Oracle GoldenGate Kafka Connect Handler/Formatter 1.0

Kafka Connect Handler/Formatter Overview

The Kafka Connect Handler/Formatter takes change data capture operations from a source trail file and generates data structs (org.apache.kafka.connect.data.Struct) as well as the associated schemas (org.apache.kafka.connect.data.Schema). The data structs are serialized via configured converters then enqueued onto Kafka topics. The topic name used corresponds to the fully qualified source table name as obtained from the GoldenGate trail file. Individual operations consist of inserts, updates, and delete operations executed on the source RDBMS. Insert and update operation data include the after change data. Delete operations include the before change data. A primary key update is a special case for an update where one or more of the primary key(s) is/are changed. The primary key update represents a special case in that without the before image data it is not possible to determine what row is actually changing when only in possession of the after change data. The default behavior of a primary key update is to ABEND in the Kafka Connect formatter. However, the formatter can be configured to simply treat these operations as regular updates or to treat them as deletes and then an insert which is probably the closest interpretation of what is actually occurring.

Truncate operations are currently unsupported. Truncates are specialized DDL operations and an edge use case.

The recommended mode to run the Kafka Handler is to run in operation (op) mode. Transaction mode caches operations and then processes them all at transaction commit.

The Kafka Connect jar file and all of the required dependencies must be included in the Java classpath of the Oracle GoldenGate replicat process hosting the Kafka Connect Handler/Formatter. The Java classpath can be controlled by the gg.classpath configuration property in the Java Adapter properties file.

Kafka Connect schema object are generated on a just in time basis. The first time an operation is encountered for given source table a schema is generated. This schema is then cached and reused for subsequent operations.

Data is enqueued to Kafka topics using asynchronous puts. The flush call is made at transaction commit to flush all outstanding operations to Kafka topics to ensure write durability. The GROUPTRANSOPS replicat configuration parameter controls the maximum number of incoming transactions that will be grouped into a single output transaction. The default setting for GROUPTRANSOPS is 1000. Using GROUPTRANSOPS is the primary to increase the performance of the Kafka Connect handler.

The Java classpath of the Java Adapter is controlled by the gg.classpath configuration parameter. The user can employ the wildcard * to indicate all jars in a directory (i.e. /usr/home/kafkaconnect/*). To access the Kafka producer properties file you must configure the gg.properties file to point to the directory containing this file without using the wildcard (i.e. /usr/home/kafkaconnect).

Kafka Connect Handler Contents

The Kafka Connect Handler/Formatter distribution contains the following:
kafka-connect/bin directory contains the binary (does not include the Kafka dependencies)
kafka-connect/dirdat directory contains a sample trail file

kafka-connect/dirprm directory contains sample configuration file

- Confluent.properties – sample kafka properties file
- Conf.prm – sample replicat config file
- Conf.props – sample Java Adapter properties file

kafka-connect/src directory contains the source code

Kafka Connect Handler Configuration

gg.handler.{name}.type=oracle.goldengate.kafkaconnect.KafkaConnectHandler
Required
Explanation: Selects the Kafka Connect handler as the instantiated handler.

gg.handler.{name}.mode
Optional
Legal values: tx | op
Default: op
Explanation: Selects operation (op) mode or transaction (tx) mode for the handler. In most scenarios operation mode is the preferred mode of operation

gg.handler.{name}.kafkaProducerConfigFile
Optional
Legal values: Any legal file name
Default: kafka-producer-default.properties
Explanation: Provides the name of the Kafka Producer configuration file to use to instantiate the Kafka producer object. This file must be accessible on the Java classpath.

gg.handler.{name}.sourceRecordGeneratorClass
Optional
Legal values: Any legal file name
Default: oracle.goldengate.kafkaconnect.DefaultSourceRecordGenerator
Explanation: Provides the implementing class that generates the SourceRecord object (org.apache.kafka.connect.source.SourceRecord) which is used to generate the Kafka ProducerRecord object (org.apache.kafka.clients.producer.ProducerRecord) which is ultimately put to Kafka. This object controls topic selection, partition selection, as well as key and value. Users can override by writing their own implementation.

Kafka Connect Formatter Configuration

gg.handler.{name}.format=oracle.goldengate.kafkaconnect.formatter.KafkaConnectFormatter
Required
Explanation: Selects the Kafka Connect formatter

gg.handler.{name}.format.insertOpKey
Optional
Legal Values: Any string
Default: I

Explanation: Indicator to be inserted into the output record to indicate an insert operation.

gg.handler.{name}.format.updateOpKey
Optional
Legal Values: Any string
Default: U
Explanation: Indicator to be inserted into the output record to indicate an update operation.

gg.handler.{name}.format.deleteOpKey
Optional
Legal Values: Any string
Default: D
Explanation: Indicator to be inserted into the output record to indicate a delete operation.

gg.handler.{name}.format.iso8601Format
Optional:
Legal Values: true | false
Default: true
Explanation: The default format for the current timestamp is ISO8601. Set to false to remove the "T" between the date and time in the current timestamp and output a space (" ") instead.

gg.handler.{name}.format.treatAllColumnsAsStrings
Optional
Legal Values: true | false
Default: false
Explanation: Controls the output typing of generated Kafka Connect messages. If set to false then the formatter will attempt to map GoldenGate types to the corresponding Kafka Connect type. If set to true then all data will be treated as Strings in the generated Kafka Connect messages and schemas.

gg.handler.{name}.format.pkUpdateHandling
Optional:
Legal values: abend | update | delete-insert
Default: abend
Explanation: Provides configuration for how the formatter should handle update operations which change a primary key. Primary key operations can be problematic for the Kafka Connect formatter and require special consideration by the user.
abend - indicates the process will abend
update – indicates the process will treat this as a normal update
delete-insert – indicates the process will treat this as a delete and an insert. Full supplemental logging with NOCOMPRESSUPDATES needs to be enabled for this to work. In order for this to properly work the user should configure the GoldenGate extract process to capture complete before and after image data for updates through RDBMS configuration and NOCOMPRESSUPDATES configured on the extract process.