

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ливанов Дмитрий Викторович
Должность: Ректор
Дата подписания: 17.06.2022
Уникальный программный ключ:
с6d909c49c1d2034fa3a0156c4eaa51e7232a3a2

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»

ВЫПИСКА ИЗ ПРОТОКОЛА № 11
заседания учебно-методического совета от «27» июня 2022 г.

ПОВЕСТКА:

Рассмотрение дополнительных общеобразовательных и профессиональных программ.
Проректор по учебной работе А. А. Воронов.

СЛУШАЛИ: Директора Физтех-школы прикладной математики и информатики
А.М. Райгородского.

ПОСТАНОВИЛИ:

Рекомендовать к утверждению в установленном порядке дополнительную программу профессиональной переподготовки «Data Engineer».

Решение принято единогласно.

Форма проведения заседания: заочная.

Председатель УМС МФТИ



А.А. Воронов

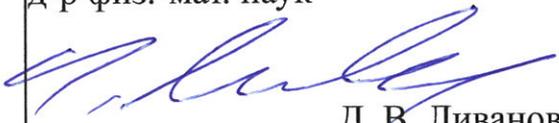
Ученый секретарь УМС МФТИ



М.В. Берзникова

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение высшего
образования

«Московский физико-технический институт
(национальный исследовательский университет)»

	Ректор д-р физ.-мат. наук  Д. В. Ливанов «27» июня 2022 г.
--	--

**Дополнительная профессиональная программа
Программа профессиональной переподготовки**

«Data Engineer»

Москва 2021

1. Рабочая группа

С целью повышения качества дополнительной профессиональной программы (ДПП) с учетом требований профессиональных стандартов (ПС) в группу разработчиков входили:

Райгородский Андрей Михайлович

Доктор физико-математических наук, директор ФПМИ МФТИ

Благодарный Евгений Владимирович

заведующий учебно-методической лабораторией инноватики

Иванова Анастасия Сергеевна

руководитель проектов учебно-методической лаборатории инноватики ФПМИ МФТИ

2. Профессиональные стандарты, учтённые в содержании программы

Настоящая дополнительная профессиональная программа (ДПП) разработана с учётом соответствующих ей профессиональных стандартов (ПС) из национального реестра профессиональных стандартов. Профессиональные стандарты (ПС), связанные с настоящей ДПП:

06.003 Архитектор программного обеспечения

06.042 Специалист по большим данным

3. Квалификационные требования, учтённые в содержании программы

Квалификационные требования, учтённые в содержании программы, указаны в описании должностей единого квалификационного справочника должностей руководителей, специалистов и служащих. Причём учтены требования должностей, указанных в дополнительных характеристиках отобранных для данной программы ОТФ (п. 5. настоящего документа), которые приведены в указанных в п. 2. настоящего документа профессиональных стандартах (ПС) в строке «ЕТКС или ЕКС». Квалификационные требования инкорпорированы в настоящий документ путём отсылки и являются его неотъемлемой частью.

4. Требования федерального государственного стандарта высшего образования, на основании которых была разработана программа

Настоящей дополнительной профессиональной программе соответствует федеральный государственный стандарт высшего образования (ФГОС ВО):

01.03.02 “Прикладная математика и информатика” (бакалавриат)

Настоящая дополнительная профессиональная программа разработана на основании требований указанного ФГОС ВО.

Полный перечень требований приведён в тексте указанного ФГОС ВО, инкорпорирован в настоящий документ путём отсылки и является его неотъемлемой частью.

5. Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и (или) уровней квалификации.

«Программа профессиональной переподготовки направлена на получение компетенции, необходимой для выполнения нового вида профессиональной деятельности, приобретение новой квалификации (часть 5 статьи 76 Федерального закона N 273-ФЗ). Для их определения и

может использоваться ПС. При этом необходимо учесть различия терминологии, используемой в образовании и в профессиональных стандартах, о которых говорилось выше. Вид профессиональной деятельности, квалификация, упоминаемые в цитируемой статье закона, в ПС в большинстве случаев соответствуют ОТФ, иногда - ТФ.»¹,

Новые квалификации (ОТФ/ТФ) и их уровни:

Вид профессиональной деятельности (ПС)	Новые квалификации (ОТФ/ТФ)	Уровень
Проектно-конструкторская деятельность в области информационных технологий	Создание вариантов архитектуры программного средства	4
	Реализация программных средств	4
	Оценка и выбор варианта архитектуры программного средства	5
	Оценка возможности создания архитектурного проекта	6
Создание и применение технологий больших данных	Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	6
	Управление этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации	7

Характеристики данных ОТФ/ТФ приведены в профессиональном стандарте (ПС), указанном в п. 2. настоящего документа. инкорпорированы в настоящий документ путём отсылки, и являются его неотъемлемой частью.

Связанные виды профессиональной деятельности (ВПД):

научно-исследовательская деятельность

проектная и производственно-технологическая деятельность

Характеристика данных ВПД приведена в в федеральном государственном стандарте высшего образования (ФГОС ВО), указанном в п. 4. настоящего документа, инкорпорирована в настоящий документ путём отсылки, и является его неотъемлемой частью. Указанные ВПД,

¹ Письмо Министерства образования и науки РФ от 22 апреля 2015 г. N ВК-1032/06 "О направлении методических рекомендаций"

освоение которых предусмотрено указанным ФГОС ВО, являются «сквозными» и служат основой овладения выбранной квалификацией (ОТФ/ТФ).

ТФ, связанные с указанными в настоящем пункте ОТФ (в случае их указания), а также их характеристики приведены в профессиональном стандарте (ПС), указанном в п. 2. настоящего документа. Инкорпорированы в настоящий документ путём отсылки, и являются его неотъемлемой частью.

6. Характеристика компетенций, подлежащих совершенствованию, и (или) перечень компетенций, формирующихся в результате освоения программы

Учитывая, что слушатели программы уже освоили или осваивают основную профессиональную образовательную программу, в том числе общие компетенции, соответствующего уровня профессионального образования, компетенции формирующиеся в результате освоения настоящей программы и компетенции подлежащие совершенствованию, а также их характеристика, представлены в ФГОС, указанном в п. 4. настоящего документа, инкорпорированы в настоящий документ путём отсылки, и являются его неотъемлемой частью.

Данные компетенции указанного ФГОС ВО служат основой определения результатов обучения по программе.

7. Цель программы

Реализация программы профессиональной переподготовки направлена на получение компетенции, необходимой для выполнения нового вида профессиональной деятельности, приобретение новой квалификации.

В соответствии с Письмом Министерства образования и науки РФ от 22 апреля 2015 г. N ВК-1032/06 «О направлении методических рекомендаций», цель представляет собой осознанное представление (предвосхищение) результата деятельности. Планируемые результаты отражены ниже.

8. Планируемые результаты обучения по программе

Планируемые результаты обучения соответствуют результатам освоения соответствующей основной профессиональной образовательной программы, а также направлены на приобретение новой квалификации, требующей изменения направленности (профиля) или специализации в рамках направления подготовки (специальности) полученного ранее профессионального образования, определены на основе профессиональных компетенций соответствующего федерального государственного образовательного стандарта, указанного в п. 4. настоящего документа.

Имеющаяся квалификация и (или) уровень образования (требования к слушателям):

Высшее, незаконченное высшее образование или среднее профессиональное.

Область профессиональной деятельности – специалисты, интересующиеся проектированием хранилищ данных, сбором, очисткой, трансформацией, хранением и анализом данных с помощью SQL и Python

Новые виды профессиональной деятельности, и ПК соответствующие им, которые планируются к получению в результате обучения по настоящей программе, а также их характеристики, представлены в ФГОС, указанном в п. 4. настоящего документа, инкорпорированы в настоящий документ путём отсылки, и являются его неотъемлемой частью.

Выпускник также должен обладать прочими компетенциями в соответствии с указанным

ФГОС, которые приведены и охарактеризованы в нём, инкорпорированы в настоящий документ путём отсылки, и являются его неотъемлемой частью.

Новые квалификации (ОТФ/ТФ), планируемые к приобретению в результате обучения по настоящей программе, приведены и охарактеризованы в п. 5. настоящего документа.

Приведённая информация о цели и результатах обучения является основой для разработки рабочих программ, оценочных материалов и иных компонентов дополнительной профессиональной программы.

1. Учебный план

№ п/п	Наименование Модулей, дисциплин	Всего, час.	В том числе:				
			Лекции	Практические занятия (семинары)	Самостоятельн ая работа	Контрольные задания	Форма контроля
1	Модуль 1 Старт в Data Engineering	128	32	32	54	10	проект
	или Модуль 1. SQL для прикладных задач	128	32	32	32	10	проект
2	Модуль 2. Python: основы языка	56	14	14	24	4	Проверочны е задания
3	Модуль 3. Инструменты работы с большими данными	128	24	24	70	10	проект
4	Итоговая квалификационна я работа	35			35		
	ИТОГО ПРОГРАММА	347	70	70	183	24	

Срок освоения настоящей программы профессиональной переподготовки превышает минимально допустимый срок освоения 250 часов (либо другой установленный актуальными нормативными документами соответствующего Министерства срок).

2. Календарный учебный график

Модуль 1 Старт в Data Engineering или Модуль 1. SQL для прикладных задач

Обучение по установленной форме: очная с применением дистанционных технологий																	
Количество академических часов: 128																	
Порядковый № занятия (по горизонтали)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Лекции	2	2	2	2	2	2	2	2	2	4	0	2	2	2	2	2	
Практические занятия	2	2	2	2	2	2	2	2	2	0	4	2	2	2	2	2	
Самостоятельная работа	2	2	4	4	4	4	2	2	2	2	2	2	2	2	2		16
Контрольные задания	0,5	0,5	0,75	0,75	0,75	0,75	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	1	1	
Итого	6,5	6,5	8,75	8,75	8,75	8,75	6,5	8,5	7	5	16						

Модуль 2. Python: основы языка

	Обучение по установленной форме: очная, с применением дистанционных технологий						
	Количество академических часов: 56						
Порядковый № занятия (по горизонтали)	1	2	3	4	5	6	7
Лекции	2	2	2	2	2	2	2
практические	2	2	2	2	2	2	2
самостоятельные	4	3	3	4	3	4	3
контрольные задания	0	1	1	0	1	0	1
Итого	8	8	8	8	8	8	8

Модуль 3 Инструменты работы с большими данными

Обучение по установленной форме: очная с применением дистанционных технологий													
Количество академических часов: 128													
Порядковый № занятия (по горизонтали)	1	2	3	4	5	6	7	8	9	10	11	12	
Лекции	2	2	2	2	2	2	2	2	2	2	2	2	
Практические занятия	2	2	2	2	2	2	2	2	2	2	2	2	
Самостоятельная работа	4	4	4	4	4	4	4	4	4	4	4	4	22
Контрольные задания	1	1	1	1	1	1	1	1	0,5	0,5	0,5	0,5	
Итого	9	8,5	8,5	8,5	8,5	22							

Итоговая квалификационная работа

Обучение по установленной форме: очная с применением дистанционных технологий	
Количество академических часов: 35	
Итоговая квалификационная работа	35
Итого	35

11. Рабочие программы (учебная программа)

Модуль 1 Старт в Data Engineering или Модуль 1. SQL для прикладных задач

№ п/п	Содержание обучения, наименование и тематика практических занятий (вебинаров), самостоятельных работы	Объем, ак.час.
1.1	Раздел I. SQL базовый	48
1.1.1	Базовый синтаксис SQL	6,5
	Лекция. Базовый синтаксис SQL (синтаксис запроса, выборка, проекция)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.1.2	Создание и заполнение таблиц	6,5
	Лекция. Создание и заполнение таблиц (понятие DDL и DML, создание таблиц, изменение таблиц, добавление данных, изменение данных, удаление данных)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	1
	Выполнение контрольных заданий.	0,5
1.1.3	Объединение данных	8,75
	Лекция. Объединение данных (Union / Union All ,Особенности Join, Типы Join и различия между ними)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	4

	Выполнение контрольных заданий.	0,75
1.1.4	Агрегация	8,75
	Лекция. Агрегация (агрегация, группировка having, применение подзапросов, оператор case)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0,75
1.1.5	Очистка данных	8,75
	Лекция. Очистка данных (строковые методы, очистка данных, отработка очистки данных)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0,75
1.1.6	Регулярные выражения	8,75
	Лекция. Регулярные выражения (синтаксис регулярных выражений, методы работы с регулярными выражениями, отработка очистки данных)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0,75
1.2	Раздел II: SQL продвинутый	80
1.2.1.	Базовый синтаксис Python 1	6,5

	Лекция. Базовый синтаксис Python 1 (переменные, типы данных, ввод и вывод данных, условные операторы, циклы)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.2	Базовый синтаксис Python 2	6,5
	Лекция. Базовый синтаксис Python 2 (работа с функциями, создание модулей и применение их)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.3	Библиотека SQLite	6,5
	Лекция. Библиотека SQLite (методы работы с SQLite библиотекой, построение процессов CRUD)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.4	Оконные функции I (область применения, синтаксис оконных функций, решение прикладных задач)	6,5
	Лекция. Оконные функции I (область применения, синтаксис оконных функций, решение прикладных задач)	4
	Практическое занятие по теме лекции	0
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.5	Оконные функции II	6,5

	Лекция.	0
	Практическое занятие по теме лекции. решение прикладных задач	4
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.6	Паттерны хранения данных	6,5
	Лекция. Паттерны хранения данных (логическое удаление, исторические данные)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.7	Инкрементальная загрузка	6,5
	Лекция. Инкрементальная загрузка (процесс инкрементальной загрузки, реализация процесса с нуля на примере данных, обсуждение сложностей в процессе)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.8	Оптимизация	6,5
	Лекция. Оптимизация (анализ запроса, алгоритмы запросов индексов, как и когда их использовать, синтаксис работы с индексами)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0,5
1.2.9	Разбор задач с собеседования	7

	Лекция. Разбор задач с собеседования (как подходить к решению задачи, какие бывают подводные камни, какие темы стоит повторить)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
1.2.10	Повторение пройденного материала	5
	Лекция. Повторение пройденного материала (повторение пройденных тем)	2
	Практическое занятие по теме лекции	2
	Самостоятельная работа.	
	Выполнение контрольных заданий.	1
1.2.11	Промежуточная аттестация	16
	Самостоятельная работа. Выполнение проекта	16
	Итого:	128

Модуль 2. Python: основы языка

№ п/п	Содержание обучения, наименование и тематика практических занятий (вебинаров), самостоятельных работы	Объем, ак.час.
1	Знакомство с Python (часть 1).	8
	Лекция. Введение в Python. работа в Jupyter Notebook и Google Colab, числовые типы данных, условия и циклы.	2
	Практическое занятие..	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2	Управление вычислениями. Контейнеры, итераторы, генераторы	8
	Лекция. Set, dict; enumerate, zip, map; list, dict, set comprehensions; itertools, collections.	2

	Практическое занятие..	2
	Самостоятельная работа.	3
	Выполнение контрольных заданий.	1
3	Функциональное программирование в Python	8
	Лекция. Базовый синтаксис: объявление/вызов, аргументы по умолчанию. Рекурсия. Передача по ссылке. Конструкции args, kwargs. Распаковка контейнеров при вызове. Генераторы. Области видимости. Замыкания. Объектная природа функций. Декораторы. Лямбда-функции	2
	Практическое занятие.	2
	Самостоятельная работа.	3
	Выполнение контрольных заданий.	1
4	Классы в Python	8
	Лекция. Основные принципы ООП. Объявление класса, создание экземпляра. Атрибуты и методы класса. Наследование. Множественное наследование. Ромбовидное наследование. Приватность атрибутов	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
5	Объектно-ориентированный подход к созданию итераторов и генераторов	8
	Лекция. Объектно-ориентированный подход к созданию итераторов и генераторов. Магические методы. Перегрузка операторов	2
	Практическое занятие.	2
	Самостоятельная работа.	3
	Выполнение контрольных заданий.	1
6	Исключения, модули, встроенный модуль	8
	Лекция. Устройство исключений, создание и работа с модулями, встроенный модуль python, работа с файловой системой	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
7	Модуль NumPy	8
	Лекция. введение в NumPy, оптимизация кода, работа со строками.	2
	Практическое занятие.	2

	Самостоятельная работа.	3
	Выполнение контрольных заданий.	1
	Итого	56

Модуль 3. Инструменты работы с большими данными

<i>№п/п</i>	Содержание обучения, наименование и тематика практических занятий (вебинаров), самостоятельных работы	Объем, ак.час.
2.1	Введение. Зачем нужны большие данные. Распределенные файловые системы.	8
	Лекция. Файловые системы HDFS. Их составляющие. Их достоинства, недостатки и сфера применения. Чтение и запись в HDFS. HDFS APIs: Web API, shell, Python API.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.2	MapReduce, начало	10
	Лекция. Основная идея. MapReduce на картах. Стадии MapReduce-задачи. Самая известная реализация MapReduce - Hadoop. Роли серверов в кластере Hadoop. API для работы с Hadoop (обзор Java API и Hadoop Streaming более подробно). Счётчики с Hadoop.	2
	Практическое занятие..	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	2
2.3	MapReduce, продолжение.	8
	Лекция. Дополнительные элементы MapReduce-задачи (Combiner, comparator, partitioner). Оптимизация MapReduce-задач, Distributed cache. Join'ы в MapReduce. Планирование задач в Hadoop.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.4	SQL поверх MapReduce.	10
	Лекция. Виды таблиц в Hive, типы данных, трансляция Hive-запросов в	2

	MapReduce-задачи. Аналитические функции в Hive. Примерные расчёты в Hive.	
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	2
2.5	Расширения Hive.	8
	Лекция. Hive streaming, User defined functions. Оптимизация запросов в Hive (партиционирование, бакетирование, работа с несбалансированными данными). Оптимизация Join в Hive. Форматы данных.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.6	Spark RDD.	10
	Лекция. Вычисления в оперативной памяти. RDD API: трансформации, действия. Типы трансформаций. Аккумуляторы, сравнение со счётчиками в Hadoop. Broadcast-переменные, сравнение с Distributed cache в Hadoop. Кеширование. Join'ы в Spark.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	2
2.7.	Spark DataFrames	8
	Лекция. Spark DF и pandas. Spark SQL и интеграция с Hive. User defined функции.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.8.	Обработка данных в реальном времени	10
	Лекция. Принципы обработки данных в реальном времени. Её отличия от "батч"-обработки. Spark Streaming. Spark Structured streaming.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	2
2.9	Чтение данных из внешних источников.	8

	Лекция. Apache Kafka. Интеграция Kafka со Spark Streaming. Разбор практических кейсов	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.10	NoSQL поверх больших данных. HBase	9
	Лекция. CAP-теорема. Google Bigtable и HBase Архитектура Hbase. Чтение, запись данных в HBase, Cassandra. Схема хранения, компактификация	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	1
2.11	NoSQL поверх больших данных. Cassandra	8
	Лекция. Amazon dynamo и Cassandra. Архитектура Cassandra, отличие от Hbase-подобных систем. Интеграция Cassandra и Spark.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	0
2.12	Администрирование экосистемы BigData.	9
	Лекция. Администрирование экосистемы BigData.	2
	Практическое занятие.	2
	Самостоятельная работа.	4
	Выполнение контрольных заданий.	1
2.13	Промежуточная аттестация.	22
	Итого	128

3.	<i>Итоговая квалификационная работа</i>	35
	Самостоятельная работа	35
	ИТОГО часов	347

Содержание дополнительной профессиональной программы направлено на достижение результатов её целей (планируемых результатов).

9. Организационно-педагогические условия реализации дополнительной профессиональной программы

9.1. Требования к квалификации педагогических кадров/представителей предприятий и организаций, обеспечивающих реализацию образовательного процесса

Требования к образованию педагогических и иных работников, а также (при наличии) требования к освоению ими дополнительных профессиональных программ, опыту работы в области профессиональной деятельности, соответствующей направленности ДПП:

высшее образование, трудовой и (или) педагогический стаж от полугода, наличие опыта практической работы в сфере деятельности, совпадающей с направлением преподавания

9.2. Требования к материально-техническим условиям

Перечень кабинетов, лабораторий, мастерских, тренажеров и др., обеспечивающих проведение всех предусмотренных программой видов занятий:

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Лекции	<p>Слушателю необходимо наличие доступа в сеть интернет, компьютер.</p> <p>Преподавателю курса необходимо наличие доступа администратора курса на LMS-платформе к материалам курса.</p>
Информационно-коммуникационная платформа дистанционных семинаров	Практические занятия (дистанционные семинары)	<p>Слушателю необходимо наличие доступа в сеть интернет, компьютер.</p> <p>Преподавателю курса необходимо оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет.</p>

Система дистанционного провайдера обучения открытых онлайн курсов	Самостоятельная работа	Наличие компьютера и доступа в сеть интернет.
Система дистанционного провайдера обучения массовых открытых онлайн курсов	Промежуточный, рубежный контроль, Итоговая аттестация	Наличие компьютера и доступа в сеть интернет.

9.3. Требования к информационным и учебно-методическим условиям

Список литературы:

МОДУЛЬ 1:

1. Bob Ward. SQL Server 2019 Revealed: Including Big Data Clusters and Machine Learning.
2. Грофф, Джеймс . SQL : Энциклопедия : пер. с англ. / Д. Р. Грофф, П. Н. Вайнберг. — 3-е изд. — СПб. : Питер. 2016
3. Дэви Силен, Арно Мейсман, Мохамед Али. Основа Data Science и BigData. Python и наука о данных. Питер. 2017

МОДУЛЬ 2:

1. Марк Лутц «Изучаем Python»
2. Эрик Мэтиз «Изучаем Python. Программирование игр, визуализация данных, веб-приложения»
3. Пол Бэрри «Изучаем программирование на Python»
4. Эл Свейгарт «Автоматизация рутинных задач с помощью Python. Практическое руководство для начинающих»
5. Майкл Доусон «Программируем на Python»
6. Зед Шоу «Легкий способ выучить Python»
7. John M. Zelle «Python Programming: An Introduction to Computer Science»
8. Дэн Бейдер «Чистый Python. Тонкости программирования для профи»

МОДУЛЬ 3:

Основная

1. Tom White, Hadoop: The Definitive Guide, 4th Edition Storage and Analysis at Internet Scale, Publisher: O'Reilly, Media Release Date: April 2015

2. Karau, H., Konwinski, A., Wendell, P., & Zaharia, M, Learning Spark: Lightning-Fast Big Data Analysis
3. Holden Karau et al. Learning Spark. Lightning-fast Data Analytics. Preview Edition. O'Reilly, 2015.

Дополнительная

1. Chuck Lam. Hadoop in Action. New York: Manning Publications co., 2011.
2. Alex Holmes , "Hadoop in Practice", 2012
3. Martin Fowler, "NoSQL", 2013
4. Eric Redmond, Jim R. Wilson, "Seven Databases in Seven Weeks", 2012
5. Jonathan Leibiusky, "Getting Started with Storm", 2012
6. Donald Miner, Adam Shook. MapReduce Design Patterns. O'Reilly, 2013.
7. Arun C. Murthy et al. Apache Hadoop YARN. Addison-Wesley, 2014.
8. Spark Core Programming. Tutorialspoint, 2015.
9. Dario Simonassi, Gabriel Eisbruch, Jonathan Leibiusky. Getting Started with Storm. O'Reilly, 2012.
10. Edward Capriolo, Dean Wampler, and Jason Rutherglen. Programming Hive. O'Reilly, 2012.
11. Lars George. HBase: The Definitive Guide. O'Reilly, 2011.
12. Eben Hewitt. Cassandra: The Definitive Guide. O'Reilly, 2011.

9.4. Общие требования к организации образовательного процесса

Форма обучения – очная с применением дистанционных технологий

Режим обучения – 2 раз(а) в неделю по 4 академических часа.

Преподаватель проводит практические занятия дистанционно в форме вебинаров с использованием платформы ZOOM (или аналогичной).

Самостоятельная работа выполняется слушателем в удобном для слушателя режиме

В Таблице ниже описаны образовательные технологии.

№ п/п	Вид занятия	Форма проведения занятий	Цель
1	Лекция	Изложение материала посредством лекций, обсуждение общих вопросов по тематике курса	Ознакомление слушателей с базовым материалом по тематике курса
2	Практическое занятие	Выполнение практических заданий, получение обратной связи от преподавателя.	Практическое освоение теоретических знаний, Разбор ситуаций, а также углубление знаний по курсу
3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы. Выполнение тренировочных тестов и заданий.	Углубление знаний по курсу. Применение знаний к своей организации / проекту
4	Выполнение	Выполнение тестов, проверочных	Практическое освоение

	контрольных заданий	заданий.	теоретических знаний, контроль освоения материалов.
5	Итоговая аттестация	Подготовка итоговой аттестационной работы.	Практическое освоение теоретических знаний, контроль освоения материалов. Получение практического результата в своих организациях / проектах

10. Формы аттестации и оценочные материалы по программе

Оценка качества освоения программы проводится по пятибалльной системе: «отлично», «хорошо», «удовлетворительно» по результатам промежуточного контроля (тестирование, проверочные задания), контроля посещаемости практических занятий (вебинаров) и результатам итоговой аттестации.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 70% баллов от итоговой оценки) в процессе обучения и успешной сдачи экзамена. При этом баллы за экзамены начисляются только при достижении 50% порога при прохождении каждого экзаменационного испытания. После аттестации слушатель получает оценку «отлично», если набрано не меньше 80% баллов от возможного максимума, «хорошо», если набрано не меньше 65%, «удовлетворительно», если набрано не меньше 40%, в противном случае слушатель курс не сдает.

Результат тестирования, решения проверочных заданий и написания кода проверяется автоматически системой на образовательной платформе. Итоговая аттестационная работа проверяется преподавателем

Составляющие процесса обучения, которые оцениваются в ходе обучения, и их вклад в итоговую оценку представлены в Таблице ниже.

Таблица – Составляющие процесса обучения

	Основные показатели оценки	Вклад в итоговую оценку
1	Основной курс обучения на образовательной платформе	50%
2	Практические занятия	10%
3	Итоговая аттестация	40%

Оценочные материалы:

1) Примеры тестовых заданий к модулю 1 “Старт в Data Engineering”

1. Имеется база данных интернет-магазина. В этой БД имеются таблицы:

Покупатели – хранит информацию по клиентам (ФИО, логин, пароль и т.п.);

Заказы – содержит номер заказа, комментарий, дату доставки, покупателя (который сделал заказ) и т.д.;

Товары – id товара, наименование, цена, остатки и т.п.;

Заказанные товары – хранит информацию о заказанных товарах (заказ, товар, количество).

- Один к одному
- Один ко многим
- Многие ко многим
- Таблицы никак не связаны

2. В какой нормальной форме находится база данных?

Имеется БД из одной таблицы «Сотрудники».

Таблица содержит столбцы:

Id – уникальный идентификатор сотрудника;

ФИО – содержит фамилию, имя и отчество сотрудника;

Дата рождения – дата рождения пользователя;

Отдел_id – идентификатор отдела, в котором пользователь работает;

Отдел – наименование отдела, в котором пользователь работает.

- БД в первой нормальной форме
- БД во второй нормальной форме
- БД в третьей нормальной форме
- Ни одна нормальная форма не соблюдена

3. Чем различаются запросы с UNION и UNION ALL?

- UNION соединить таблицы по принципу внутреннего соединения, но при условии, что строки не будут полностью совпадать. UNION ALL объединит каждую строку из таблицы слева с каждой строкой из таблицы справа
- UNION объединяет наборы строк, полностью исключая те строки, которые дублируются, т.е. строка из первого набора, которая есть во втором наборе, вообще не будет выдана ни из одного набора. UNION ALL объединяет наборы с сохранением всех строк
- UNION объединяет наборы строк, исключая дубли, т.е. каждая строка после объединения будет уникальной. UNION ALL объединяет наборы с сохранением всех строк
- Различия только в написании операции. UNION - это короткий стиль написания UNION ALL

4. Какая агрегатная функция используется для расчета суммы?

- SUM
- AVG
- COUNT

2) Примеры тестовых заданий к модулю 2 “Python: основы языка”

1. Даны четыре действительных числа: x_1, y_1, x_2, y_2 . Напишите функцию `distance(x1, y1, x2, y2)`, вычисляющая расстояние между точкой (x_1, y_1) и (x_2, y_2) . Считайте четыре действительных числа и выведите результат работы этой функции.

Если вы не знаете, как решить эту задачу, то вы, возможно, не изучали в школе теорему Пифагора. Попробуйте прочитать о ней [на Википедии](#).

2. Дано действительное положительное число a и целое неотрицательное число n . Вычислите a^n не используя циклы, возведение в степень через `**` и функцию `math.pow()`, а используя рекуррентное соотношение $a_n = a \cdot a_{n-1}$.

Решение оформите в виде функции `power(a, n)`.

3. Во входной строке записана последовательность чисел через пробел. Для каждого числа выведите слово `YES` (в отдельной строке), если это число ранее встречалось в последовательности или `NO`, если не встречалось.
4. Даны значения двух моментов времени, принадлежащих одним и тем же суткам: часы, минуты и секунды для каждого из моментов времени. Известно, что второй момент времени наступил не раньше первого. Определите, сколько секунд прошло между двумя моментами времени.

Программа на вход получает три целых числа: часы, минуты, секунды, задающие первый момент времени и три целых числа, задающих второй момент времени.

Выведите число секунд между этими моментами времени.

3) Примеры тестовых заданий к модулю 3 «Инструменты работы с большими данными»:

1. Какие семантики доставки сообщений вы знаете? Хотя бы для двух из них приведите пример реальных систем.
2. Что такое Compaction в HBase? Какие они бывают и чем отличаются?
3. Назовите основное отличие архитектуры HBase от архитектуры Cassandra. Какие плюсы и минусы имеет архитектура Cassandra по сравнению с HBase? Какие из преобразований Hive позволяют изменять количество строк в таблице? * UDF *UDAF * UDTF * PTF (оконные ф-ции)
4. В таблице HBase в качестве ключа таблицы используется доменное имя. Для каких целей удобно хранить домен в обратном порядке (market.yandex.ru —> ru.yandex.market)?
5. Подходы к обеспечению обновления кода Spark Streaming с сохранением семантики доставки. Семантика и плюсы/минусы для каждого подхода. Зачем нужны эти подходы к обновлению кода с сохранением семантики, если в Spark Streaming есть checkpoint?
6. Есть стандартный wordcount: mapper разбивает на слова, reducer суммирует. Какими способами в Hadoop можно его ускорить?