

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ливанов Дмитрий Викторович
Должность: Исполняющий обязанности ректора
Дата подписания: 19.03.2026 13:47:03
Уникальный программный ключ:
1a323b7b5f4d396b27b33b5facf7dd71194562c4

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

ВЫПИСКА ИЗ ПРОТОКОЛА № 2

заседания учебно-методического совета от «24» сентября 2021 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 общеобразовательную программу «Инструменты работы с большими данными».

Решение принято единогласно заочным голосованием.

Форма проведения заседания: заочная.

Председатель УМС МФТИ



А.А. Воронов

Ученый секретарь УМС МФТИ



М.В. Березникова

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

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

УТВЕРЖДАЮ

Исполняющий обязанности ректора

д-р физ.-мат. наук



Д. В. Ливанов

«24» сентября 2021 г.

Дополнительная профессиональная программа
Программа повышения квалификации
«Инструменты работы с большими данными»

Москва 2021

Оглавление

1. Рабочая группа	2
2. Профессиональные стандарты, учтённые в содержании программы	2
3. Квалификационные требования, учтённые в содержании программы	2
4. Описание перечня профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения	3
5. Цель программы	3
6. Планируемые результаты обучения по программе	3
7. Учебный план.	4
8. Календарный учебный график.	6
9. Рабочие программы (учебная программа)	6
10. Организационно-педагогические условия реализации дополнительной профессиональной программы	9
10.1. Требования к квалификации педагогических кадров/представителей предприятий и организаций, обеспечивающих реализацию образовательного процесса.	9
10.2. Требования к материально-техническим условиям.	10
10.3. Требованиям к информационным и учебно-методическим условиям.	11
10.4. Общие требования к организации образовательного процесса.	11
11. Формы аттестации и оценочные материалы по программе	13

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 ФПМИ МФТИ

Ивченко Олег Николаевич

Старший преподаватель кафедры АТП ФПМИ МФТИ

Иванова Анастасия Сергеевна

Руководитель проектов учебно-методической лабораторией 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:

ПС 06.042 - Специалист по большим данным

Обобщённые трудовые функции (ОТФ) профессионального стандарта (ПС), соответствующие дополнительной профессиональной программе (ДПП) и уровни их квалификации**:

Обобщённые трудовые функции (ОТФ) профессионального стандарта ПС 06.042 - Специалист по большим данным	Уровень квалификации
Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	6
Управление разработкой продуктов, услуг и решений на основе больших данных	8

Уровень квалификации отобранных ОТФ не превышает возможности ДПП, связанные, прежде всего, с уровнем квалификации деятельности, овладение или совершенствование которой предусмотрено ДПП, сроком ее освоения и исходным уровнем и направленностью (профилем) имеющегося у слушателей профессионального образования.

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 соответствующих данной программе ОТФ (п. 2. 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. Описание перечня профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения

Имеющаяся квалификация и (или) уровень образования (требования к слушателям):
Высшее и незаконченное высшее образование.

Компетенции представлены в соответствии с направлением подготовки (ФГОС ВО):

Перечень профессиональных компетенции (*описание, ПК*) представлен ниже. ПК характеризуются:

01.03.02 «Прикладная математика и информатика (уровень бакалавриата)»

Перечень профессиональных компетенций, в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения:

способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям (ПК-1)

способность понимать, совершенствовать и применять современный математический аппарат (ПК-2)

способность осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в информационно-телекоммуникационной сети "Интернет" (далее - сеть "Интернет") и в других источниках (ПК-5)

способность к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения (ПК-7)

5. Цель программы

Реализация программы повышения квалификации направлена на совершенствование/получение новой компетенции, необходимой для профессиональной деятельности/повышение профессионального уровня в рамках имеющейся квалификации.

Целью реализации дополнительной профессиональной программы повышения квалификации (далее - Программа) «Инструменты работы с большими данными» является приобретение и совершенствование профессиональных, общепрофессиональных, универсальных и общекультурных компетенций слушателей, связанных с овладением алгоритмами, парадигмами и инструментами для пакетной и потоковой обработки больших объёмов данных.

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.

6. Планируемые результаты обучения по программе

Выпускник должен обладать ПК, соответствующими видам деятельности (ВД).

п/п	Вид деятельности	Выпускник должен обладать следующими ПК:
1	Научно-исследовательская деятельность	способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям (ПК-1)
		способностью понимать, совершенствовать и применять современный математический аппарат (ПК-2)
2	Проектная и производственно-технологическая деятельность	способностью осуществлять целенаправленный поиск информации о новейших научных и технологических достижениях в информационно-телекоммуникационной сети "Интернет" (далее - сеть "Интернет") и в других источниках (ПК-5)
		способностью к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения (ПК-7)
3	Организационно-управленческая деятельность	способностью составлять и контролировать план выполняемой работы, планировать необходимые для выполнения работы ресурсы, оценивать результаты собственной работы (ПК-9)

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

7. Учебный план.

№ п/п	Наименование модулей	Всего, час.	в том числе:				Форма контроля
			лекции	практические занятия (семинары)	самостоятельная работа	контрольные задания	
1	Apache Hadoop (HDFS, MapReduce, YARN)	33	10	10	8	5	проверочное задание
2	Apache Hive	14	4	4	4	2	проверочное задание
3	Apache Spark (в том числе технологии)	35	10	10	10	5	проверочное задание

	Streaming). Apache Kafka.						
4	NoSQL в больших данных (Apache Cassandra, Apache HBase)	28	8	8	8	4	тестирование
5	Итоговая аттестация	18	0	0	18	0	проект
	Итого:	128	32	32	48	16	

Срок освоения настоящей программы повышения квалификации превышает минимально допустимый срок освоения 16 часов (либо другой установленный актуальными нормативными документами соответствующего Министерства срок) или равен ему.

8. Календарный учебный график.

	Обучение по установленной форме*: очная, с применением электронного обучения и дистанционных образовательных 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	
Лекции***	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
практические	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
самостоятельные	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	18
промежуточный контроль	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	

9. Рабочие программы (учебная программа)

№ п/п	Содержание обучения, наименование и тематика практических занятий (вебинаров), самостоятельных работы	Объем, ак.час.
	Модуль 1. Apache Hadoop (HDFS, MapReduce, YARN)	33
1	Введение. Зачем нужны большие данные. Распределённые файловые системы.	6

	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.	2
	Выполнение контрольных заданий.	0
3	MapReduce, начало	7
	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.	2
	Выполнение контрольных заданий.	1
3	MapReduce, продолжение.	7
	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.	2
	Выполнение контрольных заданий.	1
4	Дополнительные элементы MapReduce	7
	Лекция. Дополнительные элементы MapReduce-задачи (Combiner, Comparator, Partitioner). Типы Join'ов и их реализации в парадигме MR. Паттерны проектирования MR (pairs, stripes, составные ключи)..	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
5	Планирование задач в MapReduce. YARN	6
	Лекция: Планирование задач в MapReduce. YARN	2

	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	0
	Модуль 2. Apache Hive	14
6	SQL over BigData: Повторение SQL. HiveQL vs. SQL. Hive.	7
	Лекция. SQL over BigData: Повторение SQL. HiveQL vs. SQL. Hive. Виды таблиц в Hive, типы данных, трансляция Hive-запросов в MapReduce-задачи.	2
	Практическое занятие.	2
	Самостоятельная работа.	7
	Выполнение контрольных заданий.	1
7.	Оптимизация запросов в Hive	7
	Лекция. Оптимизация запросов в Hive (партиционирование, бакетирование, оптимизация Join'ов). Примерные расчеты в Hive. Расширения Hive (select-transform и UDF). Не только Hive: Apache Pig, Cloudera Impala, Presto	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
	Модуль 3. Apache Spark (в том числе технологии Streaming). Apache Kafka.	35
8.	Итеративная обработка больших данных на Apache Spark,	7
	Лекция. Итеративная обработка больших данных на Apache Spark, отличия Spark от MapReduce. Spark RDD API	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
9	Spark Dataframe API	7
	Лекция. Spark Dataframe API	2
	Практическое занятие.	2

	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
10	Принципы обработки данных в реальном времени (технологии Streaming).	7
	Лекция. Принципы обработки данных в реальном времени (технологии Streaming). Её отличия от "батч"-обработки	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
11	Распределенная очередь данных. Apache kafka.	7
	Лекция. Рспределенная очередь данных. Apache kafka. Архитектура, отличия алгоритмов репликации от HDFS. Роль лидера в Kafka	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
12.	Связь Kafka с обработкой данных в реальном времени.	7
	Лекция. Связь Kafka с обработкой данных в реальном времени. Kafka Streams	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
	Модуль 4. NoSQL в больших данных (Apache Cassandra, Apache HBase)	28
13.	NoSQL в BigData.	7
	Лекция. NoSQL в BigData. CAP-теорема. Google Bigtable и Apache HBase	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
14.	Amazon Dynamo и Apache Cassandra.	7
	Лекция. Amazon Dynamo и Apache Cassandra.	2

	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
15.	От NoSQL снова к SQL. Google Spanner и CockroachDB	7
	Лекция. От NoSQL снова к SQL. Google Spanner и CockroachDB	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
16.	Администрирование экосистемы BigData	7
	Лекция. Amazon Dynamo и Apache Cassandra.	2
	Практическое занятие.	2
	Самостоятельная работа.	2
	Выполнение контрольных заданий.	1
	Итоговая аттестация.	18
	Итого	128

** указываются модули программы и их содержание (тезисно). Лекции, практические занятия, самостоятельные работы, выполнение контрольных заданий (при наличии).*

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

10. Организационно-педагогические условия реализации дополнительной профессиональной программы

10.1. 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

10.2. 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	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо наличие доступа администратора курса на LMS-платформе к материалам курса. Удалённый кластер с Linux Ubuntu 16.04, последней версией Cloudera Manager, в который нужно встроить такие сервисы: HDFS, YARN, Hive, Spark2 on YARN, HBase, Zookeeper, Kafka
Информационно-коммуникационная платформа дистанционных семинаров	Практические занятия (дистанционные семинары)	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет. Удалённый кластер с Linux Ubuntu 16.04, последней версией Cloudera Manager, в который нужно встроить такие сервисы: HDFS, YARN, Hive,

		Spark2 on YARN, HBase, Zookeeper, Kafka
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Самостоятельная работа	Наличие компьютера и доступа в сеть интернет. Удалённый кластер с Linux Ubuntu 16.04, последней версией Cloudera Manager, в который нужно встроить такие сервисы: HDFS, YARN, Hive, Spark2 on YARN, HBase, Zookeeper, Kafka
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Рубежный контроль, Итоговая аттестация	Наличие компьютера и доступа в сеть интернет. Удалённый кластер с Linux Ubuntu 16.04, последней версией Cloudera Manager, в который нужно встроить такие сервисы: HDFS, YARN, Hive, Spark2 on YARN, HBase, Zookeeper, Kafka

10.3. Требованиям к информационным и учебно-методическим условиям.

Основная литература

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 литература

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 Leibusky, "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 Leibusky. 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.

10.4. Общие требования к организации образовательного процесса.

Форма обучения – очная, с применением электронного обучения и дистанционных образовательных технологий

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

Преподаватель проводит практические занятия: *в форме вебинаров с использованием платформы ZOOM*

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	Практические занятия	Выполнение практических заданий, получение обратной связи от преподавателя.	Практическое освоение теоретических знаний, а также углубление знаний по курсу
3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы. Выполнение тренировочных тестов и заданий.	Углубление знаний по курсу.
4	Выполнение контрольных заданий	Выполнение тестов, проверочных заданий.	Практическое освоение теоретических знаний, контроль освоения материалов.
5	Итоговая аттестация	Защита выпускной квалификационной работы.	Практическое освоение теоретических знаний, контроль освоения материалов.

11. Формы аттестации и оценочные материалы по программе

Оценка качества освоения программы проводится по пятибалльной системе по результатам текущего контроля - индивидуальная работа слушателя на практических занятиях (вебинаров), промежуточного контроля (Зачет/дифференцированный зачет), 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 аттестационной работы. 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. Экзаменационная работа проверяется преподавателем.

Составляющие процесса обучения, которые оцениваются в ходе обучения, и их вклад в итоговую оценку представлены в таблице 6

Таблица 6 – Составляющие процесса обучения

№ п/п	Основные показатели оценки	Вклад в итоговую оценку
1	Текущий контроль	30%
2	Промежуточный контроль	50%
3	Итоговая аттестация	20%

Оценочные материалы:

Пример тестового задания

1. Какие семантики доставки сообщений вы знаете? Хотя бы для двух из них приведите пример реальных систем.
2. Что такое Compression 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?

Согласовано

Директор ФПМИ, д.ф.-м.н.

_____ А.М. Райгородский

КВАЛИФИКАЦИЯ И ОПЫТ ПРИВЛЕКАЕМОГО ПРЕПОДАВАТЕЛЬСКОГО СОСТАВА ДЛЯ РЕАЛИЗАЦИИ ПРОГРАММЫ
«Инструменты работы с большими данными»

Ф.И.О. лектора, год рождения	Информация об образовании, полученном в соответствии с образовательными программами высшего профессионального образования, дополнительного профессионального образования (в т.ч. о наличие званий и ученых степеней) и т.д.	Место работы, занимаемая должность в настоящий момент, общий трудовой стаж, педагогический стаж	Опыт преподавания и консультирования по предмету, согласующемуся с направлением лота (перечислить), преподавательский стаж	Наличие опыта практической работы в отечественных и зарубежных организациях в сфере деятельности, совпадающей с направлением преподавания
Ивченко Олег Николаевич	<p>Высшее образование - специалитет, магистратура, специальность: Прикладная математика и информатика, квалификация: Магистр, ФГБОУ ВПО "Московский физико-технический институт (национальный исследовательский университет)"</p> <p>Дополнительное профессиональное образование, курс: Математическое моделирование для анализа данных, ФГБОУ ВПО "Московский физико-технический институт (национальный исследовательский университет)"</p>	<p>Должность: старший преподаватель</p> <p>Общий стаж: 4 года</p> <p>Педагогический стаж: 4 года.</p>	<p>Преподавательский стаж: 4 года</p>	<p>Старший преподаватель кафедры АТП ФПМИ</p> <p>Разработчик системы NJudge - системы автоматизированного тестирования Nadoor-приложений, системный администратор инфраструктуры больших данных в группе Яндекс-CERN</p>

Пояснительная записка
к разработке и реализации дополнительной
профессиональной программы повышения квалификации
«Инструменты работы с большими данными»

Данный курс призван дать фундаментальные знания в области хранения и обработки данных, для работы с которыми недостаточно одной машины со стандартными аппаратными характеристиками. Примерами таких данных могут быть логи пользователей web-сервиса, коллекции медиа-файлов или статей Википедии. Сейчас эти подходы активно применяются в компаниях, для которых критично провести анализ больших объёмов данных в кратчайшие сроки. Это могут быть компании, владеющие: - поисковиками (например, Google, Яндекс, Microsoft, Yahoo! и др.), - социальными сетями и блогами (Facebook, Twitter, ВКонтакте, LinkedIn и др.), - рекомендательными сервисами (например, Кинопоиск от Яндекс). Практическую часть данного курса составляют программы, разрабатываемые с использованием фреймворков экосистемы Hadoop. Будет рассмотрена как батчевая обработка данных, так и обработка в реальном времени.

Данный курс нацелен на овладение алгоритмами, парадигмами и инструментами для пакетной и потоковой обработки больших объёмов данных.

Задача этого курса - научить пользоваться навыкам проектирования архитектур, применения специализированных инструментов и разработки программных систем для работы с большими объемами данных.

В данном курсе мы познакомимся с технологиями работы с большими данными, типами хранилищ больших объёмов данных, подходами к потоковой и пакетной обработке данных, принципами трансляции высокоуровневых языков программирования (SQL-подобных и функциональных) в последовательность задач на Hadoop кластере.

На выходе слушатель сможет эффективно пользоваться распределенной файловой системой, запускать задачи на Hadoop кластере, писать задачи для запуска на Hadoop кластере с помощью нативного Java-интерфейса, писать задачи для запуска на Hadoop кластере с помощью любого другого языка программирования (с помощью инструментария Hadoop streaming), пользоваться высокоуровневыми языками программирования для BigData, для обработки большого объема данных на вычислительном кластере, решать задачи статистики, задачи поиска и индексации, задачи машинного обучения на Hadoop кластере.

Программа «Инструменты работы с большими данными» нацелена на решение следующих научно – образовательных задач:

1. Ознакомление широкого круга слушателей с распределенными файловыми системами;
2. Ознакомление широкого круга слушателей с парадигмой MapReduce;
3. Ознакомление широкого круга слушателей с управлением ресурсами Hadoop-кластера;
4. Ознакомление с применением высокоуровневых языков программирования для работы с большими данными.
5. Освоение Hadoop кластера в целях решения задач машинного обучения, задач статистики, задач поиска и индексации.

Программа будет построена по очному принципу (с применением дистанционных образовательных технологий) в формате вебинаров, которые займут 48 ак.ч.

Преподавателем курса будет являться преподаватель МФТИ, имеющий обширный опыт работы в области машинного обучения и анализа данных.

Взаимодействие со слушателями курса будет осуществляться дистанционно с помощью системы управления образовательным процессом на базе miptx (далее СДО – система дистанционного образования). Посредством СДО слушатели будут получать доступ к интерактивным лекциям в формате видеолекций. Контроль изучения этих материалов будет организован, как автоматически с помощью средств СДО, так и через тестовые задания.

Дистанционный цикл будет проводиться одновременно для всех участников. Задания с открытым вариантом ответа будут проверяться преподавателями курса.

Составители программы:

Райгородский Андрей Михайлович
Доктор физико-математических наук, директор ФПМИ МФТИ

Благодарный Евгений Владимирович
заведующий учебно-методической лабораторией инноватики ФПМИ

Ивченко Олег Николаевич
старший преподаватель кафедры АТП ФПМИ

Иванова Анастасия Сергеевна
руководитель проектов учебно-методической лаборатории инноватики ФПМИ

Директор ФПМИ, д.ф.-м.н.

_____ А.М. Райгородский