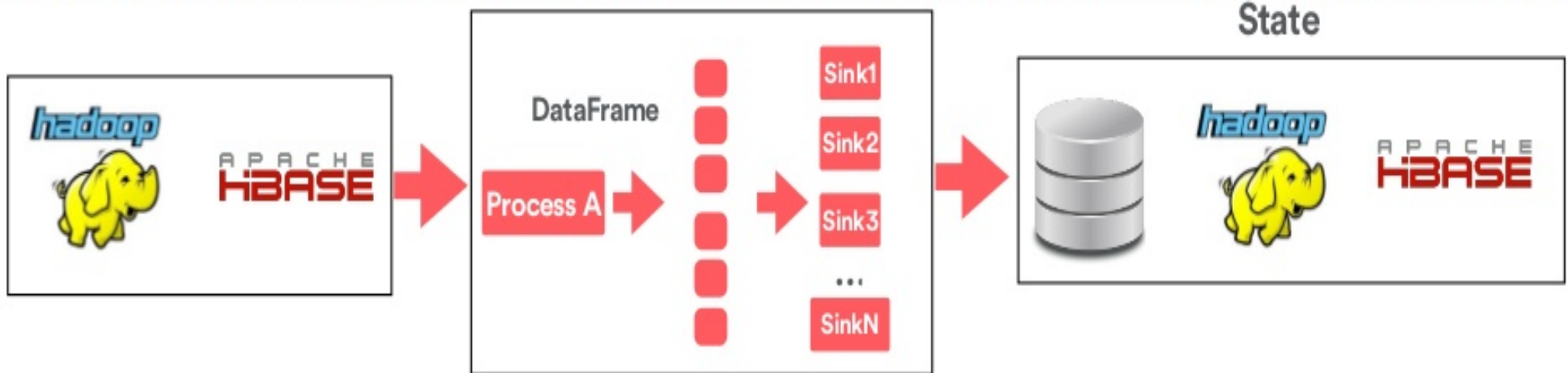
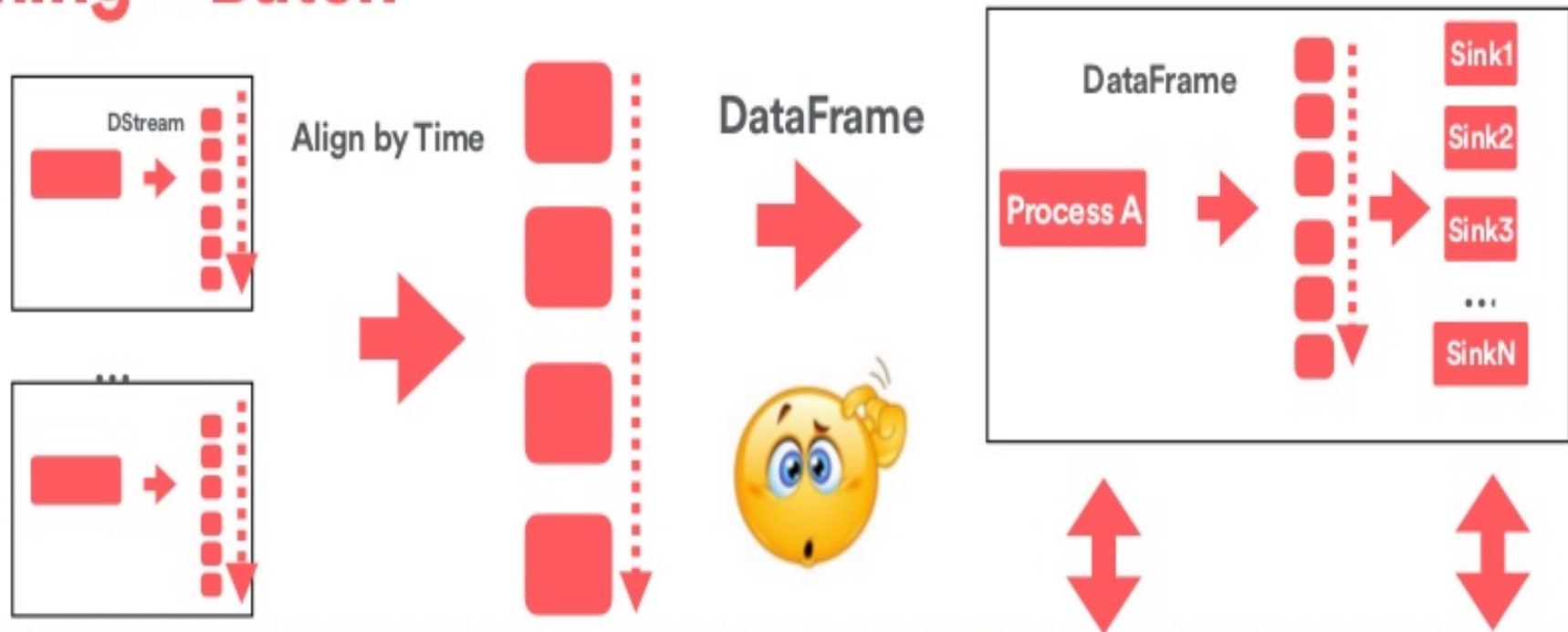
Stateful



Multiple Streams



Streaming + Batch



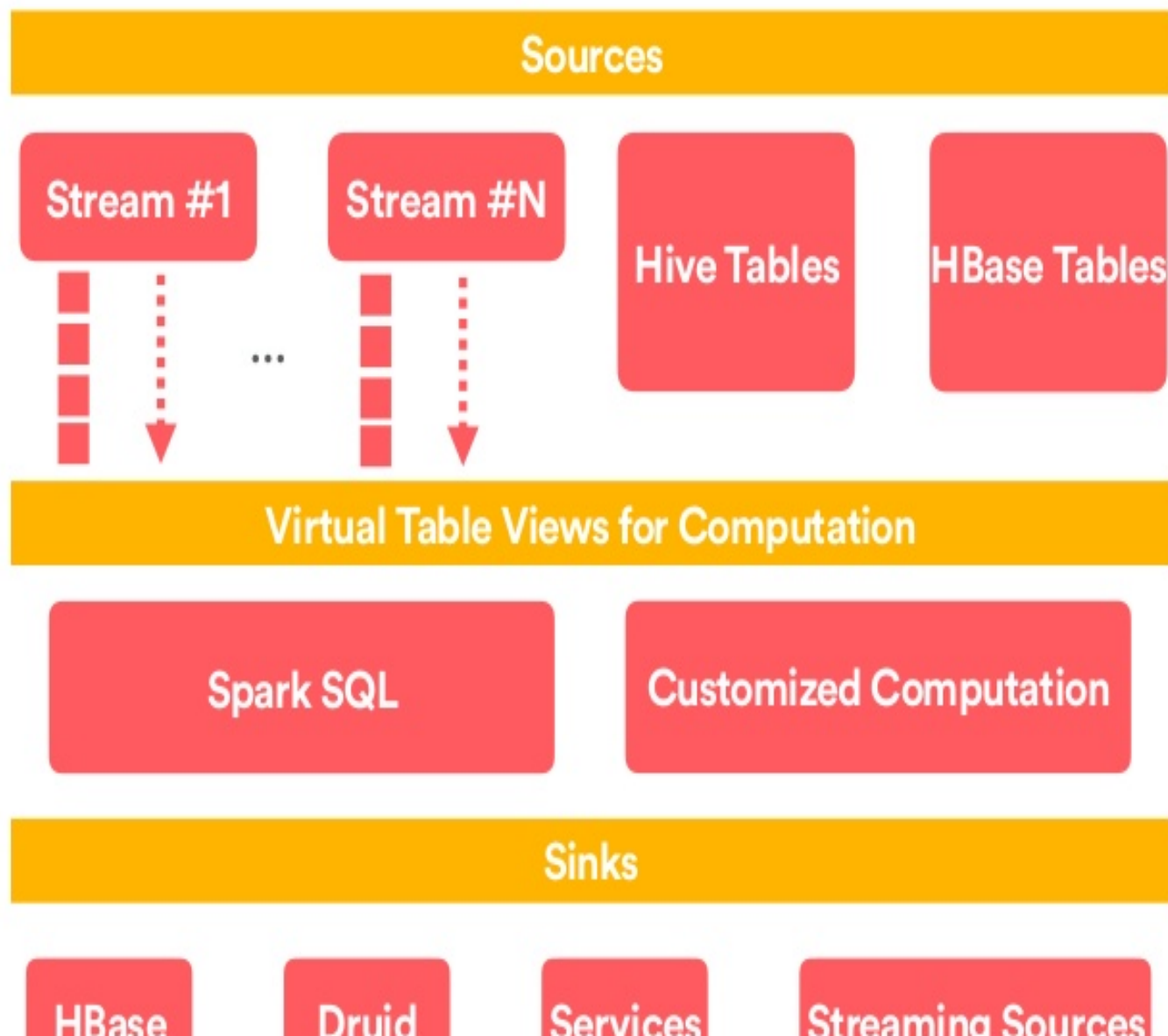
Simplify and Unify

The background of the slide is a solid red color. Overlaid on this background is a repeating pattern of stylized, hand-drawn teardrop or leaf-like shapes. These shapes are drawn with a darker red, slightly irregular line, giving them a sketchy, organic appearance. They are scattered across the entire red field, creating a textured, patterned effect.

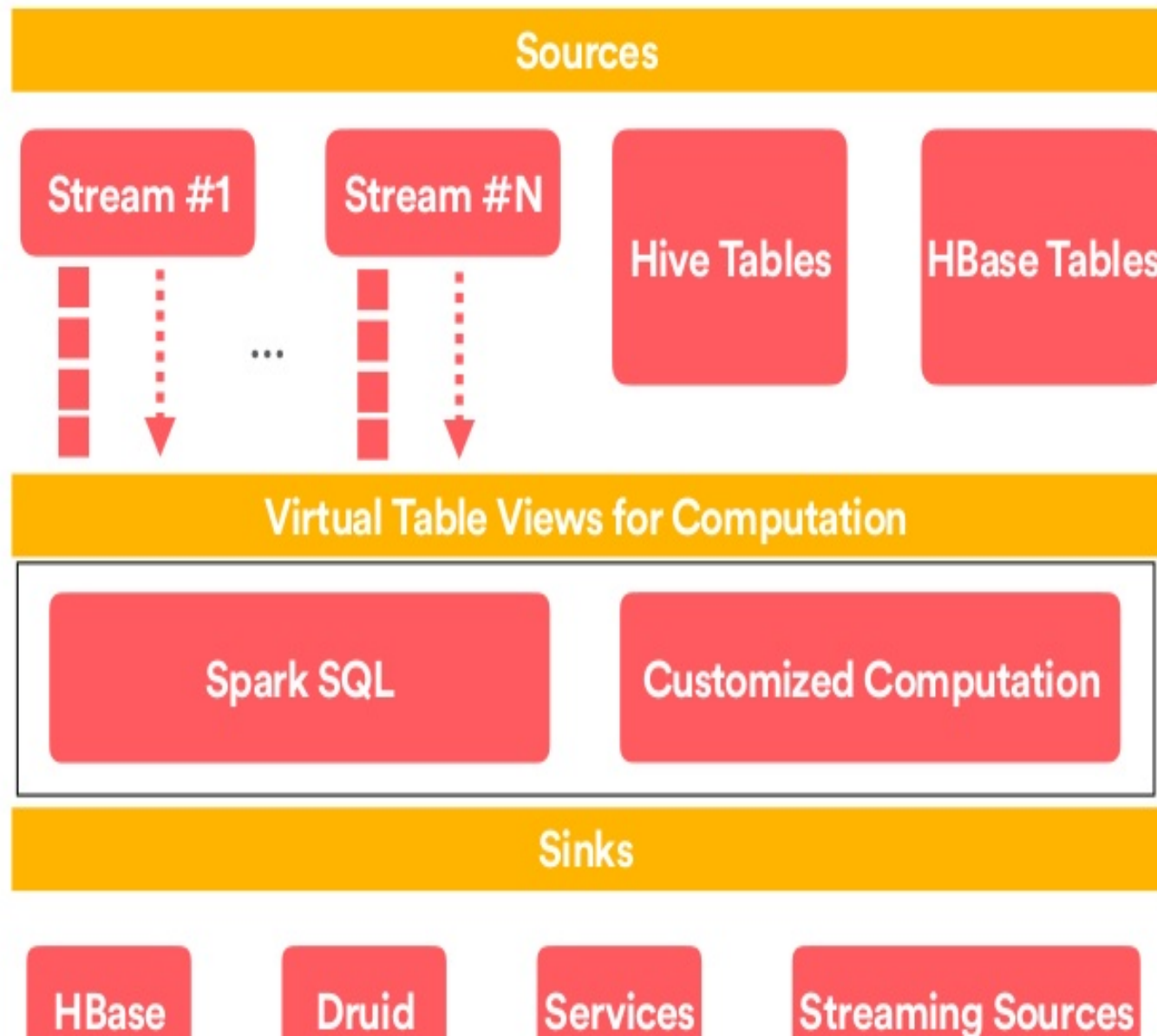
AirStream Architecture



Simple Config



AirStream Architecture



Same Computation for
Batch processing

Stateful

The background of the slide is a solid red color. Overlaid on this background is a dense, repeating pattern of stylized, hand-drawn shapes. These shapes are drawn in a lighter shade of red and resemble teardrops or abstract, elongated forms with a small loop or swirl at the bottom. The pattern is dense and covers the entire area of the slide.

State Store

- Merge changes
- Provide fast lookup
- Fast persistent storage across streaming and batch jobs

Why HBase

Reliable and Scalable

Rich Integration with Hadoop EcoSystem

Rich Functionalities

Easy Management

Strong Community

HBase State Store

Operators in Airstream



Full Table Scan



Key/Prefix Lookup



Update

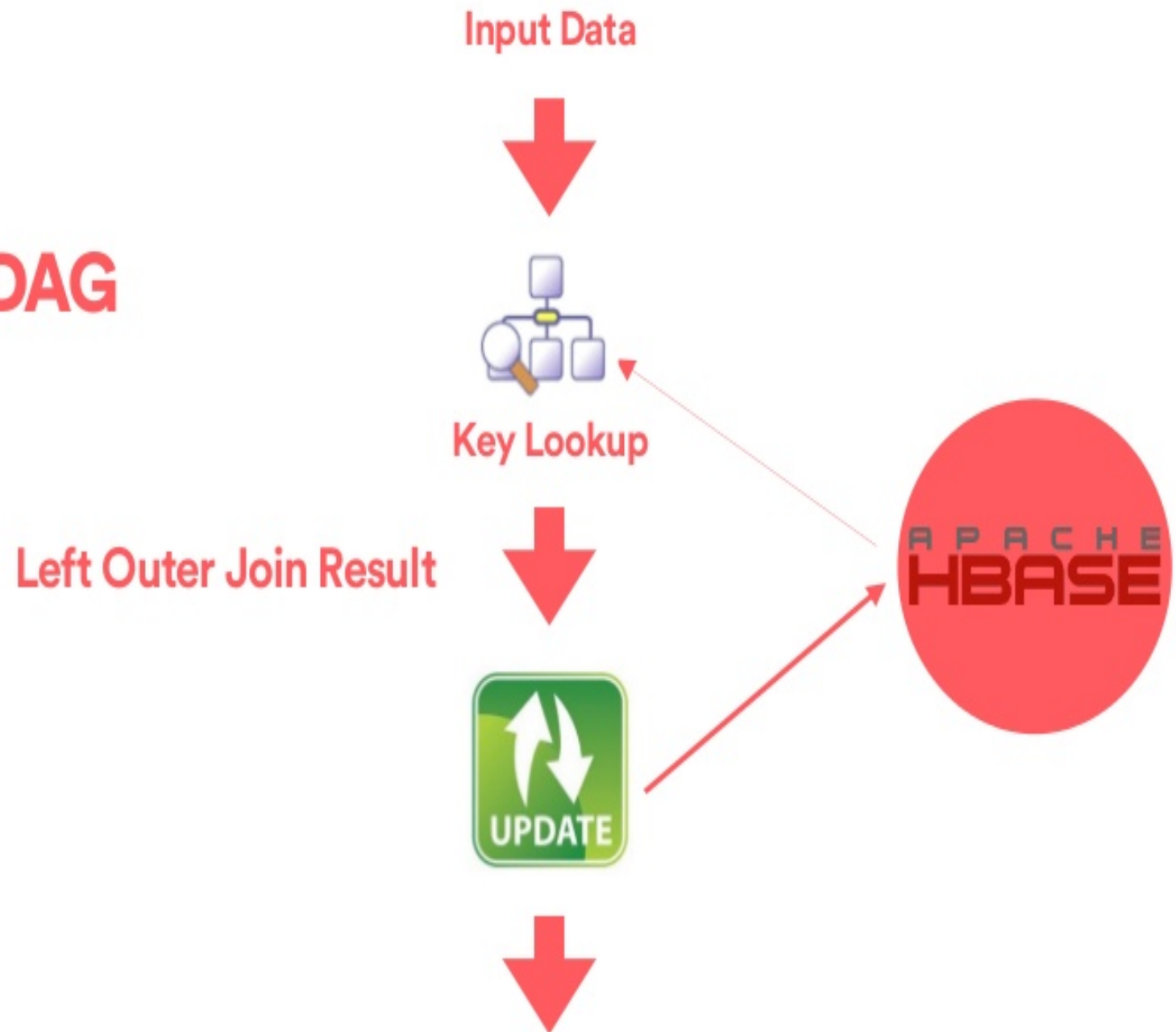


Bulk Upload



Simple Aggregation

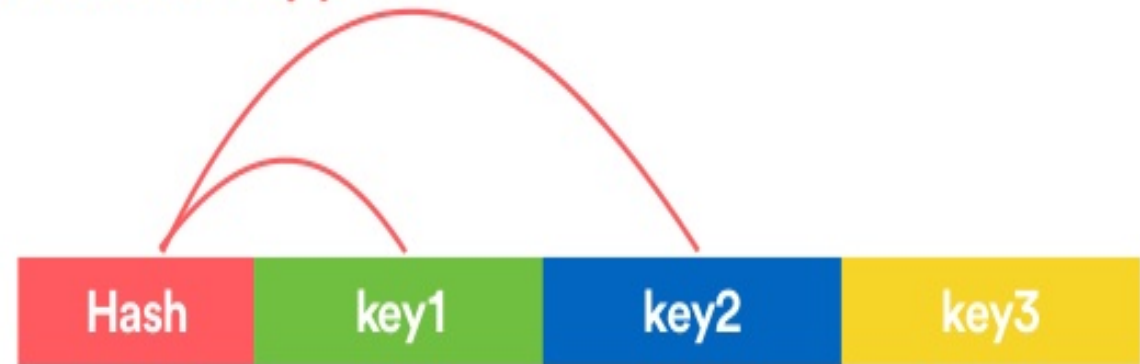
Computation DAG



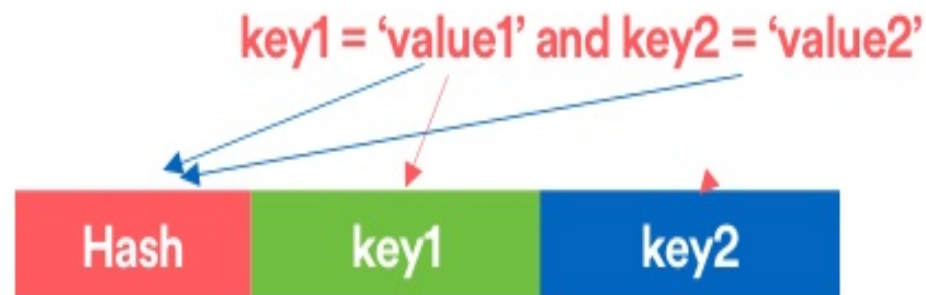
Key Space Design

- Hash partition key space for load balance
- Composite key for $K \rightarrow V$
- Support full key lookup
- Prefix lookup supported for all keys used in hash function

Hash based on key prefix



Lookup based on key prefix

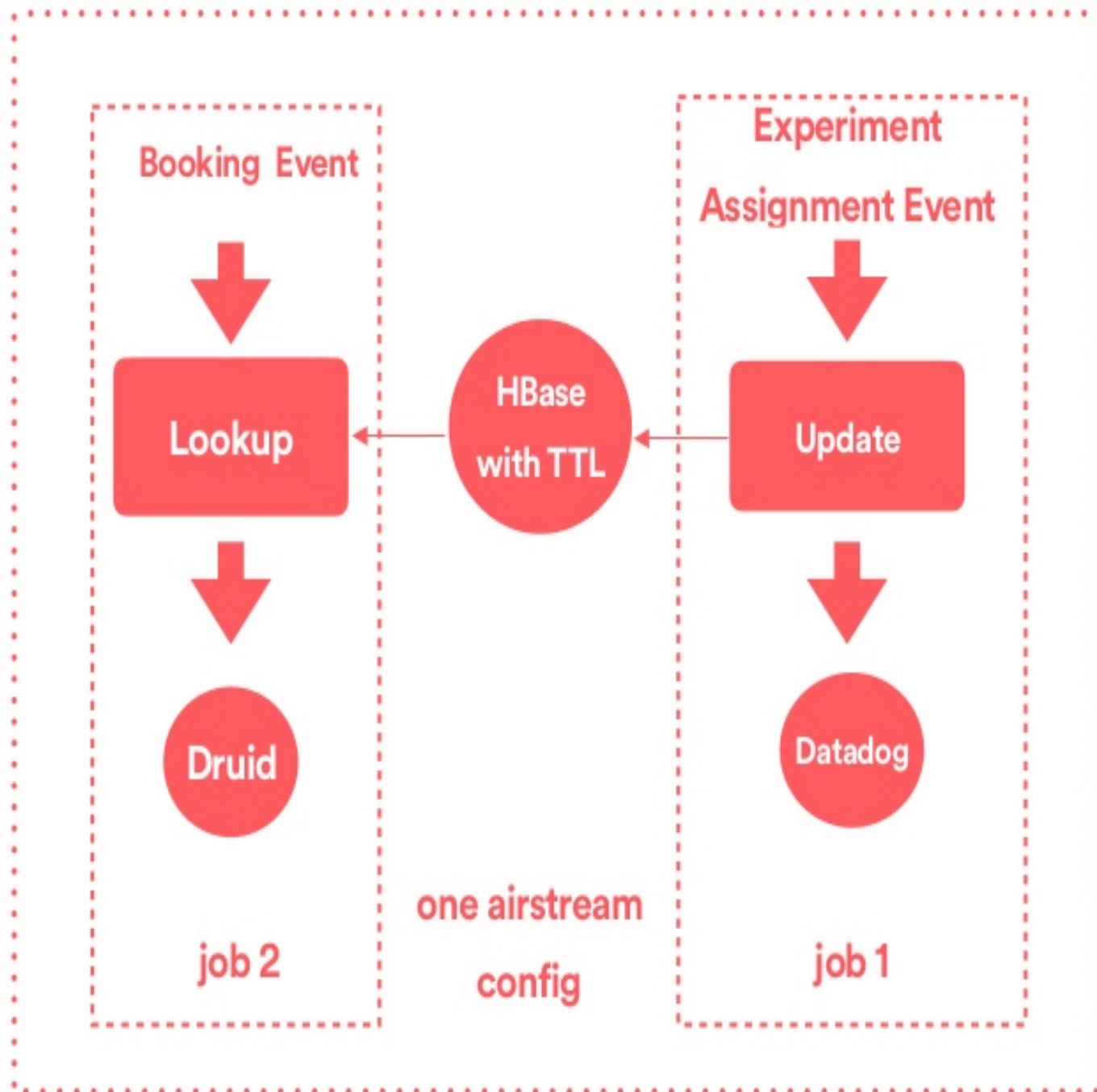


Write Performance

- Partition based on key before write
- Use bulk upload for large volume update

Case Study

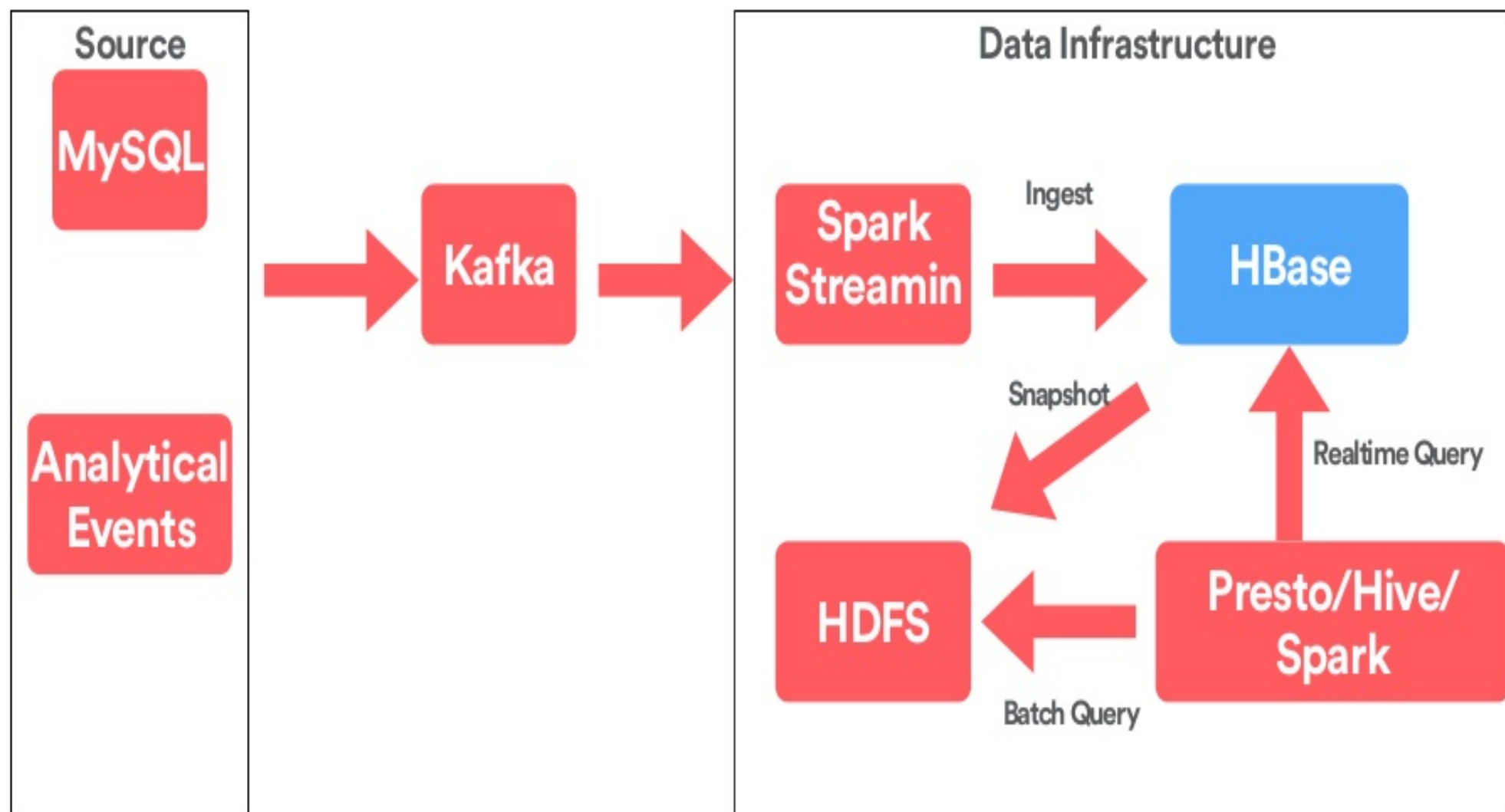
Experiment realtime feedback



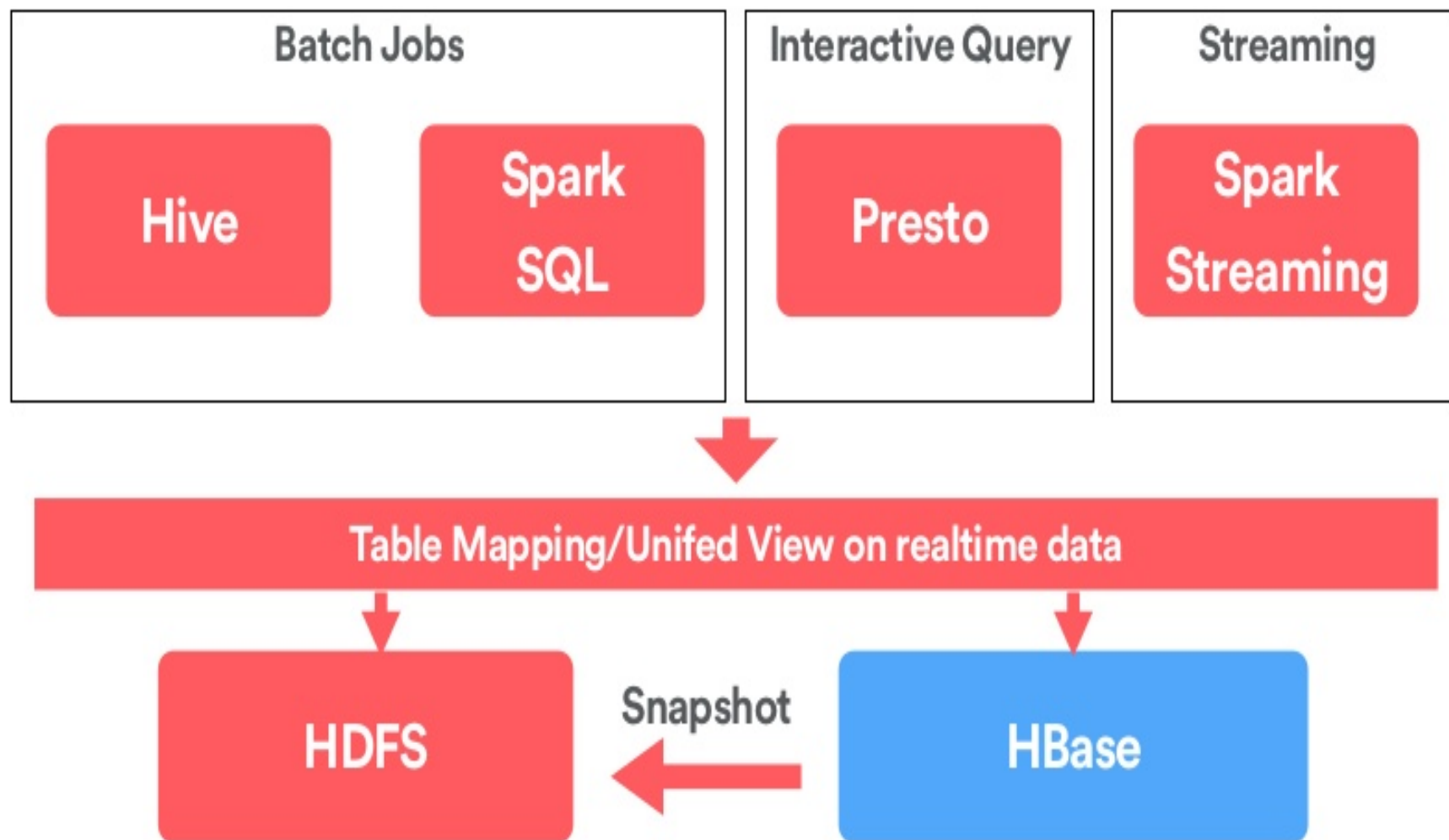
Realtime Data Ingestion

The background of the slide is a solid red color. Overlaid on this background is a repeating pattern of white, hand-drawn style teardrop or leaf-like shapes. These shapes are oriented in various directions, creating a textured, organic feel. The shapes are simple outlines with a small circular detail inside, resembling a stylized 'e' or a drop with a seed.

Realtime Ingestion on HBase



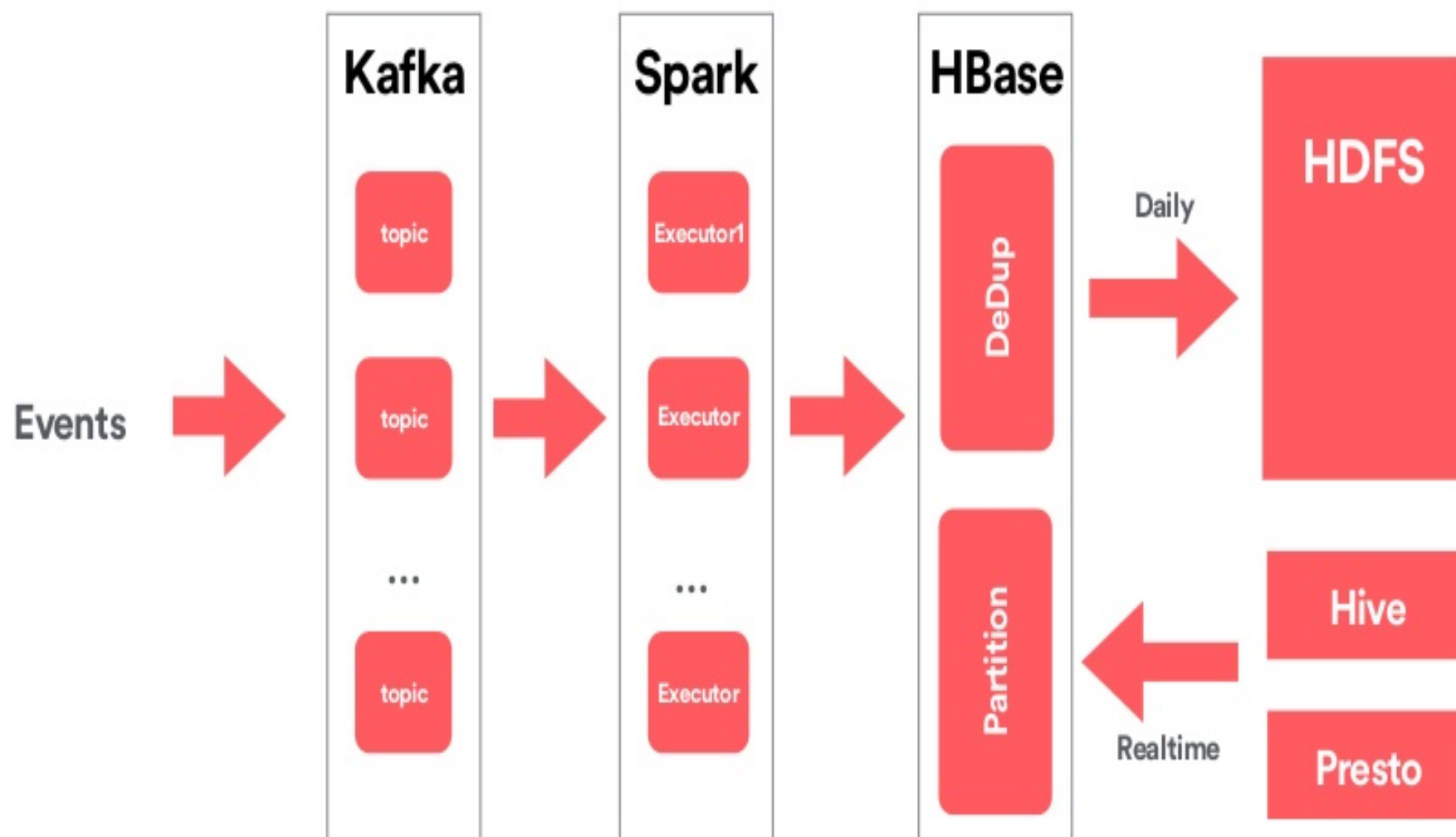
Access Data in HBase



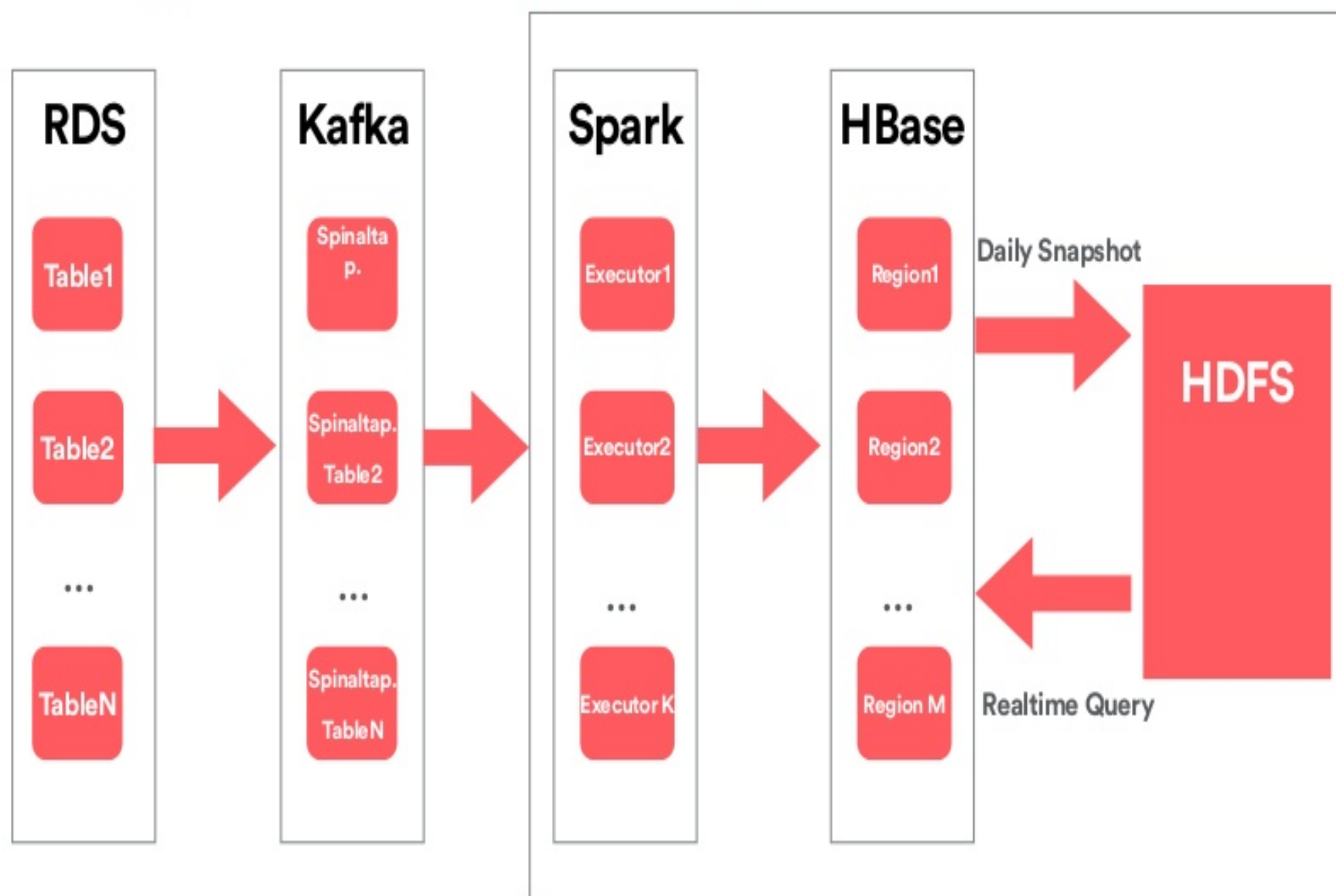
Snapshot & Reseed



Case Study 1: Events Ingestion



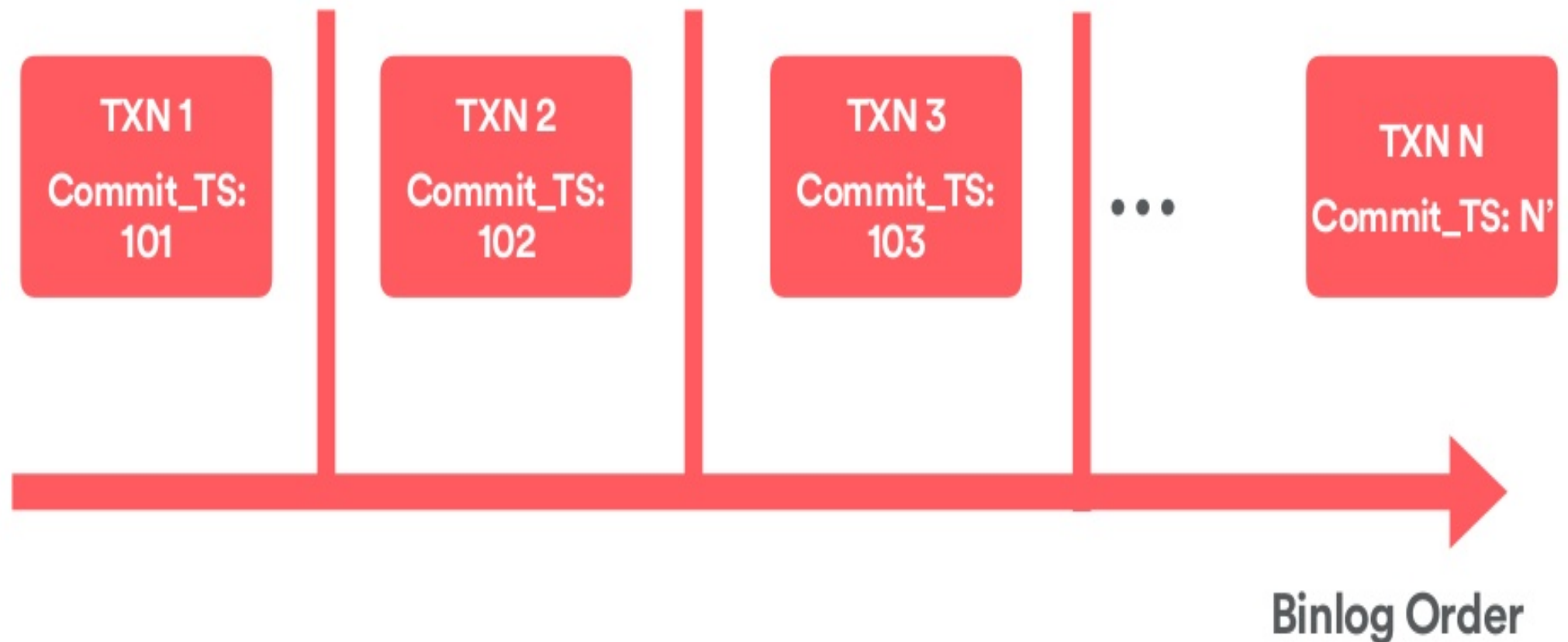
Case Study 2: Streaming DB Export



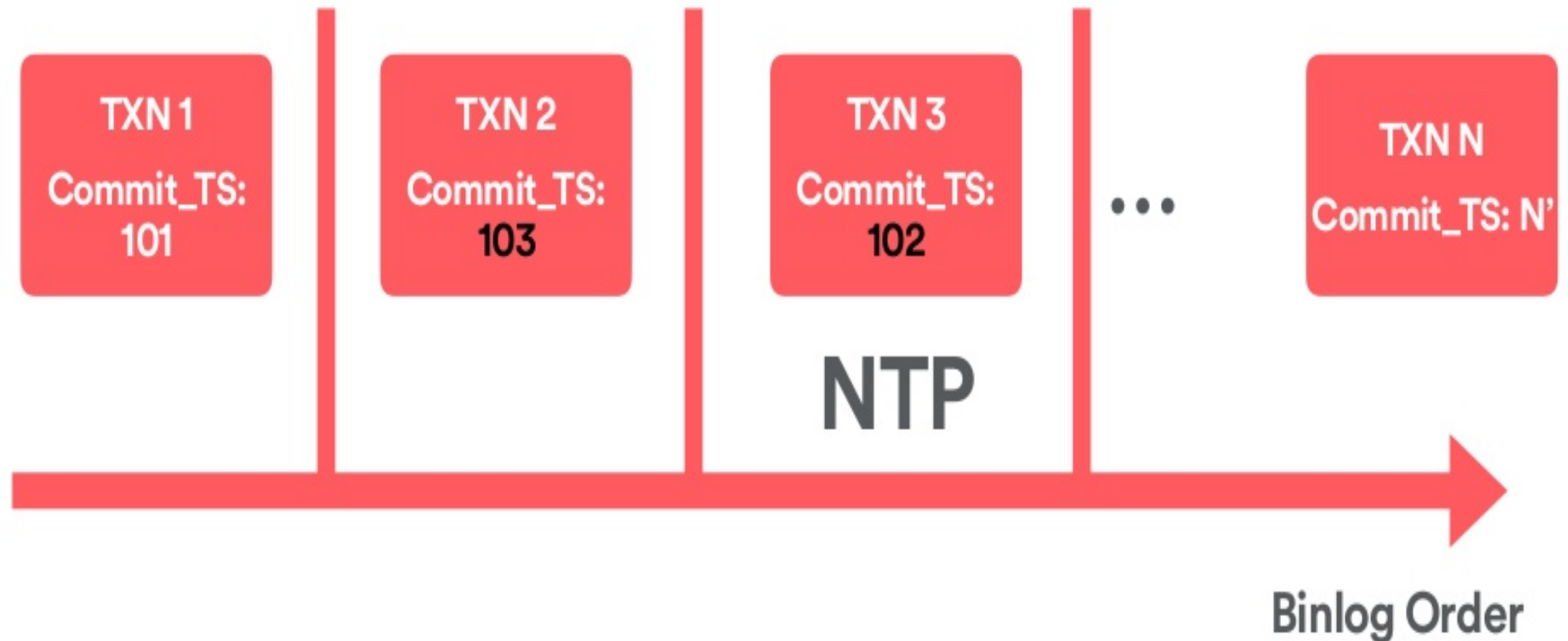
Case Study: Streaming DB Export

Rows	CF: Cols	Version	Value
<ShardKey><DB_TABLE_#1><PK_a=A>	id	Fri May 19 00:33:19 2016	101
<ShardKey><DB_TABLE_#1><PK_a=A>	city	Fri May 19 00:33:19 2016	San Francisco
<ShardKey><DB_TABLE_#1><PK_a=A>	city	Fri May 10 00:34:19 2016	New York
<ShardKey><DB_TABLE_#2><PK_a=A>	id	Fri May 19 00:33:19 2016	1

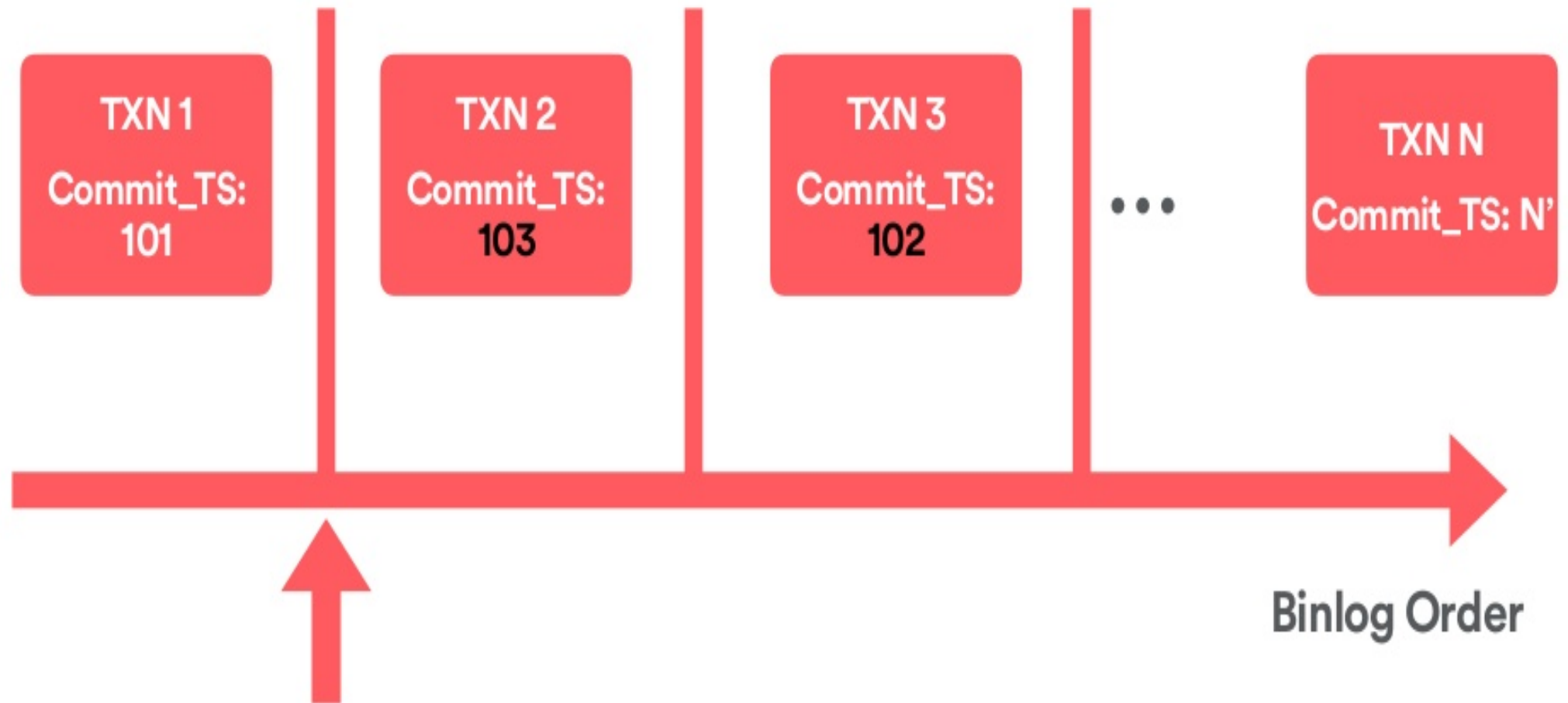
Case Study: Streaming DB Export



Case Study: Streaming DB Export



Case Study: Streaming DB Export



Point-in-Time Restore on TS 102

Case Study: Streaming DB Export

Rows	CF: Cols	Version	Value
<ShardKey><DB_TABLE_#1><PK_a=A>	id	bin100	101
<ShardKey><DB_TABLE_#1><PK_a=A>	city	bin101	San Francisco
<ShardKey><DB_TABLE_#1><PK_a=A>	city	bin102	New York
<ShardKey><DB_TABLE_#2><PK_a=A>	id	bin100	1

Case Study: Streaming DB Export

Rows	Version (Logical Offset)	Value
<ShardKey><DB_TABLE_#1><2016-05-23 23><100>	100	mysql-bin.00000:100
<ShardKey><DB_TABLE_#1><2016-05-23 23><101>	101	mysql-bin.00000:101
<ShardKey><DB_TABLE_#1><2016-05-23 23><103>	103	mysql-bin.00000:103
<ShardKey><DB_TABLE_#1><2016-05-24 00><102>	102	mysql-bin.00000:102

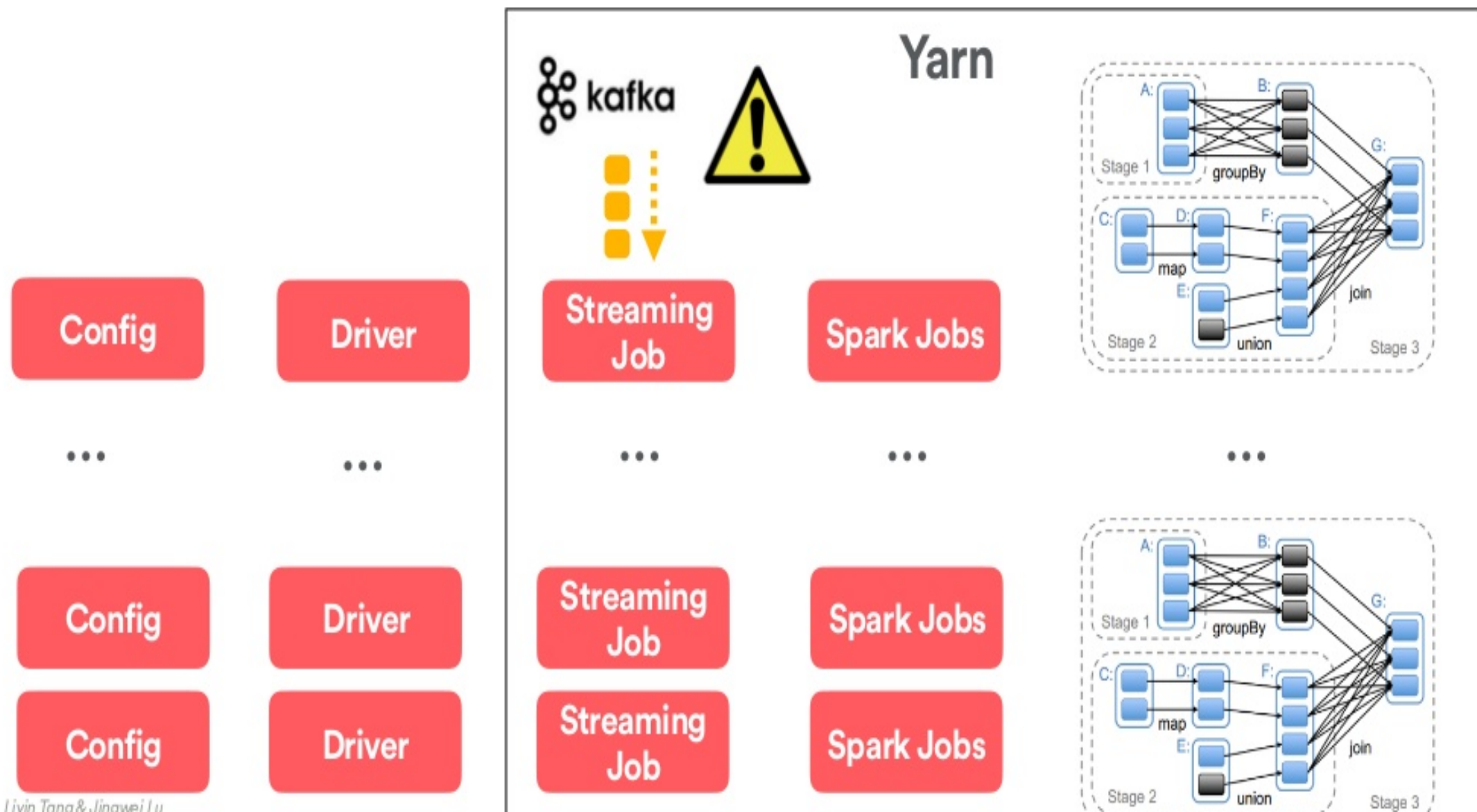
Case Study: Streaming DB Export

Rows	Version (Logical Offset)	Value
<ShardKey><DB_TABLE_#1><2016-05-23 23><100>	100	mysql-bin.00000:100
<ShardKey><DB_TABLE_#1><2016-05-23 23><101>	101	mysql-bin.00000:101
<ShardKey><DB_TABLE_#1><2016-05-23 23><103>	103	mysql-bin.00000:103
<ShardKey><DB_TABLE_#1><2016-05-24 00><102>	102	mysql-bin.00000:102

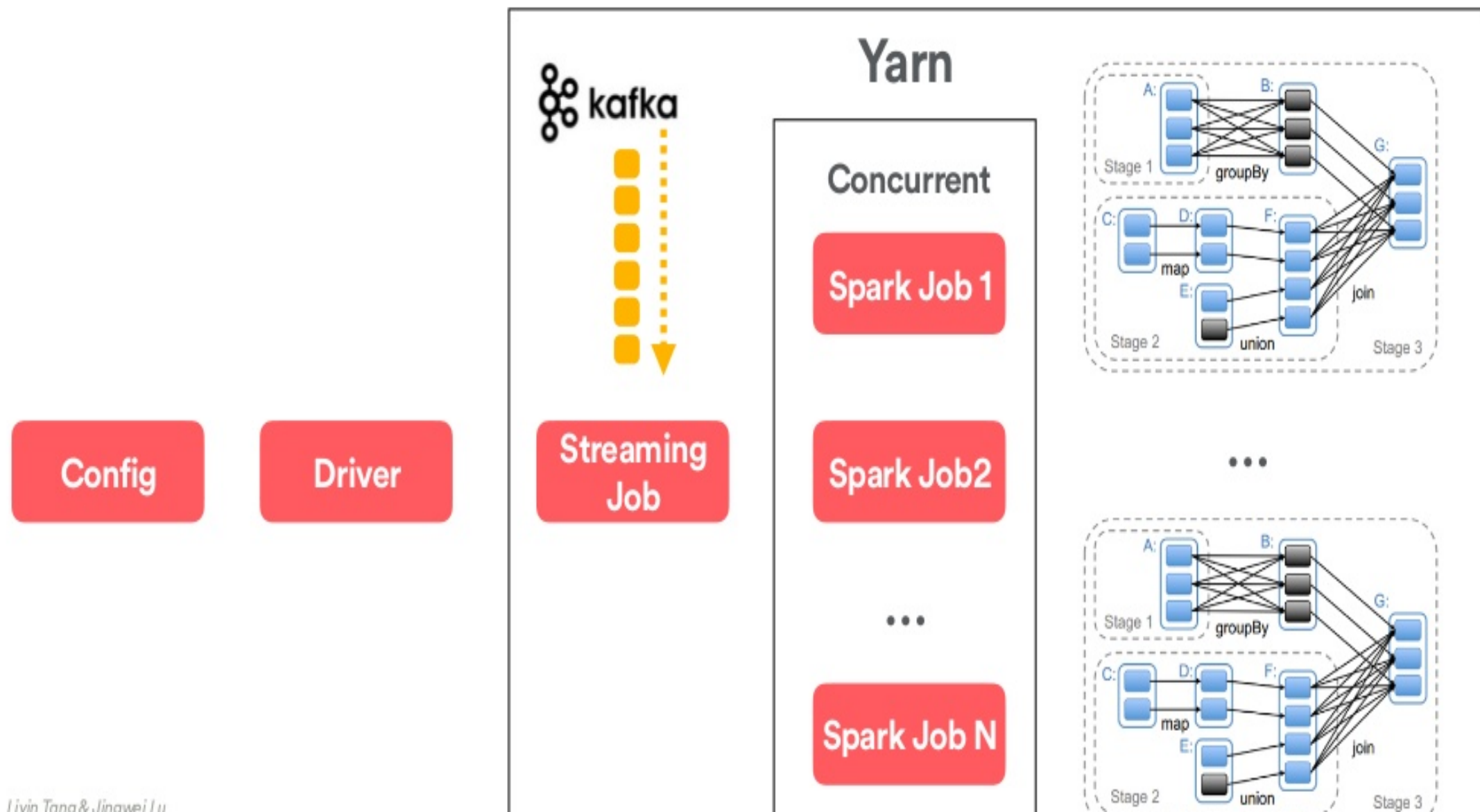
Operation

The background of the slide is a solid red color. Overlaid on this background is a repeating pattern of stylized, hand-drawn shapes that resemble trees or leaves. These shapes are drawn with a darker red, slightly textured line, giving them a sketchy, organic appearance. They are scattered across the entire surface, creating a dense, textured effect.

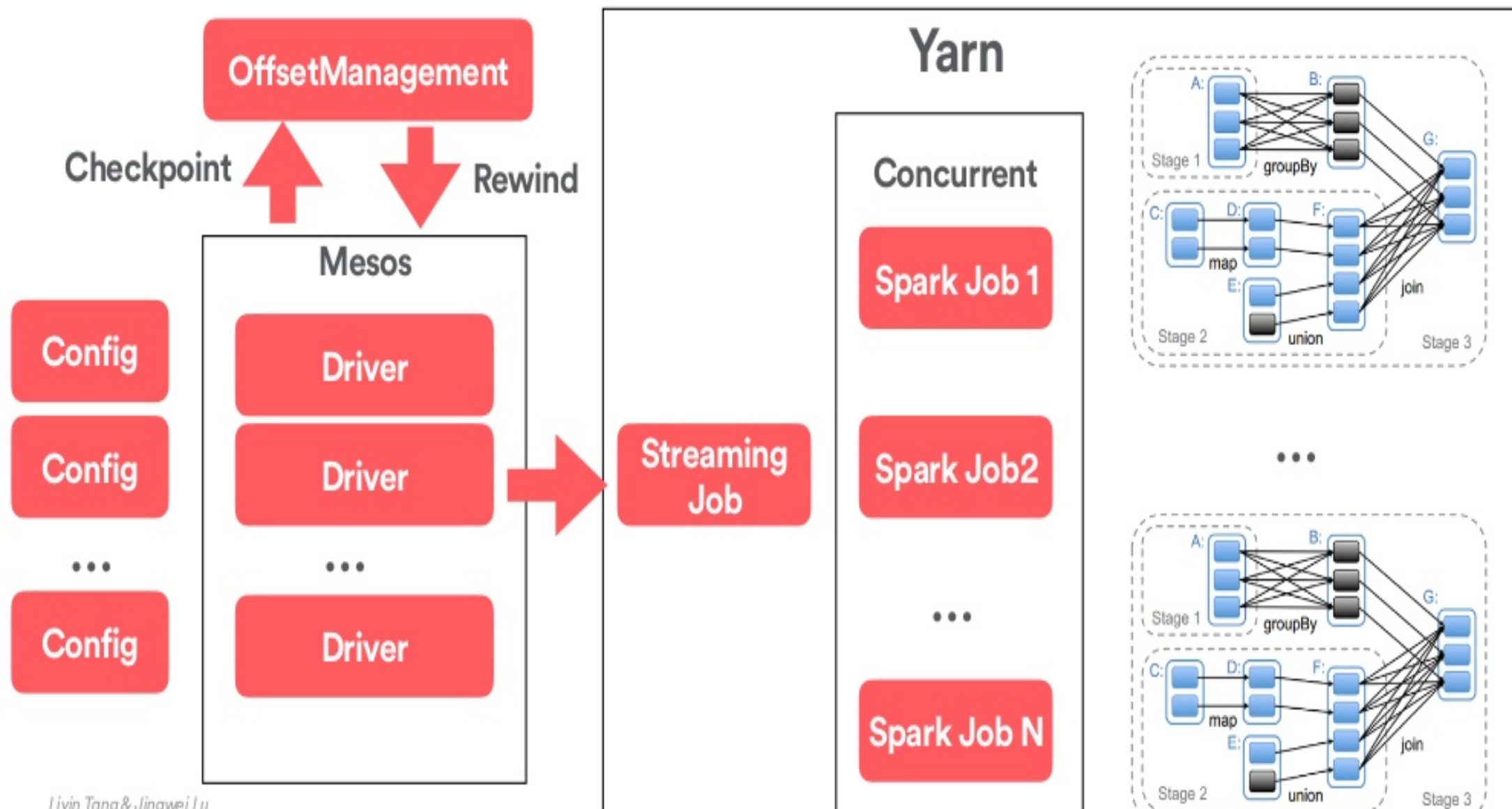
Job Management: Scaling up



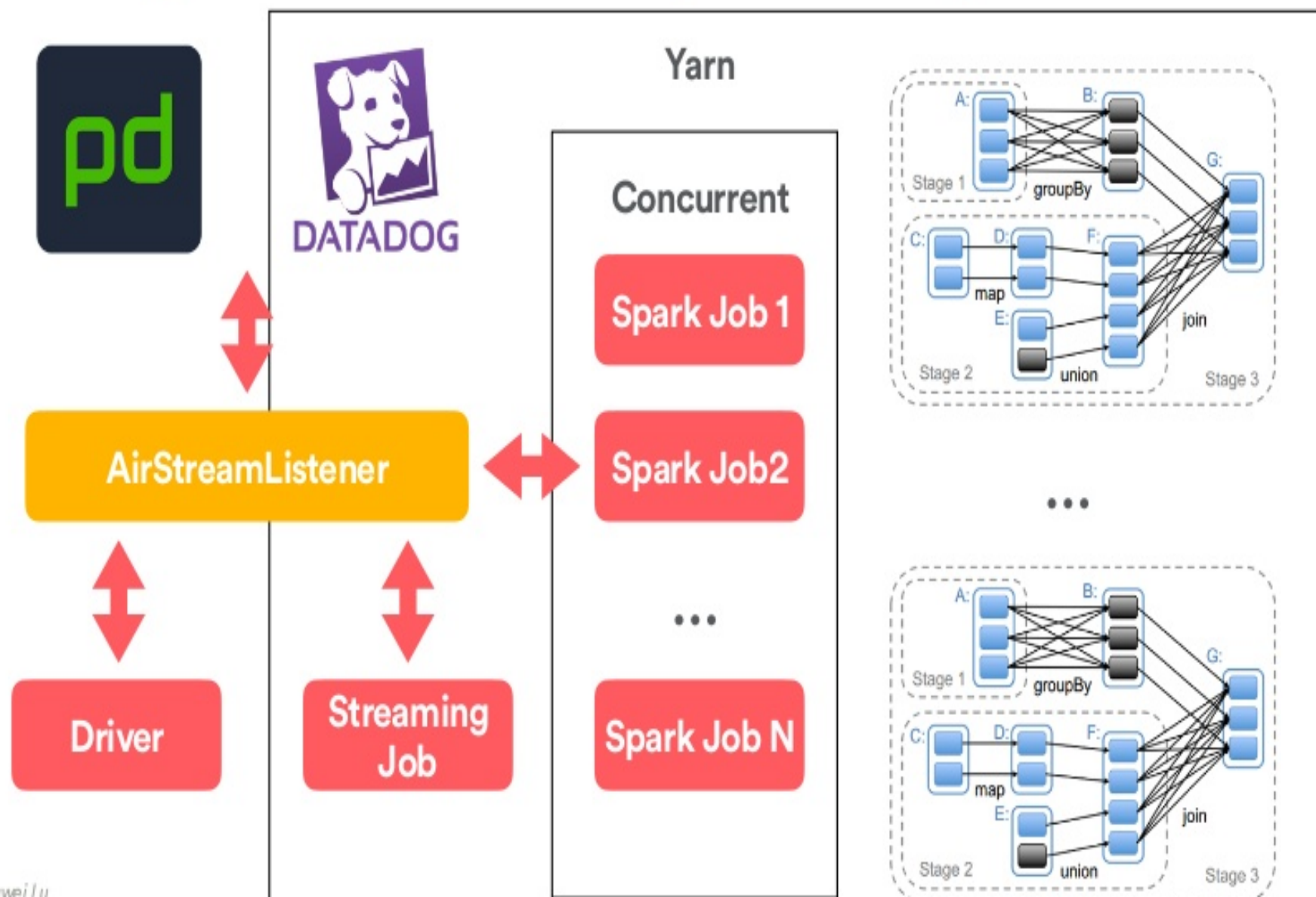
Job Management: Scaling up



Job Management: Fault Tolerant



Job Management: Monitoring & Alerting



Summary

Simplify and Unify Stream Batch Pipeline

Rich Stateful Computation

Rich Integration with Hadoop EcoSystem

Easy Operation

