

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Проректор по учебной работе
д.юр.н., доц. Васильева Н.В.



21.06.2024г.

Рабочая программа дисциплины

Б1.У.9. Технологии высокопроизводительной обработки Big Data

Направление подготовки: 09.03.03 Прикладная информатика

Направленность (профиль): Системы искусственного интеллекта

Квалификация выпускника: бакалавр

Форма обучения: очная, заочная

	Очная ФО	Заочная ФО
Курс	4	4
Семестр	42	42
Лекции (час)	14	6
Практические (сем, лаб.) занятия (час)	14	10
Самостоятельная работа, включая подготовку к экзаменам и зачетам (час)	80	92
Курсовая работа (час)		
Всего часов	108	108
Зачет (семестр)	42	42
Экзамен (семестр)		

Иркутск 2024

Программа составлена в соответствии с ФГОС ВО по направлению 09.03.03
Прикладная информатика.

Автор А.В. Родионов

Рабочая программа обсуждена и утверждена на заседании кафедры
математических методов и цифровых технологий

Заведующий кафедрой А.В. Родионов

1. Цели изучения дисциплины

Формирование у студентов знаний и навыков, необходимых для разработки архитектуры, дизайна, информационного обеспечения и прототипов информационных систем на основе технологий высокопроизводительной обработки больших данных (Big Data).

2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Компетенции обучающегося, формируемые в результате освоения дисциплины

Код компетенции по ФГОС ВО	Компетенция
ПК-4	Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Структура компетенции

Компетенция	Формируемые ЗУНы
ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС

3. Место дисциплины (модуля) в структуре образовательной программы

Принадлежность дисциплины - БЛОК 1 ДИСЦИПЛИНЫ (МОДУЛИ): Часть, формируемая участниками образовательных отношений.

Предшествующие дисциплины (освоение которых необходимо для успешного освоения данной): "Информационно-коммуникационные технологии", "Алгоритмы и структуры данных", "Разработка программного обеспечения", "Системы управления данными", "Объектно-ориентированный анализ и проектирование информационных систем", "Разработка и развертывание приложений в облачной среде", "Технологии мобильной и кросс-платформенной разработки"

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

Общая трудоемкость дисциплины составляет 3 зач. ед., 108 часов.

Вид учебной работы	Количество часов (очная ФО)	Количество часов (заочная ФО)
Контактная(аудиторная) работа		
Лекции	14	6
Практические (сем, лаб.) занятия	14	10
Самостоятельная работа, включая подготовку к экзаменам и зачетам	80	92
Всего часов	108	108

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

5.1. Содержание разделов дисциплины

Заочная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Введение в технологии Big Data	42	1	1	20		Тест по теме 1. Установка и настройка Hadoop кластера
2	Архитектура систем Big Data	42	2	3	24		Тест по теме 2
3	Дизайн и информационное обеспечение систем Big Data	42	2	4	24		Работа с NoSQL базами данных. Тест по теме 3
4	Технологии высокопроизводительной обработки данных	42	1	2	24		Обработка данных с использованием Apache Spark. Тест по теме 4
	ИТОГО		6	10	92		

Очная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Введение в технологии Big Data	42	4	4	20		Тест по теме 1. Установка и настройка Hadoop кластера
2	Архитектура систем Big Data	42	4	4	20		Тест по теме 2
3	Дизайн и информационное обеспечение систем Big Data	42	4	4	20		Работа с NoSQL базами данных. Тест по теме 3
4	Технологии высокопроизводительной обработки данных	42	2	2	20		Обработка данных с использованием Apache Spark. Тест по теме 4
	ИТОГО		14	14	80		

5.2. Лекционные занятия, их содержание

№ п/п	Наименование разделов и тем	Содержание
1	Обзор технологий	

№ п/п	Наименование разделов и тем	Содержание
	Big Data: определение, значение, основные характеристики	
2	История и эволюция технологий Big Data.	
3	Принципы построения архитектуры систем обработки Big Data.	
4	Обзор популярных архитектурных подходов: Lambda, Карра.	
5	Принципы и методы проектирования систем Big Data.	
6	Технологии и инструменты для хранения больших данных (NoSQL базы данных, распределенные файловые системы).	
7	Обзор технологий для обработки больших данных: MapReduce, Spark, Storm, Flink.	

5.3. Семинарские, практические, лабораторные занятия, их содержание

№ раздела и темы	Содержание и формы проведения
1	Примеры применения Big Data в различных отраслях.. Лабораторная работа с использованием ПК
1	Сравнительный анализ различных платформ для обработки больших данных.. Лабораторная работа с использованием ПК
2	Компоненты архитектуры: источники данных, сбор данных, хранение, обработка, анализ и визуализация.. Лабораторная работа с использованием ПК
2	Примеры архитектурных решений для различных задач.. Лабораторная работа с использованием ПК
3	Обеспечение качественного и своевременного сбора данных.. Лабораторная работа с использованием ПК
3	Сравнительный анализ различных систем хранения данных.. Лабораторная работа с использованием ПК
4	Основные алгоритмы и модели для анализа больших данных.. Лабораторная работа с использованием ПК

6. Фонд оценочных средств для проведения промежуточной аттестации по дисциплине (полный текст приведен в приложении к рабочей программе)

6.1. Текущий контроль

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п))	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
1	1. Введение в технологии Big Data	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест по теме 1	(процент правильных ответов * 12 баллов), округленных вверх до ближайшего целого числа, если процент правильных ответов меньше 50 – то 0 баллов (12)
2		ПК-4	У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Установка и настройка Hadoop кластера	Ознакомление с основными компонентами Hadoop (HDFS, YARN, MapReduce): 2 балла Установка необходимых зависимостей (Java, SSH): 1 балл Скачивание и распаковка дистрибутива Hadoop: 1 балл Конфигурация файлов (core-site.xml, hdfs-site.xml, mapred-site.xml, yarn-site.xml): 3 балла Форматирование HDFS и запуск NameNode и DataNode: 2 балла Настройка конфигурационных файлов для запуска кластера: 2 балла

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
					<p>Запуск Hadoop демонов (NameNode, DataNode, ResourceManager, NodeManager): 2 балла</p> <p>Проверка состояния кластера через веб-интерфейс: 1 балл</p> <p>Выполнение простого задания MapReduce для проверки работоспособности кластера: 2 балла (16)</p>
3	2. Архитектура систем Big Data	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест по теме 2	(процент правильных ответов * 12 баллов), округленных вверх до ближайшего целого числа, если процент правильных ответов меньше 50 – то 0 баллов (12)
4	3. Дизайн и информационное обеспечение систем Big Data	ПК-4	У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Работа с NoSQL базами данных	<p>Обзор типов NoSQL баз данных и преимуществ/недостатков NoSQL по сравнению с реляционными базами данных: 2 балла</p> <p>Скачивание и установка MongoDB: 2 балла</p> <p>Создание базы данных и коллекций в MongoDB: 2 балла</p> <p>Вставка документов в</p>

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
					коллекции: 2 балла Обновление документов в коллекциях: 2 балла Удаление документов из коллекций: 2 балла Создание и выполнение простого приложения для работы с MongoDB: 2 балла Анализ производительности MongoDB для заданного случая использования: 2 балла Анализ удобства использования MongoDB для заданного случая использования: 2 балла (18)
5		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест по теме 3	(процент правильных ответов * 12 баллов), округленных вверх до ближайшего целого числа, если процент правильных ответов меньше 50 – то 0 баллов (12)
6	4. Технологии высокопроизводительной обработки данных	ПК-4	У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного	Обработка данных с использованием Apache Spark	Введение в Apache Spark и основные концепции (RDD, DataFrame, SparkContext): 2 балла Установка и настройка Apache Spark: 2 балла

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
			обеспечения и прототипа ИС		<p>Настройка окружения для работы с Spark (Java, Scala, Python): 2 балла</p> <p>Загрузка данных в Spark из файловой системы: 2 балла</p> <p>Применение операции трансформации (map, filter) на RDD или DataFrame: 2 балла</p> <p>Применение операции действия (reduce, collect) на RDD или DataFrame: 2 балла</p> <p>Использование Spark SQL для обработки структурированных данных: 2 балла</p> <p>Оптимизация производительности обработки данных (кэширование, партиционирование): 2 балла</p> <p>Реализация аналитического задания с использованием Spark (например, анализ логов веб-сервера, расчет статистик по данным из базы данных): 2 балла</p> <p>(18)</p>
7		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест по теме 4	(процент правильных ответов * 12 баллов), округленных вверх до

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п))	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
					ближайшего целого числа, если процент правильных ответов меньше 50 – то 0 баллов (12)
				Итого	100

6.2. Промежуточный контроль (зачет, экзамен)

Рабочим учебным планом предусмотрен Зачет в семестре 42.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (40 баллов), вид вопроса: Тест/проверка знаний. Критерий: Сумма баллов за правильные ответы. Правильный ответ на каждый вопрос - 2 балла, неправильный ответ - 0 баллов за вопрос..

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Знание: Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

1. В чем заключается основное преимущество NoSQL баз данных по сравнению с реляционными базами данных?
2. Какие методы оптимизации производительности существуют в Apache Spark и как они применяются на практике?
3. Какой тип данных используется в NoSQL базе данных MongoDB?
4. Какой файл конфигурации в Hadoop содержит информацию о HDFS и что в нем указывается?
5. Какую функцию выполняет компонент YARN в системе Hadoop?
6. Назовите основные этапы установки и настройки кластера Hadoop.
7. Объясните, что такое MapReduce и каковы его основные принципы работы.
8. Что такое Big Data и какие основные характеристики определяют его (назовите хотя бы три)?
9. Что такое RDD в Apache Spark и какова его роль в обработке данных?
10. Что такое Spark SQL и как он используется в рамках Apache Spark?

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УМЕНИЙ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на умение. Критерий: Полностью выполненное задание – 30 баллов, допущены незначительные ошибки - (20-29) баллов, существенные ошибки, не искажающие сути - (10-19) баллов, грубые ошибки или ответа нет - (0-9) баллов.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Умение: Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС

Задача № 1. Реализовать систему обработки больших данных с использованием Apache Spark. Система должна загружать данные из источника, проводить необходимые трансформации, выполнять аналитические вычисления и сохранять результаты.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

3-й вопрос билета (30 баллов), вид вопроса: Задание на навыки. Критерий: Полностью выполненное задание – 30 баллов, допущены незначительные ошибки - (20-29) баллов, существенные ошибки, не искажающие сути - (10-19) баллов, грубые ошибки или ответа нет - (0-9) баллов.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Навык: Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС

Задание № 1. Реализовать аналитическую задачу по обработке и анализу больших данных с использованием технологий Apache Spark и NoSQL баз данных. Ваша задача — разработать приложение, которое будет загружать данные, выполнять их обработку и анализ, а также сохранять результаты в NoSQL базу данных.

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования
Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение
высшего образования
**«БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»
(ФГБОУ ВО «БГУ»)**

Направление - 09.03.03 Прикладная
информатика
Профиль - Системы искусственного
интеллекта
Кафедра математических методов и
цифровых технологий
Дисциплина - Технологии
высокопроизводительной обработки Big
Data

БИЛЕТ № 1

1. Тест (40 баллов).
2. Реализовать систему обработки больших данных с использованием Apache Spark. Система должна загружать данные из источника, проводить необходимые трансформации, выполнять аналитические вычисления и сохранять результаты. (30 баллов).
3. Реализовать аналитическую задачу по обработке и анализу больших данных с использованием технологий Apache Spark и NoSQL баз данных. Ваша задача — разработать приложение, которое будет загружать данные, выполнять их обработку и анализ, а также сохранять результаты в NoSQL базу данных. (30 баллов).

Составитель _____ А.В. Родионов

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

а) основная литература:

1. [Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: <http://www.iprbookshop.ru/88752.html> \(дата обращения: 29.01.2020\)](http://www.iprbookshop.ru/88752.html)
2. [Чубукова И.А. Data Mining \[Электронный ресурс\] / И.А. Чубукова. — Электрон. текстовые данные. — М. : Интернет-Университет Информационных Технологий \(ИНТУИТ\), 2016. — 470 с. — 978-5-94774-819-2. — Режим доступа: <http://www.iprbookshop.ru/56315.html>](http://www.iprbookshop.ru/56315.html)

б) дополнительная литература:

1. [Дьяков И.А. Базы данных. Язык SQL \[Электронный ресурс\] : учебное пособие / И.А. Дьяков. — Электрон. текстовые данные. — Тамбов: Тамбовский государственный технический университет, ЭБС АСВ, 2012. — 81 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/64070.html>](http://www.iprbookshop.ru/64070.html)
2. [Кайл Бэнкер - MongoDB в действии URL: <https://cafe-aristokrat.nethouse.ru/static/doc/0000/0000/0165/165988.c2f3acpbax.pdf>](https://cafe-aristokrat.nethouse.ru/static/doc/0000/0000/0165/165988.c2f3acpbax.pdf)

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая профессиональные базы данных и информационно-справочные системы

Для освоения дисциплины обучающемуся необходимы следующие ресурсы информационно-телекоммуникационной сети «Интернет»:

- Сайт Байкальского государственного университета, адрес доступа: <http://bgu.ru/>, доступ круглосуточный неограниченный из любой точки Интернет
- Электронно-библиотечная система ibooks.ru, адрес доступа: <https://ibooks.ru/>. доступ неограниченный

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Изучать дисциплину рекомендуется в соответствии с той последовательностью, которая обозначена в ее содержании. Для успешного освоения курса обучающиеся должны иметь первоначальные знания в области проектирования и разработки информационных систем, баз данных.

На лекциях преподаватель озвучивает тему, знакомит с перечнем литературы по теме, обосновывает место и роль этой темы в данной дисциплине, раскрывает ее практическое значение. В ходе лекций студенту необходимо вести конспект, фиксируя основные понятия и проблемные вопросы.

Практические (семинарские) занятия по своему содержанию связаны с тематикой лекционных занятий. Начинать подготовку к занятию целесообразно с конспекта лекций. Задание на практическое (семинарское) занятие сообщается обучающимся до его проведения. На семинаре преподаватель организует обсуждение этой темы, выступая в качестве организатора, консультанта и эксперта учебно-познавательной деятельности обучающегося.

Изучение дисциплины (модуля) включает самостоятельную работу обучающегося.

Основными видами самостоятельной работы студентов с участием преподавателей являются:

- текущие консультации;
- коллоквиум как форма контроля освоения теоретического содержания дисциплин: (в часы консультаций, предусмотренные учебным планом);
- прием и разбор домашних заданий (в часы практических занятий);
- прием и защита лабораторных работ (во время проведения занятий);
- выполнение курсовых работ в рамках дисциплин (руководство, консультирование и защита курсовых работ в часы, предусмотренные учебным планом) и др.

Основными видами самостоятельной работы студентов без участия преподавателей являются:

- формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);
- самостоятельное изучение отдельных тем или вопросов по учебникам или учебным пособиям;
- написание рефератов, докладов;
- подготовка к семинарам и лабораторным работам;
- выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и др.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

В учебном процессе используется следующее программное обеспечение:

- Python,
- HBase,
- Модули: pymongo, PyMySQL, PyQt, SQL Alchemy языка Python,
- Модули: matplotlib, Mlxtend, PrefixSpan, NumPy, pandas, scikit-learn, SciPy, TensorFlow, языка Python,
- Visual studio,
- MongoDB,

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю):

В учебном процессе используется следующее оборудование:

- Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду вуза,
- Мультимедийный класс,
- Компьютерный класс