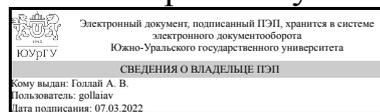


УТВЕРЖДАЮ:
Директор института
Высшая школа электроники и
компьютерных наук



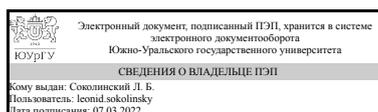
А. В. Голлай

РАБОЧАЯ ПРОГРАММА

**дисциплины 1.Ф.02 Технологии распределенной обработки данных
для направления 09.04.04 Программная инженерия
уровень Магистратура
форма обучения очная
кафедра-разработчик Системное программирование**

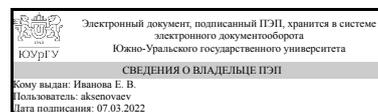
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



Л. Б. Соколинский

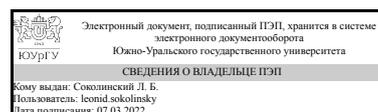
Разработчик программы,
к.физ.-мат.н., доцент



Е. В. Иванова

СОГЛАСОВАНО

Руководитель направления
д.физ.-мат.н., проф.



Л. Б. Соколинский

1. Цели и задачи дисциплины

Целью курса является изучение студентами задач, связанных с распределенным хранением и обработкой больших данных. При изучении этого курса должны быть решены следующие задачи: изучить понятие и проблематику больших данных, способы распределенного хранения больших данных, способы распределенной обработки больших данных, хранение и обработка больших данных с помощью современных программных решений.

Краткое содержание дисциплины

Понятие больших данных. Распределенная обработка больших данных. Теорема CAP. SQL, NoSQL и NewSQL-решения. Структуры для хранения больших данных. Секционирование данных. Репликация данных. Согласованность данных в базе данных. Экосистема Hadoop: HDFS, MapReduce, Pig, Apache Hive, HBase, Apache Spark, MLlib, Hadoop YARN, Zookeeper, Apache Kafka. Классификация NoSQL-решений: хранилища "ключ-значения", документо-ориентированные хранилища, хранение в виде семейства столбцов, графовые СУБД.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен разрабатывать системы хранения и обработки больших данных	Знает: основные положения и концепции в области хранения и обработки больших данных Умеет: анализировать типовые решения в области хранения и обработки больших данных, реализовывать техническое сопровождение информационных систем и баз данных, используемых для решения задач в области хранения и обработки больших данных, программировать системы хранения и обработки больших данных Имеет практический опыт: интеграции различных типов программного обеспечения в области хранения и обработки больших данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.Ф.08 Методы и системы обработки больших данных	Не предусмотрены

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
1.Ф.08 Методы и системы обработки больших	Знает: принципы и методы построения

данных	общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных Умеет: Имеет практический опыт:
--------	---

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 56,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		3	
Общая трудоёмкость дисциплины	108	108	
<i>Аудиторные занятия:</i>	48	48	
Лекции (Л)	32	32	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	16	16	
Лабораторные работы (ЛР)	0	0	
<i>Самостоятельная работа (СРС)</i>	51,5	51,5	
с применением дистанционных образовательных технологий	0		
Изучение тем и проблем, не выносимых на лекции и практические занятия	40	40	
Подготовка к экзамену	11,5	11,5	
Консультации и промежуточная аттестация	8,5	8,5	
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен	

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Понятие больших данных. Распределенное хранение и обработка больших данных	12	12	0	0
2	Экосистема Hadoop	32	16	16	0
3	NoSQL-решения	4	4	0	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение. Понятие больших данных. Проблема больших данных. Система хранения больших данных. Распределенная обработка больших данных. Введение в репликацию и секционирование данных. Теорема CAP. SQL, NoSQL и NewSQL-решения.	2
2	1	Структуры для хранения больших данных. Хеш-индексы. SS-таблицы. LSM-деревья. B-деревья.	2

3	1	Секционирование. Виды секционирования: по диапазонам значений ключа, по хешу ключа. Добавление/удаление секций, методы перебалансировки.	2
4	1	Репликация. Виды репликации: синхронная, асинхронная, полусинхронная репликация. Репликация с одним ведущим узлом. Добавление узлов в систему. Обработка сбоев узлов. Журнал репликации. Задержка репликации.	2
5	1	Репликация с несколькими ведущими узлами. Репликация без ведущего узла. Чтение и запись по кворуму. Алгоритм «векторные часы».	2
6	1	Согласованность. Виды согласованности: строгая (Strong Consistency), конечная (Eventual Consistency), согласованное префиксное чтение (Consistent Prefix), с ограниченным устареванием (Bounded Staleness), монотонные чтения (Monotonic Reads, Session guarantee), чтение своих записей (Read My Writes).	2
7	2	Введение в платформу Hadoop. Экосистема Hadoop. Примеры систем на базе Hadoop.	2
8	2	Распределенная файловая система Hadoop (HDFS).	2
9	2	Обработка данных в Hadoop. Технология MapReduce. Введение в Pig.	2
10	2	Обработка данных в Hadoop. Введение в СУБД Apache Hive и СУБД HBase.	2
11	2	Обработка данных в Hadoop. Apache Spark. Библиотека машинного обучения MLlib.	2
12	2	Архитектура MapReduce. Планирование и управление ресурсами с помощью Hadoop YARN.	2
13	2	Координация распределенных приложений с Zookeeper.	2
14	2	Брокер сообщений Apache Kafka.	2
15	3	Классификация NoSQL-решений. Хранилища "ключ-значение". Документо-ориентированные хранилища. Хранение в виде семейства столбцов.	2
16	3	Графовые СУБД. Модель графа свойств (property graph). Модель хранилища тройных кортежей (triple-store). Язык запросов Cypher. Фреймворк Pregel.	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	2	Установка платформы Hadoop	4
2	2	Разработка приложения MapReduce	4
3	2	Разработка статистических отчетов с использованием Apache Hive	4
4	2	Анализ данных в Hadoop	4

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Изучение тем и проблем, не выносимых на лекции и практические занятия	[Осн. лит., 3], Гл. 8-9, с. 273–342; [Доп. лит., 5], Гл. 1-4, с. 5–47.	3	40
Подготовка к экзамену	[Осн. лит., 1], Гл. 25, с. 458–466; [Осн. лит., 2], Гл. 4, с. 171–186; [Доп. лит., 4], Ч.3, гл. 13, с. 331–355.	3	11,5

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учи-тыва-ется в ПА
1	3	Текущий контроль	Письменный опрос 1	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 1. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
2	3	Текущий контроль	Письменный опрос 2	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 2. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
3	3	Текущий контроль	Письменный опрос 3	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 3. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности	экзамен

						обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	
4	3	Текущий контроль	Письменный опрос 4	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 4. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
5	3	Текущий контроль	Письменный опрос 5	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 1. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
6	3	Текущий контроль	Письменный опрос 6	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 6. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов,	экзамен

						разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	
7	3	Текущий контроль	Письменный опрос 7	2	2	<p>Письменный опрос проводится в виде электронного теста в конце лекции номер 7. Время, отведенное на опрос, -10 минут.</p> <p>При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179).</p> <p>Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.</p>	экзамен
8	3	Текущий контроль	Письменный опрос 8	2	2	<p>Письменный опрос проводится в виде электронного теста в конце лекции номер 8. Время, отведенное на опрос, -10 минут.</p> <p>При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179).</p> <p>Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.</p>	экзамен
9	3	Текущий контроль	Письменный опрос 9	2	2	<p>Письменный опрос проводится в виде электронного теста в конце лекции номер 9. Время, отведенное на опрос, -10 минут.</p> <p>При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179).</p> <p>Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.</p>	экзамен
10	3	Текущий	Письменный	2	2	Письменный опрос проводится в виде	экзамен

		контроль	опрос 10			электронного теста в конце лекции номер 10. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	
11	3	Текущий контроль	Письменный опрос 11	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 11. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
12	3	Текущий контроль	Письменный опрос 12	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 12. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
13	3	Текущий контроль	Письменный опрос 13	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 13. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-	экзамен

						рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	
14	3	Текущий контроль	Письменный опрос 14	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 14. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
15	3	Текущий контроль	Письменный опрос 15	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 15. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2 балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	экзамен
16	3	Текущий контроль	Письменный опрос 16	2	2	Письменный опрос проводится в виде электронного теста в конце лекции номер 16. Время, отведенное на опрос, -10 минут. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Письменный опрос оценивается в 2	экзамен

						балла. Каждый вопрос оценивается как общее количество вопросов, разделенное на 2 балла. Оценка студента - это сумма баллов за каждый вопрос, на который он дал правильный ответ.	
17	3	Текущий контроль	Практическое задание 1	17	17	17 баллов: задание полностью выполнено 1-16: задание выполнено частично 0: задание не выполнено	экзамен
18	3	Текущий контроль	Практическое задание 2	17	17	17 баллов: задание полностью выполнено 1-16: задание выполнено частично 0: задание не выполнено	экзамен
19	3	Текущий контроль	Практическое задание 3	17	17	17 баллов: задание полностью выполнено 1-16: задание выполнено частично 0: задание не выполнено	экзамен
20	3	Текущий контроль	Практическое задание 4	17	17	17 баллов: задание полностью выполнено 1-16: задание выполнено частично 0: задание не выполнено	экзамен
21	3	Бонус	Бонус-рейтинг	-	15	Студент представляет копии документов, подтверждающие победу или участие в предметных олимпиадах по темам дисциплины При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179). Максимально возможная величина бонус-рейтинга +15 %. +15 % за победу в олимпиаде международного уровня +10 % за победу в олимпиаде российского уровня +5 % за победу в олимпиаде университетского уровня +1 % за участие в олимпиаде.	экзамен
22	3	Промежуточная аттестация	Мероприятия промежуточной аттестации (компьютерное тестирование)	-	40	Промежуточная аттестация включает компьютерное тестирование. Контрольное мероприятие промежуточной аттестации проводятся во время экзамена. Тест состоит из 20 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 1 час. На экзамене происходит оценивание учебной деятельности обучающихся по дисциплине на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и	экзамен

	литература	библиотечная система издательства Лань	обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагина и А. Д. Хомоненко. — Санкт-Петербург : Лань, 2020. — 236 с. — ISBN 978-5-8114-4006-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/126938 (дата обращения: 11.10.2021).
3	Основная литература	Электронно-библиотечная система издательства Лань	Шарден, Б. Крупномасштабное машинное обучение вместе с Python : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018. — 358 с. — ISBN 978-5-97060-506-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/105836 (дата обращения: 11.10.2021).
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Григорьев, Ю. А. Реляционные базы данных и системы NoSQL : учебное пособие / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. — Благовещенск : АмГУ, 2018. — 424 с. — ISBN 978-5-93493-308-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/156492 (дата обращения: 11.10.2021).
5	Дополнительная литература	Электронно-библиотечная система издательства Лань	Бутаков, Н. А. Обработка больших данных с Apache Spark : учебно-методическое пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. — Санкт-Петербург : НИУ ИТМО, 2019. — 50 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/136573 (дата обращения: 11.10.2021).

Перечень используемого программного обеспечения:

1. РСК Технологии-Система "Персональный виртуальный компьютер" (ПВК) (MS Windows, MS Office, открытое ПО)(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	110 (3г)	Мультимедийный проектор
Экзамен	110 (3г)	Компьютерный класс с доступом к интернету
Практические занятия и семинары	110 (3г)	Компьютерный класс с доступом к интернету