

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Авторы-составители: **Никитина Елена Юрьевна
Черников Арсений Викторович
Мустакимова Яна Романовна**

Рабочая программа дисциплины
РАЗРАБОТКА ХРАНИЛИЩ BIG DATA
Код УМК 93169

Утверждено
Протокол №6
от «26» июня 2020 г.

Пермь, 2020

1. Наименование дисциплины

Разработка хранилищ Big Data

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в базовую часть Блока « М.1 » образовательной программы по направлениям подготовки (специальностям):

Направление: **01.04.02** Прикладная математика и информатика
направленность Технологии разработки программного обеспечения

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Разработка хранилищ Big Data** у обучающегося должны быть сформированы следующие компетенции:

01.04.02 Прикладная математика и информатика (направленность : Технологии разработки программного обеспечения)

ОПК.1 Способен решать актуальные задачи фундаментальной и прикладной математики

Индикаторы

ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики

4. Объем и содержание дисциплины

Направления подготовки	01.04.02 Прикладная математика и информатика (направленность: Технологии разработки программного обеспечения)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	2,3
Объем дисциплины (з.е.)	4
Объем дисциплины (ак.час.)	144
Контактная работа с преподавателем (ак.час.), в том числе:	48
Проведение лекционных занятий	12
Проведение лабораторных работ, занятий по иностранному языку	36
Самостоятельная работа (ак.час.)	96
Формы текущего контроля	Защищаемое контрольное мероприятие (8)
Формы промежуточной аттестации	Экзамен (3 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

1 триместр

Эволюция систем обработки больших данных

Основные этапы развития систем обработки больших данных, основные типы систем и их назначение, эволюция методов обработки данных.

Распределенная файловая система HDFS

Назначение распределенной файловой системы HDFS, основные принципы устройства HDFS, процедура репликации данных и обеспечение отказоустойчивости.

Технология MapReduce

Архитектура и принципы устройства ApacheZooKeeper, алгоритмы консенсуса, алгоритм PAXOS.

Обеспечение отказоустойчивости средствами ApacheZooKeeper

Введение в самоорганизующиеся карты; методы построения самоорганизующихся карт; оценка и настройка самоорганизующихся карт.

2 триместр

Менеджеры ресурсов YARN и Mesos

Назначение и задачи менеджера инфраструктуры, архитектура и принципы устройства YARN, архитектура и принципы устройства Mesos, централизованный и двухуровневый подходы к планированию.

Пакетная обработка больших данных

Принципы организации пакетной обработки данных, архитектура и принципы устройства ApacheSpark, обработка данных с помощью Spark.

Потоковая обработка больших данных

Принципы организации потоковой обработки данных, архитектура и принципы устройства ApacheKafka и ApacheFlink, обработка потоковых данных с помощью SparkStreaming и Apache Flink.

Интерактивная обработка больших данных

Принципы организации интерактивной обработки данных, архитектуры Lambda и Карра, интерактивная обработка данных с помощью SparkSQL.

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторские занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Valentina Janev. Knowledge Graphs and Big Data Processing / Valentina Janev, Damien Graux, Hajira Jabeen, Emanuel Sallinger // — 2020. — 209 p. — ISBN978-3-030-53199-7. [Электронный ресурс]. <https://link.springer.com/book/10.1007/978-3-030-53199-7>
2. Воронова Л. И. Big Data. Методы и средства анализа: Учебное пособие / Воронова Л. И. - Москва: Московский технический университет связи и информатики, 2016. - 33. <http://www.iprbookshop.ru/61463.html>

Дополнительная:

1. Petrophysics and logging big data. The 11 th UPC International Symposium on New Well Logging Techniques (24-27 september 2019). - Qingdao, 2019. - 187.

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<http://infolab.stanford.edu/~ullman/mmds/book.pdf> Leskovec J., Rajaraman A., Ullman J. D. Mining of massive datasets. – Cambridge University Press, 2014.

<http://barbie.uta.edu/~jli/Resources/MapReduce&Hadoop/MapReduce%20Design%20Patterns.pdf> Miner D., Shook A. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems. – " O'Reilly Media, Inc.", 2013.

<http://liuchengxu.org/books/src/Spark/Learning-Spark-Lightning-Fast-Data-Analysis.pdf> Karau H. et al. Learning spark: lightning-fast big data analysis. – " O'Reilly Media, Inc.", 2015

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Разработка хранилищ Big Data** предполагает использование следующего программного обеспечения и информационных справочных систем: Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения лабораторных занятий - компьютерный класс. Состав оборудования определен в Паспорте компьютерного класса.

Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения лабораторных занятий - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для групповых (индивидуальных) консультаций - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения текущего контроля - аудитория, оснащенная меловой (и) или маркерной доской.

Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью

подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Разработка хранилищ Big Data**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ОПК.1

Способен решать актуальные задачи фундаментальной и прикладной математики

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики</p>	<p>Знает основные технологии для решения прикладных и фундаментальных задач в профессиональной сфере. Умеет применять информационные технологии для решения задачи фундаментальной и прикладной задач в профессиональной сфере. Владеет навыками решения прикладных и фундаментальных задач.</p>	<p align="center">Неудовлетворител</p> <p>Не знает основные технологии для решения прикладных и фундаментальных задач в профессиональной сфере. Не умеет применять информационные технологии для решения задачи фундаментальной и прикладной задач в профессиональной сфере. Не владеет навыками решения прикладных и фундаментальных задач.</p> <p align="center">Удовлетворительн</p> <p>Знает основные технологии для решения прикладных и фундаментальных задач в профессиональной сфере. Не умеет применять информационные технологии для решения задачи фундаментальной и прикладной задач в профессиональной сфере. Не владеет навыками решения прикладных и фундаментальных задач.</p> <p align="center">Хорошо</p> <p>Знает основные технологии для решения прикладных и фундаментальных задач в профессиональной сфере. Умеет применять информационные технологии для решения задачи фундаментальной и прикладной задач в профессиональной сфере. Не владеет навыками решения прикладных и фундаментальных задач.</p> <p align="center">Отлично</p> <p>Знает основные технологии для решения прикладных и фундаментальных задач в профессиональной сфере. Умеет применять информационные технологии для решения задачи фундаментальной и прикладной задач в профессиональной сфере. Владеет навыками решения прикладных и фундаментальных задач.</p>

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : ИТМО

Вид мероприятия промежуточной аттестации : Не предусмотрено

Максимальное количество баллов : 100

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Эволюция систем обработки больших данных Защищаемое контрольное мероприятие	Задача (задание).
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Распределенная файловая система HDFS Защищаемое контрольное мероприятие	Задача (задание).
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Технология MapReduce Защищаемое контрольное мероприятие	Задача (задание).
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Обеспечение отказоустойчивости средствами ApacheZooKeeper Защищаемое контрольное мероприятие	Задача (задание).

Спецификация мероприятий текущего контроля

Эволюция систем обработки больших данных

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронной форме.	25

Распределенная файловая система HDFS

Продолжительность проведения мероприятия промежуточной аттестации: **4 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронной форме.	25

Технология MapReduce

Продолжительность проведения мероприятия промежуточной аттестации: **3 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25

Обеспечение отказоустойчивости средствами ApacheZooKeeper

Продолжительность проведения мероприятия промежуточной аттестации: **3 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 44 до 60

«неудовлетворительно» / «незачтено» менее 44 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Менеджеры ресурсов YARN и Mesos Защищаемое контрольное мероприятие	Задача (задание).
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Пакетная обработка больших данных Защищаемое контрольное мероприятие	Задача (задание).

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Потоковая обработка больших данных Защищаемое контрольное мероприятие	Задача (задание).
ОПК.1.2 Применяет информационные технологии для решения задачи фундаментальной и/или прикладной математики	Интерактивная обработка больших данных Защищаемое контрольное мероприятие	Задача (задание).

Спецификация мероприятий текущего контроля

Менеджеры ресурсов YARN и Mesos

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25

Пакетная обработка больших данных

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25

Потоковая обработка больших данных

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25

Интерактивная обработка больших данных

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **11**

Показатели оценивания	Баллы
Отчет в электронном виде.	25