

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт экономики и менеджмента

УТВЕРЖДАЮ:

Директор



Е. В. Нехода

«23» 05 2024 г.

Рабочая программа дисциплины

Большие данные и аналитика

по направлению подготовки

38.04.01 Экономика

Направленность (профиль) подготовки :

Анализ данных в экономике

Форма обучения

Очная

Квалификация

Магистр

Год приема

2024

СОГЛАСОВАНО:

Руководитель ОП

 Н.А. Скрьльникова

Председатель УМК

 М.В. Герман

Томск – 2024

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:
ПК-2 Способен разрабатывать стратегии управления изменениями в организации.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК 2.1 Определяет цели и задачи стратегических изменений в организации

2. Задачи освоения дисциплины

- Освоить понятийный аппарат и инструментарий хранения и анализа больших данных.
- Научиться применять понятийный аппарат больших данных для решения практических задач профессиональной деятельности.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплина (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Второй семестр, зачет

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Премодуль (Бизнес-аналитика)», «Премодуль (Вероятностные и статистические методы в бизнес-аналитике)», «Python и R для анализа данных», «Эконометрика».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 8 ч.

-практические занятия: 20 ч.

в том числе практическая подготовка: 20 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в Большие Данные.

Эволюция управления данными, определение больших данных, технологический стек больших данных, типы больших данных, виртуализация и распределенные вычисления при работе с большими данными.

Тема 2. Операционные базы данных.

Обзор реляционных баз данных, нереляционные базы данных и базы данных пар «ключ-значение», документальные и столбцовые базы данных, графовые и пространственные базы данных.

Тема 3. Основные принципы MapReduce.

Происхождение MapReduce, функция Map, функция Reduce, объединение функций Map и Reduce, оптимизация задач MapReduce.

Тема 4. Hadoop.

HDFS, Hadoop MapReduce, Хранение больших данных в HBase, Интеллектуальный анализ больших данных с помощью Hive, Flume, Pig, Sqoop.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр. Вклад результатов текущего контроля в итоговой оценке по дисциплине составляет – 50 баллов (50%).

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет во втором семестре проводится в комбинированной форме по билетам. Билет содержит теоретический вопрос и задачу, которую необходимо решить соответствующими программными продуктами. Продолжительность зачета 1,5 часа.

Ответ на теоретический вопрос, проверяющий ПК-2.1, дается в развернутой форме.

Результаты зачета определяются в соответствии с балльно-рейтинговой системой – максимум 50 баллов за зачет (50%):

Критерии выставления баллов за зачет:

Баллы	Характеристика
50 баллов	Дан полный и развернутый ответ на вопрос. Задача решена верно и дана обоснованная интерпретация полученных результатов.
20 баллов	Дан неполный или фрагментарный ответ на вопрос. Задача решена верно, но интерпретация полученных результатов не убедительна.
0 баллов	Не дан ответ на вопрос. Задача решена неверно

Итоговая оценка по дисциплине складывается из результатов текущего контроля (50%) и результатов промежуточной аттестации (50%) и составляет максимум 100 баллов.

Механизм перевода результатов балльно-рейтинговой системы в двухбалльную шкалу:

Баллы	Итоговая оценка
70-100 баллов	«Зачтено»
Менее 70 баллов	«Не зачтено»

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=16604>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

Примерный перечень теоретических вопросов, проверяющих ПК-2.1:

1. Эволюция управления данными.
2. Big Data.
3. Стек больших данных.
4. Реляционная база данных.
5. Баз данных NoSQL.
6. Основные принципы MapReduce.
7. Что такое Hadoop и в чем его преимущества?
8. Распределенная файловая система Hadoop.
9. Hadoop MapReduce.
10. База данных HBase.
11. Интеллектуальный анализ больших данных с помощью Hive.
12. Pig и Pig Latin.
13. Flume - инструмент для управления потоками данных.

Примерный перечень практических задач, проверяющих ПК-2.1:

Задача №1. В MySQL создайте базу данных mybase2 согласно приведенной схеме. Загрузите в созданные таблицы данные из соответствующих файлов. Создайте запрос на выборку информации о студентах указанной группы конкретного преподавателя, не сделавших ни одной задачи в каком-либо упражнении.

Задача №2. Flume Channel Selectors. Создайте селектор, с помощью которого источник дублирует события в два канала, один из которых направляет события в папку /home/b_a/flume/1, а второй - в папку /home/b_a/flume/2.

Задача №3. В MySQL есть база данных retail_db.

1. Предоставьте полный доступ к этой базе всем пользователям (чтобы не указывать пользователя и пароль):
GRANT ALL PRIVILEGES ON retail_db.* TO '@'localhost';
GRANT ALL PRIVILEGES ON retail_db.* TO '%@'localhost'.
2. Создайте в Hive базу данных sqoop_retail.
3. С помощью Sqoop импортируйте в нее клиентов (таблица customers) только из штата NY.
4. Проверьте наличие таблицы и определите сколько в ней клиентов из разных городов штата NY (по убыванию).

в) Методические указания по организации самостоятельной работы студентов:

Самостоятельная работа магистрантов включает в себя:

- самостоятельную подготовку к занятиям по заявленным темам курса в соответствии с содержанием дисциплины и литературой. Контроль выполнения производится на занятиях в блиц-опросах;
- самостоятельную работу в аудитории при ответах на вопросы, решении задач. Контроль выполнения осуществляется сразу же при оценке полученных результатов.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Макшанов А. В. Большие данные. Big Data : учебник для вузов / Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н.. - Санкт-Петербург : Лань, 2022. - 188 с.. URL: <https://e.lanbook.com/book/198599>.

– Просто о больших данных : пер. с англ. / Джудит Гурвиц, Алан Ньюджент, Ферн Халпер, Марсия Кауфман ; Сбербанк. - Москва : Эксмо, 2015. - 393, [2] с.

– Ын А. Теоретический минимум по Big Data : все, что нужно знать о больших данных / Анналин Ын, Кеннет Су. - Санкт-Петербург [и др.] : Питер, 2019. - 205 с.

б) дополнительная литература:

– Wadkar S. Pro Apache Hadoop / by Sameer Wadkar, Madhu Siddalingaiah. // Springer eBooks. URL: <http://dx.doi.org/10.1007/978-1-4302-4864-4>

– Дадян Э. Г. Методы, модели, средства хранения и обработки данных : учебник : [для бакалавров и магистрантов всех специальностей, аспирантов] / Э. Г. Дадян, Ю. А. Зеленков ; Финансовый ун-т при Правит. Рос. Фед.. - Москва : Вузовский учебник [и др.], 2017. - 167, [1] с.

– Замятин А. В. Интеллектуальный анализ данных : учебное пособие : [для студентов университетов и вузов] / А. В. Замятин ; Нац. исслед. Том. гос. ун-т. - Томск : Издательский Дом Томского государственного университета, 2020. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000722107>

в) ресурсы сети Интернет:

– Apache Hadoop - <https://hadoop.apache.org>

– Apache Flume - <https://flume.apache.org>

– Apache Hadoop Ecosystem - <https://www.cloudera.com/products/open-source/apache-hadoop.html>

– Официальный сайт Федеральной службы государственной статистики РФ - www.gsk.ru

– Официальный сайт Всемирного банка - www.worldbank.org

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

– Oracle VirtualBox;

– Cloudera CDH;

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юрайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения практических занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Авдеенко Сергей Николаевич, кандидат технических наук, доцент кафедры информационных технологий и бизнес-аналитики Института экономики и менеджмента Национального исследовательского Томского государственного университета