Day 4
Practical data streaming (cont.)

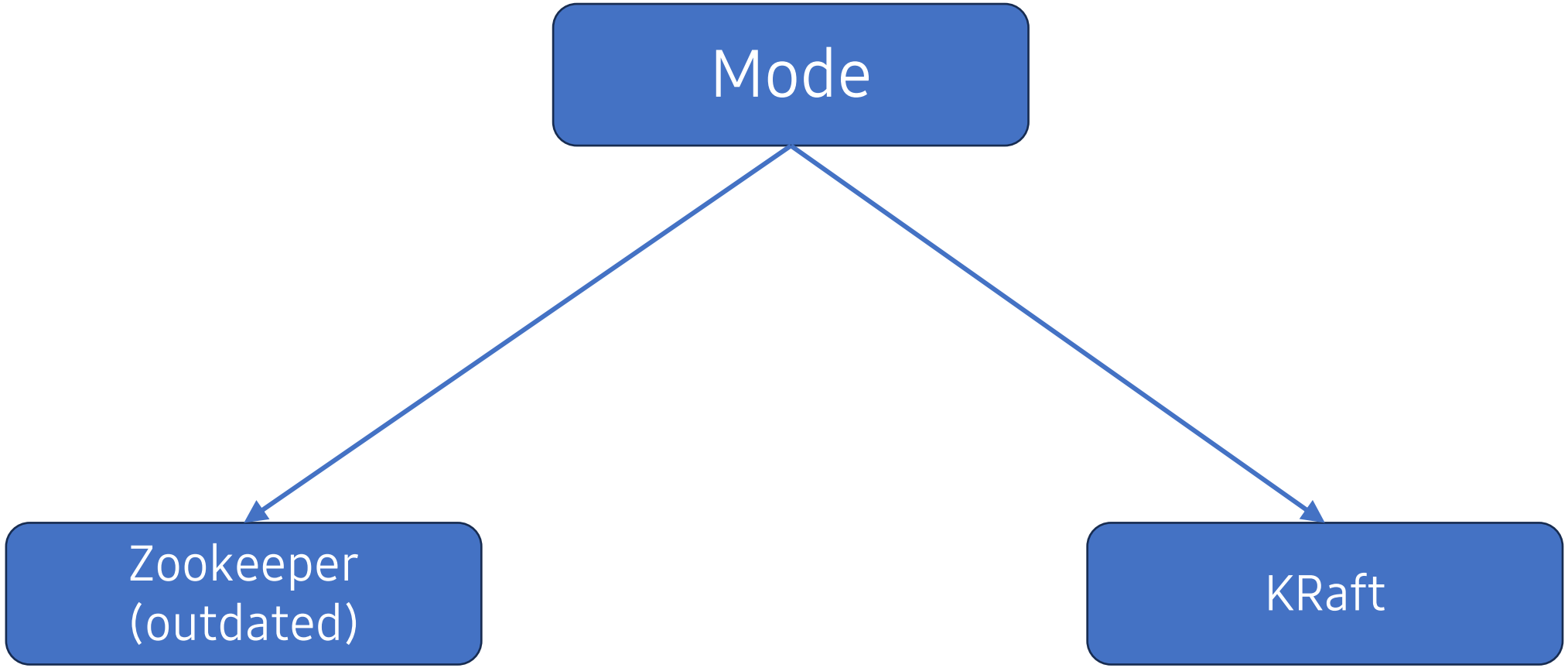
Lecturer: M.S. Le Minh Tan

Outline

I. Apache Kafka 💻
II. Apache Spark 💻

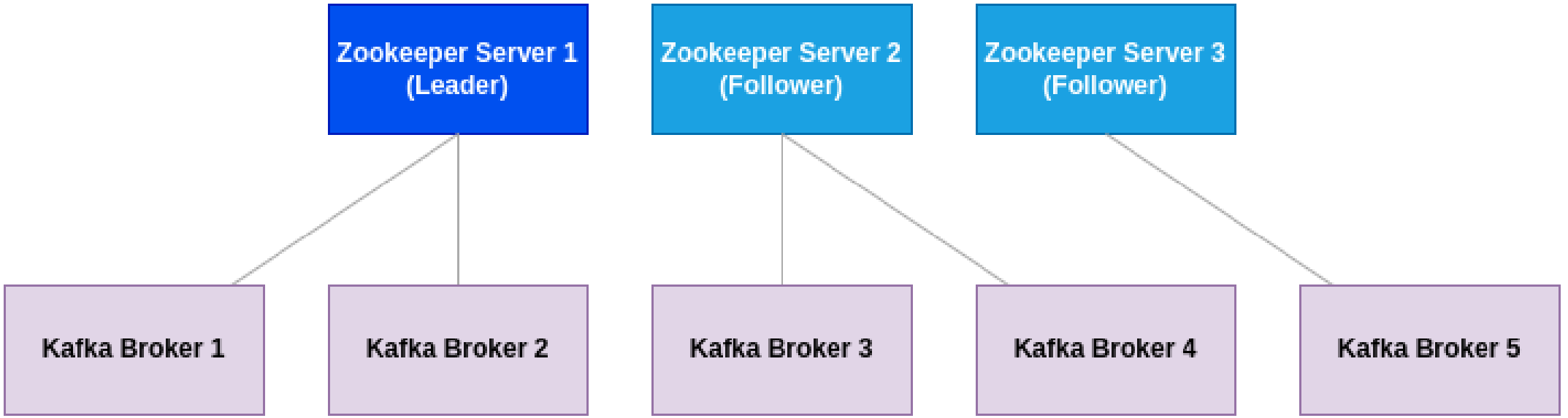
I. Kafka

- Kafka is a message broker
  - It has producer, broker, consumer.
- High throughput
- Log-based (not queue-based)
  - Messages stay in the log file(s) until retention condition is met
- Data as events, which are grouped by topics.


Mode
Zookeeper
(outdated)
KRaft

Zookeeper

- Zookeeper is a service for distribution.
  - Replicated
  - Leader – Follower model

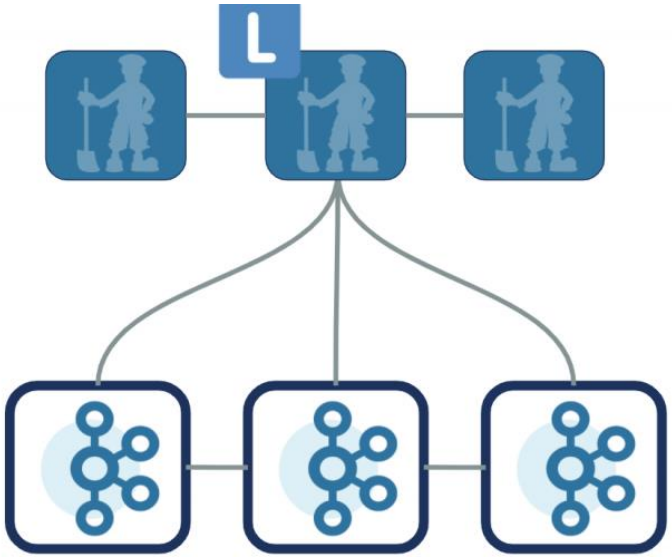


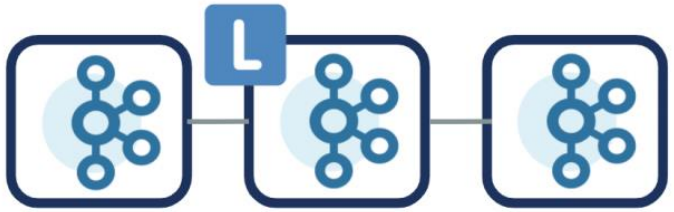
Problems

- Performance issue
  - Extra layer of management
  - Synchronous metadata propagation
- The data must be written to leader node
- Security

KRaft (Kafka Raft)

- Built for Kafka (native)
- Still leader – follower model: Leader, followers are brokers
- Some nodes (if not all) are quorum active controllers
- Quorum controllers are Zookeeper Server alternatives.





Step 1: Create environment

- Download and extract Kafka release tgz
- Generate cluster ID
    1. Can we use any name as cluster ID? Why?
- Create log directories
    2. Where's the log directories by default?
- Start Fafka server

https://kafka.apache.org/quickstart

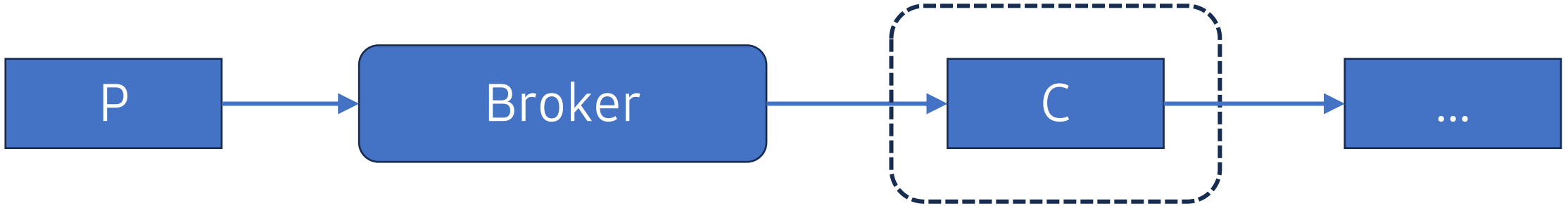
Step 2: Topic

- Create topic “quickstart-events”
- Write some events
- Read the events

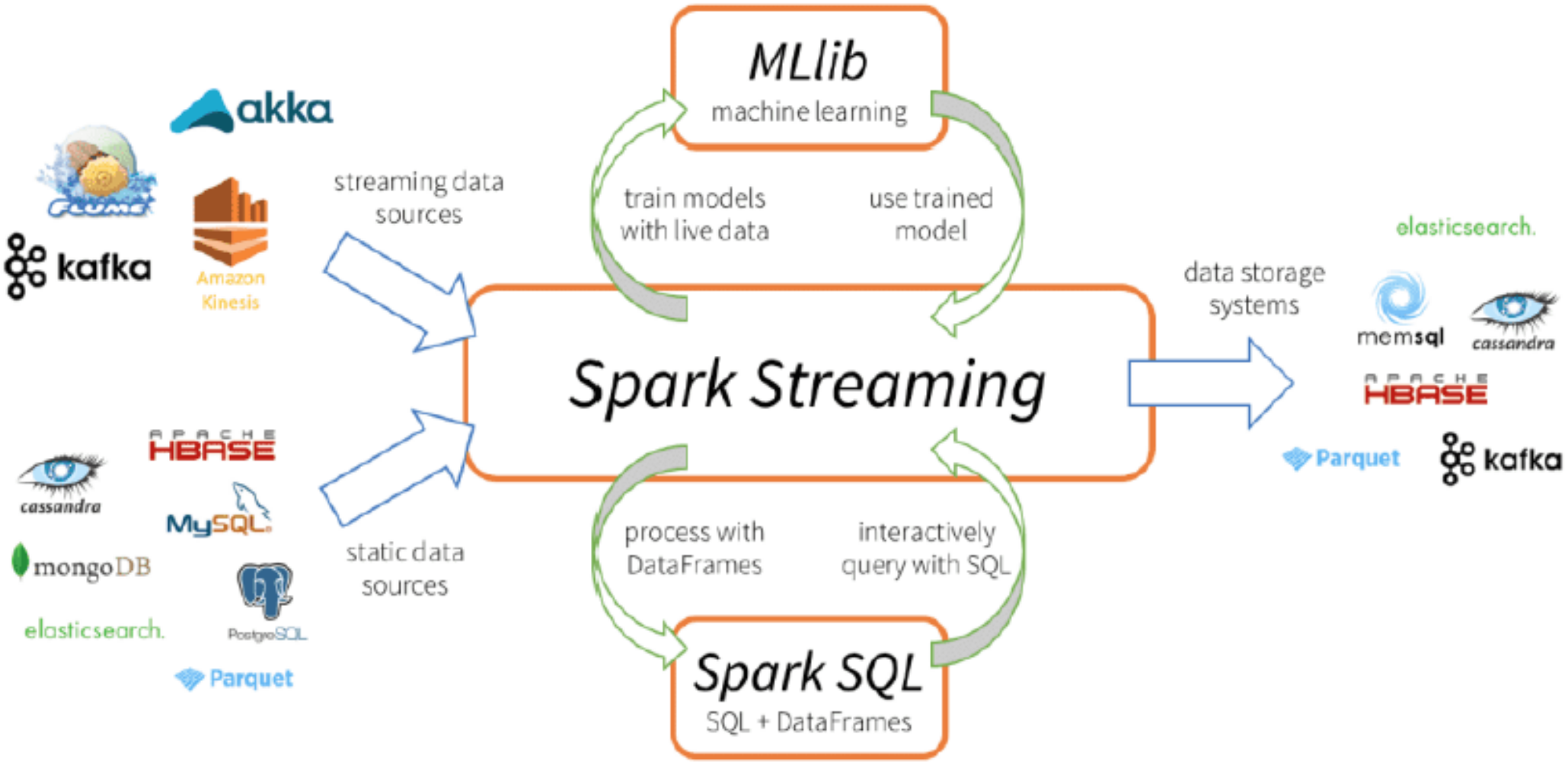
https://kafka.apache.org/quickstart

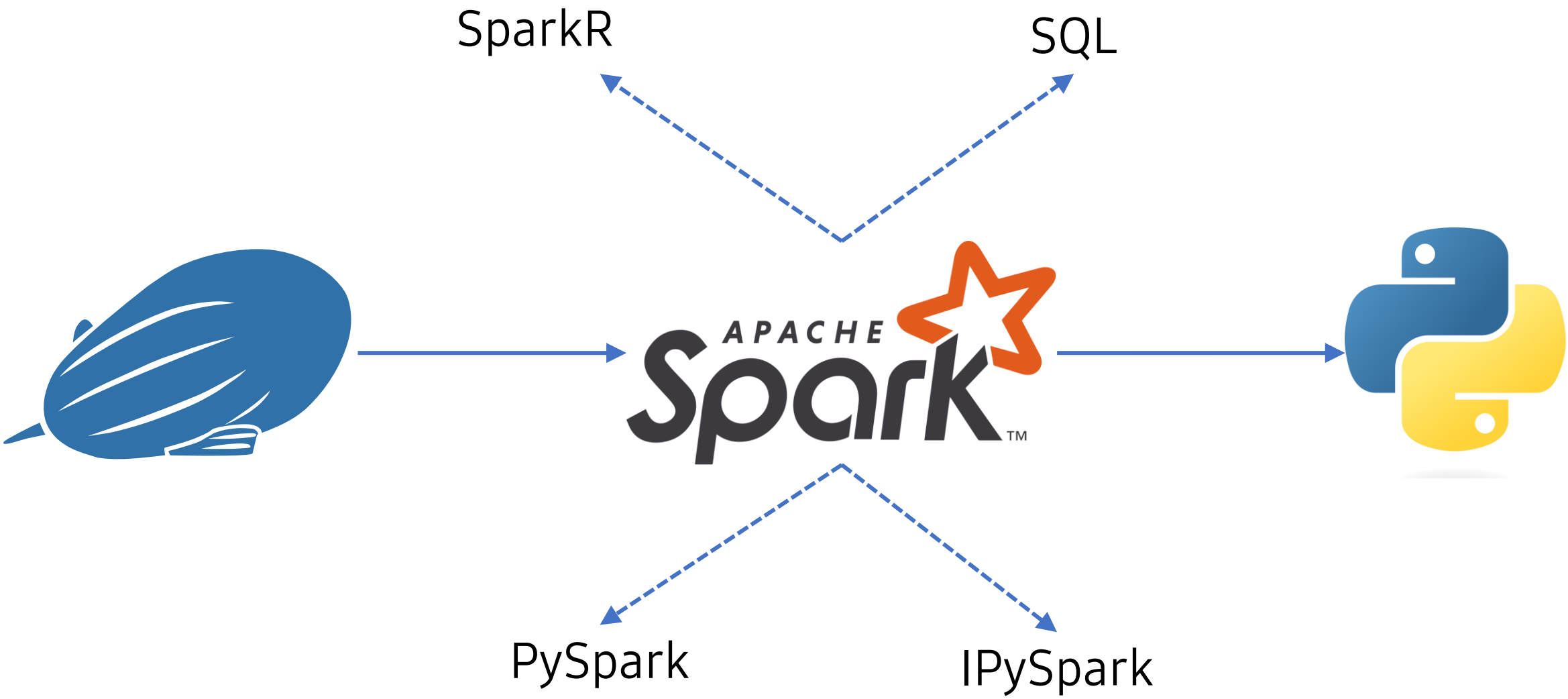
Streaming data source

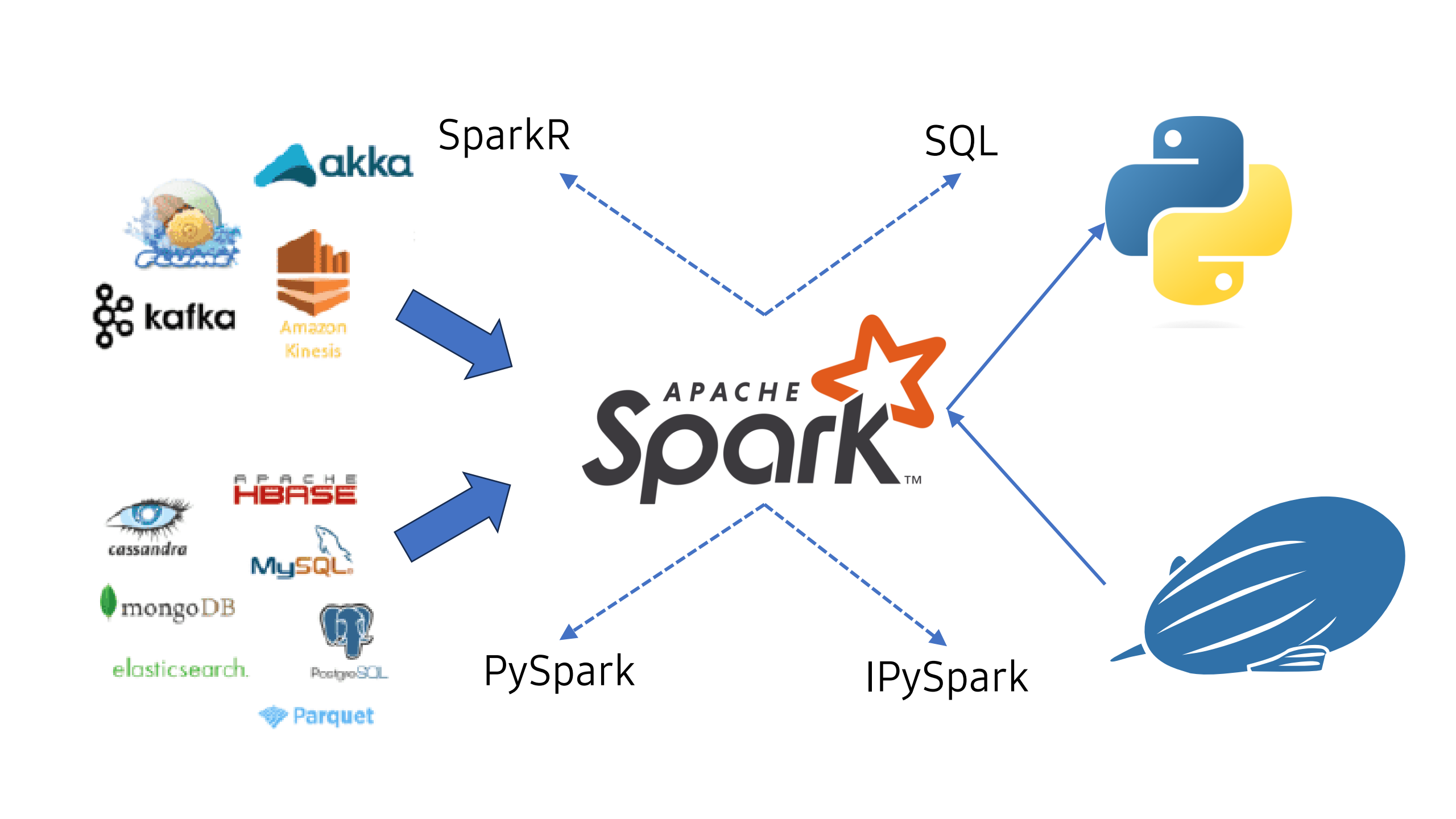
- Both RabbitMQ & Kafka are streaming platforms.
- Both producers and consumers are out of their scopes.



II. Real-time Spark

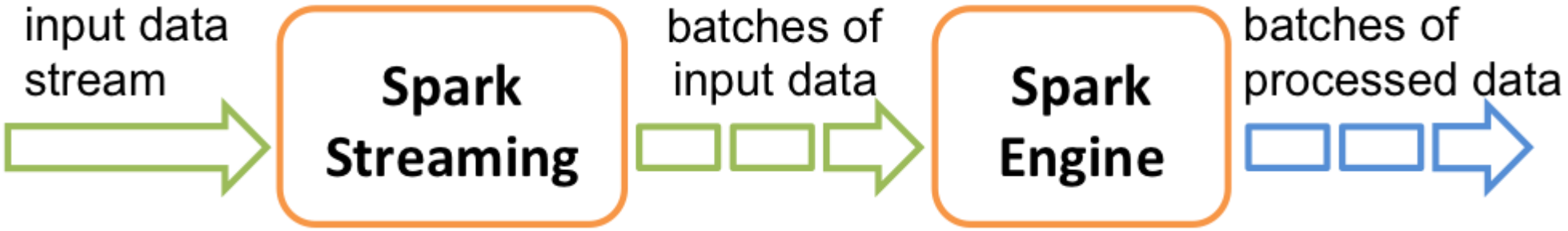


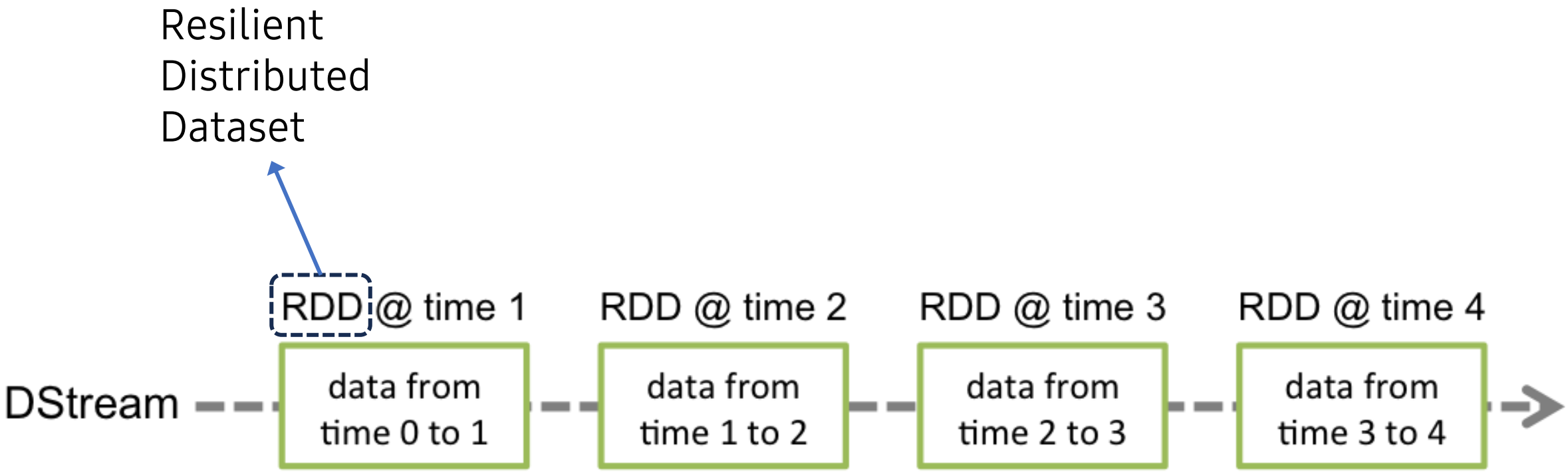

SparkR
SQL
APACHE
Spark
PySpark
IPySpark


akka
kafka
Amazon Kinesis
APACHE HBASE
cassandra
MySQL
mongoDB
elasticsearch
PostgreSQL
Parquet
SparkR
SQL
APACHE Spark™
PySpark
IPySpark

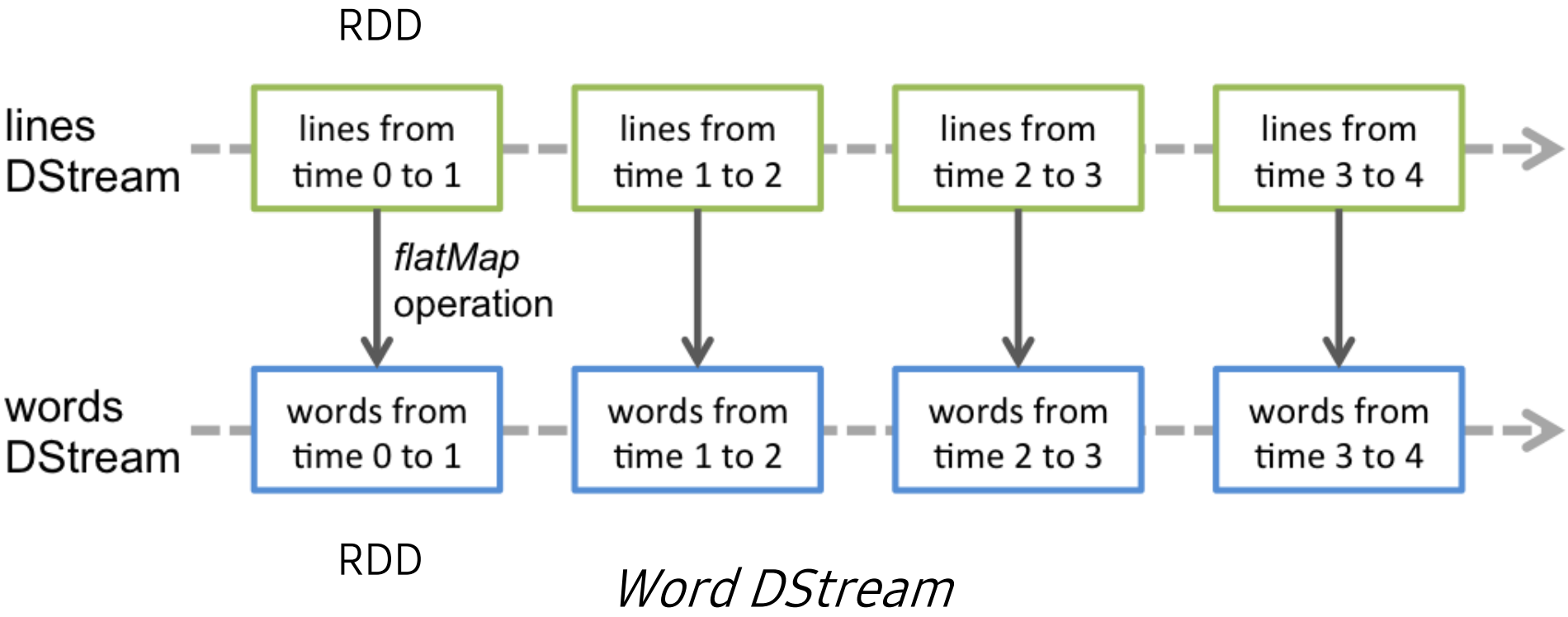
Concepts

- Spark Streaming
- Spark Engine
- SparkContext
- StreamingContext
- SQLContext
- Discretized Stream (DStream)





Abstract DStream



https://spark.apache.org/docs/latest/streaming-programming-guide.html#a-quick-example

Homework

- Do the Netcat + Spark example.
- Record a video demonstrating the example.