



Тинькофф
Банк

Операционный BI И большие данные



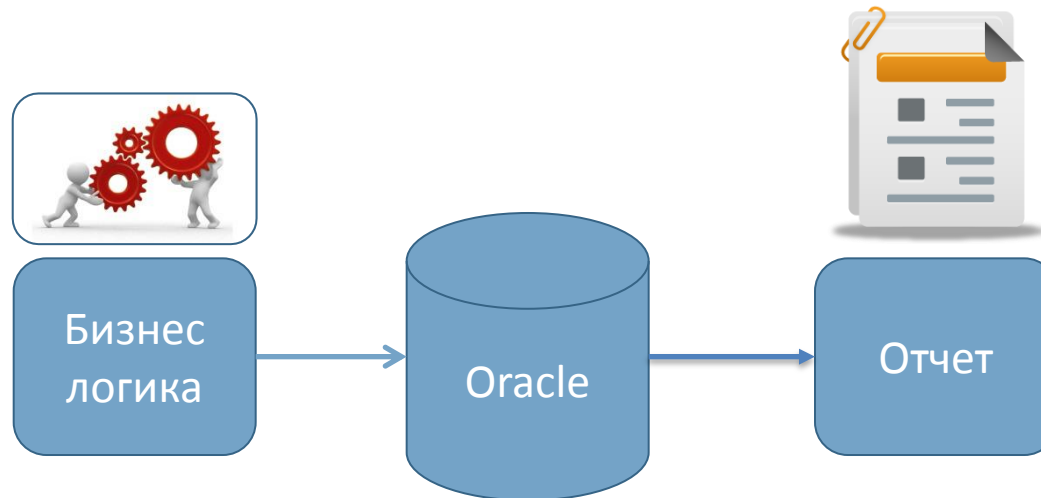
tinkoff.ru



- Аналитический
 - Комплексные бизнес-метрики
 - Сложная интеграция
 - Высокая актуальность данных не критична
- Операционный
 - Отдельные бизнес процессы
 - Простая интеграция – одна система
 - Необходима высокая актуальность данных



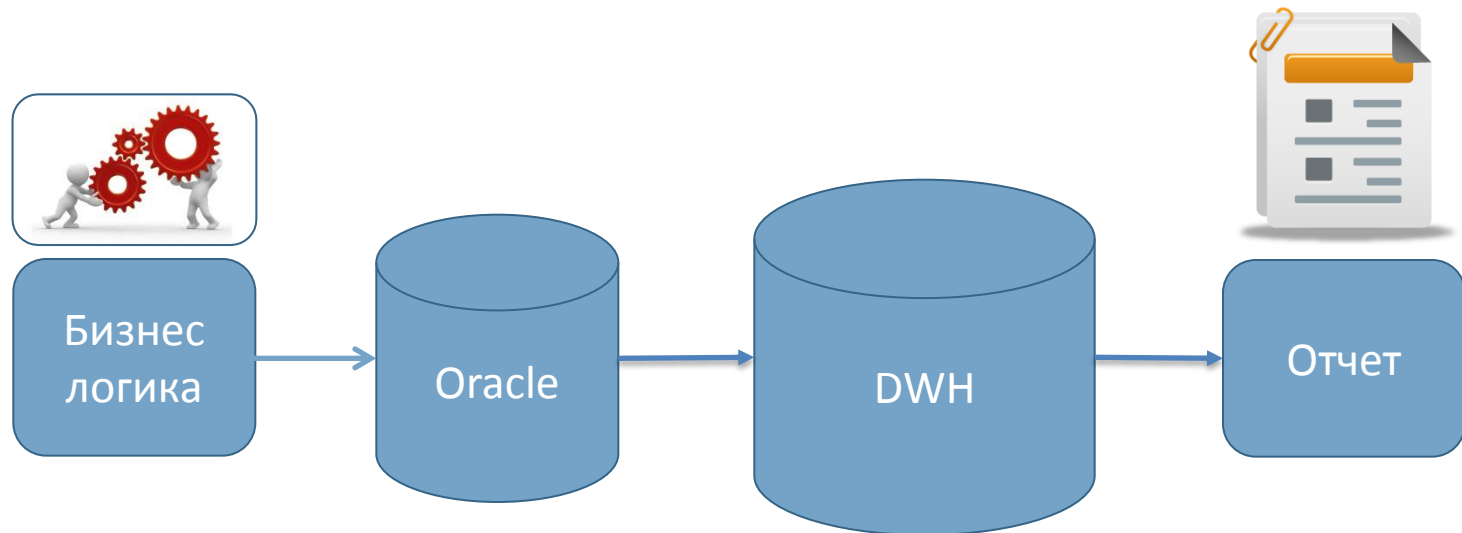
- Строим отчет напрямую на базе бизнес-системы



- Высокая нагрузка на бизнес-систему



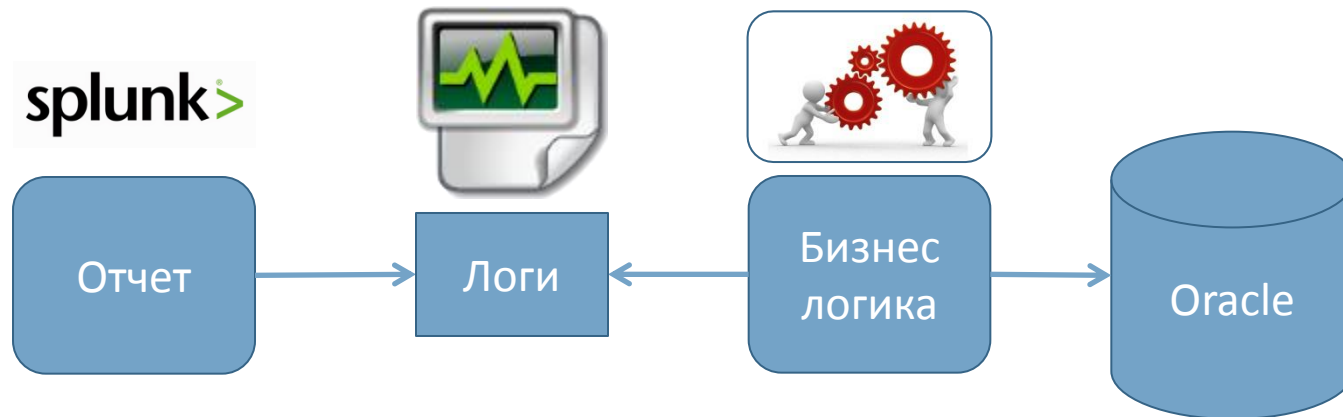
- Строим отчет на данных DWH



- Высокая латентность данных – сутки
- Недостаточная гибкость



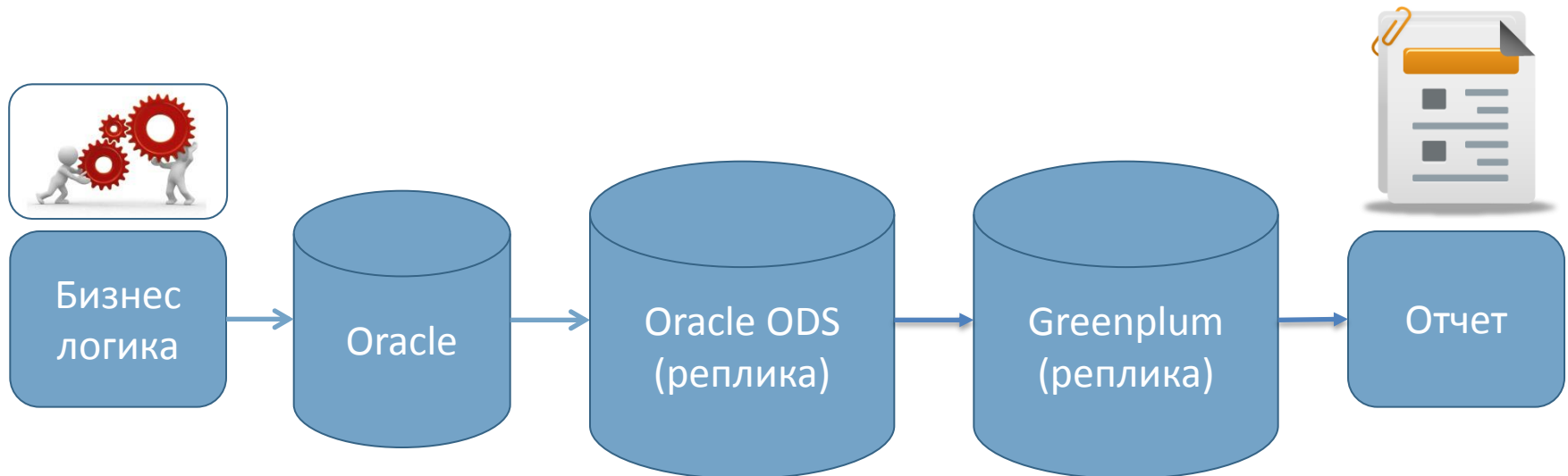
- Строим отчет на Operational Intelligence - Splunk



- Не все данные доступны
- Не вся логика реализуема



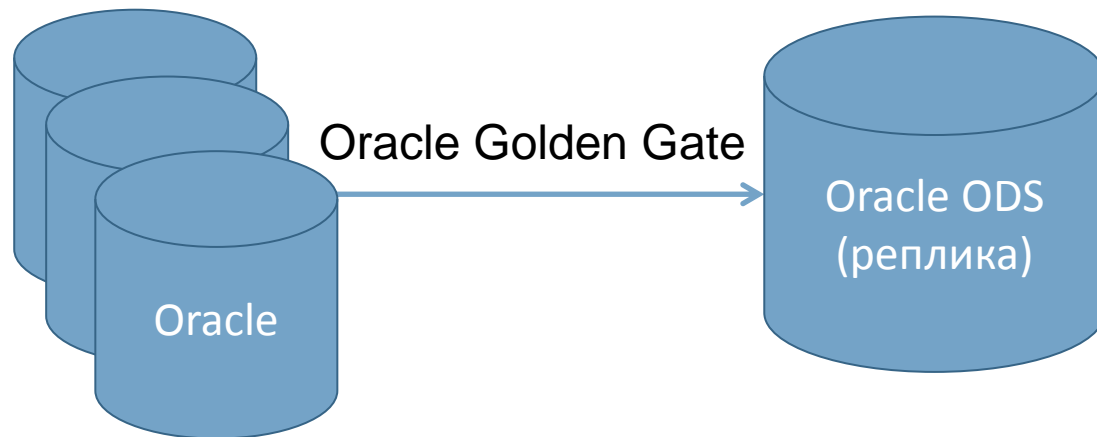
- Используем реплики



- Высокая гибкость
- Приемлемая латентность (1-3 часа)
- База-источник не нагружается



Режим «Real time data warehousing».



- Минутная латентность данных в ODS
- Журнал изменения таблиц источников в ODS



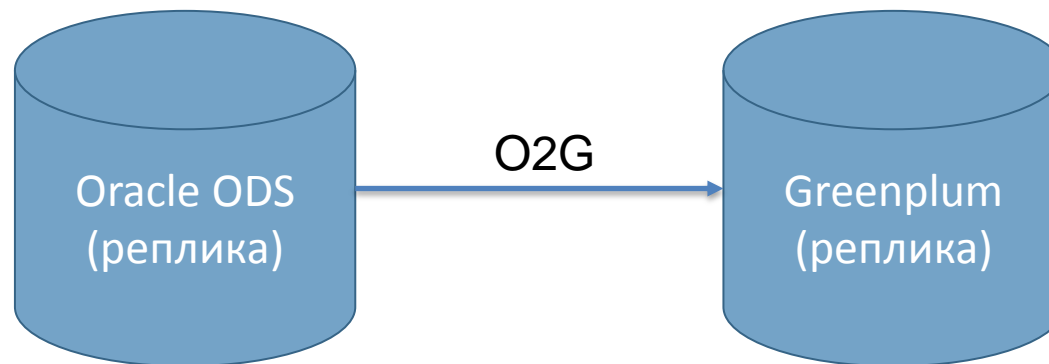
Основная задача – переместить из Oracle в Greenplum десятки миллионов изменений в час не создавая высокую нагрузку ни на Oracle, ни на Greenplum



Репликация O2G

Собственная разработка

Эффективный движок репликации из Oracle в Greenplum



- Допустимая латентность (~1-3 часа) в Greenplum
- Платформа для аналитических запросов



- Как делаем:
 - «Батчевая» выгрузка + выгружаются только измененные данные
 - Внутренний bulk-механизм загрузки Greenplum

O2G –факты



- Латентность данных не более 1-3 часов
- Невысокая нагрузка на Greenplum (~10% ресурсов max)
- Эффективное хранение данных в Greenplum – поколоночное сжатие
- Возможность выгрузки в несколько GreenPlum'ов
- Другие потребители данных ODS (DWH, шина)
- Храним все изменения по всем таблицам за все время

Большие данные – Velocity, Volume



- 14 TB – объем таблиц всех источников в ODS
- 3 TB - размер реплик ODS в GreenPlum: 3 Тб
- 500 миллионов операций в сутки
- 75 миллиардов операций за полгода

Контактная информация



tinkoff.ru