Apache Pulsar Geo-Replication

Geo-replication is a typical mechanism used to provide disaster recovery.

Generally, any database or message bus solution replicates data between two data centers.

Pulsar supports multi-datacenter replication(n-mesh) with the below strategies:

- Asynchronous Replication
- Synchronous Replication

Global clusters can be configured at namespace level to be replicated within any number of clusters.

In normal cases, when there are no connectivity issues, messages are replicated immediately, at the same time as they are dispatched to local consumers.

Typically, end-to-end delivery latency is defined by the network round-trip time (RTT) between the data centers.

Applications can create producers and consumers in any of the clusters, even when the remote clusters are not reachable (for example, during a network partition).

Asynchronous geo-replication provides lower latency but may result in weaker consistency guarantees due to the potential replication lag that some data hasn't been replicated.

From the below example, Datacenter C does not have a consumer, but it's message can still be consumed in Datacenters A or B based on the subscription model.