**基于SQL Server原生日志读取接口实现数据库事务日志解析工具概要设计**

# 1 引言

## 1.1 背景

本软件名为“SQL Reader”，是基于SQL Server原生日志读取接口实现的数据库事务日志解析工具。

开发者为来自电子科技大学的队伍“Sever三巨头”。

开发环境：SQL Server2019

开发语言：Java

## 1.2 术语和概念说明

### 1.2.1 事务日志

SQL Server数据库的日志文件也称为事务日志文件。每个 SQL Server 数据库都有至少有一个日志文件，用于记录所有事务以及每个事务对数据库所做的修改。事务日志中，数据变化被记录在一个连续的日志记录中，且每一个记录都有一个编号，叫做日志序列号(Log Sequence Number, LSN)。以下是我们的产品在解析过程中用到的日志记录：

1. Current LSN：当前LSN。事务日志中的每一条日志记录由LSN(Log Sequence Number)唯一标识。LSN是有序的，如果LSN2大于LSN1,则LSN2的日志所代表的数据修改操作发生在LSN1之后。
2. Operation：当前LSN所做的操作
3. Context：操作的上下文
4. Previous LSN：前一个LSN号
5. AllocUnitName：修改了数据的表名
6. Offset in Row：update操作修改的部分的起始位置
7. Transaction Name：事务名称
8. RowLog Contents 0/1/2：十六进制字符串，在不同的SQL操作下存储不同的数据信息

### 1.2.2 定长/变长数据类型

定长即固定长度，如int、char(n)等，该数据类型的存储是固定长度的，长度不足则会填充；变长即可变长度，如varchar(n)，varbinary(n)等，该数据类型的存储长度随具体输入而变化，不会填充。另外，定长类型和变长类型的数据在RowLog Contents会分开存储。以上特性在数据解析中至关重要。

### 1.2.3 系统元数据库（information\_schema）

可以检索数据库中的对象的元数据，包括但不限于表的主键，列名，数据类型等。我们可以直接在当前实例的数据库调用需要的数据。

### 1.2.4 检查点（checkpoint）

数据修改操作都是在内存中的数据页进行修改，每次修改后并没有立即把这些页面写入磁盘，而是等到一定时期，数据库引擎对数据库发起检查点命令，这时该命令就会创建一个检查点，把当前所有在内存中已修改的页面（脏页）即事务日志信息从内存中写入到磁盘，并且记录下有关事务日志的信息。值得注意的是，检查点的执行会导致该点之前的日志丢失，这也是我们的产品需要解决的问题。

# 2 需求分析

## 2.1 需求概述

通过SQL Server数据库原生的日志读取接口实现一个可持续运行的SQL Server增量事务日志解析工具，将用户业务系统对数据库进行的事务操作解析为统一格式的SQL语句。

## 2.2 任务和目标

1. 可以识别并解析所有的DML操作。
2. 可以将解析后的数据转化为标准的SQL语句。
3. 大对象类型text、ntext、image的解析
4. 解决checkpoint后数据丢失问题
5. 实时性，对数据库的DML操作需要在1S内解析出来。
6. 保证解析出的DML语句的正确性。
7. 保证程序的7\*24小时稳定运行

# 3 系统概要设计

## 3.1 用户图形界面模块

1. 用户需要输入用户名、密码、数据库名以及日志序列号（LSN），因此需要四个输入框以及开始解析的按钮。
2. 用户需要在窗口上看到解析后的SQL语句，因此需要重定向打印输出到GUI界面。
3. 界面简洁大方，新用户容易操作上手。

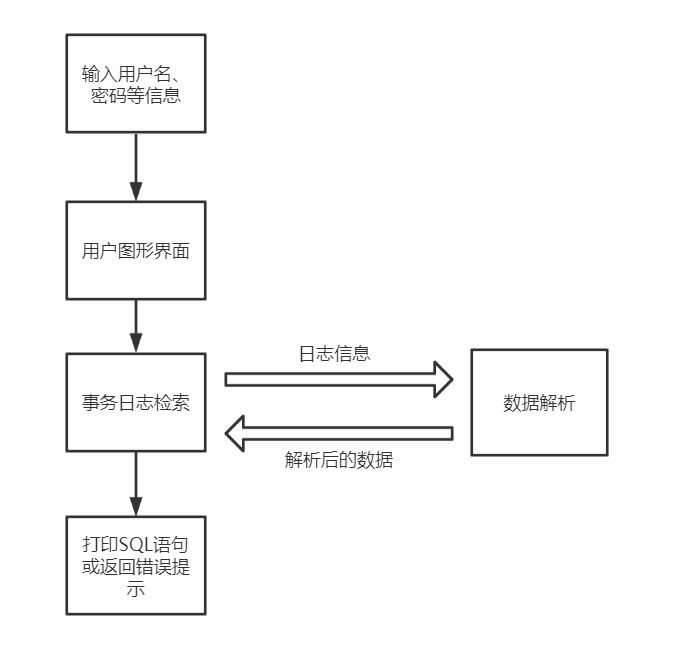
## 3.2 事务日志检索模块

1. 通过用户名、密码、数据库名建立JDBC连接，用以返回查询对象和结果集
2. 获取用户LSN号，根据LSN号对事务日志逐个检索。
3. 检索关键字段CREATE TABLE,INSERT,DELETE,UPDATE等DDL和DML操作字段确定事务类型。
4. 通过查询相关日志获得表名，含数据的十六进制字符串等关键信息。
5. 通过查询系统元数据库（information\_schema）获得数据类型，长度，主键，列数量等关键信息。
6. 利用获得的信息，调用数据解析模块相关方法，获得解析后的数据。
7. 通过关键字段，按标准SQL语句格式打印DDL和DML操作语句。
8. 处理或预防checkpoint后数据库日志丢失的问题

## 3.3 数据解析模块

为事务日志检索模块所调用，获得表的各个列的情况（数据类型，数据长度，是否为空等）以及原始数据（十六进制字符串），根据数据类型调用相应的解析方法解析出标准SQL语句中的数据并返回。需区分定长和变长类型（解析逻辑和处理不同）。

如图一，系统整体流程图一。



图一