

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Петербургский государственный университет путей сообщения
Императора Александра I»
(ФГБОУ ВО ПГУПС)

Кафедра «Информационные и вычислительные системы»

РАБОЧАЯ ПРОГРАММА

дисциплины

Б1.В.ДВ.1.2 «АРХИТЕКТУРА BIG DATA СИСТЕМ»

для направления подготовки
38.03.05 «Бизнес-информатика»

по профилю
«Цифровые технологии в экономике и бизнесе»

Форма обучения – очная

Санкт-Петербург
2025

ЛИСТ СОГЛАСОВАНИЙ

Рабочая программа рассмотрена и утверждена на заседании кафедры «Информационные и вычислительные системы»

Протокол № 7 от 19 марта 2025 г.

Заведующий кафедрой
«Информационные и вычислительные
системы»

С.Г. Ермаков

19 марта 2025 г.

СОГЛАСОВАНО

Руководитель ОПОП ВО

Т.П. Сацук

19 марта 2025 г.

1. Цели и задачи дисциплины

Рабочая программа дисциплины «Архитектура Big Data систем» (Б1.В.ДВ.1.2) (далее – дисциплина) составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.03.05 «Бизнес-информатика» (далее – ФГОС ВО), утвержденного 29.07.2020 г., приказ Министерства образования и науки Российской Федерации №838 (Редакция с изменениями №662 от 19.07.2022, №208 от 27.02.2023), с учетом профессионального стандарта (08.037) «Бизнес-аналитик», приказ Министерства труда и социальной защиты Российской Федерации от 22 ноября 2023 года, №821н. и профессионального стандарта (06.015) «Специалист по информационным системам», приказ Министерства труда и социальной защиты Российской Федерации от 13 июля, 2023 года, №586н.

Целью изучения дисциплины является изучение технологии Big Data, изучение методов обработки больших данных.

Для достижения цели дисциплины решаются следующие задачи:

- формирование практического навыка по работе с кластером хранения и обработки Big Data на примере программной платформы Quant;
- формирование практического навыка по выбору методов и технических решений в зависимости от типа решаемой задачи обработки данных и их объема;
- формирование практических навыков по использованию инструментария анализа данных и их визуализации;
- приобретение навыков управления проектом развития системы информационного обеспечения принятия управленческих решений от сбора данных, их хранения и обработки до представления результатов.

2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций

Планируемыми результатами обучения по дисциплине (модулю) является формирование у обучающихся компетенций (части компетенций). Сформированность компетенций (части компетенции) оценивается с помощью индикаторов достижения компетенций.

В рамках изучения дисциплины (модуля) осуществляется практическая подготовка обучающихся к будущей профессиональной деятельности. Результатом обучения по дисциплине является формирования у обучающихся практических навыков.

Индикаторы достижения компетенций	Результаты обучения по дисциплине (модулю)
ПК-1 Формирование возможных решений на основе разработанных для них целевых показателей	
ПК-1.2.8 Умеет пользоваться системами анализа и визуализации данных	Обучающийся умеет: – пользоваться системами анализа и визуализации больших данных.
ПК-3 Выявление требований к ИС в рамках выполнения работ и управления работами по созданию (модификации) и сопровождению ИС	
ПК-3.1.8 Знает современные стандарты информационного взаимодействия систем	Обучающийся знает: – современные стандарты информационного взаимодействия систем.

3. Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений блока 1 «Дисциплины (модули)».

4. Объем дисциплины и виды учебной работы

Вид учебной работы	Всего часов
Контактная работа (по видам учебных занятий) В том числе:	96
– лекции (Л)	32
– практические занятия (ПЗ)	32
– лабораторные работы (ЛР)	32
Самостоятельная работа (СРС) (всего)	44
Контроль	4
Форма контроля (промежуточной аттестации)	3
Общая трудоемкость: час / з.е.	144

Примечание: «Форма контроля» – зачет (3)

5. Структура и содержание дисциплины

5.1. Разделы дисциплины и содержание рассматриваемых вопросов

№ п/п	Наименование раздела дисциплины	Содержание раздела	Индикаторы достижения компетенций
1	Введение в теорию больших данных	Лекция 1. Введение в теорию больших данных.	ПК-1.2.8 ПК-3.1.8
		Лекция 2. Экономика данных.	
		Лекция 3. Управление данными.	
		Лабораторная работа 1. Проектирование структуры базы данных.	
		Самостоятельная работа. Повторение лекционного материала, подготовка к выполнению лабораторной работы № 1.	
2	Реляционная система управления базами данных Greenplum	Лекция 4. СУБД Greenplum. Принципы построения и функционирования. Основные объекты.	ПК-1.2.8 ПК-3.1.8
		Лекция 5. Улучшение производительности СУБД Greenplum: Мониторинг текущих запросов. План запроса.	
		Лабораторная работа 2. Работа в СУБД Greenplum. Обслуживание СУБД Greenplum. Обеспечение безопасности данных	

		Практическое занятие 1. Порядок инсталляции и настройки Greenplum	
		Практическое занятие 2. Улучшение производительности СУБД Greenplum.	
		Практическое занятие 3. Практическое занятие на отказоустойчивость Greenplum	
		Самостоятельная работа. Повторение лекционного материала, подготовка к выполнению лабораторных работ № 2-3 и практических работ № 1-2.	
3	SQL-запросы	Лекция 6. SQL-запросы для выборки данных: SELECT, LIMIT, OFFSET. Сочетание запросов. Списки VALUES.	ПК-1.2.8 ПК-3.1.8
		Лекция 7. SQL-запросы: DML-операции. Транзакции.	
		Практическое занятие 4. Практическое занятие с SQL запросами.	
		Практическое занятие 5. Практическое занятие с SQL запросами.	
		Практическое занятие 6. Практическое занятие с SQL запросами.	
		Самостоятельная работа. Повторение лекционного материала, подготовка к выполнению практических работ № 4-6.	
4	Методы и модели хранения и управления данными	Лекция 8. Метод хранения данных Datalake. Золотая запись данных.	ПК-1.2.8 ПК-3.1.8
		Лекция 9. Модели управления данными: MDM. Кубы данных.	
		Лекция 10. Методы и модели построения BI-систем.	
		Лабораторная работа 3. Каталог данных (Atlas). Управление метаданными.	
		Лабораторная работа 4. ETL/ELT в управлении данными (Airflow). Управление качеством данных.	
		Практическое занятие 7. Практическая работа с Liquibase	
		Самостоятельная работа. Повторение лекционного материала, подготовка к выполнению лабораторных работ № 5-7 и практической работы № 7.	
5	Программная платформа Quant	Лекция 11. Программная платформа Quant: Структура. Принципы	ПК-1.2.8

		построения. Решаемые задачи.	ПК-3.1.8
		Лекция 12. Модуль ETL/ELT (Airflow). Применение модуля Airflow в Quant.	
		Лекция 13. Модуль Data Catalog (Atlas). Модуль синхронизации. Применение в программной платформе Quant.	
		Лекция 14 Модуль обслуживания БД. Применение модуля в программной платформе Quant.	
		Лекция 15. Модуль ReportServer. Модуль Birt. Применение в программной платформе Quant.	
		Лекция 16. Направления развития Платформы Quant	
		Лабораторная работа 5. Применение Quant DPM в процессе ETL/ELT	
		Лабораторная работа 6. Управление нормативно-справочной информацией	
		Лабораторная работа 7. Работа с BIRT.	
		Лабораторная работа 8. Работа с ReportServer и Apache Superset.	
		Практическое занятие 8. Практическое занятие на формирование витрин данных.	
		Самостоятельная работа. Повторение лекционного материала, подготовка к выполнению лабораторных работ № 8-12 и практической работы №8	

5.2. Разделы дисциплины и виды занятий

№ п/п	Наименование раздела дисциплины	Л	ПЗ	ЛР	СРС	Всего
1	Введение в теорию больших данных	6	-	4	8	18
2	СУБД Greenplum	4	12	4	8	28
3	SQL-запросы	4	12	-	8	24
4	Методы и модели хранения и управления данными	6	4	8	8	26
5	Программная платформа Quant	12	4	16	12	44
	Итого	32	32	32	44	140
Контроль						4
Всего (общая трудоемкость, час.)						144

6. Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

Оценочные материалы по дисциплине является неотъемлемой частью рабочей программы и представлены отдельным документом, рассмотренным на заседании кафедры и утвержденным заведующим кафедрой.

7. Методические указания для обучающихся по освоению дисциплины

Порядок изучения дисциплины следующий:

1. Освоение разделов дисциплины производится в порядке, приведенном в разделе 5 «Содержание и структура дисциплины». Обучающийся должен освоить все разделы дисциплины, используя методические материалы дисциплины, а также учебно-методическое обеспечение, приведенное в разделе 8 рабочей программы.

2. Для формирования компетенций обучающийся должен представить выполненные задания, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, предусмотренные текущим контролем успеваемости (см. оценочные материалы по дисциплине).

3. По итогам текущего контроля успеваемости по дисциплине, обучающийся должен пройти промежуточную аттестацию (см. оценочные материалы по дисциплине).

8. Описание материально-технического и учебно-методического обеспечения, необходимого для реализации образовательной программы по дисциплине

8.1. Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой бакалавриата/специалитета/магистратуры, укомплектованные специализированной учебной мебелью и оснащенные оборудованием и техническими средствами обучения, служащими для представления учебной информации большой аудитории: настенным экраном (стационарным или переносным), маркерной доской и (или) меловой доской, мультимедийным проектором (стационарным или переносным).

Все помещения, используемые для проведения учебных занятий и самостоятельной работы, соответствуют действующим санитарным и противопожарным нормам и правилам.

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

8.2. Университет обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства:

- операционная система Windows и Linux, MS Office;
- Антивирус Касперский;
- Adobe Acrobat Reader DC (бесплатное, свободно распространяемое программное обеспечение; режим доступа <https://get.adobe.com/ru/reader/>);
- Программная платформа Quant.

8.3. Обучающимся обеспечен доступ (удаленный доступ) к современным профессиональным базам данных:

- Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] – Режим доступа: <https://www.elibrary.ru> (свободный доступ);
- Электронно-библиотечная система «Лань» [Электронный ресурс] – Режим доступа: <https://e.lanbook.com> (свободный доступ);
- Электронная библиотечная система «Юрайт» [Электронный ресурс] – Режим

доступа: <https://urait.ru> (свободный доступ).

8.4. Обучающимся обеспечен доступ (удаленный доступ) к информационным справочным системам:

– Кирюшин С., Борисов Е. и др. – Учебник 4CDTO. – Клуб топ-менеджеров СЮ, 2021;

– Макшанов А.В., Тындыкарь Л.Н., Журавлев А.Е. Большие данные: учебник / Издательство: Лань, 2022 г. – 188 с.

8.5. Перечень печатных изданий, используемых в образовательном процессе:

– Ермаков С.Г., Бухарова К.А., Мандрика О. С., Лабузов С.В. «Анализ и обработка больших данных на платформе Quant»: учебное пособие. - СПб.: ПГУПС, 2024.- 121 с.

8.6. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых в образовательном процессе:

– Личный кабинет обучающегося и электронная информационно-образовательная среда [Электронный ресурс]. - Режим доступа: <http://sdo.pgups.ru/> (для доступа к полнотекстовым документам требуется авторизация).

Разработчик рабочей программы,
заведующий кафедрой

С.Г. Ермаков

19 марта 2025 г.