KIP-411: Make default Kafka Connect worker task client IDs distinct

Status

Current state: Accepted

Discussion thread: here

Voting thread: here

JIRA: *KAFKA-5061*

Released: *2.3.0*

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

Motivation

It is currently not possible to monitor producer and consumer metrics for individual tasks in Kafka Connect due to JMX MBean naming conflicts. Standard Kafka producer and consumer clients use `client.id` in metric names to disambiguate JMX MBeans when multiple instances are running in the same JVM. In Kafka Connect, all producer and consumer instances created by a Worker inherit the same client id from Worker properties file. Since it is not possible to set a different client ID for each task, any Connector with more than one task running on a Worker node will generate JMX MBean naming conflicts. This makes it impossible to collect per-task Kafka client metrics across a Kafka Connect cluster.

Public Interfaces

Propose updating the default client.id used by Kafka clients created by Kafka Connect worker tasks. The new default will include the current connector and task id.

Proposed Changes

PR Available here: https://github.com/apache/kafka/pull/6097

Consumers created for sink tasks will have a default `client.id` of the form:

```
connector-consumer-{connectorId}-{taskId}
```

e.g. For connector "conn1", task "2" the default client ID would become: `connector-consumer-conn1-2`

Producers created for source tasks will have a default `client.id` of the form:

```
connector-producer-{connectorId}-{taskId}
```

Dead-letter queue producers created for sink tasks will have a default `client.id` of the form:

```
connector-dlq-producer-{connectorId}-{taskId}
```

Compatibility, Deprecation, and Migration Plan

The change will affect any existing cluster where `client.id` has not been over-ridden in the worker configuration. Since the current default is not useful for JMX monitoring the change should have minimal impact.

Any client IDs specified in the worker configuration via producer.client.id or consumer.client.id properties will remain unchanged, as those will take precedence.

While this change will not affect quota limits, in some cases it could have an indirect impact on resource usage by a Connector. For example, a system that was enforcing quotas using the default "consumer-[id]" client ids will need to update their configuration to enforce quota on "connector-consumer-[task-id]" instead. Note that enforcing quotas on default client ids in this way is not recommended, before or after this change. For systems that were not enforcing any quota limits on client ids, or using default quotas, no change is expected.

Rejected Alternatives

Add a worker configuration option to automatically append task ID to the client ID used by producer or consumer instances instantiated by a Worker. This ensures all JMX MBean names used within a Kafka Connect cluster are distinct.

This approach was rejected because adding new configuration options is considered a larger change to public interfaces than necessary.

Allow overriding client.id on a per-connector basis

This option does not give the required level of granularity, individual tasks would still have name conflicts without other tasks create by the same connector. This could be avoided if the connector could control the client id at the task level, but that would require a much more complex change.