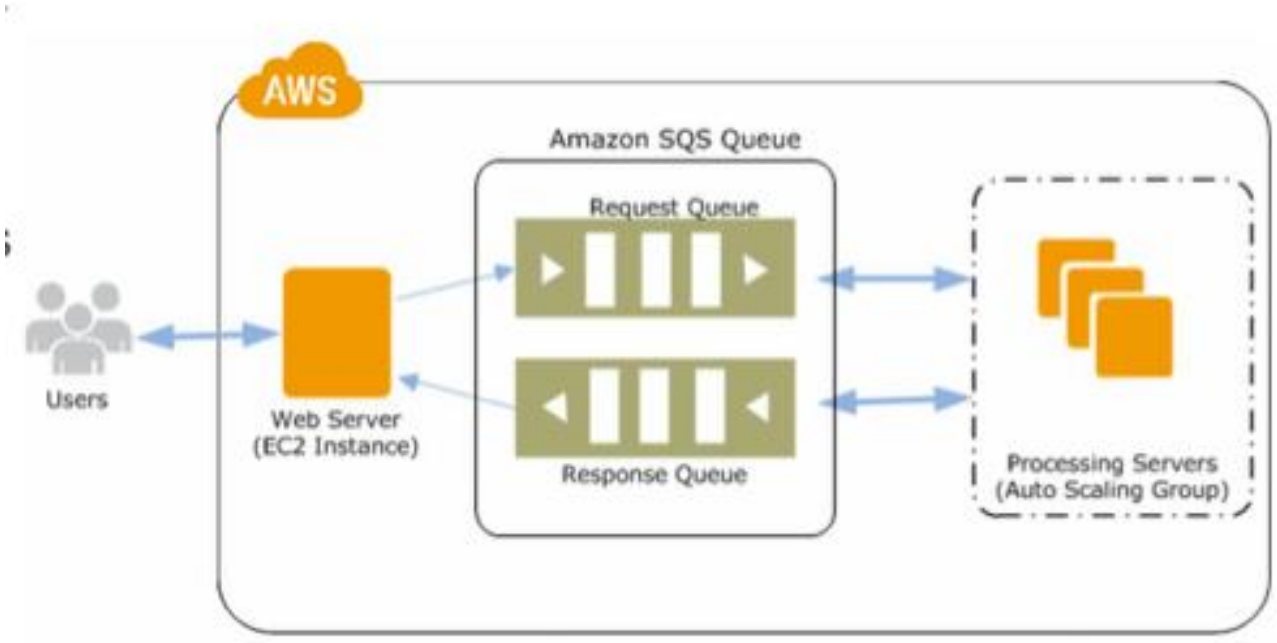
SQS (Simple Queue Service)

What is Amazon SQS

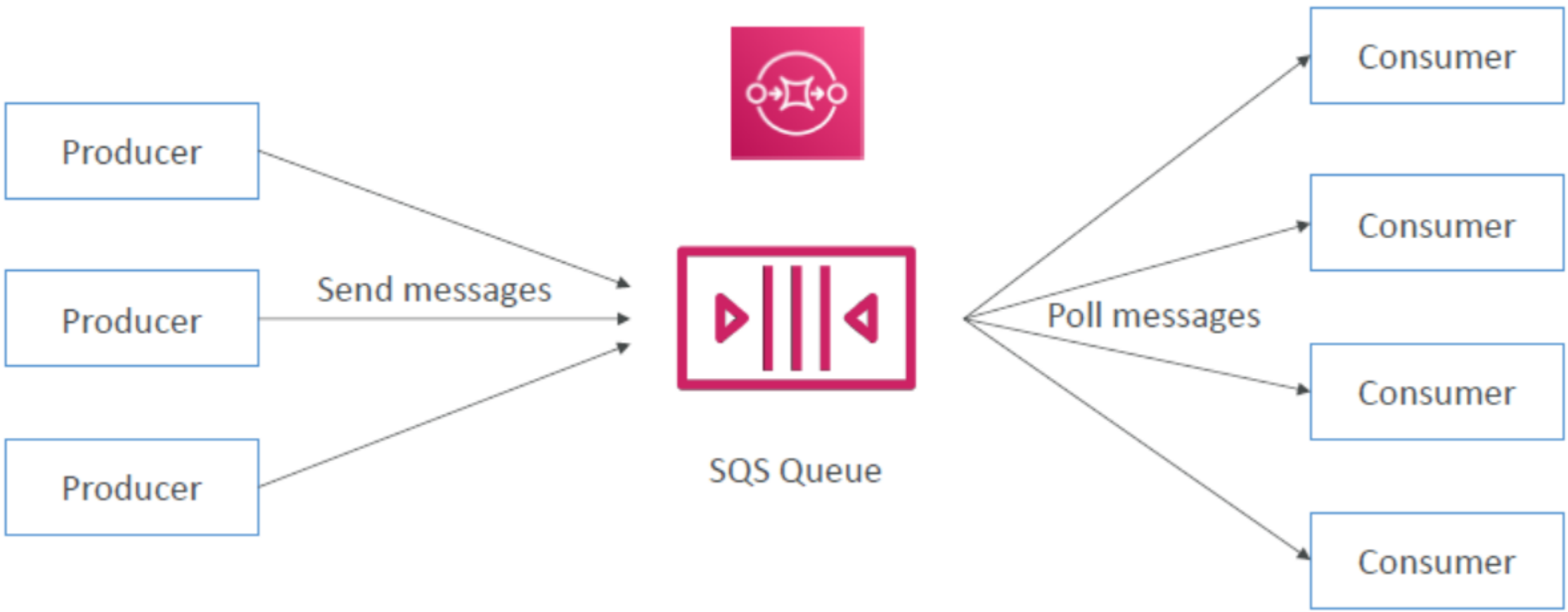
What is Amazon SQS

- Amazon Simple Queue Service (Amazon SQS) is a fully managed message queuing service that makes it easy to decouple and scale microservices, distributed systems, and serverless applications.



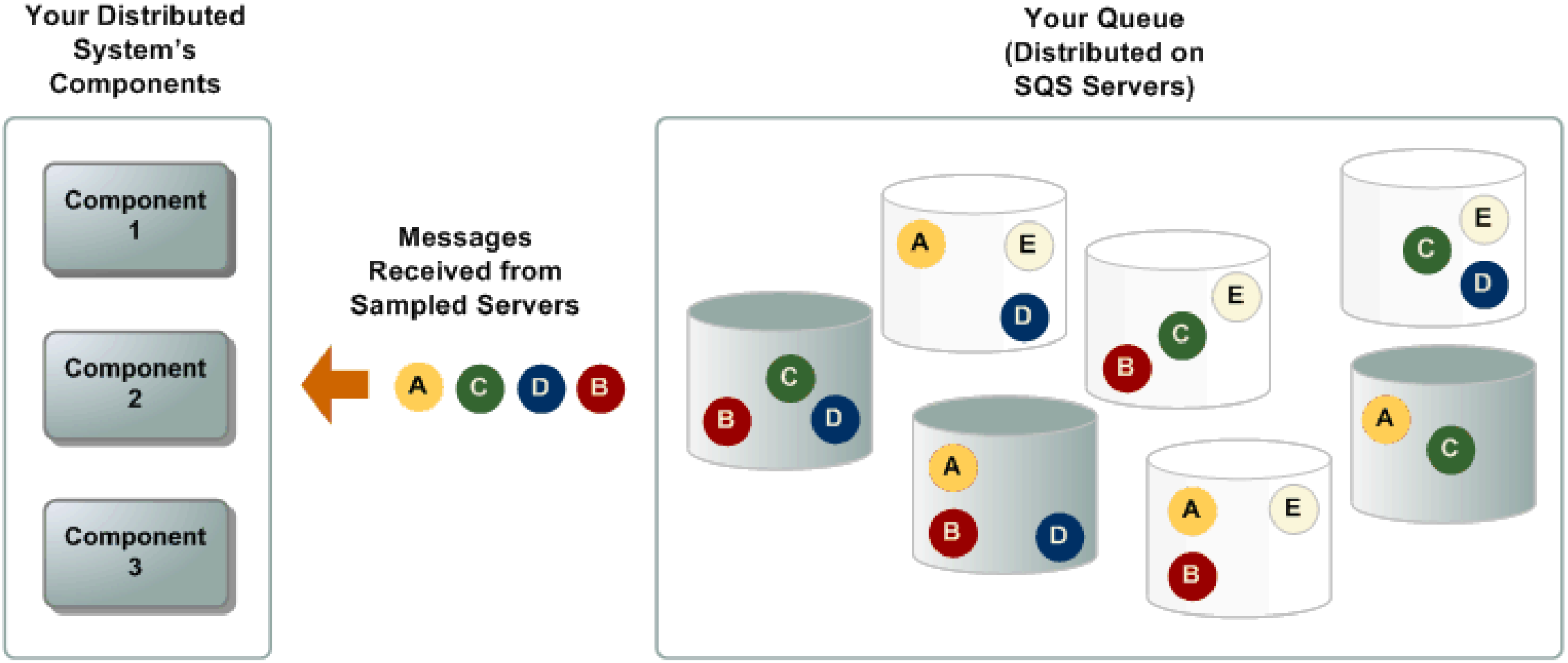
Amazon SQS

- It makes simple and cost effective to decouple the components of a cloud application.
- You can use Amazon SQS to transmit any volume of data, at any level of throughput.
- Using Amazon SQS, you can store application messages on reliable and scalable infrastructure.
- An Amazon SQS queue is basically a buffer between the application components that receive data and those components that process the data in your system.
- If your processing servers cannot process the work fast enough, the work is queued so that the processing servers can get to it when they are ready.


Producer
Producer
Producer
Send messages
SQS Queue
Poll messages
Consumer
Consumer
Consumer
Consumer

Amazon SQS – Standard Queue

- Unlimited throughput, unlimited number of messages in queue
- Default retention of messages: 4 days, maximum of 14 days
- Low latency (<10 ms on publish and receive)
- Limitation of **256KB** per message sent
- Message retention: **default 4 days, up to 14 days**
- Can have duplicate messages (at least once delivery, occasionally)
- Can have out of order messages (best effort ordering)
- Poll SQS for messages
- Process the messages (example: insert the message into an RDS database)
- Delete the messages using the DeleteMessage API

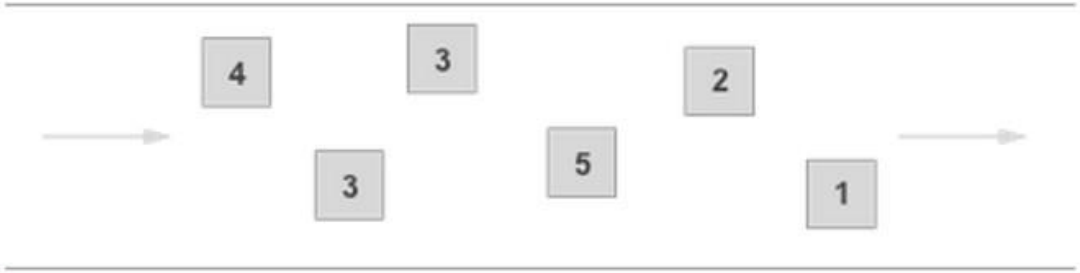

Your Distributed System's Components
Component 1
Component 2
Component 3
Messages Received from Sampled Servers
A
C
D
B
Your Queue (Distributed on SQS Servers)
A
E
D
C
B
E
C
D
E
C
D
B
A
C
A
B
D
A
E
B
A
C

Amazon SQS – FIFO Queue

- FIFO = First In First Out (ordering of messages in the queue)
- Messages are processed in order by the consumer
- Duplicates aren't introduced into the queue.

Standard Queue

- **High Throughput:** Standard queues have nearly-unlimited transactions per second (TPS).
- **At-Least-Once Delivery:** A message is delivered at least once, but occasionally more than one copy of a message is delivered.
- **Best-Effort Ordering:** Occasionally, messages might be delivered in an order different from which they were sent.

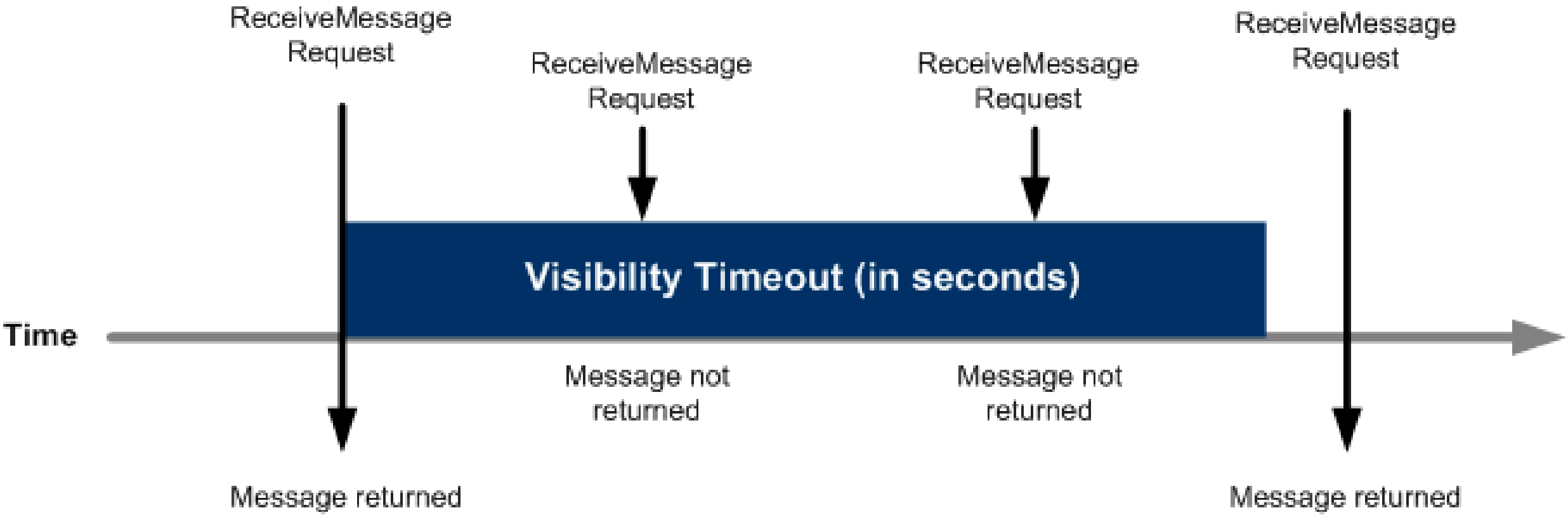


FIFO Queue

- **First-In-First-Out Delivery:** The order in which messages are sent and received is strictly preserved.
- **Exactly-Once Processing:** A message is delivered once and remains available until a consumer processes and deletes it. Duplicates are not introduced into the queue.
- **Limited Throughput:** 300 transactions per second (TPS).

SQS – Message Visibility Timeout

- When a consumer receives and processes a message from a queue, the message remains in the queue.
- Amazon SQS doesn't automatically delete the message.
- The consumer must delete the message from the queue after receiving and processing it.
- To prevent other consumers from processing the message again, Amazon SQS sets a visibility timeout, a period of time during which Amazon SQS prevents other consumers from receiving and processing the message.
  - Default visibility timeout: 30 seconds.
  - Minimum is: 0 seconds
  - Maximum is: 12 hours
- A consumer could call the **ChangeMessageVisibility** API to get more time


ReceiveMessage Request
ReceiveMessage Request
ReceiveMessage Request
ReceiveMessage Request
Visibility Timeout (in seconds)
Time
Message not returned
Message not returned
Message returned
Message returned

Dead Letter Queue

- If a consumer fails to process a message within "Visibility Timeout" the message goes back to the queue.
- We can set a threshold of how many times a message can go back to the queue
- After the MaximumReceives threshold is exceeded, the message goes into a dead letter queue (DLQ)
- Useful for debugging!
- Make sure to process the messages in the DLQ before they expire

Amazon SQS delay queues

- Delay queues let you postpone the delivery of new messages to a queue for a number of seconds
- Delay a message (consumers don’t see it immediately) up to 15 minutes
- Default is 0 seconds (message is available right away)
- Can set a default at queue level
- Can override the default on send using the DelaySeconds parameter

Amazon SQS - Long Polling

- Amazon SQS provides polling to receive messages from a queue
  - Short polling – **default**
    - **Amazon SQS sends the response right away**
  - Long polling
    - SQS sends a response after it collects at least one available message
    - SQS sends an empty response only if the **polling wait time expires**.
- LongPolling decreases the number of API calls made to SQS while increasing the efficiency and latency of your application.
- The wait time can be between 1 sec to 20 sec (20 sec preferable)
- Long Polling is preferable to Short Polling
- Long polling can be enabled at the queue level or at the API level using WaitTimeSeconds