



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)М

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА ИУ6

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОМУ ПРОЕКТУ

НА ТЕМУ:

**Разработка макета аналитической системы на
основе баз данных NoSQL (вариант № 6)**

Салон красоты

Студент

ИУ6-23М

(Группа)

(Подпись, дата)

Е.К. Матвиенко

(И.О. Фамилия)

Руководитель

курсового проекта

(Подпись, дата)

Ю.А. Григорьев

(И.О. Фамилия)

2022 г.

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ
Заведующий кафедрой _____
(Индекс)

(И.О.Фамилия)
« ____ » _____ 20 ____ г.

ЗАДАНИЕ на выполнение курсового проекта

по дисциплине «Технология параллельных систем баз данных»

Студент группы ИУ6-23М

_____ Матвиенко Елена Константиновна _____
(Фамилия, имя, отчество)

Тема курсового проекта

«Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)»

Направленность КП (учебный, исследовательский, практический, производственный, др.)
исследовательский

Источник тематики (кафедра, предприятие, НИР): кафедра

График выполнения проекта: 25% к 3 нед., 50% к 10 нед., 75% к 13 нед., 100% к 16 нед.

Задание

Установить виртуальную машину, системы NoSQL Elasticsearch, Neo4j, Hadoop+Spark. В **Elasticsearch** создать индекс с анализатором и маппингом, проиндексировать json-документы, разработать запросы с вложенной агрегацией, представить результаты в среде Kibana. В **Neo4j** по данным из Elasticsearch заполнить графовую базу данных, разработать и реализовать запрос к этой БД. В **Spark** по данным из Elasticsearch сформировать csv-файлы (с внутренней схемой) таблиц и сохранить их в файловой системе HDFS, написать запрос и реализовать его в Spark, проанализировать процесс выполнения запроса с использованием монитора.

Оформление курсового проекта:

Расчетно-пояснительная записка на 30-50 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.):

1. Название темы КП, задание, описание варианта.

2,3. По Elasticsearch: описание анализатора и маппинга; алгоритм программы индексации документов; тексты запросов, результаты выполнения.

4,5. По Neo4j: алгоритм программы создания и заполнения графовой БД; текст запроса, результат.

5,6,7. По Spark: алгоритм программы создания таблиц и их сохранения в HDFS; скрипт запроса к БД, результат выполнения; результат анализа работы монитора.

Дата выдачи задания « ____ » _____ 20 ____ г.

Руководитель курсового проекта

(Подпись, дата) **ЮА Григорьев**
(И.О.Фамилия)

Студент

(Подпись, дата) **Е.К.Матвиенко**
(И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

РЕФЕРАТ

РПЗ 139 с., 31 рис., 1 табл., 12 источн., 3 прил.

АНАЛИТИЧЕСКАЯ СИСТЕМА, БОЛЬШИЕ ДАННЫЕ, БАЗА ДАННЫХ, ВИЗУАЛИЗАЦИЯ ДАННЫХ, ПАРАЛЛЕЛЬНЫЕ СИСТЕМЫ, NOSQL ELASTICSEARCH, HADOOP, HDFS, SPARK, KIBANA, NEO4J, NOSQL, SPARK, САЛОН КРАСОТЫ, ПРОЦЕДУРА, КЛИЕНТ, МАСТЕР.

Цель курсового проекта: разработать макет аналитической системы для салона красоты с использованием баз данных NoSQL Elasticsearch, Neo4j, Hadoop и Spark.

Во время выполнения курсового проекта были реализованы следующие задачи:

- установка виртуальной машины с Ubuntu-20.04.3 и настройка необходимых программных компонентов в VirtualBox;
- установка системы NoSQL Elasticsearch, Neo4j, Hadoop и Spark;
- генерация JSON-файлов, содержащих данные о процедурах и мастерах, для заполнения базы данных;
- создание индекса с анализатором и маппингом, индексация JSON-документов;
- разработка и выполнение запросов к данным в соответствии с заданием;
- визуализация данных в среде Kibana;
- заполнение графовой базы данных и выполнение в ней запроса;
- анализ выполнения запросов с помощью монитора.

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.....	5
ВВЕДЕНИЕ.....	6
1. Задание и описание варианта курсового проекта.....	9
2. Установка используемых средств.....	11
2.1. Установка виртуальной машины.....	11
2.2. Установка Elasticsearch и Kibana.....	13
2.3. Установка Neo4j.....	16
2.4. Установка Hadoop и Spark.....	17
3. Работа с Elasticsearch.....	19
3.1. Создание JSON-документов.....	19
3.2. Индексация документов.....	21
3.3. Запросы с вложенной агрегацией к данным ES.....	28
3.3.1. Первый запрос.....	28
3.3.2. Второй запрос.....	30
3.3.3 Визуализация запросов в среде Kibana.....	32
Neo4j.....	40
3.4. Создание и заполнение графовой базы данных.....	40
3.5. Запрос к графовой базе данных.....	43
4. Spark.....	44
4.1. Создание и заполнение таблиц.....	44
4.2. Запрос в Spark.....	46
4.3. Мониторинг Spark.....	47
ЗАКЛЮЧЕНИЕ.....	49
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	50
ПРИЛОЖЕНИЕ А.....	51
ПРИЛОЖЕНИЕ Б.....	94
ПРИЛОЖЕНИЕ В.....	127

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

БД – база данных

ВМ – виртуальная машина

ОП – оперативная память

ОС – операционная система

СУБД – система управления базой данных

DAG (Directed Acyclic Graph) – направленный ациклический граф

ELK (Elasticsearch, Logstash, Kibana) – стек технологий, используемый при работе с Elasticsearch

ES – Elasticsearch

DDB – распределенная база данных

HDFS – Hadoop Distributed File System

MapReduce – модель распределенных вычислений, используемая для параллельных вычислений над очень большими объемами данных

Анализ – процесс преобразования неструктурированного текста в формат, оптимизированный для поиска

Стемминг – процесс нахождения основы слова для заданного исходного слова

Токенизация – процесс разбиения неструктурированного текста на более мелкие, доступные для поиска фрагменты, называемые токенами

Хост-машина – машина, на которой запускается другая виртуальная операционная система

Индекс – совокупность документов JSON формата

Кластер (ES) – группа из одного или нескольких связанных узлов Elasticsearch

Маппинг – схема базы данных в Elasticsearch.

Kibana – свободная тиражируемая программная панель визуализации данных

REST (Representational State Transfer) – набор архитектурных принципов взаимодействия компонентов распределённого приложения в сети

ВВЕДЕНИЕ

В настоящее время параллельные системы баз данных используются во многих ИТ-проектах и активно вытесняют компьютеры основного класса. Параллельные СУБД предоставляют возможность работы со значительно более крупными БД, при этом являются основой высокопроизводительных серверов высокой доступности и обладают более привлекательной стоимостью.

В процессе выполнения курсового проекта были использованы технологии параллельных СУБД, актуальность которых обуславливается вышесказанным.

Основной задачей курсового проекта является разработка макета аналитической системы салона красоты. При реализации данного макета использованы такие системы как NoSQL Elasticsearch, Hadoop, Spark и Neo4j.

В основе макета аналитической системы салона красоты лежат данные о процедурах и мастерах салонов. Эти данные представлены в различном виде в зависимости от используемой БД. В работе использованы БД Elasticsearch и графовая БД Neo4j. Также в макет входят файлы csv, расположенные в файловой системе HDFS.

Elasticsearch (ES) является масштабируемым поисковым движком с аналитическими функциями. Он используется, когда необходимо выполнить сложный поиск по базе документов. Например, поиск, учитывающий географические координаты или морфологию языка. Для записи данных в ES были сгенерированы соответствующие JSON-файлы.

ES – это ядро ELK-стека. ELK-стек состоит из Logstash, Kibana и соответственно Elasticsearch. Kibana позволяет взаимодействовать с данными, расположенными в индексах ES и представлять их в виде таблиц, графиков и диаграмм. В данном курсовом проекте web-интерфейс Kibana использовался для визуализации запросов, сделанных в ES.

Neo4j представляет собой графовую систему управления базами данных с открытым исходным кодом. Эта СУБД реализована на языке Java. Составляющими графа являются узлы – объекты БД и ребра – отношения между объектами. Помимо наглядного представления взаимосвязей между объектами,

данная БД удобна тем, что в ней можно использовать связи при реализации поисковых запросов. Neo4j использует собственный язык запросов - Cypher.

Поскольку Cypher предоставляет функции для графового хранилища его также можно считать языком манипулирования данными [1].

Hadoop представляет собой набор утилит, библиотек и фреймворк. Все они используются для организации и работы распределённых программ, которые размещены на различных узлах [2]. Hadoop работает по принципу MapReduce – одно задание разбивается на множество небольших задач, и каждая из этих задач выполняется на отдельном узле, входящем в определенный кластер. В итоге все эти задачи объединяются и выводится общий результат.

Для того чтобы управлять множеством файлов на разных узлах применяется распределённая файловая система HDFS (Hadoop Distributed File System). HDFS хранит файлы на различных серверах (DataNodes), адреса серверов расположены на соответствующем сервере имен (NameNode). Благодаря вышесказанному принципу HDFS способна хранить файлы значительных размеров.

Безусловно, для того чтобы обеспечить корректную работу, масштабирование и надежность распределенных приложений необходимо строго соблюдать соответствие версий продуктов, а также провести полную и правильную установку пакетов. В ходе работы над проектом для данной задачи использовался пакетный менеджер YARN (Yet Another Resource Negotiator).

При работе с Hadoop нельзя не упомянуть о MapReduce фреймворке Spark. Apache Spark – это фреймворк с открытым исходным кодом, в основном применяемый для больших данных. Он используется при необходимости распределённой пакетной и потоковой обработки неструктурированных и слабоструктурированных данных Hadoop [3].

Отличительной особенностью Spark является использование оперативной памяти для работы с данными, чего нельзя сказать о базовом Hadoop. Такая особенность непременно повышает производительность выполнения поставленных задач.

Во время выполнения курсового проекта применялись приведенные выше технологии и продукты. Процесс реализации аналитического макета салона красоты можно представить в виде следующих этапов:

- 1) Установка программного продукта виртуализации для операционной систем Linux – VirtualBox;
- 2) Установка ОС Ubuntu 20.04.3 на разворачиваемую ВМ;
- 3) Установка программных пакетов: Elasticsearch, Neo4j, Hadoop и Spark;
- 4) Генерация JSON-файлов с исходными данными;
- 5) Создание индексов с анализатором и маппингом в Elasticsearch
- 6) Реализация запросов к БД с последующей визуализацией результатов в среде Kibana;
- 7) Заполнение графовой базы данных Neo4j данными из Elasticsearch
- 8) Реализация запросов к БД Neo4j;
- 9) Формирование csv-файлов с таблицами и сохранение этих файлов в файловой системе HDFS.
- 10) Реализации запроса в Apache Spark;
- 11) Анализ процесса выполнения запроса в Apache Spark с использованием мониторинга.

1. Задание и описание варианта курсового проекта

Для успешной реализации курсового проекта необходимо выполнить следующие задачи:

1. установить VM с ОС Ubuntu 20.04.3 в VirtualBox по предлагаемому дистрибутиву. Требования к VM: имя – Ubuntu-01, тип – Linux, версия ОС – Ubuntu (64 bit), объём ОП – 4ГБ, диск – 20ГБ;
2. установить Elasticsearch, Neo4j, Hadoop+Spark;
3. автоматически создать два JSON-файла с 20 – 30 JSON-документами каждого типа для предметной области «Салон красоты» с указанными в варианте полями документов;
4. в Elasticsearch: создать индекс с анализатором и маппингом, проиндексировать JSON-документы, разработать запросы с вложенной агрегацией, представить результаты в среде Kibana;
5. в Neo4j: по данным из ES заполнить графовую базу данных, разработать и реализовать запрос к этой БД;
6. в Spark: по данным из Elasticsearch сформировать csv-файлы с таблицами и сохранить их в файловой системе HDFS, написать запрос и реализовать его в Spark, проанализировать процесс выполнения запроса с использованием монитора.

В соответствии с шестым номером варианта предметной областью курсового проекта является салон красоты. В задании требуется сгенерировать два JSON-документа со следующими полями:

- Процедура

```
{index, doc_type, id, body: {id_клиента, возраст, персональные_данные*,
id_процедуры, дата_процедуры, стоимость, id_специалиста, [услуга*],
[препарат*]}}
```

- Специалист (в дальнейшем будет использовано “мастер”)

```
{index, doc_type, id, body: {специальность, стаж_работы,
сведения_о_тренере, [отзыв_о_тренере*] } }.
```

Примечание. Квадратные скобки [] обозначает тег (может быть несколько значений).

Требование к анализатору: поля, отмеченные *, разделить на слова, убрать пунктуацию с помощью токенизатора standart (русский), перевести все токены в нижний регистр, убрать токены, находящиеся в списке стоп-слов, выполнить стемминг оставшихся токенов с помощью фильтра snowball.

Помимо генерации файлов в ES требуется реализовать два запроса с вложенной агрегацией:

- разбить процедуры по дате с периодом 1 год, для каждой «корзины» определить суммарную стоимость по каждому специалисту,
- предложить признаки отрицательного отзыва; вывести специалистов хотя бы с одним отрицательным отзывом.

Результаты данных запросов представить в среде Kibana.

Для графовой БД Neo4j необходимо сделать следующее:

1. По данным из Elasticsearch заполнить графовую базу данных Клиент(id_клиента, персональные_данные) - Посетил(дата_процедуры, стоимость) - Специалист(id_специалиста, специализация, стаж работы).

Примечание. В скобках приведены свойства узлов и отношения (связи), глагол – это отношение. Между двумя узлами может быть несколько связей одного типа.

2. Разработать и реализовать запрос: определить специалиста с наибольшей стоимостью выполненных процедур.

Для Spark необходимо выполнить следующие задачи:

1. По данным из Elasticsearch сформировать csv-файлы (с внутренней схемой) таблиц «Клиент», «Процедура», «Специалист» и сохранить их в файловой системе HDFS.
2. Написать запрос select: определить число выполненных процедур каждым специалистом.
3. Реализовать этот запрос в Spark. Построить временную диаграмму его выполнения по результатам работы монитора.

2. Установка используемых средств

2.1. Установка виртуальной машины

Первым этапом в ходе работы над курсовым проектом был установлен программный продукт виртуализации для операционной систем Linux – VirtualBox. Преимуществом данной среды является компактность, высокая производительность, а также возможность бесплатного использования

В VirtualBox была развернута ВМ в соответствии с данными в задании характеристиками: объём ОП – 4ГБ, диск – 20ГБ. В качестве дистрибутива использован образ Ubuntu 20.30.40 - ubuntu-20.04.3-desktop-amd64.iso.

На рисунке 1 представлена конфигурация развернутой ВМ Ubuntu-01 с ОС Ubuntu 20.30.40.

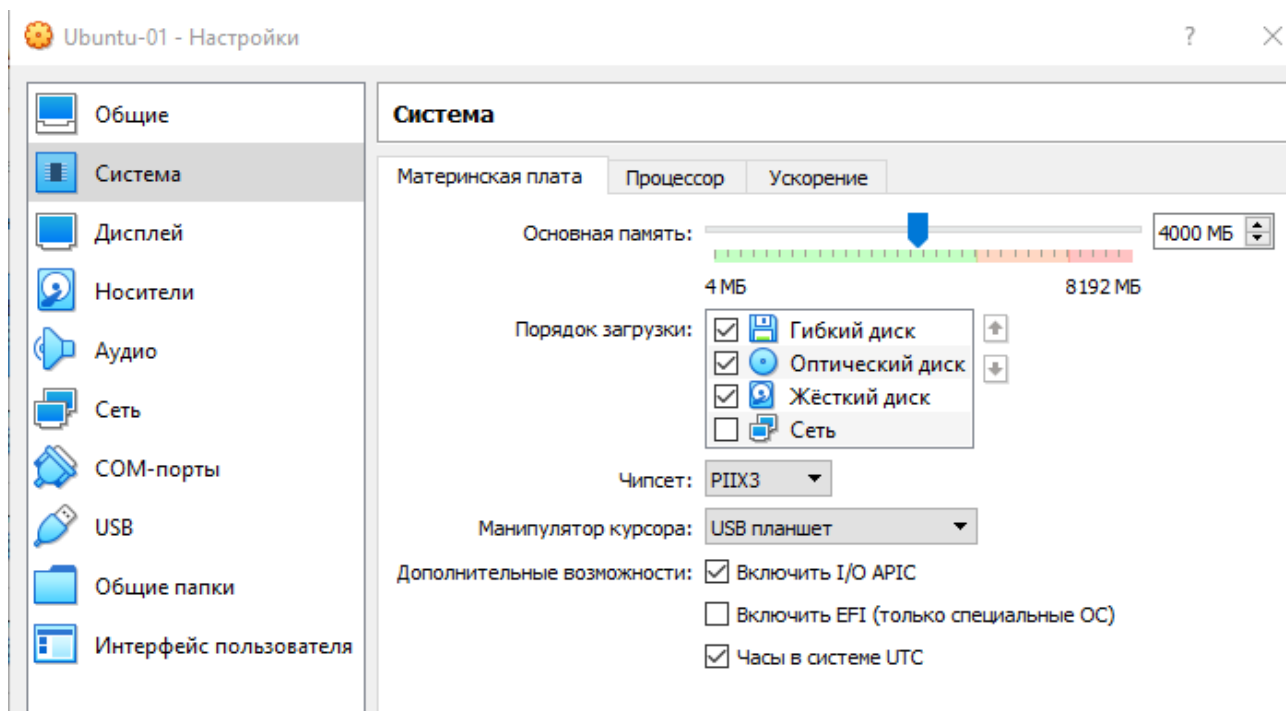


Рисунок 1 – Конфигурация ВМ Ubuntu-01

После успешной установки Ubuntu 20.30.40 в VirtualBox была создана учетная запись с логином `hadoopuser` и паролем `hadoop`.

Для удобства работы со средой были установлены функции общего буфера и Drag'n'Drop, позволяющие копировать текст и команды в основной ОС и вставить эти данные в машину на VirtualBox. В параметрах данных функций

был указан двунаправленный обмен, поэтому эти же функции работают и в обратную сторону. На рисунке 2 приведен пример данных настроек.

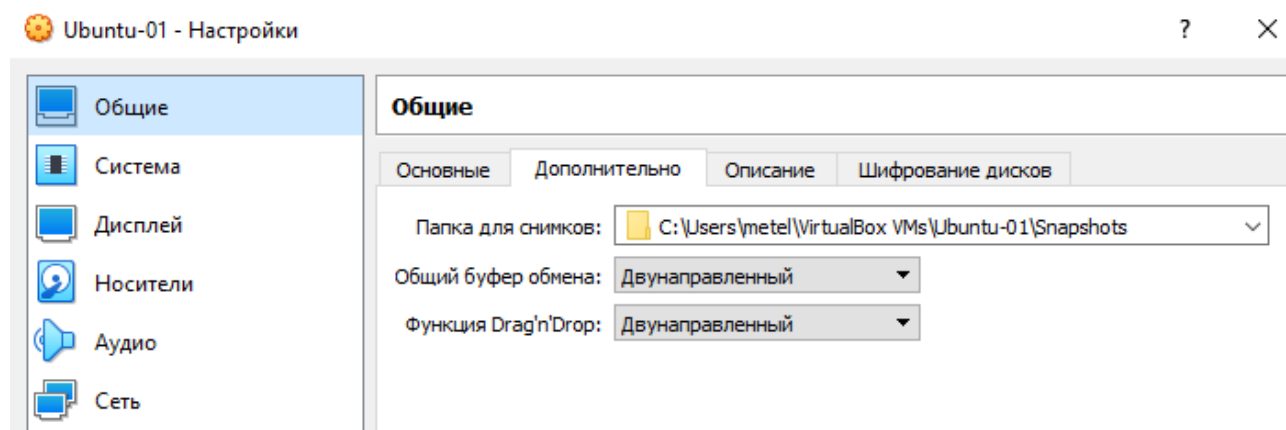
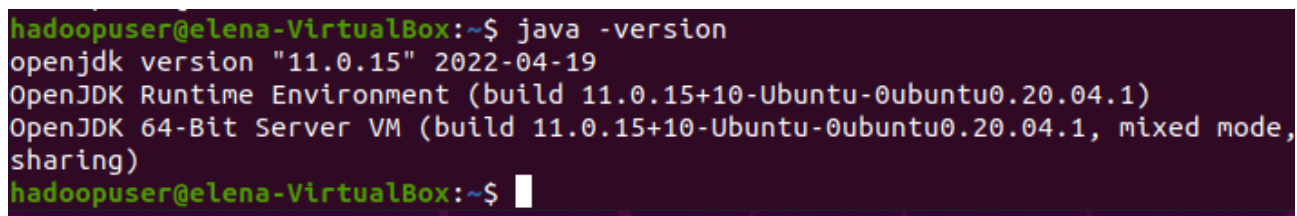


Рисунок 2 – Дополнительные функции VirtualBox

2.2. Установка Elasticsearch и Kibana

Для дальнейшей работы необходима установка Elasticsearch. Значительными преимуществами ES являются такие функции как хранение значительного объема данных, выполнение по ним быстрого поиска, но самое основное – возможность анализа больших данных практически в режиме реального времени. В Elasticsearch библиотека Lucene используется как основа для выполнения поисковых запросов. Эта библиотека реализована на языке Java, что делает ее доступной для различных платформ. Стоит отметить, что рекомендуется установка Java не ниже 8 версии. На рисунке 3 приведен результат выполнения команды «java -version», по которому видно, что версия Java 11.0.15, следовательно она соответствует требованиям.



```
hadoopuser@elena-VirtualBox:~$ java -version
openjdk version "11.0.15" 2022-04-19
OpenJDK Runtime Environment (build 11.0.15+10-Ubuntu-0ubuntu0.20.04.1)
OpenJDK 64-Bit Server VM (build 11.0.15+10-Ubuntu-0ubuntu0.20.04.1, mixed mode,
sharing)
hadoopuser@elena-VirtualBox:~$
```

Рисунок 3 – Результат выполнения команды java -version

Поскольку в стандартных репозиториях ОС Ubuntu нет компонентов стека ES, они были добавлены в каталог sources.list. Именно там менеджер apt будет осуществлять поиск новых пакетов. Чтобы исключить добавление поддельных пакетов был добавлен ключ GPG Elasticsearch.

Сначала командой «sudo apt update» были обновлены индексы пакетов, а потом установлен ES.

Для запуска ES применяется команда “sudo systemctl start elasticsearch”. Но для того, чтобы каждый раз не запускать ES этой командой была прописана другая команда - «sudo systemctl enable elasticsearch». Благодаря ей ES будет автоматически запускаться при каждом включении ВМ.

Чтобы проверить работоспособность сервера необходимо прописать команду «curl -X GET "localhost:9200"». Результат вывода данной команды представлен на рисунке 4.

```
hadoopuser@elena-VirtualBox:~$ curl -X GET 'localhost:9200'
curl: (3) Failed to convert -X to ACE; string contains a disallowed character

curl: (6) Could not resolve host: GET
{
  "name" : "elena-VirtualBox",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "5F1ncZaVSKSCUH10NEcxtA",
  "version" : {
    "number" : "7.10.1",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "1c34507e66d7db1211f66f3513706fdf548736aa",
    "build_date" : "2020-12-05T01:00:33.671820Z",
    "build_snapshot" : false,
    "lucene_version" : "8.7.0",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
hadoopuser@elena-VirtualBox:~$
```

Рисунок 4 – Вывод команды «curl -X GET "localhost:9200"»

По выводу команды видно, что ES находится в активном состоянии. Помимо информации о работоспособности системы можно получить данные о версии – ES 7.10.1, кластере – elasticsearch и ВМ – elena-VirtualBox.

С одной стороны ES можно отнести к базам данных NoSQL, поскольку неструктурированные данные хранятся в JSON-файлах, но с другой – ES позволяет делать поисковые запросы и некоторые другие функции. Файлы, хранящиеся в ES можно просмотреть в браузере. В работе используется браузер Mozilla, поскольку в нем файлы JSON отображаются более наглядно благодаря подсветке и возможности свернуть отдельные части документа. Пример начальной страницы ES в браузере Mozilla представлен на рисунке 5.

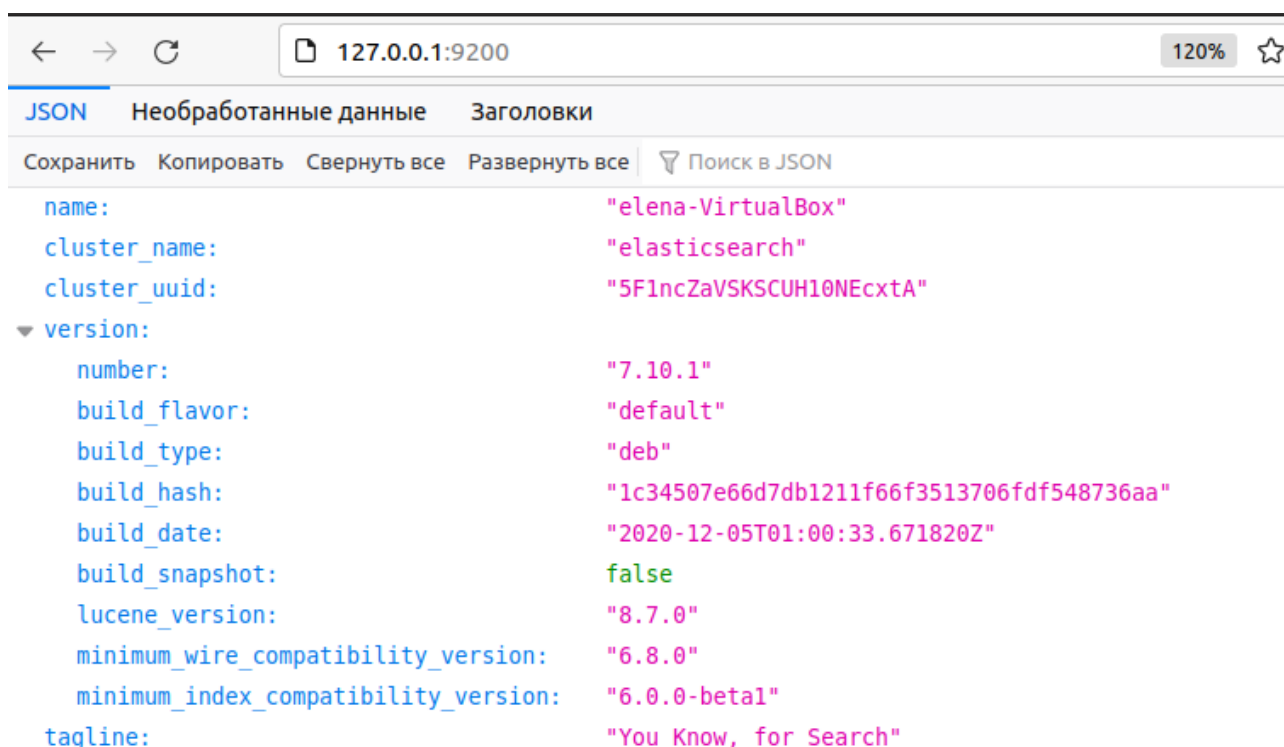


Рисунок 5 – Начальная страница ES

Для выполнения запросов используется среда Kibana. Установка данной программы происходит при помощи команды «`sudo apt-get install kibana=x`», где `x` – версия Kibana. Стоит отметить, что каждой версии ES соответствуют определенные версии Kibana. Совместимость версий можно посмотреть на официальном сайте [4]. В нашем случае была установлена версия 7.10.1 Kibana, поскольку она совместима с ES 7.10.1. После запуска сервиса была проверена работоспособность данного сервиса. Для этого в браузере был сделан переход по адресу <http://localhost:5601/app/kibana>. На рисунке 6 продемонстрирован web-интерфейс данного сервиса.

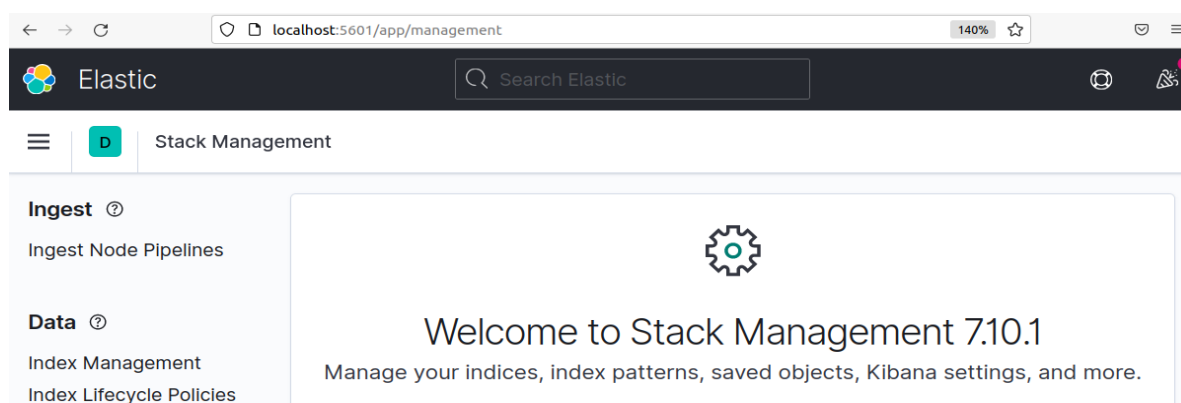


Рисунок 6 – Web-интерфейс Kibana

2.3. Установка Neo4j

Следующим шагом была установка Neo4j. По аналогии с установкой ES сначала был импортирован ключ GPG, далее к менеджеру пакетов apt был добавлен сам пакет Neo4j. Завершительными шагами стали обновление индексов пакетов и запуск команды «`sudo apt install neo4j=3.5.14`», благодаря которой и был установлен сервис.

Далее был запущен сам сервис командой «`sudo systemctl start neo4j`» и активирован автоматический запуск командой «`sudo systemctl enable neo4j`». Для проверки работоспособности графовой СУБД была использована команда «`service --status-all | grep neo4j`». На рисунке 7 приведен вывод данной команды, подтверждающий успешную установку и запуск сервиса.

```
elena@elena-VirtualBox:~$ sudo systemctl start neo4j
elena@elena-VirtualBox:~$ sudo systemctl enable neo4j
Synchronizing state of neo4j.service with SysV service script with /lib/systemd/
systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable neo4j
Created symlink /etc/systemd/system/multi-user.target.wants/neo4j.service → /lib
/systemd/system/neo4j.service.
elena@elena-VirtualBox:~$ service --status-all | grep neo4j
[ + ] neo4j
elena@elena-VirtualBox:~$
```

Рисунок 7 – Проверка успешной установки и запуска Neo4j

Также для Neo4j доступен графический web-интерфейс, доступный по адресу <http://localhost:7474>. На рисунке 8 представлен данный интерфейс.

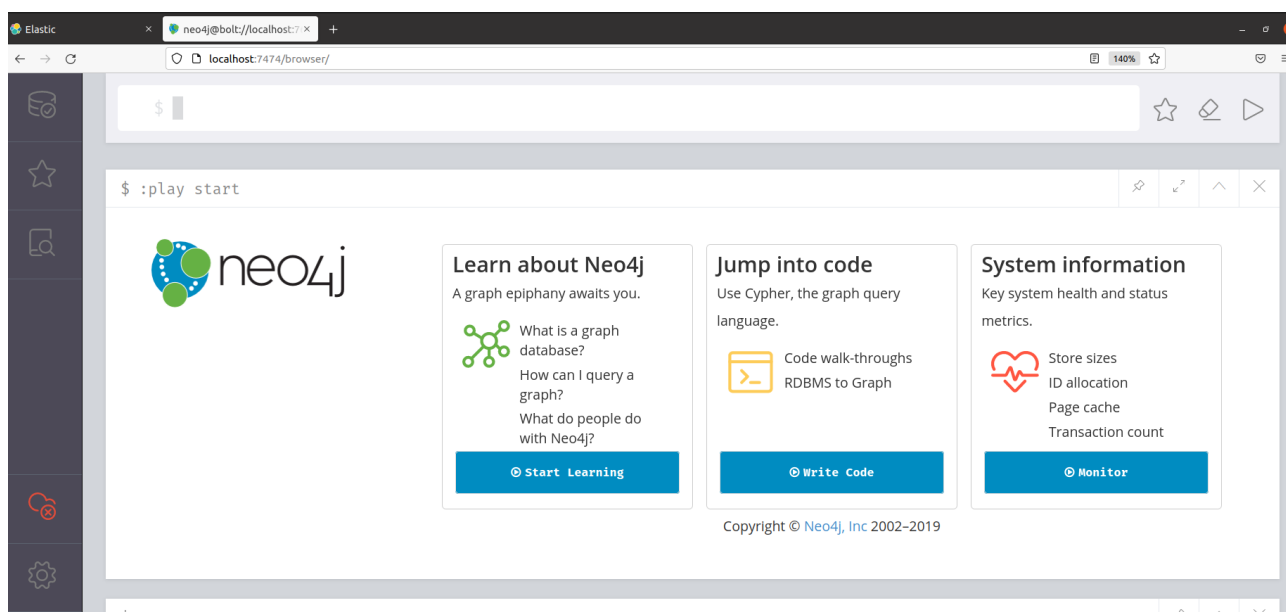


Рисунок 8 – Web-интерфейс Neo4j

2.4. Установка Hadoop и Spark

Установки Hadoop и Spark осуществлялась в соответствии с методическими указаниями лабораторных работ 7-8 по курсу «Технология параллельных систем баз данных».

Изначально были скачены архивы для установки - `hadoop-3.3.0.tar.gz` и `spark-3.2.1-bin-hadoop2.7.tgz`. Далее была создана группа `hadoop`, в которой был создан пользователь для работы с Hadoop – `hadoopuser`.

Следующим шагом были распакованы скаченные ранее пакеты и сделаны необходимые для корректной работы исправления в файлах конфигурации [5].

Таким образом, был создан однонодовый кластер Hadoop.

Для запуска Hadoop используются менеджеры HDFS и YARN. Для запуска HDFS применяется команда «`start-dfs.sh`», а для запуска YARN - «`start-yarn.sh`».

Для проверки успешного запуска Hadoop можно перейти по адресу <http://localhost:8088>. Web-интерфейс Hadoop представлен на рисунке 9.

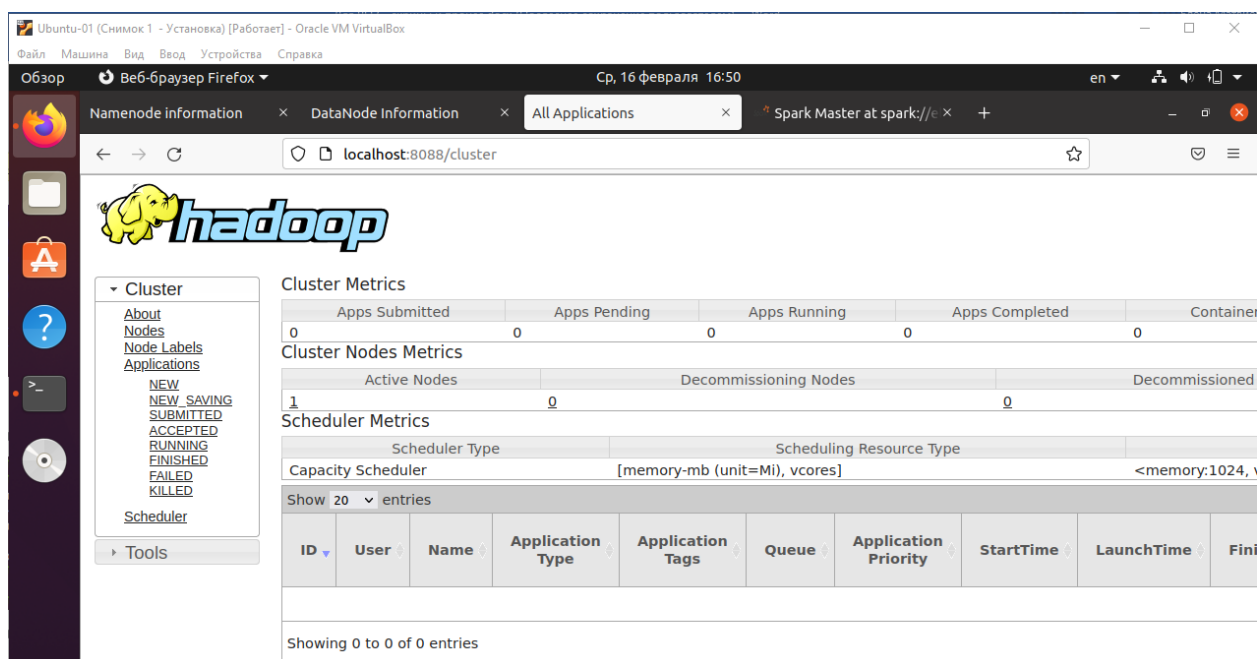


Рисунок 9 – Web-интерфейс Hadoop

При запуске Hadoop начинают работать определенные процессы и демоны. Для их просмотра используется команда «`jps`», вывод которой представлен на рисунке 10.

```
hadoopuser@elena-VirtualBox:~$ jps
10545 NodeManager
9716 NameNode
10916 Jps
9880 DataNode
9177 SparkSubmit
10377 ResourceManager
10105 SecondaryNameNode
hadoopuser@elena-VirtualBox:~$
```

Рисунок 10 – Запущенные процессы и демоны Hadoop

Для проверки успешной установки Spark можно перейти по адресу <http://localhost:8080>. Web-интерфейс Spark представлен на рисунке 10.

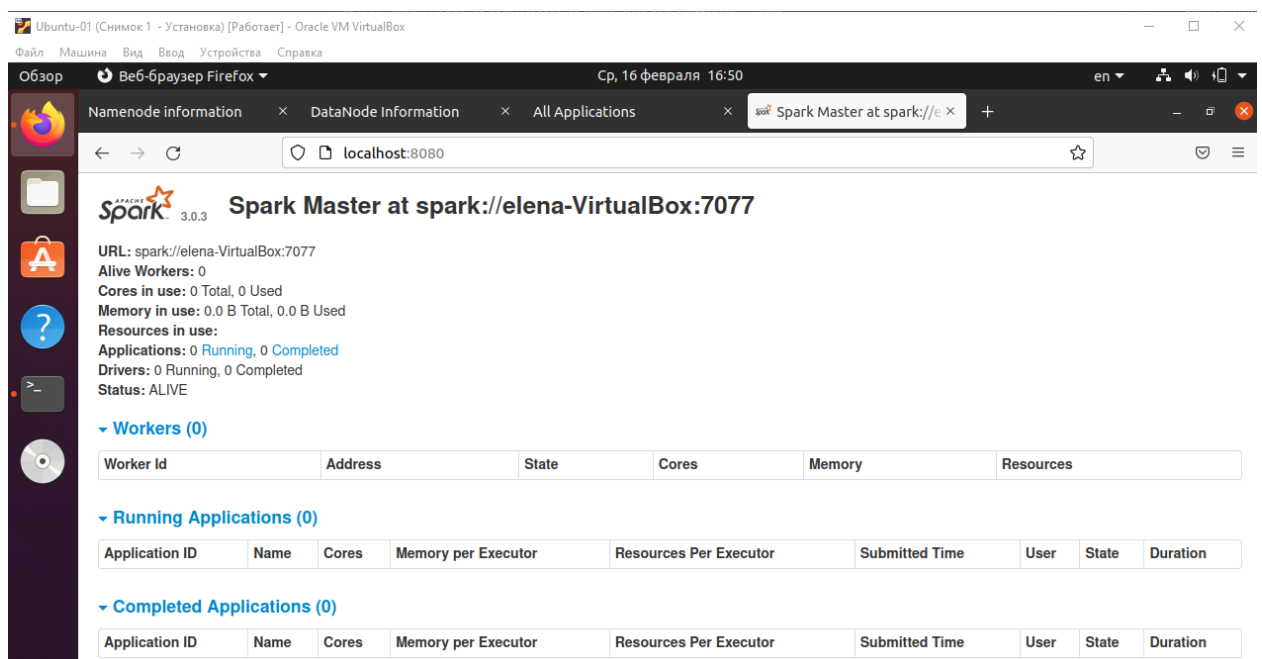


Рисунок 10 – Web-интерфейс Hadoop

3. Работа с Elasticsearch

3.1. Создание JSON-документов

Для того, чтобы наполнить кластер ES нужными данными сначала нужно создать два JSON-документа: Специалист и Процедура. Для создания этих документов были использованы различные онлайн сервисы. Сначала был создан список, содержащий фамилии, имена и отчества при помощи генератора ФИО[6]. Далее полученные данные были помещены в кавычки благодаря онлайн-сервису Webbos [7]. Чтобы использовать полученные данные их необходимо было поместить в массив, а значит после каждого ФИО поставить запятую. Для этой задачи применялся сервис Lineeditor [8]. В итоге были созданы две программы, которые были обработаны онлайн-генератором JSON-документов [9]. Программы генерации JSON-Procedure и JSON-Master представлены в приложении Б.

В итоге были сгенерированы два файла - JSON-Procedure, содержащий в себе 40 записей и JSON-Master, содержащий в себе 25 записей. Просмотреть полученные документы можно в приложении А. Стоит отметить, что при генерации документа Процедура использовались реальные отзывы, собранные с различных сайтов салонов красоты.

В документе типа Процедура содержатся следующие данные:

- `id_of_client` – уникальный набор символов и цифр, позволяющий идентифицировать клиента салона красоты;
- `client_age` – возраст клиента, который лежит в диапазоне от 20 до 80 лет;
- `client_personal_data` – данное поле содержит ФИО клиента;
- `id_of_procedure` – уникальный набор символов и цифр, позволяющий идентифицировать каждую из процедур, сделанных в салоне красоты;
- `date_of_procedure` – дата проведения процедуры;
- `price` – цена сделанной процедуры;
- `id_of_specialist` - уникальный набор символов и цифр, позволяющий идентифицировать мастера, сделавшего процедуру;

- service – услуги, которые были предоставлены клиенту в каждой процедуре;
- cosmetics – косметические средства, которые были использованы во время проведения процедуры.

При генерации файла Процедура уникальные идентификаторы мастеров салона были взяты из того же списка, который использовался для заполнения файла Специалист.

В документе типа Специалист содержатся следующие данные:

- specialisation – специализация мастера, перечисление услуг, на которых специализируется данный мастер;
- work_experience – опыт работы мастера;
- master_personal_data – ФИО мастера;
- schedule – расписание работы мастера;
- reviews – отзывы о процедурах, проведенных мастером.

3.2. Индексация документов

ES - это документоориентированная база данных, в которой все объекты должны быть денормализованы, а потом проиндексированы. При таком подходе производительность будет высокой, но изменение данных – затруднительным.

Индексация документов в Elasticsearch – это внесение документов в индексы. Для того, чтобы организовать анализ текста был осуществлен маппинг (mapping - создание схемы данных, только в терминологии стека ELK). Также были указаны поля документов, которые будут анализироваться в дальнейшем. Помимо этого, в работе применяется анализатор, отличительной особенностью которого является улучшенный полнотекстовый поиск по документам [10].

Анализатор представляет собой совокупность следующих частей:

- символьный фильтр;
- токенизатор;
- фильтра токенов.

На рисунке 11 продемонстрирована структурная схема анализатора.



Рисунок 11 – Структурная схема анализатора

По схеме видно, что в символьный фильтр подается поток строк. В нем могут быть удалены, добавлены, либо изменены некоторые символы. Далее обработанные строки поступают на вход токенизатора. Токенизатор разбивает полученные строки на различные токены (слова). Далее полученные токены попадают на вход фильтра токенов, после чего подаются на выход.

Оба JSON-документа Процедура и Специалист были помещены в соответствующие индексы - "salon_procedure" и "salon_master".

Для индексов salon_procedure и salon_master был использован анализатор, код которого представлен на рисунке 12.

```

Salon_Settings = {
  "analysis" : {
    "filter": {
      "russian_stop_words": {
        "type": "stop",
        "stopwords": "_russian_"
      },
      "filter_ru_sn": {
        "type": "snowball",
        "language": "Russian"
      }
    },
    "analyzer": {
      "analitic_for_ru": {
        "type": "custom",
        "tokenizer": "standard",
        "filter": [
          "lowercase",
          "russian_stop_words",
          "filter_ru_sn"
        ]
      }
    }
  }
}

```

Рисунок 12 – Код анализатора для индексации документов

По рисунку видно, что анализатор включает в себя токинезатор standard, который выкидывает запятые, точки и другие знаки пунктуации. в данном случае он работает с русским текстом.

Изначально все токены переходят в нижний регистр благодаря фильтру lowercase.

Далее данный анализатор применяет следующие фильтры токенов: russian_stop_words и snowball.

Благодаря фильтру russian_stop_words из строк удаляются различные стоп слова, такие как союзы, частицы и предлоги. Фильтр токенов filter_ru_sn с типом snowball позволяет выделить основу слов и отбросить окончание. Такая функция называется стемминг (stamming).

Для применения процедуры маппинга в коде были указаны названия полей для документов, их типы данных, а также осуществлена настройка анализатора относительно этих полей. Ниже приведен код маппинга для индекса `salon_procedure`:

```
ProcedureMapping = {
  "properties":{
    "id_of_client": {
      "type": "text",
      "fielddata": True
    },
    "client_age": {
      "type": "integer"
    },
    "client_personal_data":{
      "type": "text",
      "analyzer":"analitic_for_ru",
      "fielddata": True
    },
    "id_of_procedure": {
      "type": "text",
      "fielddata": True
    },
    "date_of_procedure": {
      "type": "date",
      "format": "yyyy-MM-dd"
    },
    "price": {
      "type": "integer"
    },
    "id_of_specialist": {
      "type": "text",
      "fielddata": True
    },
    "service": {
      "type": "text",
      "fielddata": True,
      "analyzer":"analitic_for_ru",
      "search_analyzer":"analitic_for_ru"
    },
    "cosmetics": {
      "type": "text",
```

```

        "analyzer":"analitic_for_ru",
        "fielddata": True,
    }
}

```

Далее приведен код маппинга для индекса `salon_master`:

```

MasterMapping = {
    "properties":{
        "specialisation": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"

        },
        "work_experience": {
            "type": "integer"
        },
        "master_personal_data": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"

        },
        "schedule": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"

        },
        "reviews": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"

        }
    }
}

```


В таблице 1 показаны типы данных, применяемые в маппингах индексов.

Таблица 1 – Типы данных для индексов

Тип данных	Описание типа данных
date	дата в заданном формате
integer	целое число
keyword	ключевые слова, структурированные данные к которым не применяется анализатор
text	неструктурированный текст, который может быть обработан анализатором

Для того, чтобы проиндексировать документы, была написана программа на языке Python. Далее представлен алгоритм работы программы:

- 1) Импортируется библиотека ES, при помощи которой и происходит подключение к имеющемуся кластеру Elasticsearch;
- 2) Производится проверка индексов. Если индексов не обнаружено, то они создаются, если обнаружены, то удаляются и создаются заново;
- 3) Создается анализатор и указываются его настройки;
- 4) Закрываются индексы, обновляются настройки анализатора;
- 5) Открываются индексы и описываются поля для маппинга;
- 6) Добавление маппинга к индексам;
- 7) Открытие JSON-файлов;
- 8) Заполнение индексов. При возникновении исключения, индексация прекращается и выводится информация об исключении.

На рисунке 13 продемонстрирована схема алгоритма программы создания индексов с маппингом и анализатором, а также индексации JSON-документов. Код данной программы размещен в приложении Б.

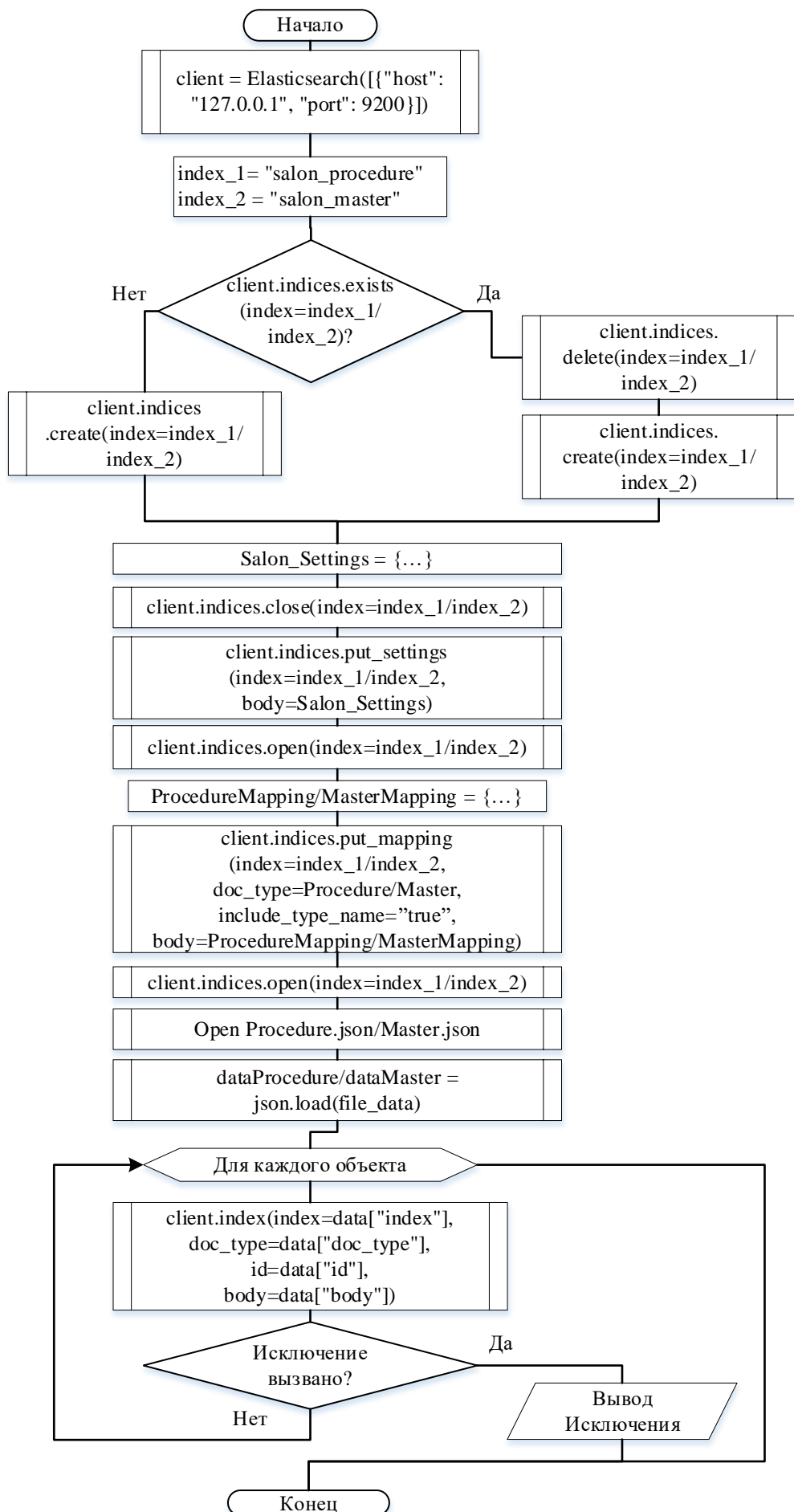


Рисунок 13 – схема алгоритма маппинга и индексации

Результат отработанной программы представлен на рисунке 14.

```
hadoopuser@elena-VirtualBox: ~/scripts
hadoopuser@elena-VirtualBox:~$ cd scripts
hadoopuser@elena-VirtualBox:~/scripts$ python3 mapp_index.py
Recreate salon_procedure index
Recreate salon_master index
Procedure_indexed
Master_indexed
hadoopuser@elena-VirtualBox:~/scripts$
```

Рисунок 14 – вывод программы маппинга и индексации

Чтобы просмотреть созданные индексы, их маппинг и анализатор необходимо перейти по следующим ссылкам: «http://localhost:9200/salon_procedure» и «http://localhost:9200/salon_master». Результат перехода по указанным адресам представлен на рисунке 15.

← → ↺ 127.0.0.1:9200/salon_procedure	← → ↺ 127.0.0.1:9200/salon_master
JSON Необработанные данные Заголовки	JSON Необработанные данные Заголовки
Сохранить Копировать Свернуть все Развернуть все 🔍 Поиск в JSON	Сохранить Копировать Свернуть все Развернуть все 🔍 Поиск в JSON
▼ salon_procedure: aliases: {} ▼ mappings: ▼ properties: ▼ client_age: type: "integer" ▼ client_personal_data: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ cosmetics: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ date_of_procedure: type: "date" format: "yyyy-MM-dd" ▼ id_of_client: type: "text" fielddata: true ▼ id_of_procedure: type: "text" fielddata: true ▼ id_of_specialist: type: "text" fielddata: true ▼ price: type: "integer" ▼ service: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ settings: ▼ index: ▼ routing: ▼ allocation: ▼ include: _tier_preference: "data_content" number_of_shards: "1"	▼ salon_master: aliases: {} ▼ mappings: ▼ properties: ▼ master_personal_data: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ reviews: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ schedule: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ specialisation: type: "text" analyzer: "analitic_for_ru" fielddata: true ▼ work_experience: type: "integer" ▼ settings: ▼ index: ▼ routing: ▼ allocation: ▼ include: _tier_preference: "data_content" number_of_shards: "1" provided_name: "salon_master" creation_date: "1645044608996" ▼ analysis: ▼ filter: ▼ filter_ru_sn: type: "snowball" language: "Russian" ▼ russian_stop_words: type: "stop" stopwords: " russian "

Рисунок 15 – Индексы с маппингом и анализатором

3.3.Запросы с вложенной агрегацией к данным ES

3.3.1. Первый запрос

По заданию выданного варианта необходимо выполнить запрос: разбить процедуры по дате с периодом 1 год, для каждой «корзины» определить суммарную стоимость по каждому специалисту.

Данный запрос был реализован следующим образом:

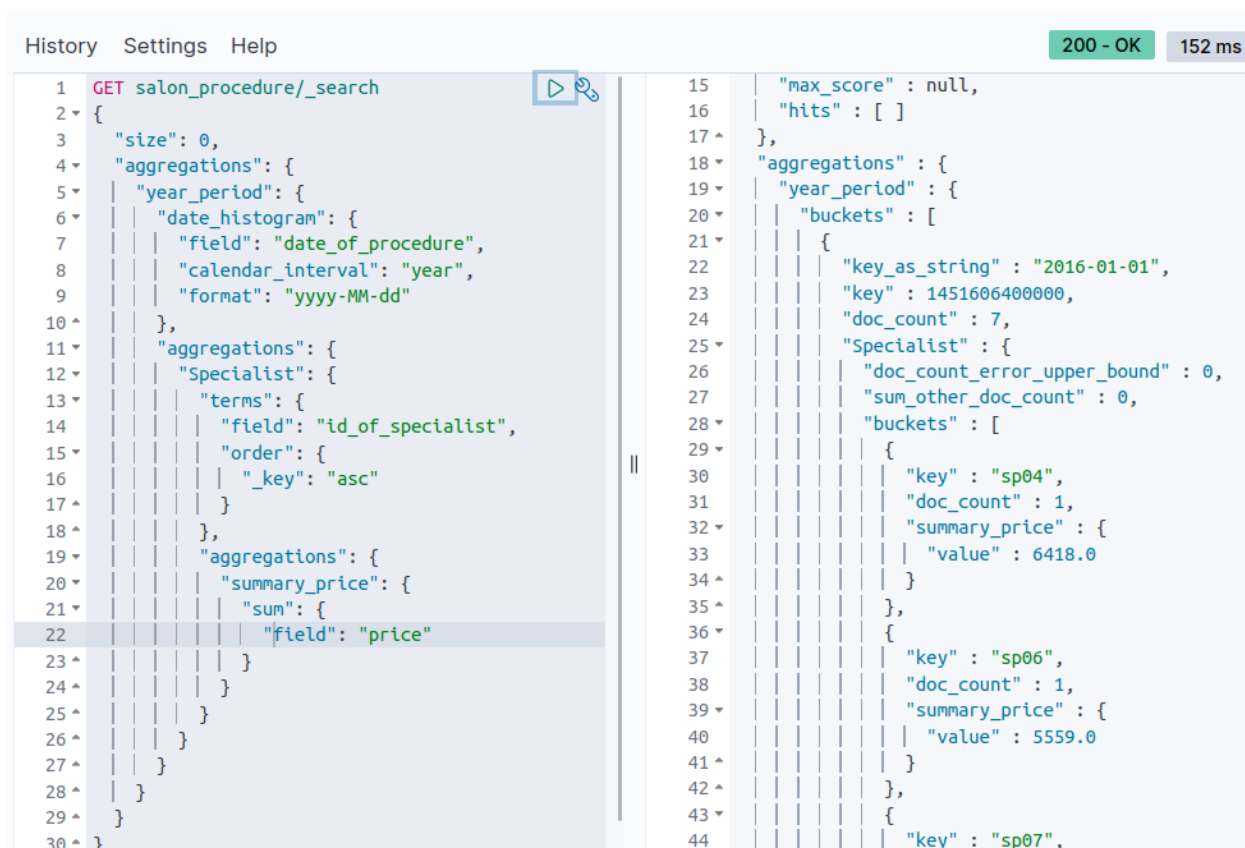
```
GET salon_procedure/_search
{
  "size": 0,
  "aggregations": {
    "year_period": {
      "date_histogram": {
        "field": "date_of_procedure",
        "calendar_interval": "year",
        "format": "yyyy-MM-dd"
      },
      "aggregations": {
        "Specialist": {
          "terms": {
            "field": "id_of_specialist",
            "order": {
              "_key": "asc"
            }
          },
          "aggregations": {
            "summary_price": {
              "sum": {
                "field": "price"
              }
            }
          }
        }
      }
    }
  }
}
```

В данном запросе применяется вложенная агрегация, где каждое вложение помечается как "aggregations". На первом уровне все документы процедур

распределяются по годам при помощи временной гистограммы "date_histogram".

На следующем уровне для каждого года осуществляется группировка по специалисту, для чего задействовано поле "id_of_specialist".

На конечном этапе агрегации определяется цена сделанных процедур по каждому специалисту. Для этого используется гистограмма "sum" по полю "price". На рисунке 16 выведен результат сформированного запроса в среде Kibana. Полный вывод запроса расположен в приложении А.



The screenshot displays the Kibana search console interface. At the top, there are tabs for 'History', 'Settings', and 'Help'. On the right side, there are two status indicators: '200 - OK' in a green box and '152 ms' in a grey box. The main area is split into two panels. The left panel shows the search query in JSON format, with line numbers 1 through 30 on the left margin. The query is a GET request to 'salon_procedure/_search' with a body containing 'size': 0, 'aggregations' with a 'date_histogram' and a nested 'aggregations' block for 'Specialist' (with 'terms' and 'order' sub-aggregations) and a 'sum' sub-aggregation for 'price'. The right panel shows the JSON response, with line numbers 15 through 44 on the left margin. The response includes 'max_score': null, 'hits': [], and a detailed 'aggregations' block. The 'date_histogram' part is partially visible, showing a bucket for '2016-01-01' with a 'doc_count' of 7. The 'Specialist' aggregation shows two buckets: 'sp04' with a 'doc_count' of 1 and a 'summary_price' of 6418.0, and 'sp06' with a 'doc_count' of 1 and a 'summary_price' of 5559.0. The 'sum' aggregation for 'price' is also visible, showing a value of 6418.0.

```
1 GET salon_procedure/_search
2 {
3   "size": 0,
4   "aggregations": {
5     "year_period": {
6       "date_histogram": {
7         "field": "date_of_procedure",
8         "calendar_interval": "year",
9         "format": "yyyy-MM-dd"
10      },
11      "aggregations": {
12        "Specialist": {
13          "terms": {
14            "field": "id_of_specialist",
15            "order": {
16              "_key": "asc"
17            }
18          },
19          "aggregations": {
20            "summary_price": {
21              "sum": {
22                "field": "price"
23              }
24            }
25          }
26        }
27      }
28    }
29  }
30 }
```

```
15   "max_score" : null,
16   "hits" : [ ]
17 },
18   "aggregations" : {
19     "year_period" : {
20       "buckets" : [
21         {
22           "key_as_string" : "2016-01-01",
23           "key" : 1451606400000,
24           "doc_count" : 7,
25           "Specialist" : {
26             "doc_count_error_upper_bound" : 0,
27             "sum_other_doc_count" : 0,
28             "buckets" : [
29               {
30                 "key" : "sp04",
31                 "doc_count" : 1,
32                 "summary_price" : {
33                   "value" : 6418.0
34                 }
35               },
36               {
37                 "key" : "sp06",
38                 "doc_count" : 1,
39                 "summary_price" : {
40                   "value" : 5559.0
41                 }
42               }
43             ]
44           }
45         }
46       ]
47     }
48   }
49 }
```

Рисунок 16 – Результат выполнения первого запроса

3.3.2. Второй запрос

По заданию выданного варианта необходимо выполнить следующее: предложить признаки отрицательного отзыва; вывести специалистов хотя бы с одним отрицательным отзывом.

В качестве признаков отрицательного отзыва были взяты следующие слова: кидает, пострадало, расстроенный, наглость, опоздал, помойка, грязно, ужасно, плохо, печально.

Для поиска приведенных выше слов в отзывах использовалась агрегация на основе "simple_query_string". В результате выводятся персональные данные специалистов с плохими отзывами и сами отзывы (в отзывах выделены найденные слова благодаря функции "highlight". Ниже представлен реализованный запрос.

```
GET salon_master/_search
{
  "query": {
    "simple_query_string": {
      "query": "кидает пострадало расстроенный наглость
опоздал помойка грязно ужасно плохо печально"
    }
  },
  "_source": [
    "master_personal_data"
  ],
  "highlight": {
    "fields": {
      "reviews": {}
    }
  }
}
```

На рисунке 17 продемонстрирован результат запроса в среде Kibana. Полный вывод запроса расположен в приложении А.

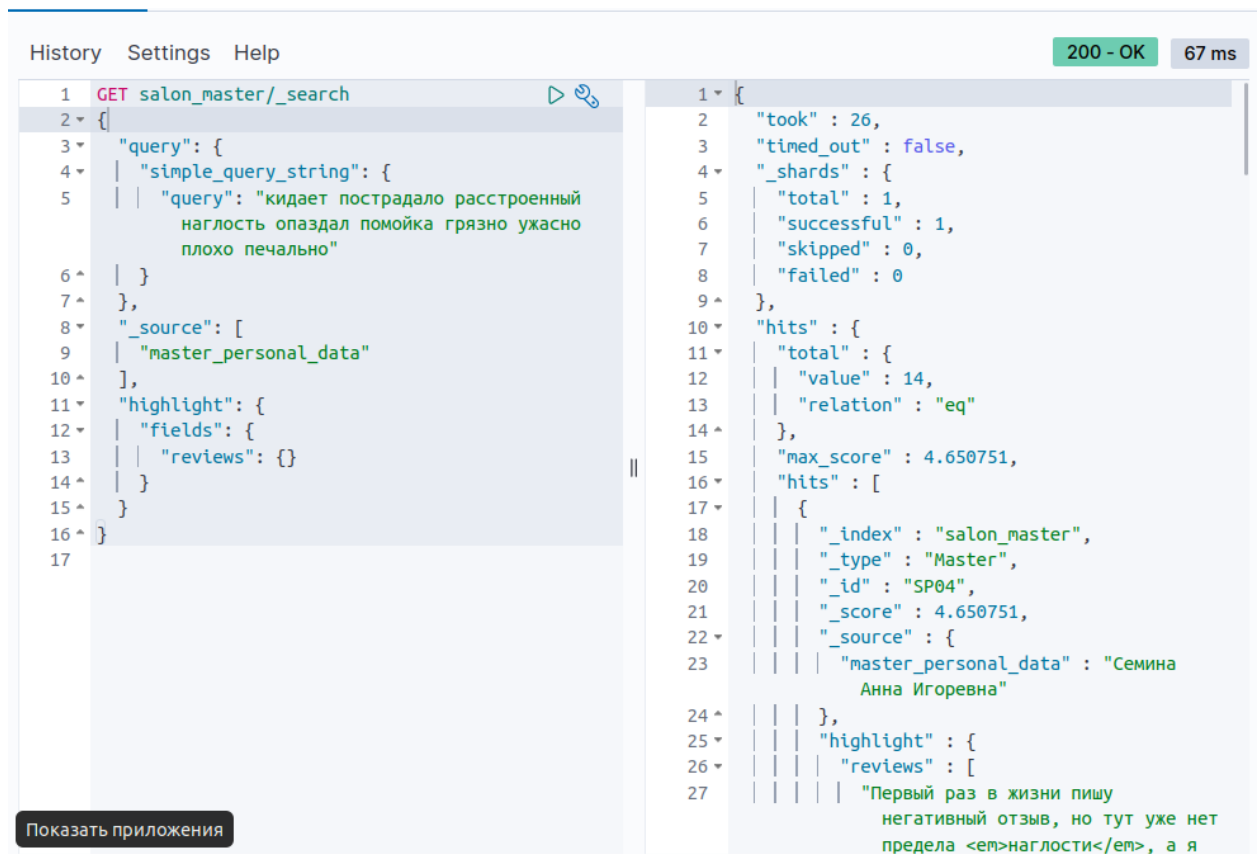


Рисунок 17 – Результат выполнения второго запроса

3.3.3 Визуализация запросов в среде Kibana

В данном курсовом проекте используется версия Kibana – 7.10.1. Стоит отметить, что версии Kibana можно разделить на две группы – в первую входят все, выпущенные до Kibana-7.0, остальные после. Несмотря на то, что речь идет об одном сервисе, платформа Kibana получила глобальную навигацию и полный редизайн интерфейса при выпуске нового релиза Elastic Stack 7.0 [11]. Таким образом, алгоритм визуализации запросов в более ранних версиях может значительно отличаться от представленного ниже.

Для того, чтобы визуализировать запросы в среде Kibana необходимо на главной странице (<http://127.0.0.1:5601>) открыть пункт меню. В открывшемся списке раскрыть более детально заголовок Kibana. Среди появившихся подзаголовков выбрать Visualize. На рисунке 18 продемонстрированы все описанные действия.

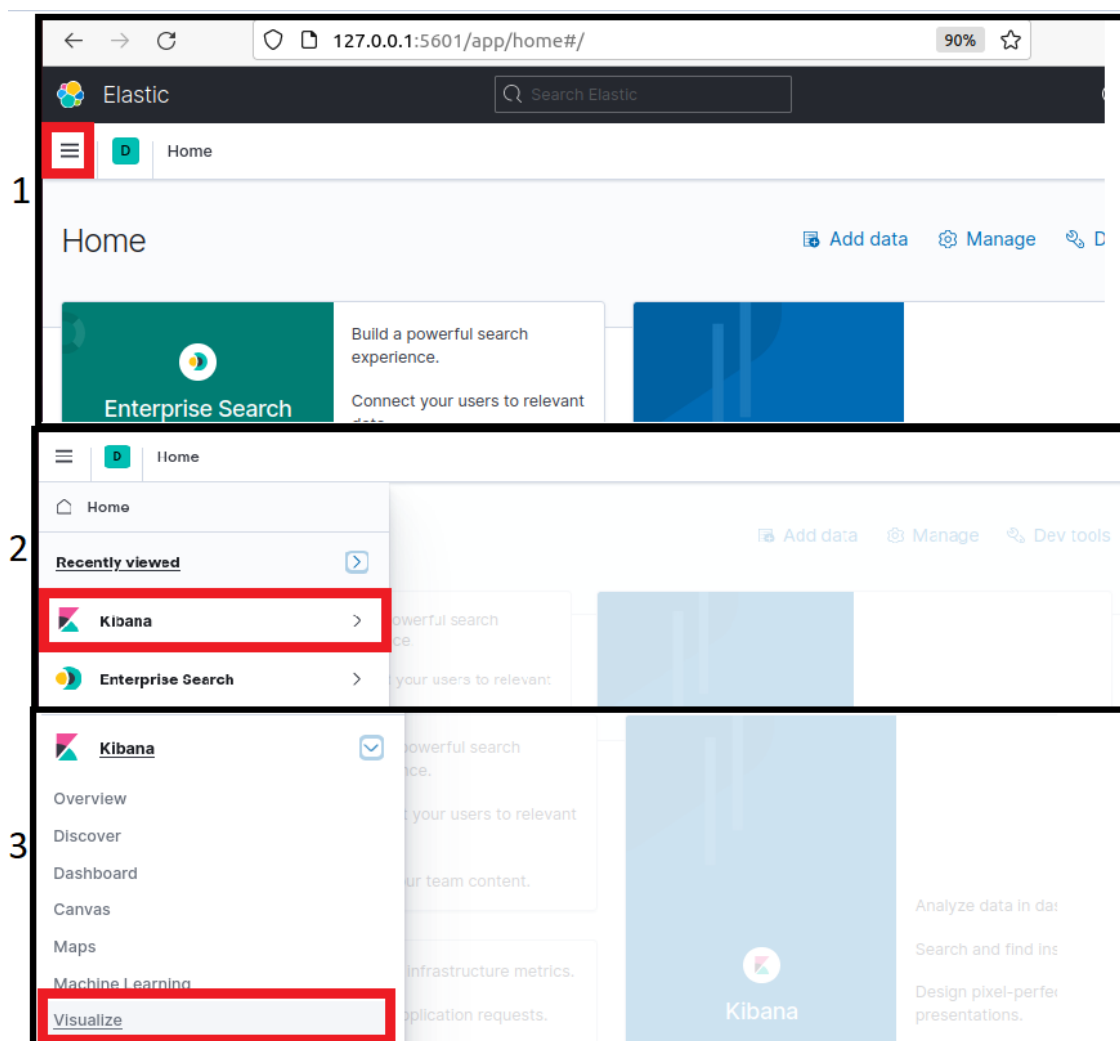


Рисунок 18 – открытие инструментов визуализации

Далее необходимо создать новую визуализацию – “+ Create visualization” (рисунок 19), либо выбрать из списка имеющуюся для демонстрации, либо продолжения работы.

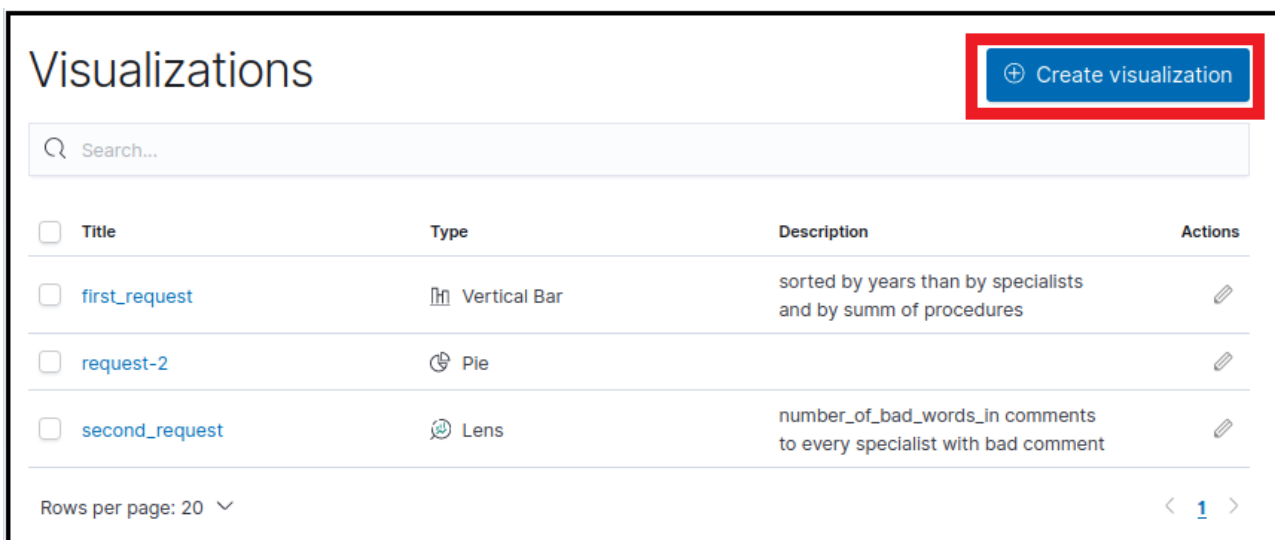


Рисунок 18 – создание нового документа визуализации

При создании нового документа высветится окно с выбором типа диаграммы (столбчатая, круговая, табличная форма, карта, ломаная и другое). Данное окно представлено на рисунке 19. Наиболее подходящей под первый запрос является “Vertical Bar”. Однако для начала работы следует выбрать “Lens”, поскольку этот тип подразумевает интуитивное создание диаграмм – оно является наиболее гибким, понятным и простым в составлении.

New Visualization

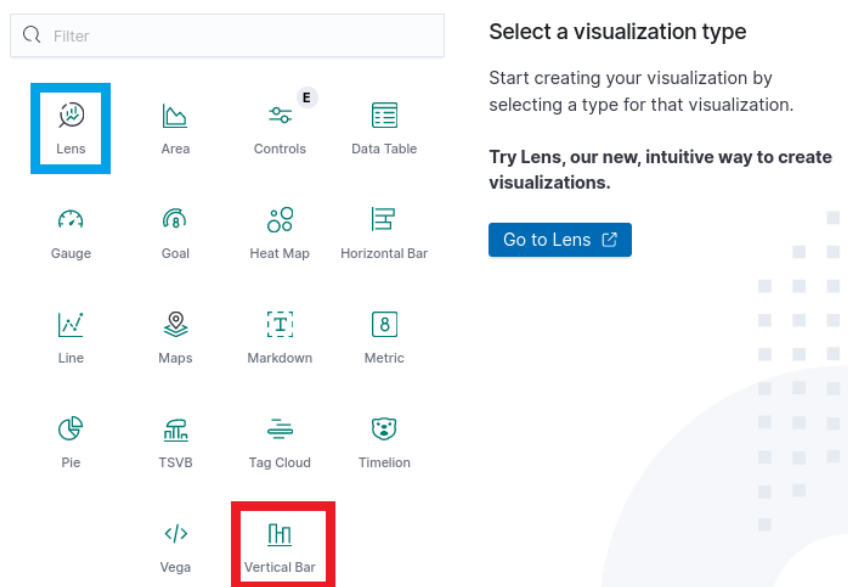


Рисунок 19 – окно с выбором типа диаграммы

Для визуализации первого запроса мною была выбрана столбчатая вертикальная диаграмма. Далее был выбран индекс `salon_procedure`, на основе которого делается визуализация (рисунок 20).

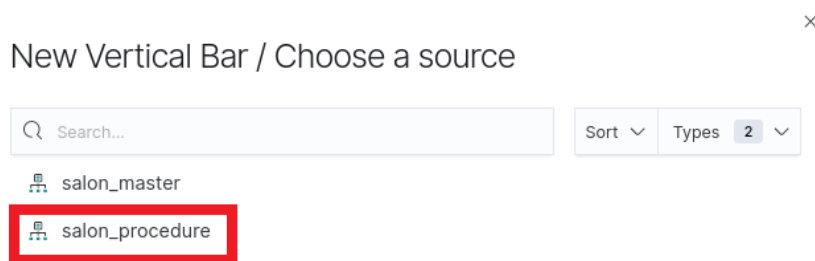


Рисунок 19 – окно с выбором индекса

После выбора индекса открылось окно непосредственно для визуализации данных. В данном окне можно выделить несколько областей:

1 – область запроса/фильтра – в ней необходимо сделать запрос к документам индекса по одному или нескольким из имеющихся полей. Если оставить это поле пустым, то никакие данные не будут участвовать в визуализации, и соответственно графиков не будет. Если необходимо, чтобы в визуализации принимали участие все документы, как это нужно в нашем случае, то следует выбрать любое поле и поставить * через двоеточие (`id_of_specialist:*`). Также через двоеточие можно ставить и другие условия фильтрации. Например, если необходимо чтобы для построения графиков использовались только те документы, `id` которых равны строго заданному значению или чтобы поле цена было выше или ниже указанного значения. Таким образом происходит начальная фильтрация данных.

2 – Область времени – указывается промежуток времени, за который берутся данные. Поскольку сейчас 2022 год, а при генерации JSON-файла Процедура был выбран диапазон с 2016 по 2021 год, то выбираем диапазон – за последние 7 лет (`Last 7 years`). Также в данном поле можно выбрать промежуток по месяцам, неделям, дням и часам.

3 – Область настройки графика – в данной области указываются поля, которые будут соответствовать оси X и оси Y.

4 – Область визуализации – в ней выводится график по заданным значениям.

5 – кнопка для сохранения графика.

6 – кнопка обновления данных после внесения изменения параметров.

На рисунке 20 представлено окно визуализации с обозначением всех используемых областей.

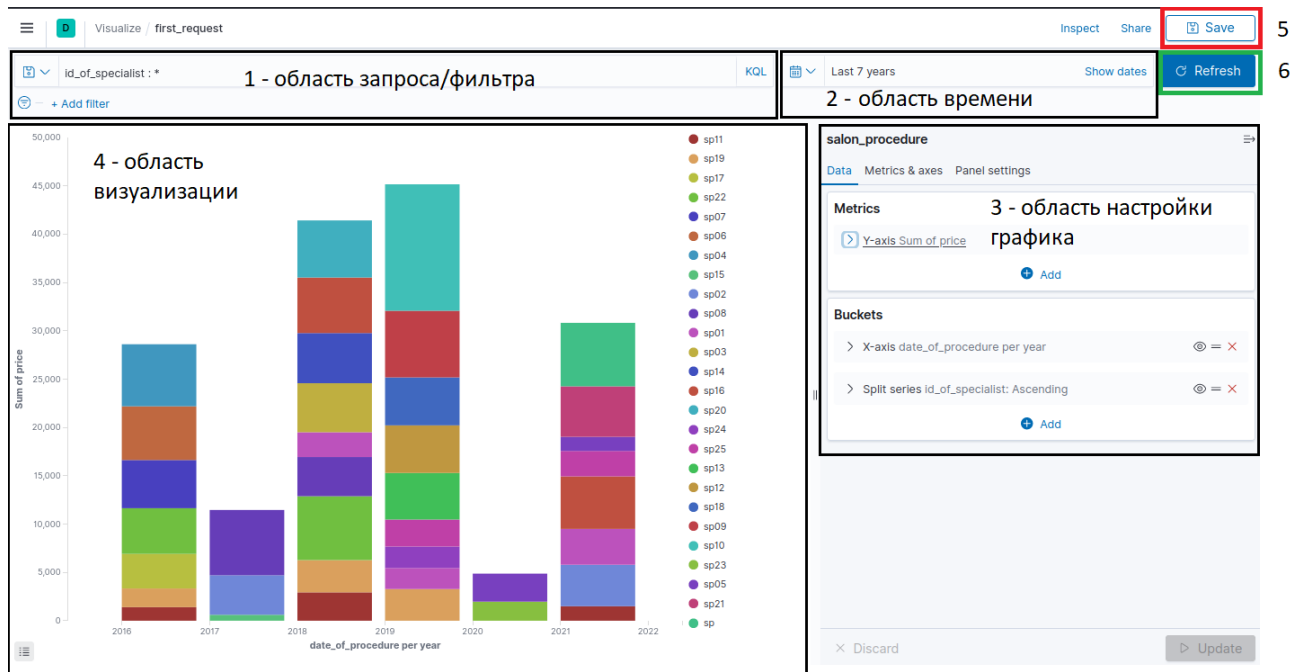


Рисунок 20 – окно визуализации данных

Разберем более подробно саму визуализацию запроса. По заданию необходимо разбить процедуры по дате с периодом 1 год, для каждой «корзины» определить суммарную стоимость по каждому специалисту.

Сначала необходимо разбить документы корзинам (Buckets) в соответствии с годом. Для этого во вкладке Buckets нажимаем Add и выбираем X-axis, поскольку разделение будет идти по оси X. Далее выбираем тип агрегации. Аналогично построенному ранее запросу было выбрано “Date Histogram”. Поле date_of_procedure подставилось автоматически, поскольку это единственное поле в индексе, тип которого – дата. Далее выбирается интервал разбиения – в нашем случае год (year).

Следующим уровнем агрегации в запросе было разделение по специалистам. По аналогии добавим ещё одно разделение в Buckets. На этот раз выбираем Split series. В качестве sub aggregation, как и в запросе, выбираем Terms, а в качестве поля – id_of_specialist. Далее выбираем порядок

расположения специалистов по столбцам – в соответствии с суммой, по возрастанию (ascending). Указываем максимальный размер специалистов в одном столбце (одном году) – 25, поскольку у нас всего 25 специалистов. Если стоит задача выделить 5 наиболее прибыльных по каждому году, то можно указать 5.

Последним этапом необходимо настроить ось Y. Точно также выбираем Add. Поскольку в задаче необходимо подсчитать сумму, то в качестве aggregation выбираем sum, а в качестве поля – price. Все сделанные настройки представлены на рисунке 21.

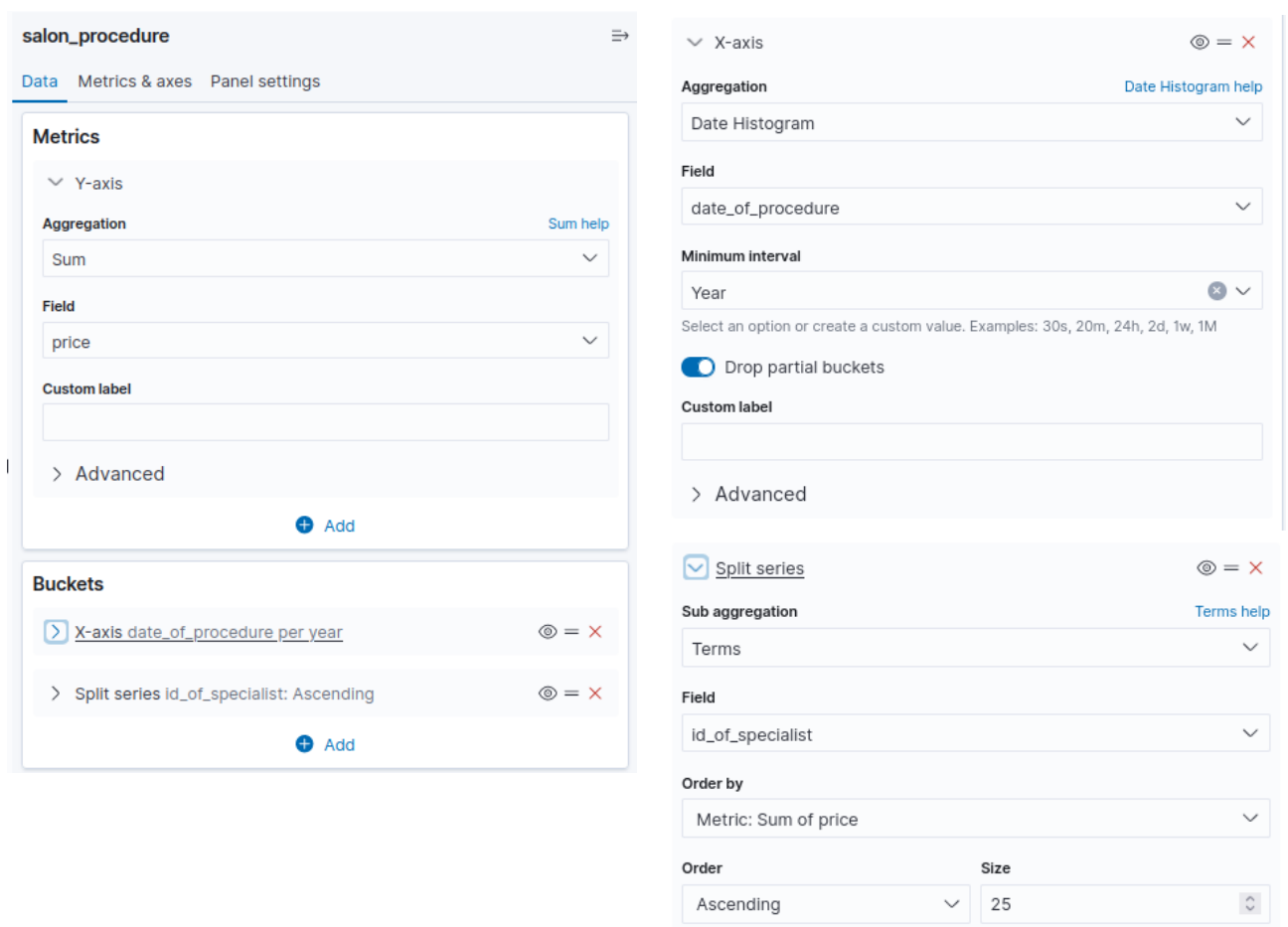


Рисунок 21 – настройка графика

Таким образом, получилась визуализация первого запроса, которая представлена на рисунке 22. По полученной диаграмме можно наглядно посмотреть наиболее прибыльные года для салона красоты, просмотреть заработок каждого специалиста по годам, а также выделить наиболее успешных специалистов по каждому году.

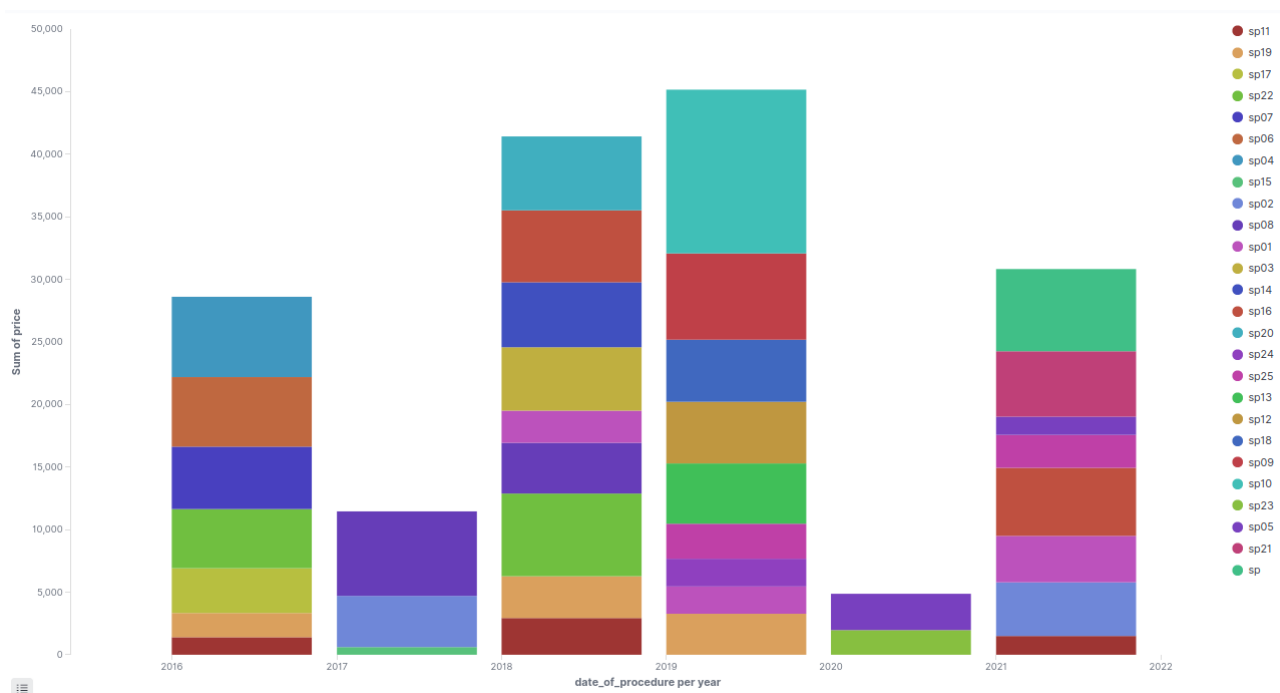


Рисунок 21 – визуализация первого запроса

После успешной реализации графика, он был сохранен при помощи кнопки Save (см. рисунок 20). Если этого не сделать, то все настройки будут потеряны.

Для визуализации второго запроса применялся шаблон Lens. Как было сказано выше данный шаблон является более гибким, а поэтому больше подходит для второго запроса, в котором необходимо отобрать специалистов с негативными отзывами.

Для начала сделаем фильтр, в котором будут отбираться документы только с негативными отзывами:

reviews: "кидает" or "пострадало" or "расстроенный" or "наглость" or "опоздал" or "помойка" or "грязно" or "ужасно" or "плохо" or "печально".

Далее, как и в первом варианте, был выбран промежуток – последние 7 лет. По оси X будут расположены специалисты, а точнее их id, поэтому в разделе X-axis configuration было выбрано “Top values” (Также среди возможных вариантов были интервал и фильтр, но по оси X нет необходимости в их использовании). Далее было выбрано поле, которому будут выводиться документы - _id. Количество возможных записей было поставлено равным 25. Далее был заполнен порядок следования. По оси Y имеется возможность выбрать функцию агрегации – в данном случае была выбрана функция count, а в качестве боля

выбрано – records. Такой выбор параметров позволит наглядно показать сколько негативных слов встречается в отзывах каждого из мастеров. И завершающим этапом является выделение негативных слов в отдельные записи – отдельные фильтры. Для этого во вкладке Break down by были выбраны Filters и при помощи кнопки Add a filter были добавлены фильтры по принципу « = reviews: признак негативного отзыва». Все настройки представлены на рисунке 22.

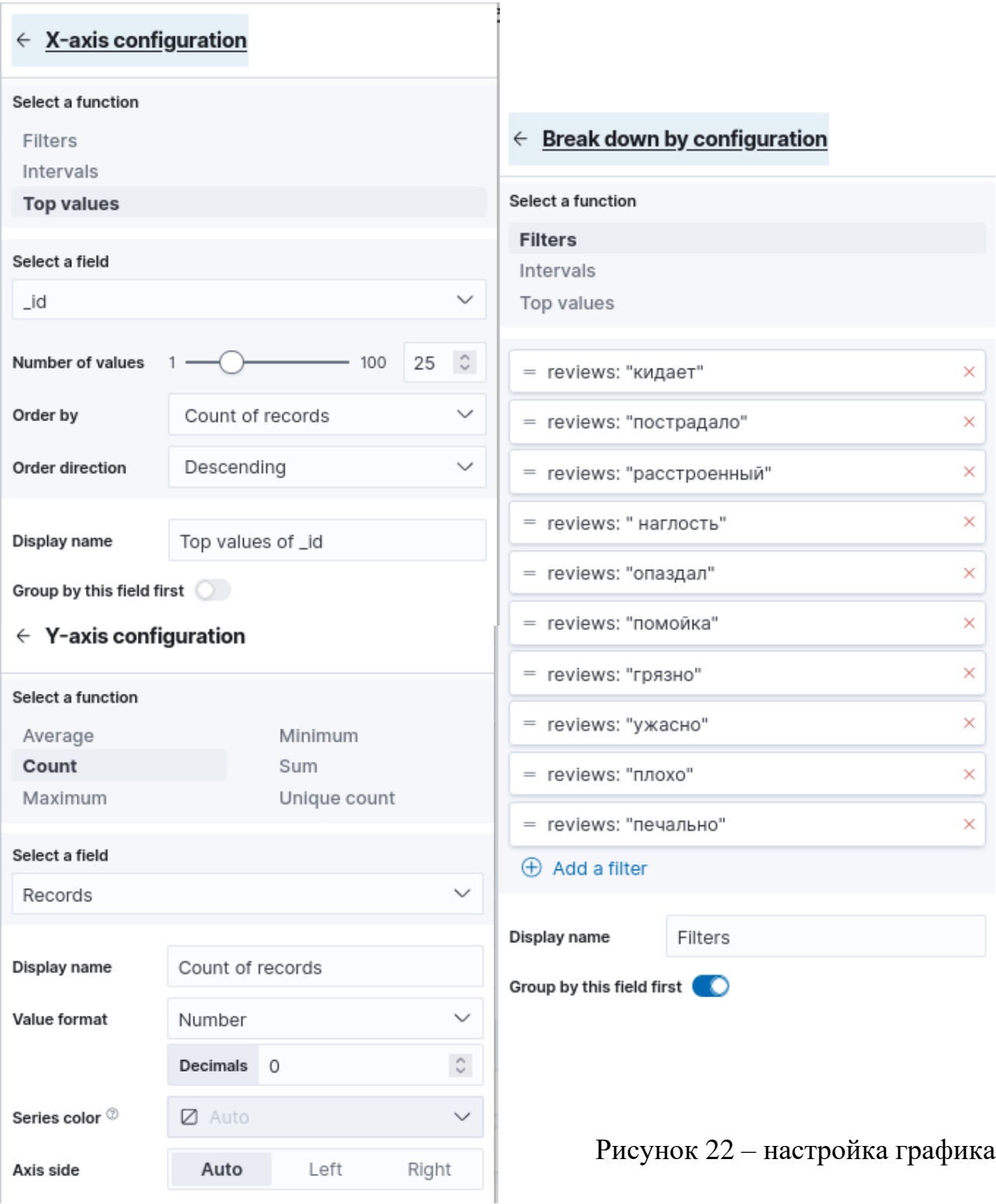


Рисунок 22 – настройка графика

На рисунке 23 представлена графическая визуализация второго запроса.

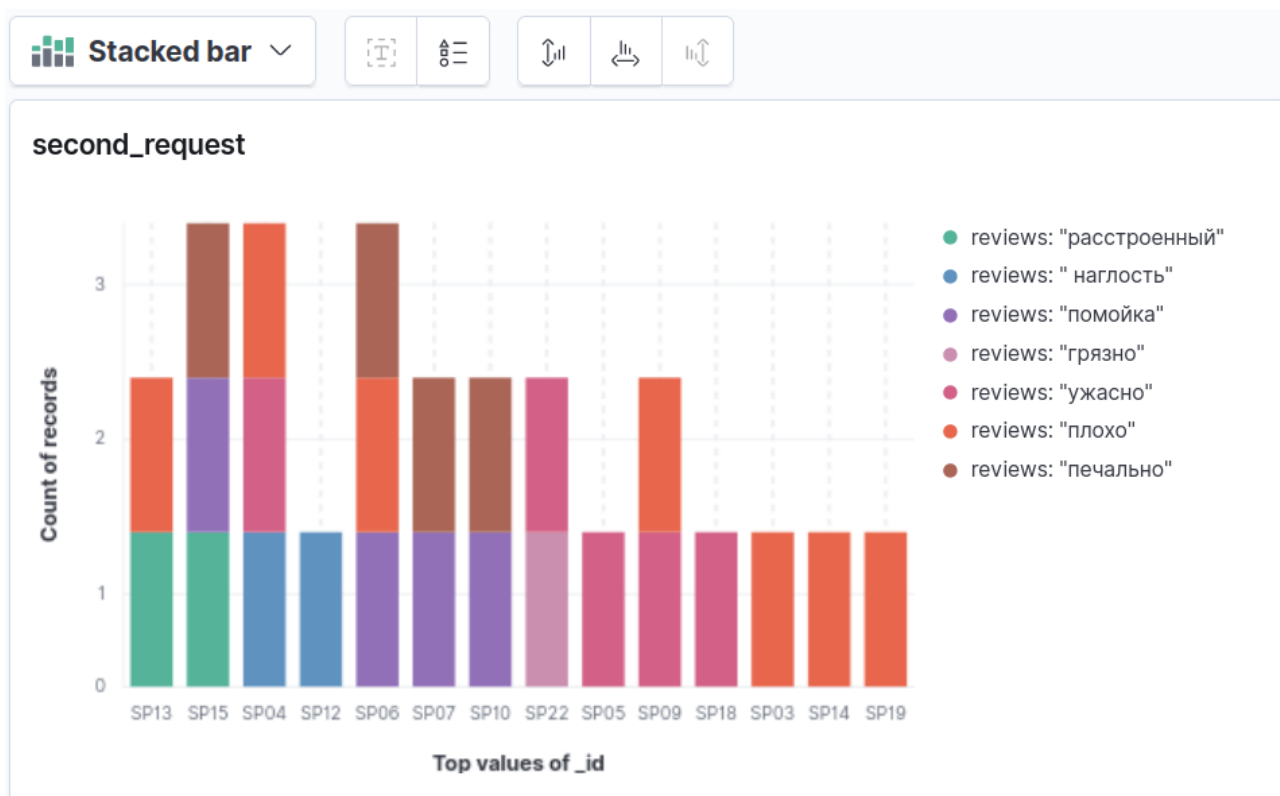


Рисунок 23 – визуализация второго запроса

Neo4j

3.4. Создание и заполнение графовой базы данных

Прежде чем заполнить графовую БД Neo4j сначала необходимо получить документы из ES. После получения данных были созданы сами узлы.

Итоговая графовая БД имеет представленные ниже сущности:

- Узел: «Client» (Клиент): «Client_id» (идентификатор посетителя), «Client_personal_data» (данные о клиенте) и «Client_age» (возраст);
- Узел: «procedure» (Процедура): «Client_id» (идентификатор посетителя), «Cosmetics» (косметика), «Date_of_procedure» (дата процедуры), «Master_id» (идентификатор мастера), «Price» (цена), «Service» (услуги);
- Узел: «Master» (Специалист): «Master_id» (идентификатор мастера), «Master_personal_data» (данные о мастере), «Specialization» (специализация), «Work_experience» (опыт работы).
- Связь: «connect_Client_Master» (Связь с клиентов с мастерами): «Date_of_Procedure» (дата процедуры), «Price» (стоимость), «Service» (услуги);
- Связь: «connect_master» (Связь процедуры с мастером): «Date_of_Procedure» (дата процедуры);
- Связь: «connect_procedure» (Связь клиента с процедурой): «Date_of_Procedure» (дата процедуры).

На языке Python была написана программа для заполнения Neo4j, которая имеет следующий алгоритм:

- 1) Подключение к кластеру ES и клиенту Neo4j;
- 2) Удаление имеющейся базы Neo4j;
- 3) Импорт из кластера данных из ES;
- 4) Создание узла Процедура в основном цикле;
- 5) Проверка на наличие в БД клиента (создание при отсутствии);
- 6) Проверка на наличие в БД мастера (создание при отсутствии) – происходит во вложенном цикле;
- 7) Создание новых связей, если идентификатор мастера во внутреннем цикле совпадает с идентификатором во внешнем.

На рисунке 24 представлена схема алгоритма программы, созданной для заполнения Neo4j. Код программы расположен в приложении Б.

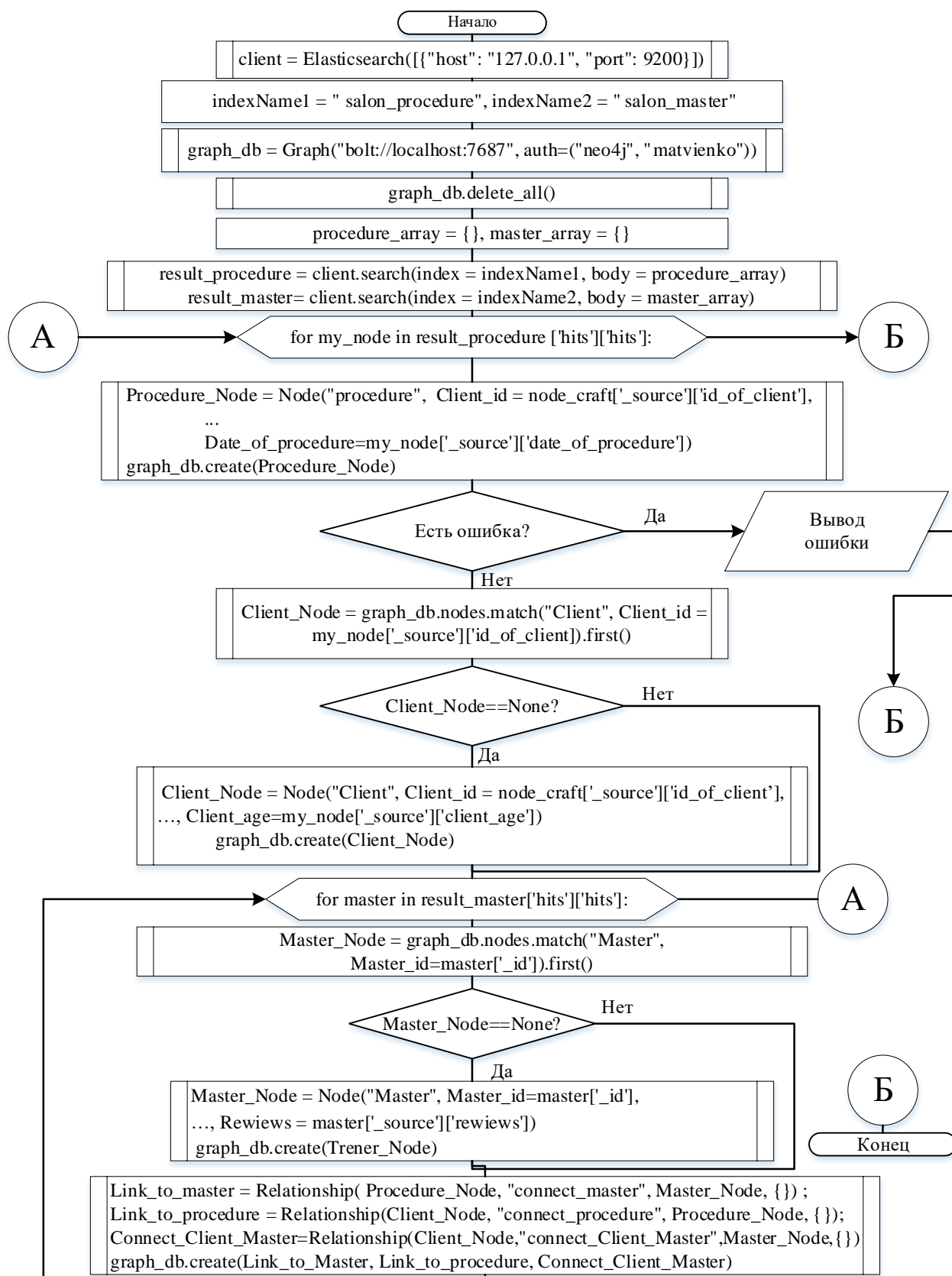


Рисунок 24 – Схема алгоритма заполнения графовой БД Neo4j

Для просмотра полученной БД перейдем по адресу <http://localhost:7474/browser>. Сделаем запрос на языке Cypher, возвращающий связь Клиент - Мастер, которую необходимо было создать по заданию. Запрос выглядит следующим образом: `MATCH p=()-[r:connect_Client_Master]->() RETURN p`. Результат выполнения запроса представлен на рисунке 25. В данном случае узлы типа Мастер представлены оранжевым, а узлы типа Клиент красным цветом.

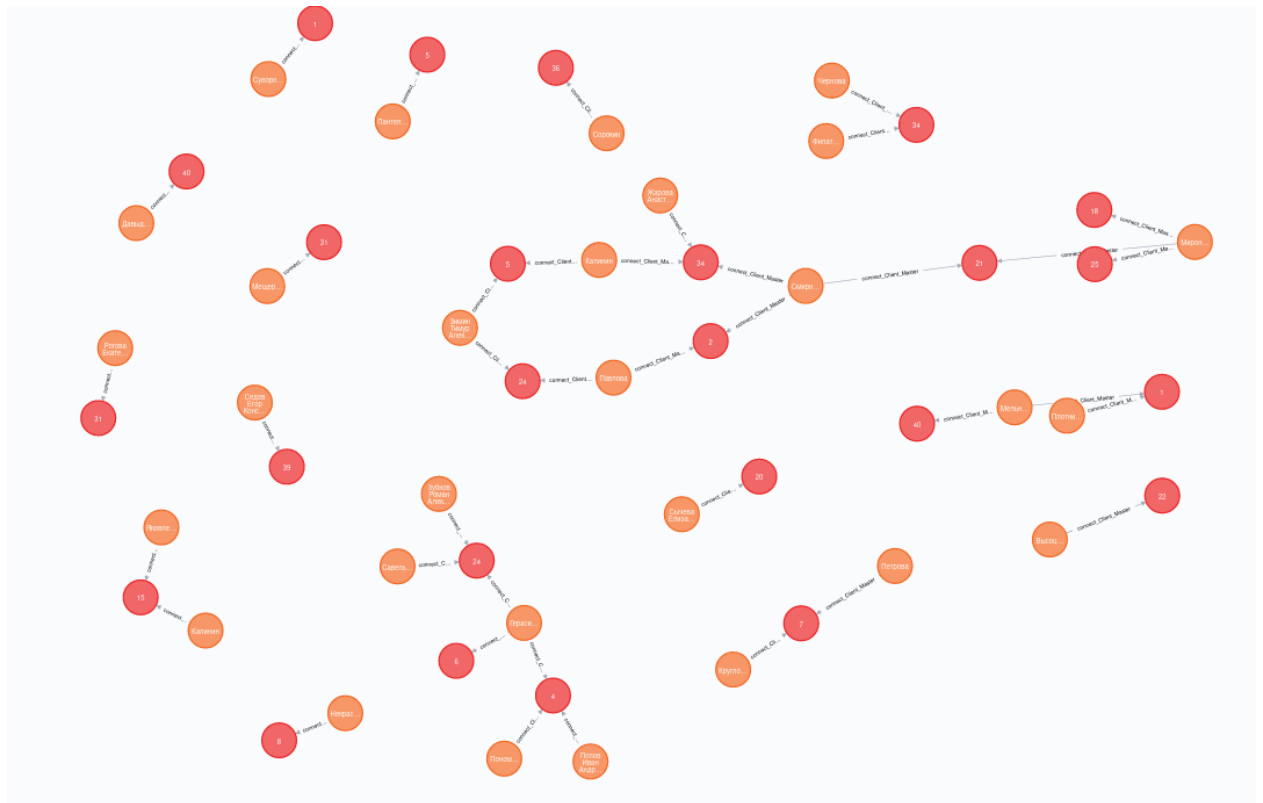


Рисунок 25 – Граф связи Мастер – Клиент БД Neo4j

3.5. Запрос к графовой базе данных

В соответствии с заданием необходимо реализовать запрос: определить специалиста с наибольшей стоимостью выполненных процедур. Запрос на языке Cypher:

```
MATCH p=(c:Client)-[r:connect_Client_Master]->(m:Master)
WITH m, sum(toInteger(r.Price)) as master_sum
ORDER BY master_sum desc
RETURN m, master_sum
LIMIT 1
```

В данном запросе берутся связи по типу Мастер – Клиент. Далее по каждому мастеру (m) рассчитывается суммарная стоимость, указанная в связи. Вывод результатов происходит по убыванию (desc). Задавая ограничение в одну запись, мы получаем мастера с наибольшей стоимостью выполненных процедур.

На рисунке 26 представлен вывод описанного выше запроса.

The screenshot displays the Neo4j Cypher query editor and its results. The query is as follows:

```
1 MATCH p=(c:Client)-[r:connect_Client_Master]->(m:Master)
2 WITH m, sum(toInteger(r.Price)) as master_sum
3 ORDER BY master_sum desc
4 RETURN m, master_sum
5 LIMIT 1
```

The results pane shows a single record with the following data:

m	master_sum
<pre>{ "Work_experience": "2", "Master_id": "SP10", "Specialization": ["Массаж рук", "Спа-маникюр", "Биопилинг кожи головы", "Гальванизация"], "Master_personal_data": "Ильинский Матвей Эмирович" }</pre>	13084

At the bottom of the interface, a status message reads: "Started streaming 1 records after 2 ms and completed after 2 ms."

Рисунок 26 – Результат выполнения запроса к БД Neo4j

4. Spark

4.1. Создание и заполнение таблиц

В соответствии с заданием необходимо создать 3 CSV-файла (с внутренней схемой) таблиц «Клиент», «Процедура», «Специалист» и сохранить их в файловой системе HDFS. Для выполнения задания была написана программа, алгоритм которой представлен ниже:

- 1) Подключение при помощи API к кластеру ES;
- 2) Подключение к сессии Spark;
- 3) Выполнение запросов Elasticsearch для получения всех документов;
- 4) Создание схемы данных, применимой для каждой из таблиц и задание типов полей в данных таблицах;
- 5) Заполнение таблиц при помощи циклов и с использованием преобразования полей документов в соответствии с созданными ранее схемами;
- 6) Создание датафреймов на основе схем и самих данных, полученных в шаге 5;
- 7) Перепись датафреймов в csv-файлы HDFS с указанием параметров, включая заголовки столбцов и разрешения перезаписи.

На рисунке 27 представлена схема алгоритма программы, написанной для создания csv-файлов. Код данной программы расположен в приложении Б.

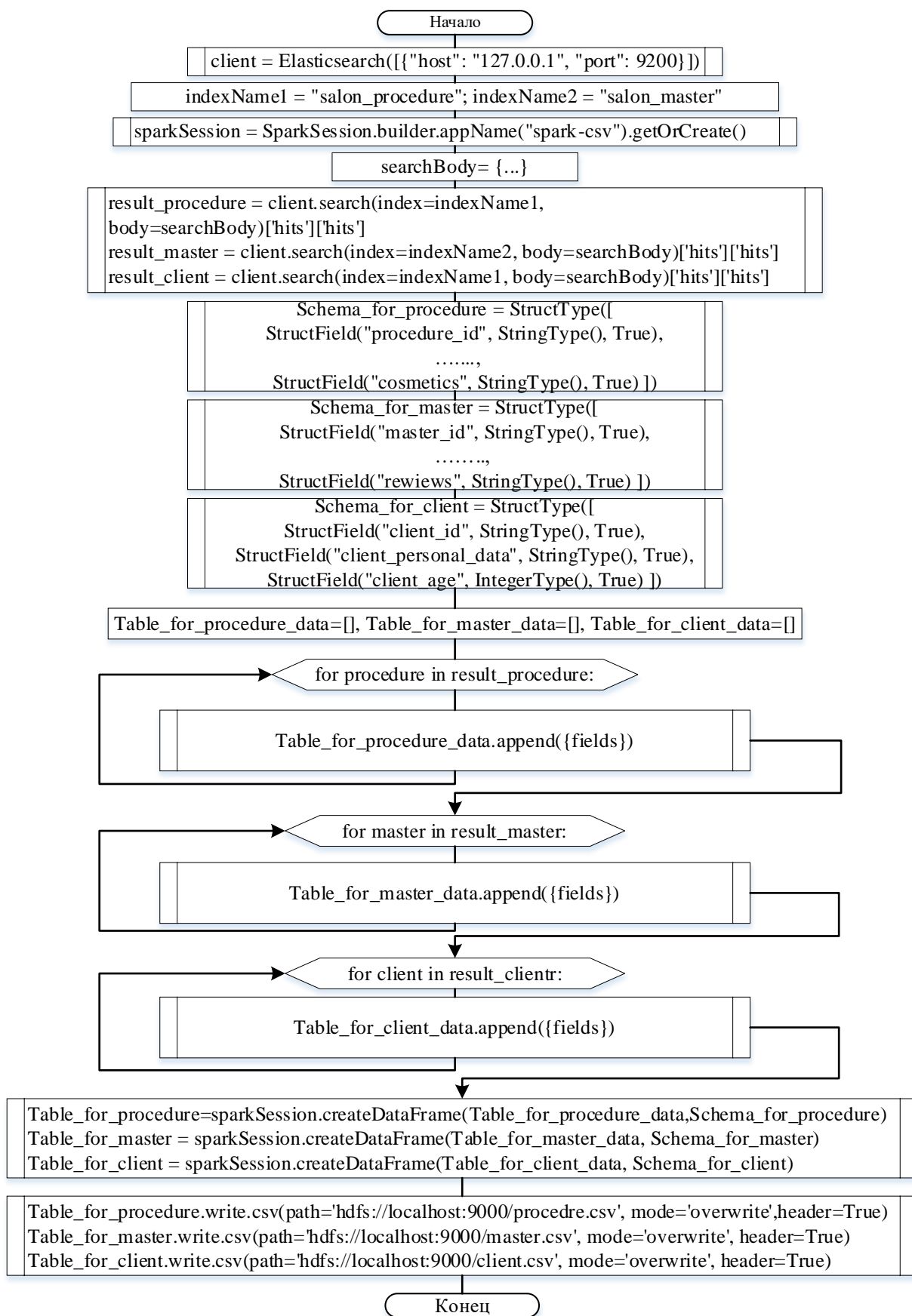


Рисунок 27 – Схема алгоритма программы создания csv-файлов
В итоге выполнения программы были выведены созданные таблицы.

4.2. Запрос в Spark

По заданию требуется реализовать запрос select: определить число выполненных процедур каждым специалистом. На данный момент в Spark присутствуют следующие таблицы «procedure», «client» и «master». Используя данные таблицы построим запрос:

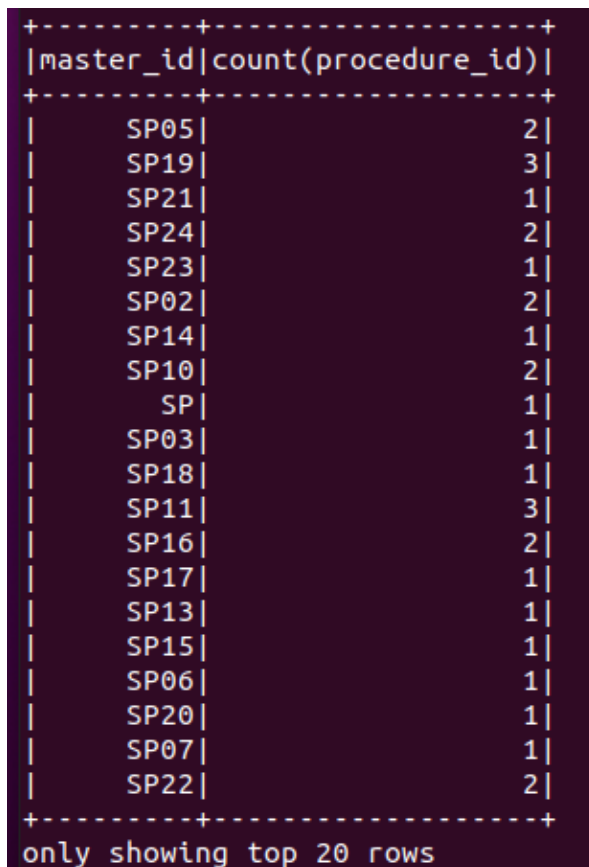
```
SELECT procedure.master_id, COUNT(procedure.id)
FROM procedure
GROUP BY procedure.master_id
```

Чтобы выполнить запрос была написана программа на Python, алгоритм работы которой представлен ниже:

- 1) Подключение к сессии Spark;
- 2) Чтение csv-файла procedure.csv;
- 3) Создание временной таблицы;
- 4) Выполнение SQL-запроса.

Код написанной программы размещен в приложении Б.

На рисунке 28 показан результат запроса.



master_id	count(procedure_id)
SP05	2
SP19	3
SP21	1
SP24	2
SP23	1
SP02	2
SP14	1
SP10	2
SP	1
SP03	1
SP18	1
SP11	3
SP16	2
SP17	1
SP13	1
SP15	1
SP06	1
SP20	1
SP07	1
SP22	2

only showing top 20 rows

Рисунок 28 – Результат выполнения запроса в Spark

4.3. Мониторинг Spark

Для мониторинга данных в Spark был осуществлен переход по адресу <http://127.0.0.1:4040/jobs/>.

При запуске программы с запросом монитор Spark показывает временные характеристики выполнения задач данного запроса. Временная диаграмма седьмого job-а (относящегося непосредственно к запросу) представлена на рисунке 29.

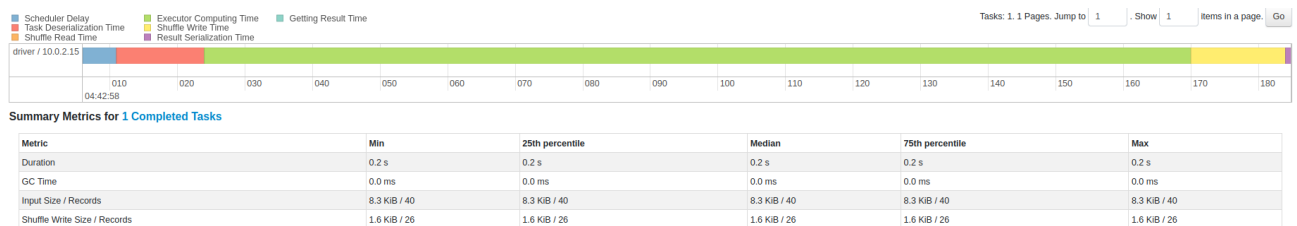


Рисунок 29 – Временная диаграмма выполнения запроса

По временной диаграмме становится понятно, что выполнение запроса заняло 0,2 секунды (Total Tasks Time). Количество выполнимых задач зависит от количества разделов с записями (partitions), для сокращения времени выполнения запроса это число было сокращено с стандартного 200 до 80, что уменьшило время выполнения с 0,5 секунд до 0,2.

Во вкладке SQL имеются данные о всех выполненных запросах. Они представлены на рисунке 30.

ID	Description	Submitted	Duration	Job IDs
6	showString at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:57	0.7 s	[6][7]
5	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:57	8 ms	
4	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:57	1 ms	
3	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:57	18 ms	
2	load at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:56	75 ms	[4]
1	load at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:56	62 ms	[2]
0	load at NativeMethodAccessorImpl.java:0	2022/05/24 04:42:52	3 s	[0]

Рисунок 30 – SQL-запросы сессии в Spark

Основным запросом задания является тот, что имеет ID=6. Разберем его более подробно. Стадии выполнения данного запроса и его временные характеристики представлены в виде DAG (Ориентированного ациклического графа). На рисунке 21 представлен DAG основного SQL-запроса. Физический план выполнения запроса представлен в приложении Б.

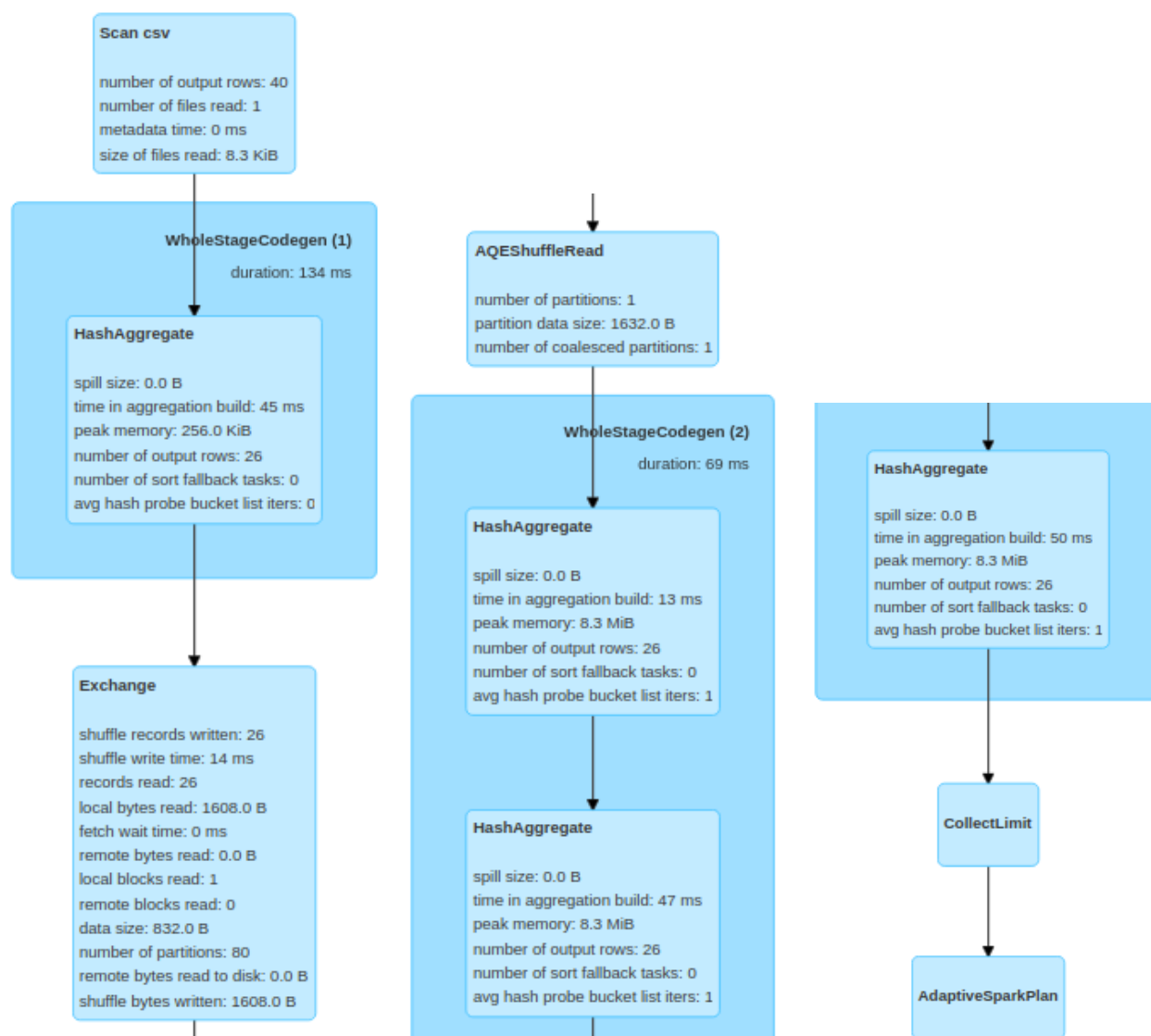


Рисунок 31 – DAG SQL-запроса

Выполнение SQL-запроса начинается с чтения таблицы «процедура» из файла procedure.csv. Было просканировано 40 строк. Далее выполняется HashAggregate. Эта функция выполняет группировку строк по специалистам. Всего выделено 26 человек. Полученный результат поступает в Exchange, где происходит слияние всех данных из разных процессов и разделов.

Далее выполняется повторная оптимизация запроса для выбора лучшей физической стратегии – AQEShuffleRead [12]. Потом в разделе WholeStageCodegen подразделы HashAggregate выполняют группировку записей.

В конце выполнения запроса Collect Limit отбирает первые 20 записей для вывода результата.

ЗАКЛЮЧЕНИЕ

В результате выполнения курсового проекта был создан макет аналитической системы салона красоты. Основой макета стали БД Elasticsearch и Neo4j совместно с сервисами Hadoop и Spark.

Все программные средства были установлены на VM Virtual Box под ОС Ubuntu-20.

Во время выполнения работы были сгенерированы два документа типа Процедура и Специалист. На VM был развернут кластер ES. В данном кластере реализованы два индекса - Salon_Procedure и Salon_Master. Для полученных индексов была реализована программа маппинга с анализатором русского текста. Также благодаря этой программе документы были проиндексированы. В среде Kibana были сформированы и выполнены запросы с вложенной агрегацией в соответствии с заданием. Помимо этого, при помощи средств Kibana имеющиеся запросы были визуализированы в качестве столбчатых диаграмм с разбиением по подгруппам.

В ходе курсового проекта была создана и заполнена графовая БД Neo4j. В данной БД были сформированы узлы "Процедура", "Мастер" и "Клиент", а также связи между всеми этими сущностями. Связь Мастер-Клиент, требуемая по заданию, была продемонстрирована в графическом виде. Опираясь на задание, был сформирован и реализован запрос к данной БД на языке Cypher.

Работая со Spark в файловой системе HDFS были созданы CSV-файлы, содержащие таблицы «Мастер», «Процедура» и «Клиент». После этого был реализован SQL-запрос, требуемый в задании. Выполнение запроса было проанализировано при помощи монитора Spark.

Таким образом, был разработан макет аналитической системы салона красоты с использованием баз данных NoSQL.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Neo4j – краткое руководство. – URL: <https://coderlessons.com/tutorials/bazy-dannykh/uznaite-neo4j/neo4j-kratkoe-rukovodstvo> (дата обращения 13.03.2022)
2. Из чего состоит Hadoop: концептуальная архитектура. – URL: <https://www.bigdataschool.ru/wiki/hadoop> (дата обращения 13.03.2022)
3. Apache Spark. – URL: <https://www.bigdataschool.ru/wiki/spark> (дата обращения 17.03.2022)
4. Elastic Matrix | Elasticsearch. – URL: https://www.elastic.co/support/matrix#matrix_compatibility (дата обращения 17.03.2022)
5. Лабораторные работы 7-8 «Работа с Hadoop и Spark» по курсу «Технологии параллельных систем баз данных» – URL: <https://disk.yandex.ru/d/NkoTT8RvISqvnw> (дата обращения 17.03.2022)
6. Сайт Генератора ФИО - URL: <https://randomus.ru> (дата обращения 17.03.2022)
7. Сайт для помещения строк в кавычки - URL: <https://service.webboss.pro> (дата обращения 17.03.2022)
8. Сайт для проставления запятых URL: <https://lineeditor.ru/index.php> (дата обращения 17.03.2022)
9. Онлайн-генератором JSON-документов URL: <https://app.json-generator.com/> (дата обращения 17.03.2022)
10. Text analysis overview. – URL: <https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-overview.html#analysis-overview> (дата обращения 15.04.2022).
11. Elastic | Kibana 7.0.0 released URL: <https://www.elastic.co/blog/kibana-7-0-0-released> (дата обращения 17.04.2022).
12. Adaptive query execution URL: <https://docs.databricks.com/spark/latest/spark-sql/ape.html> (дата обращения 15.05.2022).

ПРИЛОЖЕНИЕ А

Созданные JSON-файлы и результаты выполнения запросов

Листов 42

1 Созданные JSON документы типа «Процедура»

```
[
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR01",
    "body": {
      "id_of_client": "5H3T",
      "client_age": "37",
      "client_personal_data": "Сорокин Илья Ильич",
      "id_of_procedure": "PR01",
      "date_of_procedure": "2018-01-22",
      "price": "5067",
      "id_of_specialist": "SP03",
      "service": [
        "Ультразвуковая чистка лица",
        "Женская, мужская и детская стрижка",
        "Наращивание и коррекция ногтей гелем",
        "Наращивание и окраска волос",
        "Пилинг стоп"
      ],
      "cosmetics": [
        "Лак ART-VISAGE",
        "Ромашка"
      ]
    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR02",
    "body": {
      "id_of_client": "J4SI",
      "client_age": "49",
      "client_personal_data": "Попов Иван Андреевич",
      "id_of_procedure": "PR02",
      "date_of_procedure": "2018-02-03",
      "price": "2570",
      "id_of_specialist": "SP01",
      "service": [
        "Уход за руками",
        "Покрытие гель-лаком",
        "Брондирование и колорирование",
        "Ламинирование волос"
      ],
      "cosmetics": [
        "Чабрец",
        "Клевер"
      ]
    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR03",
    "body": {
      "id_of_client": "S747",
      "client_age": "63",
      "client_personal_data": "Филатова Елизавета Артемьевна",
      "id_of_procedure": "PR03",
      "date_of_procedure": "2021-12-30",
    }
  }
]
```

```

    "price": "4289",
    "id_of_specialist": "SP02",
    "service": [
        "Маникюр Европейский",
        "Артдизайн",
        "Завивка и выпрямление волос",
        "Каутеризация"
    ],
    "cosmetics": [
        "Крем для лица A`PIEU с гликолевой кислотой",
        "Календула",
        "Кокосовое масло"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR04",
    "body": {
        "id_of_client": "YK9N",
        "client_age": "64",
        "client_personal_data": "Миронова Кира Захаровна",
        "id_of_procedure": "PR04",
        "date_of_procedure": "2016-06-26",
        "price": "5559",
        "id_of_specialist": "SP06",
        "service": [
            "Артдизайн",
            "Парафинотерапия"
        ],
        "cosmetics": [
            "Лак OPI Nail Lacquer Classics"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR05",
    "body": {
        "id_of_client": "YK9N",
        "client_age": "24",
        "client_personal_data": "Зимин Тимур Александрович",
        "id_of_procedure": "PR05",
        "date_of_procedure": "2020-10-09",
        "price": "2901",
        "id_of_specialist": "SP05",
        "service": [
            "Лимфодренаж"
        ],
        "cosmetics": [
            "Лак Benecos Happy Nails"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR06",
    "body": {
        "id_of_client": "8LW3",

```

```

    "client_age": "37",
    "client_personal_data": "Высоцкая Виктория Андреевна",
    "id_of_procedure": "PR06",
    "date_of_procedure": "2016-02-13",
    "price": "4988",
    "id_of_specialist": "SP07",
    "service": [
        "Горячие обвертывания",
        "Spa-маникюр"
    ],
    "cosmetics": [
        "Клевер"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR07",
    "body": {
        "id_of_client": "HXAB",
        "client_age": "20",
        "client_personal_data": "Давыдова Маргарита Матвеевна",
        "id_of_procedure": "PR07",
        "date_of_procedure": "2016-10-31",
        "price": "6418",
        "id_of_specialist": "SP04",
        "service": [
            "Каутеризация",
            "Маникюр Аппаратный",
            "Укрепление натуральных ногтей",
            "Педикюр Классический"
        ],
        "cosmetics": [
            "Липа",
            "Масло семян черного тмина"
        ]
    }
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR08",
    "body": {
        "id_of_client": "QPKX",
        "client_age": "67",
        "client_personal_data": "Петрова Софья Михайловна",
        "id_of_procedure": "PR08",
        "date_of_procedure": "2017-12-03",
        "price": "6753",
        "id_of_specialist": "SP08",
        "service": [
            "Укладки и тонирование",
            "Биопилинг кожи головы"
        ],
        "cosmetics": [
            "Лак Golden Rose"
        ]
    }
}
},
{
    "index": "salon_procedure",

```

```

"doc_type": "Procedure",
"id": "PR09",
"body": {
  "id_of_client": "1AI2",
  "client_age": "66",
  "client_personal_data": "Мельникова Айлин Максимовна",
  "id_of_procedure": "PR09",
  "date_of_procedure": "2019-02-11",
  "price": "4818",
  "id_of_specialist": "SP13",
  "service": [
    "Долговременная укладка"
  ],
  "cosmetics": [
    "Крем для лица A`PIEU с гликолевой кислотой",
    "Облепиховое масло"
  ]
}
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR10",
  "body": {
    "id_of_client": "T2F3",
    "client_age": "62",
    "client_personal_data": "Сычева Елизавета Кирилловна",
    "id_of_procedure": "PR10",
    "date_of_procedure": "2019-01-13",
    "price": "6895",
    "id_of_specialist": "SP09",
    "service": [
      "Артдизайн",
      "Цветовой дизайн волос",
      "Спа-педикюр",
      "Массаж рук",
      "Биопилинг кожи головы"
    ],
    "cosmetics": [
      "Травы Горного Крым"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR11",
  "body": {
    "id_of_client": "9YY4",
    "client_age": "62",
    "client_personal_data": "Калинин Фёдор Максимович",
    "id_of_procedure": "PR11",
    "date_of_procedure": "2016-03-14",
    "price": "1401",
    "id_of_specialist": "SP11",
    "service": [
      "Женская, мужская и детская стрижка",
      "Гальванизация"
    ],
    "cosmetics": [
      "Алерана вит-мин комплекс табл №60",
      "Кунжутное масло",

```

```

        "Календула"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR12",
    "body": {
        "id_of_client": "C84B",
        "client_age": "44",
        "client_personal_data": "Герасимова Юлия Савельевна",
        "id_of_procedure": "PR12",
        "date_of_procedure": "2019-04-12",
        "price": "4927",
        "id_of_specialist": "SP12",
        "service": [
            "Наращивание и коррекция ногтей гелем",
            "Парафинотерапия",
            "Кератиновое разглаживание",
            "Брондирование и колорирование",
            "Лимфодренаж"
        ],
        "cosmetics": [
            "Лак Schwarzkopf Professional Silhouette"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR13",
    "body": {
        "id_of_client": "7HZK",
        "client_age": "46",
        "client_personal_data": "Пантелеева Милана Адамовна",
        "id_of_procedure": "PR13",
        "date_of_procedure": "2018-06-13",
        "price": "5192",
        "id_of_specialist": "SP14",
        "service": [
            "Аппаратный"
        ],
        "cosmetics": [
            "Льняное масло",
            "Клевер",
            "Витамин Е альфа-токоферола ацетат 200мг капс №10"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR14",
    "body": {
        "id_of_client": "YK9N",
        "client_age": "38",
        "client_personal_data": "Семенова Вероника Тимуровна",
        "id_of_procedure": "PR14",
        "date_of_procedure": "2017-08-12",
        "price": "612",
        "id_of_specialist": "SP15",
    }
}

```



```

    "service": [
        "Укрепление натуральных ногтей",
        "Свадебные и вечерние прически"
    ],
    "cosmetics": [
        "Кокосовое масло",
        "Аевит капс №20"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR15",
    "body": {
        "id_of_client": "INYNK",
        "client_age": "30",
        "client_personal_data": "Седов Егор Константинович",
        "id_of_procedure": "PR15",
        "date_of_procedure": "2018-12-22",
        "price": "5907",
        "id_of_specialist": "SP20",
        "service": [
            "Ламинирование волос"
        ],
        "cosmetics": [
            "Витамин Е альфа-токоферола ацетат 200мг капс №20",
            "Аевит капс №20"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR16",
    "body": {
        "id_of_client": "Z8Q2",
        "client_age": "34",
        "client_personal_data": "Савельев Василий Олегович",
        "id_of_procedure": "PR16",
        "date_of_procedure": "2016-11-21",
        "price": "1935",
        "id_of_specialist": "SP19",
        "service": [
            "Массаж рук",
            "Педикюр Классический",
            "Маникюр Аппаратный"
        ],
        "cosmetics": [
            "Кокосовое масло"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR17",
    "body": {
        "id_of_client": "8VNM",
        "client_age": "55",
        "client_personal_data": "Павлова Полина Артемьевна",
        "id_of_procedure": "PR17",

```

```

    "date_of_procedure": "2018-03-21",
    "price": "5747",
    "id_of_specialist": "SP16",
    "service": [
        "Педикюр Классический",
        "Массаж рук",
        "Маникюр Европейский"
    ],
    "cosmetics": [
        "Лак Schwarzkopf Professional Silhouette"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR18",
    "body": {
        "id_of_client": "QFNA",
        "client_age": "43",
        "client_personal_data": "Мещеряков Александр Львович",
        "id_of_procedure": "PR18",
        "date_of_procedure": "2019-02-05",
        "price": "4962",
        "id_of_specialist": "SP18",
        "service": [
            "Маникюр Аппаратный",
            "Лимфодренаж",
            "Завивка и выпрямление волос",
            "Долгосрочное покрытие гель-лаками (ши-лаками)",
            "Вакуумная чистка лица"
        ],
        "cosmetics": [
            "Витамин Е альфа-токоферола ацетат 200мг капс №20",
            "Тыквенное масло"
        ]
    }
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR19",
    "body": {
        "id_of_client": "DMSH",
        "client_age": "29",
        "client_personal_data": "Суворова Виктория Эмильевна",
        "id_of_procedure": "PR19",
        "date_of_procedure": "2016-05-17",
        "price": "3579",
        "id_of_specialist": "SP17",
        "service": [
            "Свадебные и вечерние прически",
            "Парафинотерапия",
            "Биопилинг кожи головы",
            "Массаж рук"
        ],
        "cosmetics": [
            "Увлажняющий дневной крем для чувствительной кожи Redness Solutions",
            "Ромашка"
        ]
    }
}
},

```

```

{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR20",
  "body": {
    "id_of_client": "IDT2",
    "client_age": "18",
    "client_personal_data": "Рогова Екатерина Данииловна",
    "id_of_procedure": "PR20",
    "date_of_procedure": "2021-05-19",
    "price": "5226",
    "id_of_specialist": "SP21",
    "service": [
      "Маникюр классический мужской и детский"
    ],
    "cosmetics": [
      "Лак Golden Rose",
      "Лак NailLOOK Stone Spa"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR21",
  "body": {
    "id_of_client": "62TC",
    "client_age": "50",
    "client_personal_data": "Чернова Мария Артёмовна",
    "id_of_procedure": "PR21",
    "date_of_procedure": "2021-01-26",
    "price": "6568",
    "id_of_specialist": "SP",
    "service": [
      "Лимфодренаж",
      "Долговременная укладка",
      "Дарсонваль"
    ],
    "cosmetics": [
      "Льняное масло"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR22",
  "body": {
    "id_of_client": "VJIM",
    "client_age": "31",
    "client_personal_data": "Некрасова Маргарита Константиновна",
    "id_of_procedure": "PR22",
    "date_of_procedure": "2020-11-24",
    "price": "1971",
    "id_of_specialist": "SP23",
    "service": [
      "Укрепление натуральных ногтей",
      "Ламинирование волос",
      "Ионофорез"
    ],
    "cosmetics": [
      "Лак OSIS+Session, экстрасильная фиксация"
    ]
  }
}

```

```

    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR23",
  "body": {
    "id_of_client": "RA8X",
    "client_age": "63",
    "client_personal_data": "Зимин Тимур Александрович",
    "id_of_procedure": "PR23",
    "date_of_procedure": "2021-08-14",
    "price": "2634",
    "id_of_specialist": "SP25",
    "service": [
      "Мелирование и блондирование",
      "Плетение косичек",
      "Брашинг"
    ],
    "cosmetics": [
      "Касторовое масло"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR24",
  "body": {
    "id_of_client": "6KJY",
    "client_age": "35",
    "client_personal_data": "Яковлева Вероника Матвеевна",
    "id_of_procedure": "PR24",
    "date_of_procedure": "2016-03-12",
    "price": "4726",
    "id_of_specialist": "SP22",
    "service": [
      "Наращивание и коррекция ногтей гелем",
      "Массаж рук"
    ],
    "cosmetics": [
      "Рыжиковое масло",
      "Масло черного тмина"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR25",
  "body": {
    "id_of_client": "1AI2",
    "client_age": "56",
    "client_personal_data": "Ильина Валерия Николаевна",
    "id_of_procedure": "PR25",
    "date_of_procedure": "2019-04-19",
    "price": "1584",
    "id_of_specialist": "SP24",
    "service": [
      "Уход за волосами и лечение волос",
      "Покрывтие гель-лаком",

```

```

        "Завивка и выпрямление волос",
        "Наращивание и коррекция ногтей гелем",
        "Артдизайн"
    ],
    "cosmetics": [
        "Консумед"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR26",
    "body": {
        "id_of_client": "44MD",
        "client_age": "46",
        "client_personal_data": "Пономарева Карина Андреевна",
        "id_of_procedure": "PR26",
        "date_of_procedure": "2019-01-17",
        "price": "2181",
        "id_of_specialist": "SP01",
        "service": [
            "Долговременная укладка",
            "Покрытие лаком"
        ],
        "cosmetics": [
            "Лен",
            "Лак Golden Rose"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR27",
    "body": {
        "id_of_client": "0TJQ",
        "client_age": "69",
        "client_personal_data": "Смирнов Александр Александрович",
        "id_of_procedure": "PR27",
        "date_of_procedure": "2021-06-18",
        "price": "1459",
        "id_of_specialist": "SP05",
        "service": [
            "Плетение косичек"
        ],
        "cosmetics": [
            "Лак Schwarzkopf Professional Silhouette"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR28",
    "body": {
        "id_of_client": "0TJQ",
        "client_age": "30",
        "client_personal_data": "Прохоров Александр Александрович",
        "id_of_procedure": "PR28",
        "date_of_procedure": "2019-03-23",
        "price": "6836",

```

```

    "id_of_specialist": "SP10",
    "service": [
        "Брашинг",
        "Укладки и тонирование",
        "Покрытие лаком"
    ],
    "cosmetics": [
        "Кедровое масло"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR29",
    "body": {
        "id_of_client": "RA8X",
        "client_age": "60",
        "client_personal_data": "Грачев Никита Дамирович",
        "id_of_procedure": "PR29",
        "date_of_procedure": "2021-03-28",
        "price": "5433",
        "id_of_specialist": "SP16",
        "service": [
            "Амбре-окрашивание",
            "Лимфодренаж",
            "Вакуумная чистка лица"
        ],
        "cosmetics": [
            "Льняное масло",
            "Лак Matrix Vavoom Freezing Spray"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR30",
    "body": {
        "id_of_client": "C84B",
        "client_age": "49",
        "client_personal_data": "Иванова Софья Максимовна",
        "id_of_procedure": "PR30",
        "date_of_procedure": "2018-06-13",
        "price": "3338",
        "id_of_specialist": "SP19",
        "service": [
            "Педикюр Классический",
            "Отбеливание",
            "Спа-педикюр",
            "Массаж рук"
        ],
        "cosmetics": [
            "Травы Алтая",
            "Лак WULA"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR31",

```

```

"body": {
  "id_of_client": "USY2",
  "client_age": "65",
  "client_personal_data": "Плотникова Александра Григорьевна",
  "id_of_procedure": "PR31",
  "date_of_procedure": "2019-03-31",
  "price": "635",
  "id_of_specialist": "SP24",
  "service": [
    "Женская, мужская и детская стрижка",
    "Аппаратный",
    "Мелирование и блондирование"
  ],
  "cosmetics": [
    "Липа",
    "Лак ART-VISAGE"
  ]
}
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR32",
  "body": {
    "id_of_client": "8IDC",
    "client_age": "66",
    "client_personal_data": "Чернова Мария Артёмовна",
    "id_of_procedure": "PR32",
    "date_of_procedure": "2017-10-09",
    "price": "4091",
    "id_of_specialist": "SP02",
    "service": [
      "Долгосрочное покрытие гель-лаками (ши-лаками)",
      "Spa-маникюр"
    ],
    "cosmetics": [
      "Лак WULA",
      "Миндальное масло",
      "Рыжиковое масло"
    ]
  }
},
{
  "index": "salon_procedure",
  "doc_type": "Procedure",
  "id": "PR33",
  "body": {
    "id_of_client": "C84B",
    "client_age": "43",
    "client_personal_data": "Кудряшова Агата Егоровна",
    "id_of_procedure": "PR33",
    "date_of_procedure": "2021-06-04",
    "price": "3713",
    "id_of_specialist": "SP01",
    "service": [
      "Завивка и выпрямление волос",
      "Покрытие гель-лаком",
      "Spa-маникюр"
    ],
    "cosmetics": [
      "Рыжиковое масло"
    ]
  }
}

```

```

    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR34",
    "body": {
      "id_of_client": "UTNZ",
      "client_age": "56",
      "client_personal_data": "Зубков Роман Александрович",
      "id_of_procedure": "PR34",
      "date_of_procedure": "2019-06-24",
      "price": "3275",
      "id_of_specialist": "SP19",
      "service": [
        "Брондирование и колорирование",
        "Амбре-окрашивание",
        "Вакуумная чистка лица",
        "Парафинотерапия"
      ],
      "cosmetics": [
        "Витамин Е альфа-токоферола ацетат 200мг капс №10",
        "Льняное масло"
      ]
    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR35",
    "body": {
      "id_of_client": "T9UC",
      "client_age": "36",
      "client_personal_data": "Калинин Артём Максимович",
      "id_of_procedure": "PR35",
      "date_of_procedure": "2018-07-16",
      "price": "6609",
      "id_of_specialist": "SP22",
      "service": [
        "Покрытие гель-лаком",
        "Маникюр Аппаратный",
        "Дарсонваль",
        "Ламинирование волос",
        "Горячие обвертывания"
      ],
      "cosmetics": [
        "Крем SHISEIDO"
      ]
    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR36",
    "body": {
      "id_of_client": "8VNM",
      "client_age": "62",
      "client_personal_data": "Алексеева Ольга Ярославовна",
      "id_of_procedure": "PR36",
      "date_of_procedure": "2019-01-07",
      "price": "6248",
      "id_of_specialist": "SP10",

```



```

    "service": [
        "Женская, мужская и детская стрижка",
        "Маникюр Европейский",
        "Педикюр Классический"
    ],
    "cosmetics": [
        "Аевит капс №20",
        "Облепиховое масло"
    ]
}
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR37",
    "body": {
        "id_of_client": "0TJQ",
        "client_age": "22",
        "client_personal_data": "Долгова Александра Антоновна",
        "id_of_procedure": "PR37",
        "date_of_procedure": "2018-08-08",
        "price": "2935",
        "id_of_specialist": "SP11",
        "service": [
            "Цветовой дизайн волос",
            "Отбеливание",
            "Лимфодренаж",
            "Горячие обвертывания"
        ],
        "cosmetics": [
            "Тыквенное масло",
            "Лак Essence",
            "Лак WULA"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR38",
    "body": {
        "id_of_client": "5S12",
        "client_age": "69",
        "client_personal_data": "Жарова Анастасия Ивановна",
        "id_of_procedure": "PR38",
        "date_of_procedure": "2021-05-11",
        "price": "1499",
        "id_of_specialist": "SP11",
        "service": [
            "Spa-педикюр",
            "Кератиновое разглаживание",
            "Наращивание и окраска волос",
            "Уход за волосами и лечение волос"
        ],
        "cosmetics": [
            "Календула"
        ]
    }
},
{
    "index": "salon_procedure",
    "doc_type": "Procedure",

```

```

    "id": "PR39",
    "body": {
      "id_of_client": "YR0Y",
      "client_age": "31",
      "client_personal_data": "Круглова Амина Матвеевна",
      "id_of_procedure": "PR39",
      "date_of_procedure": "2018-10-23",
      "price": "4054",
      "id_of_specialist": "SP08",
      "service": [
        "Дарсонваль",
        "Завивка и выпрямление волос",
        "Наращивание и окраска волос",
        "Spa-маникюр"
      ],
      "cosmetics": [
        "Тыквенное масло",
        "Льняное масло"
      ]
    }
  },
  {
    "index": "salon_procedure",
    "doc_type": "Procedure",
    "id": "PR40",
    "body": {
      "id_of_client": "9YY4",
      "client_age": "43",
      "client_personal_data": "Суворова Виктория Эмильевна",
      "id_of_procedure": "PR40",
      "date_of_procedure": "2019-02-02",
      "price": "2795",
      "id_of_specialist": "SP25",
      "service": [
        "Гальванизация",
        "Дарсонваль"
      ],
      "cosmetics": [
        "Биотин Форте капс №60",
        "Травы Алтая",
        "Лак Schwarzkopf Professional Silhouette"
      ]
    }
  }
]

```

2 Созданные JSON документы типа «Мастер»

```
[
  {
    "index": "/home/hadoopuser/js_files",
    "doc_type": "Master",
    "id": "SP01",
    "body": {
      "specialisation": [
        "Покрытие гель-лаком",
        "Аппаратный",
        "Отбеливание",
        "Завивка и выпрямление волос"
      ],
      "work_experience": "4",
      "master_personal_data": "Шаповалова Мария Александровна",
      "schedule": "Чт,Пт,Сб 15:00-20:00",
      "reviews": [
        "Спасибо огромное мастеру за чудесный маникюр и педикюр !!! Все просто супер
        !!!!",
        "Спасибо огромное в очередной раз... моя красота (простите за нескромность) в
        ваших руках!",
        "Была первый раз вернусь обязательно еще.",
        "Всегда выхожу от вас с позитивным настроем!!"
      ]
    }
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP02",
    "body": {
      "specialisation": [
        "Пилинг стоп",
        "Дезинкрустация",
        "Женская, мужская и детская стрижка",
        "Spa-маникюр"
      ],
      "work_experience": "34",
      "master_personal_data": "Трофимов Михаил Олегович",
      "schedule": "Ср, Пт 15:00-20:00",
      "reviews": [
        "Люди будьте бдительны",
        "Спасибо большое мастеру за очаровательное каре, золотые руки, приятный
        собеседник, восхищаюсь образом который мне подарили!",
        "Всегда выхожу от вас с позитивным настроем!!",
        "Михаил внимательный и приятный в общении мастер. Однако на выходе я получила
        немного не тот цвет, который хотела."
      ]
    }
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP03",
    "body": {
      "specialisation": [
        "Маникюр Европейский",
        "Свадебные и вечерние прически",
        "Кератиновое разглаживание"
      ],
    }
  }
]
```

```

    "work_experience": "36",
    "master_personal_data": "Белоусова Елизавета Артёмовна",
    "schedule": "Пн, Вт, Сб, Вс 15:00-20:00",
    "reviews": [
        "Делаю брови, ресницы, брови, ламинировала волосы. Качеством довольна. Минус есть-принимают позже назначенного времени. Есть куда стремиться.",
        "Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу.",
        "Постоянно хожу в салон на маникюр и педикюр к этому мастеру. Мастер все делает аккуратно, доброжелательно.",
        "Выбор гель-лаков потрясающий, глаза разбежались. А цены очень порадовали кошелек)))",
        "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти."
    ]
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP04",
    "body": {
      "specialisation": [
        "Маникюр Европейский",
        "Уход за руками",
        "Долговременная укладка",
        "Долгосрочное покрытие гель-лаками (ши-лаками)"
      ],
      "work_experience": "40",
      "master_personal_data": "Семина Анна Игоревна",
      "schedule": "Пн, Вт, Чт, Пт 10:00-15:00",
      "reviews": [
        "Всегда выхожу от вас с позитивным настроением!",
        "Первый раз в жизни пишу негативный отзыв, но тут уже нет предела наглости, а я наглых людей не люблю. Записалась на педикюр, но видите ли мастер не успевал сделать покрытие, которое я хотела. мол её не предупредили и она время не оставила. на что администратор после сказала, что она записала с покрытием, а вы мастеру просто не понравились.",
        "Не рекомендую",
        "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти.",
        "Никогда больше не пойду к этому мастеру. Работа выполнена ужасно."
      ]
    }
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP05",
    "body": {
      "specialisation": [
        "Педикюр Классический",
        "Амбре-окрашивание",
        "Дезинкрустация",
        "Ультразвуковая чистка лица"
      ],
      "work_experience": "21",
      "master_personal_data": "Матвеева Валерия Максимовна",
      "schedule": "Пн, Вт, Чт, Пт 10:00-15:00",
      "reviews": [
        "Прекрасный мастер!",

```

"Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново.Результат: испорчены ногти, настроение.",
 "Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",
 "Никогда больше не пойдк этому мастеру. Работа выполнена ужасно."

```

    ]
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP06",
    "body": {
      "specialisation": [
        "Ионофорец",
        "Цветовой дизайн волос",
        "Маникюр классический мужской и детский",
        "Свадебные и вечерние прически"
      ],
      "work_experience": "18",
      "master_personal_data": "Горелова Любовь Тимофеевна",
      "schedule": "С6, Вс 10:00-15:00",
      "reviews": [
        "За стрижку со скидкой спасибо от души!",
        "Извините, но это помойка, а не студия маникюра.То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",
        "Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!",
        "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти.",
        "Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу."
      ]
    }
  },

```

```

  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP07",
    "body": {
      "specialisation": [
        "Дезинкрустация",
        "Ультразвуковая чистка лица",
        "Ламинирование волос",
        "Амбре-окрашивание"
      ],
      "work_experience": "22",
      "master_personal_data": "Горбачева Валерия Артемьевна",
      "schedule": "С6, Вс 15:00-20:00",
      "reviews": [
        "Извините, но это помойка, а не студия маникюра.То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",
        "Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!",
        "За стрижку со скидкой спасибо от души!",
        "Спасибо большое мастеру за очаровательное каре, золотые руки, приятный собеседник, восхищаюсь образом который мне подарили!",

```

"Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново.Результат: испорчены ногти, настроение."

```

    ]
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP08",
    "body": {
      "specialisation": [
        "Покрытие лаком",
        "Отбеливание",
        "Парафинотерапия",
        "Вакуумная чистка лица"
      ],
      "work_experience": "7",
      "master_personal_data": "Никитина Софья Павловна",
      "schedule": "Пн, Вт, Чт, Пт 15:00-20:00",
      "reviews": [
        "Первый раз делала чистку лица, и уже с первого раза виден результат. Час процедуры прошел незаметно в приятной позитивной обстановке, что не маловажно) Приду скоро еще!",
        "Постоянно хожу в салон на маникюр и педикюр к этому мастеру. Мастер все делает аккуратно, доброжелательно.",
        "Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",
        "Спасибо за отличное настроение и классную работу."
      ]
    }
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP09",
    "body": {
      "specialisation": [
        "Каутеризация",
        "Гальванизация",
        "Ионофорез",
        "Электрофорез"
      ],
      "work_experience": "20",
      "master_personal_data": "Петровская Ника Кирилловна",
      "schedule": "Пн, Вт, Чт, Пт 15:00-20:00",
      "reviews": [
        "Никогда больше не пойду к этому мастеру. Работа выполнена ужасно.",
        "Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!",
        "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти."
      ]
    }
  },
  {
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP10",
    "body": {
      "specialisation": [
        "Массаж рук",
        "Spa-маникюр",

```

```

    "Биопилинг кожи головы",
    "Гальванизация"
  ],
  "work_experience": "2",
  "master_personal_data": "Ильинский Матвей Эмирович",
  "schedule": "Ср, Пт 10:00-15:00",
  "reviews": [
    "Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",
    "Не рекомендую",
    "Извините, но это помойка, а не студия маникюра. То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",
    "Спасибо большое мастеру за очаровательное каре, золотые руки, приятный собеседник, восхищаюсь образом который мне подарили!"
  ]
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP11",
  "body": {
    "specialisation": [
      "Мелирование и блондирование",
      "Долгосрочное покрытие гель-лаками (ши-лаками)",
      "Спа-педикюр",
      "Ультразвуковая чистка лица"
    ],
    "work_experience": "34",
    "master_personal_data": "Назарова Аделя Константиновна",
    "schedule": "Сб, Вс 15:00-20:00",
    "reviews": [
      "Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу.",
      "Всегда выхожу от вас с позитивным настроением!"
    ]
  }
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP12",
  "body": {
    "specialisation": [
      "Покрытие лаком",
      "Брашинг",
      "Пилинг стоп"
    ],
    "work_experience": "6",
    "master_personal_data": "Покровский Даниил Владиславович",
    "schedule": "Сб, Вс 10:00-15:00",
    "reviews": [
      "Первый раз в жизни пишу негативный отзыв, но тут уже нет предела наглости, а я наглых людей не люблю. Записалась на педикюр, но видите ли мастер не успевал сделать покрытие, которое я хотела. мол её не предупредили и она время не оставила. на что администратор после сказала, что она записала с покрытием, а вы мастеру просто не понравились.",
      "Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу.",

```

```

        "Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",
        "Выбор гель-лаков потрясающий, глаза разбежались. А цены очень порадовали кошелек)))"
    ]
}
},
{
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP13",
    "body": {
        "specialisation": [
            "Ультразвуковая чистка лица",
            "Отбеливание",
            "Биопилинг кожи головы",
            "Покрывание гель-лаком"
        ],
        "work_experience": "40",
        "master_personal_data": "Губанов Лев Максимович",
        "schedule": "Пн, Ср, Вс 10:00-15:00",
        "reviews": [
            "Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка.",
            "Делаю брови, ресницы, брови, ламинировала волосы. Качеством довольна. Минус есть-принимают позже назначенного времени. Есть куда стремиться.",
            "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти.",
            "Второй день хожу вся расстроенная от того, как не качественно подстригли. Каскад, где одна прядь короче другой на пару сантиметров и профилирована раз 5, а с другой пару."
        ]
    }
},
{
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP14",
    "body": {
        "specialisation": [
            "Педикюр Классический",
            "Наращивание и окраска волос",
            "Покрывание лаком"
        ],
        "work_experience": "5",
        "master_personal_data": "Кондратова Александра Максимовна",
        "schedule": "Пн, Вт, Ср, Вс 10:00-15:00",
        "reviews": [
            "Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново.Результат: испорчены ногти, настроение.",
            "Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти.",
            "Всегда выхожу от вас с позитивным настроением!!!"
        ]
    }
},
{
    "index": "salon_master",
    "doc_type": "Master",
    "id": "SP15",
    "body": {
        "specialisation": [

```



```

        "Амбре-окрашивание",
        "Гальванизация",
        "Дарсонваль",
        "Парафинотерапия"
    ],
    "work_experience": "25",
    "master_personal_data": "Мальцева Анна Ярославовна",
    "schedule": "Сб, Вс 10:00-15:00",
    "reviews": [
        "Второй день хожу вся расстроенная от того, как не качественно подстригли. Каскад, где одна прядь короче другой на пару сантиметров и профилирована раз 5, а с другой пару.",
        "Делаю брови, ресницы, брови, ламинировала волосы. Качеством довольна. Минус есть-принимают позже назначенного времени. Есть куда стремиться.",
        "Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",
        "Извините, но это помойка, а не студия маникюра. То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",
        "Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка."
    ]
    },
    {
        "index": "salon_master",
        "doc_type": "Master",
        "id": "SP16",
        "body": {
            "specialisation": [
                "Горячие обвертывания",
                "Ионофорез",
                "Завивка и выпрямление волос",
                "Spa-маникюр"
            ],
            "work_experience": "24",
            "master_personal_data": "Попова Анастасия Львовна",
            "schedule": "Пн, Вт, Сб, Вс 15:00-20:00",
            "reviews": [
                "Работа выполнена хорошо"
            ]
        }
    },
    {
        "index": "salon_master",
        "doc_type": "Master",
        "id": "SP17",
        "body": {
            "specialisation": [
                "Массаж рук",
                "Плетение косичек",
                "Гальванизация",
                "Брондирование и колорирование"
            ],
            "work_experience": "1",
            "master_personal_data": "Мельникова Ульяна Андреевна",
            "schedule": "Ср, Пт 10:00-15:00",
            "reviews": [
                "Всегда выхожу от вас с позитивным настроением!",
                "Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",
            ]
        }
    }
}

```

"Прекрасный врач-косметолог, решает проблемы на раз-два. Результатом всегда довольна.",

"Наконец моя мечта сбылась и я сделала Лифтинг! Это не дешёвая процедура, но результат того стоит, я в восторге!"

```
]
}
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP18",
  "body": {
    "specialisation": [
      "Горячие обвертывания",
      "Артдизайн",
      "Наращивание и коррекция ногтей гелем",
      "Укладки и тонирование"
    ],
    "work_experience": "31",
    "master_personal_data": "Майорова Екатерина Дмитриевна",
    "schedule": "Чт,Пт,Сб 15:00-20:00",
    "reviews": [
      "Работа выполнена хорошо",
      "Спасибо большое мастеру за очаровательное каре, золотые руки, приятный собеседник, восхищаюсь образом который мне подарили!",
      "Никогда больше не пойдк этому мастеру. Работа выполнена ужасно."
    ]
  }
}
```

```
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP19",
  "body": {
    "specialisation": [
      "Spa-педикюр",
      "Биопилинг кожи головы",
      "Мелирование и блондирование",
      "Маникюр классический мужской и детский"
    ],
    "work_experience": "24",
    "master_personal_data": "Морозова Алина Романовна",
    "schedule": "Пн, Ср, Вс 10:00-15:00",
    "reviews": [
      "Всегда выхожу от вас с позитивным настроем!!",
      "Была в Клинике красоты впервые, мастер плохо сделала биопилинг.",
      "Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",
      "Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!"
    ]
  }
}
```

```
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP20",
  "body": {
    "specialisation": [
      "Покрытие гель-лаком",
      "Парафинотерапия",
      "Аппаратный",

```

```

    "Дезинкрустация"
  ],
  "work_experience": "39",
  "master_personal_data": "Макарова Валерия Петровна",
  "schedule": "Сб, Вс 15:00-20:00",
  "reviews": [
    "Люди будьте бдительны",
    "Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка.",
    "Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",
    "Наконец моя мечта сбылась и я сделала Лифтинг! Это не дешёвая процедура, но результат того стоит, я в восторге!",
    "Спасибо за отличное настроение и классную работу."
  ]
}
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP21",
  "body": {
    "specialisation": [
      "Парафинотерапия",
      "Укладки и тонирование",
      "Spa-педикюр"
    ],
    "work_experience": "31",
    "master_personal_data": "Агеев Фёдор Даниилович",
    "schedule": "Чт,Пт,Сб 15:00-20:00",
    "reviews": [
      "Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка.",
      "Нестерильно. Мастер разговаривает по телефону во время маникюра и после разговора не обрабатывает руки.",
      "За стрижку со скидкой спасибо от души!"
    ]
  }
}
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP22",
  "body": {
    "specialisation": [
      "Парафинотерапия",
      "Брондирование и колорирование",
      "Наращивание и коррекция ногтей гелем",
      "Аппаратный"
    ],
    "work_experience": "15",
    "master_personal_data": "Курочкин Алексей Иванович",
    "schedule": "Пн, Вт, Чт, Пт 15:00-20:00",
    "reviews": [
      "Работа выполнена хорошо",
      "Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново.Результат: испорчены ногти, настроение.",
      "Записалась на коррекцию. Зайдя в кабинет, была сильно удивлена, насколько были грязные столы, и разбросан материал, неизвестного происхождения, и незнакомых брендов."
    ]
  }
}
}

```

Самое ужасное началось когда начали делать покрытие, небрежно, гель затёк под кутикулу, образовались бугры, ногти выглядели ужасно, разной формы и длинны.",

"Нестерильно. Мастер разговаривает по телефону во время маникюра и после разговора не обрабатывает руки."

```
]
}
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP23",
  "body": {
    "specialisation": [
      "Лимфодренаж",
      "Цветовой дизайн волос",
      "Женская, мужская и детская стрижка",
      "Педикюр Классический"
    ],
    "work_experience": "8",
    "master_personal_data": "Егоров Дмитрий Давидович",
    "schedule": "Пн, Вт, Чт, Пт 10:00-15:00",
    "reviews": [
      "Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",
      "Была в Клинике красоты впервые, мастер понравился"
    ]
  }
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP24",
  "body": {
    "specialisation": [
      "Цветовой дизайн волос",
      "Педикюр Классический",
      "Электрофорез",
      "Дарсонваль"
    ],
    "work_experience": "1",
    "master_personal_data": "Михайлова Алина Михайловна",
    "schedule": "Пн, Вт, Чт, Пт 10:00-15:00",
    "reviews": [
      "Спасибо огромное мастеру за чудесный маникюр и педикюр !!! Все просто супер !!!!",
      "Была первый раз вернусь обязательно еще.",
      "Была в Клинике красоты впервые, мастер супер"
    ]
  }
},
{
  "index": "salon_master",
  "doc_type": "Master",
  "id": "SP25",
  "body": {
    "specialisation": [
      "Кератиновое разглаживание",
      "Женская, мужская и детская стрижка",
      "Уход за руками",
      "Гальванизация"
    ],
    "work_experience": "5",
```

```
    "master_personal_data": "Иванова Елизавета Михайловна",  
    "schedule": "Пн, Вт, Чт, Пт 10:00-15:00",  
    "reviews": [  
        "Работа выполнена хорошо",  
        "Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро.  
Очень довольна.",  
        "Была первый раз вернусь обязательно еще."  
    ]  
}  
]
```

3 Результат выполнения первого запроса к кластеру ES

```
{
  "took" : 4,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 40,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "year_period" : {
      "buckets" : [
        {
          "key_as_string" : "2016-01-01",
          "key" : 1451606400000,
          "doc_count" : 7,
          "Specialist" : {
            "doc_count_error_upper_bound" : 0,
            "sum_other_doc_count" : 0,
            "buckets" : [
              {
                "key" : "sp04",
                "doc_count" : 1,
                "summary_price" : {
                  "value" : 6418.0
                }
              }
            ]
          }
        }
      ]
    }
  }
}
```

```

},
{
  "key" : "sp06",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 5559.0
  }
},
{
  "key" : "sp07",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 4988.0
  }
},
{
  "key" : "sp11",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 1401.0
  }
},
{
  "key" : "sp17",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 3579.0
  }
},
{
  "key" : "sp19",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 1935.0
  }
}

```

```

    },
    {
      "key" : "sp22",
      "doc_count" : 1,
      "summary_price" : {
        "value" : 4726.0
      }
    }
  ]
}
},
{
  "key_as_string" : "2017-01-01",
  "key" : 1483228800000,
  "doc_count" : 3,
  "Specialist" : {
    "doc_count_error_upper_bound" : 0,
    "sum_other_doc_count" : 0,
    "buckets" : [
      {
        "key" : "sp02",
        "doc_count" : 1,
        "summary_price" : {
          "value" : 4091.0
        }
      }
    ],
    {
      "key" : "sp08",
      "doc_count" : 1,
      "summary_price" : {
        "value" : 6753.0
      }
    },
    {
      "key" : "sp15",

```



```

        "doc_count" : 1,
        "summary_price" : {
            "value" : 612.0
        }
    }
]
}
},
{
    "key_as_string" : "2018-01-01",
    "key" : 1514764800000,
    "doc_count" : 9,
    "Specialist" : {
        "doc_count_error_upper_bound" : 0,
        "sum_other_doc_count" : 0,
        "buckets" : [
            {
                "key" : "sp01",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 2570.0
                }
            },
            {
                "key" : "sp03",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 5067.0
                }
            },
            {
                "key" : "sp08",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 4054.0
                }
            }
        ]
    }
}

```

```

    }
  },
  {
    "key" : "sp11",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 2935.0
    }
  },
  {
    "key" : "sp14",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 5192.0
    }
  },
  {
    "key" : "sp16",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 5747.0
    }
  },
  {
    "key" : "sp19",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 3338.0
    }
  },
  {
    "key" : "sp20",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 5907.0
    }
  }

```

```

        }
    },
    {
        "key" : "sp22",
        "doc_count" : 1,
        "summary_price" : {
            "value" : 6609.0
        }
    }
]
}
},
{
    "key_as_string" : "2019-01-01",
    "key" : 1546300800000,
    "doc_count" : 11,
    "Specialist" : {
        "doc_count_error_upper_bound" : 0,
        "sum_other_doc_count" : 0,
        "buckets" : [
            {
                "key" : "sp01",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 2181.0
                }
            },
            {
                "key" : "sp09",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 6895.0
                }
            },
            {

```

```

    "key" : "sp10",
    "doc_count" : 2,
    "summary_price" : {
      "value" : 13084.0
    }
  },
  {
    "key" : "sp12",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 4927.0
    }
  },
  {
    "key" : "sp13",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 4818.0
    }
  },
  {
    "key" : "sp18",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 4962.0
    }
  },
  {
    "key" : "sp19",
    "doc_count" : 1,
    "summary_price" : {
      "value" : 3275.0
    }
  },
  {

```

```

        "key" : "sp24",
        "doc_count" : 2,
        "summary_price" : {
            "value" : 2219.0
        }
    },
    {
        "key" : "sp25",
        "doc_count" : 1,
        "summary_price" : {
            "value" : 2795.0
        }
    }
]
}
},
{
    "key_as_string" : "2020-01-01",
    "key" : 1577836800000,
    "doc_count" : 2,
    "Specialist" : {
        "doc_count_error_upper_bound" : 0,
        "sum_other_doc_count" : 0,
        "buckets" : [
            {
                "key" : "sp05",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 2901.0
                }
            }
        ],
        {
            "key" : "sp23",
            "doc_count" : 1,
            "summary_price" : {

```

```

        "value" : 1971.0
    }
}
]
}
},
{
    "key_as_string" : "2021-01-01",
    "key" : 1609459200000,
    "doc_count" : 8,
    "Specialist" : {
        "doc_count_error_upper_bound" : 0,
        "sum_other_doc_count" : 0,
        "buckets" : [
            {
                "key" : "sp",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 6568.0
                }
            },
            {
                "key" : "sp01",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 3713.0
                }
            },
            {
                "key" : "sp02",
                "doc_count" : 1,
                "summary_price" : {
                    "value" : 4289.0
                }
            },
        ],
    },
}

```

```

{
  "key" : "sp05",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 1459.0
  }
},
{
  "key" : "sp11",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 1499.0
  }
},
{
  "key" : "sp16",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 5433.0
  }
},
{
  "key" : "sp21",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 5226.0
  }
},
{
  "key" : "sp25",
  "doc_count" : 1,
  "summary_price" : {
    "value" : 2634.0
  }
}

```


4 Результат выполнения второго запроса к кластеру ES

```
{
  "took" : 26,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 14,
      "relation" : "eq"
    },
    "max_score" : 4.650751,
    "hits" : [
      {
        "_index" : "salon_master",
        "_type" : "Master",
        "_id" : "SP04",
        "_score" : 4.650751,
        "_source" : {
          "master_personal_data" : "Семина Анна Игоревна"
        },
        "highlight" : {
          "reviews" : [
            "Первый раз в жизни пишу негативный отзыв, но тут уже нет предела  
<em>наглости</em>, а я наглых людей не люблю.",
            "Если вы читаете этот отзыв, то знайте, что мастер <em>плохо</em> подходит  
к своему делу.",
            "Работа выполнена <em>ужасно</em>."
          ]
        }
      }
    ]
  }
}
```

```

    "_index" : "salon_master",
    "_type" : "Master",
    "_id" : "SP22",
    "_score" : 4.361892,
    "_source" : {
      "master_personal_data" : "Курочкин Алексей Иванович"
    },
    "highlight" : {
      "reviews" : [
        "Зайдя в кабинет, была сильно удивлена, насколько были <em>грязные</em> столы, и разбросан материал, неизвестного",
        "Самое <em>ужасное</em> началось когда начали делать покрытие, небрежно, гель затёк под кутикулу, образовались",
        "бугры, ногти выглядели <em>ужасно</em>, разной формы и длины."
      ]
    }
  },
  {
    "_index" : "salon_master",
    "_type" : "Master",
    "_id" : "SP15",
    "_score" : 3.9118097,
    "_source" : {
      "master_personal_data" : "Мальцева Анна Ярославовна"
    },
    "highlight" : {
      "reviews" : [
        "Второй день хожу вся <em>расстроенная</em> от того, как не качественно подстригли.",
        "Извините, но это <em>помойка</em>, а не студия маникюра.То ли мне не повезло, то ли, действительно, все <em>печально</em>"
      ]
    }
  },
  {
    "_index" : "salon_master",
    "_type" : "Master",

```

```

    "_id" : "SP06",
    "_score" : 3.804809,
    "_source" : {
      "master_personal_data" : "Горелова Любовь Тимофеевна"
    },
    "highlight" : {
      "reviews" : [
        "Извините, но это <em>помойка</em>, а не студия маникюра.То ли мне не повезло, то ли, действительно, все <em>печально</em>",
        "Если вы читаете этот отзыв, то знайте, что мастер <em>плохо</em> подходит к своему делу."
      ]
    }
  },
  {
    "_index" : "salon_master",
    "_type" : "Master",
    "_id" : "SP09",
    "_score" : 3.2982922,
    "_source" : {
      "master_personal_data" : "Петровская Ника Кирилловна"
    },
    "highlight" : {
      "reviews" : [
        "Работа выполнена <em>ужасно</em>.",
        "Если вы читаете этот отзыв, то знайте, что мастер <em>плохо</em> подходит к своему делу."
      ]
    }
  },
  {
    "_index" : "salon_master",
    "_type" : "Master",
    "_id" : "SP10",
    "_score" : 3.1171489,
    "_source" : {

```

```

    "master_personal_data" : "Ильинский Матвей Эмирович"
  },
  "highlight" : {
    "reviews" : [
      "Извините, но это <em>помойка</em>, а не студия маникюра.То ли мне не повезло, то ли, действительно, все <em>печально</em>"
    ]
  }
},
{
  "_index" : "salon_master",
  "_type" : "Master",
  "_id" : "SP13",
  "_score" : 2.8708844,
  "_source" : {
    "master_personal_data" : "Губанов Лев Максимович"
  },
  "highlight" : {
    "reviews" : [
      "Если вы читаете этот отзыв, то знайте, что мастер <em>плохо</em> подходит к своему делу.",
      "Второй день хожу вся <em>расстроенная</em> от того, как не качественно подстригли."
    ]
  }
},
{
  "_index" : "salon_master",
  "_type" : "Master",
  "_id" : "SP07",
  "_score" : 2.8092537,
  "_source" : {
    "master_personal_data" : "Горбачева Валерия Артемьевна"
  },
  "highlight" : {
    "reviews" : [

```

"Извините, но это помойка, а не студия маникюра.То ли мне не повезло, то ли, действительно, все печально"

```
]
}
},
{
  "_index" : "salon_master",
  "_type" : "Master",
  "_id" : "SP18",
  "_score" : 1.9254131,
  "_source" : {
    "master_personal_data" : "Майорова Екатерина Дмитриевна"
  },
  "highlight" : {
    "reviews" : [
      "Работа выполнена <em>ужасно</em>."
    ]
  }
},
{
  "_index" : "salon_master",
  "_type" : "Master",
  "_id" : "SP12",
  "_score" : 1.8753293,
  "_source" : {
    "master_personal_data" : "Покровский Даниил Владиславович"
  },
  "highlight" : {
    "reviews" : [
      "Первый раз в жизни пишу негативный отзыв, но тут уже нет предела
      <em>наглости</em>, а я наглых людей не люблю."
    ]
  }
}
]
}}
```

ПРИЛОЖЕНИЕ Б

Исходные коды программ

Листов 33

1 Программа маппинга и индексации документов типа «Процедура» и

«Мастер»

```
import json
from elasticsearch import Elasticsearch

client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
index_1= "salon_procedure"
index_2 = "salon_master"

#client.indices.delete(index=index_1)
if client.indices.exists(index=index_1):
    print("Recreate " + index_1 + " index")
    client.indices.delete(index=index_1)
    client.indices.create(index=index_1)
else:
    print("Create " + index_1)
    client.indices.create(index=index_1)

if client.indices.exists(index=index_2):
    print("Recreate " + index_2 + " index")
    client.indices.delete(index=index_2)
    client.indices.create(index=index_2)
else:
    print("Create " + index_2)
    client.indices.create(index=index_2)

Salon_Settings = {
    "analysis" : {
        "filter": {
            "russian_stop_words": {
                "type": "stop",
                "stopwords": "_russian_"
            },
            "filter_ru_sn": {
                "type": "snowball",
                "language": "Russian"
            }
        },
        "analyzer":
        {
            "analitic_for_ru":
            {
                "type": "custom",
                "tokenizer": "standard",
                "filter": [
                    "lowercase",
                    "russian_stop_words",
                    "filter_ru_sn"
                ]
            }
        }
    }
}
```

```

    }
  }
}

```

```

client.indices.close(index=index_1)
client.indices.put_settings(index=index_1, body=Salon_Settings)
client.indices.close(index=index_2)
client.indices.put_settings(index=index_2, body=Salon_Settings)

client.indices.open(index=index_1)

```

```

ProcedureMapping = {
  "properties":{
    "id_of_client": {
      "type": "text",
      "fielddata": True
    },
    "client_age": {
      "type": "integer"
    },
    "client_personal_data":{
      "type": "text",
      "analyzer":"analitic_for_ru",
      "fielddata": True
    },
    "id_of_procedure": {
      "type": "text",
      "fielddata": True
    },
    "date_of_procedure": {
      "type": "date",
      "format": "yyyy-MM-dd"
    },
    "price": {
      "type": "integer"
    },
    "id_of_specialist": {
      "type": "text",
      "fielddata": True
    },
    "service": {
      "type": "text",
      "fielddata": True,
      "analyzer":"analitic_for_ru",
      "search_analyzer":"analitic_for_ru"
    },
    "cosmetics": {
      "type": "text",
      "analyzer":"analitic_for_ru",
      "fielddata": True,

```



```

    }
}

```

```

client.indices.put_mapping(index=index_1,
                           doc_type="Procedure",
                           include_type_name="true",
                           body=ProcedureMapping)

```

```

client.indices.open(index="salon_procedure")

```

```

with open("/home/hadoopuser/js_files/Procedure.json", 'r') as
file_Procedure:
    Procedure_data = json.load(file_Procedure)

```

```

for data in Procedure_data:
    try:
        client.index(index=data["index"],
                     doc_type=data["doc_type"],
                     id=data["id"],
                     body=data["body"]
                     )
    except Exception as e:
        print(e)
print("Procedure_indexed")

```

```

client.indices.open(index=index_2)

```

```

MasterMapping = {
    "properties":{
        "specialisation": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"
        },
        "work_experience": {
            "type": "integer"
        },
        "master_personal_data": {
            "type": "text",
            "fielddata": True,
            "analyzer":"analitic_for_ru",
            "search_analyzer":"analitic_for_ru"
        },
        "schedule": {

```

```

        "type": "text",
        "fielddata": True,
        "analyzer": "analitic_for_ru",
        "search_analyzer": "analitic_for_ru"
    },
    "reviews": {
        "type": "text",
        "fielddata": True,
        "analyzer": "analitic_for_ru",
        "search_analyzer": "analitic_for_ru"
    }
}

client.indices.put_mapping(index=index_2,
                           doc_type="Master",
                           include_type_name="true",
                           body=MasterMapping)

client.indices.open(index="salon_master")

with open("/home/hadoopuser/js_files/Master.json", 'r') as file_Master:
    data_Master = json.load(file_Master)

for data in data_Master:
    try:
        client.index(index=data["index"],
                     doc_type=data["doc_type"],
                     id=data["id"],
                     body=data["body"]
                     )
    except Exception as e:
        print(e)
print("Master_indexed")

```

2 Тело первого запроса к кластеру ES

```
GET salon_procedure/_search
{
  "size": 0,
  "aggregations": {
    "year_period": {
      "date_histogram": {
        "field": "date_of_procedure",
        "calendar_interval": "year",
        "format": "yyyy-MM-dd"
      },
      "aggregations": {
        "Specialist": {
          "terms": {
            "field": "id_of_specialist",
            "order": {
              "_key": "asc"
            }
          },
          "aggregations": {
            "summary_price": {
              "sum": {
                "field": "price"
              }
            }
          }
        }
      }
    }
  }
}
```

3 Тело второго запроса к кластеру ES

```
GET salon_master/_search
{
  "query": {
    "simple_query_string": {
      "query": "кидает пострадало расстроенный наглость опаздал помойка  
грязно ужасно плохо печально"
    }
  },
  "_source": [
    "master_personal_data"
  ],
  "highlight": {
    "fields": {
      "reviews": {}
    }
  }
}
```

4 Импорт данных в базу данных Neo4j

```
from elasticsearch import Elasticsearch
from py2neo import Graph, Node, Relationship

client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])

indexName1 = "salon_procedure"
indexName2 = "salon_master"

graph_db = Graph("bolt://localhost:7687", auth=("neo4j", "matvienko"))

graph_db.delete_all()


#array of procedure
procedure_array = {
    "size": 1000,
    "query": {
        "match_all": {}
    }
}

#array of specialists
master_array = {
    "size": 1000,
    "query": {
        "match_all": {}
    }
}

result_procedure = client.search(index = indexName1, body =
procedure_array)
result_master = client.search(index = indexName2, body = master_array)

# cycle for my_node
for my_node in result_procedure ['hits']['hits']:
    Procedure_Node = Node("procedure",
        Client_id = my_node['_source']['id_of_client'],
        Master_id=my_node['_source']['id_of_specialist'],
        Service = my_node['_source']['service'],
        Cosmetics=my_node['_source']['cosmetics'],
        Price = my_node['_source']['price'],
        Date_of_procedure=my_node['_source']['date_of_procedure'
    ])
    try:
        Client_Node = graph_db.nodes.match("Client",
            Client_id = my_node['_source']['id_of_client']).first()
        if Client_Node==None:
```

```

#Create node for Client
Client_Node = Node("Client",
    Client_id=my_node['_source']['id_of_client'],
    Client_personal_data=my_node['_source']['client_personal_data'],
    Client_age=my_node['_source']['client_age'])
graph_db.create(Client_Node)
except Exception as e:
    print(e)
# Create cycle for Master node and Link
try:
    for master in result_master['hits']['hits']:
        Master_Node = graph_db.nodes.match("Master",
            Master_id=Master['_id']).first()
        if Master_Node==None:
            Master_Node = Node("Master",
                Master_id = master['_id'],
                Specialization = master['_source']['specialisation'],
                Work_experience = master['_source']['work_experience'],
                Rewiews = master['_source']['rewiews'])
            graph_db.create(Master_Node)

        if master['_id'] == my_node['_source']['Master_id']:
            Link_to_master = Relationship( Procedure_Node,
                "connect_master", Master_Node,
                Date_begin=my_node['_source']['date_of_procedure'])
            graph_db.create(Link_to_Master)

            Link_to_procedure = Relationship(Client_Node,
                "connect_procedure", Procedure_Node,
                Date_begin=my_node['_source']['date_of_procedure'])
            graph_db.create(Link_to_procedure)

except Exception as e:
    print(e)

```

5 Тело запроса запроса к базе данных Neo4j

```

MATCH p=(c:Client)-[r:connect_Client_Master]->(m:Master)
WITH m, sum(toInteger(r.Price)) as master_sum
ORDER BY master_sum desc
RETURN m, master_sum
LIMIT 1

```

6 Программа создания CSV-файлов с таблицами

```
from elasticsearch import Elasticsearch
from pyspark.sql import SparkSession
from pyspark.sql.types import *
from datetime import datetime

client = Elasticsearch([{"host": "127.0.0.1", "port": 9200}])
indexName1 = "salon_procedure"
indexName2 = "salon_master"

sparkSession = SparkSession.builder.appName("spark-csv").getOrCreate()
searchBody = {
    "size": 60,
    "query": {
        "match_all": {}
    }
}

result_procedure = client.search(index=indexName1,
body=searchBody)['hits']['hits']
result_master = client.search(index=indexName2,
body=searchBody)['hits']['hits']
result_client = client.search(index=indexName1,
body=searchBody)['hits']['hits']

# schema for procedure, master and client
Schema_for_procedure = StructType([
    StructField("procedure_id", StringType(), True),
    StructField("client_id", StringType(), True),
    StructField("procedure_date", DateType(), True),
    StructField("master_id", StringType(), True),
    StructField("price", IntegerType(), True),
    StructField("service", StringType(), True),
    StructField("cosmetics", StringType(), True)
])

Schema_for_master = StructType([
    StructField("master_id", StringType(), True),
    StructField("master_specialization", StringType(), True),
    StructField("master_experience", IntegerType(), True),
    StructField("master_personal_data", StringType(), True),
    StructField("schedule", StringType(), True),
    StructField("reviews", StringType(), True)
])

Schema_for_client = StructType([
    StructField("client_id", StringType(), True),
    StructField("client_personal_data", StringType(), True),
    StructField("client_age", IntegerType(), True)])
# procedure table
Table_for_procedure_data = []
```

```

for procedure in result_procedure:
    Table_for_procedure_data.append((
        procedure['_source']['id_of_procedure'],
        procedure['_source']['id_of_client'],
        datetime.strptime(procedure['_source']['date_of_procedure'], "%Y-
%m-%d"),
        procedure['_source']['id_of_specialist'],
        int(procedure['_source']['price']),
        procedure['_source']['service'],
        procedure['_source']['cosmetics']
    ))

# master table
Table_for_master_data = []
for master in result_master:
    Table_for_master_data.append((
        master['_id'],
        master['_source']['specialisation'],
        int(master['_source']['work_experience']),
        master['_source']['master_personal_data'],
        master['_source']['schedule'],
        master['_source']['reviews']
    ))

# client table
Table_for_client_data = []
for client in result_client:
    Table_for_client_data.append((
        client['_source']['id_of_client'],
        client['_source']['client_personal_data'],
        int(client['_source']['client_age'])
    ))

#Creating data frame
Table_for_procedure =
sparkSession.createDataFrame(Table_for_procedure_data,
Schema_for_procedure)
Table_for_master = sparkSession.createDataFrame(Table_for_master_data,
Schema_for_master)
Table_for_client = sparkSession.createDataFrame(Table_for_client_data,
Schema_for_client)

# making csv
Table_for_procedure.write.csv(path='hdfs://localhost:9000/procedure.csv',
mode='overwrite', header=True)
Table_for_master.write.csv(path='hdfs://localhost:9000/master.csv',
mode='overwrite', header=True)
Table_for_client.write.csv(path='hdfs://localhost:9000/client.csv',
mode='overwrite', header=True)

```

```

df_load = sparkSession.read.load(path='hdfs://localhost:9000/procedure.csv',
format='csv', sep=',', inferSchema="true", header="true")
df_load.show()
df_load = sparkSession.read.load(path='hdfs://localhost:9000/master.csv',
format='csv', sep=',', inferSchema="true", header="true")
df_load.show()
df_load = sparkSession.read.load(path='hdfs://localhost:9000/client.csv',
format='csv', sep=',', inferSchema="true", header="true")
df_load.show()

```

7 Программа запроса в Spark

```

#!/usr/bin/env python
# -*- coding: utf-8 -*- from pyspark.sql import SparkSession

from pyspark.sql import *

sparkSession=SparkSession.builder.appName("Python Spark SQL basic
example").config("spark.some.co>

Procedure_Table = sparkSession.read.load(path='hdfs://localhost:9000/procedure.csv',
format='csv')
Master_Table = sparkSession.read.load(path='hdfs://localhost:9000/master.csv',
format='csv', sep=>
Client_Table = sparkSession.read.load(path='hdfs://localhost:9000/client.csv',
format='csv', sep=>

Procedure_Table.registerTempTable("procedure")
Master_Table.registerTempTable("master")
Client_Table.registerTempTable("client")

df = sparkSession.sql("SELECT DISTINCT procedure.master_id,
COUNT(procedure.procedure_id) FROM pr>

input('Ctrl C')

```


8 Программа генерации файла JSON-Procedure

```
JG.repeat(40, 40, {  
  index: "salon_master",  
  doc_type: "Master",  
  id: "SP",  
  body: {  
    specialisation: _uniq(JG.repeat(4, JG.random("Свадебные и вечерние прически",  
"Завивка и выпрямление волос",  
"Кератиновое разглаживание",  
"Долговременная укладка",  
"Брондирование и колорирование",  
"Мелирование и блондирование",  
"Каутеризация",  
"Ламинирование волос",  
"Амбре-окрашивание",  
"Укладки и тонирование",  
"Цветовой дизайн волос",  
"Женская, мужская и детская стрижка",  
"Уход за волосами и лечение волос",  
"Наращивание и окраска волос",  
"Биопилинг кожи головы",  
"Плетение косичек",  
"Маникюр Аппаратный",  
"Маникюр классический мужской и детский",  
"Маникюр Европейский",  
"Долгосрочное покрытие гель-лаками (ши-лаками)",  
"Горячие обертывания",  
"Наращивание и коррекция ногтей гелем",  
"Парафинотерапия",  
"Артдизайн",  
"Спа-маникюр",  
"Отбеливание",  
"Укрепление натуральных ногтей",  
"Уход за руками",  
"Массаж рук",
```

"Педикюр Классический",
 "Аппаратный",
 "Парафинотерапия",
 "Спа-педикюр",
 "Покрытие лаком",
 "Покрытие гель-лаком",
 "Пилинг стоп",
 "Брашинг",
 "Вакуумная чистка лица",
 "Ультразвуковая чистка лица",
 "Дарсонваль",
 "Гальванизация",
 "Электрофорез",
 "Ионофорез",
 "Дезинкрустация",
 "Лимфодренаж"))),
 work_experience:`\${JG.integer(1, 40)}`,
 master_personal_data() {
 i=JG.integer(0, 199);
 person_mas=["Новикова Валерия Константиновна",
 "Соколова Диана Ивановна",
 "Черных Мария Александровна",
 "Мельникова Ульяна Андреевна",
 "Сергеева Александра Савельевна",
 "Захарова Софья Дмитриевна",
 "Александрова Мария Николаевна",
 "Трофимова Алиса Львовна",
 "Трифоновна Екатерина Тимуровна",
 "Мельникова Анна Владиславовна",
 "Некрасова Ника Александровна",
 "Пономарева Полина Александровна",
 "Лебедева Стефания Максимовна",
 "Тимофеева Мирослава Тимофеевна",
 "Пономарева Елизавета Данисовна",
 "Семина Полина Руслановна",

"Терехова Екатерина Михайловна",
"Гришина Виктория Кирилловна",
"Иванова Кира Никитична",
"Горбачева Полина Тимуровна",
"Козлова Анастасия Тимуровна",
"Яковлева Алиса Вячеславовна",
"Назарова Аделя Константиновна",
"Еремина Кира Максимовна",
"Горбачева Валерия Артемьевна",
"Попова Анастасия Львовна",
"Осипова Виктория Данииловна",
"Фадеева Анна Богдановна",
"Фокина Елизавета Данииловна",
"Попова Яна Адамовна",
"Демидова Фатима Марковна",
"Климова Анна Александровна",
"Яковлева Мелания Степановна",
"Егорова София Владимировна",
"Балашова Кристина Марковна",
"Трофимова Софья Олеговна",
"Алексеева Алиса Данииловна",
"Герасимова Виктория Гордеевна",
"Воронова Карина Александровна",
"Иванова Каролина Адамовна",
"Фомина Варвара Михайловна",
"Субботина Анна Александровна",
"Соколова Мария Артёмовна",
"Попова Полина Фёдоровна",
"Петровская Ника Кирилловна",
"Овсянникова Марина Алексеевна",
"Назарова Арина Артёмовна",
"Чумакова Ева Ярославовна",
"Михайлова Алина Михайловна",
"Быкова Алиса Максимовна",
"Кузнецова Анна Кирилловна",

"Березина Наталья Романовна",
"Серова Алёна Дмитриевна",
"Колосова Милана Васильевна",
"Макарова Валерия Петровна",
"Ковалева Мирослава Львовна",
"Калашникова Мария Павловна",
"Курочкина Ульяна Антоновна",
"Гаврилова Мария Сергеевна",
"Емельянова Арина Кирилловна",
"Егорова Алиса Ивановна",
"Суворова Анна Тимофеевна",
"Кошелева Кира Александровна",
"Гусева Софья Георгиевна",
"Морозова Полина Артёмовна",
"Ильина Зоя Мироновна",
"Филатова Ульяна Елисеевна",
"Пахомова Яна Романовна",
"Баранова Анастасия Артуровна",
"Терентьева Александра Дмитриевна",
"Мальцева Анна Ярославовна",
"Воронина Мария Михайловна",
"Пахомова Арина Вячеславовна",
"Куликова Юлия Дмитриевна",
"Лапшина Анна Евгеньевна",
"Титова Сафия Дмитриевна",
"Михайлова Диана Арсентьевна",
"Кондратова Александра Максимовна",
"Козлова Ирина Никитична",
"Яковлева Василиса Андреевна",
"Морозова Алина Романовна",
"Демьянова Алиса Викторовна",
"Шаповалова Мария Александровна",
"Болдырева Софья Ивановна",
"Кондратьева Алия Тимофеевна",
"Савицкая Александра Егоровна",

"Смирнова Полина Тимофеевна",
"Данилова Виктория Матвеевна",
"Лаптева Ксения Артемьевна",
"Романова Маргарита Максимовна",
"Крюкова Диана Савельевна",
"Усова Мария Никитична",
"Архипова Алиса Богдановна",
"Матвеева Валерия Максимовна",
"Сидорова Арина Арсентьевна",
"Кононова Евгения Максимовна",
"Филиппова Анна Владимировна",
"Белова Анна Максимовна",
"Зорина Мария Артёмовна",
"Козлова Виктория Дмитриевна",
"Баранова Амалия Данииловна",
"Пахомов Александр Матвеевич",
"Макарова Зоя Владиславовна",
"Туманов Николай Максимович",
"Хохлов Ярослав Александрович",
"Потапов Владимир Александрович",
"Колпакова Арина Артёмовна",
"Зимин Руслан Михайлович",
"Белова Варвара Антоновна",
"Рыбакова Злата Яковлевна",
"Осипов Арсений Тихонович",
"Киселев Пётр Максимович",
"Попова Елизавета Вячеславовна",
"Широков Максим Ярославович",
"Ларина Милана Николаевна",
"Губанов Лев Максимович",
"Смирнова Елизавета Дмитриевна",
"Орлов Александр Арсентьевич",
"Кузнецов Михаил Владимирович",
"Глебова Полина Данииловна",
"Майорова Екатерина Дмитриевна",

"Рыбаков Степан Андреевич",
"Матвеева Василиса Евгеньевна",
"Медведев Тимофей Львович",
"Севастьянов Кирилл Эмирович",
"Белоусова Елизавета Артёмовна",
"Орлова Дарина Никитична",
"Щукина Ксения Фёдоровна",
"Николаев Михаил Антонович",
"Шилова Нина Константиновна",
"Максимова Дарья Фёдоровна",
"Демидова Светлана Дмитриевна",
"Зубкова Софья Витальевна",
"Корнеев Антон Никитич",
"Чернова Александра Ильинична",
"Игнатов Даниил Александрович",
"Софронов Георгий Тимурович",
"Ильинский Матвей Эмирович",
"Кузьмина Полина Владимировна",
"Никитина Софья Павловна",
"Назаров Виктор Денисович",
"Грачева Ульяна Арсентьевна",
"Иванов Илья Ярославович",
"Мухин Владимир Святославович",
"Тарасова Александра Максимовна",
"Грачев Давид Никитич",
"Попов Арсен Владиславович",
"Овсянникова София Сергеевна",
"Иванова Елизавета Михайловна",
"Устинова Алина Александровна",
"Максимов Даниил Алиевич",
"Аникина Полина Михайловна",
"Логинова Алиса Михайловна",
"Чернышева Светлана Дмитриевна",
"Нефедова Анастасия Арсентьевна",
"Антонова Арина Николаевна",

"Никитина Софья Глебовна",
"Плотникова Екатерина Тимофеевна",
"Евдокимов Артём Глебович",
"Панов Мирон Яковлевич",
"Богданов Михаил Тимофеевич",
"Гусева Ксения Елисеевна",
"Николаев Михаил Александрович",
"Зайцева Виктория Ивановна",
"Ефимова Сафия Тимофеевна",
"Егоров Дмитрий Давидович",
"Головин Андрей Григорьевич",
"Волошина Надежда Максимовна",
"Курочкин Алексей Иванович",
"Смирнова Валерия Романовна",
"Агафонова Мирослава Фёдоровна",
"Постников Максим Стефанович",
"Зубова Татьяна Михайловна",
"Львова Аделина Дмитриевна",
"Леонов Матвей Павлович",
"Колесов Лев Ильич",
"Федоров Иван Андреевич",
"Гусев Михаил Владимирович",
"Головина Диана Максимовна",
"Меркулова Арина Сергеевна",
"Семенова Елизавета Георгиевна",
"Цветков Кирилл Сергеевич",
"Смирнов Роман Андреевич",
"Сергеева Вероника Никитична",
"Покровский Даниил Владиславович",
"Харитонова Оливия Ярославовна",
"Киреева Дарья Вячеславовна",
"Шестаков Максим Александрович",
"Худякова Мария Дмитриевна",
"Степанова Есения Ильинична",
"Муратов Кирилл Матвеевич",

"Васильев Григорий Глебович",
 "Горохов Андрей Семёнович",
 "Семина Анна Игоревна",
 "Суханов Елисей Ильич",
 "Горелова Любовь Тимофеевна",
 "Белоусов Пётр Маркович",
 "Ковалева Николь Романовна",
 "Попова Алия Андреевна",
 "Агеев Фёдор Даниилович"];

return (person_mas[i]);

},

schedule(){

i=JG.integer(0, 11);

schedule_mas = ["Пн, Вт, Чт, Пт 10:00-15:00",

"Пн, Вт, Сб, Вс 10:00-15:00",

"Ср, Пт 10:00-15:00",

"Чт,Пт,Сб 10:00-15:00",

"Пн, Ср, Вс 10:00-15:00",

"Сб, Вс 10:00-15:00",

"Пн, Вт, Чт, Пт 15:00-20:00",

"Пн, Вт, Сб, Вс 15:00-20:00",

"Ср, Пт 15:00-20:00",

"Чт,Пт,Сб 15:00-20:00",

"Пн, Ср, Вс 15:00-20:00",

"Сб, Вс 15:00-20:00"];

master_sched = schedule_mas[i];

return(master_sched);

},

reviews:_.uniq(JG.repeat(5,JG.random(

"Наконец моя мечта сбылась и я сделала Лифтинг! Это не дешёвая процедура, но результат того стоит, я в восторге!",

"Постоянно хожу в салон на маникюр и педикюр к этому мастеру. Мастер все делает аккуратно, доброжелательно.",

"Люди будьте бдительны",

"Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",

"Прекрасный врач-косметолог, решает проблемы на раз-два. Результатом всегда довольна.",

"За стрижку со скидкой спасибо от души!",

"Делаю брови, ресницы, брови, ламинировала волосы. Качеством довольна. Минус есть - принимают позже назначенного времени. Есть куда стремиться.",

"Была в Клинике красоты впервые, мастер",

"Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка.",

"Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!",

"Была первый раз вернусь обязательно еще.",

"Спасибо большое мастеру за очаровательное каре, золотые руки, приятный собеседник, восхищаюсь образом который мне подарили!",

"Первый раз делала чистку лица, и уже с первого раза виден результат. Час процедуры прошел незаметно в приятной позитивной обстановке, что не маловажно) Приду скоро еще!",

"Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",

"Спасибо огромное мастеру за чудесный маникюр и педикюр !!! Все просто супер !!!",

"Всегда выхожу от вас с позитивным настроением!",

"Второй день хожу вся расстроенная от того, как не качественно подстригли. Каскад, где одна прядь короче другой на пару сантиметров и профилирована раз 5, а с другой пару.",

"Первый раз в жизни пишу негативный отзыв, но тут уже нет предела наглости, а я наглых людей не люблю. Записалась на педикюр, но видите ли мастер не успевал сделать покрытие, которое я хотела. мол её не предупредили и она время не оставила. на что администратор после сказала, что она записала с покрытием, а вы мастеру просто не понравились.",

"Михаил внимательный и приятный в общении мастер. Однако на выходе я получила немного не тот цвет, который хотела.",

"Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу.",

"Извините, но это помойка, а не студия маникюра. То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",

"Записалась на коррекцию. Зайдя в кабинет, была сильно удивлена, насколько были грязные столы, и разбросан материал, неизвестного происхождения, и незнакомых брендов. Самое ужасное началось когда начали делать покрытие, небрежно, гель затёк под кутикулу, образовались бугры, ногти выглядели ужасно, разной формы и длины.",

"Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново. Результат: испорчены ногти, настроение.",

"Нестерильно. Мастер разговаривает по телефону во время маникюра и после разговора не обрабатывает руки.",

"Выбор гель-лаков потрясающий, глаза разбежались. А цены очень порадовали кошелек))))",

"Работа выполнена хорошо",

"Не рекомендую",

"Прекрасный мастер!",

"Никогда больше не пойду к этому мастеру. Работа выполнена ужасно.",

"Спасибо за отличное настроение и классную работу.",

"Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти."))),

},

});

9 Программа генерации файла JSON-Master

```
JG.repeat(20, 20, {
  index: "salon_master",
  doc_type:"Master",
  id: "SP",
  body: {

    specialisation:_uniq(JG.repeat(4, JG.random("Свадебные и вечерние прически",
"Завивка и выпрямление волос",
"Кератиновое разглаживание",
"Долговременная укладка",
"Брондирование и колорирование",
"Мелирование и блондирование",
"Каутеризация",
"Ламинирование волос",
"Амбре-окрашивание",
"Укладки и тонирование",
"Цветовой дизайн волос",
"Женская, мужская и детская стрижка",
"Уход за волосами и лечение волос",
"Наращивание и окраска волос",
"Биопилинг кожи головы",
"Плетение косичек",
"Маникюр Аппаратный",
"Маникюр классический мужской и детский",
"Маникюр Европейский",
"Долгосрочное покрытие гель-лаками (ши-лаками)",
"Горячие обертывания",
"Наращивание и коррекция ногтей гелем",
"Парафинотерапия",
"Артдизайн",
"Спа-маникюр",
"Отбеливание",
"Укрепление натуральных ногтей",
"Уход за руками",
```

"Массаж рук",
 "Педикюр Классический",
 "Аппаратный",
 "Парафинотерапия",
 "Спа-педикюр",
 "Покрытие лаком",
 "Покрытие гель-лаком",
 "Пилинг стоп",
 "Брашинг",
 "Вакуумная чистка лица",
 "Ультразвуковая чистка лица",
 "Дарсонваль",
 "Гальванизация",
 "Электрофорез",
 "Ионофорез",
 "Дезинкрустация",
 "Лимфодренаж"))),
 work_experience:`\${JG.integer(1, 40)}`,
 master_personal_data() {
 i=JG.integer(0, 199);
 person_mas=["Новикова Валерия Константиновна",
 "Соколова Диана Ивановна",
 "Черных Мария Александровна",
 "Мельникова Ульяна Андреевна",
 "Сергеева Александра Савельевна",
 "Захарова Софья Дмитриевна",
 "Александрова Мария Николаевна",
 "Трофимова Алиса Львовна",
 "Трифопова Екатерина Тимуровна",
 "Мельникова Анна Владиславовна",
 "Некрасова Ника Александровна",
 "Пономарева Полина Александровна",
 "Лебедева Стефания Максимовна",
 "Тимофеева Мирослава Тимофеевна",
 "Пономарева Елизавета Данисовна",

"Семина Полина Руслановна",
"Терехова Екатерина Михайловна",
"Гришина Виктория Кирилловна",
"Иванова Кира Никитична",
"Горбачева Полина Тимуровна",
"Козлова Анастасия Тимуровна",
"Яковлева Алиса Вячеславовна",
"Назарова Аделя Константиновна",
"Еремина Кира Максимовна",
"Горбачева Валерия Артемьевна",
"Попова Анастасия Львовна",
"Осипова Виктория Данииловна",
"Фадеева Анна Богдановна",
"Фокина Елизавета Данииловна",
"Попова Яна Адамовна",
"Демидова Фатима Марковна",
"Климова Анна Александровна",
"Яковлева Мелания Степановна",
"Егорова София Владимировна",
"Балашова Кристина Марковна",
"Трофимова Софья Олеговна",
"Алексеева Алиса Данииловна",
"Герасимова Виктория Гордеевна",
"Воронова Карина Александровна",
"Иванова Каролина Адамовна",
"Фомина Варвара Михайловна",
"Субботина Анна Александровна",
"Соколова Мария Артёмовна",
"Попова Полина Фёдоровна",
"Петровская Ника Кирилловна",
"Овсянникова Марина Алексеевна",
"Назарова Арина Артёмовна",
"Чумакова Ева Ярославовна",
"Михайлова Алина Михайловна",
"Быкова Алиса Максимовна",

"Кузнецова Анна Кирилловна",
"Березина Наталья Романовна",
"Серова Алёна Дмитриевна",
"Колосова Милана Васильевна",
"Макарова Валерия Петровна",
"Ковалева Мирослава Львовна",
"Калашникова Мария Павловна",
"Курочкина Ульяна Антоновна",
"Гаврилова Мария Сергеевна",
"Емельянова Арина Кирилловна",
"Егорова Алиса Ивановна",
"Суворова Анна Тимофеевна",
"Кошелева Кира Александровна",
"Гусева Софья Георгиевна",
"Морозова Полина Артёмовна",
"Ильина Зоя Мироновна",
"Филатова Ульяна Елисеевна",
"Пахомова Яна Романовна",
"Баранова Анастасия Артуровна",
"Терентьева Александра Дмитриевна",
"Мальцева Анна Ярославовна",
"Воронина Мария Михайловна",
"Пахомова Арина Вячеславовна",
"Куликова Юлия Дмитриевна",
"Лапшина Анна Евгеньевна",
"Титова Сафия Дмитриевна",
"Михайлова Диана Арсентьевна",
"Кондратова Александра Максимовна",
"Козлова Ирина Никитична",
"Яковлева Василиса Андреевна",
"Морозова Алина Романовна",
"Демьянова Алиса Викторовна",
"Шаповалова Мария Александровна",
"Болдырева Софья Ивановна",
"Кондратьева Алия Тимофеевна",

"Савицкая Александра Егоровна",
"Смирнова Полина Тимофеевна",
"Данилова Виктория Матвеевна",
"Лаптева Ксения Артемьевна",
"Романова Маргарита Максимовна",
"Крюкова Диана Савельевна",
"Усова Мария Никитична",
"Архипова Алиса Богдановна",
"Матвеева Валерия Максимовна",
"Сидорова Арина Арсентьевна",
"Кононова Евгения Максимовна",
"Филиппова Анна Владимировна",
"Белова Анна Максимовна",
"Зорина Мария Артёмовна",
"Козлова Виктория Дмитриевна",
"Баранова Амалия Данииловна",
"Пахомов Александр Матвеевич",
"Макарова Зоя Владиславовна",
"Туманов Николай Максимович",
"Хохлов Ярослав Александрович",
"Потапов Владимир Александрович",
"Колпакова Арина Артёмовна",
"Зимин Руслан Михайлович",
"Белова Варвара Антоновна",
"Рыбакова Злата Яковлевна",
"Осипов Арсений Тихонович",
"Киселев Пётр Максимович",
"Попова Елизавета Вячеславовна",
"Широков Максим Ярославович",
"Ларина Милана Николаевна",
"Губанов Лев Максимович",
"Смирнова Елизавета Дмитриевна",
"Орлов Александр Арсентьевич",
"Кузнецов Михаил Владимирович",
"Глебова Полина Данииловна",

"Майорова Екатерина Дмитриевна",
"Рыбаков Степан Андреевич",
"Матвеева Василиса Евгеньевна",
"Медведев Тимофей Львович",
"Севастьянов Кирилл Эмирович",
"Белоусова Елизавета Артёмовна",
"Орлова Дарина Никитична",
"Щукина Ксения Фёдоровна",
"Николаев Михаил Антонович",
"Шилова Нина Константиновна",
"Максимова Дарья Фёдоровна",
"Демидова Светлана Дмитриевна",
"Зубкова Софья Витальевна",
"Корнеев Антон Никитич",
"Чернова Александра Ильинична",
"Игнатов Даниил Александрович",
"Софронов Георгий Тимурович",
"Ильинский Матвей Эмирович",
"Кузьмина Полина Владимировна",
"Никитина Софья Павловна",
"Назаров Виктор Денисович",
"Грачева Ульяна Арсентьевна",
"Иванов Илья Ярославович",
"Мухин Владимир Святославович",
"Тарасова Александра Максимовна",
"Грачев Давид Никитич",
"Попов Арсен Владиславович",
"Овсянникова София Сергеевна",
"Иванова Елизавета Михайловна",
"Устинова Алина Александровна",
"Максимов Даниил Алиевич",
"Аникина Полина Михайловна",
"Логинова Алиса Михайловна",
"Чернышева Светлана Дмитриевна",
"Нефедова Анастасия Арсентьевна",

"Антонова Арина Николаевна",
"Никитина Софья Глебовна",
"Плотникова Екатерина Тимофеевна",
"Евдокимов Артём Глебович",
"Панов Мирон Яковлевич",
"Богданов Михаил Тимофеевич",
"Гусева Ксения Елисеевна",
"Николаев Михаил Александрович",
"Зайцева Виктория Ивановна",
"Ефимова Сафия Тимофеевна",
"Егоров Дмитрий Давидович",
"Головин Андрей Григорьевич",
"Волошина Надежда Максимовна",
"Курочкин Алексей Иванович",
"Смирнова Валерия Романовна",
"Агафонова Мирослава Фёдоровна",
"Постников Максим Стефанович",
"Зубова Татьяна Михайловна",
"Львова Аделина Дмитриевна",
"Леонов Матвей Павлович",
"Колесов Лев Ильич",
"Федоров Иван Андреевич",
"Гусев Михаил Владимирович",
"Головина Диана Максимовна",
"Меркулова Арина Сергеевна",
"Семенова Елизавета Георгиевна",
"Цветков Кирилл Сергеевич",
"Смирнов Роман Андреевич",
"Сергеева Вероника Никитична",
"Покровский Даниил Владиславович",
"Харитоновна Оливия Ярославовна",
"Киреева Дарья Вячеславовна",
"Шестаков Максим Александрович",
"Худякова Мария Дмитриевна",
"Степанова Есения Ильинична",

```

"Муратов Кирилл Матвеевич",
"Васильев Григорий Глебович",
"Горохов Андрей Семёнович",
"Семина Анна Игоревна",
"Суханов Елисей Ильич",
"Горелова Любовь Тимофеевна",
"Белоусов Пётр Маркович",
"Ковалева Николь Романовна",
"Попова Алия Андреевна",
"Агеев Фёдор Даниилович"];
    return (person_mas[i]);
},
schedule(){
    i=JG.integer(0, 11);
    schedule_mas = ["Пн, Вт, Чт, Пт 10:00-15:00",
"Пн, Вт, Сб, Вс 10:00-15:00",
"Ср, Пт 10:00-15:00",
"Чт,Пт,Сб 10:00-15:00",
"Пн, Ср, Вс 10:00-15:00",
"Сб, Вс 10:00-15:00",
"Пн, Вт, Чт, Пт 15:00-20:00",
"Пн, Вт, Сб, Вс 15:00-20:00",
"Ср, Пт 15:00-20:00",
"Чт,Пт,Сб 15:00-20:00",
"Пн, Ср, Вс 15:00-20:00",
"Сб, Вс 15:00-20:00"];
    master_sched = schedule_mas[i];
    return(master_sched);
},
reviews:_.uniq(JG.repeat(5,JG.random(
    "Наконец моя мечта сбылась и я сделала Лифтинг! Это не дешёвая процедура, но
результат того стоит, я в восторге!",
    "Постоянно хожу в салон на маникюр и педикюр к этому мастеру. Мастер все делает
аккуратно, доброжелательно.",
    "Люди будьте бдительны",

```

"Мне сделали отличное кератиновое выпрямление. Волосы теперь ухоженные прямые и блестящие, в общем ожидания оправданы полностью!",

"Прекрасный врач-косметолог, решает проблемы на раз-два. Результатом всегда довольна.",

"За стрижку со скидкой спасибо от души!",

"Делаю брови, ресницы, брови, ламинировала волосы. Качеством довольна. Минус есть-принимают позже назначенного времени. Есть куда стремиться.",

"Была в Клинике красоты впервые, мастер",

"Добрый день! Хочу выразить благодарность мастеру. Внимательно отнеслась ко всем моим пожеланиям относительно выбора цвета для волос. Осталась очень довольна результатом!!! Теперь я Ваша постоянная клиентка.",

"Спасибо огромное в очередной раз... моя красота (простите за нескромность) в ваших руках!",

"Была первый раз вернусь обязательно еще.",

"Спасибо большое мастеру за очаровательное каре, золотые руки, приятный собеседник, восхищаюсь образом который мне подарили!",

"Первый раз делала чистку лица, и уже с первого раза виден результат. Час процедуры прошел незаметно в приятной позитивной обстановке, что не маловажно) Приду скоро еще!",

"Огромное спасибо замечательному мастеру! Делает все очень аккуратно и быстро. Очень довольна.",

"Спасибо огромное мастеру за чудесный маникюр и педикюр !!! Все просто супер !!!!",

"Всегда выхожу от вас с позитивным настроением!",

"Второй день хожу вся расстроенная от того, как не качественно подстригли. Каскад, где одна прядь короче другой на пару сантиметров и профилирована раз 5, а с другой пару.",

"Первый раз в жизни пишу негативный отзыв, но тут уже нет предела наглости, а я наглых людей не люблю. Записалась на педикюр, но видите ли мастер не успевал сделать покрытие, которое я хотела. мол её не предупредили и она время не оставила. на что администратор после сказала, что она записала с покрытием, а вы мастеру просто не понравились.",

"Михаил внимательный и приятный в общении мастер. Однако на выходе я получила немного не тот цвет, который хотела.",

"Уберите из прайса укрепление гелем, если вы его делать не умеете! Не тратьте свое время и время своих клиентов! По поводу Шеллака сказать ничего не могу.",

"Извините, но это помойка, а не студия маникюра. То ли мне не повезло, то ли, действительно, все печально. Посетила салон... обстановка так себе 3 кв/м. Это было мое первое и последнее посещение. После обработки кожи вокруг ногтей заусенцы встали дыбом на следующий день.",

"Записалась на коррекцию. Зайдя в кабинет, была сильно удивлена, насколько были грязные столы, и разбросан материал, неизвестного происхождения, и незнакомых брендов. Самое ужасное началось когда начали делать покрытие, небрежно, гель затёк под кутикулу, образовались бугры, ногти выглядели ужасно, разной формы и длины.",

"Делала педикюр. Увидев конечный результат, я была шокирована и сообщила, что их теперь нужно будет переделывать заново. Результат: испорчены ногти, настроение.",

"Нестерильно. Мастер разговаривает по телефону во время маникюра и после разговора не обрабатывает руки.",

"Выбор гель-лаков потрясающий, глаза разбежались. А цены очень порадовали кошелек)))",

"Работа выполнена хорошо",

"Не рекомендую",

"Прекрасный мастер!",

"Никогда больше не пойдк этому мастеру. Работа выполнена ужасно.",

"Спасибо за отличное настроение и классную работу.",

"Если вы читаете этот отзыв, то знайте, что мастер плохо подходит к своему делу. Не советую к нему идти."))),

},

});

10 Физический план в мониторе Spark

== Physical Plan ==

AdaptiveSparkPlan (16)

+-- == Final Plan ==

CollectLimit (9)

+-- * HashAggregate (8)

+-- * HashAggregate (7)

+-- * HashAggregate (6)

+-- AQEShuffleRead (5)

+-- ShuffleQueryStage (4)

+-- Exchange (3)

+-- * HashAggregate (2)

+-- Scan csv (1)

+-- == Initial Plan ==

CollectLimit (15)

+-- HashAggregate (14)

+-- HashAggregate (13)

+-- HashAggregate (12)

+-- Exchange (11)

+-- HashAggregate (10)

+-- Scan csv (1)

(1) Scan csv

Output [2]: [procedure_id#16, master_id#19]

Batched: false

Location: InMemoryFileIndex [hdfs://localhost:9000/procedure.csv]

ReadSchema: struct<procedure_id:string,master_id:string>

(2) HashAggregate [codegen id : 1]

Input [2]: [procedure_id#16, master_id#19]

Keys [1]: [master_id#19]

Functions [1]: [partial_count(procedure_id#16)]

Aggregate Attributes [1]: [count#90L]

Results [2]: [master_id#19, count#91L]

(3) Exchange

Input [2]: [master_id#19, count#91L]

Arguments: hashpartitioning(master_id#19, 80), ENSURE_REQUIREMENTS, [id=#100]

(4) ShuffleQueryStage

Output [2]: [master_id#19, count#91L]

Arguments: 0

(5) AQEShuffleRead

Input [2]: [master_id#19, count#91L]

Arguments: coalesced

(6) HashAggregate [codegen id : 2]

Input [2]: [master_id#19, count#91L]

Keys [1]: [master_id#19]

Functions [1]: [count(procedure_id#16)]

Aggregate Attributes [1]: [count(procedure_id#16)#80L]

Results [2]: [master_id#19, count(procedure_id#16)#80L AS count(procedure_id)#81L]

(7) HashAggregate [codegen id : 2]

Input [2]: [master_id#19, count(procedure_id)#81L]

Keys [2]: [master_id#19, count(procedure_id)#81L]
Functions: []
Aggregate Attributes: []
Results [2]: [master_id#19, count(procedure_id)#81L]

(8) HashAggregate [codegen id : 2]
Input [2]: [master_id#19, count(procedure_id)#81L]
Keys [2]: [master_id#19, count(procedure_id)#81L]
Functions: []
Aggregate Attributes: []
Results [2]: [master_id#19, cast(count(procedure_id)#81L as string) AS count(procedure_id)#87]

(9) CollectLimit
Input [2]: [master_id#19, count(procedure_id)#87]
Arguments: 27

(10) HashAggregate
Input [2]: [procedure_id#16, master_id#19]
Keys [1]: [master_id#19]
Functions [1]: [partial_count(procedure_id#16)]
Aggregate Attributes [1]: [count#90L]
Results [2]: [master_id#19, count#91L]

(11) Exchange
Input [2]: [master_id#19, count#91L]
Arguments: hashpartitioning(master_id#19, 80), ENSURE_REQUIREMENTS, [id=#87]

(12) HashAggregate
Input [2]: [master_id#19, count#91L]
Keys [1]: [master_id#19]
Functions [1]: [count(procedure_id#16)]
Aggregate Attributes [1]: [count(procedure_id#16)#80L]
Results [2]: [master_id#19, count(procedure_id#16)#80L AS count(procedure_id)#81L]

(13) HashAggregate
Input [2]: [master_id#19, count(procedure_id)#81L]
Keys [2]: [master_id#19, count(procedure_id)#81L]
Functions: []
Aggregate Attributes: []
Results [2]: [master_id#19, count(procedure_id)#81L]

(14) HashAggregate
Input [2]: [master_id#19, count(procedure_id)#81L]
Keys [2]: [master_id#19, count(procedure_id)#81L]
Functions: []
Aggregate Attributes: []
Results [2]: [master_id#19, cast(count(procedure_id)#81L as string) AS count(procedure_id)#87]

(15) CollectLimit
Input [2]: [master_id#19, count(procedure_id)#87]
Arguments: 27

(16) AdaptiveSparkPlan
Output [2]: [master_id#19, count(procedure_id)#87]
Arguments: isFinalPlan=true

ПРИЛОЖЕНИЕ В

Копии листов графической части

Листов 12

В графическую часть курсового проекта входят следующие листы:

- 1) Задание на курсовой проект;
- 2) Индексация документов Elasticsearch. Маппинг и анализатор;
- 3) Индексация документов Elasticsearch. Алгоритмы индексации;
- 4) Elasticsearch. Запросы;
- 5) Elasticsearch. Визуализация;
- 6) Neo4j. Алгоритм создания и заполнения графовой БД;
- 7) Neo4j. Графическое представление данных и запрос;
- 8) Spark. Алгоритм создания CSV-файлов с таблицами;
- 9) Spark. Запрос;
- 10) Spark. Мониторинг выполнения запроса;
- 11) Spark. DAG SQL-запроса.

Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)

Задание:

- 1) Установить виртуальную машину с ОС Ubuntu в VirtualBox.
- 2) Установить Elasticsearch, Neo4j, Nadoop, Spark.
- 3) Вручную создать JSON-файл с 20-30 JSON-документами для предметной области, указанной в варианте.
- 4) В Elasticsearch создать индекс с анализатором и маппингом, проиндексировать JSON-документы, разработать запросы с вложенной агрегацией, представить результаты в среде Kibana.
- 5) В Neo4j по данным из Elasticsearch заполнить графовую базу данных, разработать и реализовать запрос к этой БД.
- 6) В Spark по данным из Elasticsearch сформировать csv-файлы с таблицами и сохранить их в файловой системе HDFS, написать запрос и реализовать его в Spark, проанализировать процесс выполнения запроса с использованием монитора Spark.

Предметная область – Салон красоты

Elasticsearch:

1. Типы JSON-документов:

Процедура:

```
{index, doc_type, id, body: {id_клиента, возраст, персональные_данные*, id_процедуры, дата_процедуры, стоимость, id_специалиста, [препарат*]}}
```

Специалист:

```
{index, doc_type, id, body: {специализация, стаж_работы, сведения_о_специалисте, график_работы*, [отзыв_о_специалисте*]}}
```

2. Требования к анализатору: для полей, отмеченных символом *, должен быть разработан анализатор со следующими требованиями: разделить текст на слова, убрать пунктуацию с помощью токенизатора standard (русский), перевести все токены в нижний регистр, убрать токены, находящиеся в списке стоп-слов, выполнить стемминг оставшихся токенов с помощью фильтра snowball.

3. Запросы с вложенной агрегацией:

- разбить процедуры по дате с периодом 1 год, для каждой "корзины" определить суммарную стоимость по каждому специалисту,
- предложить признаки отрицательного отзыва, вывести специалистов хотя бы с одним отрицательным отзывом.

Neo4j:

1. По данным из Elasticsearch заполнить графовую базу данных Клиент(id_клиента, персональные_данные) - Посетил(дата_процедуры, стоимость) - Специалист(id_специалиста, специализация, стаж работы).

2. Разработать и реализовать запрос: определить специалиста с наибольшей стоимостью выполненных процедур.

Spark:

1. По данным из Elasticsearch сформировать csv-файлы (с внутренней схемой) таблиц "Клиент", "Процедура", "Специалист" и сохранить их в файловой системе HDFS.
2. Написать запрос select: определить число выполненных процедур каждым специалистом.
3. Реализовать этот запрос в Spark. Построить временную диаграмму его выполнения по результатам работы монитора.

					Разработка макета аналитической системы на основе баз Данных NoSQL (вариант № 6)			
					Задание на курсовой проект			
Изм.	Лист	№ документа	Подпись	Дата	Литер.		Масса	Масштаб
Разраб.		Матвиенко Е.К.						
Руковод.		Григорьев Ю. А.						
					Лист 1		Листов 11	
					МГТУ им. Н. Э. Баумана Группа ИУ6-23М			
Н. Контр.								

Индексация документов Elasticsearch

Анализатор для индексов

```
"analysis": {
  "filter": {
    "russian_stop_words": {
      "type": "stop",
      "stopwords": "_russian_"
    },
    "filter_ru_sn": {
      "type": "snowball",
      "language": "Russian"
    }
  },
  "analyzer": {
    "analtic_for_ru": {
      "type": "custom",
      "tokenizer": "standard",
      "filter": [
        "lowercase",
        "russian_stop_words",
        "filter_ru_sn"
      ]
    }
  }
}
```

Фрагмент маппинга для индекса процедура

```
ProcedureMapping = {
  "properties": {
    "id_of_client": {
      "type": "text",
      "fielddata": True
    },
    "client_age": {
      "type": "integer"
    },
    "client_personal_data": {
      "type": "text",
      "analyzer": "analtic_for_ru",
      "fielddata": True
    },
    "date_of_procedure": {
      "type": "date",
      "format": "yyyy-MM-dd"
    }
  }
}
```

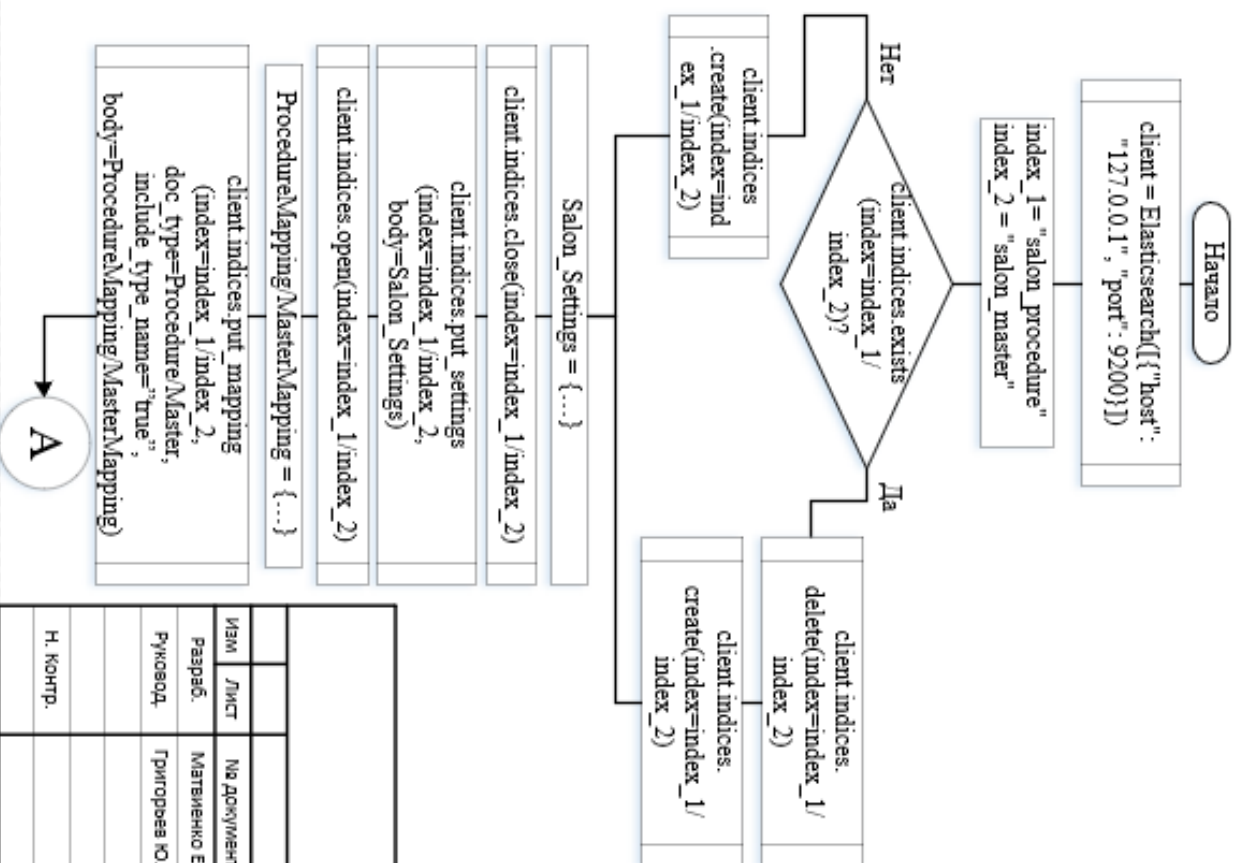
Фрагмент маппинга для индекса мастер

```
MasterMapping = {
  "properties": {
    "specialisation": {
      "type": "text",
      "fielddata": True,
      "analyzer": "analtic_for_ru"
    },
    "search_analyzer": "analtic_for_ru"
  },
  "reviews": {
    "type": "text",
    "fielddata": True,
    "analyzer": "analtic_for_ru",
    "search_analyzer": "analtic_for_ru"
  }
}
```

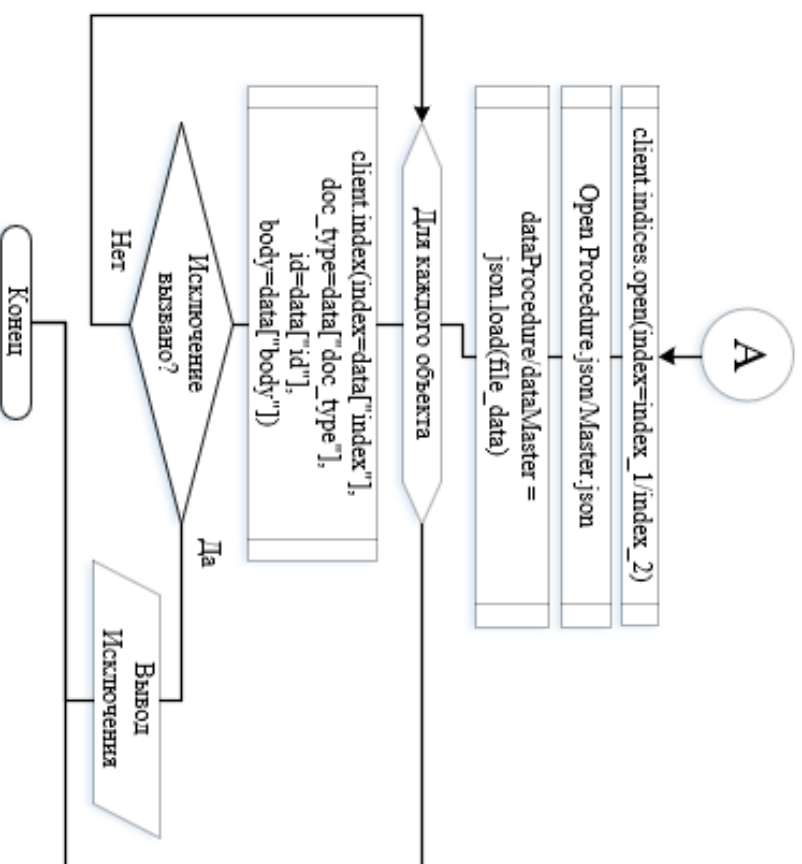
Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)				Индексация документов Elasticsearch. Маппинг и анализатор			
Изм	Лист	№ документа	Подпись	Дата	Литер.	Масштаб	
Разраб.		Матвиенко Е.К.					
Руковод.		Григорьев Ю. А.					
					Лист 2	Листов 11	
				Н. Контр.	МГТУ им. Н. Э. Баумана Группа ИУ6-23М		

Индексация документов Elasticsearch

Алгоритм добавления маппинга и настройки анализатора



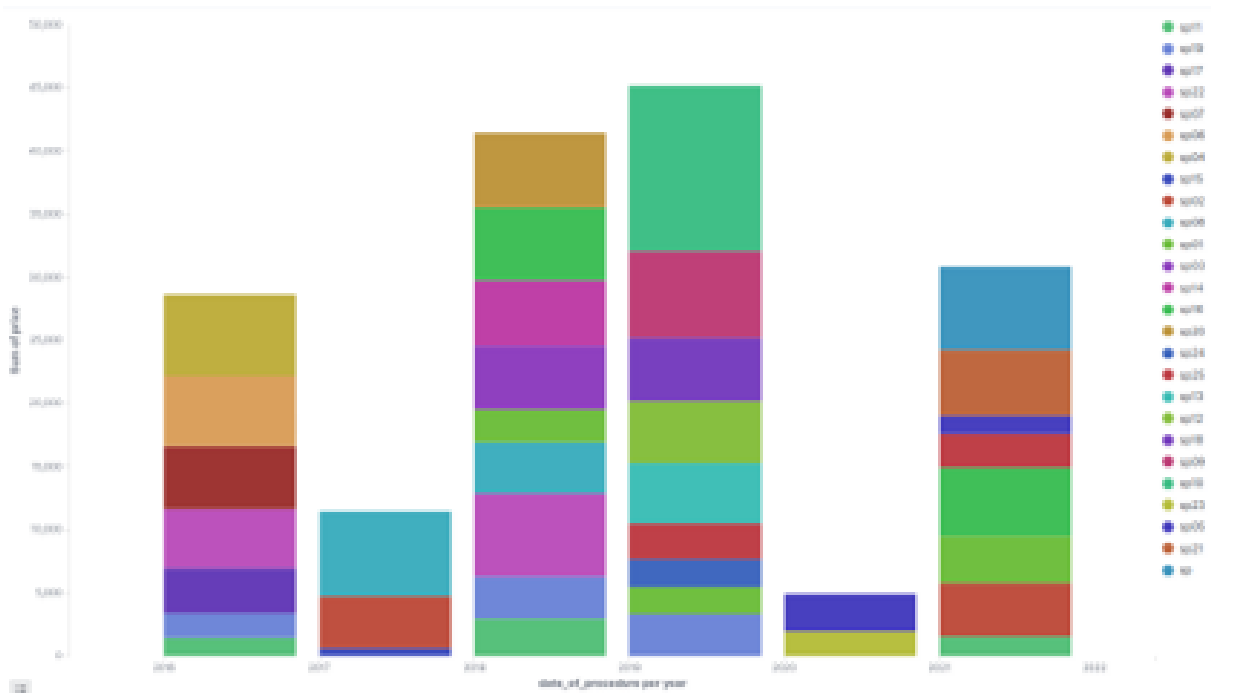
Алгоритм индексации



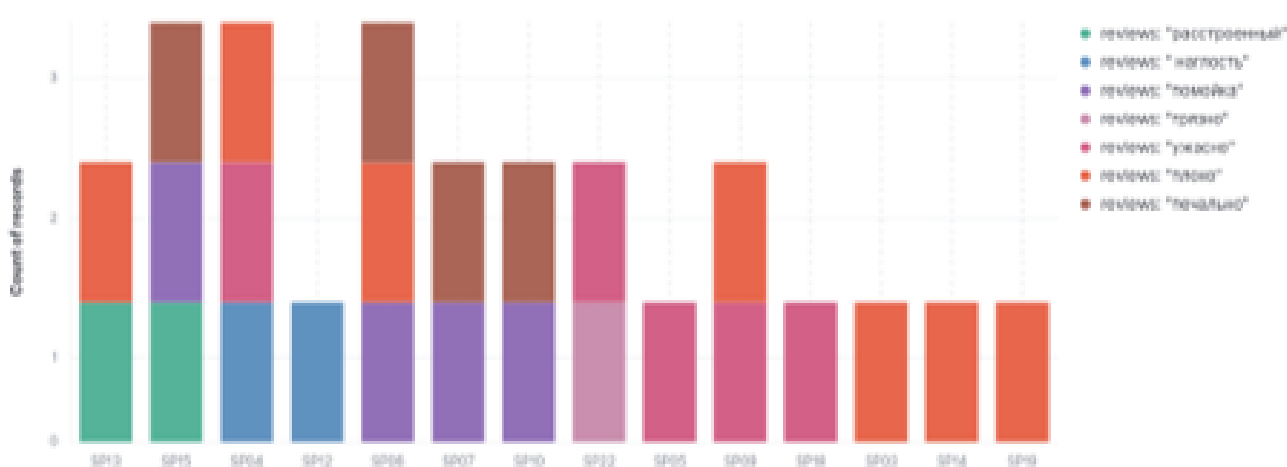
Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)					Индексация документов Elasticsearch. Алгоритмы индексации			МГТУ им. Н. Э. Баумана Группа ИУ6-23М		
Изм.	Лист	№ документа	Подпись	Дата				Листр.	Масса	Масштаб
Разраб.	Матвиенко Е.К.									
Руковод.	Григорьев Ю. А.									
Н. Контр.								Лист 3	Листов 11	

Elasticsearch-Kibana. Визуализация запросов

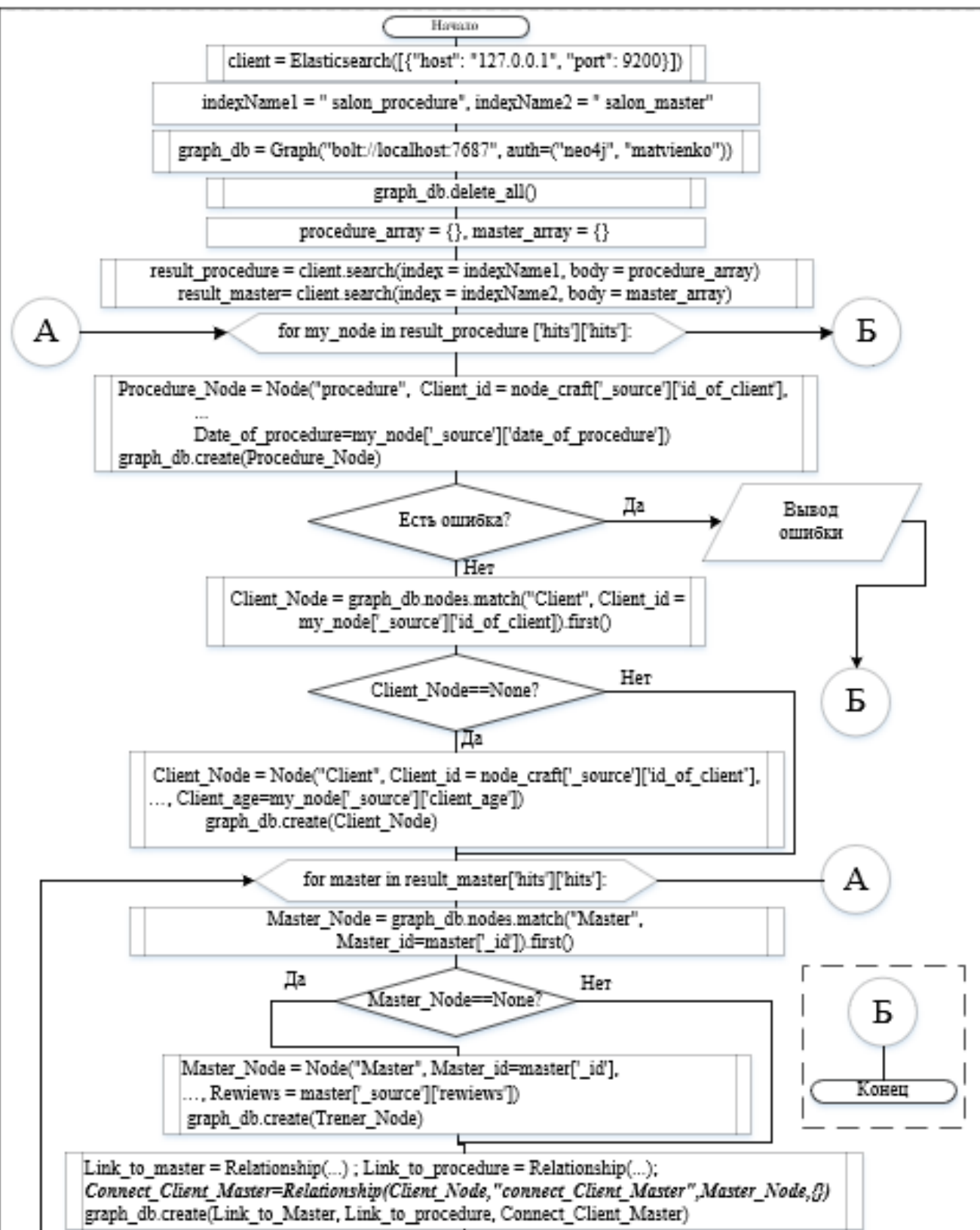
Визуализация первого запроса



Визуализация второго запроса



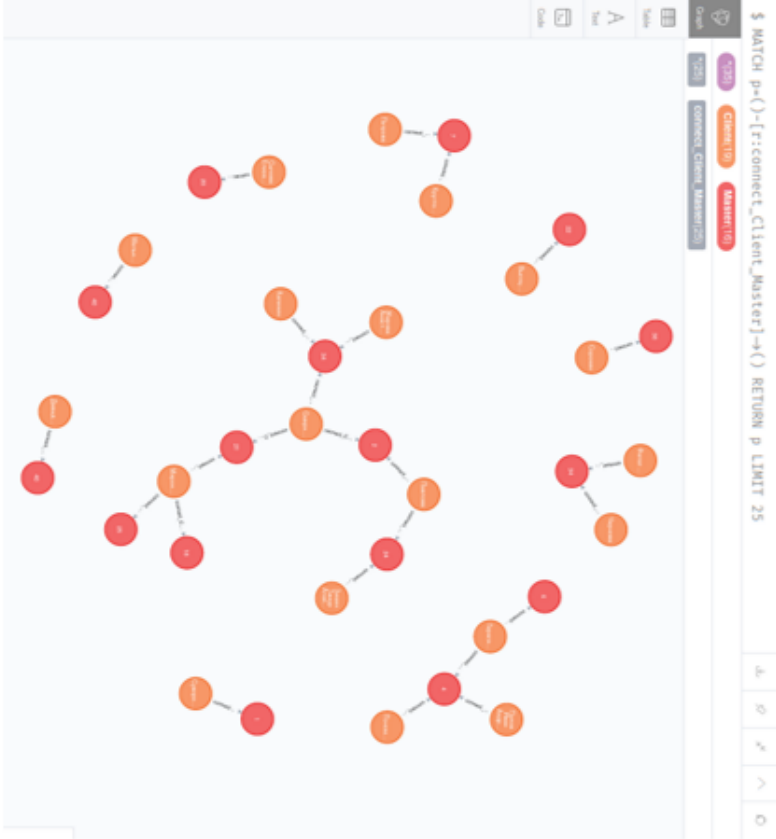
					Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)				
					Elasticsearch. Визуализация запросов	Литер.		Масса	Масштаб
Имя	Лист	№ документа	Подпись	Дата					
Разраб.		Матвеев Е.А.							
Руковод.		Григорьев Ю. А.							
						Лист 5		Листов 11	
Н. Контр.						МГТУ им. Н. Э. Баумана Группа ИУ6-23М			



					Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)				
					Neo4j. Алгоритм создания и заполнения графовой БД	Литер.		Масса	Масштаб
Изм	Лист	№ документа	Подпись	Дата					
Разраб.		Матвиенко Е.К.							
Руковод.		Григорьев Ю. А.							
						Лист 6		Листов 11	
Н. Контр.						МГТУ им. Н. Э. Баумана Группа ИУ6-23М			

Neo4j. Запрос

Связь «клиент посетил специалиста» = Запрос: определить специалиста с наибольшей стоимостью выполненных процедур = Connect_Client_Master



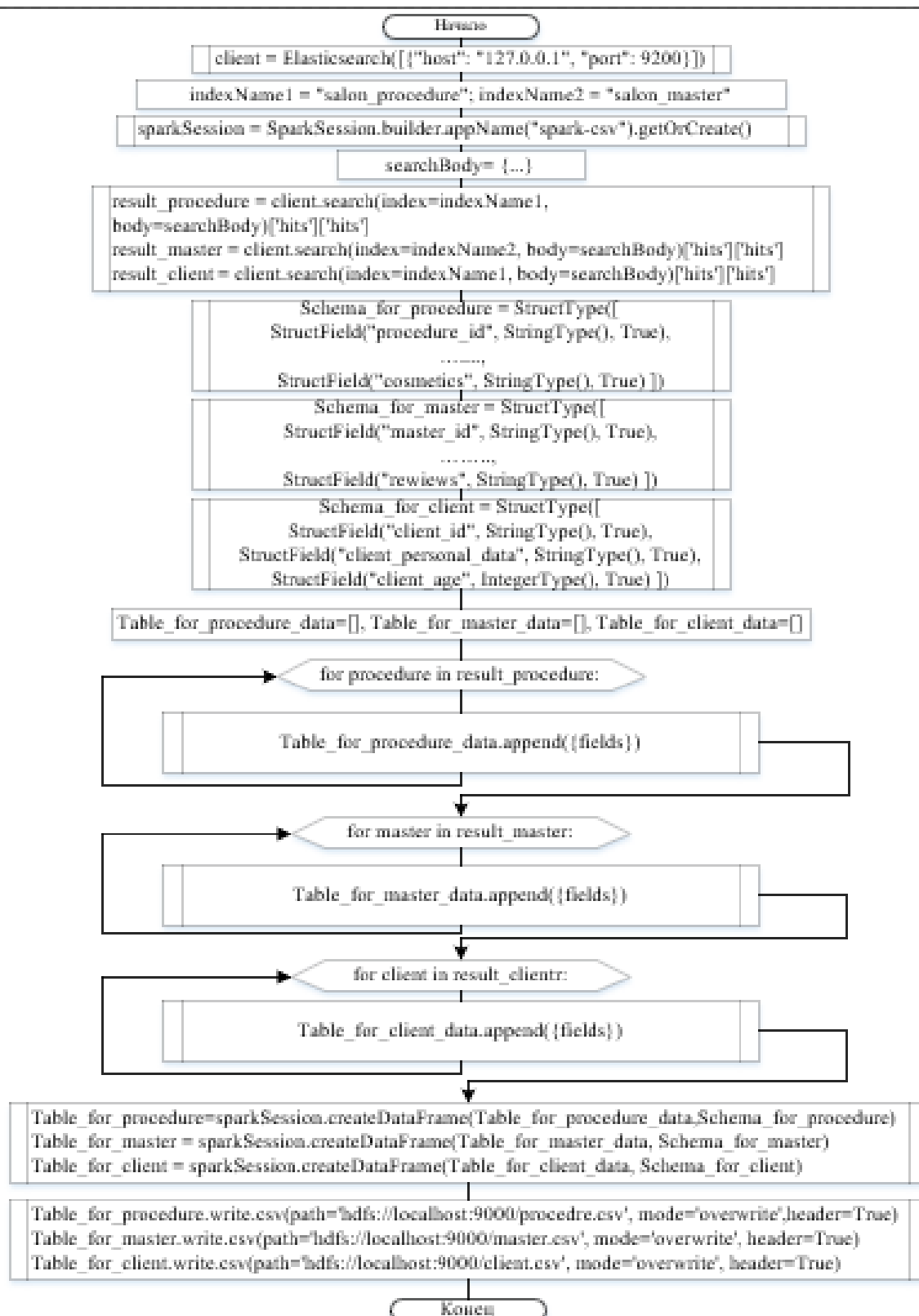
```
{
  "work_experience": "40",
  "master_id": "SP04",
  "specialization": [
    "Маникюр Европейский",
    "Уход за руками",
    "Долговременная укладка",
    "Долгосрочное покрытие геле-лаками (им-лаками)"
  ],
  "master_personal_data": "Семина Анна Игоревна"
},
{
  "client_personal_data": "Давыдова Маргарита Матвеевна",
  "client_age": "20",
  "client_id": "ИХАВ"
}
```

Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)							
Изм	Лист	№ документа	Подпись	Дата			
Разраб.		Матвиенко Е.К.					
Руковод.		Григорьев Ю. А.					
Н. Контр.							
Neo4j. Запрос							
					Литер.	Масса	Масштаб
					Лист 7	Листов 11	
МГТУ им. Н. Э. Баумана Группа ИУ6-23М							

```
1 MATCH p=(c:Client)-[r:connect_Client_Master]-(m:Master)
2 WITH m, sum(toInteger(r.price)) as master_sum
3 ORDER BY master_sum desc
4 RETURN m, master_sum
5 LIMIT 1
```

```
{
  "work_experience": "2",
  "master_id": "SP10",
  "specialization": [
    "Массаж рук",
    "Спа-маникюр",
    "Биопилинг кожи головы",
    "Гальванизация"
  ],
  "master_personal_data": "Ильинский Матвей Эмирович"
}
```

m	master_sum
	13084



					Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)			
					Spark. Алгоритм создания CSV-файлов с таблицами	Литер.	Масса	Масштаб
Изм	Лист	№ документа	Подпись	Дата				
Разраб.		Митяченко Е.Н.						
Руковод.		Григорьев Ю. А.						
						Лист 8	Листов 11	
						МГТУ им. Н. Э. Баумана Группа ИУ6-23М		
Н. Контр.								

Spark. Мониторинг выполнения запроса

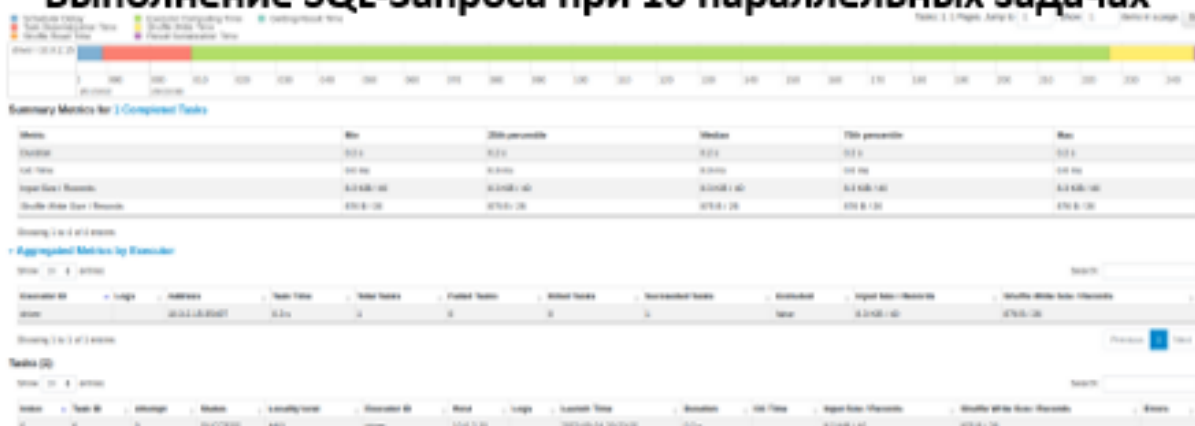
Выполненные SQL-запросы

ID	Description	Submitted	Duration	Job IDs
6	showString at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	0.9 s	[6](7)
5	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	4 ms	
4	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	1 ms	
3	createOrReplaceTempView at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	22 ms	
2	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	84 ms	[4]
1	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:01	0.1 s	[2]
0	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:22:57	2 s	[0]

Spark Jobs

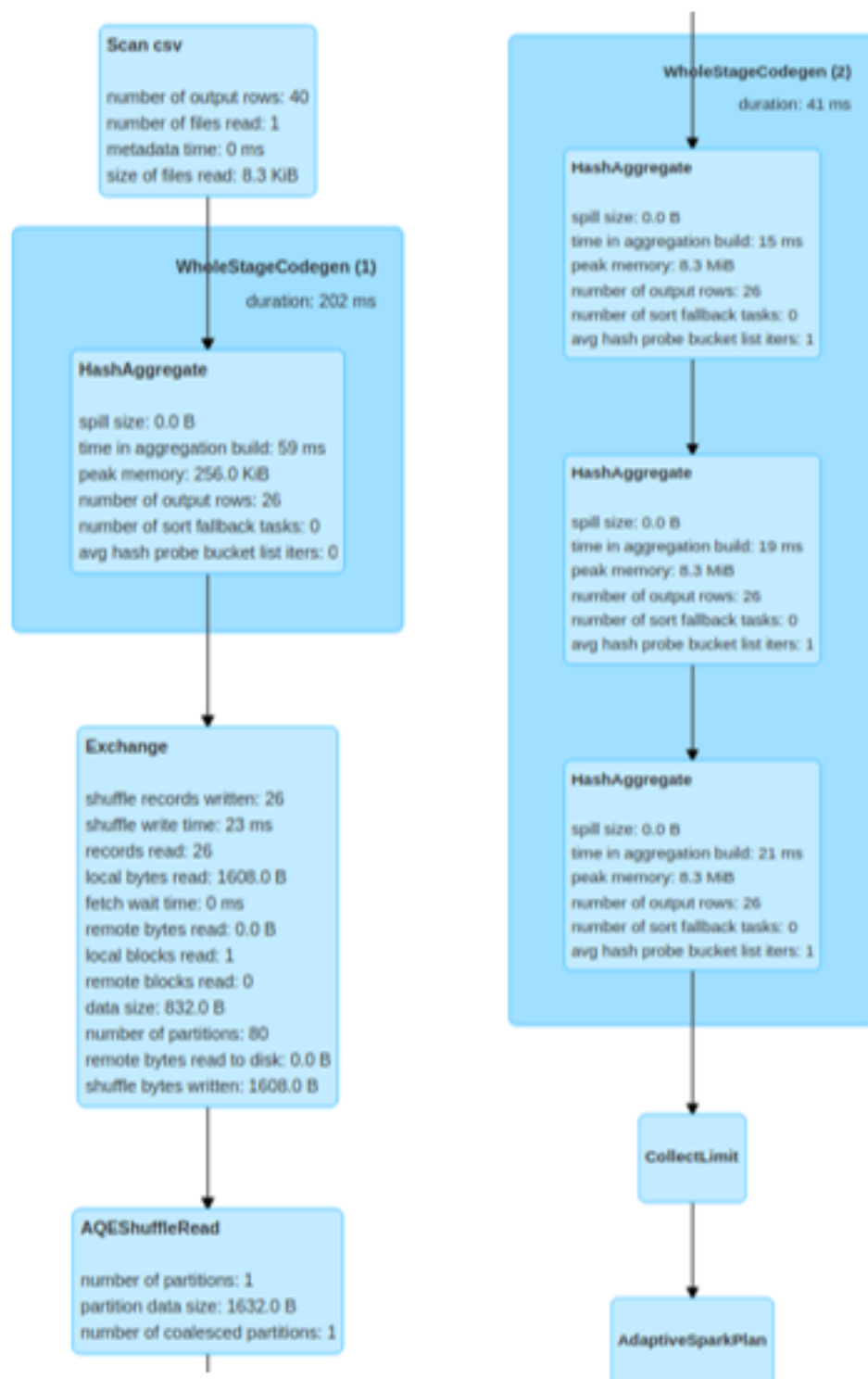
Job ID	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
7	showString at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:03	0.1 s	1/1 (1 skipped)	1/1 (1 skipped)
6	showString at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	0.3 s	1/1	1/1
5	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	41 ms	1/1	1/1
4	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:02	39 ms	1/1	1/1
3	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:01	96 ms	1/1	1/1
2	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:01	45 ms	1/1	1/1
1	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:23:00	0.9 s	1/1	1/1
0	load at NativeMethodAccessorImpl.java:0	2022/05/24 20:22:58	1 s	1/1	1/1

Выполнение SQL-Запроса при 10 параллельных задачах



					Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)			
					Spark. Мониторинг выполнения запроса	Литер.	Масса	Масштаб
Изм	Лист	№ документа	Подпись	Дата				
Разраб.		Матвеев Е.К.						
Руковод.		Григорьев Ю. А.						
						Лист 10		Листов 11
Н. Контр.						МГТУ им. Н. Э. Баумана Группа ИУ6-23М		

Spark. DAG SQL-запроса



Разработка макета аналитической системы на основе баз данных NoSQL (вариант № 6)					Литер.			Масса	Масштаб
Имя	Лист	№ документа	Подпись	Дата	Spark. DAG SQL-запроса				
Разраб.		Матвеев Е.К.							
Руковод.		Григорьев Ю. А.							
Н. Контр.									
					Лист 11				
					Листов 11				
					МГТУ им. Н. Э. Баумана Группа ИУ6-23М				

