



**СИГМА**

Миграция данных из Oracle  
в PostgreSQL с  
использованием  
инструмента Pentaho

Чувашов Иван



# Инструменты по миграции данных

Расширения под PostgreSQL, позволяющие коннектиться к СУБД Oracle

- oracle\_fdw ([https://github.com/laurenz/oracle\\_fdw](https://github.com/laurenz/oracle_fdw))
- ora\_migrator ([https://github.com/cybertec-postgresql/ora\\_migrator](https://github.com/cybertec-postgresql/ora_migrator))

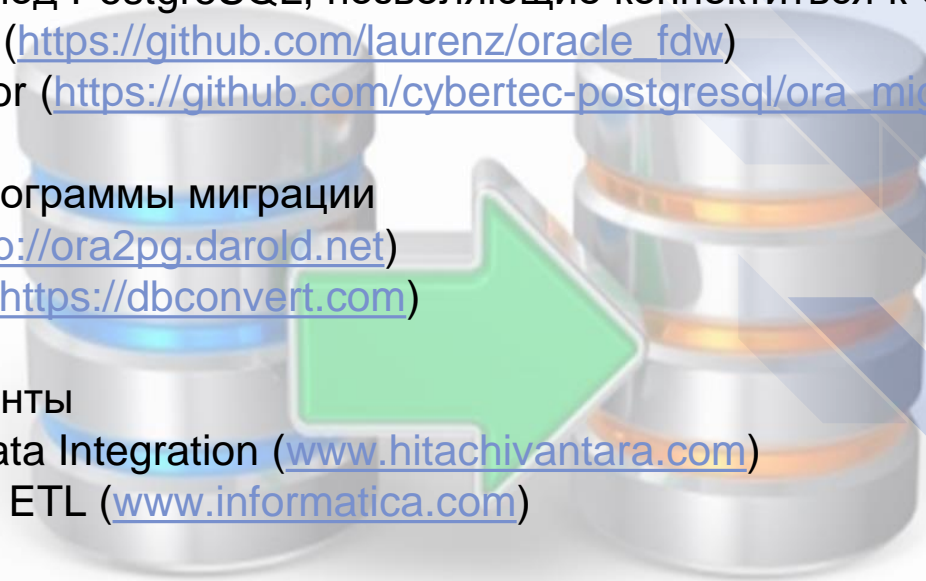
Сторонние программы миграции

- ora2pg (<http://ora2pg.darold.net>)
- dbconvert (<https://dbconvert.com>)

ETL инструменты

- Pentaho Data Integration ([www.hitachivantara.com](http://www.hitachivantara.com))
- Informatica ETL ([www.informatica.com](http://www.informatica.com))

Самописные программные модули





# Pentaho Data Integration

Spoon - dic\_okpd2

File Edit View Action Tools Help

www.BANDICAM.COM

Connect

View Design

Search

Jobs

- dic\_okpd2
  - Run configurations
  - Database connections
  - Job entries
  - Slave server
  - Hadoop clusters
  - VFS Connections

dic\_okpd2

100%

```
graph LR; Start[Start] --> Step1[Вставляем данные в batch]; Step1 --> Step2[Получение последней даты вставки данных]; Step2 --> Step3[Переносим справочник "ОКПД2"]; Step3 --> Step4[Обновляем данные в batch]; Step4 --> Step5[Success]; Step3 -.-> Step6[Записываем ошибку]; Step6 --> Step7[Abort job];
```

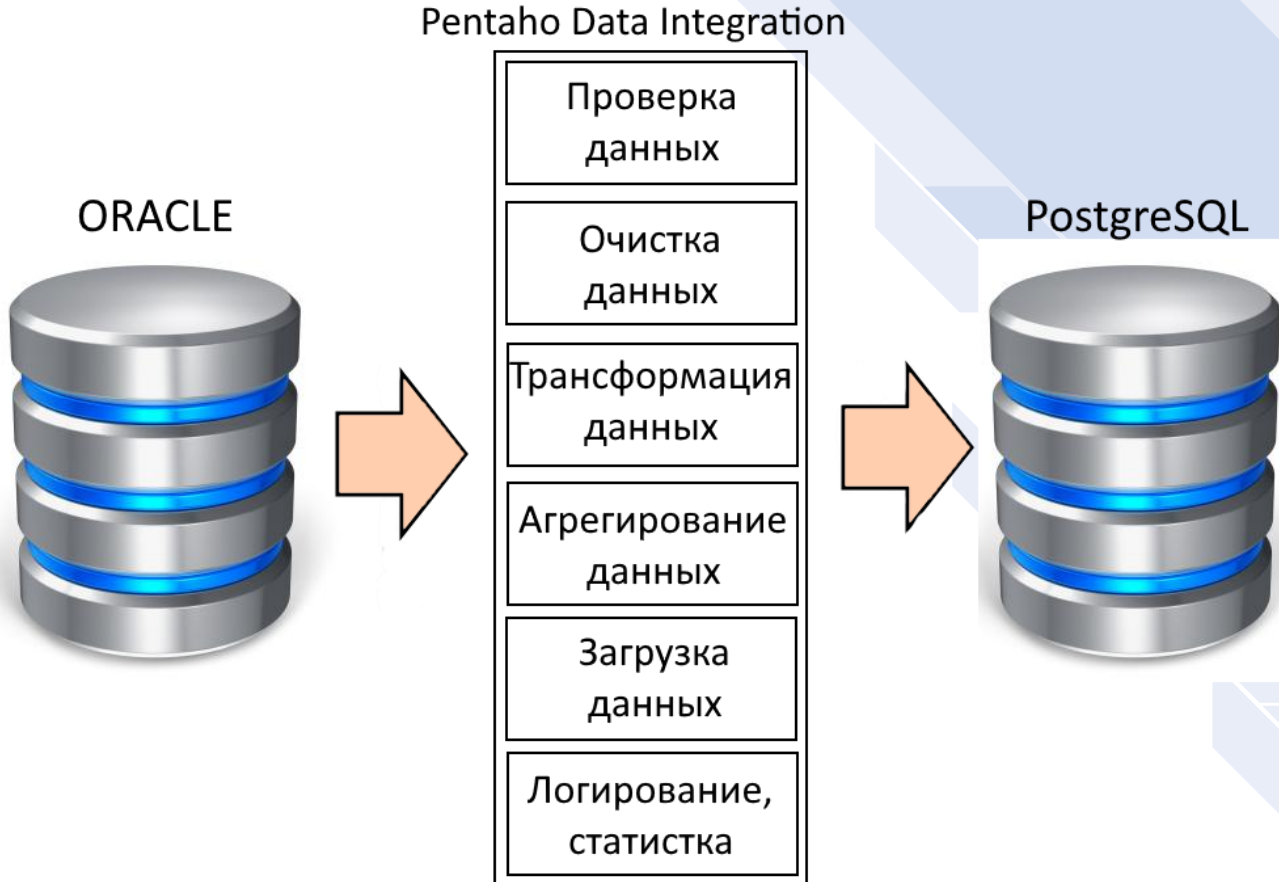
Execution Results

Logging History Job metrics Metrics

2019/11/10 14:12:55 - Текущее время.0 - Finished processing (I=0, O=0, R=1, W=1, U=0, E=0)  
2019/11/10 14:12:55 - Получаем переменные.0 - Finished processing (I=0, O=0, R=1, W=1, U=0, E=0)  
2019/11/10 14:12:55 - НЕ изменять дату обновления данных в источнике.7.0 - Finished processing (I=0, O=0, R=1, W=1, U=0, E=0)  
2019/11/10 14:12:55 - Выставляем статус "Завершен" и новую дату в таблице BATCH.0 - Finished processing (I=1, O=0, R=1, W=1, U=1, E=0)  
2019/11/10 14:12:55 - dic\_okpd2 - Starting entry [Success]  
2019/11/10 14:12:55 - dic\_okpd2 - Finished job entry [Success] (result=[true])  
2019/11/10 14:12:55 - dic\_okpd2 - Finished job entry [Обновляем данные в batch] (result=[true])  
2019/11/10 14:12:55 - dic\_okpd2 - Finished job entry [Переносим справочник "ОКПД2"] (result=[true])  
2019/11/10 14:12:55 - dic\_okpd2 - Finished job entry [Получение последней даты вставки данных] (result=[true])  
2019/11/10 14:12:55 - dic\_okpd2 - Finished job entry [Вставляем данные в batch] (result=[true])  
2019/11/10 14:12:55 - dic\_okpd2 - Job execution finished



# ETL кейс





# Миграция

Что нужно знать прежде чем приступить к миграции данных

- что переносим?
  - документация
  - разработчики
- когда и как переносим?
  - полный
  - инкрементальный
- как проверяем данные?
  - количественные
  - качественные
- как отслеживать ошибки переноса данных?
  - логирование
  - статистика
  - алармы



# Что переносим?

Аналитики

Разработчики

Пользователи старой системы

Архитектор новой ИТ системы



## Когда и как переносим?

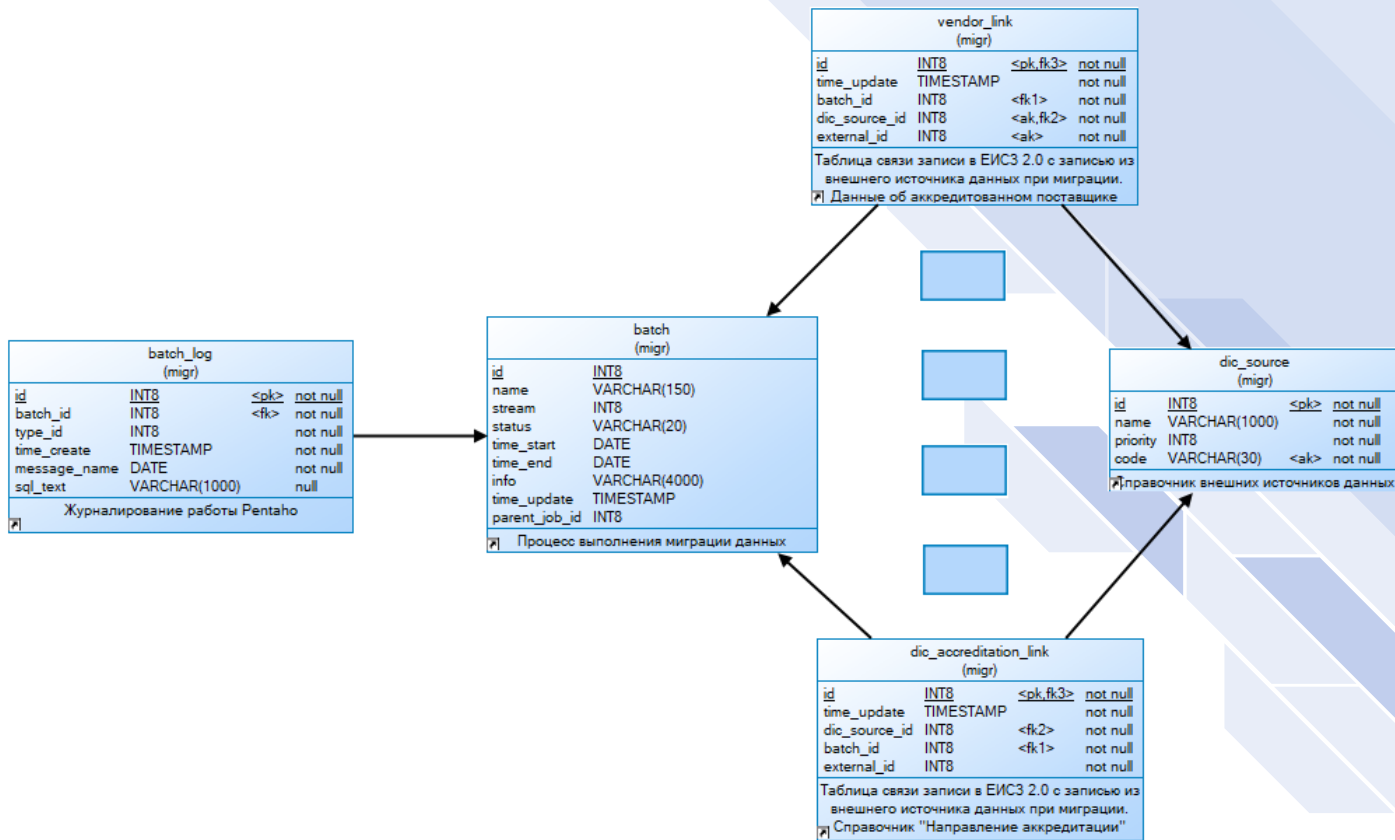
Когда переносить данные - жесткие временные рамки переноса.  
На полный перенос данных – выходные/праздники.  
На инкрементальный – 8 часов ночью.

Как переносить данные

1. опираться на первичный ключ, запоминать его у себя
2. дату модификации записи в таблице, запоминая для каждой таблицы последнюю дату переноса у себя



# Когда и как переносим?







## Как проверяем данные?

Pentaho ETL дает только количественные показатели, но качественные.

Как проверять качество данных?

1. Формирование отчетов в двух системах
2. Ручное тестирование



# Как отслеживать ошибки переноса данных?

Алармы  
письма  
сообщения

Логирование



## Ошибка в миграции dic\_post.

База данных	EISZ_STAND
Тип миграции	incremental из ветки master
Статус	stop
Общее затраченное время	00:00:00
Идентификатор канала	836dfb53-6ac9-4ee8-879d-00230c15a4be
Идентификатор миграции	68

Статистика

Затраченное время:

Миграция	Время
req.marketing_price -- "Справочник цен"	03:37:15
req.attachment -- Метаданные документов из UCM	01:58:25
req.requirement -- "Потребность"/"Позиции потребностей"	01:04:35
req.item -- "Номенклатурных позиций"	00:51:38
req.offer -- "Коммерческое предложение"	00:34:31
req.attachment -- Метаданные документов из LOB	00:29:04



## Типичные ошибки переноса

1. Строгая типизация данных.

Нужна дополнительная конвертация данных.

2. «Подвисание» программы Pentaho ETL.

Параметр в файле Spoon.bat PENTAHO\_DI\_JAVA\_OPTIONS

3. Перенос медленно работает. Параметр на любой компоненте «Change Number of Copies to Start»

4. Использование общих трансформации



## Есть вопросы?

Найдете меня так:

Чувашов Иван

[chuvashovin@gmail.com](mailto:chuvashovin@gmail.com)

8-963-311-07-73