

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Утверждена на заседании
Ученого совета ИМФКН
«___» _____ 20__ г.
Протокол №__

Рабочая программа дисциплины

Технологии сбора и обработки больших данных

Направление подготовки
02.04.03 Математическое обеспечение и администрирование
информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ
2025

Пояснительная записка

Цели освоения дисциплины

Цель дисциплины - освоение современных методов и технологий сбора и обработки больших данных.

Место дисциплины в структуре образовательной программы

Дисциплина изучается во 2 и 3 семестрах и является дисциплиной по выбору, входит в вариативную часть блока Б1 ОП по направлению подготовки 02.04.03 «Математическое обеспечение и администрирование информационных систем» (магистратура). Основывается на курсах: "Методы и технологии глубокого обучения", "Методы оптимизации".

Планируемые результаты обучения по дисциплине и индикаторы достижения компетенций.

В результате освоения дисциплины студент должен:

Знать:

Общие принципы работы с большими данными; Основные концепции вычислительных технологий больших данных; Типовые задачи обработки больших данных

Уметь:

Осуществлять сбор и хранение больших данных; применять основные концепции вычислительных технологий больших данных; решать типовые задачи обработки больших данных с применением современного инструментария

Владеть:

Навыками работы со стеком технологий Hadoop; навыками работы с конкретными инструментами стека технологий Hadoop; навыками решения типовых задач

Планируемые результаты освоения образовательной программы:

- | | |
|--------|---|
| ПК-1 | Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования |
| ПК-1.1 | Выбирает и разрабатывает программные компоненты систем искусственного интеллекта |
| ПК-2 | Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач |
| ПК-2.1 | Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области |
| ПК-3 | Способен руководить проектами по созданию комплексных систем искусственного интеллекта |
| ПК-3.1 | Применяет методы и средства разработки систем искусственного интеллекта |

Объем дисциплины в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 11 зачетные единицы, 396 часа.

№	Название разделов дисциплины	Лабораторная работа	Самостоятельная работа	Лекция
	Семестр 2	16	184	16
1	Фреймворк Apache Spark и иной инструментарий обработки данных	16	184	16
	Семестр 3	16	148	16
1	Стек технологий Hadoop и методы хранения больших данных-2	6	70	10
2	Стек технологий Hadoop и методы хранения больших данных-1	10	78	6

Тематическое планирование курса

Темы

Фреймворк Apache Spark и иной инструментарий обработки данных

Семестр 2

Apache Spark

Лабораторная работа. 6(0) ч. Модель вычислений Resilient Distributed Dataset (RDD). Spark MLlib

Самостоятельная работа. 14(0) ч. Spark SQL

Самостоятельная работа. 14(0) ч. SparkSQL и Hive

Лабораторная работа. 6(0) ч. Spark GraphX

Самостоятельная работа. 14(0) ч. Spark Streaming

Самостоятельная работа. 14(0) ч. Операции над RDD

Самостоятельная работа. 14(0) ч. Платформа Kaggle

Самостоятельная работа. 16(0) ч. Cloudera's Distribution including Apache Hadoop (CDH) и подобные. Абстракция DataFrame. Работа с Spark MLlib

Самостоятельная работа. 14(0) ч. Основные возможности Spark MLlib

Самостоятельная работа. 14(0) ч. Необходимые сведения из Scala. Основные структуры данных: Списки, Наборы, Кортеж, Карты

Лабораторная работа. 4(0) ч. Необходимые сведения из Scala. Функциональные Комбинаторы: map, foreach, filter, zip, partition, find, drop и dropWhile, foldRight и foldLeft, flatten, flatMap, Обобщенные функциональные комбинаторы

Самостоятельная работа. 14(0) ч. Платформа Kaggle. Spark MLlib

Самостоятельная работа. 14(0) ч. Работа с Spark GraphX

Самостоятельная работа. 14(0) ч. Работа с Spark Streaming

Самостоятельная работа. 14(0) ч. Hadoop Cloudera

Самостоятельная работа. 14(0) ч. Решение задач Kaggle

Лекция. 6(0) ч. Модель вычислений Resilient Distributed Dataset (RDD).Spark MLlib

Лекция. 4(0) ч. Необходимые сведения из Scala.

Лекция. 6(0) ч. Функциональные Комбинаторы: map, foreach, filter, zip, partition, find, drop и dropWhile, foldRight и foldLeft, flatten, flatMap, Обобщенные функциональные комбинаторы

Стек технологий Hadoop и методы хранения больших данных-2

Семестр 3

Обработка данных

Лабораторная работа. 2(0) ч. Принципы работы с Apache Hive. Принципы работы с Apache Pig

Самостоятельная работа. 10(0) ч. HQL (Hive query language)

Самостоятельная работа. 10(0) ч. Язык Pig Latin. Структуры данных Tuple и Bag. Базовые функции Pig Latin

Самостоятельная работа. 10(0) ч. Язык запросов Hive query language

Самостоятельная работа. 10(0) ч. Pig Latin

Самостоятельная работа. 10(0) ч. Настройка и запуск Hadoop и Pig. Типы данных в Pig Latin

Лабораторная работа. 2(0) ч. Типы данных в Pig Latin

Лекция. 4(0) ч. Принципы работы с Apache Hive. Принципы работы с Apache Pig

Лекция. 4(0) ч. Типы данных в Pig Latin

Иные технологии

Самостоятельная работа. 10(0) ч. Vowpal Wabbit. Фреймворк Caffe и концепция deep learning

Лабораторная работа. 2(0) ч. Vowpal Wabbit

Самостоятельная работа. 10(0) ч. Vowpal Wabbit

Лекция. 2(0) ч. Vowpal Wabbit

Стек технологий Hadoop и методы хранения больших данных-1

Семестр 3

Сбор и хранение данных

Лабораторная работа. 6(0) ч. модель вычислений Map-Reduce. Файловая система HDFS

Самостоятельная работа. 12(0) ч. Посредник YARN и планировщик Oozie
Самостоятельная работа. 12(0) ч. Планировщик Oozie
Самостоятельная работа. 12(0) ч. NoSQL базы данных. Импорт с помощью Apache Sqoop
Самостоятельная работа. 10(0) ч. NoSQL база данных Cassandra
Лабораторная работа. 2(0) ч. Импорт данных
Самостоятельная работа. 12(0) ч. Импорт с помощью Apache Flume
Самостоятельная работа. 12(0) ч. Импорт данных в HDFS
Самостоятельная работа. 8(0) ч. NoSQL база данных HBase
Лабораторная работа. 2(0) ч. Реализация Map-Reduce в Hadoop. Файловая система HDFS
Лекция. 2(0) ч. Модель вычислений Map-Reduce. Файловая система HDFS
Лекция. 2(0) ч. Импорт данных
Лекция. 2(0) ч. Реализация Map-Reduce в Hadoop. Файловая система HDFS Действия

БРС

Семестр	Контрольные точки	Баллы
2	Текущий контроль в разделе «Фреймворк Apache Spark и иной инструментарий обработки данных»	
	Разработка проекта	40
	Составление структурно-логической схемы	20
2	Зачет	
	Разработка проекта	40
Итого за семестр 2:		100
3	Текущий контроль в разделе «Стек технологий Hadoop и методы хранения больших данных-2»	
	Разработка проекта	20
	Составление структурно-логической схемы	10
3	Текущий контроль в разделе «Стек технологий Hadoop и методы хранения больших данных-1»	

Семестр	Контрольные точки	Баллы
	Разработка проекта	20
	Составление структурно-логической схемы	10
3	Зачет	
	Разработка проекта	40
Итого за семестр 3:		100

Учебно-методическое и информационное обеспечение учебного процесса

Образовательные технологии (в том числе на занятиях, проводимых в интерактивных формах).

При изучении данного курса применяются как традиционные (лекции, практические занятия, экзамен), так и инновационные образовательные технологии, которые реализуются в учебном процессе в активных и интерактивных формах проведения занятий, из которых можно выделить следующие:

- 1) лекция-дискуссия при рассмотрении тем "NoSQL базы данных", "Платформа Kaggle".
- 2) метод группового решения задач при изучении тем "Работа с Caffe", "Работа с Spark MLlib".
- 3) перекрестная самопроверка при изучении тем "Настройка и запуск Hadoop и Pig", "Hadoop Cloudera".

Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины

К современному специалисту общество предъявляет достаточно широкий перечень требований, среди которых немаловажное значение имеет наличие у выпускников определенных способностей и умения самостоятельно добывать знания из различных источников, систематизировать полученную информацию, давать оценку конкретной ситуации. Формирование такого умения происходит в течение всего периода обучения через участие студентов в лабораторных, практических занятиях, выполнение заданий и тестов, написание курсовых и выпускных квалификационных работ. При этом самостоятельная работа студентов играет решающую роль в ходе всего учебного процесса.

- [СРС маг_010401 Математика_2015 Технологии сбора и обработки больших данных_05151m.doc](https://my.bsu.ru/content/file/3/36/363/103424_srs_mag_010401_matematika_2015_tehnologii-sbora-i-obrabotki-bolshih-dannih_05151m.doc)
Режим доступа:
https://my.bsu.ru/content/file/3/36/363/103424_srs_mag_010401_matematika_2015_tehnologii-sbora-i-obrabotki-bolshih-dannih_05151m.doc

Учебно-методические материалы, в том числе методические указания для обучающихся по освоению дисциплины

— На лекционных занятиях студент слушает доклад преподавателя, составляет конспект лекции. Во время лекции студенту рекомендуется делать отметки на полях тетради, касающиеся того теоретического материала, который вызвал затруднения в понимании.

После лекции трудности необходимо устранить путем консультации у преподавателя или самостоятельной работы с рекомендованной учебной литературой.

— На лабораторных занятиях студенту предлагается ряд заданий по теме, прослушанной на лекции. По заданию необходимо решить поставленную задачу, пользуясь персональным компьютером и необходимым программным обеспечением. Контроль знаний, умений и навыков студентов осуществляется путём проведения на лабораторных занятиях компьютерных тестов, сдачи проектов. Студенту, выполнившему то или иное задание на низкое количество баллов, по согласованию с преподавателем, необходимо выполнить работу над ошибками для улучшения результатов.

— Предусмотрена самостоятельная работа студентов, в рамках которой студентом должны выполняться работа над проектами, анализ требований, ошибок, перспектив и т.п.. Теоретический материал студентом должен быть проработан с использованием конспектов лекций (если в плане предусмотрены лекции) или рекомендуемой учебной литературы.

Советы по планированию и организации времени, необходимого для изучения дисциплины.

Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины:

В день после лекционного занятия в течение 1 астрономического часа изучить лекцию, выявить моменты, где имеются вопросы и недостаточное понимание. По этим вопросам изучить рекомендованную литературу, устранить непонимание. В день перед проведением практического занятия в течение 30 минут повторить пройденный материал, предшествующий теме занятия.

При изучении дисциплины очень полезно самостоятельно изучать материал, который еще не прочитан на лекции, не применялся на лабораторном или практическом занятии. Тогда предстоящее занятие будет гораздо понятнее.

Рекомендации по работе с литературой. Теоретический материал курса становится более понятным, когда дополнительно к прослушиванию лекций, изучению конспектов, изучаются и книги по изучаемой дисциплине. Литературу по дисциплине рекомендуется изучать в библиотеке. Полезно использовать несколько учебников по дисциплине.

Рекомендуется, кроме «заучивания» материала, добиться состояния понимания изучаемой темы дисциплины. С этой целью рекомендуется после изучения очередного параграфа выполнить несколько простых упражнений на данную тему.

Советы по подготовке к экзамену:

1) Ознакомиться с процедурой проведения экзамена, ознакомиться со списком вопросов для подготовки к экзамену. Повторить весь материал пройденных лекций (если в плане были лекции), проработать конспекты, рекомендованную учебную литературу. Составить конспекты ответов на экзаменационные вопросы.

2) Собрать итоговые выполненные лабораторные и проекты (если в плане были лабораторные и проекты), если произведено их улучшение согласно замечаниям преподавателя. В рамках экзамена/зачета возможен пересмотр преподавателем баллов (по усмотрению преподавателя, с учетом общей работы студента во время курса), выставленных ранее за выполненные задания, при условии получения улучшения результатов

Оценочные средства

По данной дисциплине разработаны оценочные средства, критерии их оценивания, а также методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций (в приложении).

Список литературы

Перечень основной и дополнительной литературы, необходимой для освоения дисциплины.

Основная

1. [ОСНОВЫ ИСПОЛЬЗОВАНИЯ И ПРОЕКТИРОВАНИЯ БАЗ ДАННЫХ](#): Учебник/Илюшечкин В.М.. —М.: Издательство Юрайт, 2016. —213 с.
Режим доступа: <http://www.biblio-online.ru/book/1C650A7F-DC7D-4834-998E-42D06FC8EF33>
2. [ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ](#): Учебник и практикум/Миркин Б.Г.. —М.: Издательство Юрайт, 2016. —174 с.
Режим доступа: <http://www.biblio-online.ru/book/E486A3B0-1AE9-4179-8D48-FA24B626C7C9>
3. [АНАЛИЗ ДАННЫХ](#): Учебник/Мхитарян В.С. - Отв. ред.. —М.: Издательство Юрайт, 2016. —490 с.
Режим доступа: <http://www.biblio-online.ru/book/AF1D197F-1759-422E-9593-8B43E2D1093B>

Дополнительная

1. [Применение искусственных нейронных сетей и системы остаточных классов в криптографии](#)/ [Н.И. Червяков и др.]. —Москва: Физматлит, 2012. —279 с.
Режим доступа: http://e.lanbook.com/books/element.php?pl1_cid=25&pl1_id=5300
2. [Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных](#): научно-популярная литература/П. Флах ; пер. с англ. А. А. Слинкин. —Москва: ДМК Пресс, 2015. —400 с.
Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=69955
3. [Scala для нетерпеливых](#)/К. Хостманн. —Москва: ДМК Пресс, 2013. —408 с.
Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=66473
4. [Python и анализ данных](#): научно-популярная литература/У. Маккинли ; [пер. с англ. А. А. Слинкин]. —Москва: ДМК Пресс, 2015. —482 с.
Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=73074
5. [Python на практике](#)/Саммерфилд М.. —Москва: ДМК Пресс, 2014
Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=66480

Перечень ресурсов информационно-коммуникационной сети «Интернет», необходимых для освоения дисциплины

<https://www.ibm.com/developerworks/ru/library/os-log-process-hadoop/>

<https://www.kaggle.com/c/digit-recognizer>

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

<http://hadoop.apache.org/>

Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Учебная аудитория для проведения лекционных занятий; Учебная аудитория для проведения занятий лабораторного типа с доступом в Интернет; Помещение для самостоятельной работы с доступом в Интернет;

Учебная аудитория для проведения индивидуальных и групповых консультаций; Учебная

аудитория для проведения текущей и промежуточной аттестации; Требуемый перечень программного обеспечения: ОС Windows/Ubuntu; Интегрированная среда разработки Code::Blocks 13.12 и выше.

ПРИЛОЖЕНИЕ

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет имени Доржи Банзарова»
Институт математики, физики и компьютерных наук
Кафедра информационных систем и методов искусственного интеллекта

Фонд оценочных средств по учебной дисциплине

Технологии сбора и обработки больших данных

Направление подготовки
02.04.03 Математическое обеспечение и администрирование информационных систем

Квалификация
Магистр

Форма обучения
Очная

Улан-Удэ
2025

**Паспорт фонда оценочных средств (ФОС) по дисциплине
«Технологии сбора и обработки больших данных»**

Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

ПК-1	Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования
ПК-1.1	Выбирает и разрабатывает программные компоненты систем искусственного интеллекта
ПК-2	Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач
ПК-2.1	Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области
ПК-3	Способен руководить проектами по созданию комплексных систем искусственного интеллекта
ПК-3.1	Применяет методы и средства разработки систем искусственного интеллекта

Этапы формирования компетенции

Семестр	Вид контроля	Оценочные средства
1 семестр	Текущий	Разработка проекта Составление структурно-логической схемы
	Итоговый (зачет)	Разработка проекта
2 семестр	Текущий	Разработка проекта Составление структурно-логической схемы
	Итоговый (зачет)	Разработка проекта

Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Фонд оценочных средств сформирован на основе ключевых принципов оценивания:

- валидность: объекты оценки должны соответствовать поставленным целям обучения;
- определенность: оценочные средства должны быть понятны каждому обучающемуся;
- однозначность: одинаковость оценки качества оценочного средства;
- надежность: использование единообразных показателей и критериев для оценивания достижений.

Описание показателей и критериев оценивания компетенций, а также шкал оценивания

Показатели оценивания компетенций	Уровень сформированности компетенций	Шкала оценивания	Официальный цифровой эквивалент оценки
Знать: Общие принципы работы с большими данными; Основные концепции вычислительных технологий больших данных;	Высокий	85 – 100 баллов	5 (отлично)
	Базовый	70 – 84 баллов	4 (хорошо)

Типовые задачи обработки больших данных Уметь: Осуществлять сбор и хранение больших данных; применять основные концепции вычислительных технологий больших данных; решать типовые задачи обработки больших данных с применением современного инструментария Владеть: Навыками работы со стеком технологий Hadoop; навыками работы с конкретными инструментами стека технологий Hadoop; навыками решения типовых задач	Пороговый	60-69 баллов	3 (удовлетворительно)
---	-----------	--------------	--------------------------

Балльно-рейтинговая система

Для текущего и итогового контроля качества обучения студентов и магистрантов применяется балльно-рейтинговая система, разработанная в соответствии с «Положением об организации учебного процесса с применением кредитно-модульной системы обучения», утвержденным Учебно-методическим советом ФГБОУ ВО «Бурятский государственный университет» от 20 февраля 2012 г. Целью БРС является определение уровня успешности освоения (завершения изучения) обучающимися учебных дисциплин (модулей, циклов) через балльные оценки и рейтинги качества сформированных знаний, умений, профессиональных компетенций, накапливаемые в соответствии с измеряемыми в зачетных единицах трудоемкостями каждого цикла (модуля, дисциплины) и основной образовательной программы в целом.

1. Общая максимальная сумма баллов, которую студент может набрать по дисциплине в течение семестра – 100 баллов: 60 баллов текущий контроль и 40 баллов итоговый контроль (экзамен).

2. Минимальная сумма баллов, при которой студент допускается к экзамену (итоговому контролю), равна 20 баллам.

3. Минимальная сумма баллов, при которой студент получает положительную итоговую оценку по дисциплине равна 60 баллам (60% от 100 баллов).

4. Максимальная оценка за выполнение одной лабораторной работы – 10 баллов.

Связь между четырехбалльной и стобалльной системами оценки качества обучения студентов

Оценка	Рейтинговые баллы
Отлично	80-100
Хорошо	70-80
Удовлетворительно	60-70
Неудовлетворительно	<60

ЗАДАНИЯ НА РАЗРАБОТКУ ПРОЕКТОВ И КРИТЕРИИ ИХ ОЦЕНИВАНИЯ

Задания на разработку проектов по разделу «Стек технологий Hadoop и методы хранения больших данных-1»

1 задание. Создать готовый для работы образ виртуальной машины (Virtual Box) с установленной (свободно распространяемой) операционной системой и инструментарием из стека технологий Hadoop – текущий контроль

2 задание. Разработать программный комплекс на базе стека технологий Hadoop, который производит извлечение полезных данных из журналов в типичных Linux-системах <https://www.ibm.com/developerworks/ru/library/os-log-process-hadoop/> – зачет.

Задания на разработку проектов по разделу «Стек технологий Hadoop и методы хранения больших данных-2»

Задание 1. Разработать программный комплекс на базе стека технологий Hadoop и фреймворка Vowpal Wabbit, решающий задачу предсказания выживания пассажиров Титаника <https://www.kaggle.com/c/titanic> – текущий контроль

Задание 2. Разработать программный комплекс на базе стека технологий Hadoop и фреймворка Caffe, решающий задачу распознавания рукописных цифр из выборки MNIST (Mixed National Institute of Standards and Technology database) <https://www.kaggle.com/c/digit-recognizer> – зачет.

Задание на разработку проектов по разделу «Фреймворк Apache Spark и иной инструментарий обработки данных»

Задание 1. Разработать программный комплекс на базе стека технологий Hadoop и фреймворка Apache Spark (в частности, Spark MLib), решающий задачу предсказания выживания пассажиров Титаника <https://www.kaggle.com/c/titanic> – текущий контроль

Задание 2. Разработать программный комплекс на базе стека технологий Hadoop и фреймворка Apache Spark (в частности, Spark MLib), решающий задачу распознавания рукописных цифр из выборки MNIST (Mixed National Institute of Standards and Technology database) <https://www.kaggle.com/c/digit-recognizer> – зачет.

Критерии оценивания проектов

Программные комплексы (ПК) представляются в электронном виде. Выбор языка разработки (Python, Java, Scala, R и др.), сторонних решений, фреймворков и т.п. осуществляется пользователем, помимо указанных ниже требований не предъявляется.

ПК должны удовлетворять следующим требованиям:

1. ПК должен быть реализован на стеке технологий Hadoop, ПК должен быть работоспособным и должен решать указанную задачу.
2. ПК должен сопровождаться пользовательской документацией, достаточной для того, чтобы пользователь, незнакомый ранее с данным ПК мог с ним работать.
3. ПК должен сопровождаться документацией разработчика, достаточной для того, чтобы разработчик, незнакомый ранее с данным ПК мог вносить необходимые изменения.

ПК	Количество баллов
В качестве ПК использовано полностью стороннее решение или ПК не представлен	0-5

Не выполнено два пункта требований к ПК и/или не выполнен п.1 требований к ПК	6-15
Не выполнен один пункт требований к ПК, при этом выполнен п.1. требований к ПК	16-29
Выполнены все пункты требований к ПК	30-40

Разработанный проект представляется студентом публично или преподавателю. После представления преподаватель задает вопросы по функционированию или реализации проекта. Итоговое количество баллов выставляется в рамках вышеприведенной таблицы в зависимости от представления и ответов на вопросы.

ЗАДАНИЯ НА РАЗРАБОТКУ СТРУКТУРНО-ЛОГИЧЕСКИХ СХЕМ И КРИТЕРИИ ИХ ОЦЕНИВАНИЯ

Задание на разработку структурно-логической схемы по разделу «Стек технологий Hadoop и методы хранения больших данных-1»

Разработать СЛС «Стек технологий Hadoop: состав и использование»

СЛС должна включать:

- все программные решения, относимые к Hadoop, с небольшим описанием их назначения и функциональности
- описание типовых кейсов использования структурных элементов схемы

Задание на разработку структурно-логической схемы по разделу «Стек технологий Hadoop и методы хранения больших данных-2»

Разработать СЛС «Pig и Hive»

СЛС должна включать:

- структуру и обзор возможностей Pig
- структуру и обзор возможностей Hive

Задание на разработку структурно-логической схемы по разделу «Фреймворк Apache Spark и иной инструментарий обработки данных»

Разработать СЛС «Фреймворк Apache Spark: состав и использование»

СЛС должна включать:

- все программные решения, относимые к Apache Spark, с небольшим описанием их назначения и функциональности
- описание типовых кейсов использования структурных элементов схемы

Критерии оценивания структурно-логических схем по разделам «Программное обеспечение вычислительного эксперимента» и «Фреймворк Apache Spark и иной инструментарий обработки данных»

Характеристика СЛС	Количество баллов
Структурно-логическая схема не соответствует требованиям задания на разработку; Включено не более 50 структурных единиц описания (для схем 3.1 3.2)	0-7
Структурно-логическая схема соответствует большей части требований задания на разработку; Включено более 50, но менее 100 структурных единиц описания (для	8-15

схем 3.1, 3.2)	
Структурно-логическая схема не соответствует требованиям задания на разработку; Включено более 100 структурных единиц описания (для схем 3.1, 3.2)	16-20

Разработанная структурно-логическая схема представляется студентом публично или преподавателю. После представления преподаватель задает вопросы по содержанию структурно-логических схем. Итоговое количество баллов выставляется в рамках вышеприведенной таблицы в зависимости от представления и ответов на вопросы.