

АВТОМАТИЗАЦИЯ ПРОЦЕССА РЕПЛИКАЦИИ ORACLE GOLDENGATE

Гуляев С.Г. Email: Gulyaev629@scientifictext.ru

Гуляев Сергей Германович – бакалавр,
кафедра компьютерных интеллектуальных технологий,
Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург

Аннотация: современным организациям необходимы актуальные данные для проведения различных работ с ними. Для того чтобы не нагружать основные сервера при анализе данных или при сохранении резервных копий, используется непрерывная репликация данных. На данный момент достаточно много компаний предлагают свои решения по репликации данных, в том числе и компания Oracle, предлагающая нам продукт GoldenGate. GoldenGate предоставляет спектр возможностей по репликации данных, в том числе и DDL-репликации. Но по соображениям безопасности компании не всегда прибегают к ее возможностям. В связи с этим встает вопрос о выравнивании структуры базы данных приемника вручную, что занимает достаточно длительное время. Об автоматизации этого процесса и пойдет речь в данной статье.

Ключевые слова: базы данных (БД), система управления базами данных (СУБД), репликация, автоматизация.

AUTOMATION OF THE REPLICATION PROCESS OF ORACLE GOLDENGATE

Gulyaev S.G.

Gulyaev Sergey Germanovich – bachelor,
DEPARTMENT COMPUTER INTELLIGENT TECHNOLOGIES,
ST. PETERSBURG POLYTECHNIC UNIVERSITY OF PETER THE GREAT, ST. PETERSBURG

Abstract: modern organizations need vital data sets for carrying out various works with them. In order to not burden the main servers when analyzing data or when saving backups, continuous data replication is used. At the moment, quite a lot of companies offer their solutions for replicating data, including Oracle, which offers us the GoldenGate product. GoldenGate provides a range of capabilities for replicating data, including DDL replication. But for security reasons, companies do not always resort to its capabilities. In this connection, the question arises of aligning the database structure of the receiver manually, which takes quite a long time. The automation of this process will be discussed in this article.

Keywords: databases, database management system (DBMS), replication, automation.

УДК 004.658.3

Введение

Современным организациям жизненно необходимы актуальные массивы данных для проведения различных работ с ними. Технологии не стоят на месте, и, поскольку твердотельная память сегодня продвинулась до такой степени, что ее стоимость для организаций стала относительно незначительным фактором, теперь организации могут позволить себе работать с системами, которые хранят сотни терабайты данных. Но возникает другой вопрос: каким образом анализировать актуальные данные, при этом не затрагивая работу основной производственной БД компании? Здесь очевидным решением является репликация данных.

Репликация – одна из техник масштабирования баз данных. Техника состоит в постоянном копировании (реплицировании) данных небольшого объема с одного сервера на другие в режиме реального времени, что позволяет распределить нагрузку, тем самым минимизировать нагрузку системы-источника данных [5].

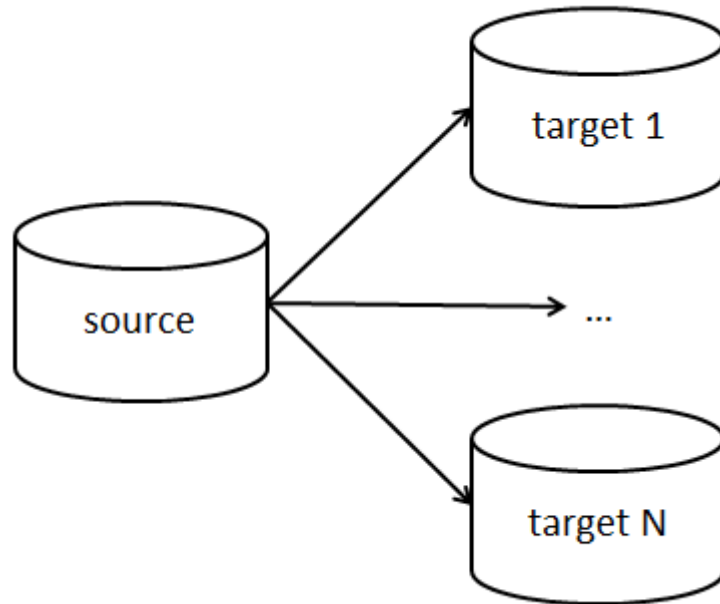


Рис. 1 Схема репликации данных

Описание процесса репликации

Процесс репликации данных систем-источников состоит из следующих функций:

1. Захват изменений данных из журналов СУБД системы-источника по транзакционным журналам изменений указанных объектов;
2. Доставка изменений данных системы-источника на целевую БД. Захваченные изменения по отслеживаемым объектам передаются на сервер-приемник для применения в целевой БД;
3. Применение изменений на приемнике данных. Доставленные изменения объектов системы-источника применяются (сохраняются) в специально созданных объектах целевой БД.

В классическом виде репликация GoldenGate состоит из 3 основных процессов:

- EXTRACT – процесс, извлекающий изменения данных из транзакционного журнала БД источника и складывающий их в trail-файлы;
- DATA PUMP – процесс, передающий trail-файлы по сети из файловой системы источника на файловую систему приемника;
- REPLICAT применяет данные содержащиеся в trail-файлах в БД приемника [1].

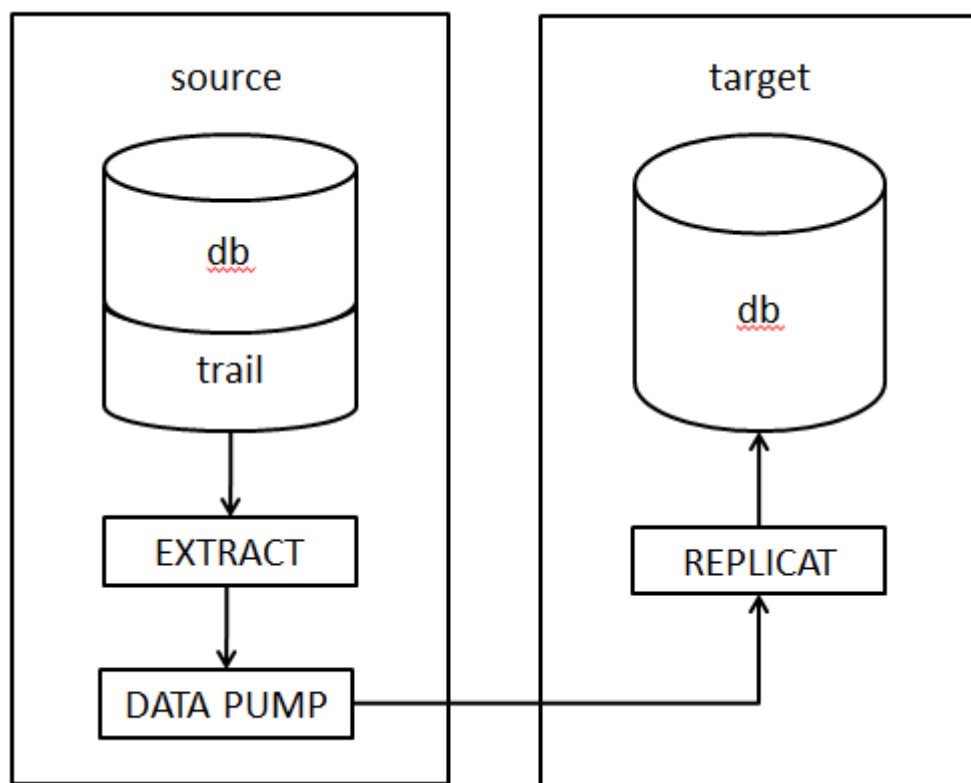


Рис. 2. Процесс репликации Oracle GoldenGate

Цель работы

Поскольку при репликации данных не используется DDL-репликация, а структуры БД-источника достаточно часто изменяется, было принято решение о реализации системы, которая бы выравнявала структуры БД-приемника на основе этих изменений. Сложность также заключается в том, что у разработчика нет доступа к БД-источнику.

Таким образом, были поставлены следующие задачи:

- Автоматизация генерации скриптов процесса репликации;
- Автоматизация сборки процесса репликации в соответствии с контурами установки;
- Разработка скриптов проверки работоспособности и проверки изменения структуры представлений.

Реализация

В самом начале необходимо вычислить список изменений, которые произошли на приемнике. Поскольку у разработчика нет доступа к БД-источнику, было решено использовать выгрузку из системного представления `all_tab_columns` в виде файла Excel.

В связи с тем, что нужно было сравнивать списки данных, для сравнения и генерации скриптов изменения структуры было решено использовать язык программирования Python как наиболее удобный и гибкий для данных операций [3].

Данные скрипты обрабатывают следующие возможные изменения:

- Добавление/удаление таблиц;
- Добавление/удаление столбцов из таблиц;
- Модификация столбцов в таблицах;
- Изменение структуры секционирования таблиц.

Далее выполняется тестирование скриптов средствами Oracle SQL Developer [4]. Тестирующая БД была заполняется небольшим количеством данных так, чтобы размер данных в каждом поле был равен максимальному возможному объему данного поля.

За непрерывную интеграцию отвечает система Jenkins [1]. Она выполняет следующие задачи:

- запуск генерации скриптов выравнивания структуры;
- подготовка тестирующей системы на основе БД-приемника;
- установка на тестирующую систему сгенерированных скриптов выравнивания структуры;
- запуск Unit-тестирования средствами;

- в случае успешного тестирования сборка текущих скриптов в поставку для дальнейшей установки на БД-приемник.

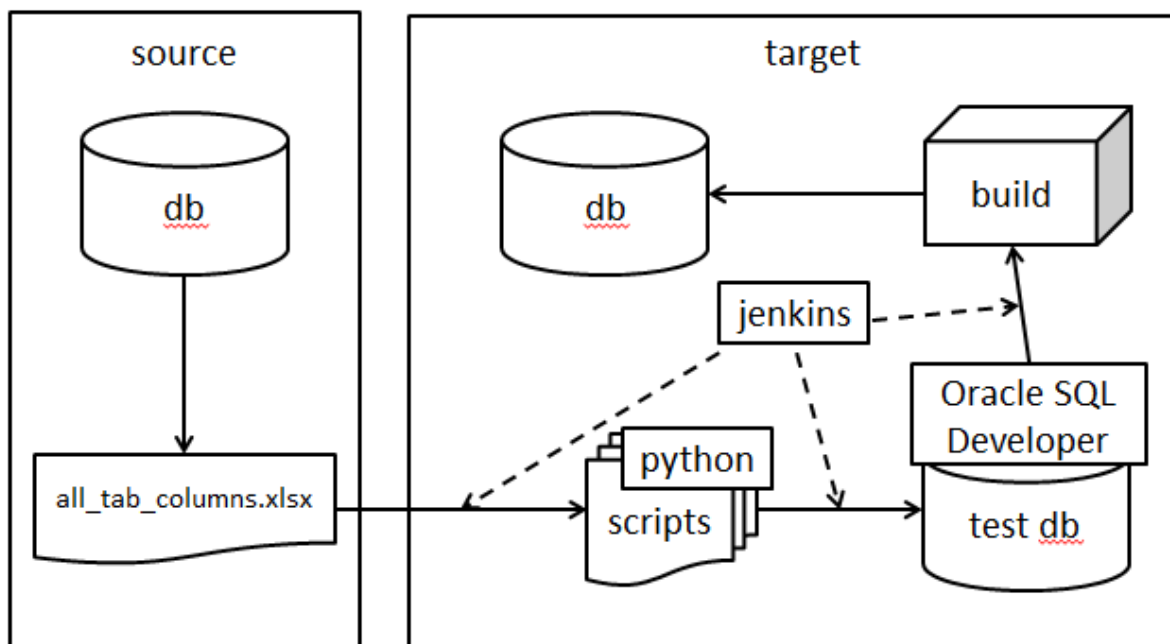


Рис. 3. Структура системы автоматизации

Итоги

Автоматизация процесса репликации значительно ускорила разработку и поддержку процесса репликации данных, что очень важно в условиях большого объема данных и часто меняющейся структуры. Если вначале на разработку уходили часы, то теперь на это затрачиваются секунды.

Список литературы / References

1. Jenkins Quick Guide // TutorialsPoint – Simply Easy Learning. [Электронный ресурс]. Режим доступа: https://www.tutorialspoint.com/jenkins/jenkins_quick_guide.htm/ (дата обращения: 13.01.2017).
2. Prusinski B., Phillips S. Expert Oracle GoldenGate. Apress, 2011. С. 35-39.
3. Using Python With Oracle Database 11g // Oracle | Integrated Cloud Applications and Platform Services [Электронный ресурс]. Режим доступа: <http://www.oracle.com/technetwork/articles/dsl/python-091105.html> (дата обращения: 20.10.2016).
4. Unit Testing with SQL Developer // Oracle | Integrated Cloud Applications and Platform Services [Электронный ресурс]. Режим доступа: https://docs.oracle.com/cd/E15846_01/doc.21/e15222/unit_testing.htm (дата обращения: 21.11.2016).
5. Репликация (вычислительная техника) // Википедия — Свободная энциклопедия [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BF%D0%BB%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F_\(%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D1%82%D0%B5%D1%85%D0%BD%D0%B8%D0%BA%D0%B0\)](https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BF%D0%BB%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F_(%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D1%82%D0%B5%D1%85%D0%BD%D0%B8%D0%BA%D0%B0)) (дата обращения: 13.02.2017).