GoldenGate插件接口

|  |  |
| --- | --- |
| 修订记录 | |
| GoldenGate chenzichuan | Init |
|  |  |

## Story编号和名称

Story.GoldenGate插件接口

## Story价值和功能介绍

Oracle GoldenGate 是一款专门针对各种关系型数据库的数据同步工具，它支持直接解析关系型数据库的事务日志文件，包括：Oracle、DB2、SQL Server、Mysql等，然后通过它自身的程序对事务日志进行数据解析，然后再将数据同步到远端的其他数据库上。

在Oracle GoldenGate功能特性上，主要包含3个模块，分别是：extract、dpe(Distributed Processing Environment)、replica。它能够帮助用户对已有的数据库环境部署一个冗余数据库，并且数据同步效率能够保证在秒级单位。

在Oracle GoldenGate 12.3 版本后，replica 模块支持向bigdata 相关产品同步数据，如：HDFS、HBase、Kafka、Flume等。

在之前的的解决方案中，如果希望将关系型数据库的数据实时地同步到SequoiaDB上，只能够通过以下方法进行数据同步

DB->OGG->Kafka->SparkStreaming(or Store)->SequoiaDB

或者

DB->OGG->Flume->Kafka->SparkStreaming(or Store)->SequoiaDB

以上解决方案由于涉及的产品多，数据处理流程复杂，并且在数据同步频繁情况下，容易造成处理时间过长，同步效率不高的问题。

所以本连接器主要解决的问题就是移除从DB 数据同步到SequoiaDB中多余的步骤，让数据同步流程变成

DB->OGG->SequoiaDB

通过简化流程的方法，达到简化开发、提高数据同步效率的目的。

## Story 对外功能接口描述

说明: 简单描述Story 的对外接口,包括命令行/界面, 客户可见配置文件的参数描述;

GoldenGate 插件的接口参数列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 名称 | 说明 | 取值范围 | 默认值 | 是否必填 |
| hosts | 数据库连接地址 | Null | localhost:11810 | 否 |
| password | 数据库鉴权密码 | Null | “” | 否 |
| username | 数据库鉴权用户 | null | “” | 否 |
| bulkWrite | 是否使用批量写入接口 | 【true|false】 | true | 否 |
| bulkSize | 向数据库批量写入数据时，每个批次的记录数 | 1-∞ | 1024 | 否 |
| checkMaxRowSizeLimit | 是否检查写入数据库的记录大小，如果打开，则超过16MB的BSON记录将会报错 | 【true|false】 | false | 否 |
| ignoreMissingColumns | 写入记录时，是否要忽略值为null的字段 | 【true|false】 | true | 否 |
| changeFieldToLowCase | 连接器在获取源端的表结构时，是否要对其转换为小写字母 | 【true|false】 | true | 否 |
| changeTableToLowCase | 连接器在获取源端的表名时，是否要对其转换成小写字母 | 【true|false】 | true | 否 |
| isPrintInfo | GoldenGate的replica程序在运行时是否要打印OGG-SequoiaDB连接器的日志信息 | 【true|false】 | false | 否 |

OGG配置SequoiaDB连接器的常规配置文件如下

gg.handlerlist=sequoiadb

gg.handler.sequoiadb.type=oracle.goldengate.handler.sequoiadb.SequoiaDBHandler

#The following handler properties are optional.

#Please refer to the Oracle GoldenGate for BigData documentation

#for details about the configuration.

#gg.handler.sequoiadb.hosts=sdb1:11810,sdb2:11810,sdb3:11810

gg.handler.sequoiadb.hosts=localhost:11810

#gg.handler.sequoiadb.password=sdbadmin

#gg.handler.sequoiadb.username=sdbadmin

#gg.handler.sequoiadb.bulkWrite=<true|false>

#gg.handler.sequoiadb.bulkSize=1024

#gg.handler.sequoiadb.checkMaxRowSizeLimit=<true|false>

#gg.handler.sequoiadb.ignoreMissingColumns=<true|false>

#gg.handler.sequoiadb.changeFieldToLowCase=<true|false>

#gg.handler.sequoiadb.changeTableToLowCase=<true|false>

#gg.handler.sequoiadb.isPrintInfo=<true|false>

goldengate.userexit.timestamp=utc

goldengate.userexit.writers=javawriter

javawriter.stats.display=TRUE

javawriter.stats.full=TRUE

gg.log=log4j

gg.log.level=INFO

gg.report.time=30sec

#Path to SequoiaDB Java driver.

# maven co-ordinates

# <dependency>

# <groupId>com.sequoiadb</groupId>

# <artifactId>sequoiadb-driver</artifactId>

# <version>2.8.1</version>

# </dependency>

gg.classpath=./oggapp/sequoiadb\_libs/sequoiadb-driver-\*.jar:./oggapp/sequoiadb\_libs/sequoiadb-ogg-driver.jar

javawriter.bootoptions=-Xmx512m -Xms32m -Djava.class.path=.:ggjava/ggjava.jar:./dirprm

## Story约束和依赖

说明：简单描述Story对其它功能的约束，以及对某些功能的依赖

OGG for bigdata 12.3.1.1.0.013版本、JDK 1.8、SequoiaDB 2.8.2

## Story 简单设计

### 相关模块架构图

说明: 简单的绘制以下与该Story 相关的模块, 特别是新增模块;

SequoiaDB for OGG 的连接器入口类为SequoiaDBHandler。

程序在加载SequoiaDBHandler类时，会首先执行HandlerProperties.init() 方法，以便把相关的配置参数加载进来。

源端数据库做的任何数据操作，都会由OGG解析后通过 SequoaDBHandler.operationAdded() 方法传递给连接器。

而DBOperationFactory.getInstance()方法则会判断此次操作为何种数据操作。

OperationHandler.process方法则是对操作的解析和执行，如果是选择了bulkWrite=true，则在做insert 操作时，需要等待满足bulkSize条记录后再向数据做做一次bulkInsert操作，并且在记录不满足大于等于bulkSize时，最终会在SequoiaDBHandler.trunsactionCommit()方法中集中向数据写入记录。

### 主要操作流程

说明: 使用时序图或者活动图,描述一下主要流程的处理过程

UML图



## 进度评估跟踪

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 汇总信息 | | | | | |
| 总工作量（KLOC） | |  | 开发周期（人天） | |  |
| 开始时间 | |  | 结束时间 | |  |
| 周计划 | | | | | |
| 时间（第X周） | 计划 | | | 完成情况 | |
| 第一周 |  | | |  | |
| 第二周 |  | | |  | |
| 第三周 |  | | |  | |
| 第X周 |  | | |  | |

## 遗留问题

说明：该特性中暂时未彻底解决或有待改进优化的问题

|  |  |  |
| --- | --- | --- |
| **编号** | **标题** | **描述** |
| 1 | 写入优化 | 考虑将批量写入部分调整为多线程同时写入 |

## 验收测试用例

说明: 验收用例是用于表明开发基本功能OK, 满足启动测试最基本需要的一组用例. 验收用例通过,表明开发完成,可启动正式测试;

|  |  |  |  |
| --- | --- | --- | --- |
| **用例编号** | **用例名称** | **用例操作步骤** | **期望结果** |
|  |  |  |  |

## Checklist

|  |  |  |
| --- | --- | --- |
| 检查项 | 结论  （NA/N/Y） | 备注 |
| 代码是否全部提交SVN，且符合代码规范 |  |  |
| 原有用例是否全部通过 |  |  |
| Story文档是否更新 |  |  |
| 验收用例是否全部通过 |  |  |
| 资料是否变更 |  |  |
| 接口和设计是否通过评审 |  |  |
| 相关驱动是否变更或添加 |  |  |
| 周边关联产品或模块是否通知 |  |  |