



**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ИНСТИТУТ МЕЖДУНАРОДНЫХ ЭКОНОМИЧЕСКИХ СВЯЗЕЙ»**

**INSTITUTE OF INTERNATIONAL ECONOMIC RELATIONS**

Принята на заседании  
Учёного совета ИМЭС  
(протокол от 28 марта 2024 г. № 8)

**УТВЕРЖДАЮ**  
Ректор ИМЭС Ю.И. Богомолова  
28 марта 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
ВВЕДЕНИЕ В ТЕХНОЛОГИИ BIG DATA**

по направлению подготовки  
38.03.05 Бизнес-информатика

Направленность (профиль)  
«Информационные технологии в бизнесе»

*Приложение 4*  
*к основной профессиональной образовательной программе*  
*по направлению подготовки 38.03.05 Бизнес-информатика,*  
*направленность (профиль) «Информационные технологии в бизнесе»*

**Программу составил:** Казаков М.Ю.

Рабочая программа дисциплины «Введение в технологии Big Data» входит в состав основной профессиональной образовательной программы высшего образования по направлению подготовки 38.03.05 Бизнес-информатика, направленность (профиль) «Информационные технологии в бизнесе» и предназначена для обучающихся очной и очно-заочной форм обучения.

## СОДЕРЖАНИЕ

1. Цель и задачи дисциплины .....	4
2. Место дисциплины в структуре основной профессиональной образовательной программы высшего образования .....	4
3. Объем дисциплины в зачетных единицах и академических часах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся	4
4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения основной профессиональной образовательной программы .....	5
5. Содержание дисциплины .....	6
6. Структура дисциплины по темам с указанием отведенного на них количества академических часов и видов учебных занятий .....	7
7. Перечень учебно-методического обеспечения для самостоятельной работы и текущего контроля обучающихся по дисциплине .....	8
8. Перечень вопросов и типовые задания для подготовки к промежуточной аттестации .....	9
9. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины .....	10
10. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины и информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости) .....	10
11. Методические рекомендации по организации изучения дисциплины .....	11
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине .....	13

## 1. Цель и задачи дисциплины

**Целью** дисциплины «Введение в технологии Big Data» является получение теоретических знаний и практических навыков, необходимых для понимания основ и принципов технологии Big Data, применения Big Data на практике в различных областях, включая хранение, обработку и анализ маркетинговых и финансовых данных.

### Задачи изучения дисциплины:

- приобретение понимания основных принципов организации хранения Big Data;
- приобретение понимания современных технологий обработки и анализа Big Data;
- изучение основных инструментов, реализующих технологии Big Data на практике.

## 2. Место дисциплины в структуре основной профессиональной образовательной программы высшего образования

Дисциплина «Введение в технологии Big Data» входит в часть учебного плана, формируемую участниками образовательных отношений, по направлению подготовки 38.03.05 Бизнес-информатика, направленность (профиль) «Информационные технологии в бизнесе».

## 3. Объем дисциплины в зачетных единицах и академических часах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоёмкость дисциплины составляет 3 зачётные единицы, всего – 108 часа.

Вид учебной работы	Всего часов	
	очная форма обучения	очно-заочная форма обучения
<b>Контактная работа с преподавателем (всего)</b>	<b>28</b>	<b>8</b>
В том числе:		
Занятия лекционного типа	14	6
Занятия семинарского типа (семинары)	14	2
<b>Самостоятельная работа (всего)</b>	<b>80</b>	<b>100</b>
Форма контроля	Зачет с оценкой	
<b>Общая трудоёмкость дисциплины</b>	<b>108</b>	

**4. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения основной профессиональной образовательной программы**

Код и наименование компетенции(ий) выпускника	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
<p><b>ПК-4</b> Способен осуществлять проектирование и дизайн информационных систем</p>	<p><b>ИПК-4.1</b> Ориентируется в современных структурных языках программирования</p> <p><b>ИПК-4.2</b> Осуществляет проектирование и дизайн информационных систем с использованием современных объектно-ориентированных языков программирования</p>	<p><b>Знать:</b> классификацию современных структурных языков программирования в части их использования при работе с Big Data.</p> <p><b>Уметь:</b> анализировать информационные системы, использующие Big Data и их дизайн.</p> <p><b>Владеть:</b> навыками проектирования информационных систем, использующих Big Data и создания их дизайна.</p>

## 5. Содержание дисциплины

Наименование тем (разделов)	Содержание тем (разделов)
Тема 1. Введение в Big Data	История возникновения технологий Big Data. Распределённое хранение и обработка данных. Горизонтальное масштабирование. Системы управления данными: Apache Hadoop, Spark. Экосистема Hadoop.
Тема 2. Apache Hadoop	Основные компоненты: Hadoop Distributed File System (HDFS), MapReduce, YARN. HDFS: ноды, работа с файлами, репликация данных и устойчивость к сбоям. MapReduce: фазы обработки данных, создание и запуск задач. YARN: мониторинг и управление кластером, конфигурирование ресурсов и настройка задач в кластере Hadoop.
Тема 3. Apache Spark	Причины популярности Apache Spark. Архитектура. Работа с RDD (Resilient Distributed Datasets). Структурированные данные и DataFrame. Создание DataFrame и операции с ним. SQL в Spark. Обработка потоковых данных DStream. Машинное обучение с использованием MLlib.

**6. Структура дисциплины по темам с указанием отведенного на них количества академических часов и видов учебных занятий**

**Очная форма обучения**

№ п/п	Наименование тем (разделов) дисциплины	Контактная работа, час.		Самостоятельная работа, час.	Всего, час.
		Занятия лекционного типа	Практические занятия		
1.	Введение в Big Data	1	1	2	3
2.	Apache Hadoop	6	6	40	52
3.	Apache Spark	7	7	38	53
<b>Итого:</b>		<b>14</b>	<b>14</b>	<b>80</b>	<b>108</b>

**Очно-заочная форма обучения**

№ п/п	Наименование тем (разделов) дисциплины	Контактная работа, час.		Самостоятельная работа, час.	Всего, час.
		Занятия лекционного типа	Практические занятия		
1.	Введение в Big Data	1	-	2	3
2.	Apache Hadoop	2	1	50	53
3.	Apache Spark	3	1	48	52
<b>Итого:</b>		<b>6</b>	<b>2</b>	<b>100</b>	<b>108</b>

## 7. Перечень учебно-методического обеспечения для самостоятельной работы и текущего контроля обучающихся по дисциплине

Самостоятельная работа является одним из основных видов учебной деятельности, составной частью учебного процесса и имеет своей целью: глубокое усвоение материала дисциплины, совершенствование и закрепление навыков самостоятельной работы с литературой, рекомендованной преподавателем, умение найти нужный материал и самостоятельно его использовать, воспитание высокой творческой активности, инициативы, привычки к постоянному совершенствованию своих знаний, к целеустремленному научному поиску.

Контроль самостоятельной работы, является важной составляющей текущего контроля успеваемости, осуществляется преподавателем во время лекционных и практических (семинарских) занятий и обеспечивает оценивание хода освоения изучаемой дисциплины.

### Вопросы для самостоятельного изучения

1. Основные методы сбора данных.
2. Методы хранения данных.
3. Масштабирование инфраструктуры для работы с Big Data.
4. Подходы к параллельной обработке данных.
5. Алгоритмы машинного обучения для анализа Big Data.
6. Визуализация и интерпретация результатов анализа.
7. Big Data в медицине и здравоохранении.
8. Big Data в финансовой сфере.
9. Big Data в маркетинге и рекламе.
10. Big Data в государственном управлении и общественной политике.
11. Проблемы и вызовы, возникающие при обработке и анализе Big Data.
12. Тенденции и будущие направления развития Big Data.
13. Обеспечение безопасности и конфиденциальности данных.
14. Прозрачность и справедливость в обработке данных.
15. Вопросы конфиденциальности и анонимности.
16. Этика сбора и использования Big Data.

### Распределение самостоятельной работы

Виды, формы и объемы самостоятельной работы студентов при изучении данной дисциплины определяются ее содержанием и отражены в следующей таблице:

№ п/п	Наименование тем (разделов) дисциплины	Вид самостоятельной работы	Объем самостоятельной работы	
			очная форма обучения	очно-заочная форма обучения
1.	Введение в Big Data	подготовка к аудиторным занятиям	2	2
2.	Apache Hadoop	подготовка к аудиторным занятиям	40	50
3.	Apache Spark	подготовка к аудиторным занятиям	38	48
<b>ИТОГО:</b>			<b>80</b>	<b>100</b>



## **8. Перечень вопросов и типовые задания для подготовки к промежуточной аттестации**

### **8.1 Перечень вопросов для подготовки к зачету с оценкой**

1. Определение Big Data и его особенности.
2. Проблемы, возникающие при обработке и анализе больших объемов данных.
3. Технологии и инструменты, использующиеся для обработки и анализа Big Data.
4. Apache Hadoop и его основные компоненты.
5. Экосистема Apache Hadoop.
6. Основные компоненты Hadoop Distributed File System (HDFS).
7. Задачи, решаемые Apache Hadoop MapReduce.
8. Процесс обработки данных в Hadoop.
9. YARN: мониторинг и управления кластером Hadoop.
10. Особенности Apache Spark по сравнению с Apache Hadoop.
11. Компоненты экосистемы Apache Spark.
12. Spark RDD.
13. Spark DataFrame.
14. Обработка потоковых данных DStream.
15. Apache Spark для машинного обучения и анализа данных.

### **8.2. Типовые задания для оценки знаний**

1. Какой из компонентов Hadoop отвечает за обработку больших объемов данных:
  - а) Hadoop Distributed File System (HDFS);
  - б) Apache Spark;
  - в) MapReduce;
  - г) Apache Hive.
2. Какое значение имеет термин ETL в контексте Big Data:
  - а) Extract, Transform, Load;
  - б) Efficient Testing and Learning;
  - в) Enhanced Technical Logistics;
  - г) Elastic Task List.
3. Какой из языков программирования чаще всего используется для написания MapReduce-задач в Hadoop:
  - а) Java;
  - б) Python;
  - в) C++;
  - г) Scala.

### **8.3. Типовые задания для оценки умений**

**Задание 1.** Напишите команду для отправки MapReduce-задачи на выполнение в Hadoop.

**Задание 2.** Укажите какой конфигурационный файл необходимо использовать для определения параметров выполнения задачи в Hadoop.

**Задание 3.** Укажите параметр конфигурации Hadoop, указывающий количество Reducer-задач в задаче MapReduce.

## 8.4. Типовые задания для оценки навыков

**Задание 1.** Определите оптимальное количество Reducer-задач для 100 уникальных ключей, полученных в результате фазы Map.

**Задание 2.** По умолчанию, размер блока HDFS составляет 128 Мбайт. Опишите подходы к выбору оптимального размера блока для условий конкретной задачи.

**Задание 3.** Сформулируйте правила выбора количества параллельных задач, которые могут выполняться одновременно в каждом исполнителе (executor) Apache Spark.

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

### 9.1. Основная литература

1. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2024. — 121 с. — (Высшее образование). — ISBN 978-5-534-09837-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/539330>

### 9.2. Дополнительная литература

1. Стружкин, Н. П. Базы данных: проектирование : учебник для вузов / Н. П. Стружкин, В. В. Годин. — Москва : Издательство Юрайт, 2024. — 477 с. — (Высшее образование). — ISBN 978-5-534-00229-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536006>

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2024. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536007>

## 10. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины и информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

1. <https://urait.ru> - ЭБС «Образовательная платформа Юрайт»
2. <https://coderwall.com/> - Coderwall «Сообщество программистов»

### Лицензионное программное обеспечение:

- Windows (зарубежное, возмездное);
- MS Office (зарубежное, возмездное);
- Adobe Acrobat Reader (зарубежное, свободно распространяемое);
- КонсультантПлюс: «КонсультантПлюс: Студент» (российское, свободно распространяемое);
- 7-zip – архиватор (зарубежное, свободно распространяемое);
- Comodo Internet Security (зарубежное, свободно распространяемое);

- Apache NetBeans – свободная интегрированная среда разработки приложений (IDE) на языках программирования Java, PHP, JavaScript, C, C++, Ада и ряда других (зарубежное, свободно распространяемое).

## **11. Методические рекомендации по организации изучения дисциплины**

Организация образовательного процесса регламентируется учебным планом и расписанием учебных занятий. Язык обучения (преподавания) — русский.

При формировании своей индивидуальной образовательной траектории обучающийся имеет право на перезачёт соответствующих дисциплин и профессиональных модулей, освоенных в процессе предшествующего обучения, который освобождает обучающегося от необходимости их повторного освоения.

### **Образовательные технологии**

Учебный процесс при преподавании курса основывается на использовании традиционных, инновационных и информационных образовательных технологий. Традиционные образовательные технологии представлены занятиями семинарского и лекционного типа. Инновационные образовательные технологии используются в виде широкого применения активных и интерактивных форм проведения занятий. Информационные образовательные технологии реализуются путем активизации самостоятельной работы студентов в информационной образовательной среде.

### **Занятия лекционного типа**

Лекционный курс предполагает систематизированное изложение основных вопросов учебного плана.

На первой лекции лектор обязан предупредить студентов, применительно к какому базовому учебнику (учебникам, учебным пособиям) будет прочитан курс.

Лекционный курс должен давать наибольший объем информации и обеспечивать более глубокое понимание учебных вопросов при значительно меньшей затрате времени, чем это требуется большинству студентов на самостоятельное изучение материала.

### **Занятия семинарского типа**

Семинарские (практические занятия) представляют собой детализацию лекционного теоретического материала, проводятся в целях закрепления курса и охватывают все основные разделы.

Основной формой проведения семинаров и практических занятий является обсуждение наиболее проблемных и сложных вопросов по отдельным темам, а также решение задач и разбор примеров и ситуаций в аудиторных условиях. В обязанности преподавателя входят: оказание методической помощи и консультирование студентов по соответствующим темам курса.

Активность на практических занятиях оценивается по следующим критериям:

- ответы на вопросы, предлагаемые преподавателем;
- участие в дискуссиях;
- выполнение проектных и иных заданий;
- ассистирование преподавателю в проведении занятий.

Доклады и оппонирование докладов проверяют степень владения теоретическим материалом, а также корректность и строгость рассуждений.

Оценивание практических заданий входит в накопленную оценку.

### **Самостоятельная работа обучающихся**

Самостоятельная работа студентов – это процесс активного, целенаправленного приобретения студентом новых знаний, умений без непосредственного участия преподавателя, характеризующийся предметной направленностью, эффективным контролем и оценкой результатов деятельности обучающегося.

Цели самостоятельной работы:

- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную и справочную документацию, специальную литературу;
- развитие познавательных способностей, активности студентов, ответственности и организованности;
- формирование самостоятельности мышления, творческой инициативы, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений и академических навыков.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, уровня сложности, конкретной тематики.

Технология организации самостоятельной работы студентов включает использование информационных и материально-технических ресурсов образовательного учреждения.

Перед выполнением обучающимися самостоятельной работы преподаватель может проводить инструктаж по выполнению задания. В инструктаж включается:

- цель и содержание задания;
- сроки выполнения;
- ориентировочный объем работы;
- основные требования к результатам работы и критерии оценки;
- возможные типичные ошибки при выполнении.

Инструктаж проводится преподавателем за счет объема времени, отведенного на изучение дисциплины.

Контроль результатов самостоятельной работы студентов может проходить в письменной, устной или смешанной форме.

Студенты должны подходить к самостоятельной работе как к наиболее важному средству закрепления и развития теоретических знаний, выработке единства взглядов на отдельные вопросы курса, приобретения определенных навыков и использования профессиональной литературы.

Помещения для самостоятельной работы обучающихся должны быть оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

При самостоятельной проработке курса обучающиеся должны:

- просматривать основные определения и факты;
- повторить законспектированный на лекционном занятии материал и дополнить его с учетом рекомендованной по данной теме литературы;
- изучить рекомендованную литературу, составлять тезисы, аннотации и конспекты наиболее важных моментов;
- самостоятельно выполнять задания, аналогичные предлагаемым на занятиях;
- использовать для самопроверки материалы фонда оценочных средств;
- выполнять домашние задания по указанию преподавателя.

### **Рекомендации по обучению инвалидов и лиц с ОВЗ**

Освоение дисциплины инвалидами и лицами с ОВЗ может быть организовано как совместно с другими обучающимися, так и в отдельных группах. Предполагаются специальные условия для получения образования инвалидами и лицами с ОВЗ.

Профессорско-педагогический состав знакомится с психолого-физиологическими особенностями обучающихся инвалидов и лиц с ОВЗ, индивидуальными программами реабилитации инвалидов (при наличии). При необходимости осуществляется

дополнительная поддержка преподавания тьюторами, психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. № АК-44/05вн) в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе. Подбор и разработка учебных материалов производятся с учетом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем.

Медиа материалы также следует использовать и адаптировать с учетом индивидуальных особенностей обучения инвалидов и лиц с ОВЗ.

Освоение дисциплины инвалидами и лицами с ОВЗ осуществляется с использованием средств обучения общего и специального назначения (персонального и коллективного использования). Материально-техническое обеспечение предусматривает приспособление аудиторий к нуждам инвалидов и лиц с ОВЗ.

Форма проведения аттестации для студентов-инвалидов и лиц с ОВЗ устанавливается с учетом индивидуальных психофизических особенностей. Для инвалидов и лиц с ОВЗ предусматривается доступная форма предоставления заданий оценочных средств, а именно:

- в печатной или электронной форме (для лиц с нарушениями опорно-двигательного аппарата);
- в печатной форме или электронной форме с увеличенным шрифтом и контрастностью (для лиц с нарушениями слуха, речи, зрения);
- методом чтения ассистентом задания вслух (для лиц с нарушениями зрения).

Студентам с инвалидностью и лицам с ОВЗ увеличивается время на подготовку ответов на контрольные вопросы. Для таких студентов предусматривается доступная форма предоставления ответов на задания, а именно:

- письменно на бумаге или набором ответов на компьютере (для лиц с нарушениями слуха, речи);
- выбором ответа из возможных вариантов с использованием услуг ассистента (для лиц с нарушениями опорно-двигательного аппарата);
- устно (для лиц с нарушениями зрения, опорно-двигательного аппарата).

При необходимости для обучающихся с инвалидностью и лиц с ОВЗ процедура оценивания результатов обучения может проводиться в несколько этапов.

## **12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

**Учебная аудитория** «Лаборатория информационно-коммуникационных технологий», предназначенная для проведения учебных занятий, предусмотренных настоящей рабочей программой дисциплин, оснащенная оборудованием и техническими средствами обучения, в состав которых входят: комплекты специализированной учебной мебели, доска классная, мультимедийный проектор, экран, принтер, компьютер преподавателя и компьютеры для работы обучающихся с установленным лицензионным программным обеспечением, с выходом в сеть «Интернет» и доступом в электронную информационно-образовательную среду.

**Помещение для самостоятельной работы** обучающихся – аудитория, оснащенная следующим оборудованием и техническими средствами: специализированная мебель для преподавателя и обучающихся, доска учебная, мультимедийный проектор, экран, звуковые колонки, компьютер (ноутбук), персональные компьютеры для работы обучающихся с установленным лицензионным программным обеспечением, с выходом в сеть «Интернет» и доступом в электронную информационно-образовательную среду.