

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Забайкальский государственный университет»  
(ФГБОУ ВО «ЗабГУ»)

Энергетический факультет  
Кафедра Прикладной информатики и математики

УТВЕРЖДАЮ:

Декан факультета

Энергетический факультет

Батухтин Андрей  
Геннадьевич

«\_\_\_\_\_» \_\_\_\_\_ 20\_\_\_\_  
г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

Б1.В.ДВ.02.01 Технологии обработки больших данных  
на 108 часа(ов), 3 зачетных(ые) единиц(ы)  
для направления подготовки (специальности) 09.04.01 - Информатика и вычислительная  
техника

составлена в соответствии с ФГОС ВО, утвержденным приказом  
Министерства образования и науки Российской Федерации от  
«\_\_\_\_\_» \_\_\_\_\_ 20\_\_\_\_ г. №\_\_\_\_\_

Профиль – Искусственный интеллект в автоматизированных системах обработки  
информации и управления (для набора 2024)  
Форма обучения: Очная

# 1. Организационно-методический раздел

## 1.1 Цели и задачи дисциплины (модуля)

Цель изучения дисциплины:

формирование у студентов знаний и навыков в области современных технологий обработки больших объёмов данных, включая проектирование, реализацию и оптимизацию распределённых систем хранения и обработки данных. Студенты осваивают методы и инструменты для эффективного сбора, хранения, обработки и анализа больших данных, а также учатся создавать и администрировать высокопроизводительные системы для поддержки принятия решений на основе больших данных.

Задачи изучения дисциплины:

1. Формирование у студентов навыков проектирования и разработки распределённых систем хранения и обработки данных, включая использование современных технологий и архитектурных решений.
2. Освоение студентами методов и инструментов для эффективной работы с большими объемами данных, включая сбор, хранение, обработку и анализ данных.
3. Развитие у студентов компетенции в создании и администрировании высокопроизводительных систем для поддержки принятия решений на основе больших данных.

## 1.2. Место дисциплины (модуля) в структуре ОП

Дисциплина входит в дисциплины по выбору вариативной части блока Б1 «Дисциплины (модули)» образовательной программы магистратуры по направлению подготовки 09.04.01 «Информатика и вычислительная техника». Изучение дисциплины предполагает предварительное освоение следующих дисциплин учебного плана: • Теоретические основы реляционной алгебры; • Надежность и достоверность автоматизированных систем обработки информации и управления; • Постреляционные базы данных. Освоение данной дисциплины необходимо как предшествующее для следующей дисциплины образовательной программы: • Подготовка и защита ВКР. Освоение учебной дисциплины связано с формированием компетенций с учетом матрицы компетенций ОПОП по направлению подготовки 09.04.01 «Информатика и вычислительная техника».

## 1.3. Объем дисциплины (модуля) с указанием трудоемкости всех видов учебной работы

Общая трудоемкость дисциплины (модуля) составляет 3 зачетных(ые) единиц(ы), 108 часов.

Виды занятий	Семестр 4	Всего часов
Общая трудоемкость		108
Аудиторные занятия, в т.ч.	44	44
Лекционные (ЛК)	11	11

Практические (семинарские) (ПЗ, СЗ)	0	0
Лабораторные (ЛР)	33	33
Самостоятельная работа студентов (СРС)	28	28
Форма промежуточной аттестации в семестре	Экзамен	36
Курсовая работа (курсовой проект) (КР, КП)		

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

Планируемые результаты освоения образовательной программы		Планируемые результаты обучения по дисциплине
Код и наименование компетенции	Индикаторы достижения компетенции, формируемые в рамках дисциплины	Дескрипторы: знания, умения, навыки и (или) опыт деятельности
УК-6	УК-6.3. Владеть: - технологиями и навыками управления своей познавательной деятельностью и ее совершенствования на основе самооценки, самоконтроля и принципов самообразования в течение всей жизни, в том числе с использованием здоровьесберегающих подходов и методик	Знать: Технологии управления своей познавательной деятельностью и ее совершенствования на основе самооценки, самоконтроля и принципов самообразования  Уметь: Управлять своей познавательной деятельностью и ее совершенствования на основе самооценки, самоконтроля и принципов самообразования  Владеть: Технологиями и навыками управления своей познавательной деятельностью и ее совершенствования на основе самооценки, самоконтроля и принципов самообразования в течение всей жизни, в том числе с использованием здоровьесберегающих подходов и методик
ПК-4	ПК-4.1. Руководит работами по	Знать: Особенности оценки и

	оценке и выбору моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи.	<p>выбора моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи.</p> <p>Уметь: Выполнять оценку искусственных нейронных сетей и инструментальных средств для решения поставленной задачи.</p> <p>Владеть: Навыками оценки искусственных нейронных сетей и инструментальных средств для решения поставленной задачи.</p>
ПК-4	ПК-4.2. Руководит созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств.	<p>Знать: Особенности руководства созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств.</p> <p>Уметь: Руководить созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств.</p> <p>Владеть: Навыками руководства созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств.</p>
ПК-4	ПК-4.3. Руководит проектами по разработке систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов.	<p>Знать: Особенности руководства проектами по разработке систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов.</p> <p>Уметь: Руководить проектами по разработке систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов.</p> <p>Владеть: Навыками руководства проектами по разработке систем искусственного интеллекта на</p>

		основе моделей глубоких нейронных сетей и нечетких моделей и методов.
--	--	---

### 3. Содержание дисциплины

#### 3.1. Разделы дисциплины и виды занятий

##### 3.1 Структура дисциплины для очной формы обучения

Модуль	Номер раздела	Наименование раздела	Темы раздела	Всего часов	Аудиторные занятия			С Р С
					Л К	П З (С З)	Л Р	
1	1.1	Базовые понятия систем обработки больших данных	Базовые понятия Big Data. MapReduce. Средства потоковой обработки данных.	38	7	0	17	14
	1.2	Nosql- и sql-системы баз данных, использующие шардирование	Базы данных NoSQL SQL СУБД для работы с большими данными	34	4	0	16	14
Итого				72	11	0	33	28

#### 3.2. Содержание разделов дисциплины

##### 3.2.1. Лекционные занятия, содержание и объем в часах

Модуль	Номер раздела	Тема	Содержание	Трудоемкость (в часах)
1	1.1	Базовые понятия Big Data	Роль дисциплины «Технологии обработки больших данных». Понятие Big Data. Задачи Big Data. Возникновение технологий Big Data. MPI. Распределённые файловые системы для хранения больших объемов данных. NoSQL Data Bases. Hadoop. История разработки Hadoop. Принципы работы Hadoop.	3

			<p>Современное состояние Hadoop. Экосистема Hadoop. Инструменты Big Data - Платформы на основе дистрибутивов Hadoop. Hortonworks Data Platform. Cloudera Enterprise. MapR. Pivotal. Инструменты Big Data - Интегрированные платформы. Возникновение и функциональность интегрированных платформ Big Data. GridGain In-Memory Data Fabric. Data Lake – основной способ организации больших данных. Архитектурные решения с использованием инструментов Big Data.</p> <p>Использование инструментов Big Data в интернет-компаниях. Лямбда-архитектура для BI-проектов. Подход и рекомендуемая архитектура SAP для использования инструментов Big Data в составе корпоративных системных ландшафтов. Новые направления, возникающие в результате применения и дальнейшего развития инструментария Big Data, в научных дисциплинах, использующих моделирование. Big Calculation. Big Simulation. Big Management. Big Optimal Control</p>	
	1.1	MapReduce	<p>Детальная схема работы MapReduce. YARN. HDFS. Apache Avro. Apache Pig. Apache Oozie. Apache Tez. Платформа OpenStack. OpenStack Sahara. Платформа Microsoft Azure. Платформа IBM Bluemix. Экосистема SAP. Платформа SAP HANA. Amazon EC2. Apache Spark. Apache Flink.</p>	2
	1.1	Средства потоковой обработки данных.	<p>Средства потоковой обработки данных. Apache Storm. Trident поверх Storm. Apache Kafka. Apache Flume. Spark Streaming. Потоковая обработка в Apache Flink. Apache Beam. MapR Streams. SAP HANA Smart Data Streaming. Декларативные языки для обработки потоковых данных. SAP HANA CCL. Интернет вещей. Архитектура приложений IoT.</p>	2

			Apache NiFi. Microsoft Azure IoT Reference Architecture. SAP Leonardo	
	1.2	Базы данных NoSQL	<p>Базы данных NoSQL. Причины появления систем NoSQL баз данных и их роль в развитии технологий обработки больших данных. 4 этапа развития СУБД и 3 революции, произошедшие в ходе этого развития. Типы систем NoSQL баз данных и тенденции развития систем NoSQL баз данных. HBase. Hipertable. Accumulo. Cassandra. MongoDB. CouchDB как предшественник Couchbase. RavenDB. Redis. Riak. Kudu. Введение в Search Engines. Apache Lucene. Apache Nutch и Apache Solr как развитие Apache Lucene. Elasticsearch. Splunk. Графовые СУБД и граф-энджины. Общий обзор графовых СУБД. Neo4j. OrientDB. Titan. Общий обзор Graph Engines. Pregel. Giraph. Sufer, Signal/Collect и Mizan. Семейство продуктов GraphLab. PEGASUS. Apache Spark GraphX. Совмещение графовых СУБД и Graph Engine в одном продукте. SAP HANA Graph. Microsoft Graph Engine</p>	2
	1.2	SQL СУБД для работы с большими данными	<p>SQL СУБД для работы с большими данными. История и тенденции развития SQL СУБД, способных обрабатывать большие данные. Couchbase N1QL похожий на SQL. HiveQL - почти как SQL. Couchbase. Apache Phoenix. Apache Drill. Cloudera Impala. Apache HAWQ. HP Vertica. IBM BigSQL. SAP HANA Vora</p>	2

### 3.2.2. Практические занятия, содержание и объем в часах

Модуль	Номер раздела	Тема	Содержание	Трудоемкость (в часах)

### 3.2.3. Лабораторные занятия, содержание и объем в часах

Модуль	Номер раздела	Тема	Содержание	Трудоемкость (в часах)
1	1.1	Базовые понятия Big Data	Знакомство с Hadoop	8
	1.1	Средства потоковой обработки данных	Знакомство со Spark	8
	1.2	Базы данных NoSQL	Знакомство с SAP HANA	6
	1.2	SQL СУБД для работы с большими данными	Взаимодействие с SAP HANA через Web	6
	1.2	SQL СУБД для работы с большими данными	Обработка больших графов в Microsoft Graph Engine.	5

### 3.3. Содержание материалов, выносимых на самостоятельное изучение

Модуль	Номер раздела	Содержание материалов, выносимого на самостоятельное изучение	Виды самостоятельной деятельности	Трудоемкость (в часах)
1	1.1	Базовые понятия систем обработки больших данных	Проработка материала лекций. Подготовка к лабораторным работам. Подготовка к рубежному контролю. Выполнение домашнего задания.	14
	1.2	Nosql- и sql-системы баз данных, использующие шардирование	Проработка материала лекций Подготовка к лабораторным работам Подготовка к рубежному контролю Выполнение домашнего задания	14

### 4. Фонд оценочных средств для проведения текущей и промежуточной аттестации обучающихся по дисциплине

Фонд оценочных средств текущего контроля и промежуточной аттестации по итогам

освоения дисциплины представлен в приложении.

[Фонд оценочных средств](#)

## 5. Учебно-методическое и информационное обеспечение дисциплины

### 5.1. Основная литература

#### 5.1.1. Печатные издания

1. 1. Лэм Ч. Hadoop в действии. – М.:ДМК Пресс, 2012. – 424с.: ил. Режим доступа: <https://e.lanbook.com/reader/book/39997/#1> (дата обращения: 17.12.2016). 2. Карау Х., Конвински Э., Венделл П., Захария М. Изучаем Spark: молниеносный анализ данных. [М.: ДМК Пресс, 2015. [ ] 304 с.: ил. Режим доступа: <https://e.lanbook.com/reader/book/90118/#1> (дата обращения: 17.12.2016).

#### 5.1.2. Издания из ЭБС

1.

### 5.2. Дополнительная литература

#### 5.2.1. Печатные издания

1. 1. Big Data Now. 2015-edition. O'Reilly Media, Inc., 2016 Dequenne F. Why HSMs still have a place at ECMWF Tape is dead! Or is it?, 2. Рахматуллин Д. Я. Введение в MPI.– ИМБИЦ УНЦ РАН, Уфа Уайт Т. Hadoop: Подробное руководство. [ ] СПб.: Питер, 2013. [ ] 672 с.: ил. [ ] (Серия «Бестселлеры O'Reilly»). 3. Oozie / coreservlets.com and Dima May, 2012, 22 p. Islam M.K., Srinivasan A. Apache Oozie.– O'Reilly Media, Inc., 2015

#### 5.2.2. Издания из ЭБС

1.

### 5.3. Базы данных, информационно-справочные и поисковые системы

Название	Ссылка
Ресурс по теме «Понятие больших данных и технологий обработки больших данных	<a href="https://www.nist.gov/itl/big-data-nist">https://www.nist.gov/itl/big-data-nist</a>
Ресурс по Apache Avro:	<a href="https://www.tutorialspoint.com/avro/index.htm">https://www.tutorialspoint.com/avro/index.htm</a>

## 6. Перечень программного обеспечения

Программное обеспечение общего назначения: ОС Microsoft Windows, Microsoft Office, АБВУУ FineReader, ESET NOD32 Smart Security Business Edition, Foxit Reader, АИБС "МегаПро".

Программное обеспечение специального назначения:

1) Apache OpenOffice

## 7. Материально-техническое обеспечение дисциплины

Наименование помещений для проведения учебных занятий и для самостоятельной работы обучающихся	Оснащенность специальных помещений и помещений для самостоятельной работы
Учебные аудитории для проведения занятий лекционного типа	Состав оборудования и технических средств обучения указан в паспорте аудитории, закрепленной расписанием по факультету
Учебные аудитории для проведения лабораторных занятий	
Учебные аудитории для промежуточной аттестации	
Учебные аудитории для проведения групповых и индивидуальных консультаций	Состав оборудования и технических средств обучения указан в паспорте аудитории, закрепленной расписанием по кафедре
Учебные аудитории для текущей аттестации	

## 8. Методические рекомендации по организации изучения дисциплины

Приступая к работе, каждый студент должен принимать во внимание следующие положения.

Дисциплина построена по модульному принципу, каждый модуль представляет собой логически завершённый раздел курса. Дисциплина делится на три модуля, включая экзамен. На первом занятии каждый студент получает в электронном виде полный комплекс учебно-методических материалов по дисциплине.

Лекционные занятия посвящены рассмотрению ключевых, базовых положений курса и разъяснению учебных заданий, выносимых на самостоятельную проработку.

Лабораторные работы предназначены для приобретения опыта практической реализации основной профессиональной образовательной программы. Техническая литература по теме лабораторной работы прорабатывается студентами во время самостоятельной подготовки. Необходимый уровень подготовки контролируется перед проведением лабораторных работ. Самостоятельная работа студентов включает проработку лекционного курса, подготовку к лабораторным работам, подготовку к рубежным контролям, выполнение домашних заданий, подготовку к экзамену.

Текущий контроль проводится в течение каждого модуля, его итоговые результаты складываются из оценок по следующим видам контрольных мероприятий:

- ☐ домашние задания;
- ☐ защита лабораторных работ;
- ☐ рубежные контроли.

Освоение дисциплины, ее успешное завершение на стадии промежуточной аттестации возможно только при регулярной работе во время семестра и планомерном прохождении текущего контроля.

Для завершения работы в семестре студент должен выполнить все контрольные

мероприятия.

Промежуточная аттестация по результатам семестра по дисциплине проходит в форме экзамена, контролирующего освоение ключевых, базовых положений дисциплины, составляющих основу остаточных знаний по ней.

Методика оценки по рейтингу

Студент, выполнивший все предусмотренные учебным планом задания и сдавший все контрольные мероприятия, получает итоговую оценку по дисциплине за семестр в соответствии со шкалой:

Рейтинг Оценка на экзамене

85 – 100 отлично

71 – 84 хорошо

60 – 70 удовлетворительно

0 – 59 неудовлетворительно

Оценивание дисциплины ведется в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов ЗабГУ.

Разработчик/группа разработчиков:  
Лидия Леонидовна Яковлева

**Типовая программа утверждена**

Согласована с выпускающей кафедрой  
Заведующий кафедрой

\_\_\_\_\_ «\_\_\_» \_\_\_\_\_ 20\_\_\_ г.