**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**

Фрагменты JSON и результаты выполнения запросов

Листов 9

**1. Фрагменты JSON с сущностями “Мастер” (первые 10 записей)**

|  |
| --- |
| [  {  "index": "master",  "id": 563268,  "body": {  "master\_id": 563268,  "master\_desc": "Милица Руслановна Фомина, Тип Образования: Среднее.",  "master\_feedbacks": [  "ворчливый, медлительный.",  "ворчливый, медлительный.",  "медлительный, ворчливый."  ]  }  },  {  "index": "master",  "id": 871908,  "body": {  "master\_id": 871908,  "master\_desc": "Козлов Юрий Фомич, Тип Образования: Среднее.",  "master\_feedbacks": [  "ворчливый, аккуратный.",  "медлительный, аккуратный.",  "аккуратный, ворчливый."  ]  }  },  {  "index": "master",  "id": 164475,  "body": {  "master\_id": 164475,  "master\_desc": "Юлий Дмитриевич Константинов, Тип Образования: Среднее.",  "master\_feedbacks": [  "аккуратный, ворчливый."  ]  }  },  {  "index": "master",  "id": 464209,  "body": {  "master\_id": 464209,  "master\_desc": "Ия Тарасовна Степанова, Тип Образования: Высшее.",  "master\_feedbacks": [  "аккуратный, медлительный."  ]  }  },  {  "index": "master",  "id": 22653,  "body": {  "master\_id": 22653,  "master\_desc": "Акулина Рудольфовна Никитина, Стаж Работы: 15 Л./Г..",  "master\_feedbacks": [  "ворчливый, аккуратный.",  "ворчливый, медлительный."  ]  }  },  {  "index": "master",  "id": 282806,  "body": {  "master\_id": 282806,  "master\_desc": "Маслов Емельян Дмитриевич, Тип Образования: Высшее.",  "master\_feedbacks": [  "ворчливый, медлительный."  ]  }  },  {  "index": "master",  "id": 736774,  "body": {  "master\_id": 736774,  "master\_desc": "Владимирова Анжелика Эдуардовна, Тип Образования: Среднее.",  "master\_feedbacks": [  "аккуратный, ворчливый."  ]  }  },  {  "index": "master",  "id": 300714,  "body": {  "master\_id": 300714,  "master\_desc": "Вероника Петровна Силина, Стаж Работы: 12 Л./Г..",  "master\_feedbacks": [  "ворчливый, медлительный.",  "ворчливый, аккуратный."  ]  }  },  {  "index": "master",  "id": 405064,  "body": {  "master\_id": 405064,  "master\_desc": "Хохлов Олег Харлампьевич, Стаж Работы: 3 Л./Г..",  "master\_feedbacks": [  "ворчливый, медлительный.",  "ворчливый, аккуратный.",  "ворчливый, аккуратный."  ]  }  },  {  "index": "master",  "id": 257458,  "body": {  "master\_id": 257458,  "master\_desc": "Шестакова Наина Владиславовна, Стаж Работы: 9 Л./Г..",  "master\_feedbacks": [  "аккуратный, ворчливый.",  "ворчливый, аккуратный."  ]  }  },  ...  ] |

**2. Фрагменты JSON с сущностями “Заказ” (первые 10 записей)**

|  |
| --- |
| [  {  "index": "order",  "id": 131919,  "body": {  "order\_id": 131919,  "order\_date": "2024-01-06",  "order\_customer\_id": 227431,  "order\_customer\_desc": "Имя: Маргарита Валентиновна Чернова; премиальный, постоянный.",  "order\_details\_desc": "хрупкий, ценный.",  "order\_due\_date": "2024-04-10",  "order\_fact\_completion\_date": "2024-04-27",  "order\_parts": [  "аккумулятор"  ],  "repair\_types": [  "замена компонентов"  ],  "order\_price": 2296.89215138,  "order\_master\_id": 257458  }  },  {  "index": "order",  "id": 664730,  "body": {  "order\_id": 664730,  "order\_date": "2024-02-11",  "order\_customer\_id": 656432,  "order\_customer\_desc": "Имя: Комарова Дарья Борисовна; постоянный, премиальный.",  "order\_details\_desc": "хрупкий, тяжелый.",  "order\_due\_date": "2024-04-30",  "order\_fact\_completion\_date": "2024-04-20",  "order\_parts": [  "дисплей",  "дисплей",  "usb-разъем"  ],  "repair\_types": [  "мелкие детали",  "пайка"  ],  "order\_price": 3866.441,  "order\_master\_id": 785701  }  },  {  "index": "order",  "id": 141550,  "body": {  "order\_id": 141550,  "order\_date": "2024-01-10",  "order\_customer\_id": 683623,  "order\_customer\_desc": "Имя: Лора Борисовна Жукова; должник, постоянный.",  "order\_details\_desc": "хрупкий, тяжелый.",  "order\_due\_date": "2024-04-03",  "order\_fact\_completion\_date": "2024-04-27",  "order\_parts": [  "дисплей"  ],  "repair\_types": [  "анализ неисправности",  "мелкие детали"  ],  "order\_price": 4879.1271000384,  "order\_master\_id": 814777  }  },  {  "index": "order",  "id": 99676,  "body": {  "order\_id": 99676,  "order\_date": "2023-12-25",  "order\_customer\_id": 444528,  "order\_customer\_desc": "Имя: Василий Эдуардович Жданов; должник, премиальный.",  "order\_details\_desc": "тяжелый, ценный.",  "order\_due\_date": "2024-04-20",  "order\_fact\_completion\_date": "2024-04-04",  "order\_parts": [  "usb-разъем"  ],  "repair\_types": [  "замена компонентов",  "замена компонентов"  ],  "order\_price": 2103.6634502,  "order\_master\_id": 397809  }  },  {  "index": "order",  "id": 277607,  "body": {  "order\_id": 277607,  "order\_date": "2024-02-22",  "order\_customer\_id": 972535,  "order\_customer\_desc": "Имя: Гурий Анисимович Евдокимов; премиальный, постоянный.",  "order\_details\_desc": "хрупкий, ценный.",  "order\_due\_date": "2024-04-22",  "order\_fact\_completion\_date": "2024-04-10",  "order\_parts": [  "дисплей",  "корпус",  "корпус"  ],  "repair\_types": [  "замена компонентов",  "пайка"  ],  "order\_price": 5652.6299329,  "order\_master\_id": 163543  }  },  {  "index": "order",  "id": 832420,  "body": {  "order\_id": 832420,  "order\_date": "2024-03-01",  "order\_customer\_id": 872571,  "order\_customer\_desc": "Имя: Тимофеев Савва Григорьевич; постоянный, премиальный.",  "order\_details\_desc": "тяжелый, ценный.",  "order\_due\_date": "2024-04-14",  "order\_fact\_completion\_date": "2024-04-11",  "order\_parts": [  "usb-разъем",  "аккумулятор"  ],  "repair\_types": [  "мелкие детали"  ],  "order\_price": 5427.53678234973,  "order\_master\_id": 405064  }  },  {  "index": "order",  "id": 880172,  "body": {  "order\_id": 880172,  "order\_date": "2024-02-08",  "order\_customer\_id": 460465,  "order\_customer\_desc": "Имя: Лыткин Игнатий Валерьянович; должник, премиальный.",  "order\_details\_desc": "тяжелый, объемный.",  "order\_due\_date": "2024-04-23",  "order\_fact\_completion\_date": "2024-04-20",  "order\_parts": [  "usb-разъем"  ],  "repair\_types": [  "пайка",  "замена компонентов"  ],  "order\_price": 7003.855612,  "order\_master\_id": 871908  }  },  {  "index": "order",  "id": 993858,  "body": {  "order\_id": 993858,  "order\_date": "2024-01-16",  "order\_customer\_id": 433504,  "order\_customer\_desc": "Имя: Лукьян Венедиктович Богданов; премиальный, должник.",  "order\_details\_desc": "объемный, ценный.",  "order\_due\_date": "2024-04-08",  "order\_fact\_completion\_date": "2024-04-27",  "order\_parts": [  "дисплей",  "корпус"  ],  "repair\_types": [  "пайка"  ],  "order\_price": 7472.5287,  "order\_master\_id": 871908  }  },  {  "index": "order",  "id": 683490,  "body": {  "order\_id": 683490,  "order\_date": "2024-01-22",  "order\_customer\_id": 269146,  "order\_customer\_desc": "Имя: Алина Оскаровна Орехова; постоянный, должник.",  "order\_details\_desc": "ценный, тяжелый.",  "order\_due\_date": "2024-04-15",  "order\_fact\_completion\_date": "2024-04-20",  "order\_parts": [  "корпус",  "корпус"  ],  "repair\_types": [  "анализ неисправности",  "анализ неисправности"  ],  "order\_price": 2959.11938941,  "order\_master\_id": 300714  }  },  {  "index": "order",  "id": 243483,  "body": {  "order\_id": 243483,  "order\_date": "2024-02-01",  "order\_customer\_id": 408481,  "order\_customer\_desc": "Имя: Филатов Карл Алексеевич; должник, премиальный.",  "order\_details\_desc": "объемный, тяжелый.",  "order\_due\_date": "2024-04-05",  "order\_fact\_completion\_date": "2024-04-27",  "order\_parts": [  "дисплей",  "корпус",  "дисплей"  ],  "repair\_types": [  "анализ неисправности",  "пайка"  ],  "order\_price": 2323.37152,  "order\_master\_id": 736774  }  },  ...  ] |

**3. Результат запроса №1 к ElasticSearch**

|  |
| --- |
| [{'key\_as\_string': '2023-12-01',  'key': 1701388800000,  'doc\_count': 13,  'over\_parts': {'doc\_count\_error\_upper\_bound': 0,  'sum\_other\_doc\_count': 0,  'buckets': [{'key': 'usb-разъем', 'doc\_count': 5},  {'key': 'дисплей', 'doc\_count': 4},  {'key': 'корпус', 'doc\_count': 4},  {'key': 'аккумулятор', 'doc\_count': 3}]}},  {'key\_as\_string': '2024-01-01',  'key': 1704067200000,  'doc\_count': 48,  'over\_parts': {'doc\_count\_error\_upper\_bound': 0,  'sum\_other\_doc\_count': 0,  'buckets': [{'key': 'usb-разъем', 'doc\_count': 23},  {'key': 'дисплей', 'doc\_count': 22},  {'key': 'корпус', 'doc\_count': 18},  {'key': 'аккумулятор', 'doc\_count': 16}]}},  {'key\_as\_string': '2024-02-01',  'key': 1706745600000,  'doc\_count': 37,  'over\_parts': {'doc\_count\_error\_upper\_bound': 0,  'sum\_other\_doc\_count': 0,  'buckets': [{'key': 'дисплей', 'doc\_count': 19},  {'key': 'usb-разъем', 'doc\_count': 14},  {'key': 'аккумулятор', 'doc\_count': 14},  {'key': 'корпус', 'doc\_count': 14}]}},  {'key\_as\_string': '2024-03-01',  'key': 1709251200000,  'doc\_count': 2,  'over\_parts': {'doc\_count\_error\_upper\_bound': 0,  'sum\_other\_doc\_count': 0,  'buckets': [{'key': 'usb-разъем', 'doc\_count': 1},  {'key': 'аккумулятор', 'doc\_count': 1},  {'key': 'корпус', 'doc\_count': 1}]}}] |

**4. Результат запроса №2 к ElasticSearch**

|  |
| --- |
| {'total': {'value': 8, 'relation': 'eq'},  'max\_score': 0.8435577,  'hits': [{'\_index': 'master',  '\_type': '\_doc',  '\_id': '22653',  '\_score': 0.8435577,  '\_source': {'master\_id': 22653,  'master\_desc': 'Акулина Рудольфовна Никитина, Стаж Работы: 15 Л./Г..',  'master\_feedbacks': ['ворчливый, аккуратный.',  'ворчливый, медлительный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '300714',  '\_score': 0.8435577,  '\_source': {'master\_id': 300714,  'master\_desc': 'Вероника Петровна Силина, Стаж Работы: 12 Л./Г..',  'master\_feedbacks': ['ворчливый, медлительный.',  'ворчливый, аккуратный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '405064',  '\_score': 0.8435577,  '\_source': {'master\_id': 405064,  'master\_desc': 'Хохлов Олег Харлампьевич, Стаж Работы: 3 Л./Г..',  'master\_feedbacks': ['ворчливый, медлительный.',  'ворчливый, аккуратный.',  'ворчливый, аккуратный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '257458',  '\_score': 0.8435577,  '\_source': {'master\_id': 257458,  'master\_desc': 'Шестакова Наина Владиславовна, Стаж Работы: 9 Л./Г..',  'master\_feedbacks': ['аккуратный, ворчливый.', 'ворчливый, аккуратный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '814777',  '\_score': 0.8435577,  '\_source': {'master\_id': 814777,  'master\_desc': 'Шарова Жанна Аркадьевна, Стаж Работы: 1 Л./Г..',  'master\_feedbacks': ['ворчливый, медлительный.',  'медлительный, аккуратный.',  'ворчливый, медлительный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '847121',  '\_score': 0.8435577,  '\_source': {'master\_id': 847121,  'master\_desc': 'Жуков Ладимир Всеволодович, Стаж Работы: 12 Л./Г..',  'master\_feedbacks': ['медлительный, ворчливый.',  'аккуратный, медлительный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '163543',  '\_score': 0.8435577,  '\_source': {'master\_id': 163543,  'master\_desc': 'Максимильян Виленович Гурьев, Стаж Работы: 11 Л./Г..',  'master\_feedbacks': ['аккуратный, медлительный.']}},  {'\_index': 'master',  '\_type': '\_doc',  '\_id': '785701',  '\_score': 0.8435577,  '\_source': {'master\_id': 785701,  'master\_desc': 'Панфилова Пелагея Игоревна, Стаж Работы: 8 Л./Г..',  'master\_feedbacks': ['медлительный, аккуратный.',  'медлительный, ворчливый.']}}]} |

# **ПРИЛОЖЕНИЕ Б**

# **ПРИЛОЖЕНИЕ В**