KIP-735: Increase default consumer session timeout

Status

Current state: Adopted

Discussion thread: here

JIRA: KAFKA-12874 - Getting issue details... STATUS

Please keep the discussion on the mailing list rather than commenting on the wiki (wiki discussions get unwieldy fast).

Motivation

The consumer session timeout is a crucial configuration for group stability. When there is a member failure, the group must pause in order to rebalance partitions. If the failure is spurious, then we often get two rebalances: one for the member failure and one for the failed member when it rejoins. Spurious failures may be unlikely when the cluster has dedicated resources and has been properly tuned with a relatively static load requirement. However, as explained in KIP-537, multi-tenant cloud environments with dynamically changing workloads are becoming the norm. In practice, we find that transient network/load failures are much more common than genuine client failures, so we propose to increase the consumer's session timeout from 10s to 45s to avoid the likelihood of spurious failures.

A second related issue concerns the consistency of configurations. By default, the consumer uses a 30s `request.timeout.ms`, which is consistent with both the producer and the admin client. However, it does not work well with the 10s default for `session.timeout.ms`. When the connection to the coordinator is not closed cleanly, the consumer will wait 30s before it disconnects and retries. By the time that happens, the group often has already kicked the member out and moved on. By increasing to 45s, we allow enough time to reconnect and retry after one request timeout.

Public Interfaces

We propose to increase `session.timeout.ms` in the consumer from 10s to 45s. We are also making changes to the behavior of `group.min.session.timeout.ms` and `group.max.session.timeout.ms` as described below.

Note that we are not making any change to `heartbeat.interval.ms`. Although the increased session timeout allows for less frequent heartbeats, the heartbeat also serves the purpose of discovering that a rebalance is in progress.

Proposed Changes

There are no changes here beyond the change to the default session timeout.

Compatibility, Deprecation, and Migration Plan

We don't foresee any complications with compatibility. New clients will take the new default and old clients will continue to use the old value.

Rejected Alternatives

An earlier iteration of this proposal considered a change to the behavior of `group.min.session.timeout.ms` and `group.max.session.timeout.ms`. The idea was to let the coordinator automatically adjust the session timeout provided by the client to be within this range. This would give operators a way to change the allowed session timeout range without causing existing clients to fail. Unfortunately, this does not work gracefully with all clients. In particular, librdkafka-based consumers enforce the session timeout locally. If the session timeout is reached without response from the coordinator, then partitions are automatically revoked and the consumer rejoins the group as a new member. The coordinator, on the other hand, would still enforce the adjusted session timeout, which means that the rebalance would get delayed until the old member could be expired. For this reason, we decided to reject this option for now. We may reconsider it in a separate proposal which allows the coordinator to propagate the adjusted session timeout to the client.