

Компонент ОПОП 09.04.01 Информатика и вычислительная техника
Направленность (профиль) Компьютерный анализ и интерпретация данных.
Data Science.

Б1.О.13
шифр дисциплины

РАБОЧАЯ ПРОГРАММА

Дисциплины
(модуля)

Технология больших данных

Разработчик:

Золотов О.В.

ФИО

доцент

должность

канд. физ.-мат. наук

ученая степень,
звание

Утверждено на заседании кафедры
Цифровых технологий, математики
и экономики

наименование кафедры

протокол № 8 от 23.05.2023

И.о. заведующего кафедрой
ЦТМиЭ

_____ Мотина Т.Н.
подпись _____ ФИО

Мурманск
2023

Пояснительная записка

Объем дисциплины 4 з.е.

1. Результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций, установленными образовательной программой

Компетенции	Индикаторы достижения компетенций	Результаты обучения по дисциплине (модулю)
<p>ОПК-6. Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p>	<p>ИД-1опк-6 Понимает принципы работы программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-2опк-6 Анализирует существующие подходы к разработке компонент программно-аппаратных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-3опк-6 Использует инструментальные средства разработки компонент программно-аппаратных комплексов</p>	<p>Знать:</p> <ul style="list-style-type: none"> - принципы работы программно-аппаратных комплексов обработки информации и автоматизированного проектирования, ориентированных на технологии работы с большими данными; - специфику зарубежных комплексов обработки информации и автоматизированного проектирования, применяющих технологии работы с большими данными. <p>Уметь:</p> <ul style="list-style-type: none"> - анализировать существующие подходы к разработке компонент программно-аппаратных комплексов обработки информации и автоматизированного проектирования, ориентированных на технологии работы с большими данными; - анализировать цифровые потребности отечественных предприятий в области технологии работы с большими данными.
<p>ОПК-7 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий</p>	<p>ИД-1опк-7 Понимает специфику зарубежных комплексов обработки информации и автоматизированного проектирования</p> <p>ИД-2опк-7 Анализирует цифровые потребности отечественных предприятий</p> <p>ИД-3опк-7 Адаптирует комплексы обработки информации и автоматизированного проектирования к нуждам конкретного предприятия</p>	<p>Владеть:</p> <ul style="list-style-type: none"> - инструментальными средствами разработки компонент программно-аппаратных комплексов, ориентированных на применение технологий работы с большими данными; - навыком адаптации комплексов обработки информации в соответствии с нуждами в сфере технологии работы с большими данными.

2. Содержание дисциплины (модуля)

Тема 1. *Введение в понятие больших данных (BigData).*

Тема 2. *Организация распределенных вычислений на базе Apache Hadoop (I/II)*

Тема 3. *Организация распределенных вычислений на базе Apache Hadoop (II/II)*

Тема 4. *Apache Pig как средство доступа к данным Hadoop*

Тема 5. *Apache Hive и HiveQL как средство доступа к данным Hadoop*

Тема 6. *Нереляционное распределенное хранилище (NoSQL СУБД) Apache HBase*

Тема 7. *Apache Phoenix как средство доступа к данным HBase (I/II)*

Тема 8. *Apache Phoenix как средство доступа к данным HBase (II/II)*

Тема 9. *NoSQL СУБД ClickHouse (I/II)*

Тема 10. *NoSQL СУБД ClickHouse (II/II)*

3. Перечень учебно-методического обеспечения дисциплины (модуля)

- мультимедийные презентационные материалы по дисциплине (модулю) представлены в электронном курсе в ЭИОС МАУ;

- методические указания к выполнению лабораторных работ представлены в электронном курсе в ЭИОС МАУ;

- методические материалы для обучающихся по освоению дисциплины (модуля) представлены на официальном сайте МАУ в разделе «Информация по образовательным программам, в том числе адаптированным».

4. Фонд оценочных средств по дисциплине (модулю)

Является отдельным компонентом образовательной программы, разработан в форме отдельного документа, включает в себя:

- перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины (модуля);

- задания текущего контроля;

- задания промежуточной аттестации;

- задания внутренней оценки качества образования.

5. Перечень основной и дополнительной учебной литературы (печатные издания, электронные учебные издания и (или) ресурсы электронно-библиотечных систем)

Основная литература:

1. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/182452>. — Режим доступа: для авториз. пользователей.

2. Кобзаренко, Д. Н. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес» : учебное пособие / Д. Н. Кобзаренко, А. Г. Мустафаев ; составитель Д. Н. Кобзаренко. — Махачкала : ДГУНХ, 2019. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/246542>. — Режим доступа: для авториз. пользователей.

3. Григорьев, Ю. А. Реляционные базы данных и системы NoSQL / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. - Благовещенск: Амурский государственный университет, 2018. - 424 с. — URL: <https://www.elibrary.ru/item.asp?id=35284192>. — Режим доступа: для авториз. пользователей.

Дополнительная литература:

4. Zolotov, O. On Definition of BigData / O. Zolotov, Yu. Romanovskaya, V. Rzhannikova // EPJ Web Conf. –2019. – V. 224. – URL: <https://doi.org/10.1051/epjconf/201922404011>
5. Apache Hadoop. – Text : electronic // The Apache Software Foundation : [site]. – [S. I.], 2022. – URL: <https://hadoop.apache.org/>
6. Welcome to Apache Pig!. – Text : electronic // The Apache Software Foundation : [site]. – [S. I.], 2022. – URL: <https://pig.apache.org/>
7. Pig Latin Basics // Apache Hadoop : [site]. – 2017. – URL: <https://pig.apache.org/docs/r0.17.0/basic.html>
8. Apache Hive. – Text : electronic // The Apache Software Foundation : [site]. – [S. I.], 2022. – URL: <https://hive.apache.org/>
9. GettingStarted - Apache Hive - Apache Software Foundation / Created by Confluence Administrator, Edited by Karen Coppage. – 2020. – URL: <https://cwiki.apache.org/confluence/display/Hive/GettingStarted>
10. Tutorial - Apache Hive - Apache Software Foundation / Created by Confluence Administrator, Edited by user 77895. – 2021. – URL: <https://cwiki.apache.org/confluence/display/Hive/Tutorial>
11. Mrjob // Mrjob v0.7.4 documentation : [site]. – 2009-2018. – URL: <https://mrjob.readthedocs.io/en/latest/>
12. Welcome to Apache HBase™. – Text : electronic // The Apache Software Foundation : [site]. – [S. I.], 2022. – URL: <https://hbase.apache.org/>
13. Apache HBase™ Reference Guide. – 2023. – URL: <https://hbase.apache.org/book.html>
14. Apache Phoenix : OLTP and operational analytics for Apache Hadoop : [site]. – URL: <https://phoenix.apache.org>
15. Installation // Apache Phoenix : data platform : [site]. – URL: <https://phoenix.apache.org/installation.html>
16. F.A.Q. // Apache Phoenix : data platform : [site]. – URL: <https://phoenix.apache.org/faq.html>
17. ClickHouse – высокопроизводительная аналитическая СУБД с открытым исходным кодом. – Текст : электронный // Технологии Яндекса : [сайт]. – [S. I.], 2022. – URL: <https://yandex.ru/dev/clickhouse/>
18. Advanced Tutorial // ClickHouse : database : [site]. – URL: <https://clickhouse.com/docs/en/tutorial>
19. ClickHouse не тормозит : Telegram-канал / Alexey Milovidov, Nikolai Kochetov, Alexander Tokmakov [и др.]. – 2016- . – URL: https://t.me/clickhouse_ru

6. Профессиональные базы данных и информационные справочные системы

1. eLIBRARY.RU : научная электронная библиотека. – Москва, 2000- . – URL: <https://www.elibrary.ru/>. – Режим доступа: научная электронная библиотека eLIBRARY.RU, после регистрации.
2. Kaggle : AI & ML community : site. – 2024. – URL: <https://www.kaggle.com/>
3. КонсультантПлюс : справочная правовая система : сайт. – Москва, 1997- . – URL: <https://www.consultant.ru>

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

- 1) Средство запуска виртуальных машин Virtual Box или Hyper-V.
- 2) Свободное программное обеспечение Rocky Linux v. 9.4+.
- 3) Свободное программное обеспечение Apache Hadoop.
- 4) Свободное программное обеспечение Apache Pig.
- 5) Свободное программное обеспечение Apache Hive.

- 6) Свободное программное обеспечение Apache HBase.
- 7) Свободное программное обеспечение Apache Phoenix.
- 8) Свободное программное обеспечение NoSQL СУБД ClickHouse.
- 9) Свободное программное обеспечение gcc (GNU compiler collection) v. 12+.
- 10) Свободное программное обеспечение Python 3.9+, Python-библиотека MRJob.

8. Обеспечение освоения дисциплины лиц с инвалидностью и ОВЗ

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечиваются печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

9. Материально-техническое обеспечение дисциплины (модуля) представлено в приложении к ОПОП «Материально-технические условия реализации образовательной программы» и включает:

- учебные аудитории для проведения учебных занятий, предусмотренных программой магистратуры, оснащенные оборудованием и техническими средствами обучения;

- помещения для самостоятельной работы обучающихся, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде МАУ.

Допускается замена оборудования его виртуальными аналогами.

Перечень лабораторных работ по формам обучения

№ п/п	Темы лабораторных работ
1	2
	Очная форма
1	Изучение возможностей Apache Hadoop (I)
2	Изучение возможностей Apache Hadoop (II)
3	Изучение возможностей Apache Pig (I)
4	Изучение возможностей Apache Pig (II)
5	Изучение возможностей Apache Hive и языка запросов HiveQL (I)
6	Изучение возможностей Apache Hive и языка запросов HiveQL (II)
7	Изучение возможностей NoSQL хранилища Apache HBase (I)
8	Изучение возможностей NoSQL хранилища Apache HBase (II)
9	Изучение возможностей Apache Phoenix (I)
10	Изучение возможностей Apache Phoenix (II)
11	Изучение возможностей NoSQL СУБД ClickHouse (I)
12	Изучение возможностей NoSQL СУБД ClickHouse (II)
	Заочная форма
1	Заочная форма обучения не реализуется

Перечень практических занятий по формам обучения

№ п/п	Темы практических занятий
1	2
	Очная форма
1	Практические занятия не предусмотрены учебным планом
	Заочная форма
1	Заочная форма обучения не реализуется

Перечень примерных тем курсовой работы /курсового проекта

№ п/п	Темы курсовой работы /проекта
1	2
1	Курсовая работа / проект не предусмотрены учебным планом