

A Survey of Distributed Message Broker Queues

Vineet John
University of Waterloo
v2john@uwaterloo.ca

Xia Liu
University of Waterloo
x397liu@uwaterloo.ca

ABSTRACT

This paper surveys the message brokers that are in vogue today for distributed communication. Their primary goal is to facilitate the construction of decentralized topologies without single points of failure, enabling fault tolerance and high availability. These characteristics make them optimal for usage within distributed architectures. However, there are multiple protocols built to achieve this, and it would be beneficial to have an empirical comparison between their features and performance to determine their real-world applicability.

This paper focuses on two popular protocols (Kafka and AMQP) and explores the divergence in their features as well as their performance under varied testing workloads.

KEYWORDS

distributed message broker, message queue, kafka, amqp, rabbitmq

1 INTRODUCTION

Distributed Message Brokers are typically used to **decouple** separate stages of a software architecture. They permit communication between these stages **asynchronously**, by using the publish-subscribe paradigm.[1]. These message brokers are also finding new applications in the domain of IoT devices and may also be used as method to implement an event-driven processing architecture.

A few of the current offerings of this architectural paradigm include Apache Kafka[2], AMQP[6], ActiveMQ[4].

The rest of this paper is structured as follows:

- Section 2: Research questions investigated
- Section 3: Kafka - architecture and overview
- Section 4: AMQP - architecture and overview
- Section 5: Benchmarking results
- Section 6: Kafka vs. AMQP - performance comparison
- Section 7: Conclusion and final remarks
- Section 9: Appendix

2 RESEARCH QUESTIONS

The following research questions are enumerated for further study:

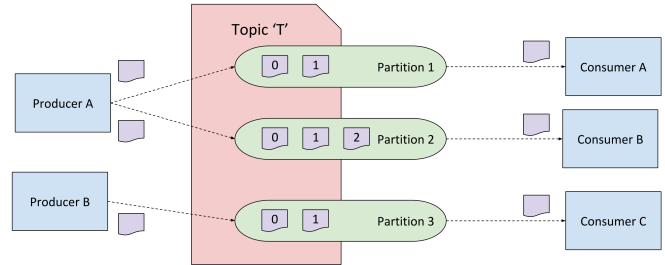


Figure 1: Kafka Architecture

RQ0 What are the message broker implementations commonly in use today?

RQ1 What are the common requirements for the implementation of message queues?

RQ2 What are the divergent functionalities in the current message queue offerings?

RQ3 How do each of the implementations offer reliability, partitioning and fault tolerance?

3 KAFKA

Kafka was developed at LinkedIn and primarily used for **log processing**. This worked well for Kafka's user engagement metrics collection use-case. The fundamental features behind Kafka are **performance over reliability** Tradeoff and it offers high throughput, low latency message queuing. Kafka can be used either for online or offline log processing.

The fact that reliability of message reception is traded off to some extent implies the loss of a single record among a multitude is not a huge deal-breaker. The rationale behind this is, for log aggregated data, delivery guarantees are unnecessary.

3.1 Kafka - Architecture

The general architecture for Kafka is illustrated in Figure 1.

Data is divided into topics, which resemble streams of messages. Topics are further divided into partitions and each broker can possess one or more such partitions. Producers publish to topics and brokers store messages received from the producers. Messages are payloads of bytes that consumers use iterators to consume. Kafka employs the pull model for consumer message receipt,

which permits consumers to consume messages at their own pace.

Kafka supports 2 models for topic subscription:

- Point-to-point: A single message copy is read by an arbitrary consumer.
- Pub-sub: Each consumer has it's own copy of each message.

4 AMQP

It is an asynchronous message queuing protocol, aiming to create an open standard for passing messages between applications and systems regardless of internal design. It was initially designed for financial transaction processing systems, such as trading and banking systems, which require high guarantees of reliability, scalability, and manageability. This use-case greatly influences the design of AMQP.

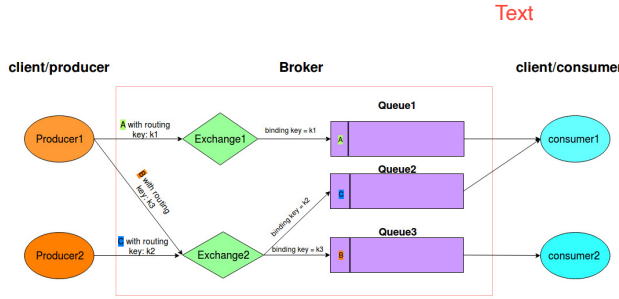


Figure 2: AMQP Workflow

4.1 AMQP Architecture

Figure 2 shows the architecture of AMQP. There are several concepts in AMQP need to be listed and explained before illustrating how an AMQP broker works. These terms are elaborated upon in the Appendix (Section 9C.1).

Instead of sending messages to queues directly, producers send messages to **exchanges**. Exchanges then send messages to queues whose **binding key** matches the messages' **routing key**. The queue, in turn, sends messages to consumers who are subscribed to it. This design of exchanges and queues in AMQP makes the message passing transparent. A producer does not need to maintain any state about its messages' consumers. Similarly, a consumer does not need to maintain state about the producers. The decisions of which queue the message should be routed to is decided by the exchange instead of the application, which makes changing message-routing and delivery logic easier compared to a messaging system where the application makes such decisions.

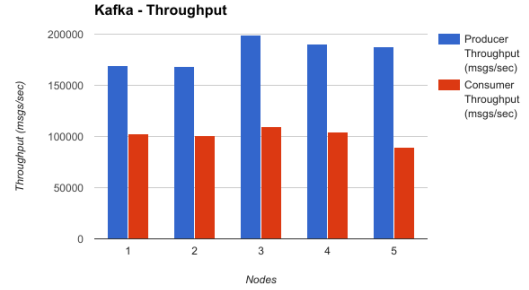


Figure 3: Kafka Throughput (Single Producer/Consumer per node)

5 EXPERIMENTAL RESULTS

5.1 Testbed

The testbed utilized to perform the benchmarking is comprised of 5 nodes, each with a similar hardware configuration: 12 cores @ 2300.13Mhz, 16GB memory, with an HDD for persistent storage, and a network interface capacity of 1Gbps.

There were 2 types of tests run for Kafka and RabbitMQ (which is the implementation used to proxy the behavior of AMQP). The single producer/consumer test keeps the overall test workload constant (1 million messages, 50B each) and scales the queue deployment from 1 to 5 nodes. The multiple producer/consumer setup keeps the number of nodes constant and scales the number of producers/consumers connecting from each node.

All the benchmarks were run using a modified version of Flotilla¹, which is a message broker benchmarking tool written in Go.

Figures 3, 4, 5 and 6 illustrate the throughput and latency of Kafka under the aforementioned workloads. Similar benchmark results for RabbitMQ are shown in figures 7, 8 and 9.

5.2 Benchmarking Results

The results of benchmarking Kafka while keeping the same payload and scaling to multiple nodes (Fig. 3, Fig. 5) shows a latency drop by a factor of about 3x, in comparison to only a nominal penalty on throughput (1.06x). This just goes to demonstrate the effectiveness of Kafka being deployed in a cluster setup in handling large log-aggregation-like workloads.

The corresponding results while keeping the number of nodes constant and increasing the payload by using additional producers and consumers causes the performance metrics to degrade in a more pronounced manner

¹<https://github.com/v1n337/flotilla>

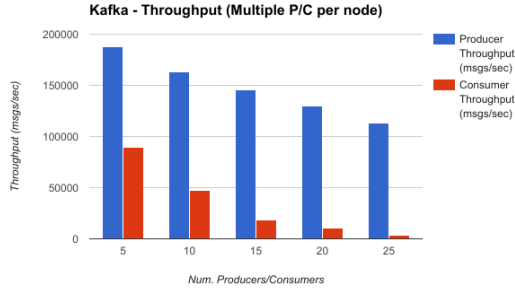


Figure 4: Kafka Throughput (Multiple Producers/Consumers per node)

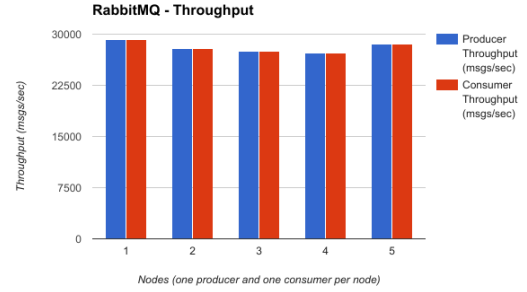


Figure 7: RabbitMQ Throughput (Single Producer/Consumer per node)

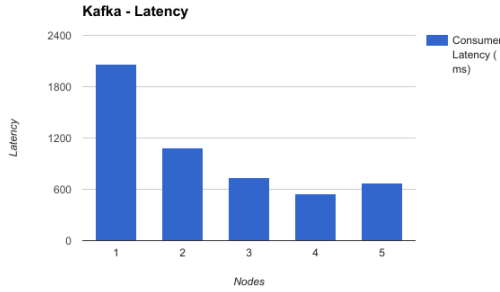


Figure 5: Kafka Latency (Single Producer/Consumer per node)

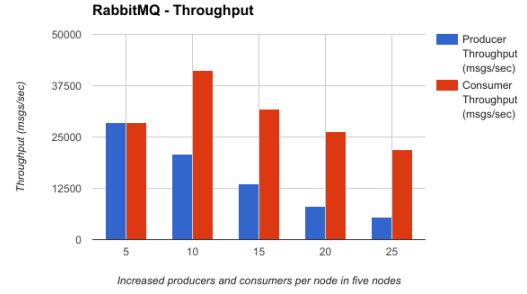


Figure 8: RabbitMQ Throughput (Multiple Producers/Consumers per node)

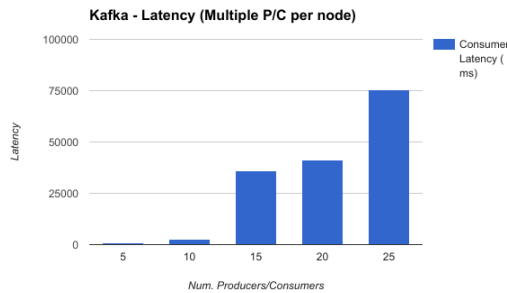


Figure 6: Kafka Latency (Multiple Producers/Consumers per node)

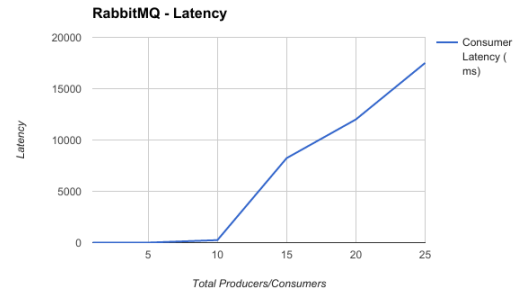


Figure 9: RabbitMQ Latency

(Fig. 4, Fig. 6). In particular, the latency increases by about 2 orders of magnitude with a 5x increase in payload coming from a 5x increase in publishers and consumers. This also corroborates the 29x drop in consumer throughput, meaning that messages stay in the Kafka log partitions (queues) a lot longer and remain unconsumed, than if there was only a single publisher/subscriber per

node. This is especially visible due to the huge disparity between the producer throughput and the consumer throughput. This could be attributed to an increased number of accesses of the metadata layer i.e. Zookeeper becoming a focus point of resource contention on a single node, causing context switching and resulting in additional overhead.

Figure 7 describes RabbitMQ clients' throughput when each node runs single producer and consumer, similar

to the setup for Kafka. The reason for this stable producer throughput could be attributed to a high exchange routing speed. The rate at which the exchange processes messages is faster than the rate of producers producing and sending messages. Another cause could be that the broker is capable of matching this rate resulting in a consumer throughput that is almost the same as producer throughput. Consumers consume messages faster than producers are able to publish messages to the exchanges. Therefore, queues can be empty, and consumers have to wait for messages from producers. As a result, producer throughput is a ceiling for consumer throughput in this experiment.

Figure 8 shows RabbitMQ clients' throughput with the rapid growth of clients. In general, scaling the number of producers and consumers resulting in a reduction in throughput, while consumers throughput is not limited by a single producer anymore and can consume messages sourced from multiple producers at a higher rate. The overall reduction in throughput can be attributed to resource contention, similar to what Kafka encounters.

Figure 9 illustrates the consumer latency with the increasing number of total clients. The latency remain low for benchmarks using a single producer and consumer per node. The main reason for low latency at the beginning and sharp increased latency is due to the queues being backed-up with messages from multiple producers as opposed to almost empty. RabbitMQ, and all distributed queues in general, operate at optimal latency when queues are relatively empty. When the number of producers increase to 15, the exchange is saturated with message traffic, and messages start queuing for longer durations, thereby increasing the latency.

6 COMPARISON

The results of Kafka and AMQP benchmarking test reveal that Kafka has higher throughput while AMQP has lower latency.

There are a few potential reasons for why Kafka provides a higher throughput, explained by its inherent optimizations:

- The **SendFile API** in Kafka helps bypass kernel buffers to write data directly from the file channel to the socket.
- Writing to a message broker in Kafka uses sequential disk writes and **page-level caching** also help to maximize throughput.
- **Batching** in Kafka is available out-of-the-box, and this helps minimize communication overhead.

In terms of latency, the consumer **push-model** option results better mean latency in AMQP; while Kafka uses the **pull-model** where consumers have to fetch messages from broker. Another reason is that AMQP's default configuration **doesn't persist messages to disk** at the broker, but Kafka does, resulting in a longer intermediate delay before message consumption.

7 CONCLUSION

Based on the experimental results and comparisons in sections 5 and 6, a distinct contrast between each of the message queue offerings can be made.

Two key aspects should be considered making a choice of message broker protocol, namely **throughput** and **reliability**. If reliability for the application is not critical, Kafka is the better choice, as evidenced by the experimental results. For example, for an application that needs to process a large amount of messages is tolerant to losing a few. Usage examples include page view counters, advertisement clicks and social media reactions.

On the other hand, if messages are important, such as financial transactions, the cost of losing any of messages is far higher than not achieving an optimal throughput and the application is encouraged to use AMQP. In addition, AMQP can encrypt messages out-of-the-box. Therefore, applications requiring security should also consider using AMQP. Moreover, four types of exchanges in AMQP allow AMQP broker work in peer-to-peer, publisher/consumer and broadcasting models. The last two models are suitable for instant messaging and notification services.

This concludes the study of two popular distributed message broker queuing protocols, Kafka and AMQP. Future work would entail exploring other implementations such as ZeroMQ, ActiveMQ and Qpid in depth. For the purposes of this paper, these implementations have been summarized at a high level in the Appendix (Section 9D).

8 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Samer Al-Kiswany, for his guidance and supervision on this project, as well as for his constant encouragement.

REFERENCES

- [1] Guruduth Banavar, Tushar Chandra, Robert Strom, and Daniel Sturman. 1999. A case for message oriented middleware. In *International Symposium on Distributed Computing*. Springer, 1–17.
- [2] Jay Kreps, Neha Narkhede, Jun Rao, and others. 2011. Kafka: A distributed messaging system for log processing. In *Proceedings of the NetDB*. 1–7.

- [3] Inc. or its affiliates Amazon Web Services. 2017. What is Amazon Simple Queue Service? (2017). <http://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/Welcome.html>
- [4] Bruce Snyder, Dejan Bosanac, and Rob Davies. 2011. Introduction to apache activemq. (2011).
- [5] Martin Sústrik. 2015. ZeroMQ. (2015).
- [6] Steve Vinoski. 2006. Advanced message queuing protocol. *IEEE Internet Computing* 10, 6 (2006).

9 APPENDIX

A FEATURE COMPARISON

Table 1 summarizes, at a high-level, the differences between Kafka and AMQP. These differences illustrate that the goal for Kafka is to focus on high throughput while AMQP tries to achieve better reliability. For example, AMQP can use consumer ACKs to ensure reliability; while, Kafka’s highest guarantee is an ACK for whether the message was written persistently to the broker log. This option increases AMQP reliability, but decreases its throughput.

AMQP acks is one of the reasons that reduce its throughput

B KAFKA

B.1 Kafka - Storage

- Each topic is a logical log. Physically, each log is a segment of a partition, each segment being of roughly the same size.
- Flow of Data:
 - Producer publishes message
 - Broker appends message to the latest segment
 - Messages are buffered/batched before being flushed to the segment
 - Message is visible to the consumers after flushing
- No message IDs used. A message is identified as it’s logical offset from the start of the logical log.
- Kafka maintains in-memory, a list of the start of the segments that mark the start of each segment file.
- Consumers are responsible for calculating the offset, and determine the next offset to request. Consumers can do a batched-pull if needed. Batched-pull would be done in terms of data-size, not number of messages.
- No broker level caching is used. Kafka relies on OS-level page caching. ‘Write-through’ and ‘Read-ahead’ are the caching strategies that work well-enough. This is because, typically, a consumer only lags producers by a small amount of messages, hence, most messages are served from the broker.

Characteristic	Kafka	AMQP
Message Format	Bytes - Easier to develop for	Binary - better compression, boosts throughput
Message Routing	No intermediaries. Messages are sent to brokers	Exchanges used to route messages using bindings
Message Reliability	Unreliable - the sender doesn’t receive an ACK	Reliable - ACKs are sent on receipt of messages
Message Batching	Available out-of-box	Difficult to implement
Virtual Hosts	Not present	Used to ensure segregation of clients
Basic Distributed Unit	Topic	Queue
Message Delivery Model	Only pull model available	Both push and pull models are available
Consumer Subscription Model	Point-to-point and Pub-sub models available	Depending the type of exchange, both Point-to-point and Pub-sub models are implementable
Message Persistence	Writes to a persistent file system log using the page-cache	Durability is a configuration option while creating a queue
Application Domain	Log Aggregation, Big Data Analytics - high throughput, weaker durability	Financial services, Stock, Banking - strong durability guarantees

Table 1: Message Broker Comparison

- The SendFile API is used to bypass the application and kernel buffers to write data directly from the file channel to the socket,
- Brokers don’t maintain state of any consumers. The data longevity on a broker is a time-based SLA configuration parameter. Consumers need to maintain their own current offset. This allows consumers to replay messages from a previous offset if needed.

B.2 Kafka - Co-ordination

- Consumer Groups are logically a single consumer, and can jointly consume messages published to a topic. Each message is read by only 1 consumer in the group.
- The minimum unit of parallelism in Kafka is defined to be a partition.
- There is always at least 1 consumer per partition of a topic. This implies that there is no co-ordination needed between consumers within a consumer group for the last offset read from a partition.
- Typically, $partitionCount > consumerCount$
- Kafka doesn't use the concept of a leader/master node.
- Zookeeper is used to provide a consensus service. The consensus to be arrived at is which consumers are the nodes that are participating as brokers and consumers, the topics that are currently active, and the offset of the last consumed message of each partition. Zookeeper is also used to trigger the re-balancing process for consumers when any change to the list of brokers and consumers are detected for a given topic.
- Zookeeper uses a file-system like API, with data stored at paths. Paths can either be persistent or ephemeral (liveness depends on the creating client). It is fully replicated across the hosts chosen as Zookeeper nodes.
- There are 4 types of registries store in Zookeeper, along with
 - **Broker**: host name, port, set of topics and partitions stored
 - **Consumer**: Consumer group name, topics subscribed to.
 - **Ownership**: Mapping from a partition path to the path of the consumer that 'owns' a particular partition.
 - **Offset**: Mapping from a partition path to the offset of the last message that was read from the partition.
- Each consumer registers a Zookeeper watcher on the brokers and consumers that correspond to the topics that it is currently subscribed to. If there is any change, then the re-balancing process is triggered.

Re-balancing Process: This is initiated when there is a change in the broker and consumer node list for a given topic.

- Remove the currently subscribed partitions for each of the consumers in the consumer group from the Ownership registry.
- List all the partitions for the topics that the consumer group is subscribed to.
- Divide the partitions equally amongst the consumers, and update the Ownership registry.
- Read the current offset for each of the assigned partition and resume reading and processing the messages from the point where it was left off.
- Since there are situations in which the change notifications arrive at different times at different consumers in a group, a consumer might attempt to take ownership of a partition that another consumer has not relinquished ownership of yet. In this case, the first partition wait and retries.

B.3 Kafka - Delivery Guarantees

- Messages are relayed 'at least once'.
- 2PC is required if 'exactly once' semantics are needed, which is deemed to be overkill for Kafka.
- If needed, the application needs to handle deduplication.
- In-order delivery of messages is guaranteed from a single partition, due to the monotonically increasing offset of each log message.
- CRCs are stored for each message to validate against data corruption. These CRCs are logged.

C AMQP

C.1 AMQP Components

- **Broker**: it is the server side of AMQP, receiving from producers and sending messages to right consumers.
- **Producer**: Client of AMQP who sends messages to a broker.
- **Consumer**: Client of AMQP who receives messages from brokers.
- **Exchange**: Part of the broker, receiving messages from producers, and routing them to queues. It is created by clients.
- **Queue**: Part of the broker, receiving messages from exchanges and sending them to consumers. It is created by clients.
- **Virtual host**: Namespace for broker to specify the entities, such as exchanges and queues, application refers to. It is used to achieve access controls in broker. Clients need user account and password to get access to a virtual host in order to receive or send messages to broker.

- Binding: Relationship between exchange and queue, created by clients with specified binding keys.
- Routing key: Key specified by a producer and included in messages.

C.2 AMQP - Features

AMQP has some features that make it appealing to application developers. The most critical features include a choice of exchange types, message persistence, and distributed brokers.

There are four types of exchanges defined in AMQP, including direct exchange, topic exchange, fanout exchange, and header exchange.

- Direct exchange sends messages to different queues according to the routing key. Messages are load balanced between consumers instead of queues.
- Topic exchange sends messages to all queues whose binding key matches routing key. It is used to implement publish/subscribe model.
- Fanout exchange acts like a broadcaster and does not consider routing keys. The exchange just sends received messages to all queues which are bound to the exchange.
- Header exchange uses message headers to route message to queues instead of routing key.

AMQP also offers strong reliability guarantees. For each exchange and queue in the broker, messages can be persisted in disk by toggling the durability flag.

In terms of distributed brokers, AMQP provides cluster, federation and shovel methods to achieve high availability and message brokers distribution. Nodes within the cluster are mirrors of each other, except queues which are local to each node. Federation and shovel architectures are used to allow messages across a WAN.

C.3 AMQP - Failure Handling

When broker goes down and restarts, persistent queue, whose durable feature is set as true, will be recovered. However, only the persistent messages in the durable queue can be recovered. When the message is sent to the exchange, producer can assign delivery mode as 2 which means the message is persistent.

D OTHER MESSAGE BROKER IMPLEMENTATION

D.1 Apache ActiveMQ

- Built on top of the JMS specification, which an enterprise messaging standard included in Java EE.

- Supports a variety of message broker protocols, including Stomp, AMQP and MQTT, in addition to traditional communications protocols like TCP, UDP and XMPP and also multiple topologies. The architecture is illustrated in Figure 10.

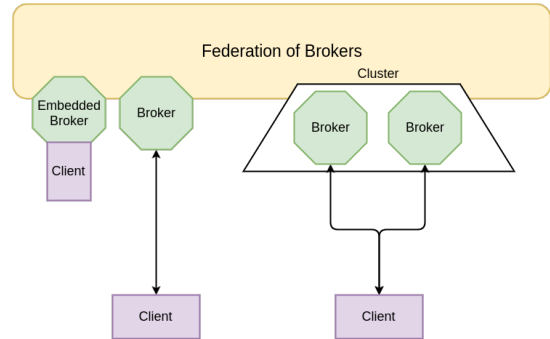


Figure 10: ActiveMQ Architecture

- ActiveMQ brokers can also be set up like a federated cluster, similar to RabbitMQ.
- Active MQ is under active development after being incubated at Apache. There have been 79 commits to the source in the past 3 months. (beginning December 17, 2016).²

D.2 ZeroMQ

- Unlike traditional message queuing systems, ZeroMQ is broker-less.
- "Smart endpoint, dumb network", was the philosophy behind ZeroMQ[5]
- It was originally designed for stock-trading messages, with the objective of circumventing a central broker as a means of routing the messages.
- Similar to ActiveMQ, it is a library, not a protocol.
- Figure 11 illustrates an example ZeroMQ architecture with node communicating between themselves without an arbiter, but merely a directory service to facilitate lookups.
- Connections between ZeroMQ clients can be established with patterns like fan-out, pub-sub, task distribution, and request-reply.
- ZeroMQ is under active development. There have been 143 commits to the source in the past 3 months. (beginning December 17, 2016).³

²<https://git-wip-us.apache.org/repos/asf?p=activemq.git>

³[git://github.com/zeromq/libzmq](https://github.com/zeromq/libzmq)

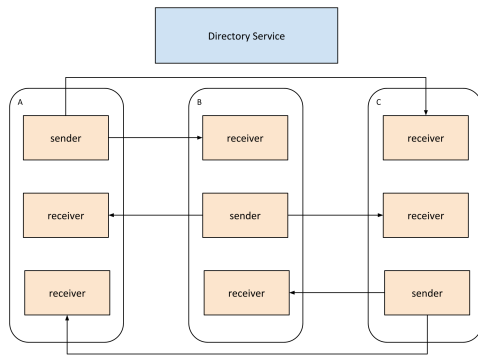


Figure 11: ZeroMQ Architecture

D.3 RabbitMQ

- RabbitMQ is a popular implementation of AMQP. It is developed by VM-Ware in Erlang.
- Compared to Apache Qpid, it is easy to use and deploy, and the code is short and clear.
- Some well-known RabbitMQ users include Mozilla, AT&T and UIDAI.
- RabbitMQ community and supports are active. There have been 294 commits to the source in the past 3 months. (beginning December 17, 2016). footnote <https://github.com/rabbitmq/rabbitmq-server>

D.4 Apache Qpid

- It is an implementation of AMQP, and provides two messaging broker version, one is in Java, the other is in C++.
- It uses Corosync to achieve high availability. Similar to RabbitMQ, Qpid uses broker federation to achieve scaling of capacity.
- It stores its queues in database or memory.
- There have been 52 commits to the source in the past 3 months. (beginning December 17, 2016).

D.5 Amazon SQS

Amazon SQS is a message queuing service which provides scalable and highly-available queues to store messages so that they can be produced and consumed between components over the distributed system. It runs within Amazon high-availability data centers, so queues will be available whenever applications needs. All messages are stored redundantly across multiple servers and data centers, providing high availability. Amazon SQS provides standard queues in all regions and FIFO queue

only in US West and East. Therefore, for most users, queue type they can use is standard queue, which delivers messages at least once and might deliver messages in different order from which messages were sent (best-effort ordering).

The official website recommends four cases to use Amazon SQS[3]:

- Tracking the success of each component of an application independently.
- Scheduling jobs in the queue with a delay individually. (up to 15 minutes with standard queues)
- Increasing concurrency or throughput dynamically at read time with no pre-provisioning.
- Scaling transparently as requests and load changes with no provisioning instructions.

D.6 Microsoft Message Queuing (MSMQ)

- Message Queuing (MSMQ)⁵ technology enables applications running at different times to communicate across heterogeneous networks and systems that may be temporarily offline.
- Applications send messages to queues and read messages from queues. The following illustration shows how a queue can hold messages that are generated by multiple sending applications and read by multiple receiving applications.
- Message Queuing provides guaranteed message delivery, efficient routing, security, and priority-based messaging.
- It can be used to implement solutions to both asynchronous and synchronous scenarios.

⁴<https://git-wip-us.apache.org/repos/asf/qpid-jms.git>

⁵<https://msdn.microsoft.com/en-us/library/ms711472.aspx>