

KIP-876: Time based cluster metadata snapshots

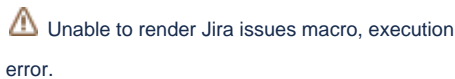
- [Status](#)
- [Motivation](#)
- [Public Interfaces](#)
 - [Configuration](#)
- [Proposed Changes](#)
- [Compatibility, Deprecation, and Migration Plan](#)
- [Test Plan](#)
- [Rejected Alternatives](#)

Status

Current state: Accepted

Discussion thread: <https://lists.apache.org/thread/ww67h9d4xvgw1f7jn4zxwydmt8x1mq72>

JIRA:



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

Motivation

For the cluster metadata topic partition, snapshots are used to compact the log and remove redundant records. Kafka supports generating snapshots based on the amount of bytes appended to the log since the last snapshot. This works well to amortize the cost of generating snapshots.

In addition to compaction, snapshot can also be used as backups to the cluster metadata partition. To better support this case we should allow Kafka to generate snapshot based on time.

Public Interfaces

Configuration

The property `metadata.log.max.snapshot.interval.ms` will be added to the Kafka server configuration. This is the maximum amount of time that Kafka will wait to generate a snapshot if there are records in the log that are not included in the latest snapshot. The default value for this property will be an hour. A value of zero, disables time based snapshot generation.

The default value for the `metadata.max.retention.bytes` will be changed from `-1` (disabled) to `100 * 1024 * 1024` (100MB).

Proposed Changes

Both the KRaft brokers and controllers will generate a snapshot if there is a record in the log that is newer than the latest snapshot and has an append time older than `metadata.log.max.snapshot.interval.ms` milliseconds ago.

This change and implementation needs to consider the two existing reasons for generating a snapshot. First, [KIP-630: Kafka Raft Snapshot](#) added support for generating snapshots when there are `metadata.log.max.record.bytes.between.snapshots` bytes in the log that are not included in the latest snapshot. Second, [KIP-778: KRaft to KRaft Upgrades](#) added support for generating snapshot when the metadata version changes. If a snapshot is generated for any of these reasons, the counter for time-based and byte-based snapshots should be reset to avoid generating extra snapshots.

Compatibility, Deprecation, and Migration Plan

There are no compatibility, deprecation or migration implications. This feature only affects the local replica and doesn't have any effect on remote replicas.

If the user wants to keep the behavior before this KIP was implemented, they would need to disable time based snapshots by setting `metadata.log.max.snapshot.interval.ms` to zero (0).

Test Plan

We will add both unit and integration tests for this feature.

Rejected Alternatives

No alternative design or implementation were considered.