✓   100 XP  ▶

# Summary

2 minutes

- Message queues are communication mechanisms used to enable indirect, asynchronous communications by partitioning and storing messages on brokers. This allows easy horizontal scaling of the messaging layer.
- Kafka is a multi-subscriber message queue developed at LinkedIn. Consumers of this queue can choose to subscribe to topic(s), and the queue is guaranteed to receive messages in the order sent.
- Stream processing systems operate on an infinitely long, often fast-moving set of input records, such as the output of a message queue. To reduce latency, there is a set of simple rules that this kind of system could follow.
- Stream processing jobs could be stateless (simply applying pre-defined rules to an input) or stateful (applying continuously changing rules based on past data and current status).
- Samza is a stream processing framework developed at LinkedIn. By default, Samza runs cgroups containers scheduled over YARN and reads from a Kafka stream, allowing programmers to use a custom API to define streaming tasks. When local state is needed, an embedded RocksDB instance is used.
- The Lambda and Kappa architectures are two methods of working with data pipelines with different latency requirements.

## Module complete:

Unlock achievement