



Welcome to Spark Streaming



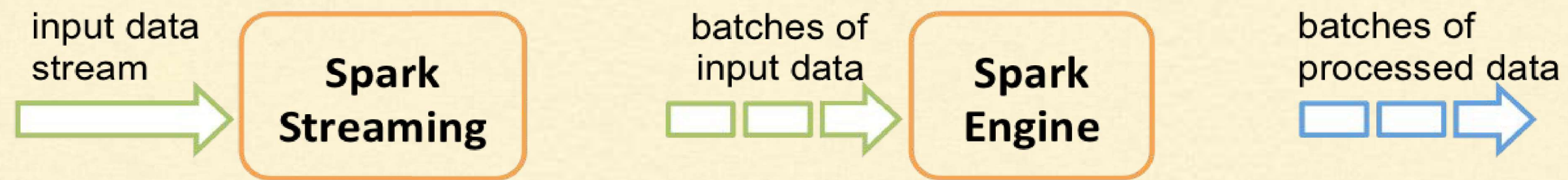
Spark Streaming - Introduction

- Extension of the core Spark API
- Enable Stream Processing
 - Scalable
 - High-throughput
 - Fault-tolerant

Spark Streaming - Introduction



Spark Streaming - Workflow

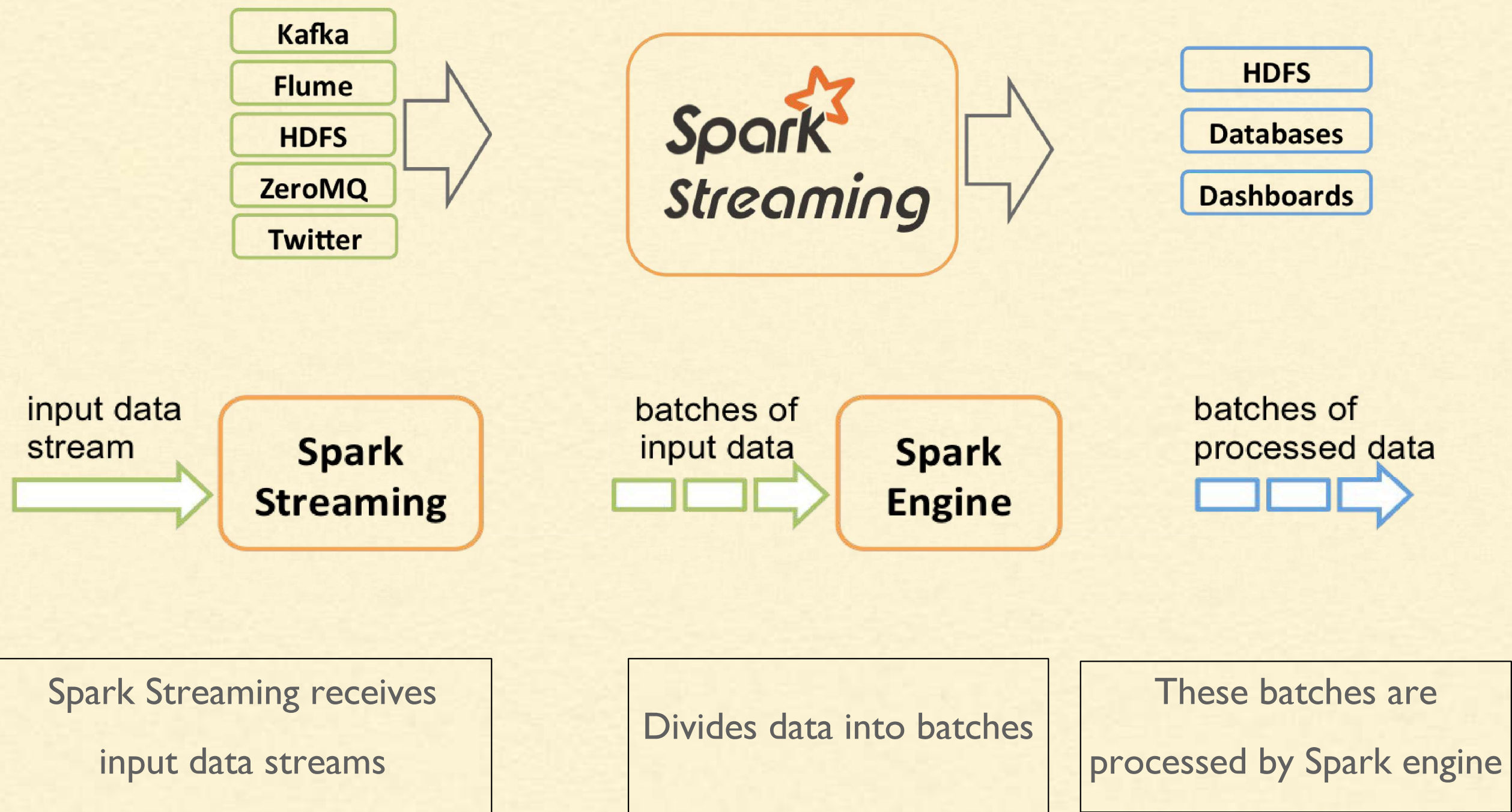


Spark Streaming
receives
input data streams

Divides data into
batches

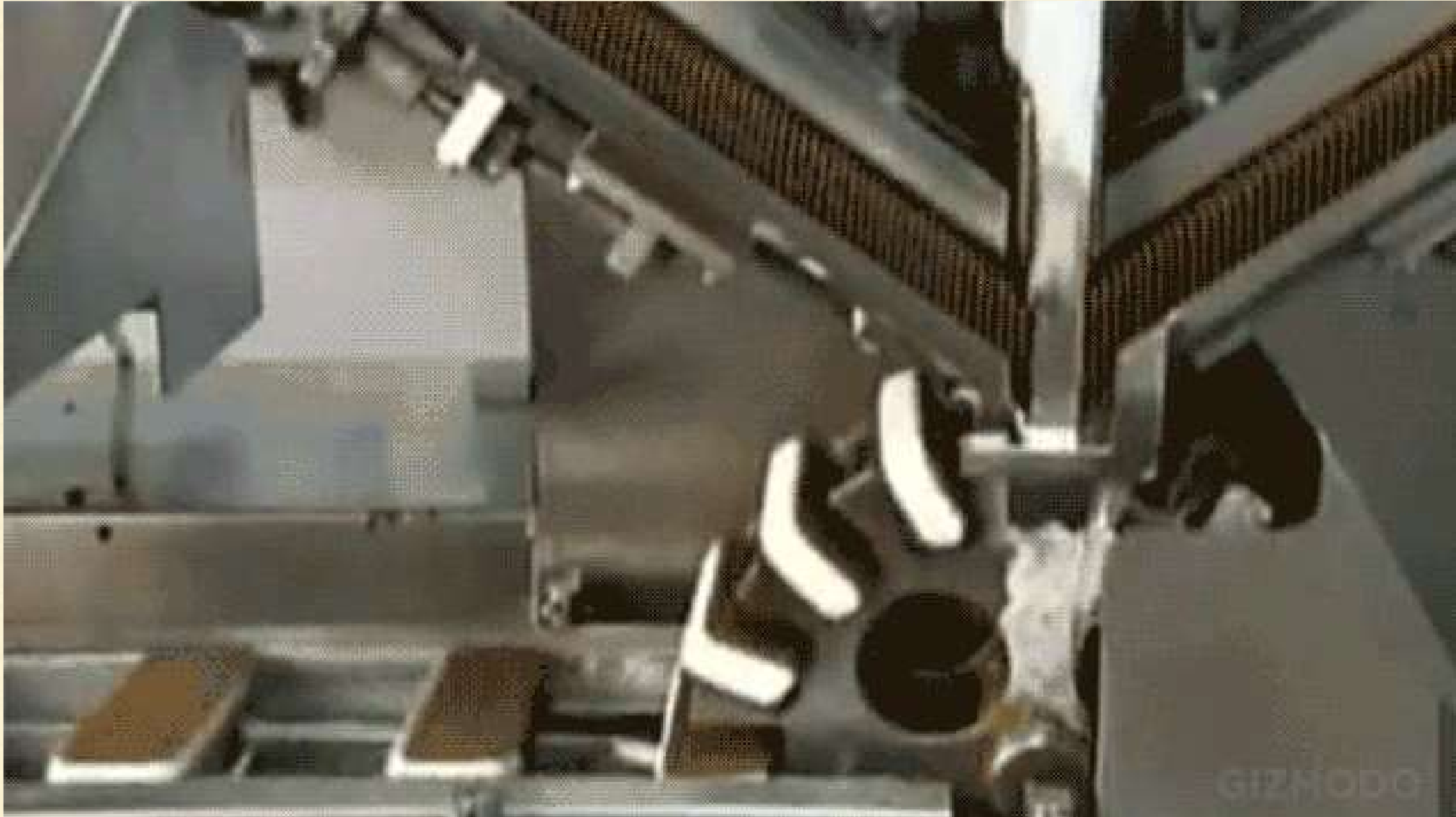
These batches are
processed by Spark
engine

Spark Streaming - Workflow



Spark Streaming - Workflow

Continuous Stream



Batches of Data

Spark Streaming - Use Case



Real-time Analytics

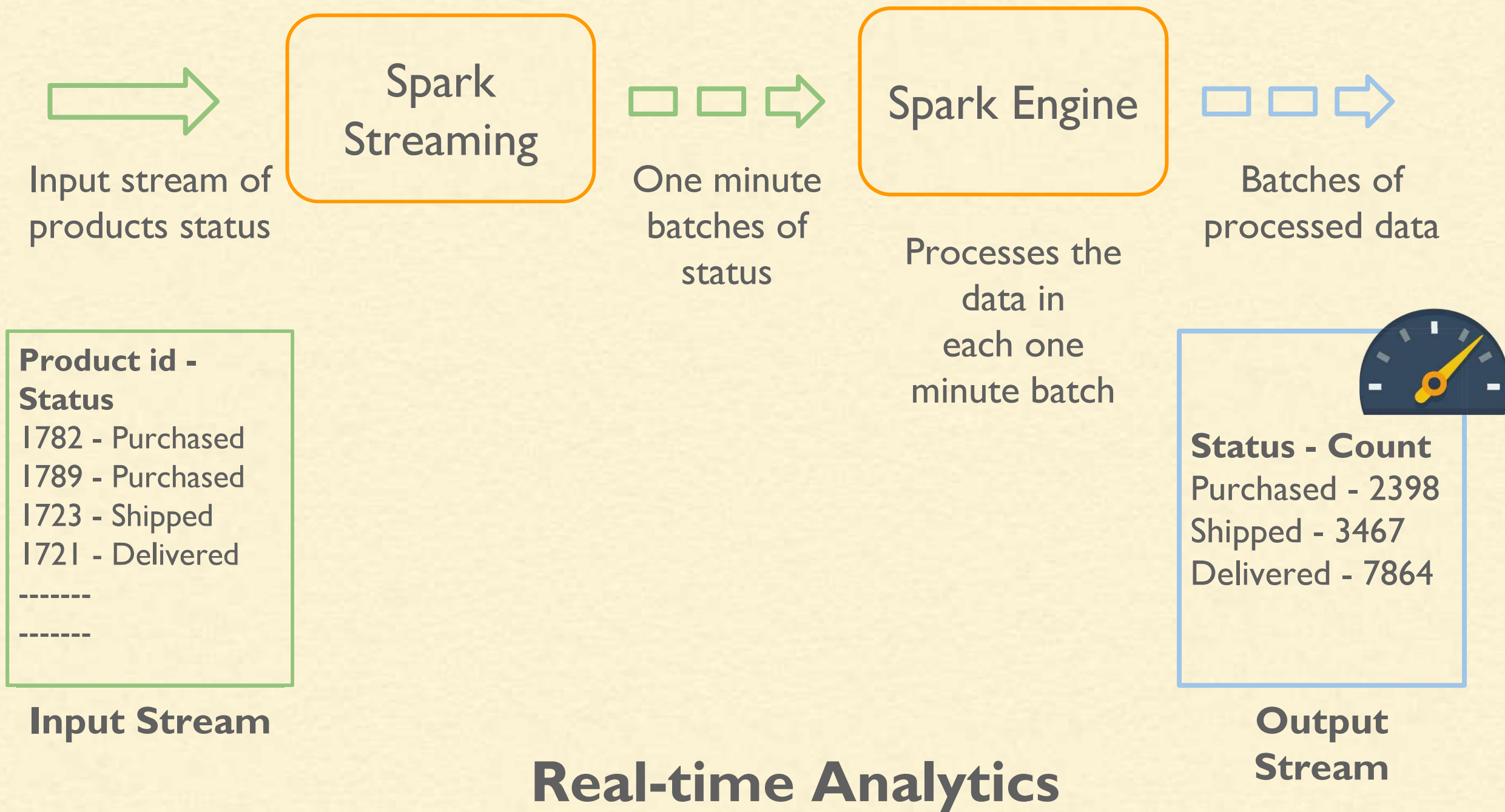
Spark Streaming - Use Case

Real-time Analytics

Problem: Build real-time analytics dashboard with information on how many products are getting

- Purchased
- Shipped
- Delivered every minute

Spark Streaming - Use Case - Ecommerce



Spark Streaming - Use Case



Negative



Neutral



Positive

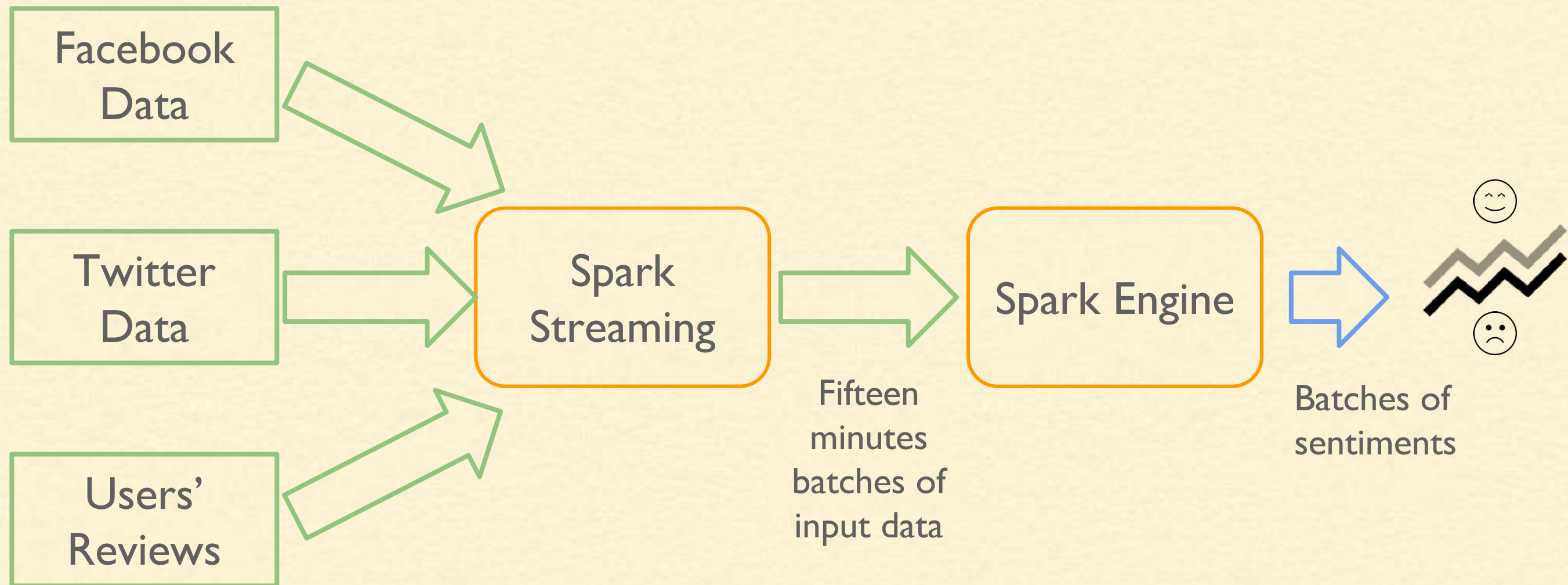
Real-time Sentiment Analysis

Spark Streaming - Use Case

Real-time Sentiment Analysis

Problem: Build Real-time sentiment analytics system to find out sentiment of users every fifteen minute by analyzing data from various sources such as Facebook, Twitter, users' feedback, comments and reviews

Spark Streaming - Use Case



Real-time Sentiment Analysis

Spark Streaming - Use Case

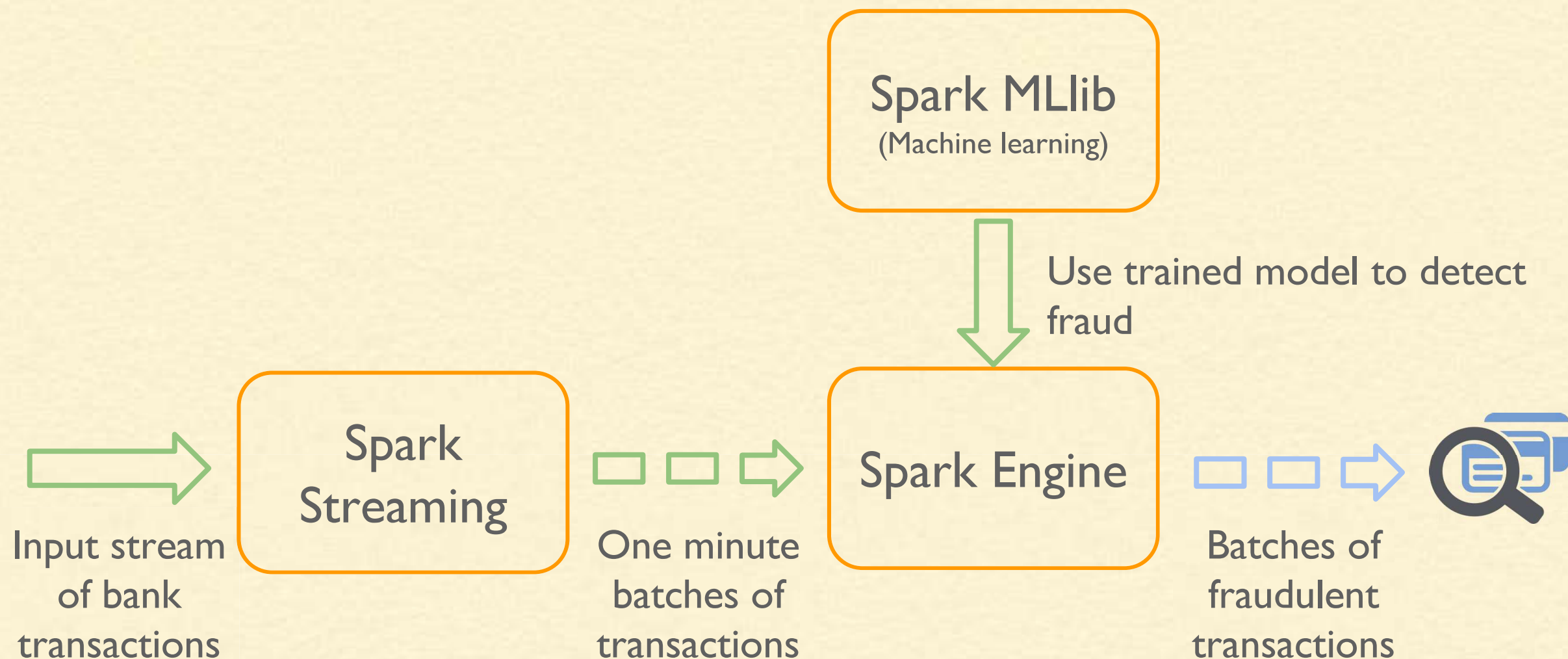
Real-time Fraud Detection

Problem: Build a real-time fraud detection system for a bank to find out the fraudulent transactions



Spark Streaming - Use Case

Real-time Fraud Detection



Spark Streaming - Use Cases - Uber



U B E R

Uber uses Spark Streaming for real-time telemetry analytics

Spark Streaming - Use Cases - Pinterest



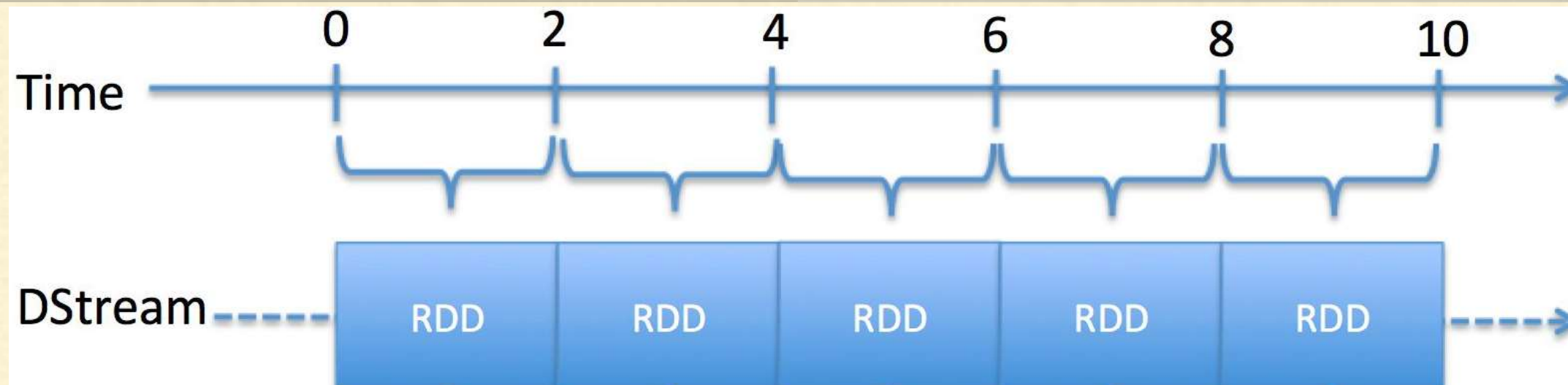
Pinterest uses Spark Streaming to provide immediate insight into how users are engaging with pins across the globe in real-time

Spark Streaming - Use Cases - Netflix

NETFLIX

Netflix uses Spark Streaming to provide movie recommendations to its
users

Spark Streaming - DStream



Discretized stream or DStream:

- Represents continuous stream of data
- Can be created
 - From input data streams
 - Or from other DStreams
- Is represented as a sequence of RDDs internally
- Is used to write Spark Streaming programs

Spark Streaming - Hands-on - Word Count

Problem: Count the number of words in the text data received from a server listening on a host and a port

Spark Streaming - Hands-on - Word Count

