

ORACLE®

甲骨文



ORACLE®

Oracle Data Integrator 技术介绍

- 数据集成中的变化数据捕获(CDC)
- ODI 的 CDC 解决方案
- CDC练习
 - 简单 CDC
 - 一致性 CDC
 - Oracle Logminer CDC
 - 双向同步 CDC



数据集成中的增量问题

- 对大数据量需要增量处理
- 如何捕获增量（CDC）
 - 基于时间/序号字段的方法
 - 基于日志的方法
- 数据同步的实时性/一致性
- 如何在复杂的 IT 环境之间实现同步



基于时间戳的 CDC

- 基于交易时间/序号

- 源表存在交易时间/序号
- 只存在数据插入
- 或者源表的变化受到交易时间的限制
- 适用于
 - 如移动通话数据
 - 总账数据

TransDate↕	...↕	↕
05/01/01↕	↕	冻结↕
...↕	↕	
07/02/28↕	↕	
07/03/01↕	↕	变化数据↕
07/03/02↕	↕	
...↕	↕	
今天↕	↕	



- 基于修改时间

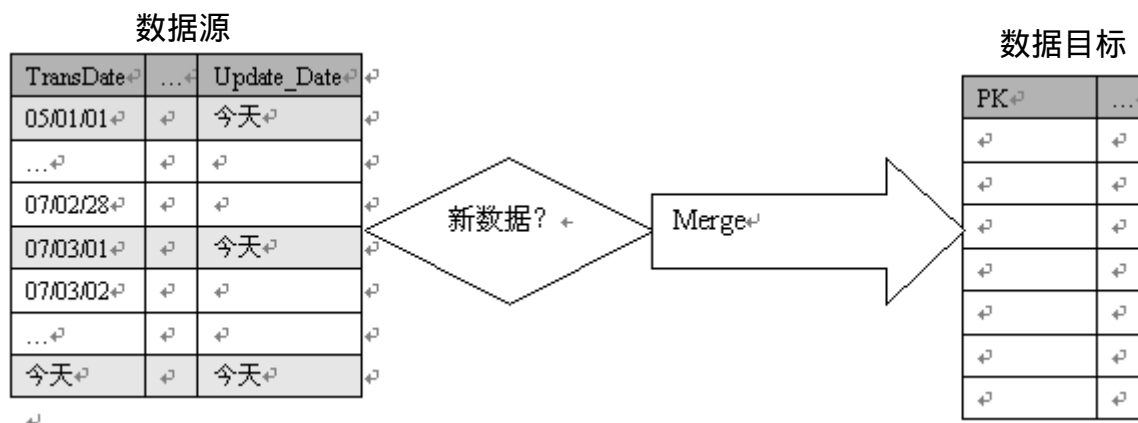
- 源表存在时间戳字段
- 每次修改源表都会更新时间戳字段
- 适用于
 - 如 ERP 的订单数据

TransDate↕	...↕	Update_Date↕
05/01/01↕	↕	今天↕
...↕	↕	↕
07/02/28↕	↕	↕
07/03/01↕	↕	今天↕
07/03/02↕	↕	↕
...↕	↕	↕
今天↕	↕	今天↕

基于时间戳 CDC 的实现

- 实现简单

- 根据时间字段过滤
- 根据系统日期过滤，或者每次完成抽取后记录最大的处理日期
- 需要重复抽取一些数据，然后 Update/Insert 到目标表



基于时间戳 CDC 的缺点

适用的情况少

- 只有少数情况符合第一种方法的前提
- 很多表没有时间戳

容易丢失数据

- 无法处理数据删除的情况
- 一些人为修改不会修改时间戳

满足不了事实同步的要求

- 如果时间戳上不能建索引，性能存在问题
- 长事务的时间戳不准确



序号	事务	时间	SQL
1	事务 A	18:00	INSERT...UPDATE_DATE=18:00
2	ETL-1	18:01	ETL 开始
3	ETL-1	18:02	ETL 完成
4	事务 A	18:03	COMMIT;
6	ETL-2	17:01	ETL 开始
...			

下一次将只抽取
18:01 以后的数据

序号 1 插入
的数据将丢

基于日志的 CDC

- 将变化写到日志表

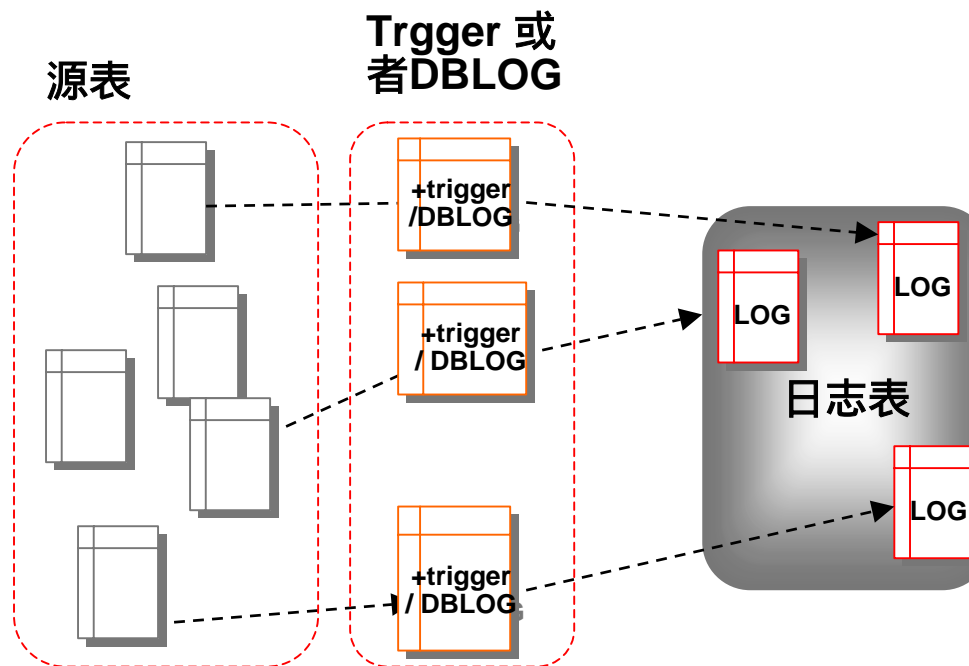
- 同步模式
- 异步模式

- 同步模式

- 一般需要创建 Trigger
- 对业务系统有一些影响

- 异步模式

- 一般通过 Logminner
- 对业务系统影响小
- 可以做到很小的延迟



数据集成对实时性/一致性的要求

- 对实时性要求越来越高

- 数据仓库项目对实时数据的要求
- 主数据管理要求实时同步

- 数据一致性问题

- 多张表可能存在主外键引用
- 多张表可能存在因果关系
- 需要保证一次性被捕获并同步的多张表的数据的一致性



订阅者

序号	事务	SQL
1	ETL 表 ORDERS	ETL 开始
2	事物 A	INSERT INTO ORDER ORDER_ID=1000
3	事物 A	INSERT INTO ORDER_LINE ORDER_ID=1000...
4	ETL ORDER	ETL 完成
5	事务 A	COMMIT;
6	ETL 表 ORDER_LINE	ETL 开始
7	ETL 表 ORDER_LINE	ETL 完成
...		

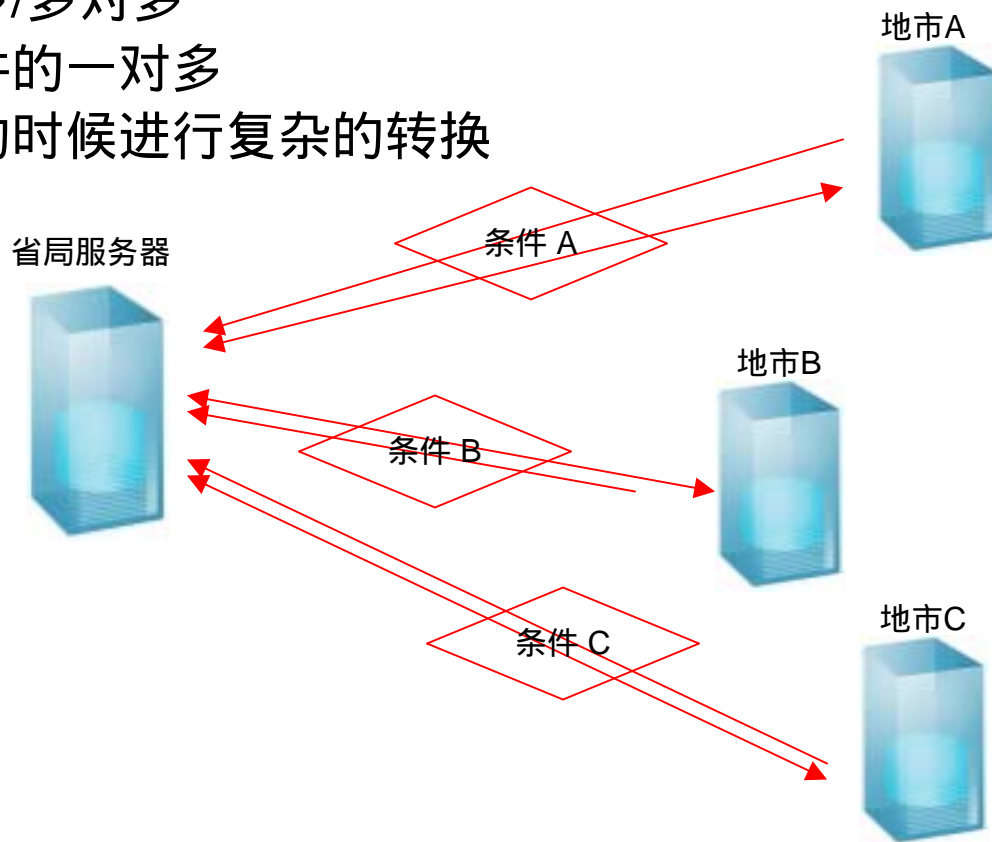
ORDER_ID=1000
不会同步到目标
ORDERS 表

ORDER_ID=1000
会同步到目标
ORDER_LINES
表。违反约束!

复杂的 IT 环境

• 多个 站点/IT 系统之间的集成

- 双向同步
- 一对多/多对多
- 带条件的一对多
- 同步的时候进行复杂的转换



ODI 的 CDC 解决方案

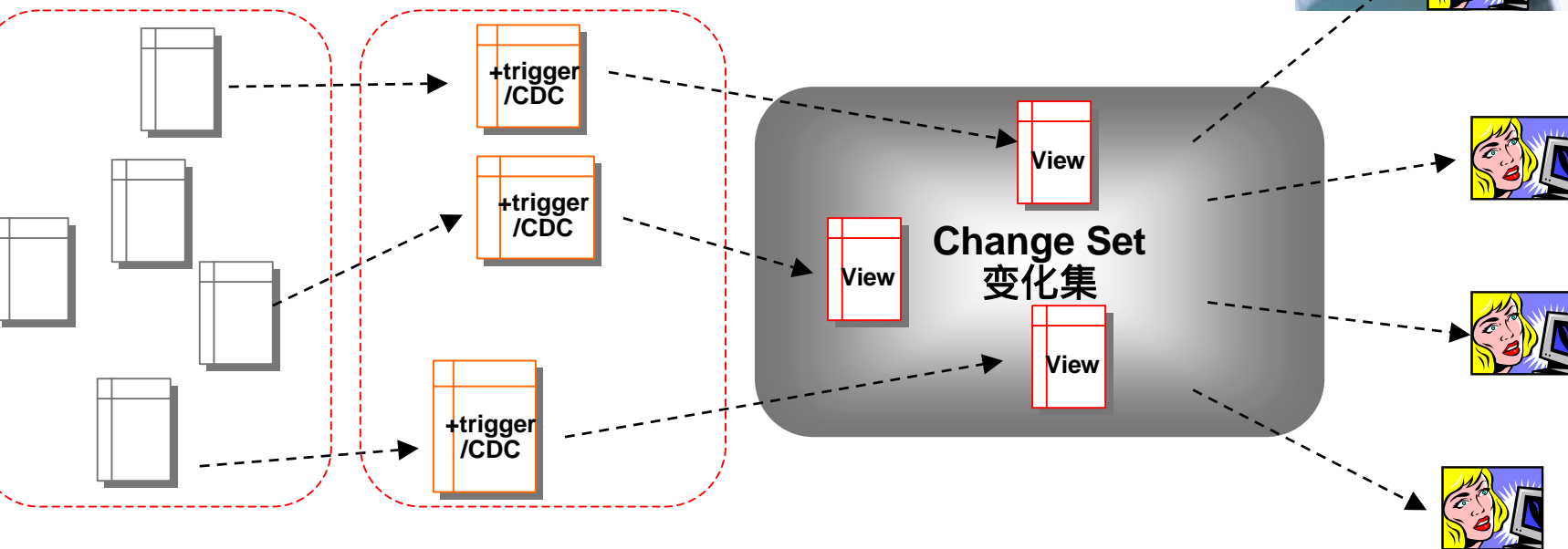
- 完全可以支持基于时间的变化捕获
 - 根据时间字段过滤
 - 保存当前抽取状态
 - 增量数据 Update/Insert
- 基于日记的 CDC 方案
 - 简单的 CDC JKM
 - 一致性的 CDC JKM
 - Oracle Logminer CDC JKM
 - 自定义/客户化 JKM



ODI CDC框架

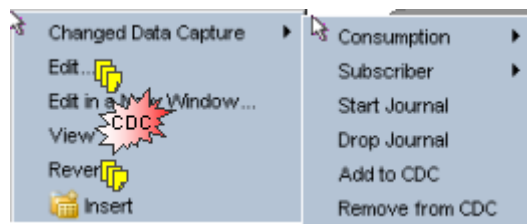
源表

源表上添加Trigger
或者CDC



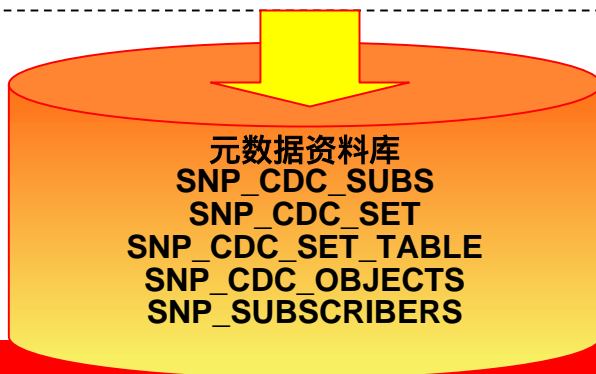
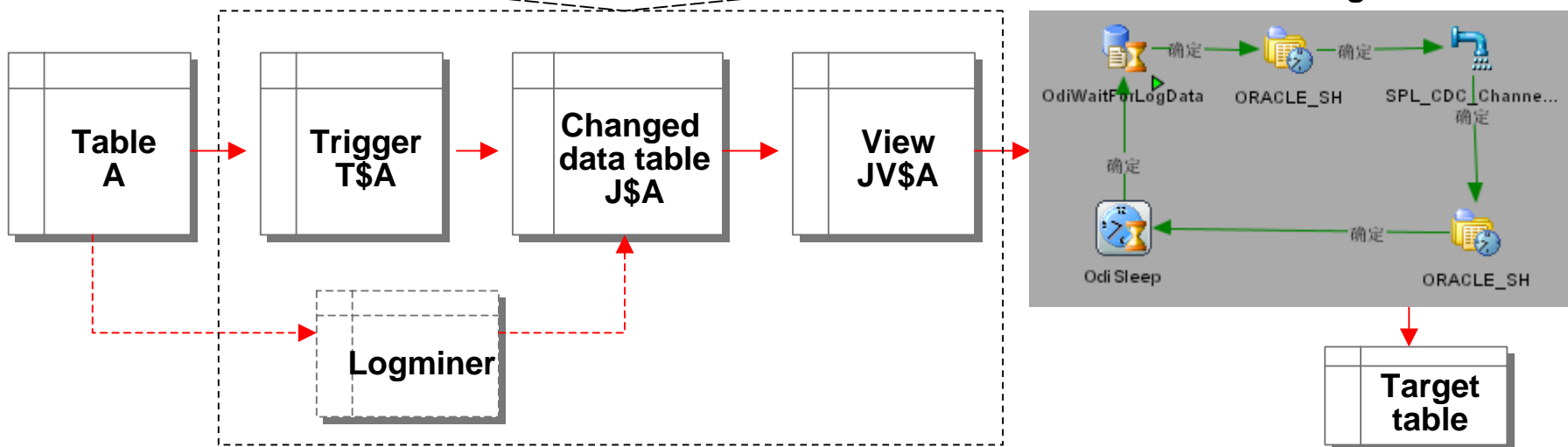
订阅者

ODI CDC 框架细节

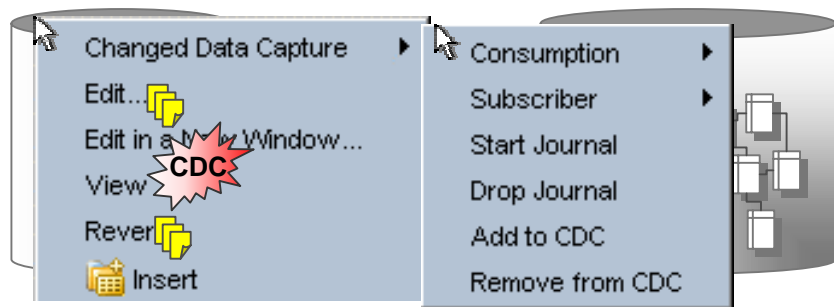


JKM

Package



ODI 变化数据捕获过程



- 发布和订阅的CDC框架
 - 数据库日志
 - 触发器
 - 第三方解决方案
- 确保多表读事务的一致性



1. 设置CDC参数
2. 添加增量存储到CDC
3. 添加订阅者
4. 启动日志

- 四个练习
 - 简单 CDC
01_Simple CDC.doc
 - 一致性 CDC
02_Consistent CDC.doc
 - Oracle Logminer CDC
03_Logminer CDC.doc
 - 客户化的双向 CDC 方案
04_Logminer BI-CDC.doc