Оленчикова Т.Ю.

**Методические указания и задачи к выполнению лабораторной работы по дисциплине «**Программирование на языке Java для анализа данных**»**

**Лабораторная работа 8 Обработка реляционных данных с помощью Hive, Pig, MapReduce**

Цель: Ознакомиться с фреймворком Apache Hadoop для распределенного хранения и обработки данных, освоить модель программирования MapReduce для обработки больших данных на кластерных вычислительных системах

**Теоретические материалы**

# 1) Методы и модели исследования сложных систем и обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ;. — Санкт-Петербург : Лань, 2020. глава 6, С,203-211. – Текст : электронный // Лань : электронно-библиотечная система. – URL: https://e.lanbook.com/book/126938 (дата обращения: 23.08.2023). – Режим доступа: для авториз. пользователей.

# 2) Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. раздел 2, С.45-68. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452 (дата обращения: 23.08.2023). — Режим доступа: для авториз. пользователей.

# 3) Hive vs Pig. На что мне столько ETL? [Электронный ресурс]. – URL: <https://habr.com/ru/articles/223217/> Доступ свободный

Задание.

Обработка реляционных данных с применением Hive, Pig, MapReduce

1. Разработать базу данных (не менее 3 сущностей). Разработать не менн 5 заппросов к БД, шз них не менее 3 с применением JOIN
2. Реализовать БД и запросы на Hive и Pig.
3. Раелизовать запросы с применением паттерна MapReduce.
4. Продемонстрировать, что результаты выполнения запросов на Hive, Pig совпадают с результатами, полученными с помощью собственной реализации MapReduce
5. Написать ответ на контрольный вопрос согласно вашего варианта.

Пример работы с Hive см. в приложении 1, с Pig – в приложении 2.

**Индивидуальные задания для лабораторной работы.**

*Лабораторная работа содержит только 1 вариант*

**Контрольные вопросы**

1. Что такое Hadoop Map Reduce?
2. Как работает Hadoop MapReduce?
3. Объясните, что такое спекулятивное исполнение?
4. Объясните, в чем разница между входным разделением и блоком HDFS?
5. Укажите, в чем разница между СУБД и Hadoop?
6. Укажите, какие компоненты данных использует Hadoop?
7. Укажите, какие наиболее распространенные форматы ввода определены в Hadoop?
8. Перечислите три файла конфигурации Hadoop.
9. Что такое большие данные?
10. Объясните пять признаков больших данных
11. Сколько форматов ввода существует в Hadoop?
12. Что такое Thread? Кто использует Hadoop?
13. Каковы основные функции Hadoop?
14. В каких трех режимах может работать Hadoop?
15. В каких операционных системах работает Hadoop?
16. Каковы компоненты HDFS?

**Критерии оценивания.**

В задании 3 задачи: а) программа на Hive; б) Pig; в) MapReduce

За решение каждой задачи вы можете получить до 2 баллов: задание выполнено полностью и правильно – 2 балла; имеются незначительные ошибки – 1 балл; приложение не работоспособно – 0 баллов;

Ответ на контрольный вопрос должен продемонстрировать понимание архитектуры Hadoop, за ответ вы можете получить оценку: ответ полный и правильный – 2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов.

Итого, максимальная оценка – 8 баллов

**Внимание!** Полученная оценка автоматически снижается на 2% за каждую полную неделю задержки сдачи отчета по работе, но не более, чем на 40%

**Библиографический список**

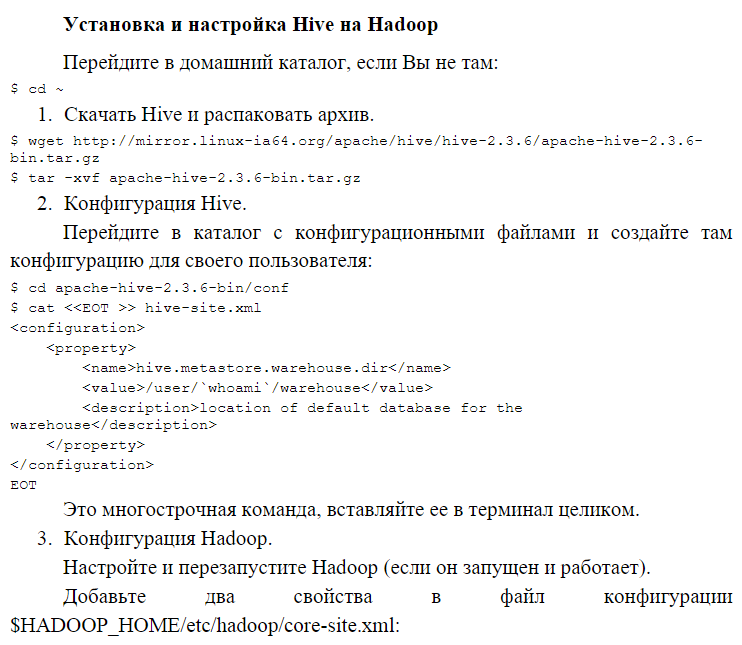
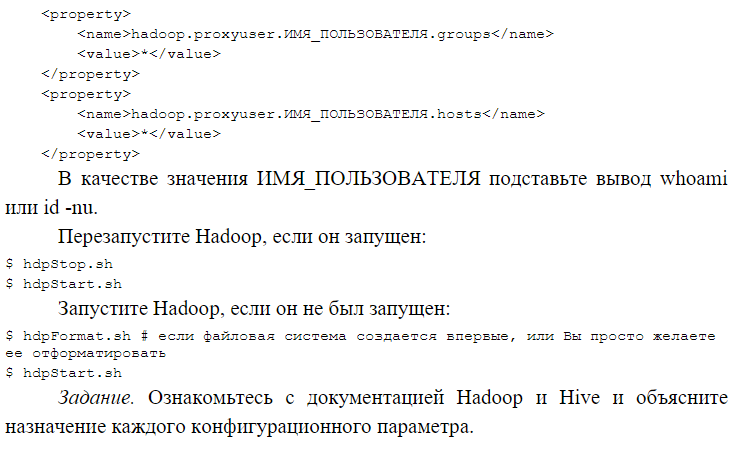
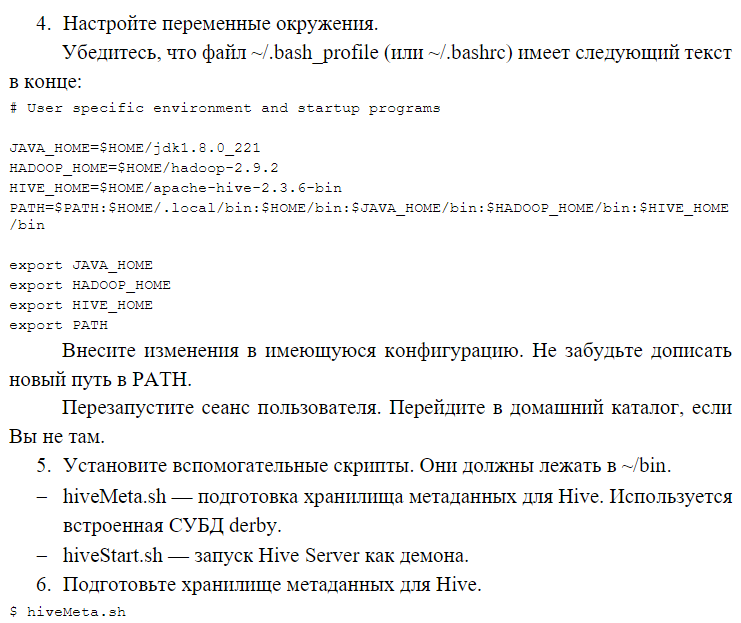
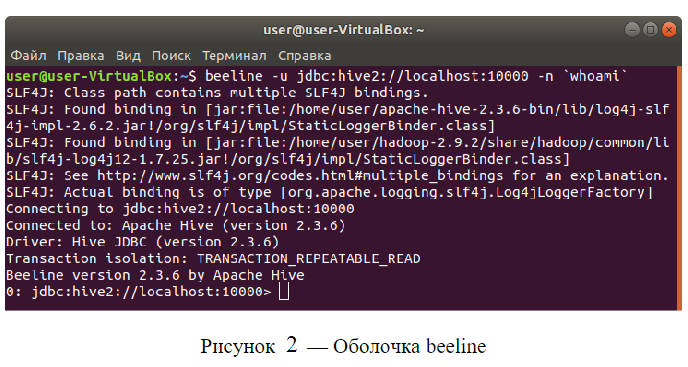
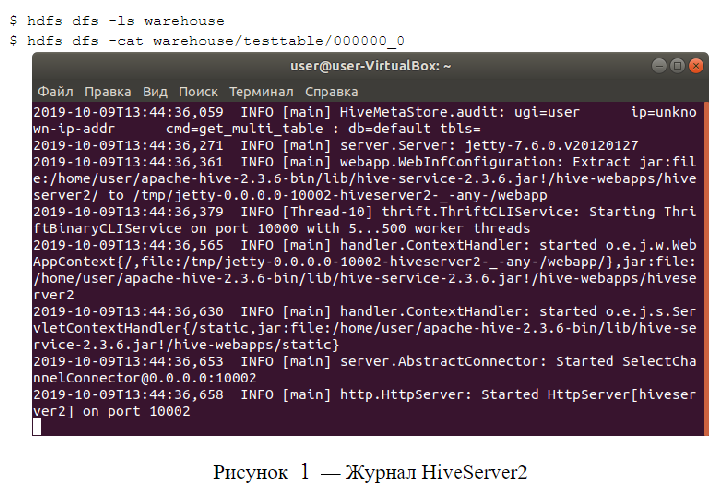
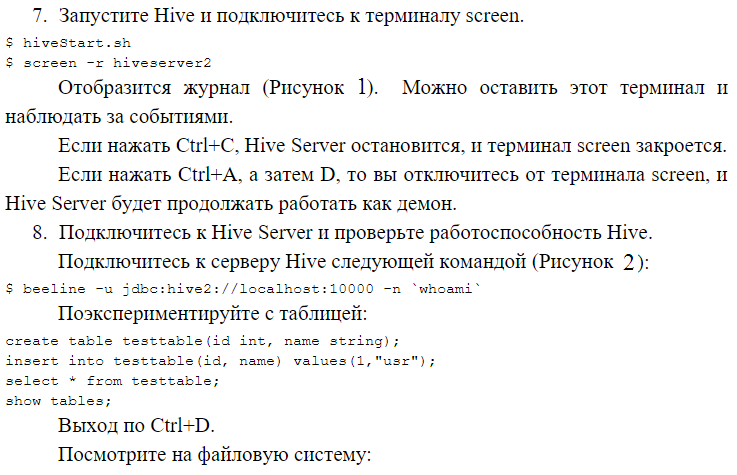
1. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. – Москва : РТУ МИРЭА, 2021. – 91 с. // Лань : электронно-библиотечная система. – URL: https://e.lanbook.com/book/182452 (дата обращения: 24.08.2023). – Режим доступа: для авториз. пользователей.

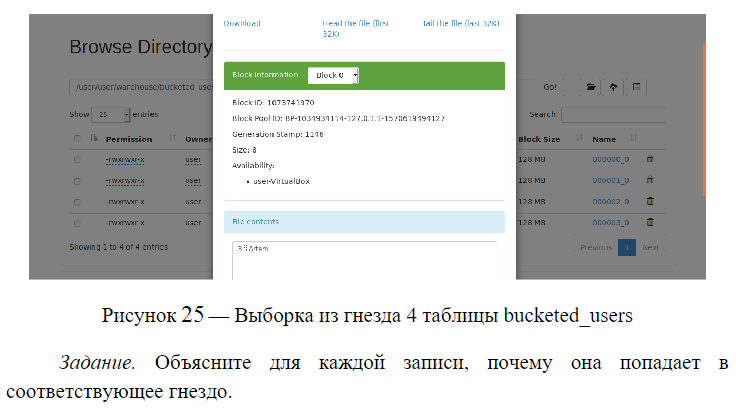
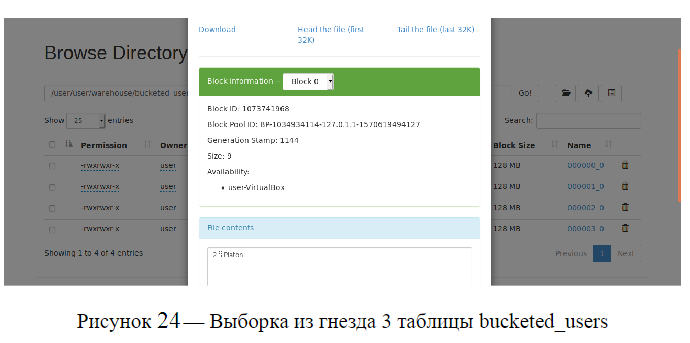
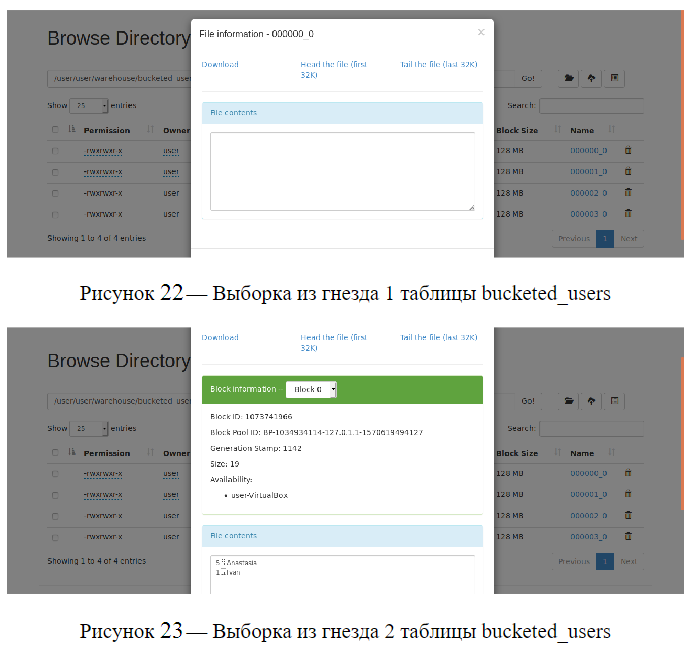
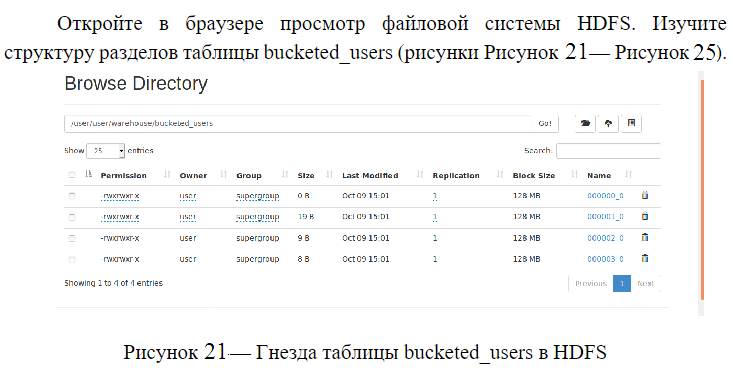
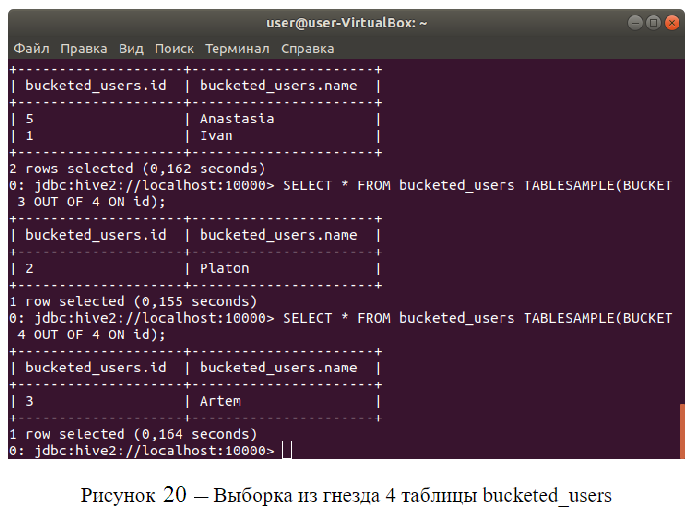
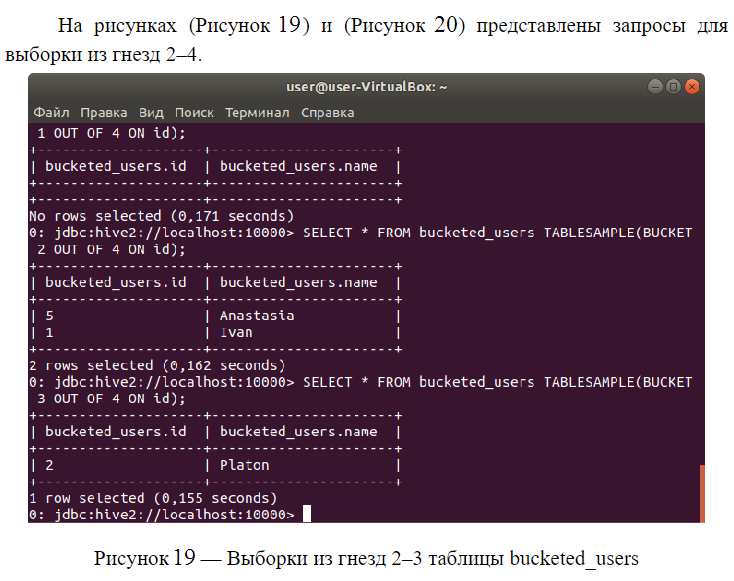
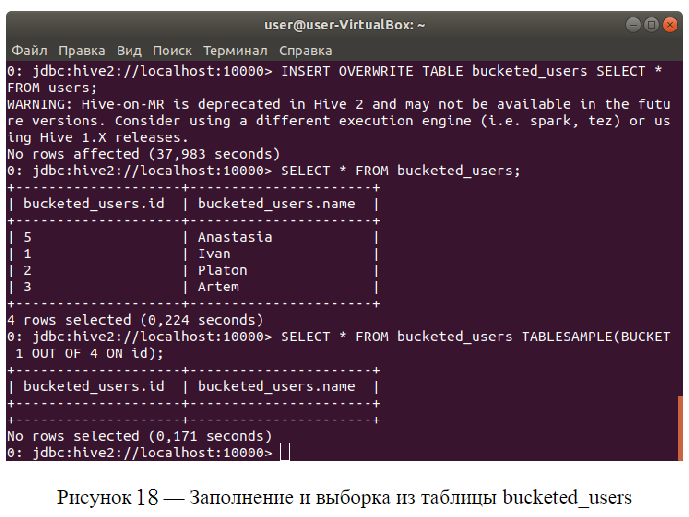
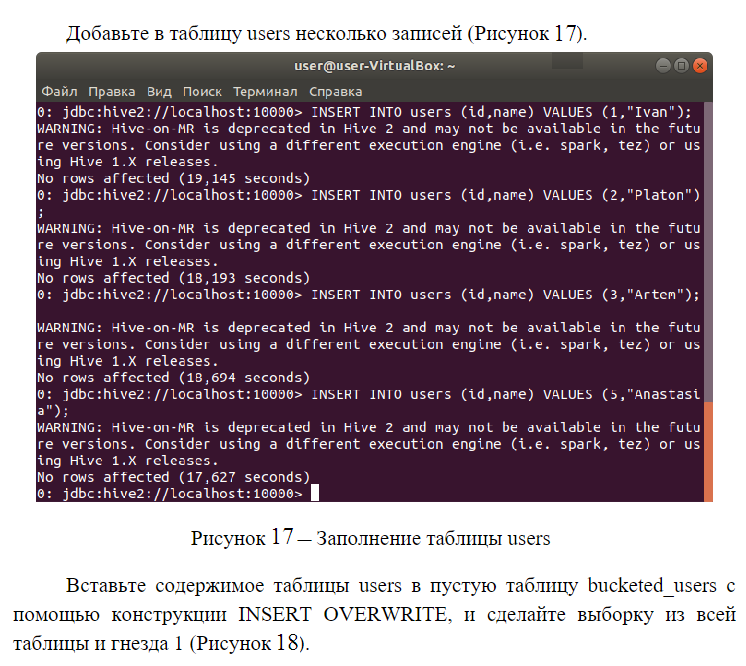
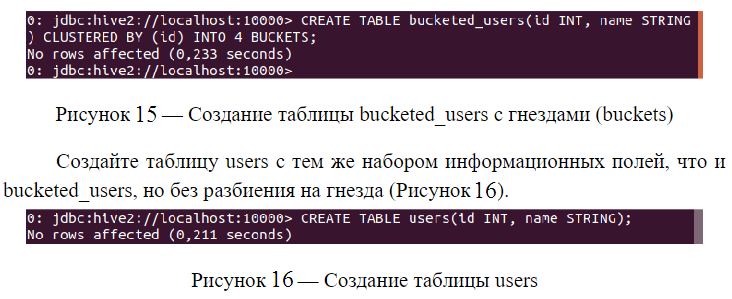
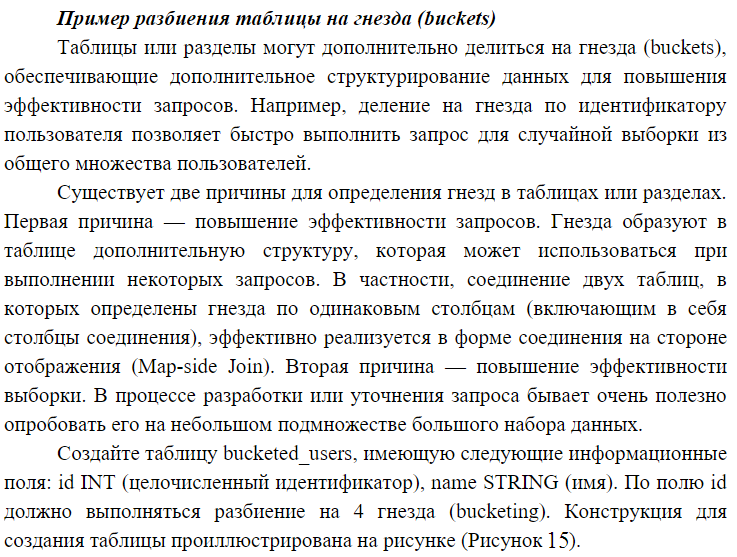
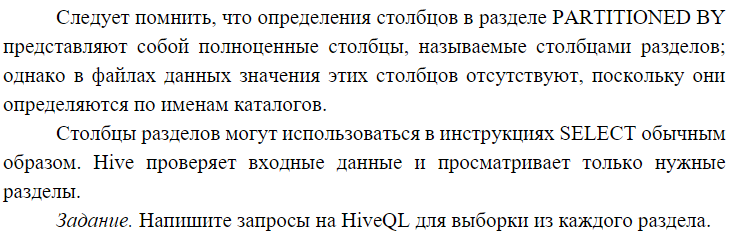
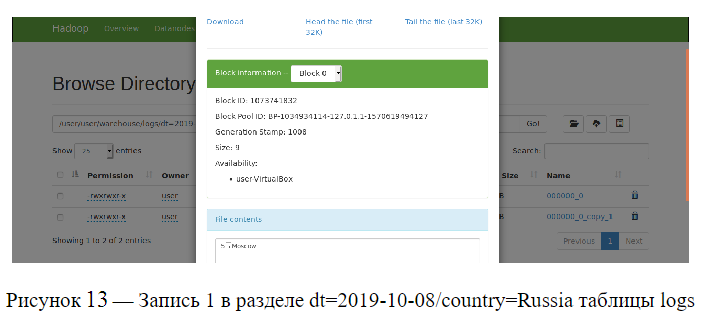
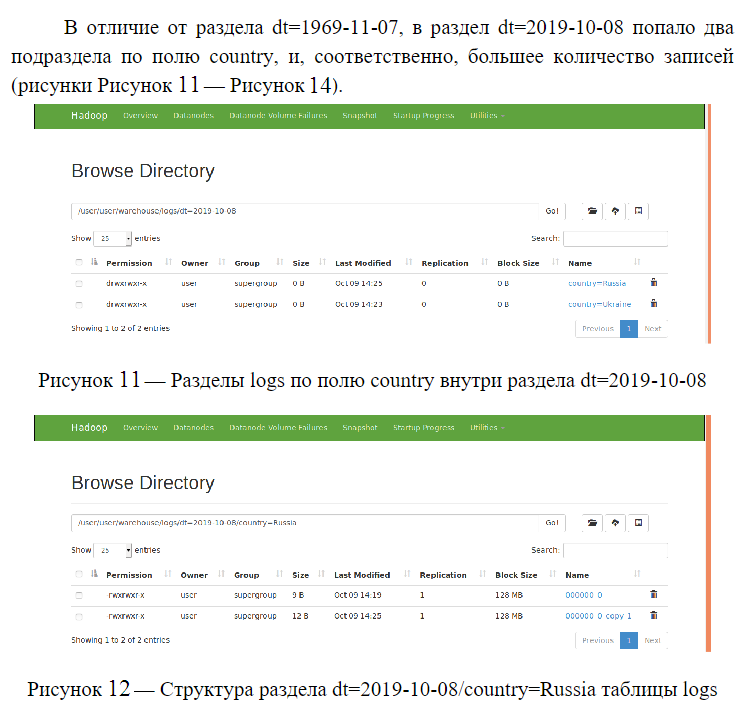
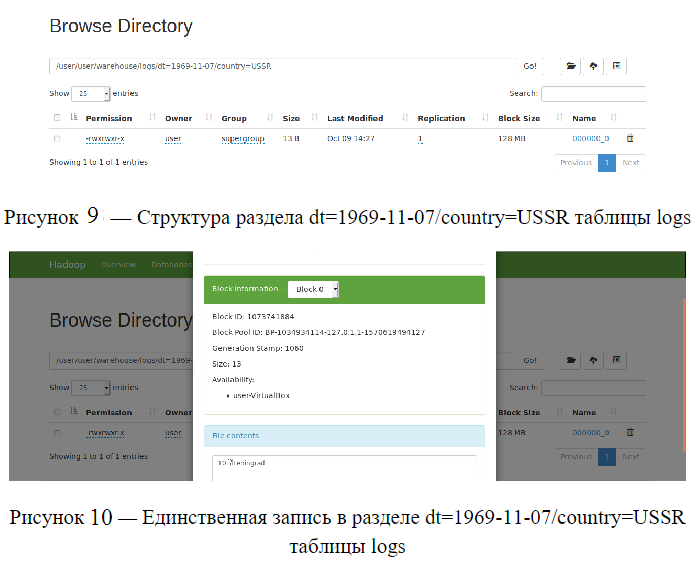
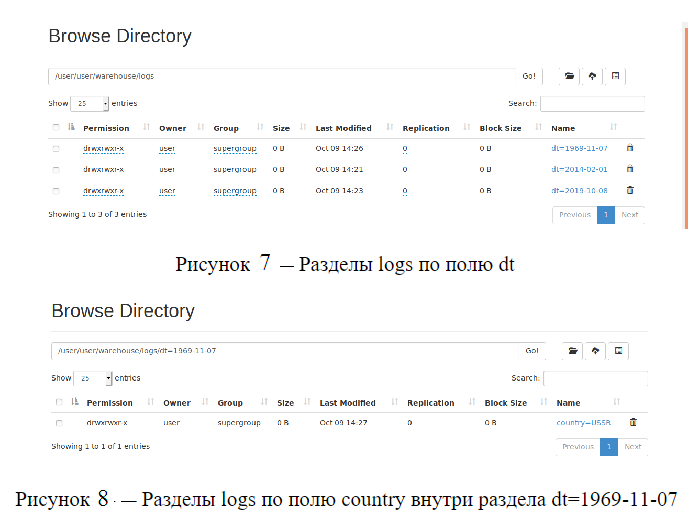
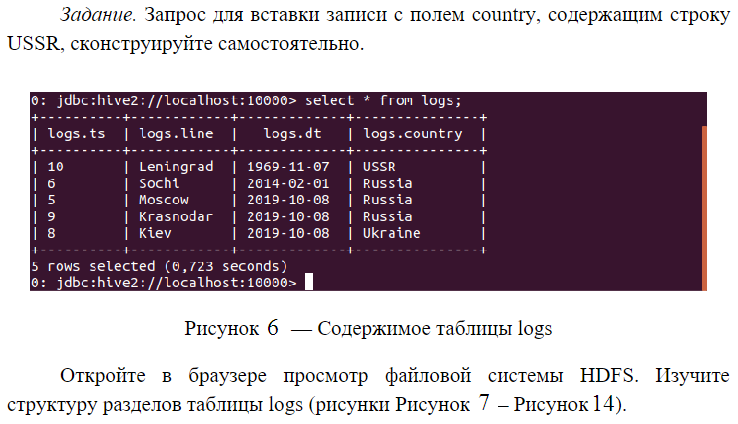
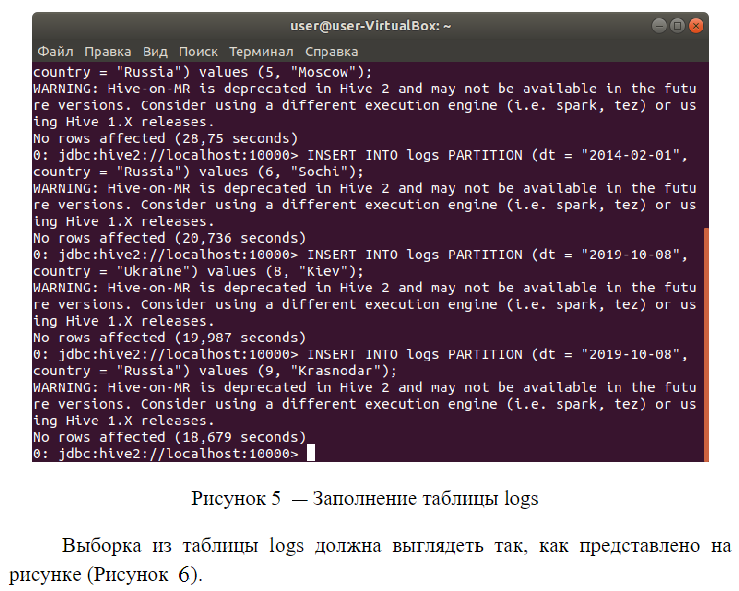
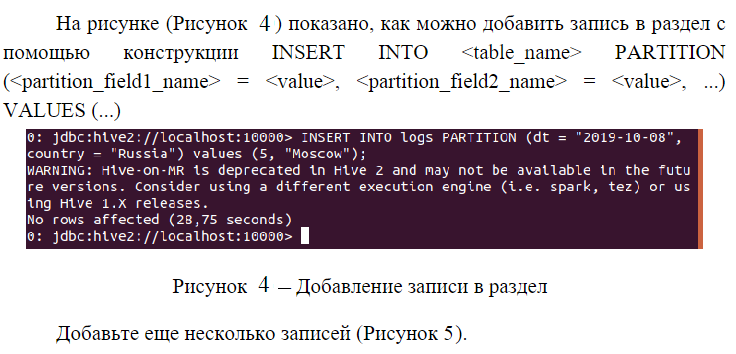
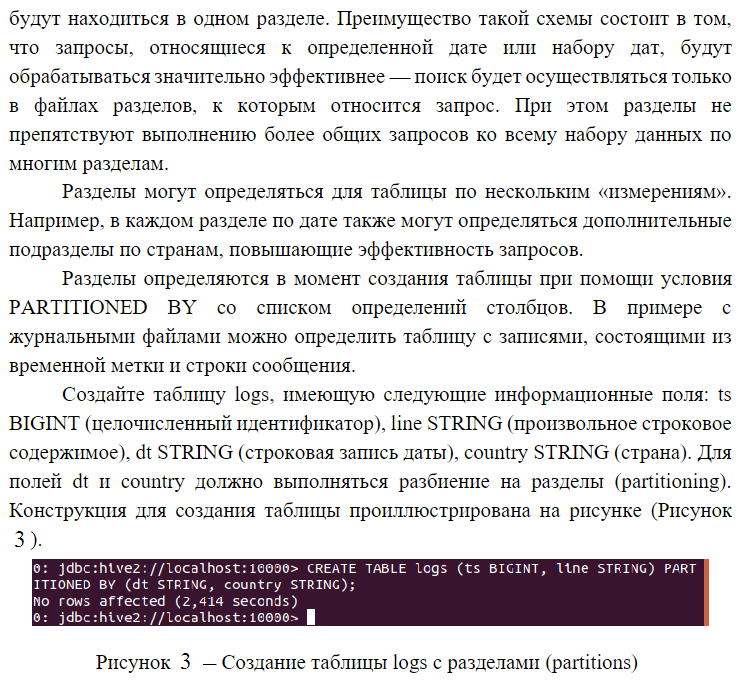
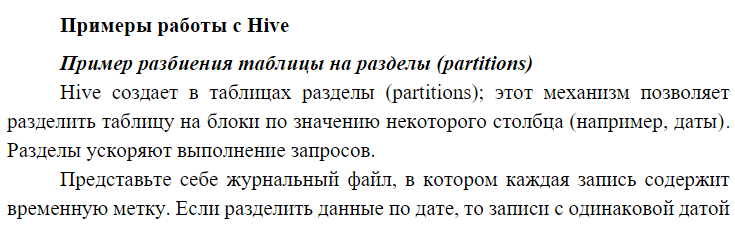
**Электронные ресурсы**

# Hadoop – MapReduce [Электроный ресурс]. – URL: <https://coderlessons.com/tutorials/bolshie-dannye-i-analitika/uchitsia-hadoop/hadoop-mapreduce> Доступ свободный

**Приложение 1.**

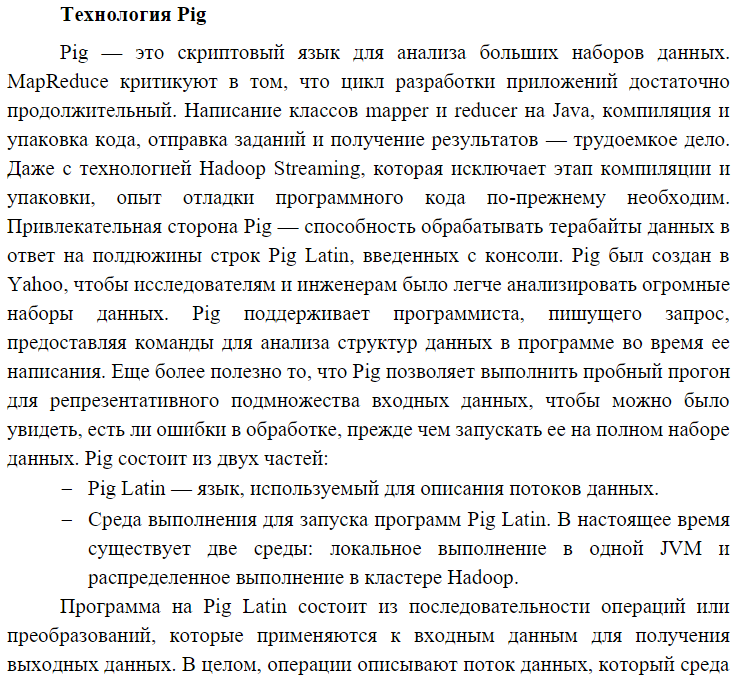
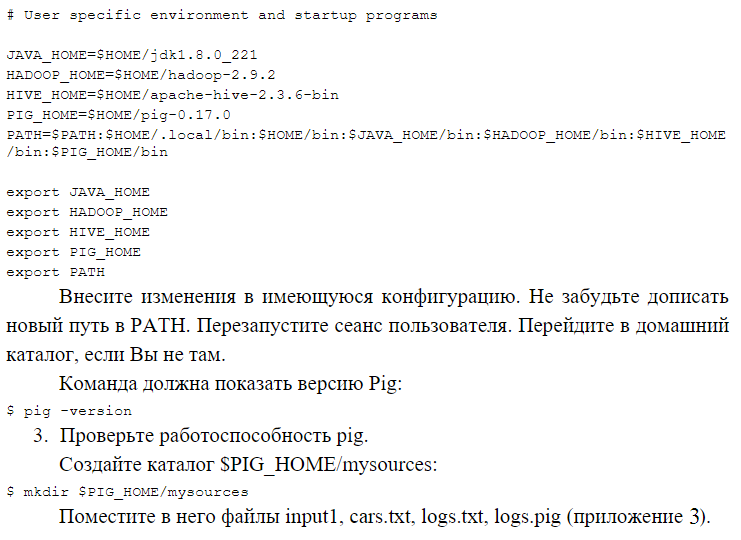
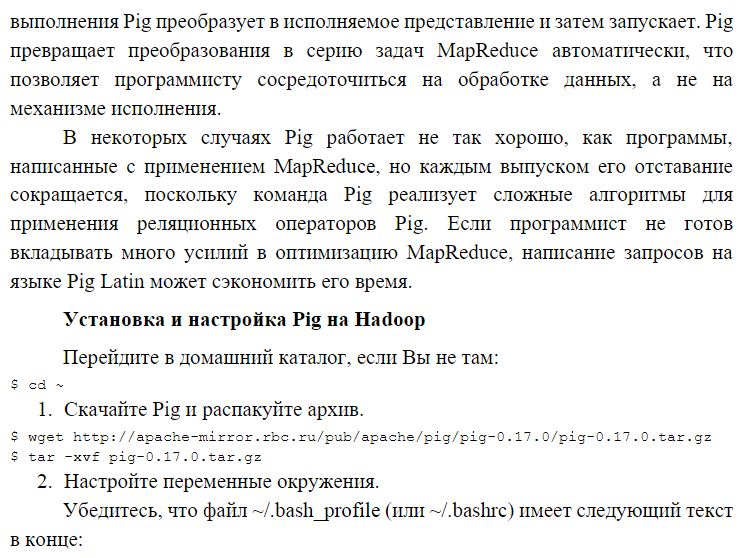
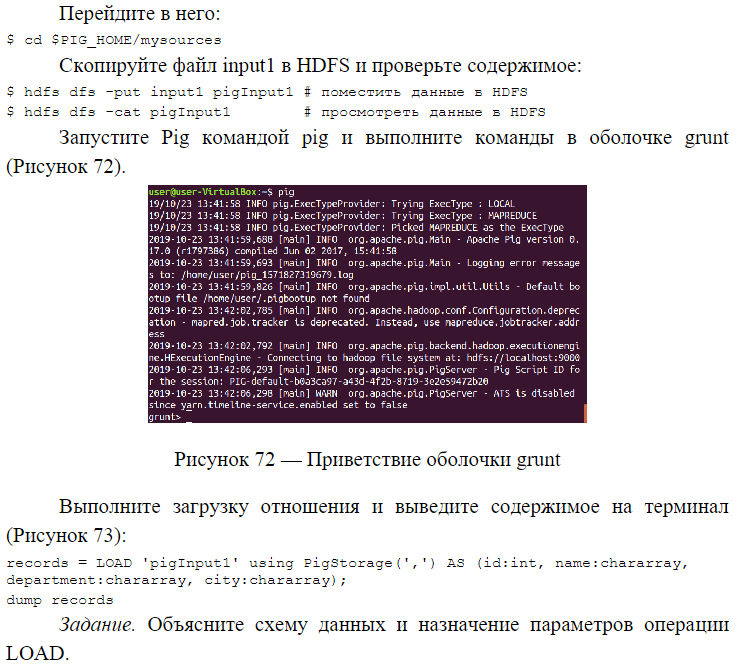
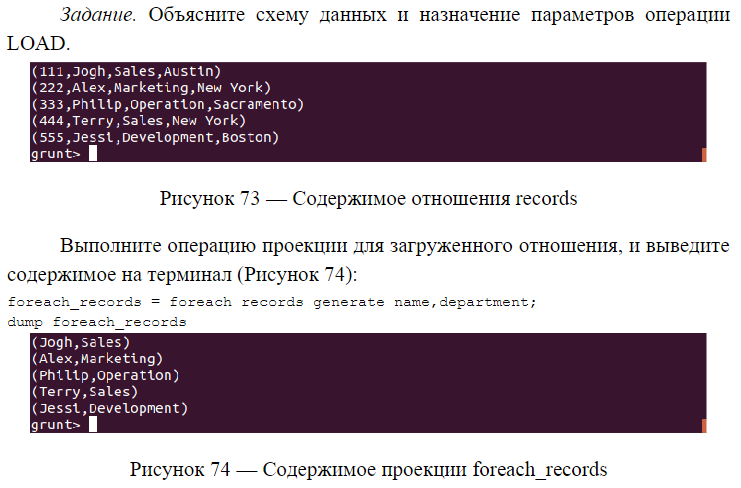
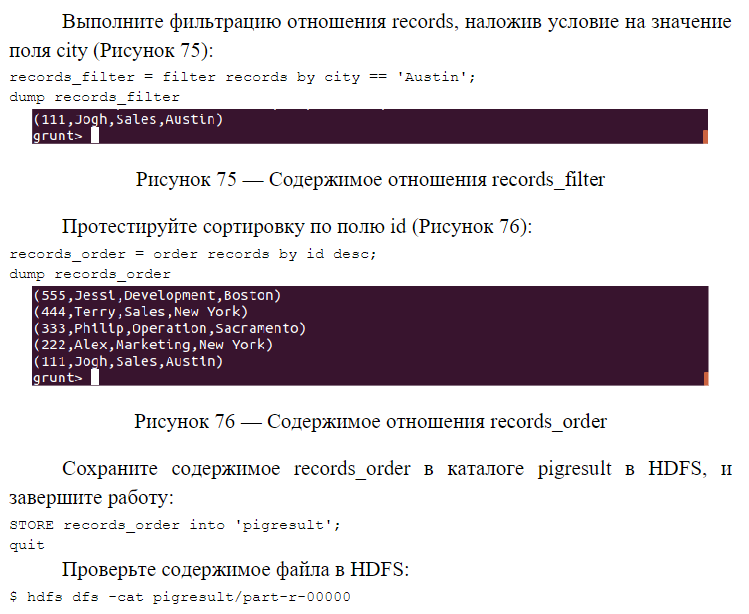
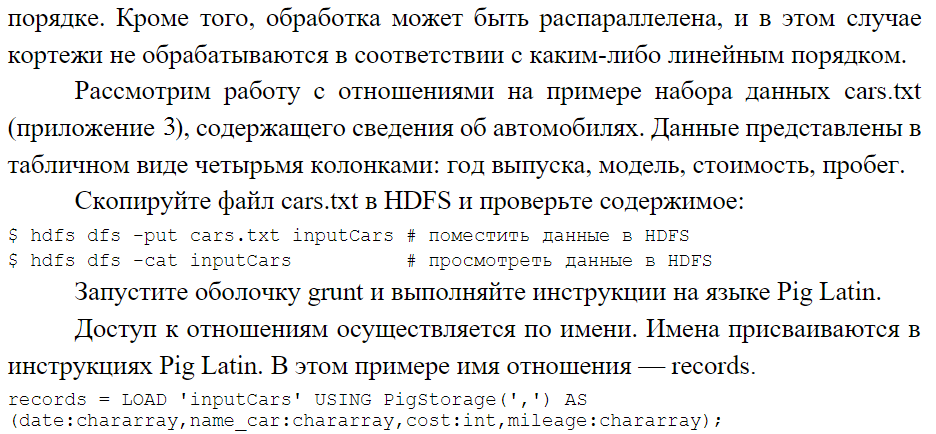
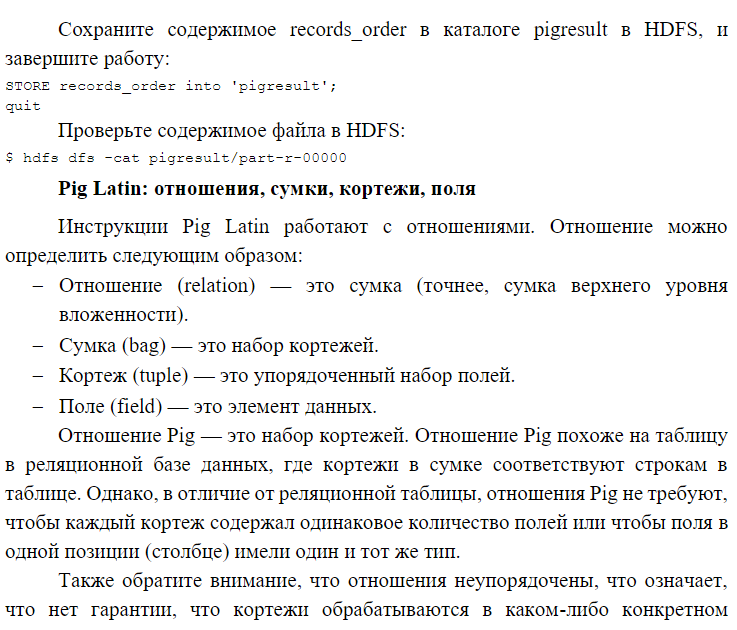
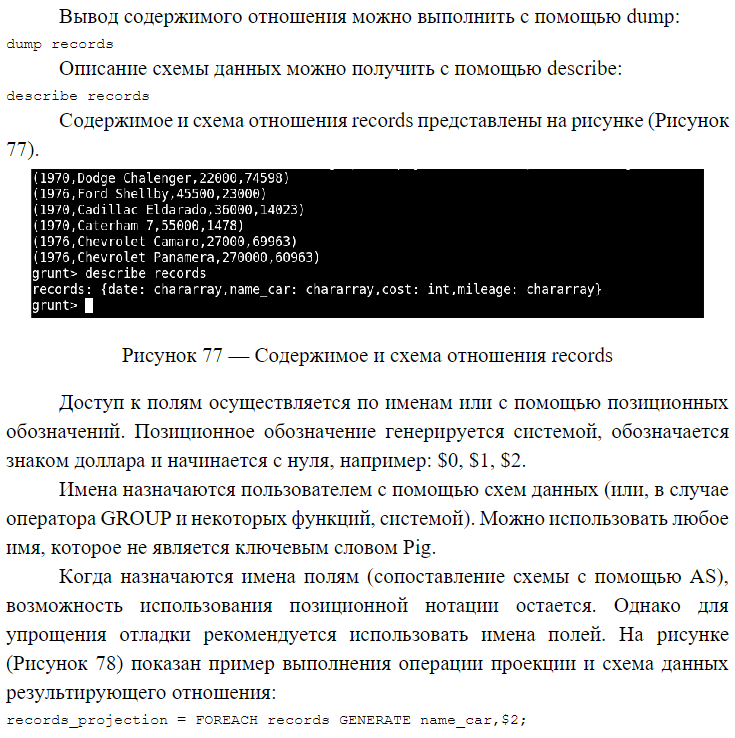
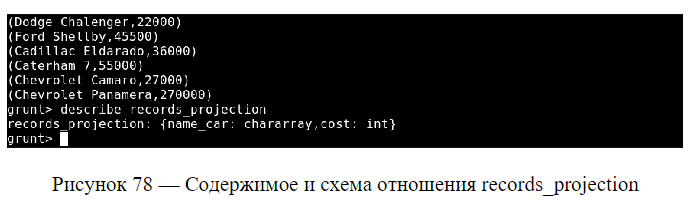
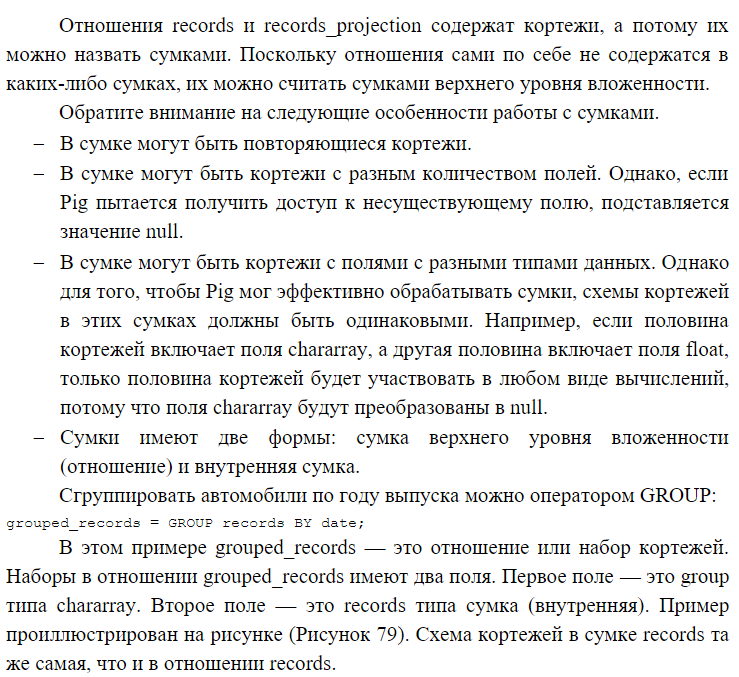
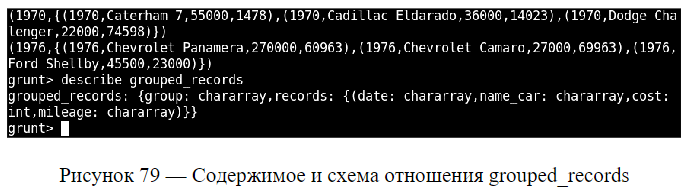
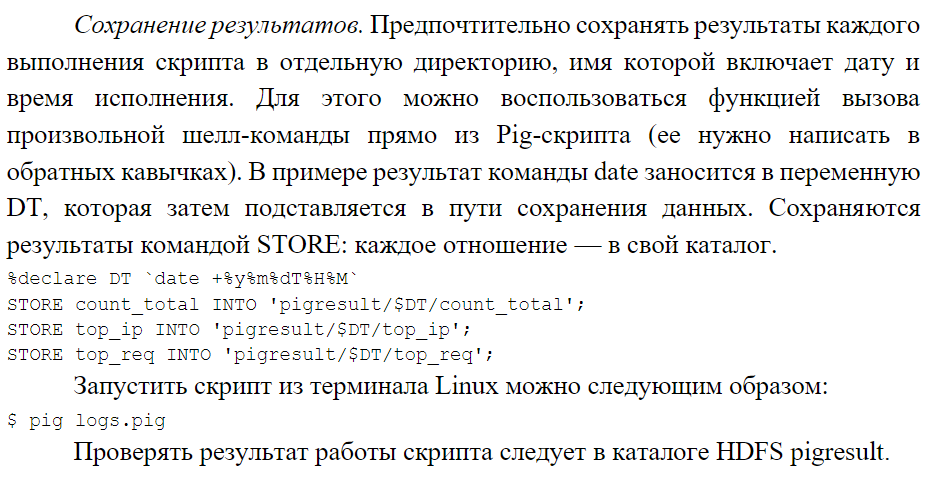
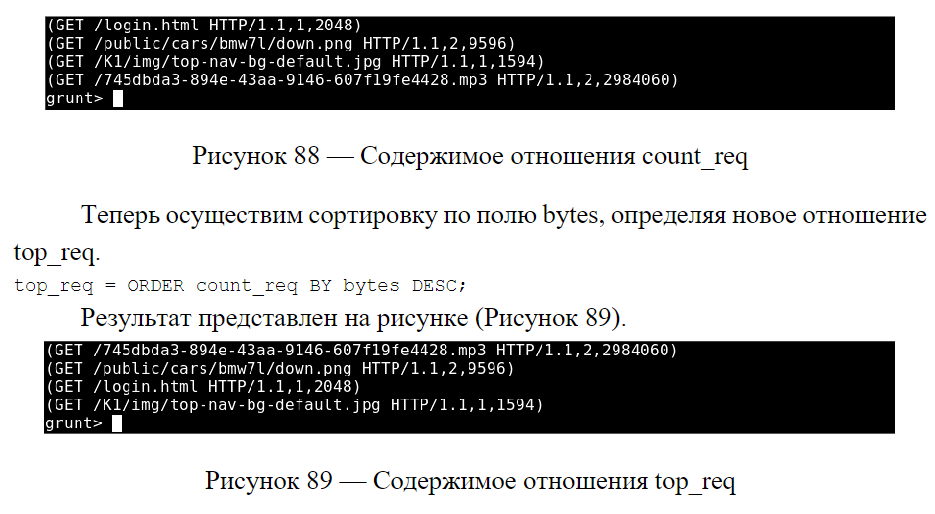
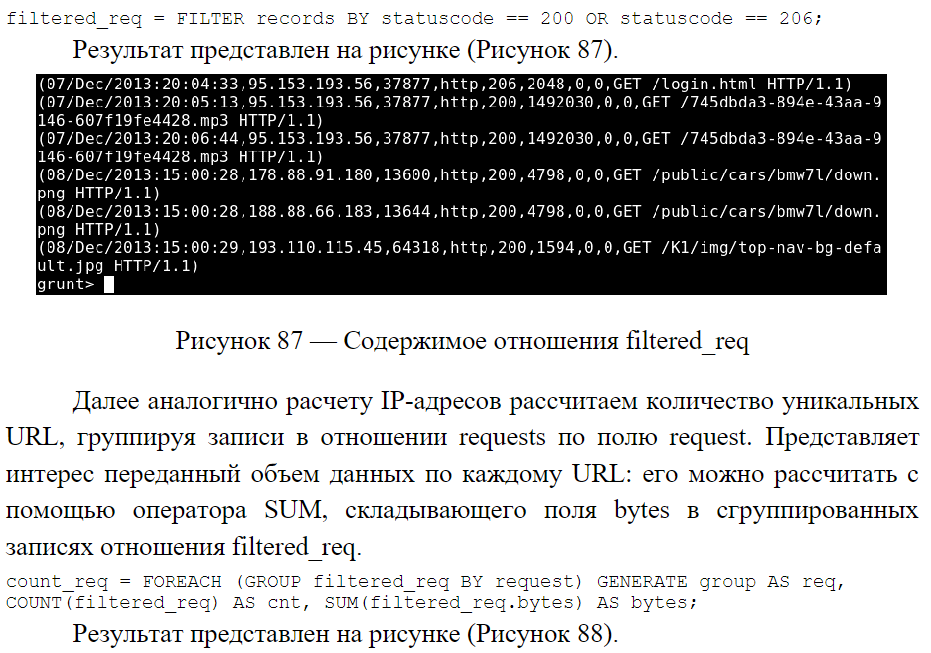
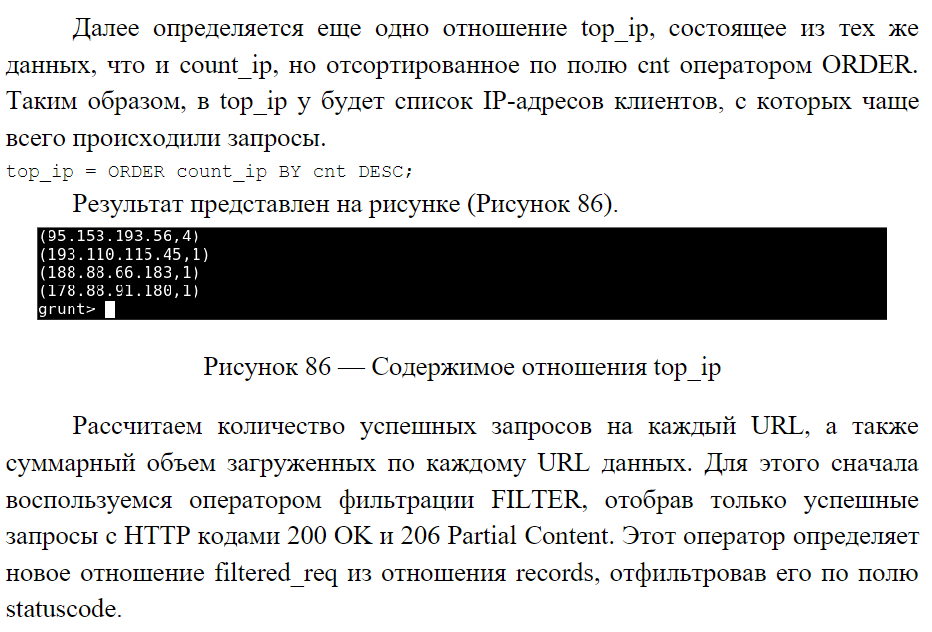
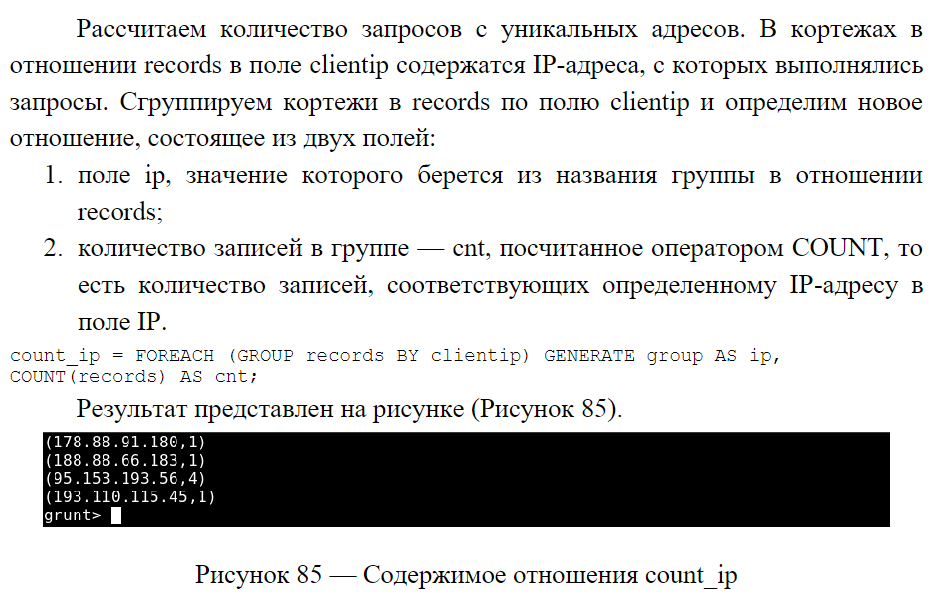
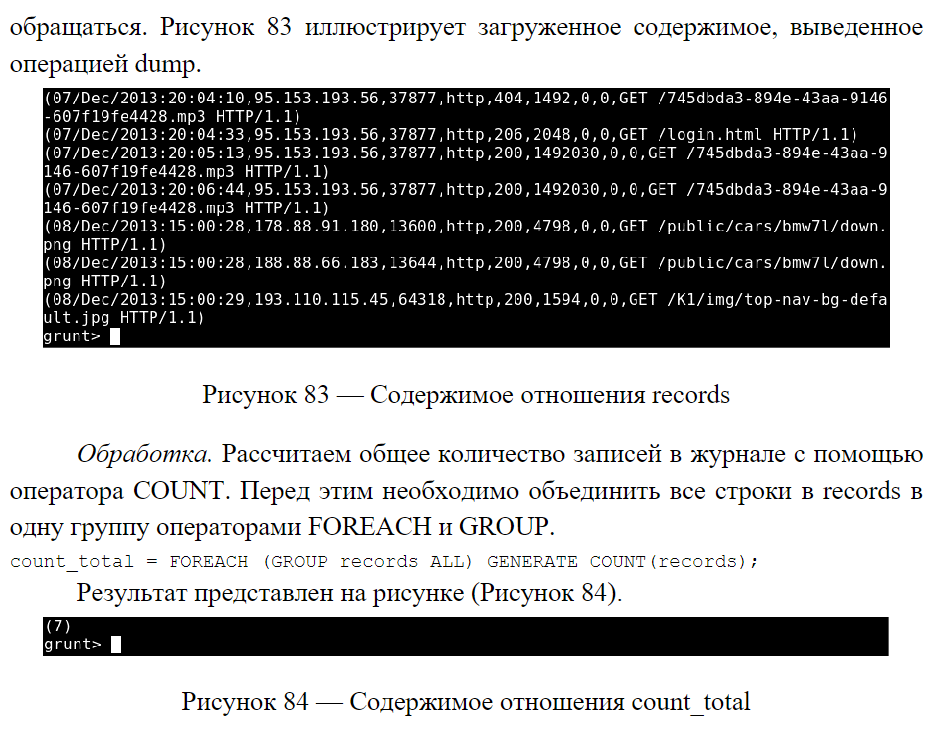
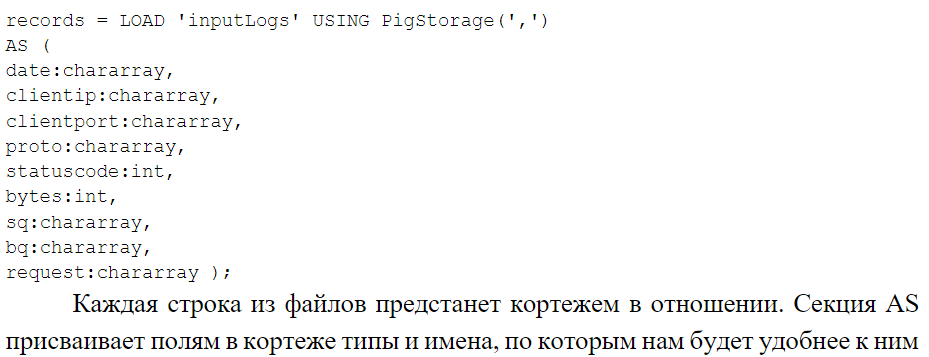
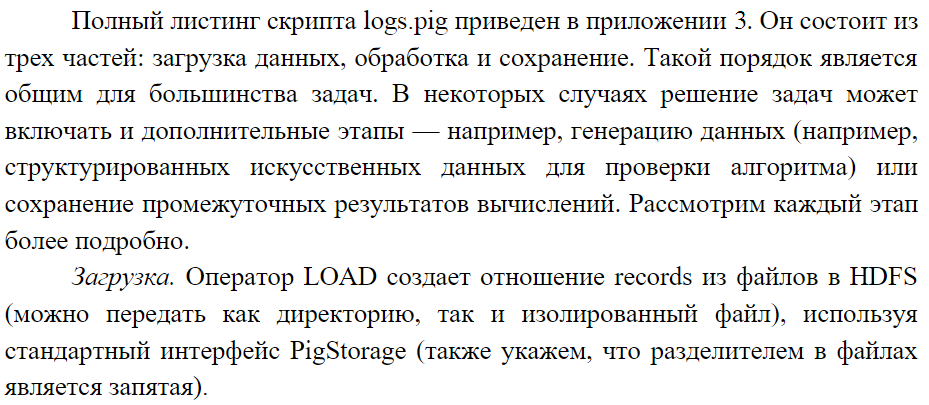
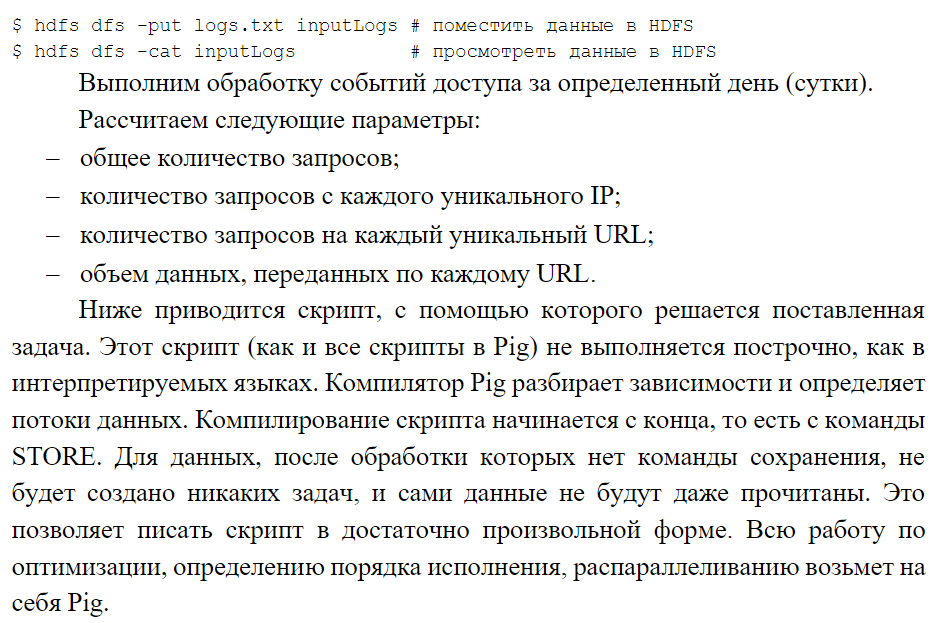
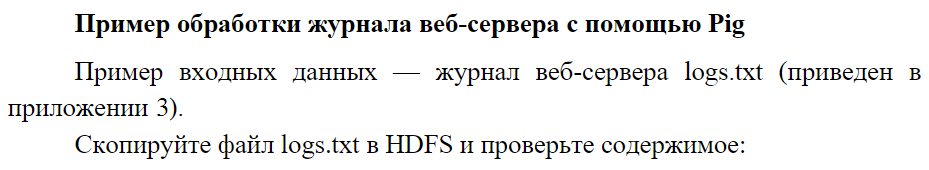
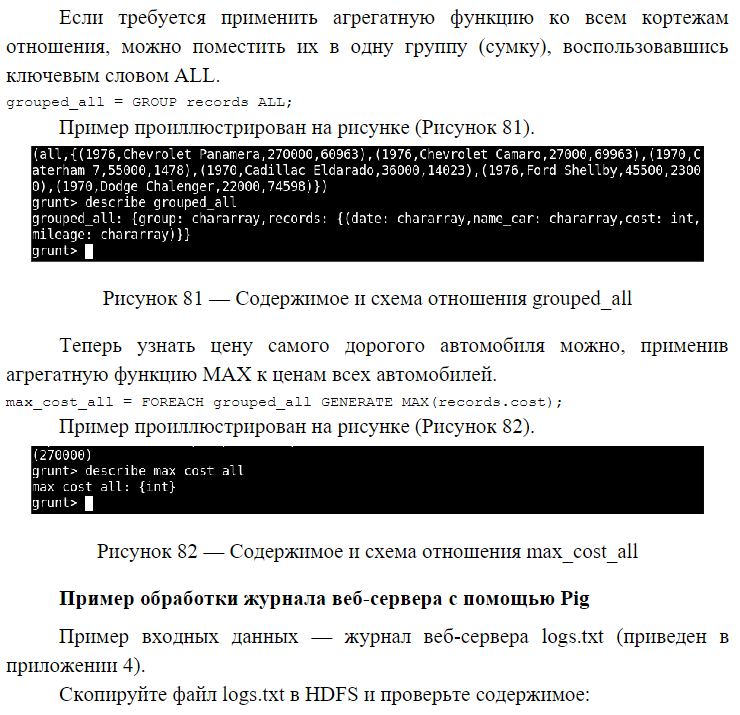
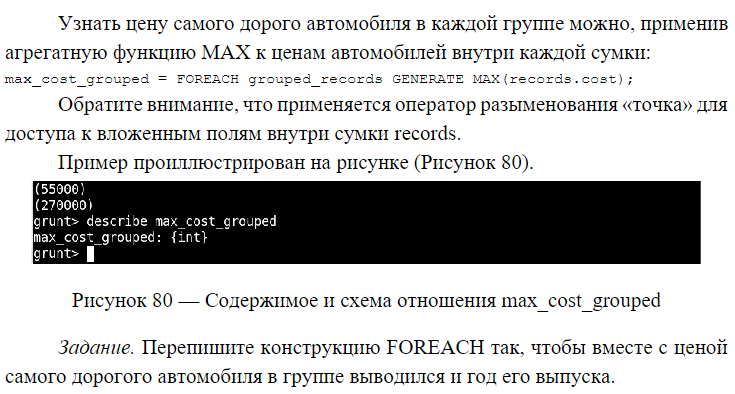
**Пример разработки программы с помощью паттерна Hive**

**   **

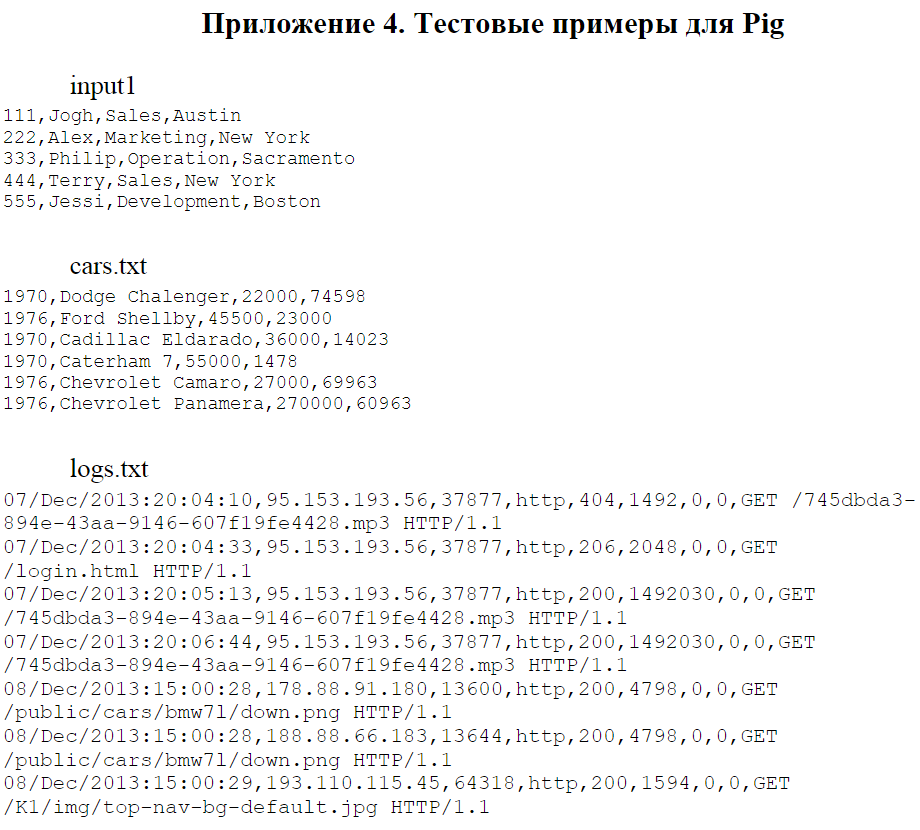
****

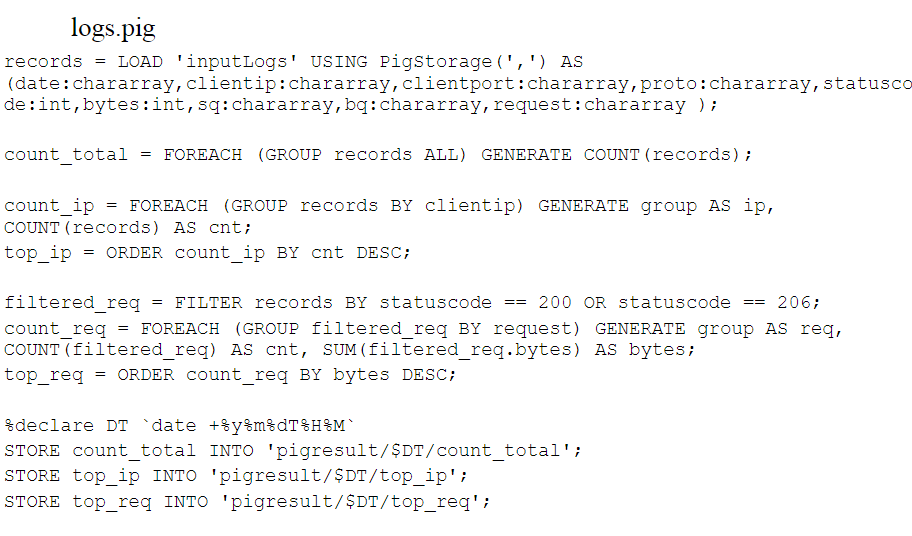
**Приложение 2.**

**Пример разработки программы с помощью паттерна Pig**

**          **

**Приложение 3. Тестовые примеры для Pig**

****

****