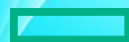




VERTICA



Hewlett Packard  
Enterprise



# Загрузка данных в ХД Vertica

Технологии захвата и загрузки данных из первичных источников

**Алексей Константинов**

# Виды первоисточников



Выполнение SQL  
запросов  
получения  
записей таблиц

Сохранение  
результатов в CSV  
файлы требуемого  
формата

Выполнение  
пакетной загрузки  
CSV файлов в  
таблицы Vertica

# NoSQL

Выполнение  
запросов  
получения данных

Сохранение  
результатов в JSON  
файлы заданной  
структуры

Выполнение  
пакетной загрузки  
файлов в Flex  
таблицы Vertica

# Web Services

Получение  
данных web  
services

Сохранение  
результатов в  
топики Apache  
Kafka

Выполнение  
загрузки данных  
из топиков Kafka  
в таблицы Vertica

# Структурированные файлы

Чтение файлов

Сохранение  
результатов в CSV  
требуемого  
формата

Выполнение  
пакетной загрузки  
CSV файлов в  
таблицы Vertica

# Неструктурированные файлы

Парсинг файлов

Сохранение  
результатов в CSV  
требуемого  
формата

Выполнение  
пакетной загрузки  
CSV файлов в  
таблицы Vertica

# Преимущества использования промежуточных файлов

- Независимые процессы выгрузки данных из источников и загрузки в приемники, что не приводит к длинным транзакциям загрузки данных в приемник при медленной отдаче данных источником;
- Архивное хранение файлов после их загрузки в приемник позволяет произвести их повторную загрузку в случае аварийного отката ХД без повторной выгрузки из источника;
- Выгрузка данных в файлы порциями установленного размера позволяет оптимизировать их загрузку в ХД и в случае сбоя загрузки продолжить её с точки падения;
- Пакетная загрузка файлов Vertica позволяет на лету в ходе загрузки производить обработку и трансформацию данных (ETL), что эффективнее, чем загрузка в промежуточные таблицы и затем дальнейшая трансформация данных (ELT).



# EasyLoader: работа с файлами

## Захват данных из таблиц РСУБД в CSV файлы

Получение данных из таблиц по запросам с условиями, по списку заданных партиций, по нарастающему значению поля.

## Загрузка файлов в Vertica

Формирование команд пакетной загрузки по списку найденных файлов для таблиц ХД по заданным условиям, с включением выражений трансформации значений полей.

## Работа с файлами

Хранение успешно загруженных файлов, а также выявленных ошибок загрузки в виде файлов за данный интервал времени, упаковка в ZIP или удаление после истечения срока хранения.

# EasyLoader: парсинг файлов

Захват и доставка  
файлов на сервер



Вызов внешнего  
парсера на каждый  
файл из списка



Сохранение  
результатов парсинга  
в CSV файлы

# EasyLoader: загрузка в Vertica



# Спасибо за внимание!

Алексей Константинов



+7 499 390 88 92



easydata.ru



info@easydata.ru