**ТИТУЛ**

**ЗАДАНИЕ**

**РЕФЕРАТ**

**СОДЕРЖАНИЕ**

[Определения, обозначения и сокращения 5](#_Toc1)

[Введение 6](#_Toc2)

[1 Задание и описание варианта КП 7](#_Toc3)

[2 Установка виртуальной машины, установка пакетов Elasticsearch, Neo4j, Hadoop+Spark 8](#_Toc4)

[2.1 Развертывание виртуальной машины 9](#_Toc5)

[2.2 Установка ElasticSearch 10](#_Toc6)

[2.3 Установка Neo4j 11](#_Toc7)

[2.4 Установка Hadoop+Spark 12](#_Toc8)

[3 Elasticsearch 13](#_Toc9)

[3.1 Индексация документов 14](#_Toc10)

[3.2 Запросы 15](#_Toc11)

[4 Neo4j 16](#_Toc12)

[4.1 Создание и заполнение графовой БД 17](#_Toc13)

[4.2 Запрос 18](#_Toc14)

[5 Spark 19](#_Toc15)

[5.1 Создание и заполнение таблиц 20](#_Toc16)

[5.2 Запрос 21](#_Toc17)

[5.3 Монитор 22](#_Toc18)

[Заключение 23](#_Toc19)

[Список использованных источников 24](#_Toc20)

[ПРИЛОЖЕНИЕ А 25](#_Toc21)

[ПРИЛОЖЕНИЕ Б 26](#_Toc22)

[ПРИЛОЖЕНИЕ В 27](#_Toc23)

# **Определения, обозначения и сокращения**

# **Введение**

# **1 Задание и описание варианта КП**

# **2 Установка виртуальной машины, установка пакетов Elasticsearch, Neo4j, Hadoop+Spark**

## **2.1 Развертывание виртуальной машины**

## **2.2 Установка ElasticSearch**

## **2.3 Установка Neo4j**

## **2.4 Установка Hadoop+Spark**

# **3 Elasticsearch**

## **3.1 Индексация документов**

Прежде чем выполнить индексацию документов было необходимо сгенерировать исходные данные. Для этого была использована python-библиотека Faker, позволяющая генерировать ФИО, тексты, даты и другие данные с заданными параметрами [ссылка]**.** В результате были сгенерированы данные со структурами, приведенными ниже**.**

Структура для записи о мастере:

* master\_id — уникальный номер мастера (пример: 163543);
* master\_desc — описание мастера — ФИО и данные об образовании/опыте (Пример: “Максимильян Виленович Гурьев, Стаж Работы: 11 Л./Г..”);
* master\_feedbacks — массив отзывов о мастере (пример: [“медлительный, аккуратный.”, “медлительный, ворчливый.”]).

Структура записи о заказе:

* order\_id — уникальный номер заказа (пример: 405064);
* order\_date — дата получения заказа (пример: 2024-03-15);
* order\_customer\_id — уникальный номер заказчика (пример: 257458);
* order\_customer\_desc — описание заказчика — ФИО и доп. информация (пример: “Имя: Комарова Дарья Борисовна; постоянный, премиальный.”);
* order\_details\_desc — свойства заказа (пример: “хрупкий, тяжелый.”);
* order\_due\_date — срок выполнения заказа (пример: 2024-04-20);
* order\_fact\_completion\_date — фактическая дата исполнения заказа (пример: 2024-04-25);
* order\_parts — массив со списком необходимых запчастей (пример: [“дисплей”, “usb-разъем”]);
* repair\_types — тип работ (пример: “мелкие детали”, “пайка”);
* order\_price — стоимость заказа (пример: 3866.44);
* order\_master\_id — уникальный номер мастера-исполнителя (пример: 163543);

**ПРОДОЛЖЕНИЕ СЛЕДУЕТ**

## **3.2 Запросы**

# **4 Neo4j**

## **4.1 Создание и заполнение графовой БД**

## **4.2 Запрос**

# **5 Spark**

## **5.1 Создание и заполнение таблиц**

## **5.2 Запрос**

## **5.3 Монитор**

# **Заключение**

# **Список использованных источников**

# **ПРИЛОЖЕНИЕ А**

# **ПРИЛОЖЕНИЕ Б**

# **ПРИЛОЖЕНИЕ В**