

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

Обнинский институт атомной энергетики –

филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный
исследовательский ядерный университет «МИФИ»

(ИАТЭ НИЯУ МИФИ)

Одобрено УМС ИАТЭ НИЯУ МИФИ,
Протокол №2-8/2024 От 30.08.2024

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ
ПО УЧЕБНОЙ ДИСЦИПЛИНЕ**

Обработка больших данных с помощью Spark

Направление подготовки	01.04.02 – Прикладная математика и информатика
Профиль	Математическое моделирование и прикладной анализ данных
Квалификация (степень)	магистр
Форма	очная

Обнинск 2024г.

Программа составлена в соответствии с требованиями образовательного стандарта высшего образования национального исследовательского ядерного университета «МИФИ» по направлению подготовки 01.04.02 – Прикладная математика и информатика. (квалификация (степень) магистр).

Программу составил:

_____ С.В. Ермаков, доцент, к.ф.-м.н, доцент

Рецензент:

_____ Г.Е. Деев, доцент, к.ф.-м.н, доцент

Программа рассмотрена на заседании ОИКС

(протокол № 5/7 от «30» июля от 2024 г.)

Руководитель направления подготовки 01.04.02
«Прикладная математика и информатика»

_____ Ермаков С.В.

« ____ » _____ 2024 г.

Область применения

Фонд оценочных средств (ФОС) – является неотъемлемой частью учебно-методического комплекса учебной дисциплины «Обработка больших данных с помощью Spark» и предназначен для контроля и оценки образовательных достижений обучающихся, освоивших программу данной дисциплины.

Цели и задачи фонда оценочных средств

Целью Фонда оценочных средств является установление соответствия уровня подготовки обучающихся требованиям федерального государственного образовательного стандарта.

Для достижения поставленной цели Фондом оценочных средств по дисциплине «Обобщённые функции» решаются следующие задачи:

- контроль и управление процессом приобретения обучающимися знаний, умений и навыков предусмотренных в рамках данного курса;
- контроль и оценка степени освоения компетенций, предусмотренных в рамках данного курса;
- обеспечение соответствия результатов обучения задачам будущей профессиональной деятельности через совершенствование традиционных и внедрение инновационных методов обучения в образовательный процесс в рамках данного курса.

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

1.1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения ООП магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенций	Результаты освоения ООП Содержание компетенций*	Перечень планируемых результатов обучения по дисциплине**
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	З-УК-1 Знать: методы системного и критического анализа; методики разработки стратегии действий для выявления и решения проблемной ситуации У-УК-1 Уметь: применять методы системного подхода и критического анализа проблемных ситуаций; разрабатывать стратегию действий, принимать конкретные решения для ее реализации В-УК-1 Владеть: методологией системного и критического анализа проблемных ситуаций; методиками постановки цели, определения способов ее достижения, разработки стратегий действий.
ПК-3	способен развивать инновационный потенциал новых научных и научно-технологических разработок	З-ПК-3 Знать основы планирования и организации научных исследований в профессиональной области; методику постановки задач по решению теоретических и прикладных исследовательских проблем; методы и средства научных исследований в профессиональной области, правила и принципы научной этики, методы математического моделирования. У-ПК-3 Уметь оценивать и развивать инновационный потенциал новых научных и научно-технологических разработок, осуществлять постановку задач по решению теоретических и прикладных исследовательских проблем; составить план научных исследований; выдвинуть гипотезы по направлению исследований и соотнести их с полученными результатами; организовать свою научно-исследовательскую работу; определять методы и средства научных исследований для решения конкретных задач в своей предметной области; оценивать результаты исследований, использовать методы математического моделирования

		В-ПК-3 Владеть навыками постановки задач по решению теоретических и прикладных исследовательских проблем; навыками выбора и использования методов и средств научных исследований задач в своей предметной области; навыками методами работы с литературными источниками; методами анализа результатов научных исследований; методами обобщения результатов научных исследований для развития инновационного потенциала новых научных и научно- технологических разработок
--	--	---

1.2. Этапы формирования компетенций в процессе освоения ООП магистратура

Компоненты компетенций, как правило, формируются при изучении нескольких дисциплин, а также в немалой степени в процессе прохождения практик, НИР и во время самостоятельной работы обучающегося. Выполнение и защита ВКР являются видом учебной деятельности, который завершает процесс формирования компетенций.

Место дисциплины и соответствующий этап формирования компетенций в целостном процессе подготовки по образовательной программе можно определить по матрице компетенций.

Этапы формирования компетенции в процессе освоения дисциплины:

- **начальный** этап – на этом этапе формируются знаниевые и инструментальные основы компетенции, осваиваются основные категории, формируются базовые умения. Студент воспроизводит термины, факты, методы, понятия, принципы и правила; решает учебные задачи по образцу;
- **основной** этап – знания, умения, навыки, обеспечивающие формирование компетенции, значительно возрастают, но еще не достигают итоговых значений. На этом этапе студент осваивает аналитические действия с предметными знаниями по дисциплине, способен самостоятельно решать учебные задачи, внося коррективы в алгоритм действий, осуществляя коррекцию в ходе работы, переносит знания и умения на новые условия;
- **завершающий** этап – на этом этапе студент достигает итоговых показателей по заявленной компетенции, то есть осваивает весь необходимый объем знаний, овладевает всеми умениями и навыками в сфере заявленной компетенции. Он способен использовать эти знания, умения, навыки при решении задач повышенной сложности и в нестандартных условиях.

Этапы формирования компетенций в ходе освоения дисциплины отражаются в тематическом плане (см.п. 4 рабочей программы дисциплины).

1.3. Паспорт фонда оценочных средств по дисциплине

Экзамен

В билете два теоретических вопроса и один практический.

Теоретические вопросы билета:

1. Какова архитектура Apache Spark и чем она отличается от Hadoop?
2. Какие шаги необходимо предпринять для настройки и установки Apache Spark? Как использовать Pandas или PySpark в Jupyter Notebook?
3. Что такое функциональное программирование и как оно может быть применено в Spark?
4. Объясните концепцию RDD и DataFrame в Spark. В чем их отличия и преимущества?
5. Как работает Spark SQL и какие возможности он предоставляет для обработки данных?
6. Что такое пользовательские функции (UDF) в Spark и как их можно использовать?
7. В чем разница между Spark Streaming и Spark Structured Streaming? Когда стоит использовать каждую из них?
8. Что такое GraphX и GraphFrame, и каковы их основные применения?
9. Как осуществляется работа с внешними данными в Apache Spark?

6.2.2. Контрольная работа № 1

Датасет. Данные об активности пользователей (на сайте/где-нибудь еще) в формате .csv (user_id, session_duration, activity_start_date). Данные будут сгенерированы скриптом.

Нужно агрегировать по общему времени, проведенным пользователем за последние полгода (к примеру), и затем найти самых неактивных пользователей (к примеру, если пользователь провел на сайте менее 10 минут за полгода).

6.2.2. Контрольная работа № 2

Датасет. Данные об оценках учеников в формате .csv (student_id, subject, grade(0-100)). Датасет будет сгенерированным по нормальному распределению со смещением.

udf:

```
def classify_grade(avg_grade):  
    if avg_grade >= 85: return "High"  
    elif avg_grade >= 70: return "Medium"  
    else: return "Low"
```

Используя заранее написанную функцию, применить её с помощью df.withColumn(), классифицировать студентов по трем уровням успеваемости ('high', 'medium', 'low')

б) критерии оценивания компетенций (результатов) – правильная работа кода программы, понимание алгоритма метода оптимизации, умение вывести необходимые

в) описание шкалы оценивания:
 Каждая задача оценивается по шкале от 0 до 10 баллов.
 Контрольная работа считается выполненной успешно при суммарной оценке не ниже 18 баллов.

6.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Форма аттестации	Наименование оценочного средства	Баллы
Зачет (100 баллов)	Контрольная работа № 1	30
	Контрольная работа № 2	30
	Ответы на билет	40

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная учебная литература:

1. С. Г. Дьяков — "Анализ данных с помощью Apache Spark" — 2020 — 320 с.
2. А. И. Хватов — "Apache Spark для профессионалов" — 2019 — 400 с.
3. Е. В. Неверов — "Spark и Hadoop. Обработка больших данных" — 2020 — 350 с.
4. А. А. Баранов — "Параллельные вычисления в Apache Spark" — 2022 — 240 с.
5. Билл Чианг — "Learning Spark: Lightning-Fast Data Analytics" — 2020 — 336 с.
6. Элиас Д. и С. Хан — "Apache Spark for Data Science" — 2021 — 300 с.

б) дополнительная учебная литература:

1. С. А. Рубцов — "Spark в действии" — 2021 — 380 с.
2. В. К. Степанов — "Функциональное программирование в Apache Spark" — 2021 — 300 с.
3. Н. С. Ермолаев — "Построение распределенных приложений с Apache Spark" — 2022 — 290 с.
4. Д. Н. Лапшин — "Проектирование приложений на Apache Spark" — 2020 — 310 с.

8. Перечень ресурсов* информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), необходимых для освоения дисциплины

-

9. Методические указания для обучающихся по освоению дисциплины

Вид учебного занятия	Организация деятельности студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.
Практические	Проработка рабочей программы, уделяя особое внимание целям и задачам,

занятия	структуре и содержанию дисциплины. Работа с конспектом лекций, просмотр рекомендуемой литературы. Изучение выбранной предметной области на примерах решения задач семинарских занятий, индивидуальных домашних заданий.
Курсовая работа	Не предусмотрена
Контрольная работа	Ознакомиться с основной и дополнительной литературой, включая справочные издания, зарубежные источники, основополагающие термины. Попрактиковаться в решении аналогичных домашних задач по всем темам контрольных работ.
Лабораторная работа	Не предусмотрена.
Подготовка к зачету	При подготовке к зачету необходимо ориентироваться на конспекты лекций и рекомендуемую литературу.

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Конечными результатами освоения программы дисциплины являются сформированные когнитивные дескрипторы «знать», «уметь», «владеть», расписанные по отдельным компетенциям, которые приведены в п.1.1. Формирование этих дескрипторов происходит в процессе изучения дисциплины по этапам в рамках различного вида учебных занятий и самостоятельной работы.

Выделяются три уровня сформированности компетенций на каждом этапе: пороговый, продвинутый и высокий.

Уровни	Содержательное описание уровня	Основные признаки выделения уровня	БРС, % освоения	ECTS/Пятибалльная шкала для оценки экзамена/зачета
Высокий <i>Все виды компетенций сформированы на высоком уровне в соответствии с целями и задачами дисциплины</i>	Творческая деятельность	<i>Включает низестоящий уровень.</i> Студент демонстрирует свободное обладание компетенциями, способен применить их в нестандартных ситуациях: показывает умение самостоятельно принимать решение, решать проблему/задачу теоретического или прикладного характера на основе изученных методов, приемов, технологий	90-100	A/ Отлично/ Зачтено
Продвинутый <i>Все виды компетенций сформированы на продвинутом уровне в соответствии с целями и задачами дисциплины</i>	Применение знаний и умений в более широких контекстах учебной и профессиональной деятельности, нежели по образцу, большей долей самостоятельности и инициативы	<i>Включает низестоящий уровень.</i> Студент может доказать владение компетенциями: демонстрирует способность собирать, систематизировать, анализировать и грамотно использовать информацию из самостоятельно найденных теоретических источников и иллюстрировать ими теоретические положения или обосновывать практику применения.	85-89	B/ Очень хорошо/ Зачтено
			75-84	C/ Хорошо/ Зачтено
Пороговый <i>Все виды компетенций сформированы на пороговом уровне</i>	Репродуктивная деятельность	Студент демонстрирует владение компетенциями в стандартных ситуациях: излагает в пределах задач курса теоретически и практически контролируемый материал.	65-74	D/Удовлетворительно/ Зачтено
			60-64	E/Посредственно /Зачтено
Ниже порогового	Отсутствие признаков порогового уровня: компетенции не сформированы. Студент не в состоянии продемонстрировать обладание компетенциями в стандартных ситуациях.		0-59	Неудовлетворительно/ Зачтено

Оценивание результатов обучения студентов по дисциплине осуществляется по регламенту текущего контроля и промежуточной аттестации.

Критерии оценивания компетенций на каждом этапе изучения дисциплины для каждого вида оценочного средства и приводятся в п. 4 ФОС. Итоговый уровень сформированности компетенции при изучении дисциплины определяется по таблице. При этом следует понимать, что граница между уровнями для конкретных результатов освоения образовательной программы может смещаться.

Уровень сформированности компетенции	Текущий контроль	Промежуточная аттестация
высокий	высокий	высокий
	<i>продвинутый</i>	<i>высокий</i>
	<i>высокий</i>	<i>продвинутый</i>
продвинутый	<i>пороговый</i>	<i>высокий</i>
	<i>высокий</i>	<i>пороговый</i>
	продвинутый	продвинутый
	<i>продвинутый</i>	<i>пороговый</i>
	<i>пороговый</i>	<i>продвинутый</i>
пороговый	пороговый	пороговый
ниже порогового	пороговый	ниже порогового
	ниже порогового	-

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков или опыта деятельности, характеризующих этапы формирования компетенций.

Рейтинговая оценка знаний является интегральным показателем качества теоретических и практических знаний и навыков студентов по дисциплине и складывается из оценок, полученных в ходе текущего контроля и промежуточной аттестации.

Текущий контроль в семестре проводится с целью обеспечения своевременной обратной связи, для коррекции обучения, активизации самостоятельной работы студентов.

Промежуточная аттестация предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины.

Текущий контроль осуществляется два раза в семестр: контрольная точка № 1 (КТ № 1) и контрольная точка № 2 (КТ № 2).

Результаты текущего контроля и промежуточной аттестации подводятся по шкале балльно-рейтинговой системы.

Вид контроля	Этап рейтинговой системы Оценочное спелство	Балл	
		Минимум	Максимум
Текущий	Контрольная точка № 1	17	28
	Контрольная работа	12	20
	Работа на семинаре	2	3
	Доклад	3	5
	Контрольная точка № 2	18	33
	Контрольная работа	14	25
	Работа на семинаре	4	8
Промежуточный	Зачет	25	40
	Теоретический вопрос	10	20
	Теоретический вопрос	15	20
ИТОГО по дисциплине за 2 семестр		60 (зачтено)	100 (зачтено)

Бонусы и штрафы. По Положению бонус (премиальные баллы) не может превышать 5 баллов. Поощрительные баллы студент получает к своему рейтингу в конце семестра: за активную и регулярную работу на занятиях, за выдающийся доклад, за досрочную сдачу домашнего задания. Штрафы студент получает к своему рейтингу в конце семестра: за несвоевременную сдачу домашнего задания (- 3 балла); за пропуск контрольной работы по неуважительной причине (- 2 балла); за нарушение расписания докладов (-3 балла).

Нормативные сроки проведения текущего контроля:

Этап рейтинговой системы, оценочное средство	Неделя семестра
Контрольная точка № 1	7
Контрольная работа	13
Работа на семинаре	1-14
Контрольная точка № 2	17
Контрольная работа	16
Работа на семинаре	8-16

Контрольная работа проводится и оценивается лектором.

Работа на семинаре, домашнее задание оцениваются преподавателем, ведущим практические занятия.

Доклад оценивается группой, окончательный балл выставляется преподавателем, ведущим практические занятия.

Зачет ставится преподавателем, ведущим практические занятия по итогам ответа на вопросы зачета и набранных баллов в семестре.

4. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков

**ОБНИНСКИЙ ИНСТИТУТ АТОМНОЙ ЭНЕРГЕТИКИ
ИАТЭ НИЯУ МИФИ**

БИЛЕТ № 1

По курсу «**Обработка больших данных с помощью Spark**»

1. Какова архитектура Apache Spark и чем она отличается от Hadoop?
2. Что такое GraphX и GraphFrame, и каковы их основные применения?
3. Датасет. Данные об активности пользователей (на сайте/где-нибудь еще) в формате .csv (user_id, session_duration, activity_start_date). Данные будут сгенерированы скриптом. Нужно агрегировать по общему времени, проведенным пользователем за последние полгода (к примеру), и затем найти самых неактивных пользователей (к примеру, если пользователь провел на сайте менее 10 минут за полгода).

**ОБНИНСКИЙ ИНСТИТУТ АТОМНОЙ ЭНЕРГЕТИКИ
ИАТЭ НИЯУ МИФИ**

БИЛЕТ № 2

По курсу «**Обработка больших данных с помощью Spark**»

1. Что такое функциональное программирование и как оно может быть применено в Spark?
2. Как осуществляется работа с внешними данными в Apache Spark?
3. Датасет. Данные об оценках учеников в формате .csv (student_id, subject, grade(0-100)). Датасет будет сгенерированным по нормальному распределению со смещением.

udf:

```
def classify_grade(avg_grade):  
    if avg_grade >= 85: return "High"  
    elif avg_grade >= 70: return "Medium"  
    else: return "Low"
```

Используя заранее написанную функцию, применить её с помощью df.withColumn(), классифицировать студентов по трем уровням успеваемости ('high', 'medium', 'low')

Описание шкалы оценивания:

Билет содержит 2 вопроса по всем разделам дисциплины. БИЛЕТ оценивается в 40 баллов, по 20 баллов за вопрос.

Оценка	Шкала оценивания	Характеристика	Критерии оценки
Отлично	36-40 баллов (90-100%)	Полный ответ на 2 вопроса	Студент должен: - продемонстрировать глубокое усвоение материала; - исчерпывающе, последовательно, грамотно и логически стройно изложить теоретический материал; - правильно формулировать определения
Хорошо	30-35 баллов (75-89%)	Неполный ответ	Студент должен: - продемонстрировать достаточное знание материала; - продемонстрировать знание основных теоретических понятий; - достаточно последовательно, грамотно и логически стройно излагать материал; - продемонстрировать умение ориентироваться в литературе
Удовлетворительно	24-29 баллов (60-74%)	Частичный ответ	Студент должен: - продемонстрировать общее знание материала; - показать общее владение понятийным аппаратом дисциплины; - уметь строить ответ в соответствии со структурой излагаемого вопроса
Неудовлетворительно	0-23 баллов (0-59%)	Неудовлетворительный ответ, отсутствие ответа	Студент демонстрирует: - незнание значительной части материала; - не владение понятийным аппаратом дисциплины; - существенные ошибки при изложении материала - не способен привести корректный пример ни по одному вопросу

Контрольная работа № 1 оценивается в 20 баллов по 2 балла за каждый вопрос. Контрольная работа № 2 оценивается в 25 баллов по 5 баллов за каждый вопрос.
по первой работе

Оценка	Шкала	Характеристика	Критерии оценки
Отлично	Количество набранных баллов: 20-17	Полный правильный ответ на все вопросы	Дан полный исчерпывающий ответ на вопросы, все понятия проиллюстрированы корректными примерами
Хорошо	Количество набранных баллов: 16-13	Неполный ответ	Дан достаточно полный исчерпывающий ответ на вопросы, имеются незначительные погрешности
Удовлетворительно	Количество набранных баллов: 12-10	Частичный ответ	Ответ на вопросы носит общий характер, отсутствует детализация важных понятий, имеются несущественные ошибки
Неудовлетворительно	Количество набранных баллов: <10	Неудовлетворительный ответ, отсутствие ответа	Ответ неверный, определения не приведены

по второй работе

Оценка	Шкала	Характеристика	Критерии оценки
Отлично	Количество набранных баллов: 25-20	Полный правильный ответ на все вопросы	Дан полный исчерпывающий ответ на вопросы, все понятия проиллюстрированы корректными примерами
Хорошо	Количество набранных баллов: 19-16	Неполный ответ	Дан достаточно полный исчерпывающий ответ на вопросы, имеются незначительные погрешности
Удовлетворительно	Количество набранных баллов: 15-12	Частичный ответ	Ответ на вопросы носит общий характер, отсутствует детализация важных понятий, имеются несущественные ошибки
Неудовлетворительно	Количество набранных баллов: <12	Неудовлетворительный ответ, отсутствие ответа	Ответ неверный, определения не приведены