Gateways and Stream Replication

When replicating streams, MapR-ES replicates messages that are published to a source stream. Gateways are services that receive messages from source streams and publish them in replica streams.

You configure gateways on nodes that are in destination clusters. On source clusters, you list the destination clusters and the gateways that are running on them.

For information on configuring and managing gateways, see:
- Configuring Gateways for Table and Stream Replication (ConfiguringMapRGatewaysForTRAndI.html)
- Managing Gateways (Gateways-managing.html)

During replication, MapR-ES sends messages from source streams to the gateways on the destination clusters, where the replicas of those source streams are located. Gateways batch the messages and then apply them to replicas. All messages from a source stream arrive at a replica after having been authenticated at a gateway. Therefore, access control expressions on the replica that control permission to publish messages are irrelevant; gateways have the implicit authority to publish messages to replicas.

MapR-ES distributes messages to a destination cluster's gateways in round-robin fashion. If a gateway is down or unreachable, MapR-ES chooses another gateway. If all of the gateways are down, MapR-ES retries the operation periodically until a gateway comes online.

You must configure gateways in destination clusters. If the destination cluster is remote from the cluster in which a source stream is located, then the gateways must be in the remote cluster. If the destination cluster is the source cluster, meaning that a source stream and its replica are located in a single cluster, then the gateways must be in the local cluster.

For more information about replicating streams, see Stream Replication (../MapR_Streams/replicating_streams.html).
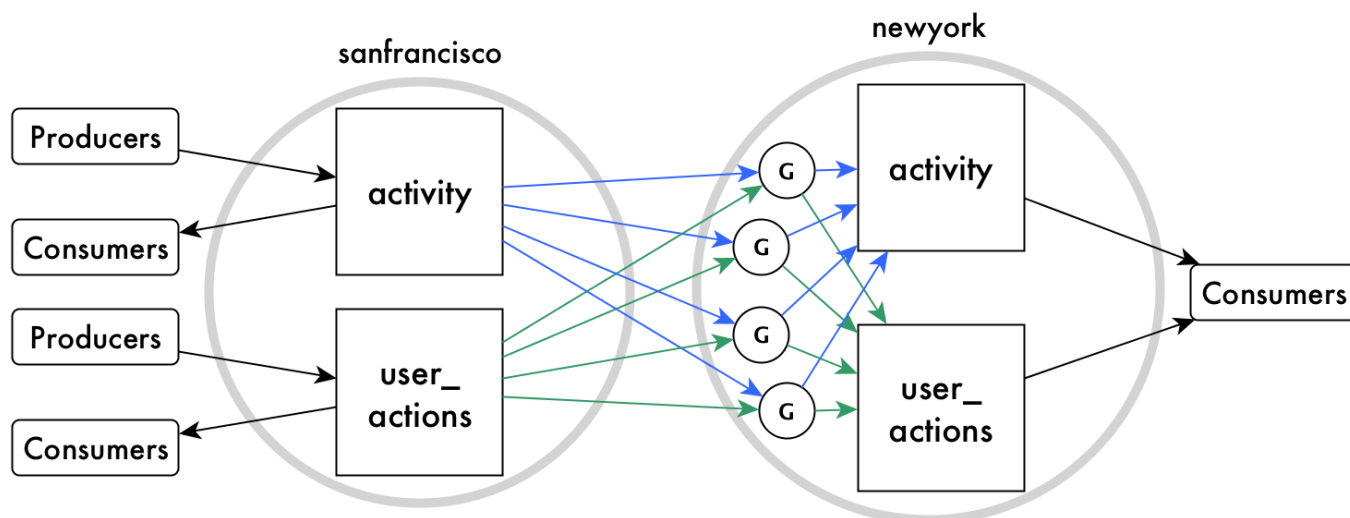
⌄ Gateways on nodes in remote destination MapR clusters

In this type of topology, gateways receive messages that are published to source streams, authenticate with the destination cluster on behalf of the source cluster, and publish the messages to the corresponding streams.

This diagram of basic intercluster master-slave replication shows messages from the `activity` stream in the cluster `sanfrancisco` being sent to gateways. The gateways then publish the messages to the replica stream that is in the cluster `newyork` .
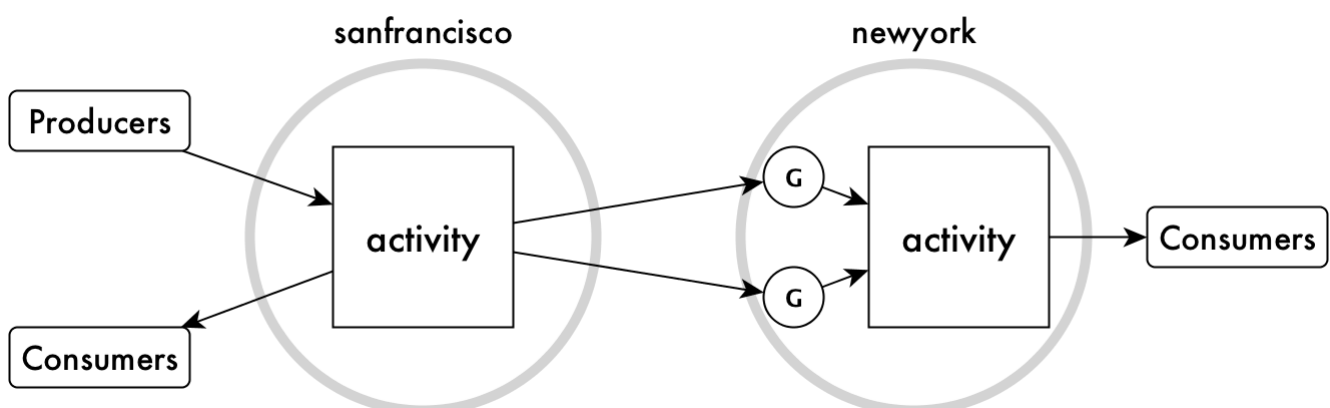
The gateways on a destination cluster are not assigned to particular replicas. They publish messages to all replicas on the destination cluster. For example, in this diagram messages from two source streams in the cluster `sanfrancisco` are being replicated to two replicas in the cluster `newyork` . There are four gateways. Each gateway receives messages from both source streams, and each gateway applies those messages to the corresponding replicas.

˅ Gateways on nodes within a MapR cluster serving as source and destination

In this type of topology, gateways again receive messages that are published to source streams and publish the streams to the replicas. However, all of this activity takes place within a single MapR cluster.

This schematic diagram of basic intracluster master-slave replication shows messages from the `activity1` stream in the cluster `sanfrancisco` being sent to gateways. The gateways then publish the messages to the stream `activity2` .