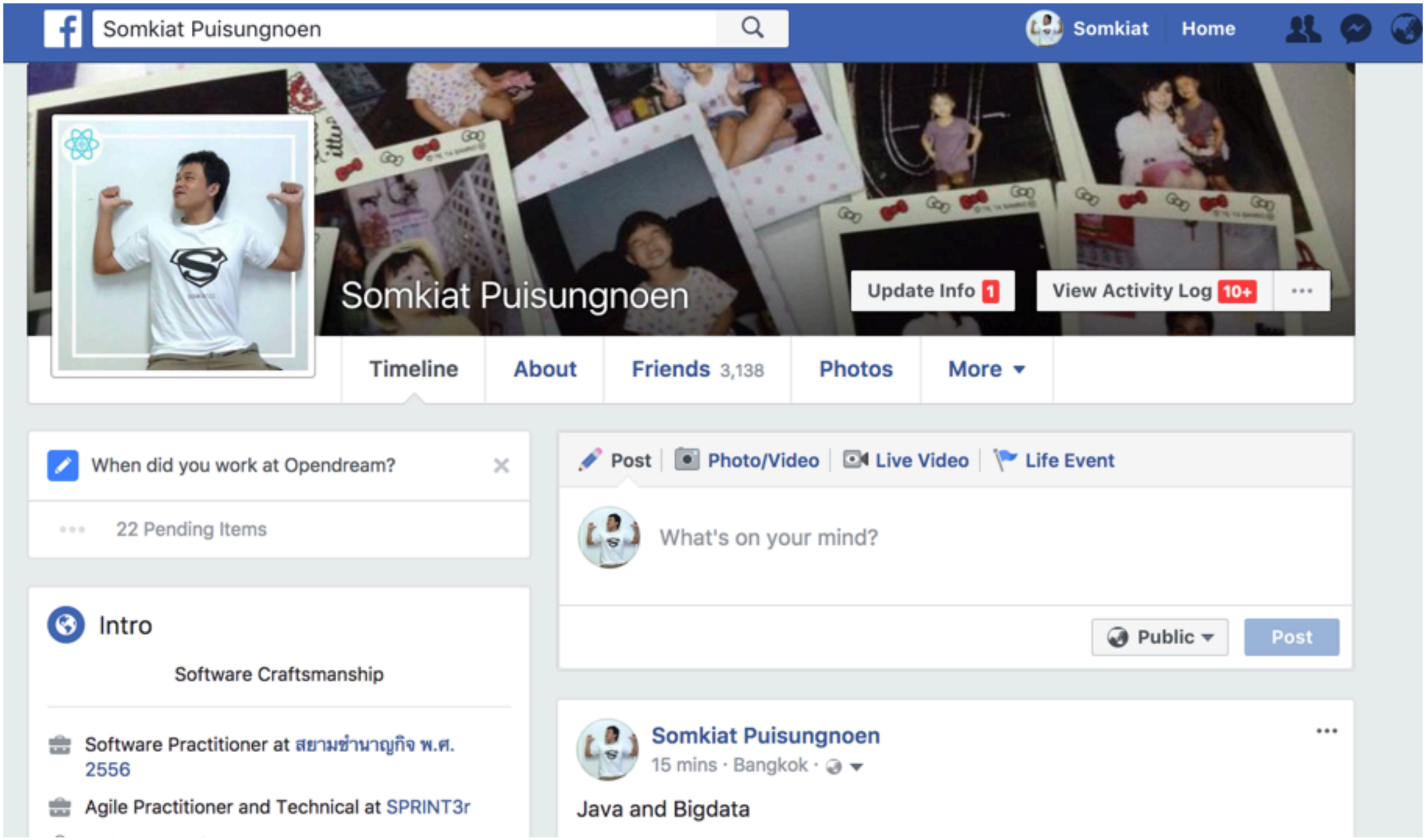
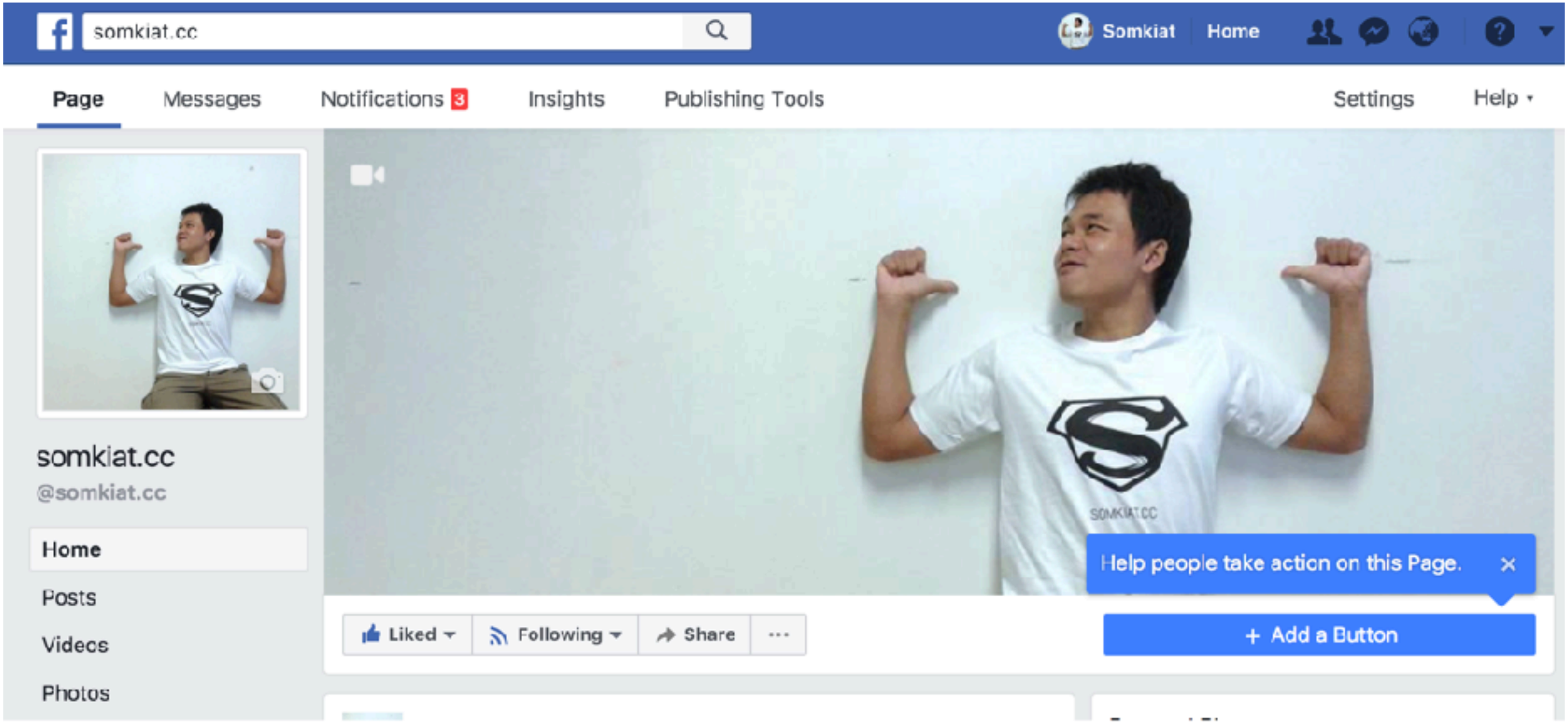
สยามชำนาญกิจ
SIAM CHAMNAN KIT

Solution Architect





https://github.com/up1/course-rabbitmq-2025

https://www.rabbitmq.com/

Topics

Messaging queue
Introduction to RabbitMQ
Use cases
Architecture of RabbitMQ
RabbitMQ patterns
Clustering
Monitoring and observability
Development

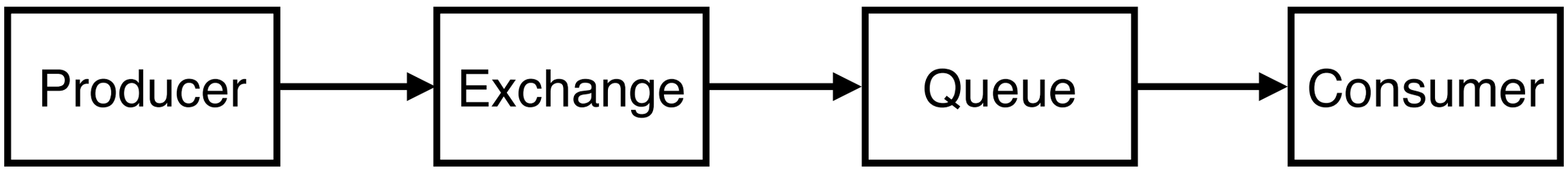
Messaging Queue

Introduction to RabbitMQ

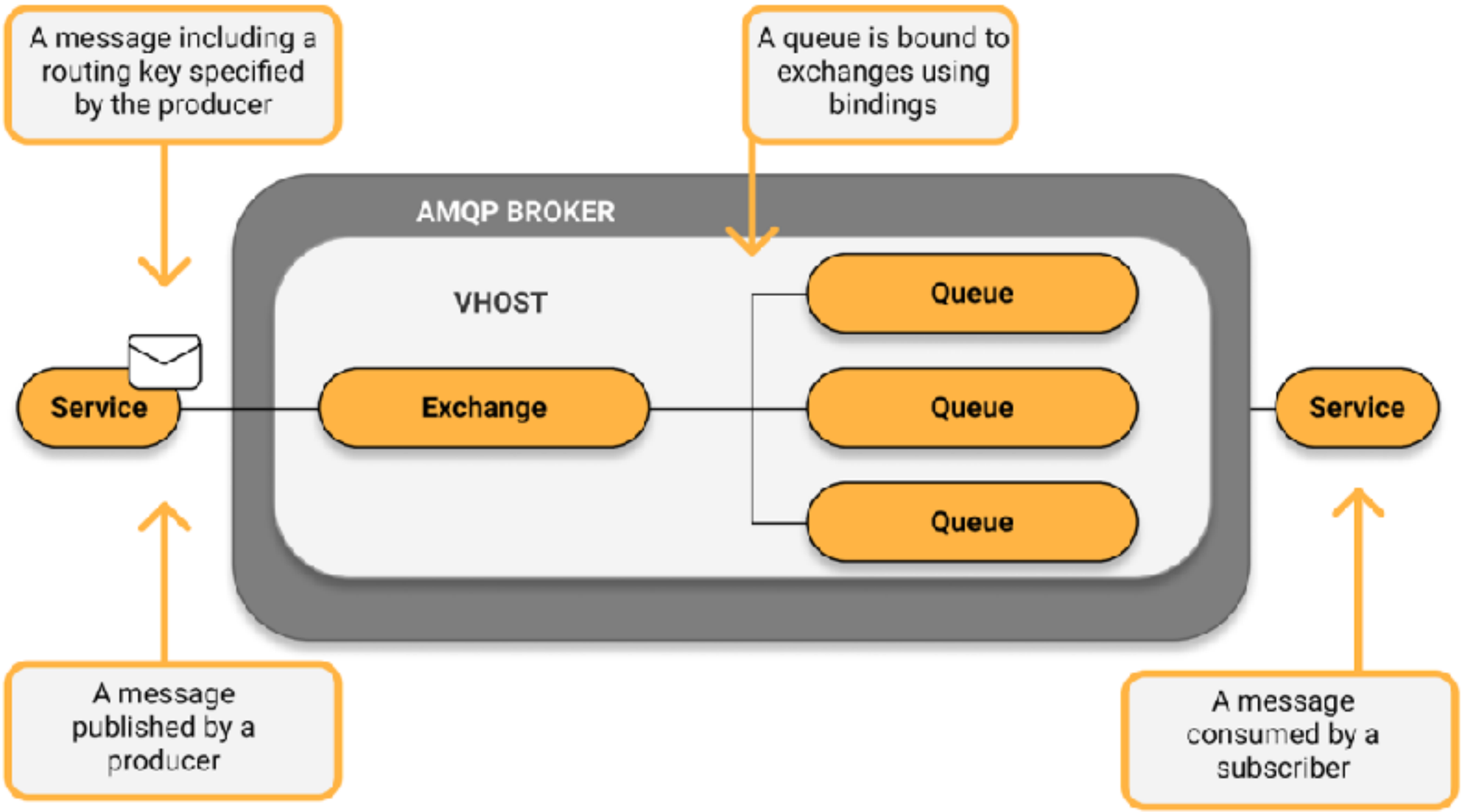
How RabbitMQ works ?

RabbitMQ works

Queue-based messaging model
Messages are published by **producers**
Routed by **exchange**
Delivery to **consumers** through **queues**
Queue is a buffer that stores messages



RabbitMQ works



Queue ?

Storage to store **ordered** collection of messages
Allow enqueue and dequeue
FIFO (First-in First-out)
Stored messages until a consumer is available to process

Durable

Exclusive

Auto-delete

Queue parameters

Parameter	Description	Use case
durable	**true** = queue will survive a broker restart	Persistent work queue where don't want data loss
exclusive	**true** = queue is restricted to the connection that declare it will delete when connection loss	One queue per connection
auto-delete	**true** = queue is automatically deleted when consumer unsubscribed	Temporary notification queue

Queue arguments (optional)

Name	Description	Use case
x-queue-mode	**default = in-memory** lazy = pages to disk	Very large queue Memory pressure Long running queue
x-queue-mode	Time(in ms) for queue may be unused Before auto-deleted	Cleanup unused queue
x-max-length	Maximum number of messages in queue	Keep only the latest N datas
x-message-ttl	Time(in ms) for message live in queue, before discarded or dead-lettered	Message expiration Retry message
x-max-priority	Enable prioritized queue (0 to N)	High priority jobs Critical alert

Queue Types ?

Classic (default)
Quorum
Stream

Classic queue

Classic (default)
FIFO (First-in, First-out) message delivery
Suitable for non-critical application
General purpose and background processing

Data loss is not major concern

https://www.rabbitmq.com/docs/quorum-queues

Quorum queue

Provide high availability and data safety
Replication using Raft consensus algorithm
Replicated data across multiple nodes

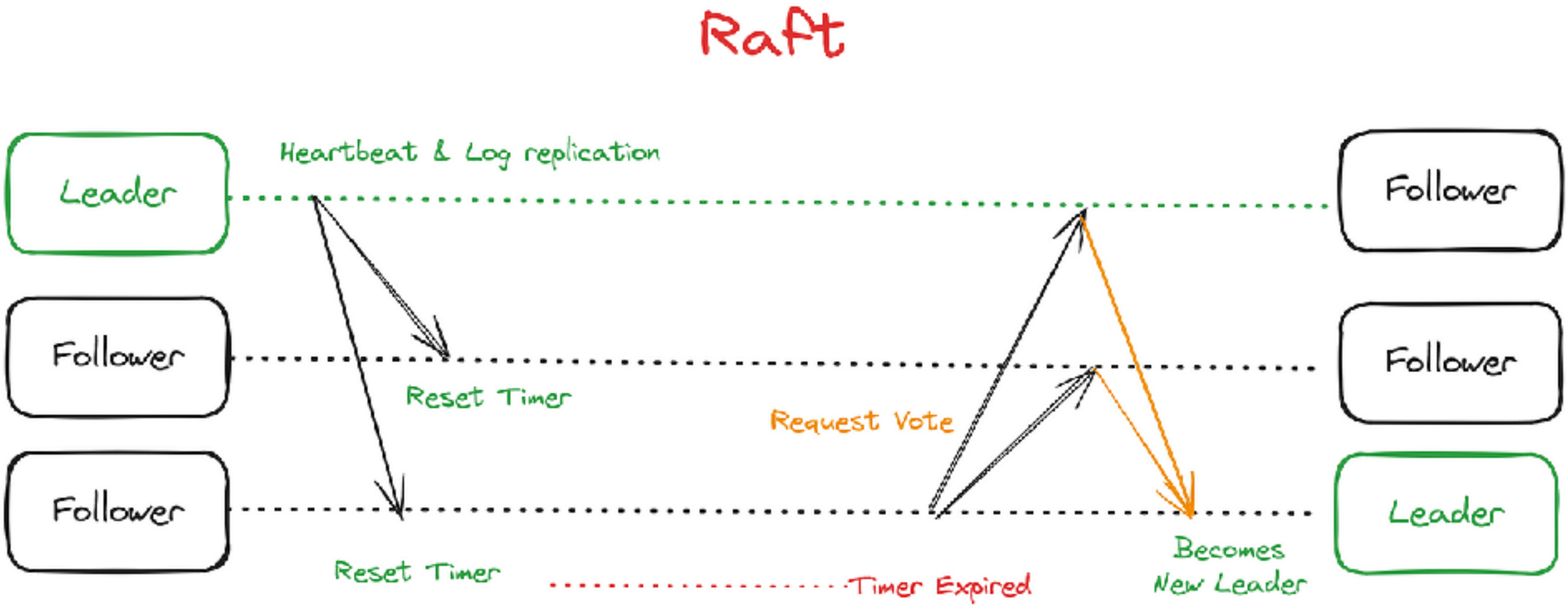
High latency !!

Data loss is unacceptable
(Financial data and sensitive data)

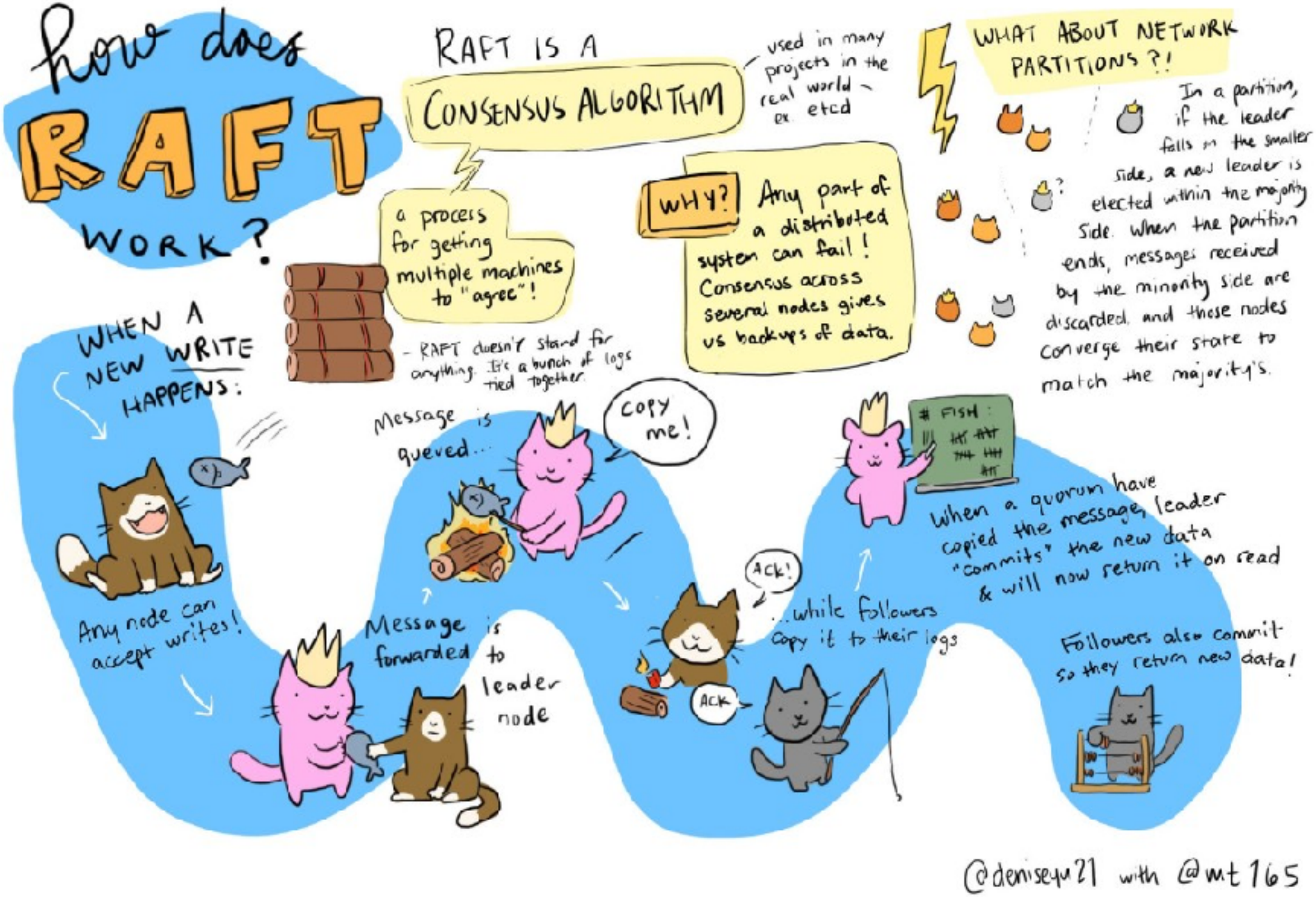
(N/2)+1

https://www.rabbitmq.com/docs/quorum-queues

Raft consensus ?



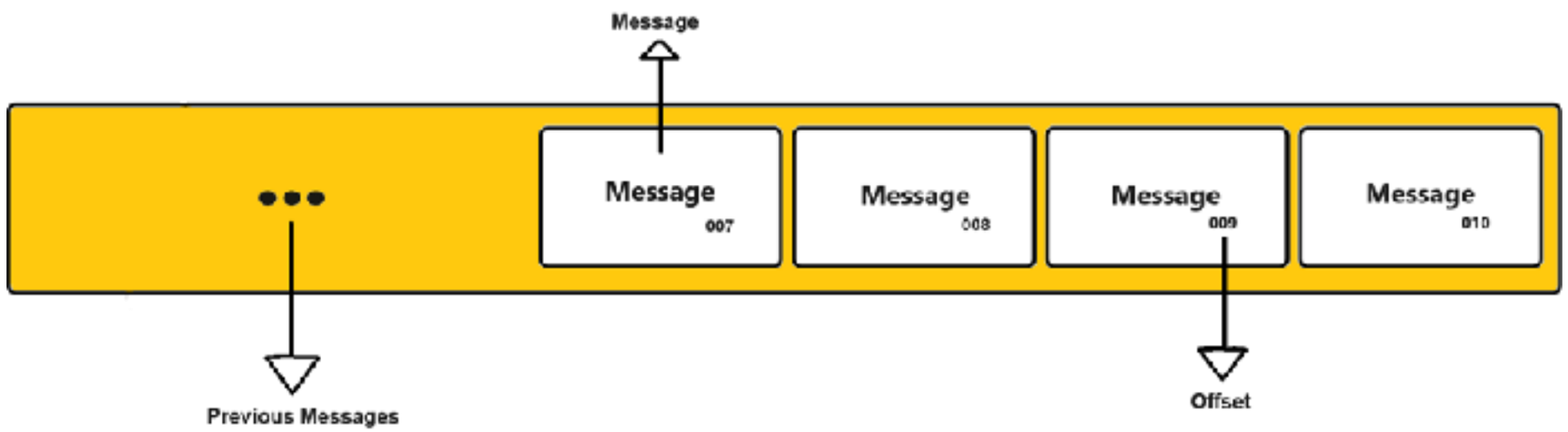
https://en.wikipedia.org/wiki/Raft_(algorithm)



https://en.wikipedia.org/wiki/Raft_(algorithm)

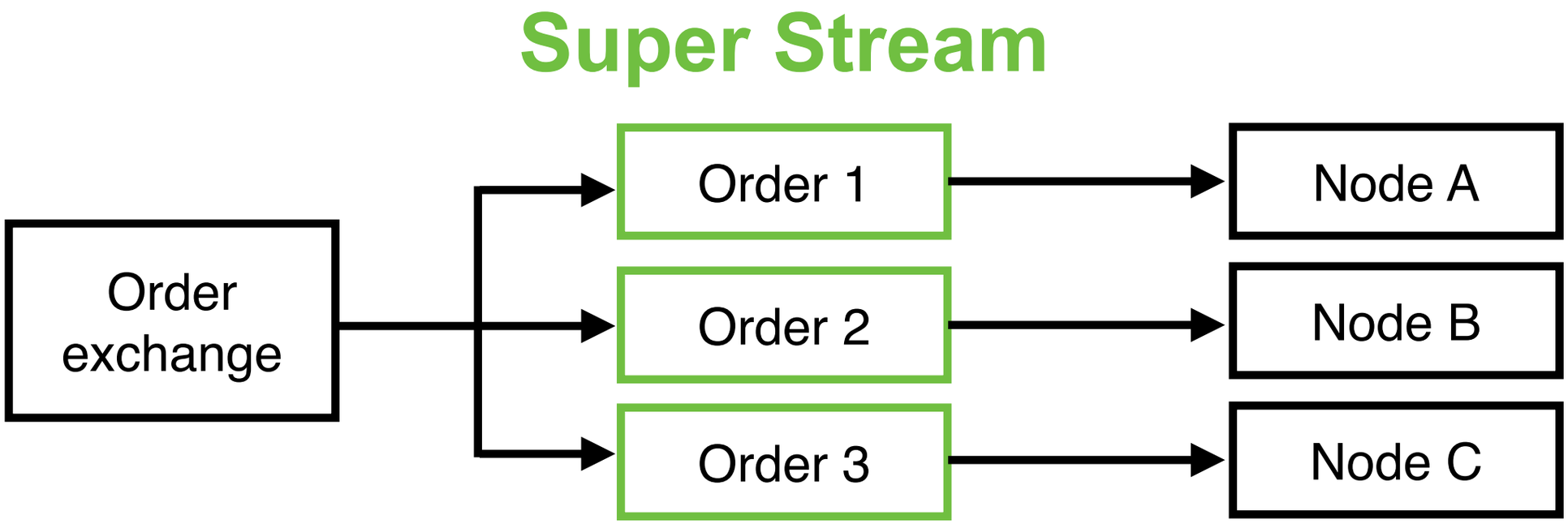
Stream

Design for high throughput
Real-time data stream
Message replay
Append-only log



Super Streams

Way to scale out by partitioning a **large stream** to small
Split storage and traffic on multiple nodes in cluster

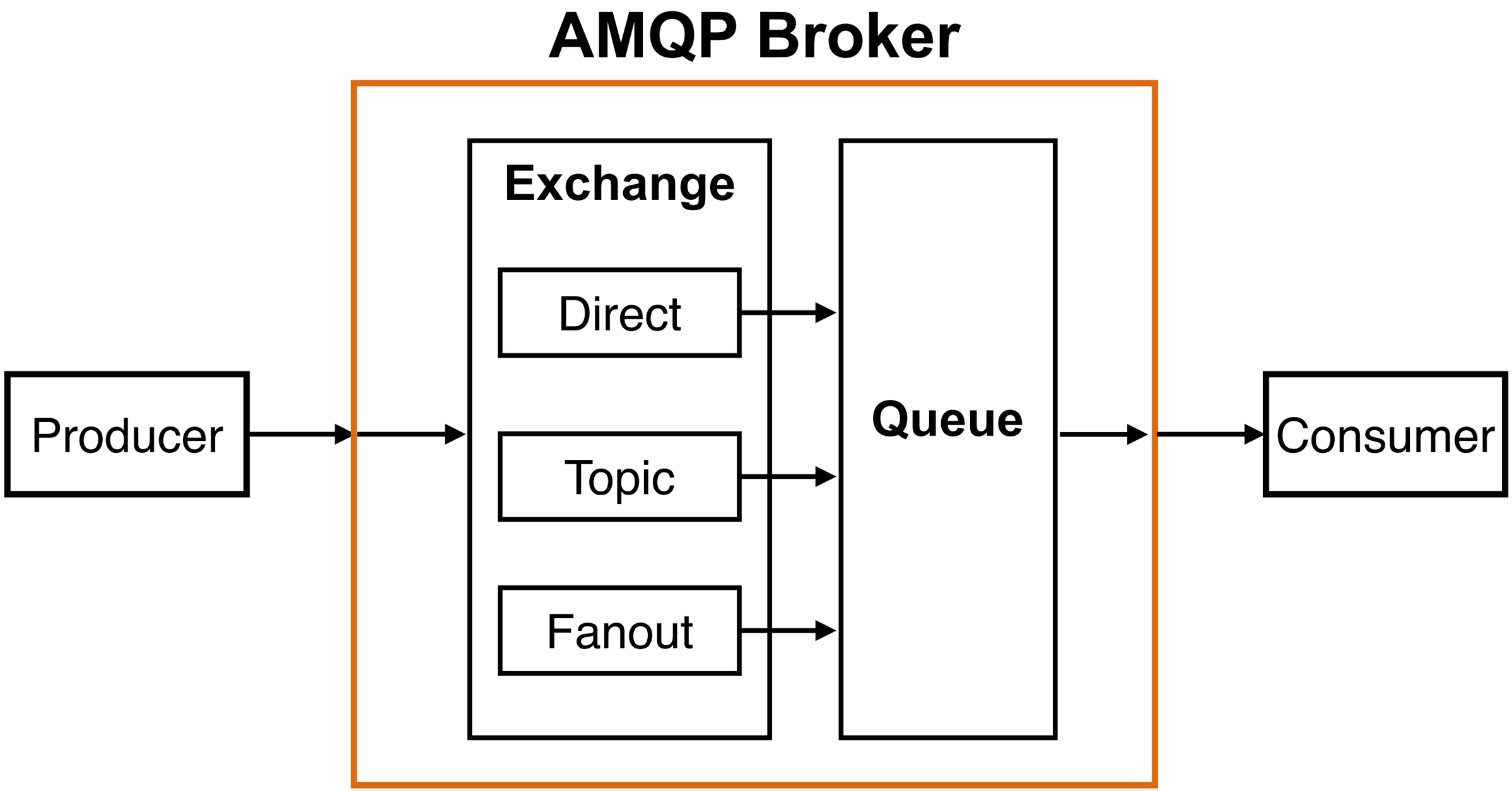


https://www.rabbitmq.com/blog/2022/07/13/rabbitmq-3-11-feature-preview-super-streams

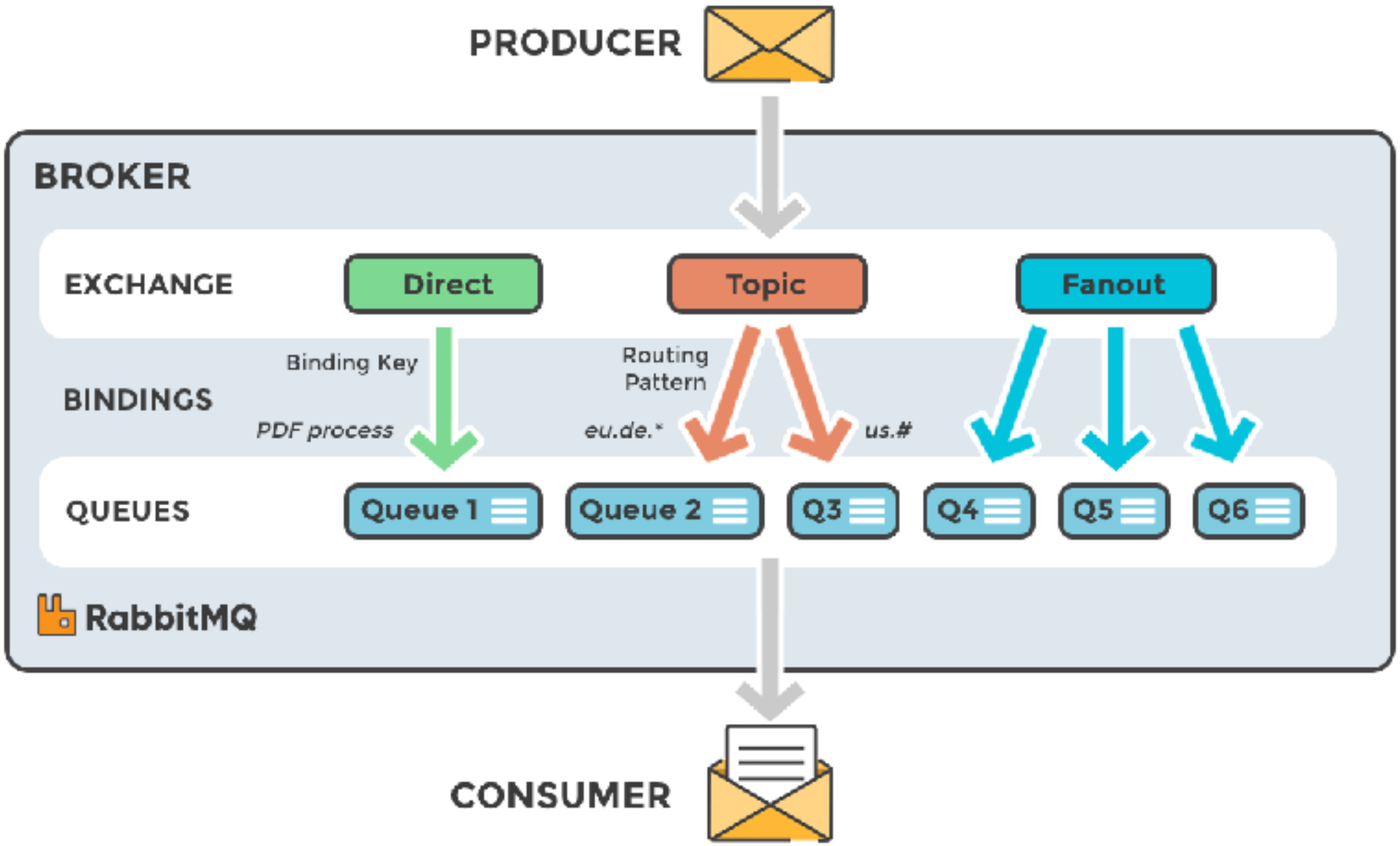
Working with Classic queue

RabbitMQ works

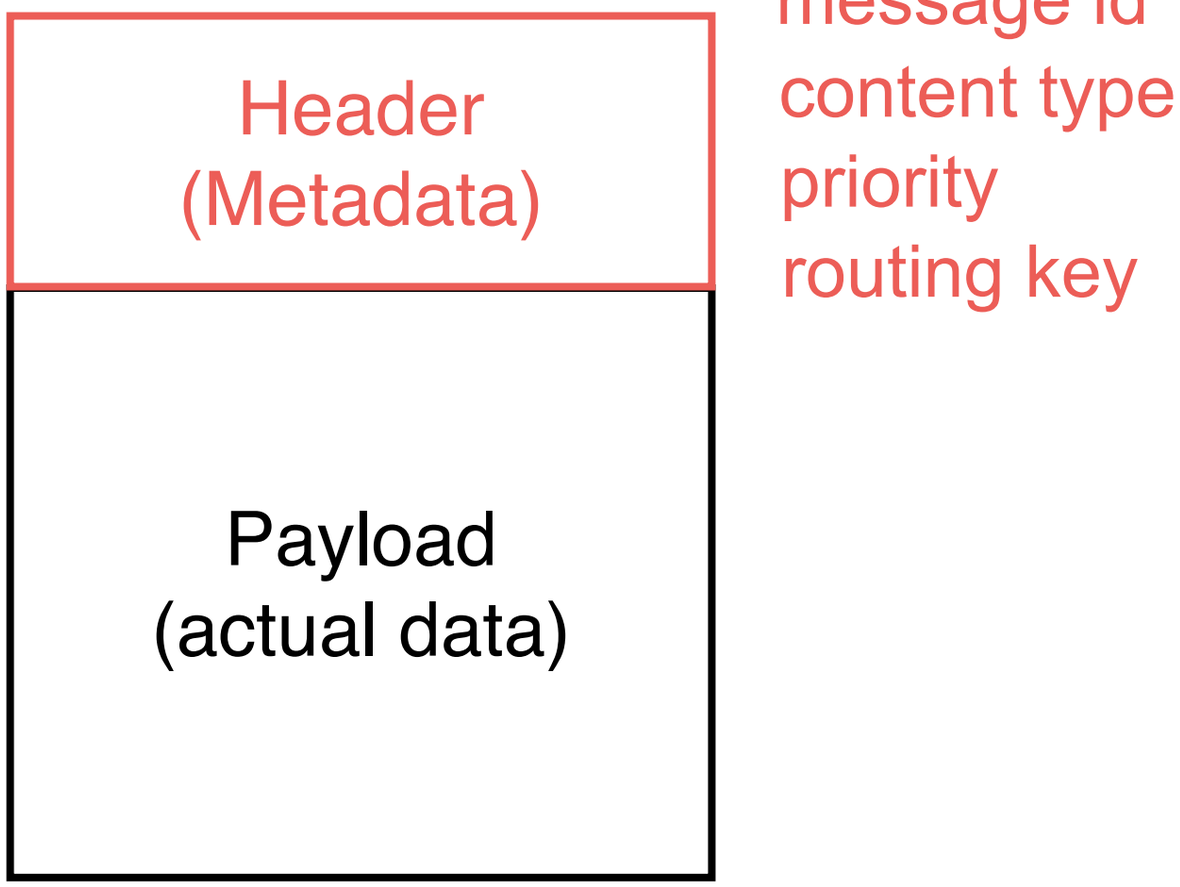
RabbitMQ works



RabbitMQ works

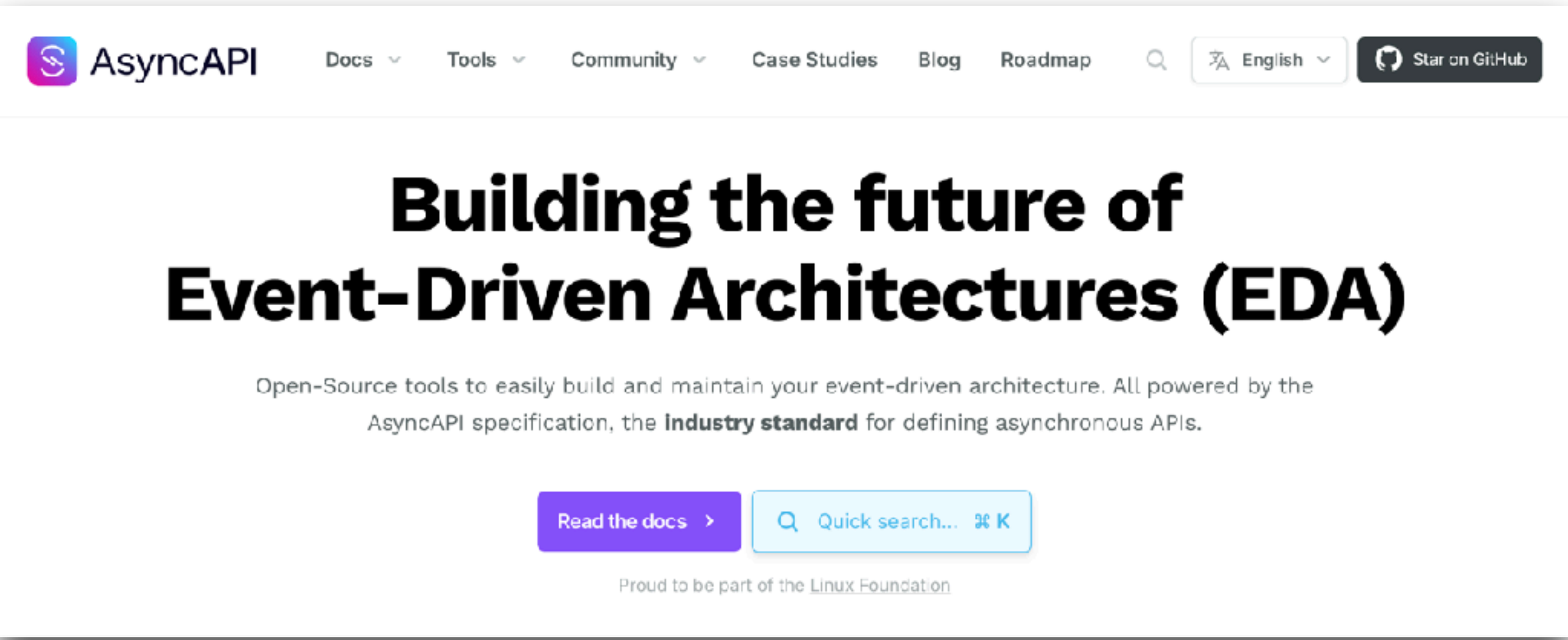


Message structure



https://www.rabbitmq.com/tutorials/amqp-concepts#messages

AsyncAPI



https://www.asyncapi.com/en

Exchange ?

Message routing agent within RabbitMQ
Receive message from producer and route to queue
Different exchange types handle routing key
in different way

Exchange types

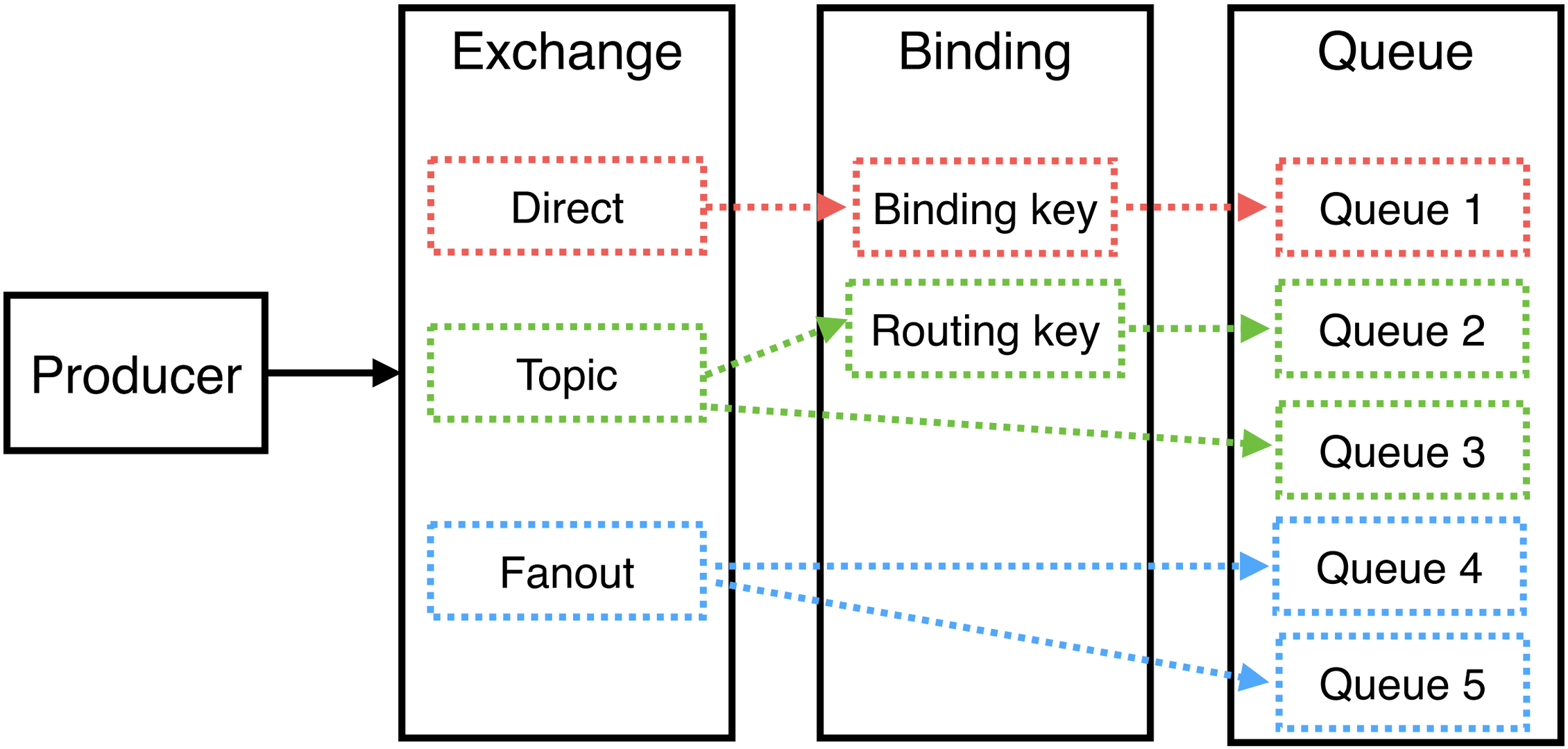
Direct

Fanout

Topic

Header

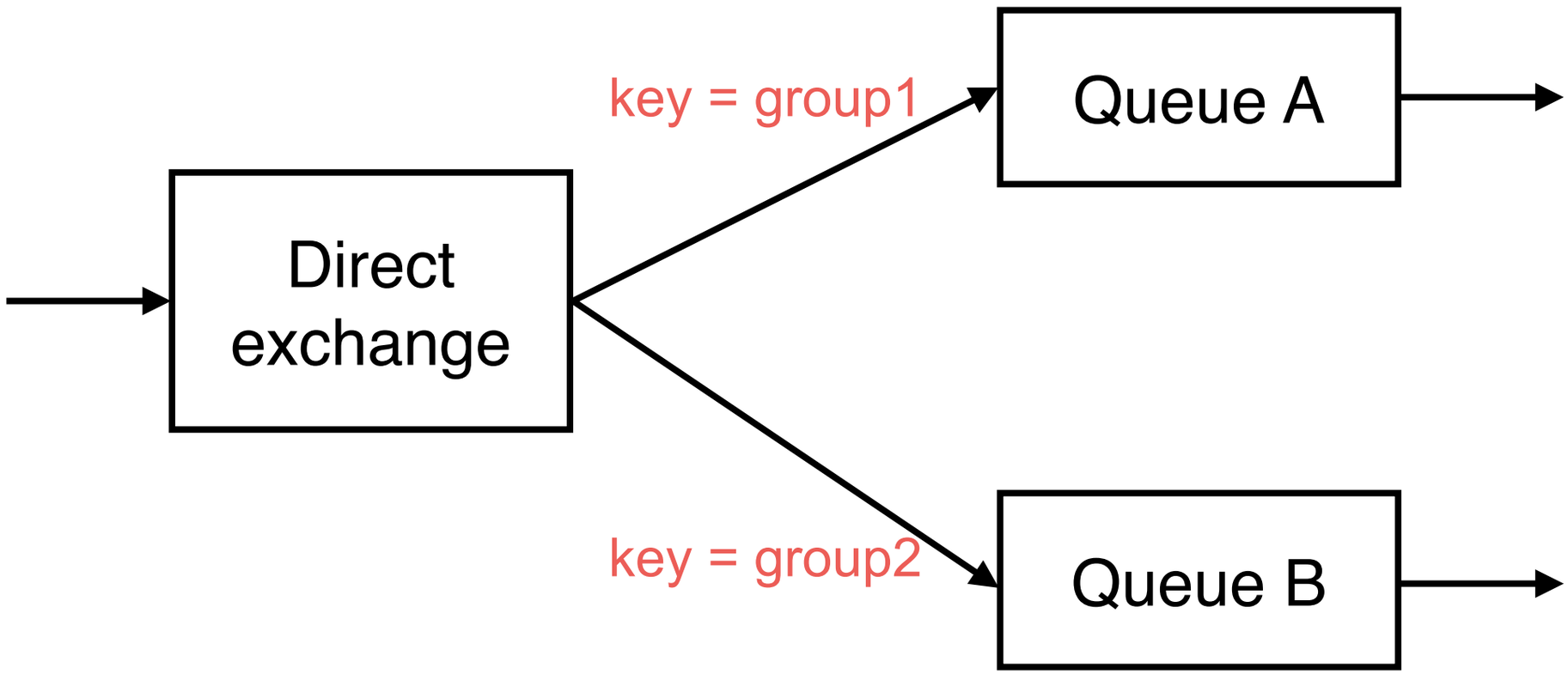
Exchange ?



Direct Exchange

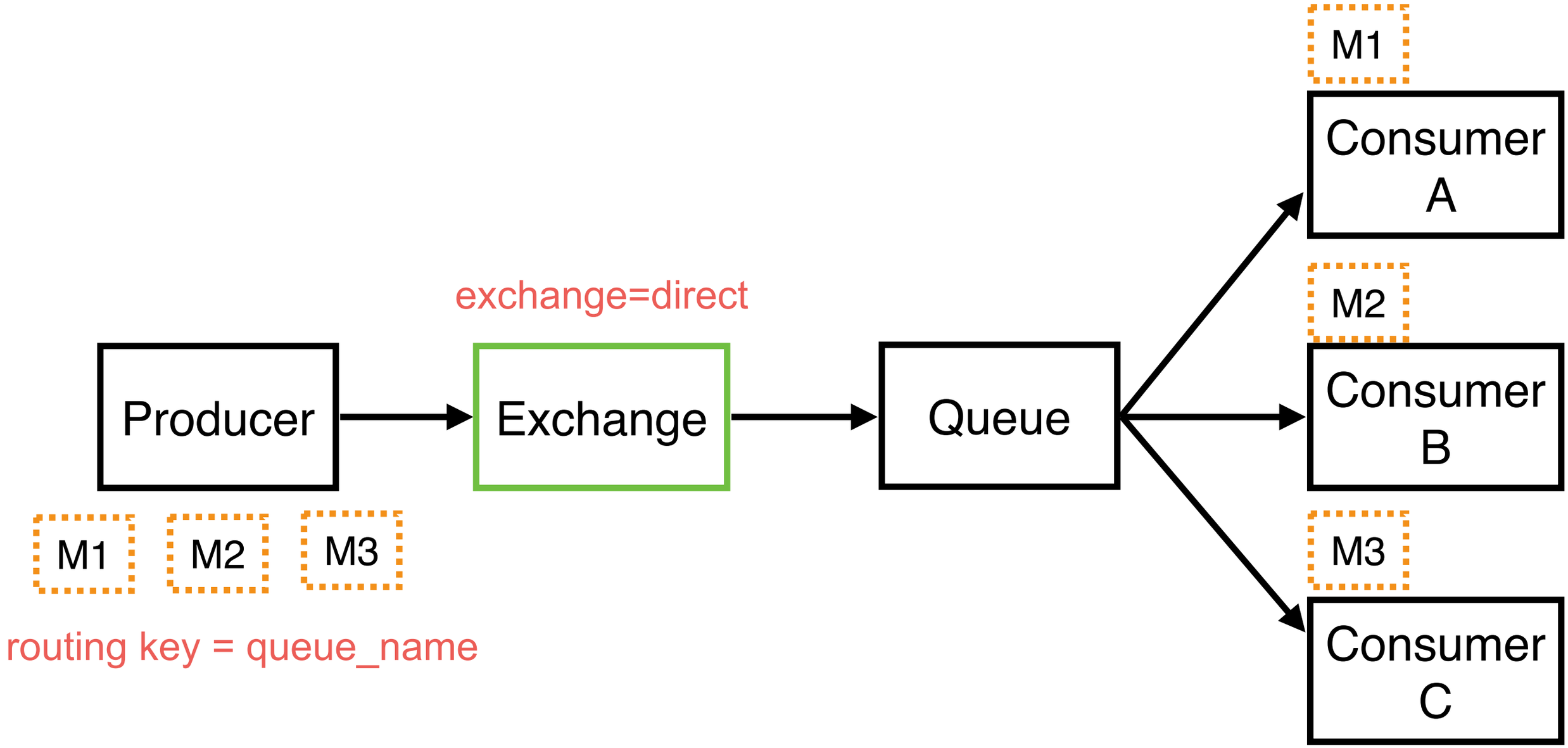
Direct exchange

Deliver message to queue based on **routing key**



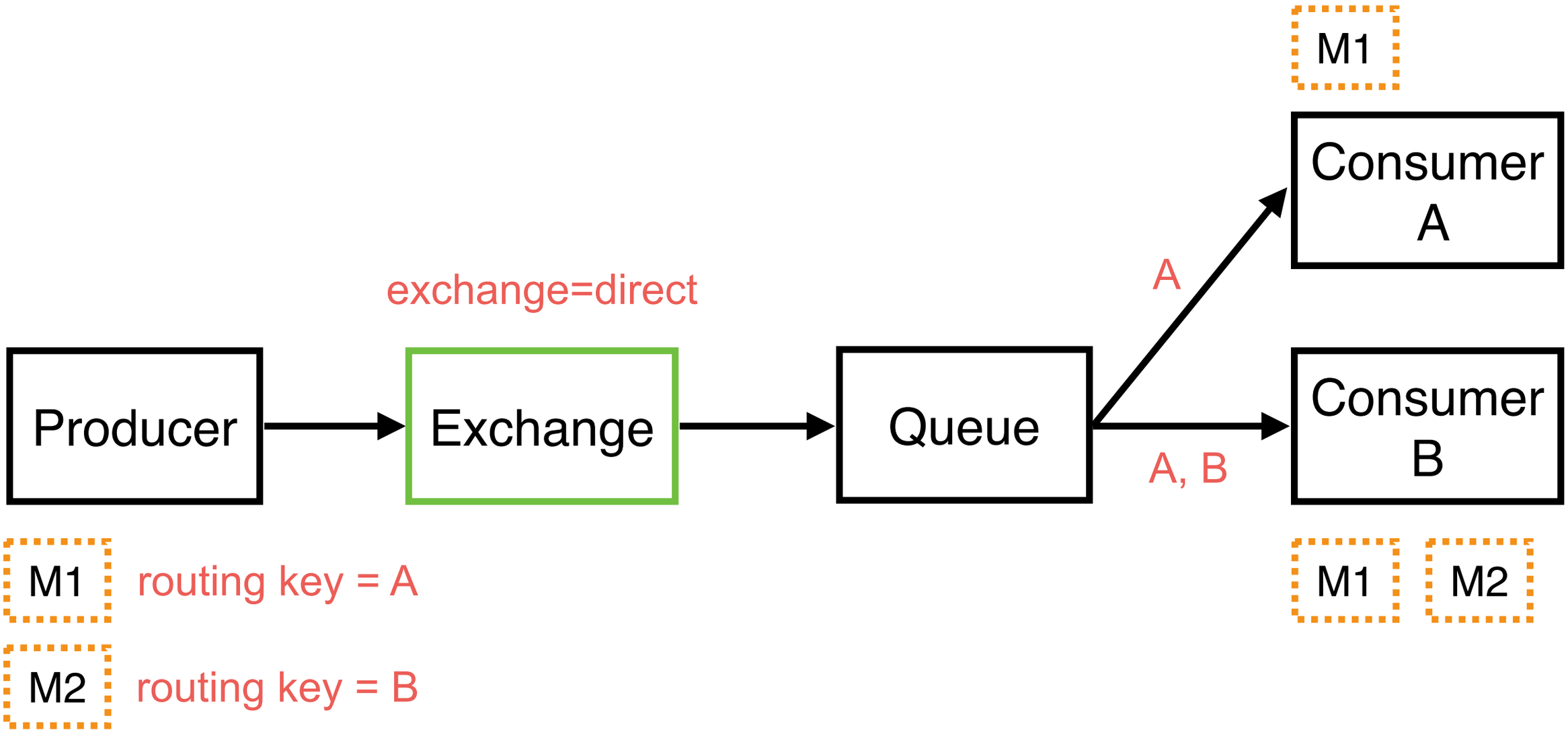
1. Work Queues

Exchange = empty = direct



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/work-queue

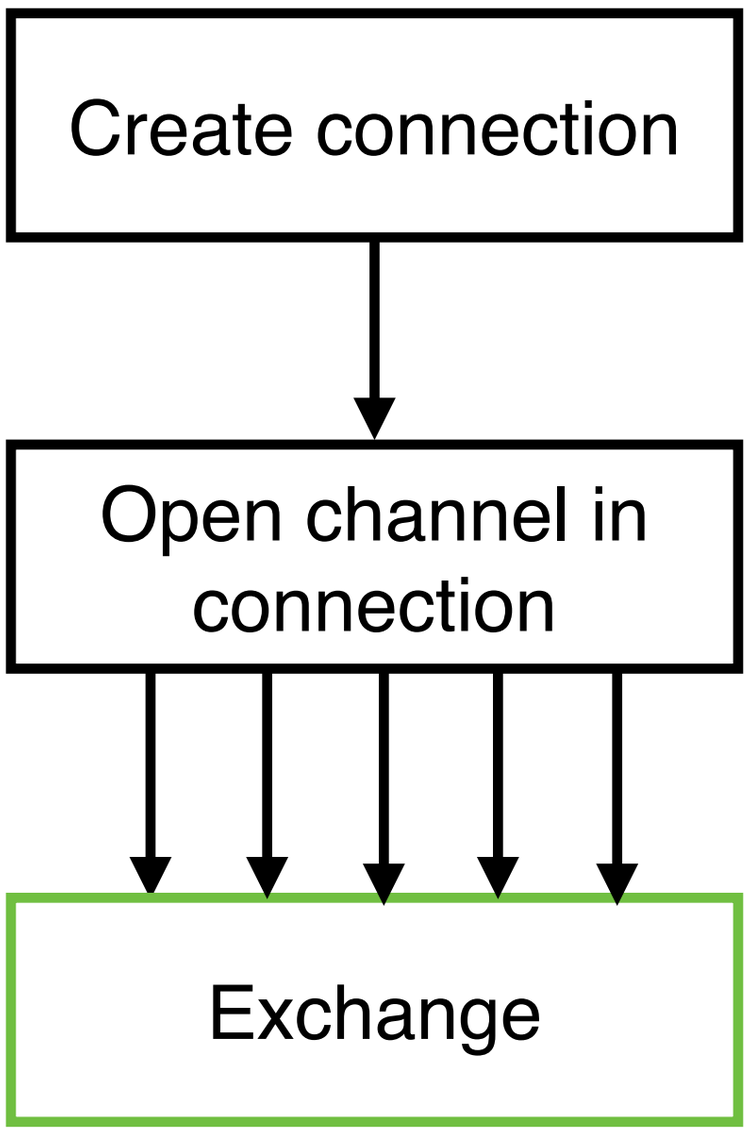
2. Routing by key



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/route-by-key

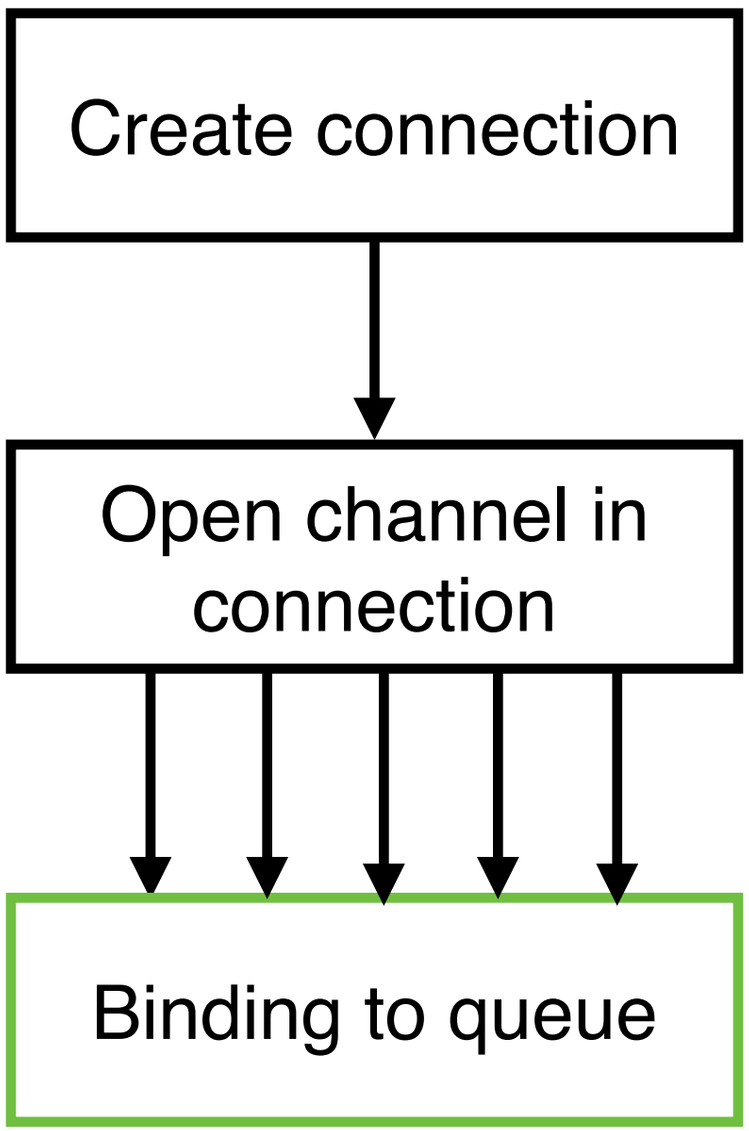
Producer

Step to publish message to exchange



Consumer

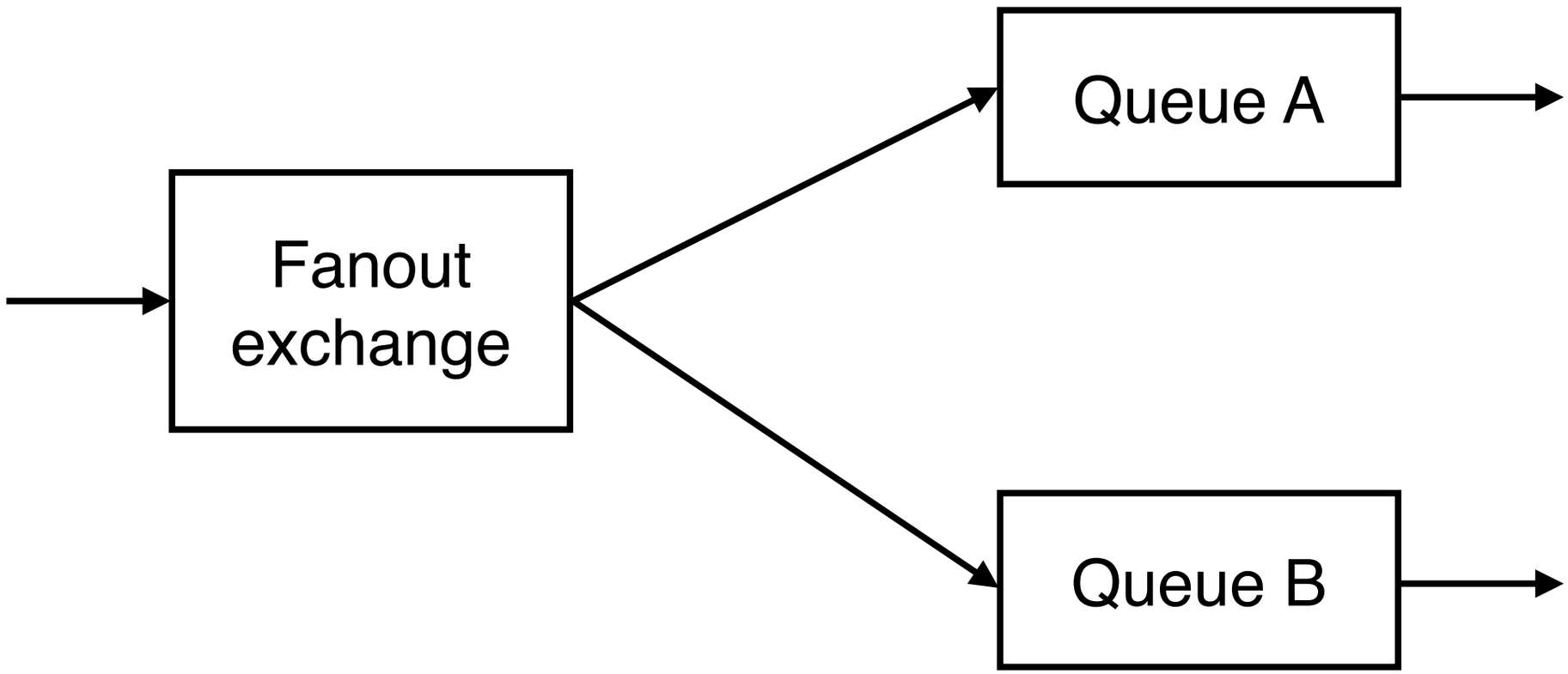
Step to read message from queue



Fanout Exchange

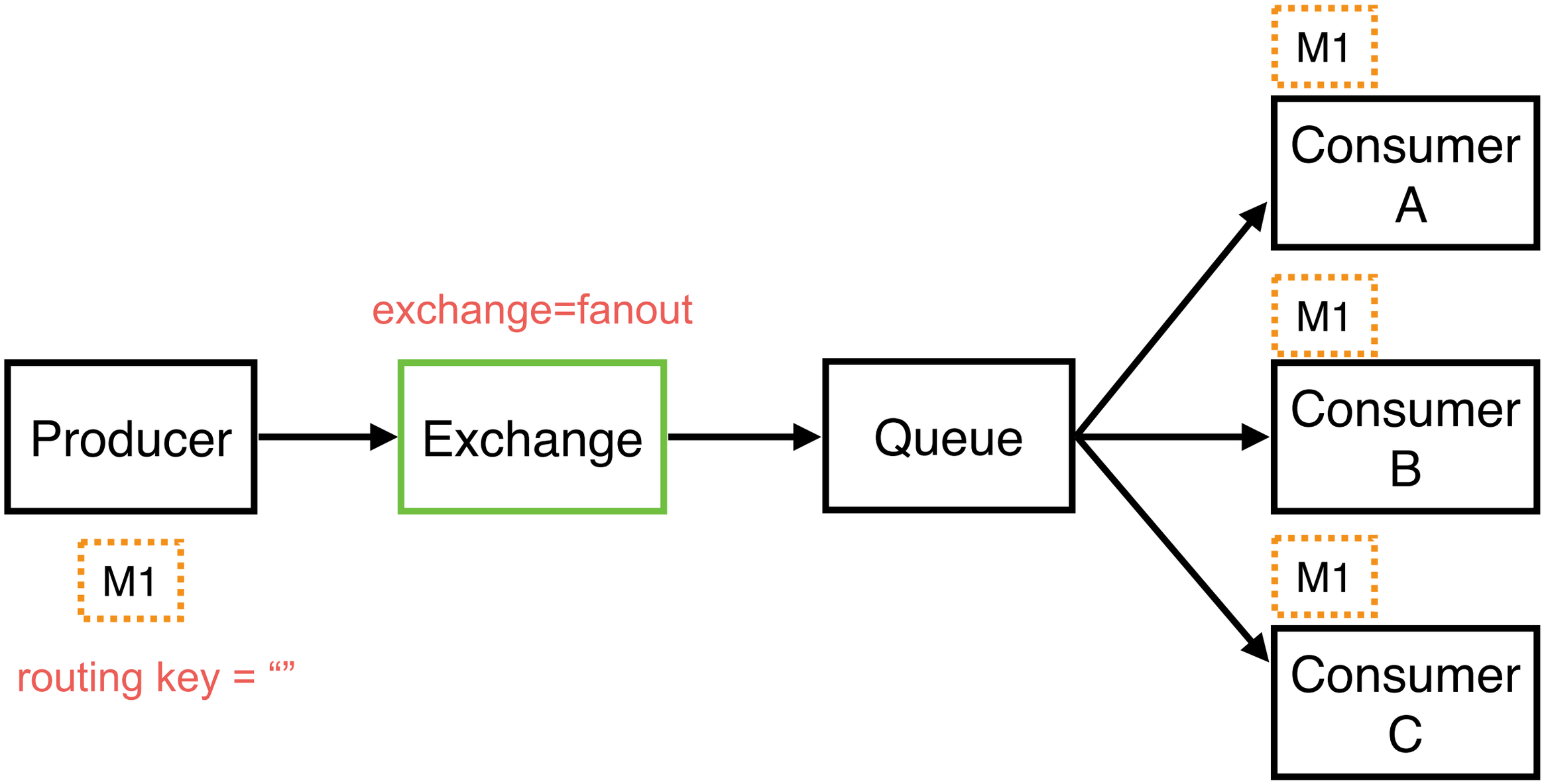
Fanout exchange

Broadcast messages to all bound queues
Ignore header, routing key in message



Publish/Subscribe

Exchange = fanout



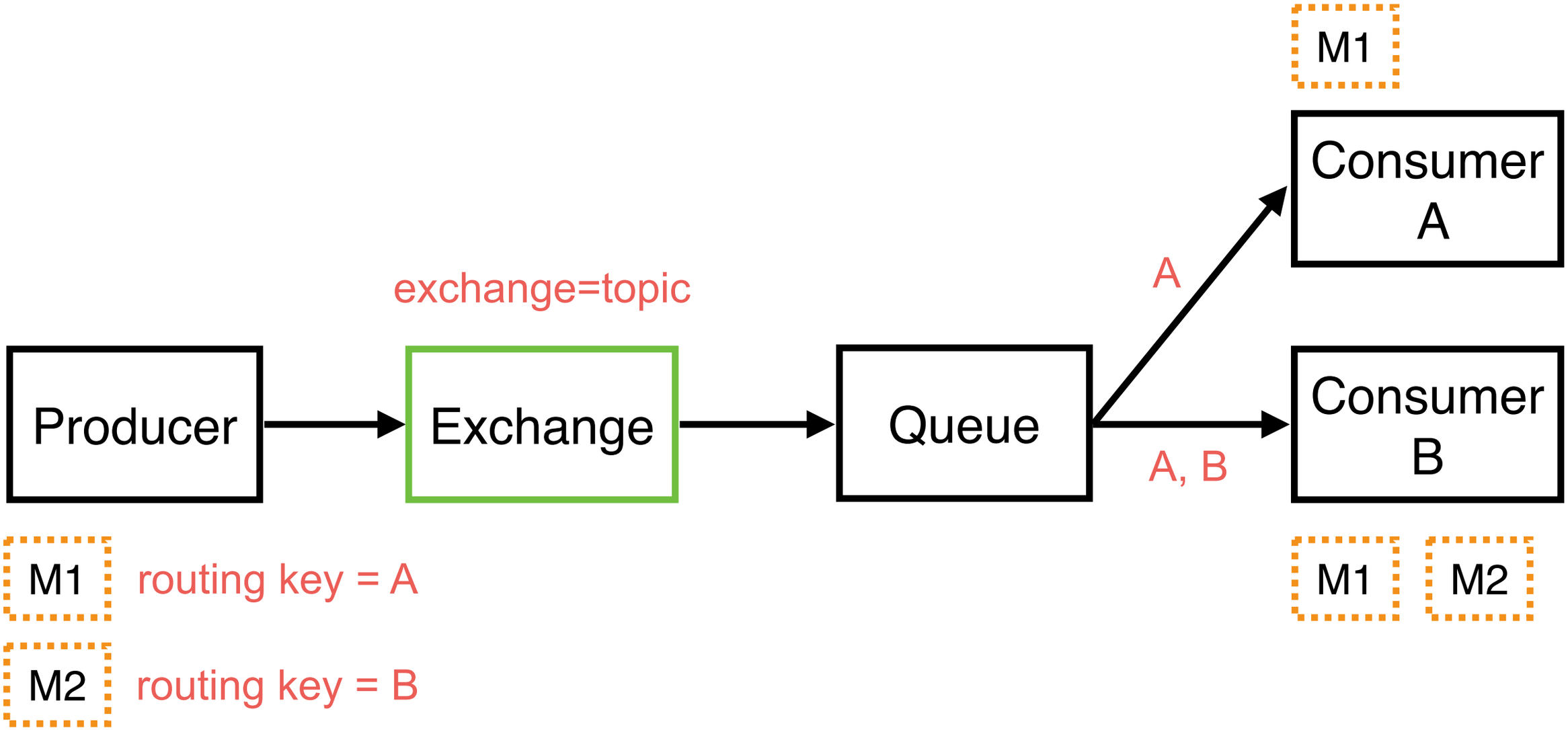
https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/pubsub

Topic Exchange

Topic exchange

Enables pattern-based routing
Using wildcard *, # in routing keys
Routing on multiple criteria

Routing by key*



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/topic

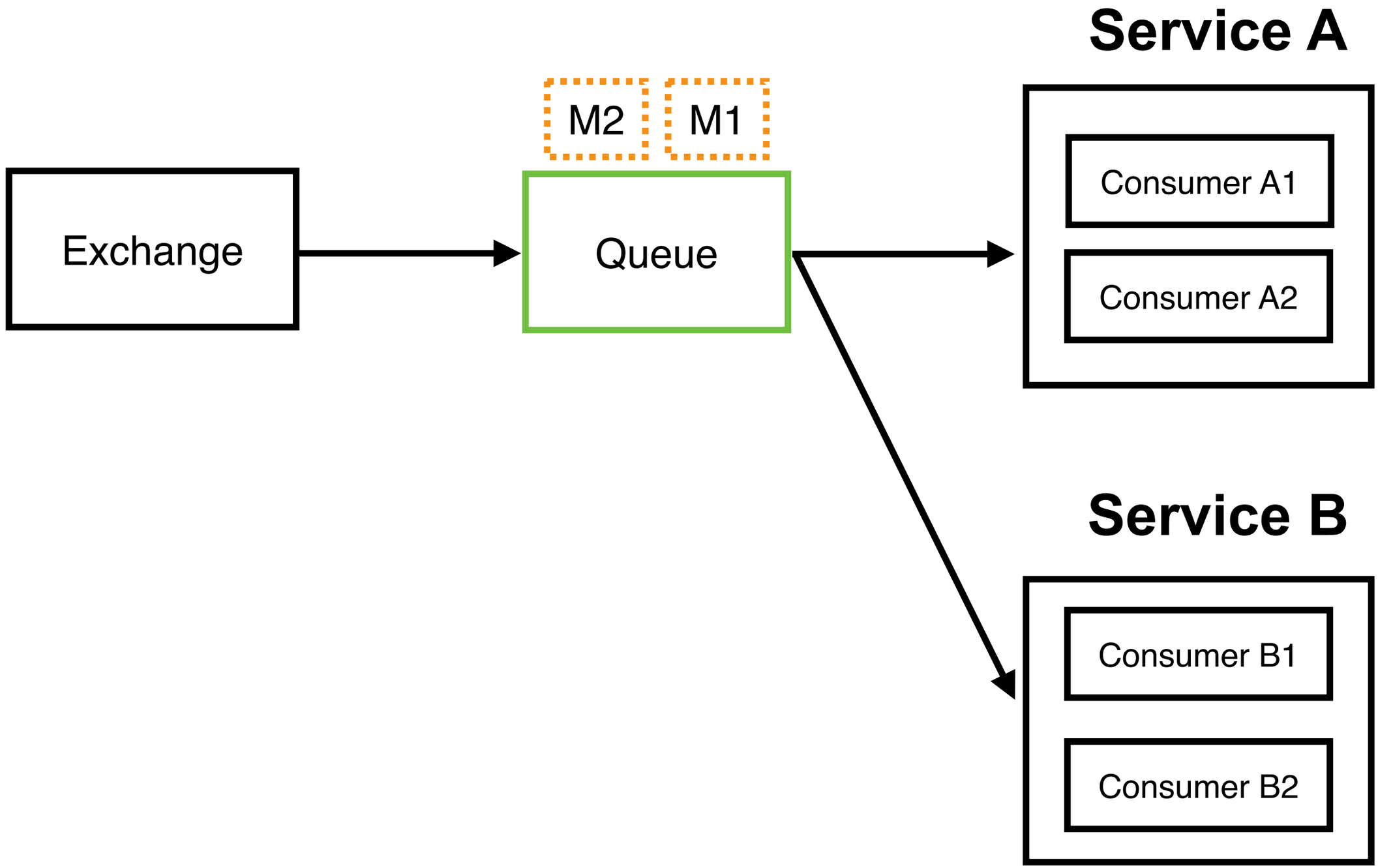
Header Exchange

Header exchange

Using message’s header to routing
Criteria based on header

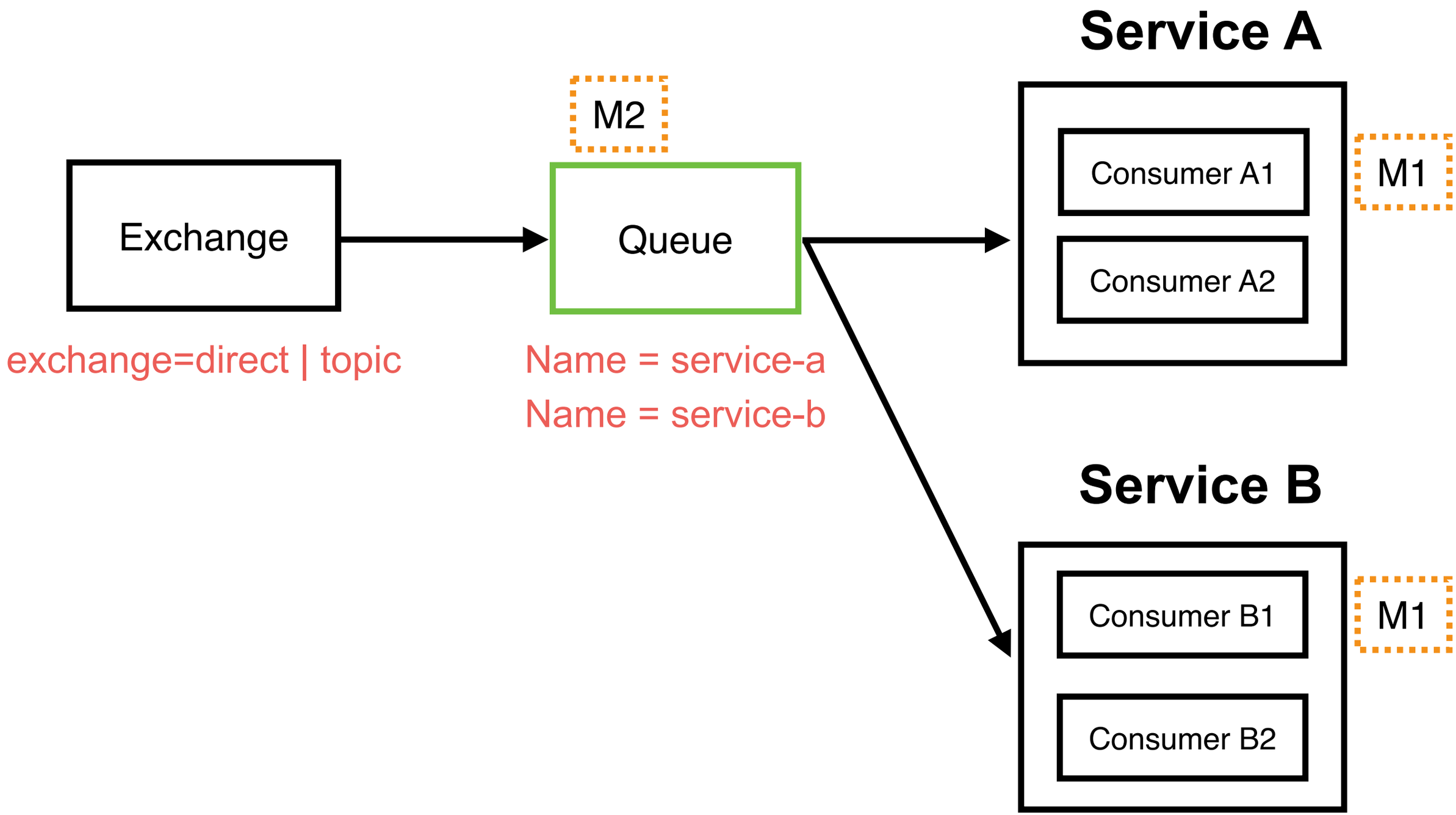
Use Cases

Scaling per services (1)



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/topic

Scaling per services (2)

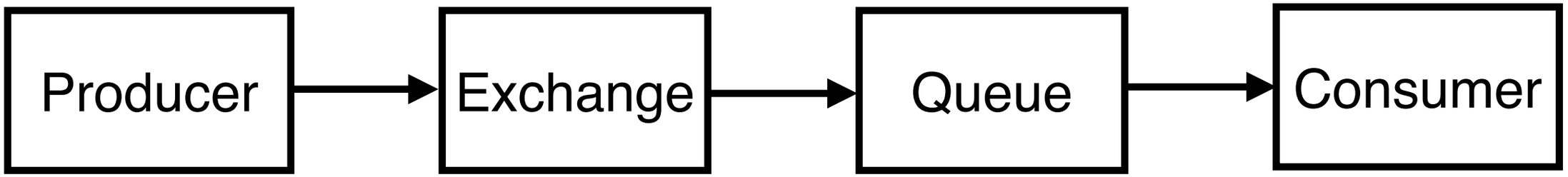


https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/topic

Failure Use Cases

Consumer crash (re-deliver)
Broker restart/crash
Retry with delay (N time)
Exceeding retry attempts
High-priority Dead Letter Queue (DLQ)

Consumer crash (re-deliver)



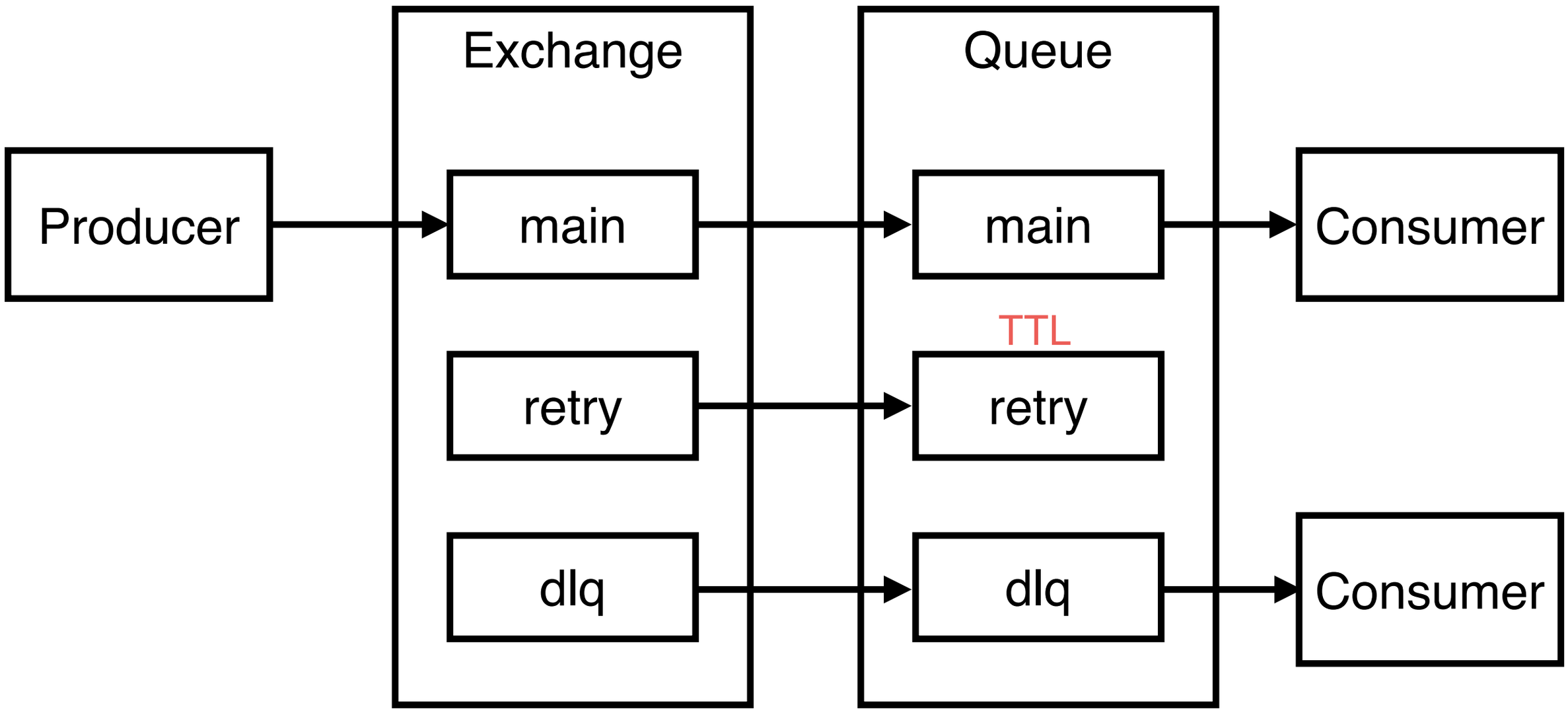
Delivery mode = Persistent

Durable = true
Delete = false

Auto ack = false

https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/redeliver

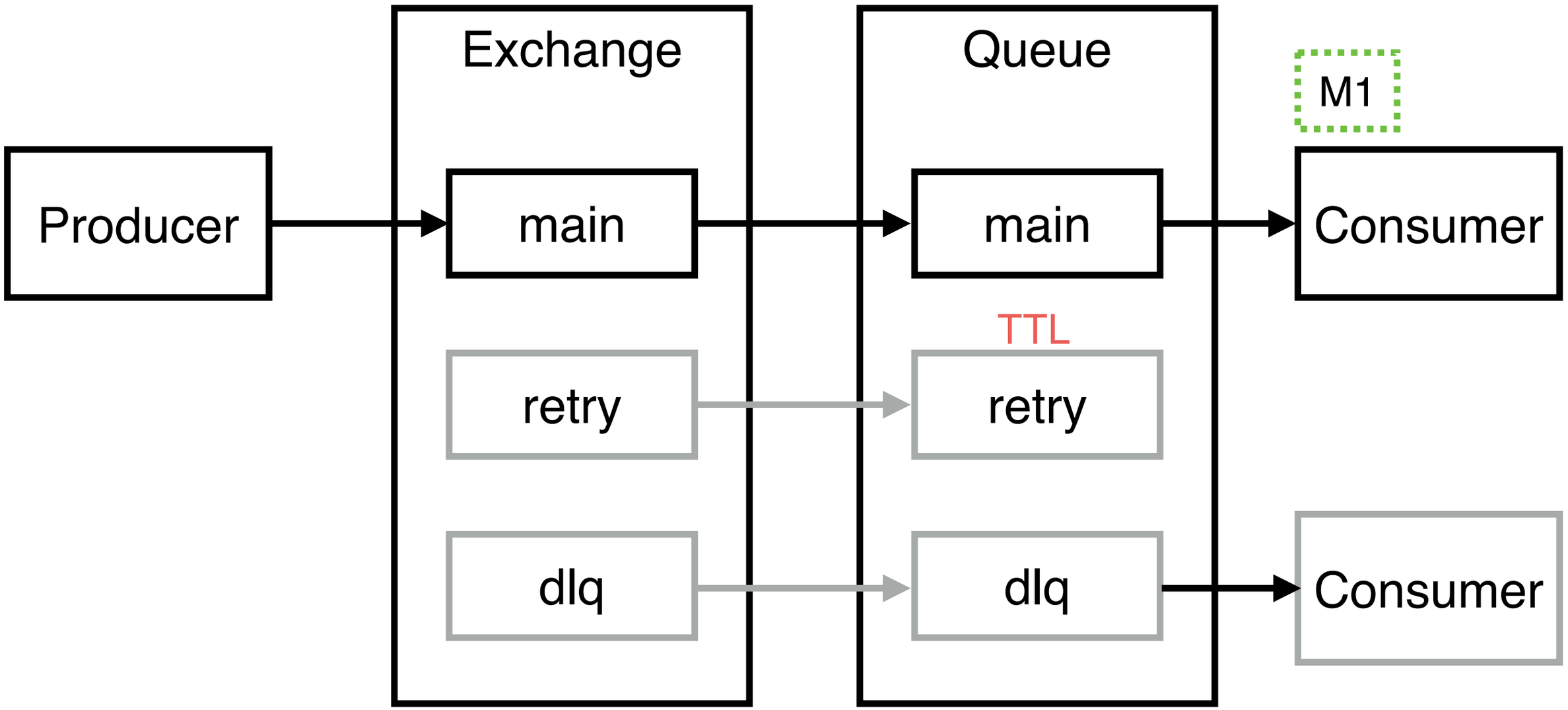
Retry and Dead Letter Queue



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/dlq

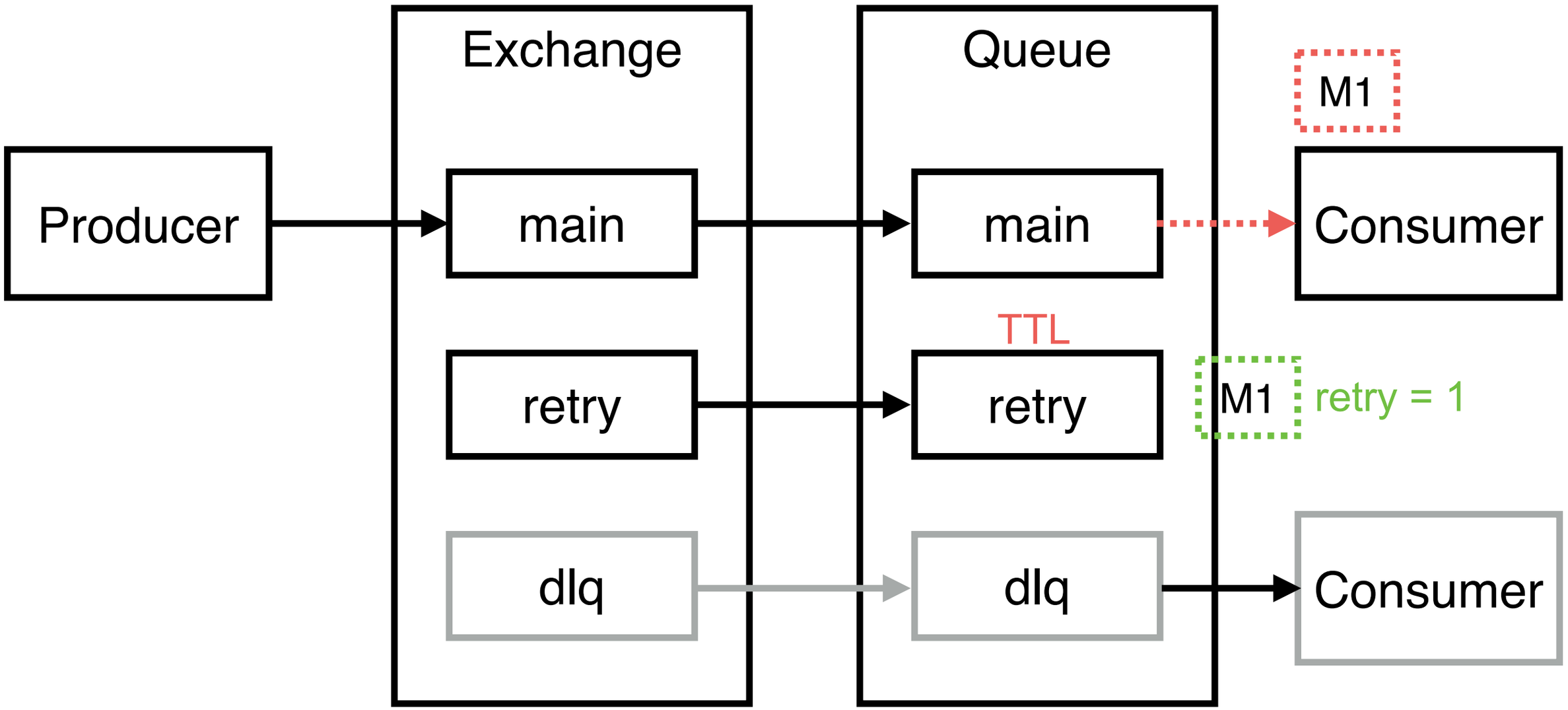
Retry and Dead Letter Queue

Normal process



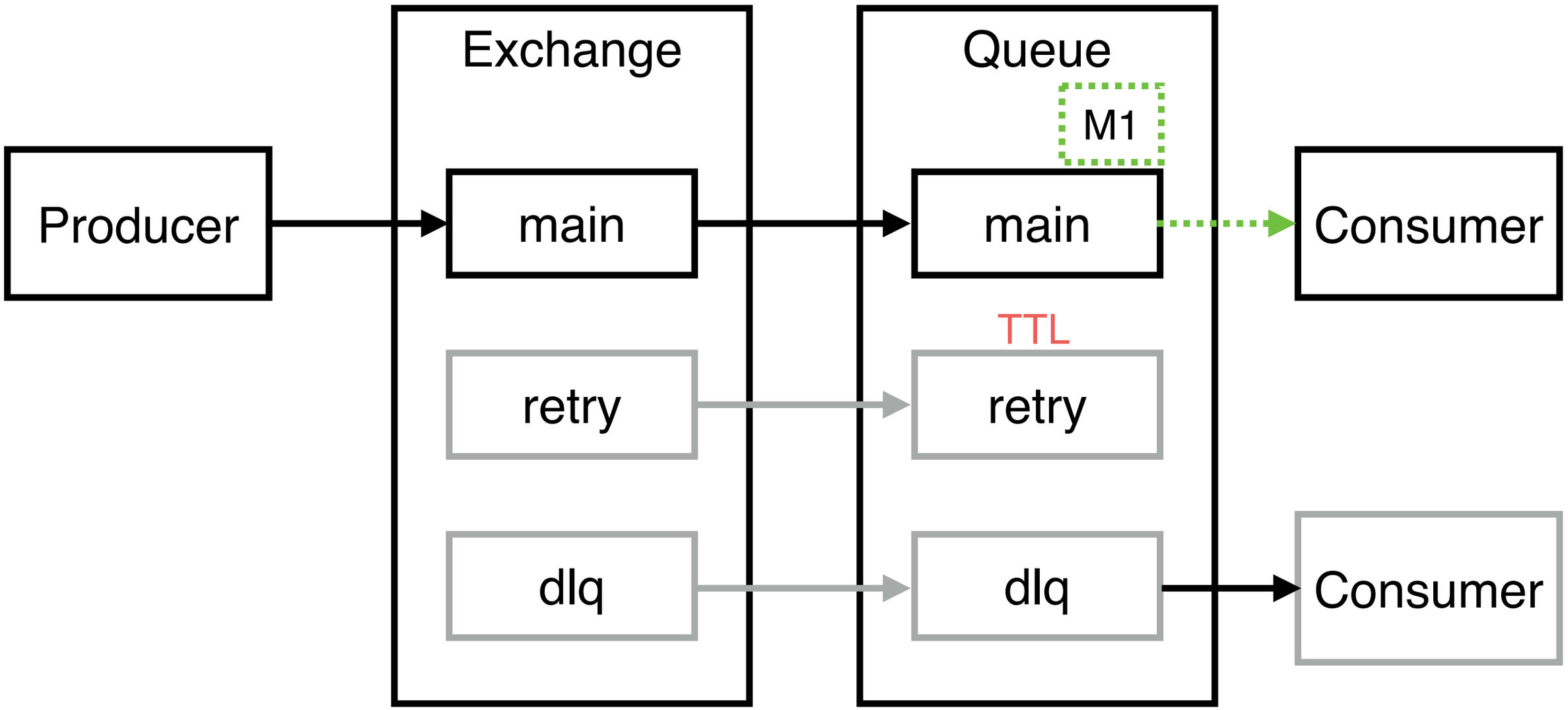
Retry and Dead Letter Queue

Problem in consumer, move message to **retry** exchange



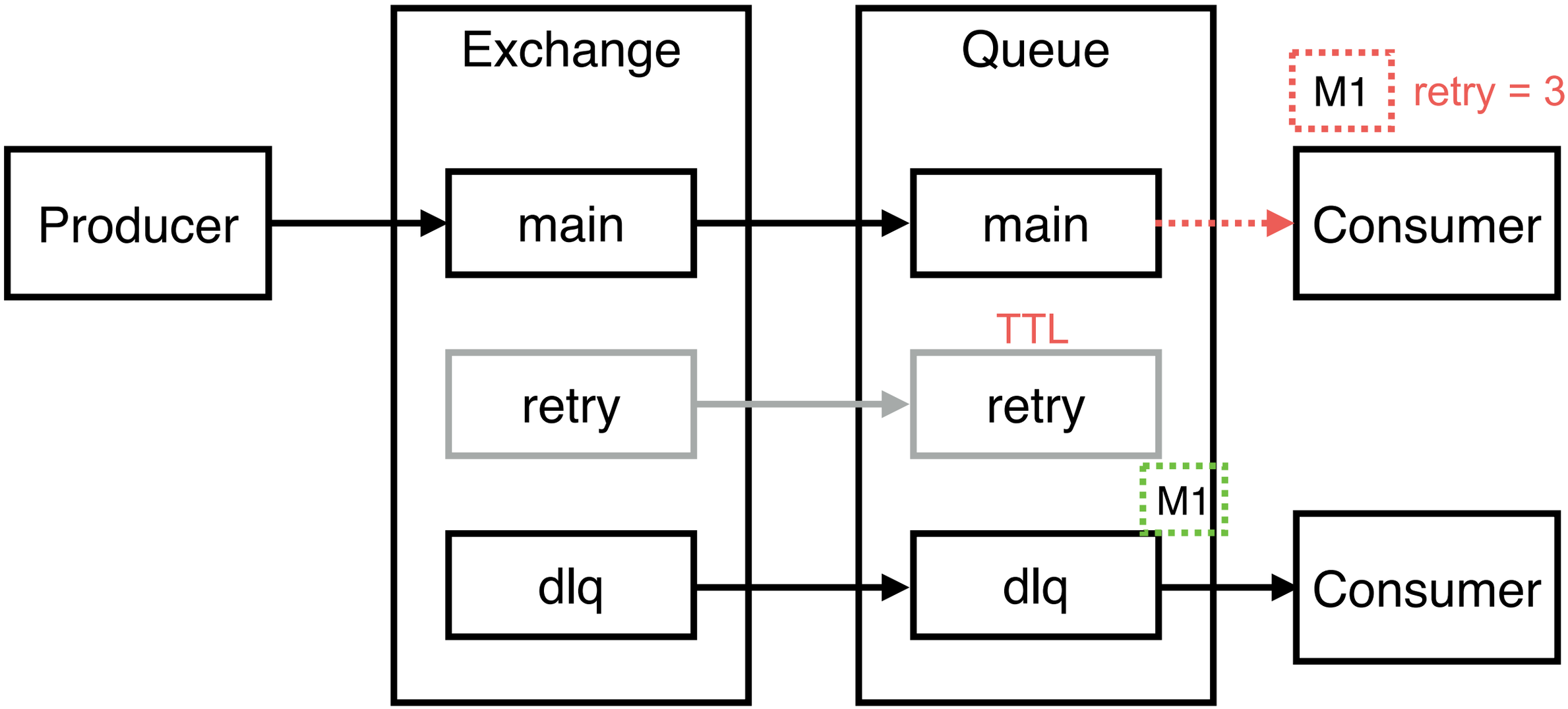
Retry and Dead Letter Queue

When TTL expired, move message to **main** exchange



Retry and Dead Letter Queue

When retry > 3 then move message to **dlq** exchange



RabbitMQ Stream

https://www.rabbitmq.com/docs/streams

Stream ?

Always persistent and replicated
Stream model an append-only log of message
Repeatedly read until message expire

Performance

High Throughput

Low Latency

Key features

Append-only log
Non-destructive consumer semantics
Persistent and replicated
Retention policies (size-based, time-based)

Use Cases of Stream ?

Large fan-out

Replay

Large volume of message

High Throughput

Event sourcing

Time-series data

Resources usage ?

All data is store on **disk**
Use disk I/O heavy
Lower CPU and memory than quorum
Tuning kernel page cache

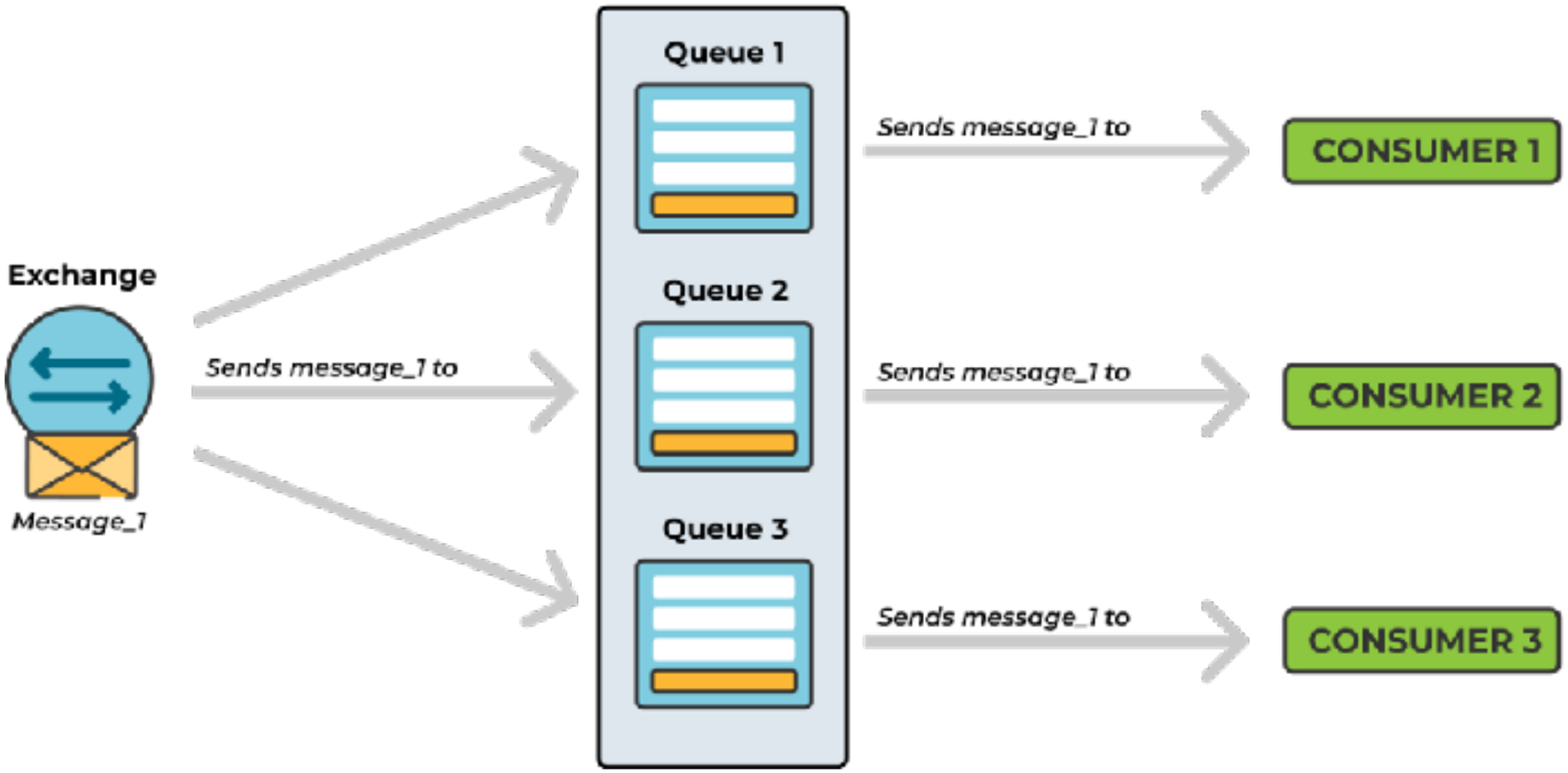
https://www.rabbitmq.com/docs/streams#resource-use

Queue vs Stream

Feature	Queue	Stream
Message handling	FIFO destructive read	Append-only log non-destructive read
Persistence	Optional	Always persistent and replicated
Retention	TTL (Time To Live) Queue length limits	Size-based, time-based
Performance	Optimize for point-to-point	Optimize for large fanout and replay

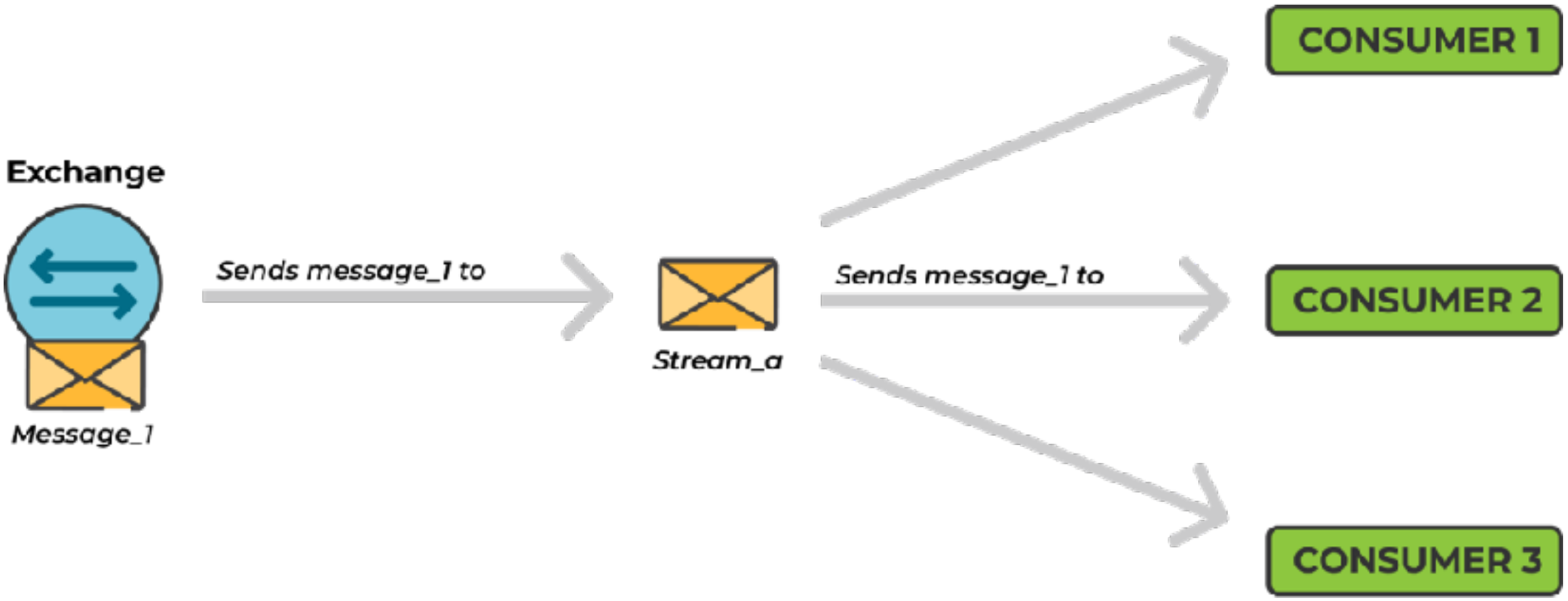
Fanout with Queue ?

Multiple consumers read the same message



https://www.cloudamqp.com/blog/rabbitmq-streams-and-replay-features-part-1-when-to-use-rabbitmq-streams.html

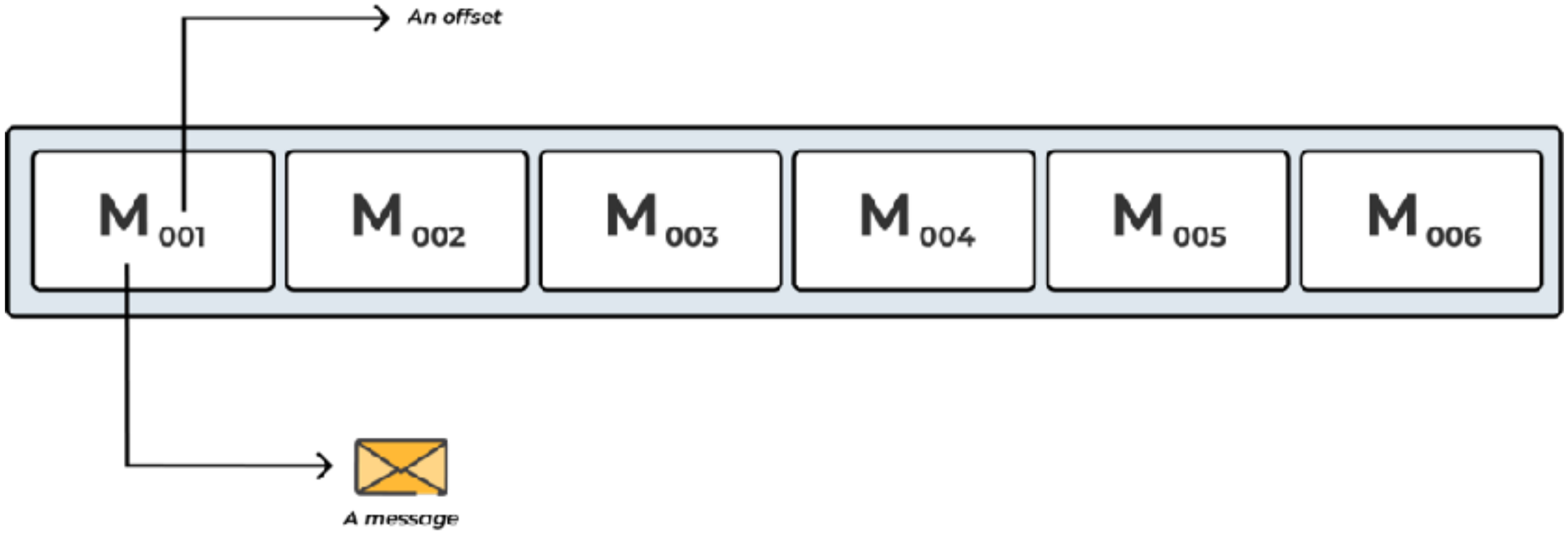
Fanout with Stream



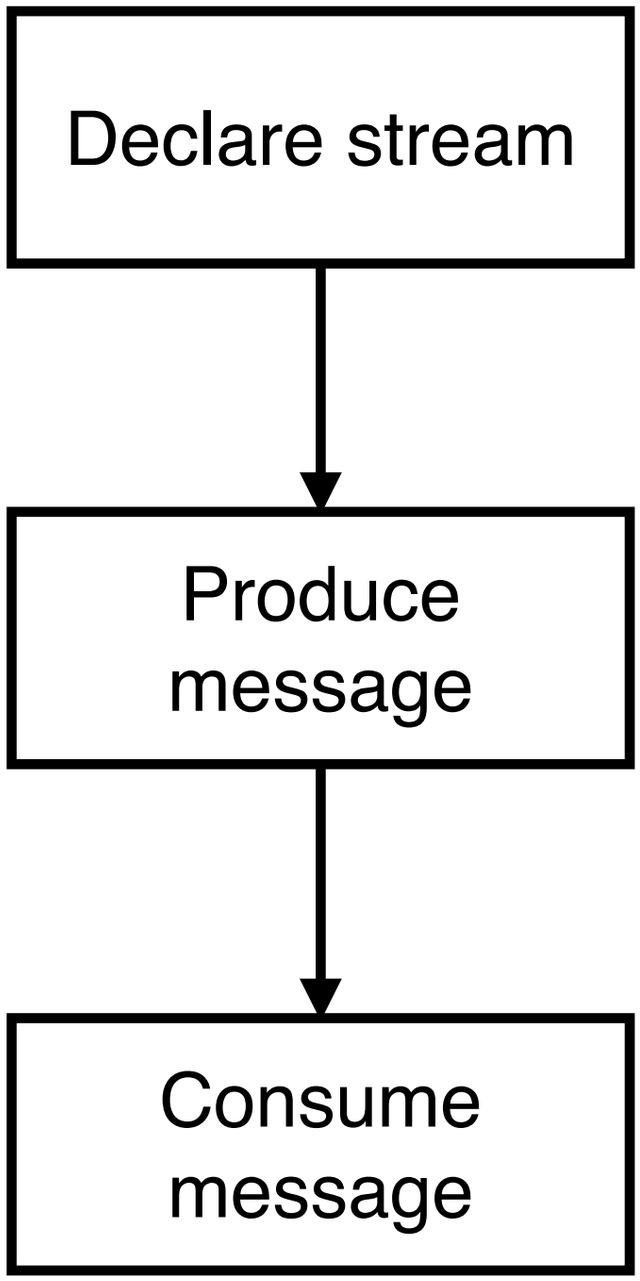
https://www.cloudamqp.com/blog/rabbitmq-streams-and-replay-features-part-1-when-to-use-rabbitmq-streams.html

Replay with offset

When consumer needs to re-read the same message



Steps to use Stream ?



Workshop
Fanout with Stream

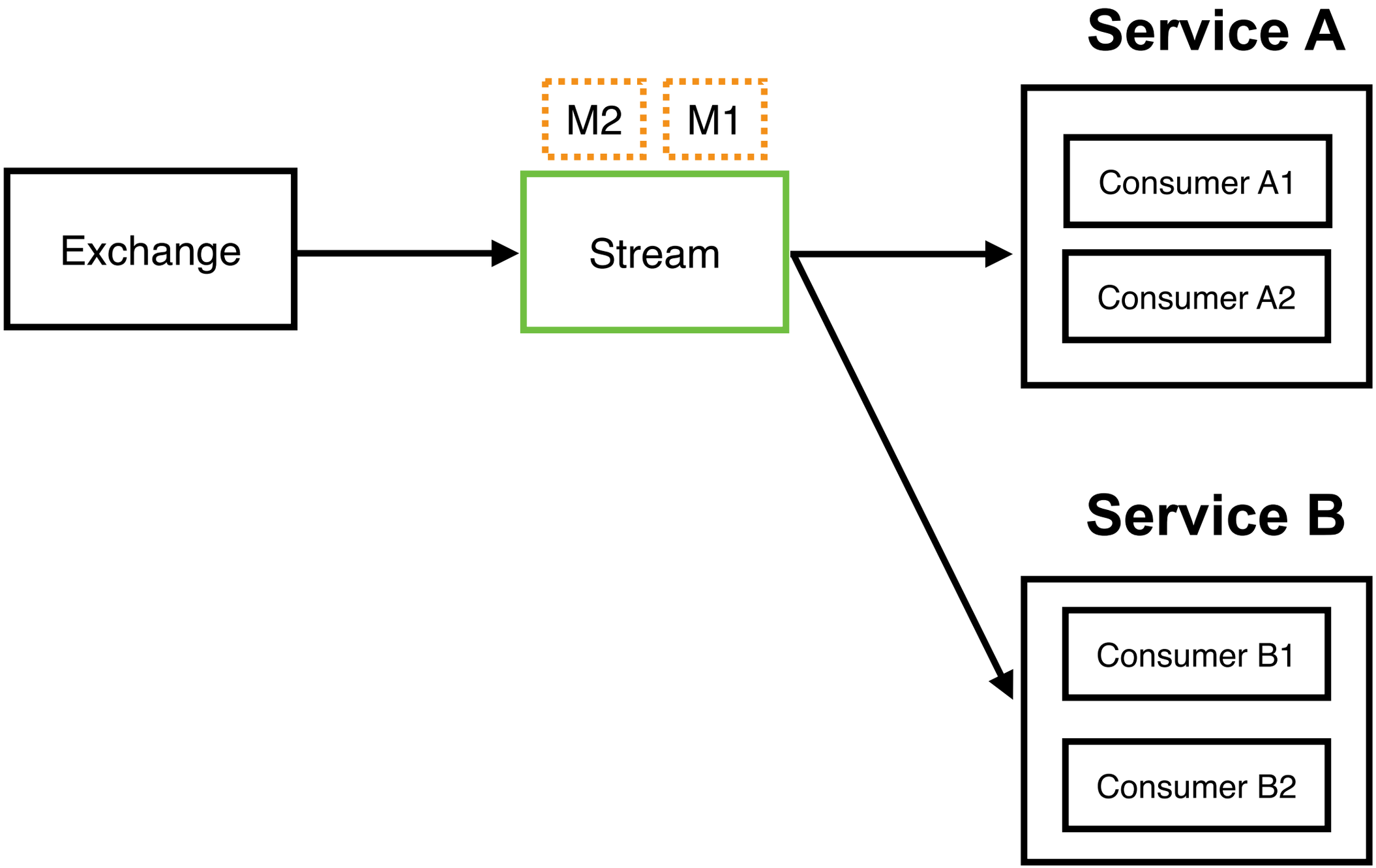
Fanout

Tracking offset

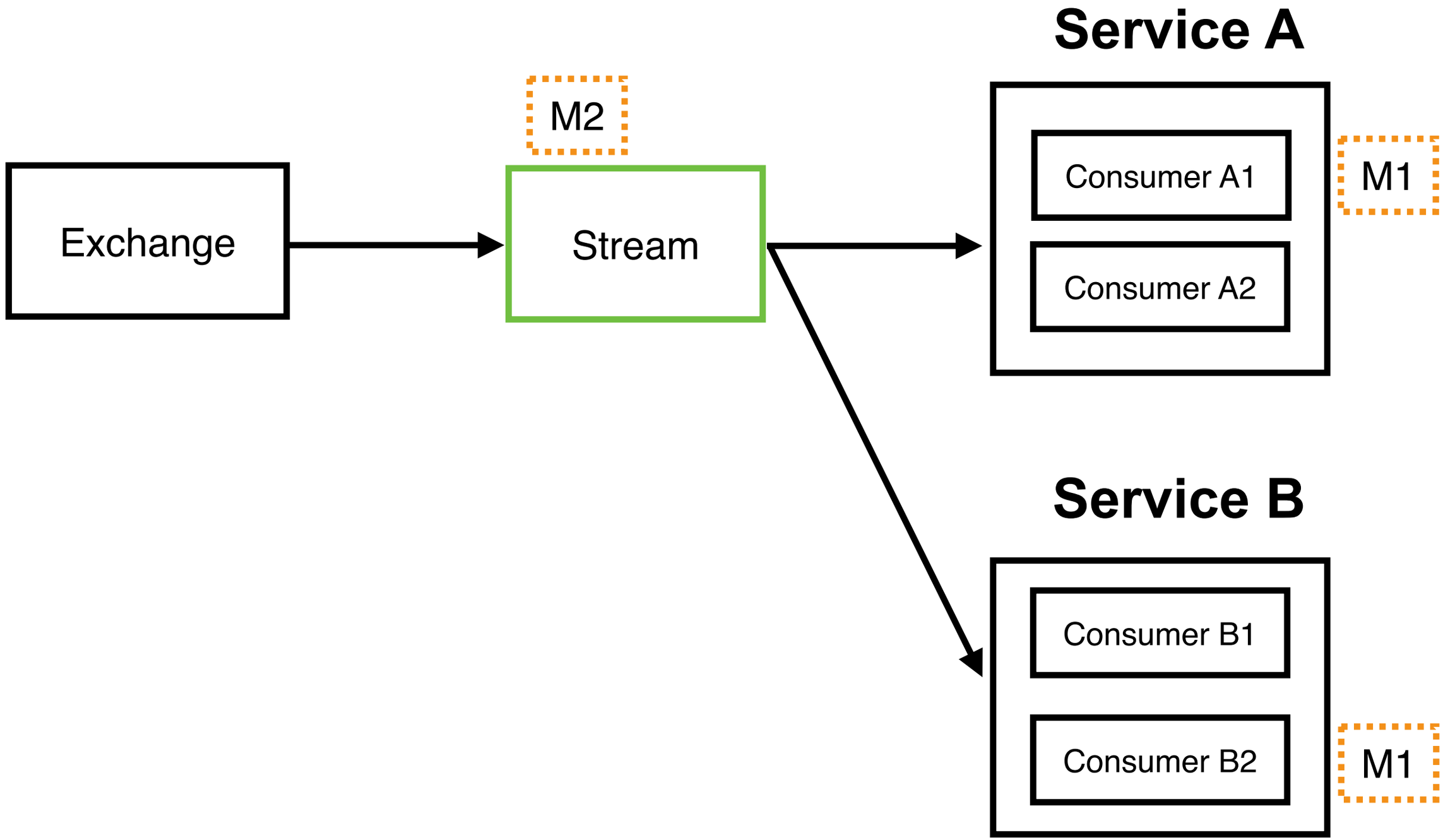
Filter message

https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/demo-stream/basic

Requirement (1)



Requirement (2)

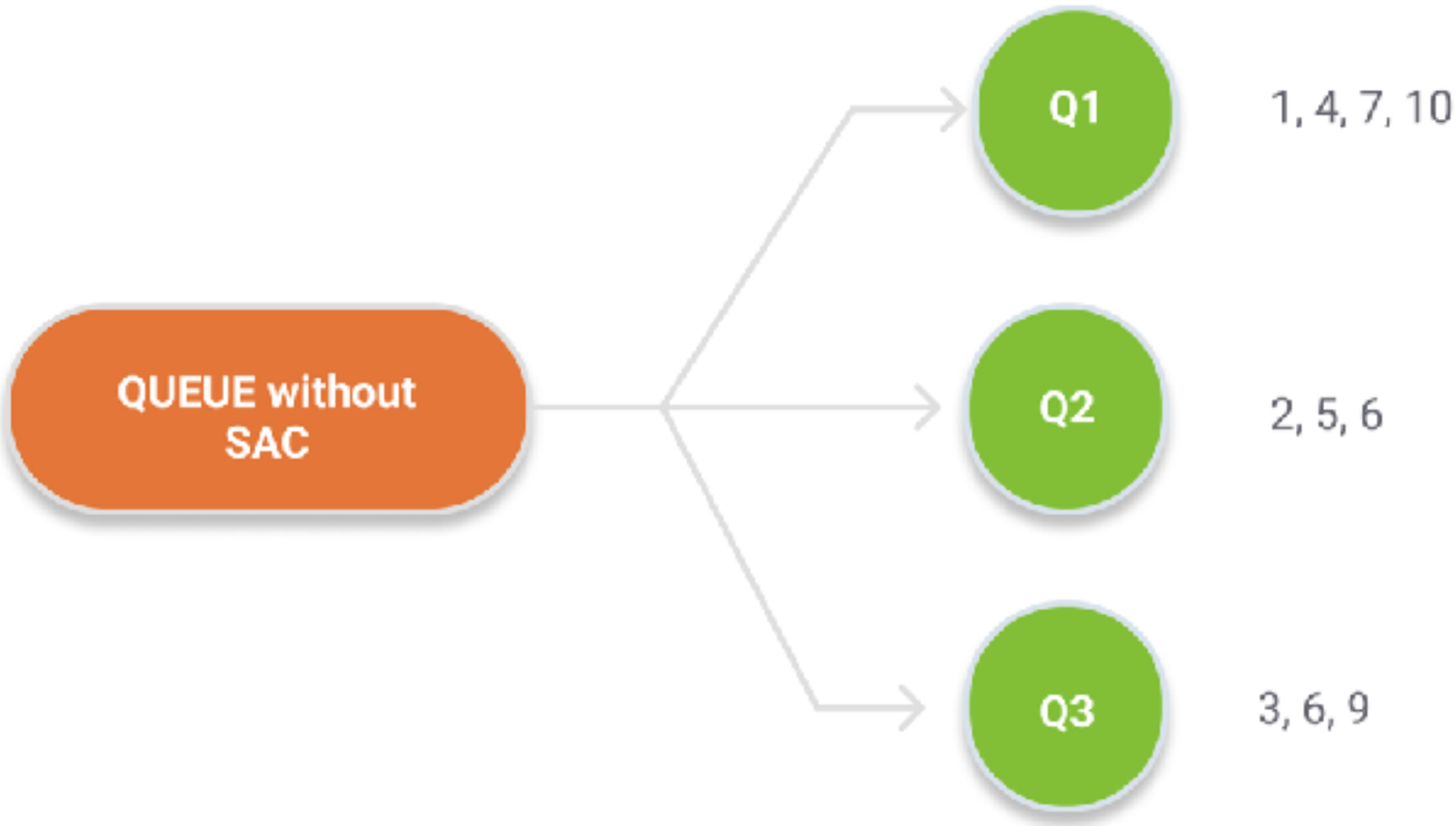


SAC
Single Active Consumer

https://www.rabbitmq.com/docs/consumers#single-active-consumer

https://www.rabbitmq.com/docs/streams#single-active-consumer

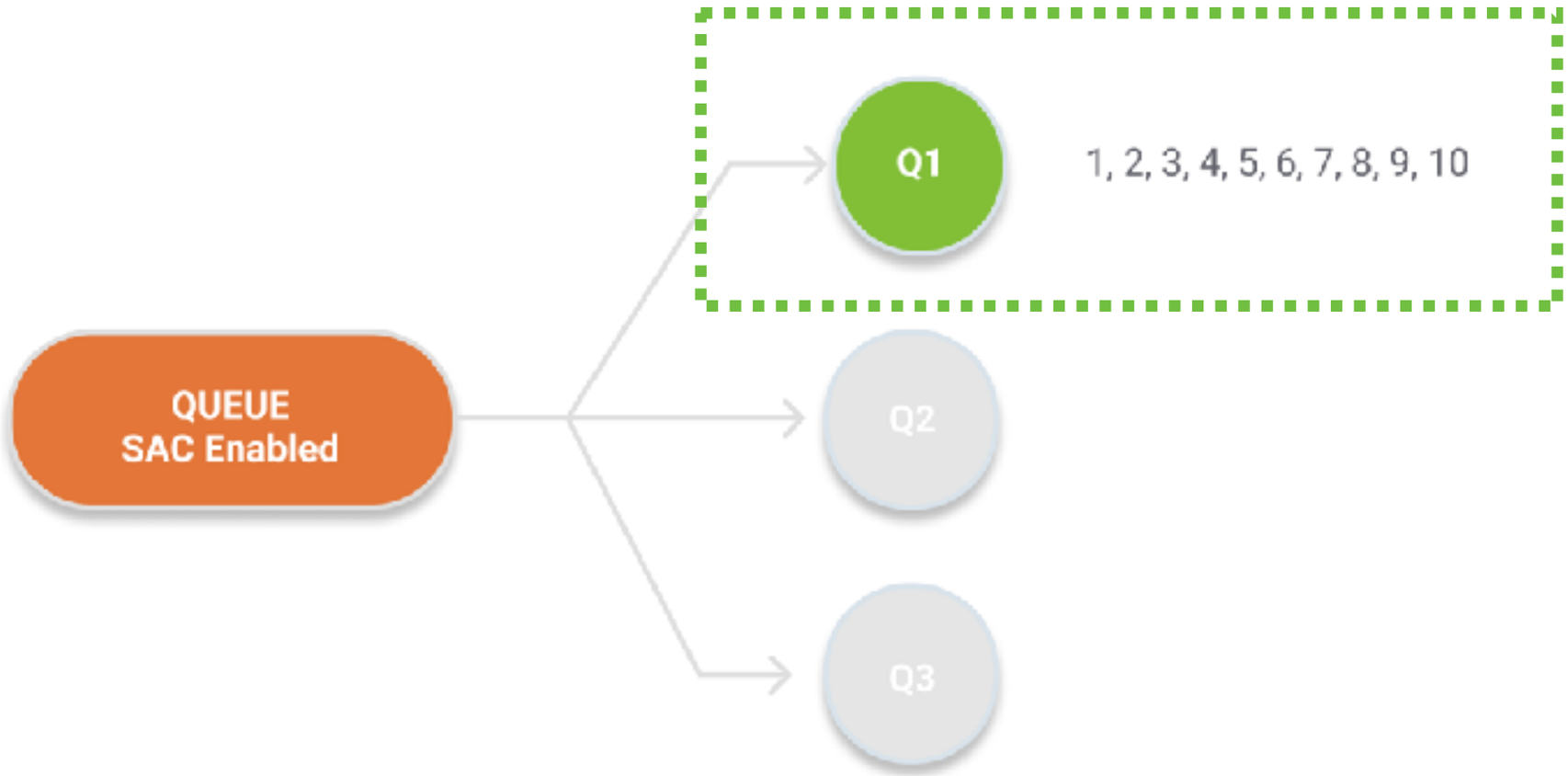
Queue without SAC



https://www.cloudamqp.com/blog/rabbitmq-3-8-feature-focus-single-active-consumer.html

Single Active Consumer

All messages get delivered to a **active consumer**
Auto switch to other consumers when active failed



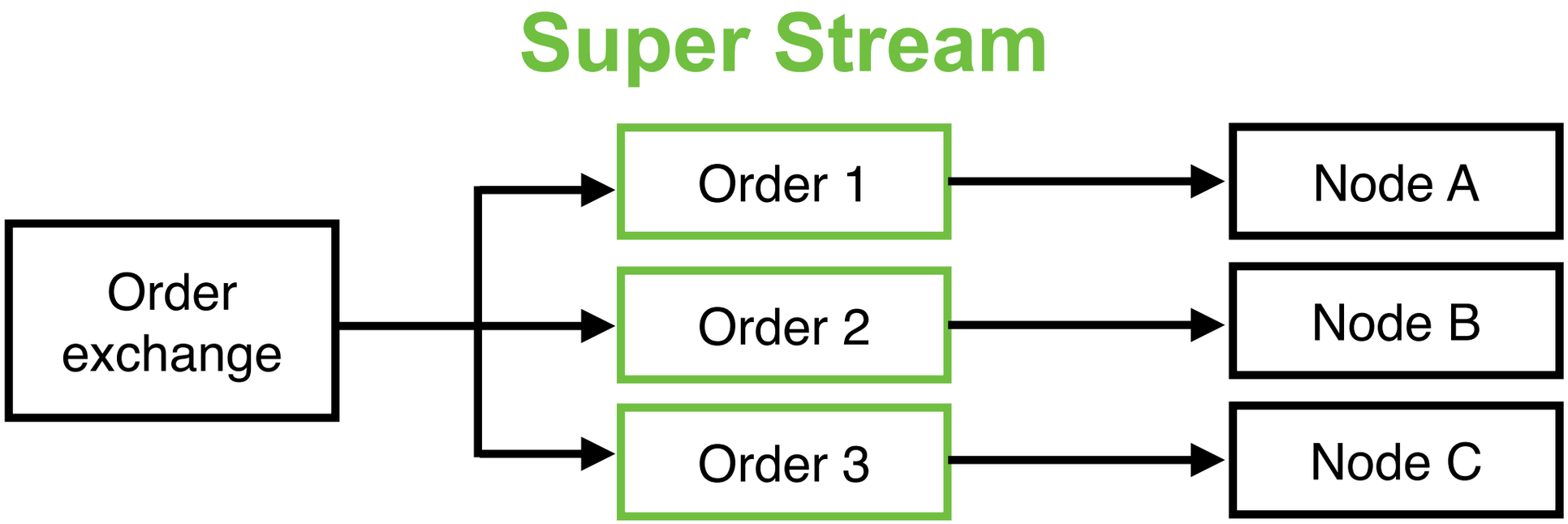
https://www.cloudamqp.com/blog/rabbitmq-3-8-feature-focus-single-active-consumer.html

Super Streams

(Partition the stream)

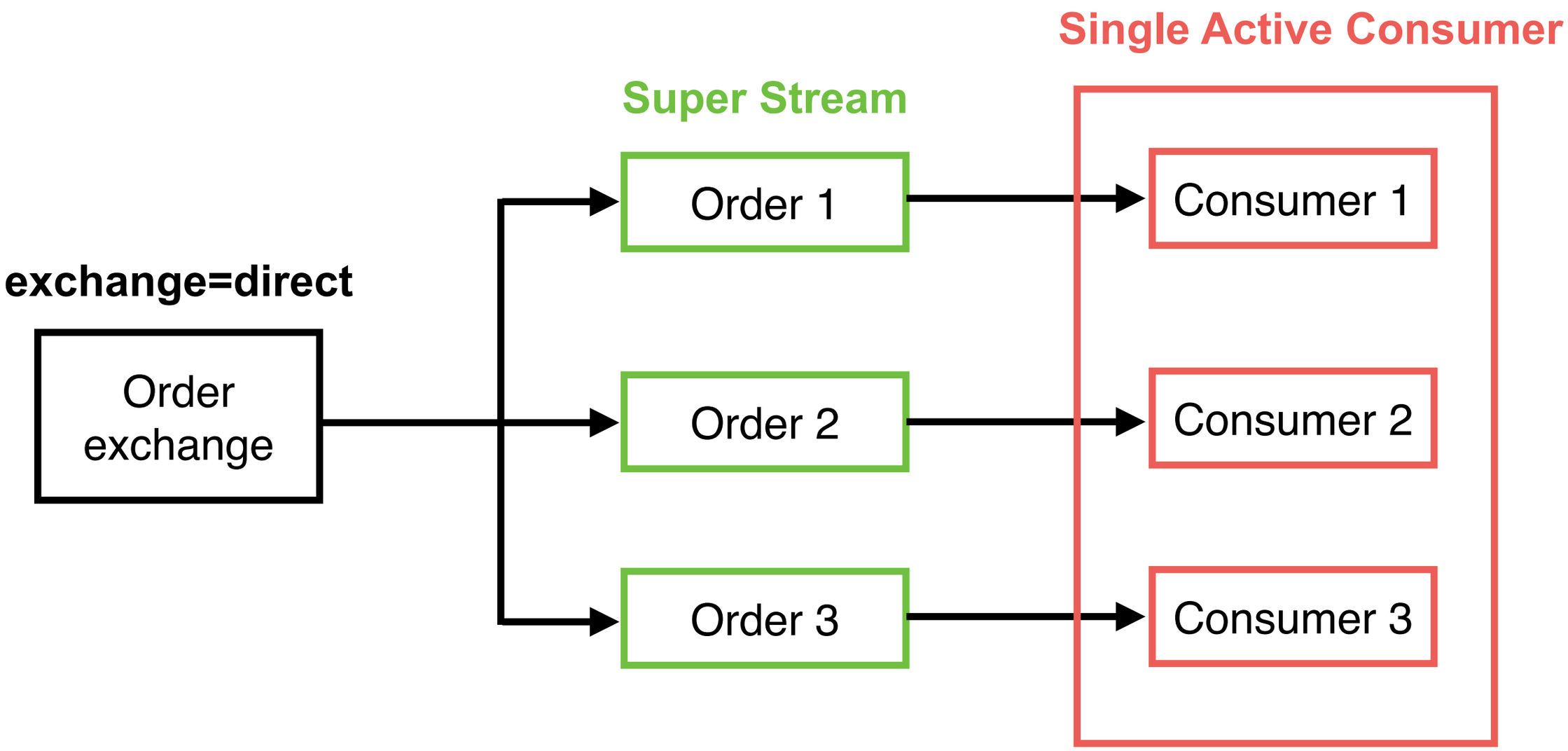
Super Streams ?

Way to scale out by partitioning a **large stream** to small
Split storage and traffic on multiple nodes in cluster



https://www.rabbitmq.com/blog/2022/07/13/rabbitmq-3-11-feature-preview-super-streams

Workshop with Orders



https://github.com/up1/course-rabbitmq-2025/tree/main/workshop/demo-go/demo-stream/basic

Q/A