

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

**Департамент бизнес-информатики
Факультета информационных технологий и анализа больших данных**

УТВЕРЖДАЮ

**Проректор по учебной
и методической работе**

_____ **Е.А. Каменева**

11.01. 2023 г.

П.А. Сахнюк, А.А. Соломахин

Управление данными предприятия

**Рабочая программа дисциплины
для студентов, обучающихся по направлению подготовки
38.03.05 Бизнес-информатика,
образовательная программа
«Цифровая трансформация управления бизнесом»**

*Рекомендовано Ученым советом Факультета информационных
технологий и анализа больших данных
(протокол № 27 от 15.12.2022 г.)*

*Одобрено Советом учебно-научного Департамента бизнес-информатики
(протокол № 3 от 08.12.2022 г.)*

Москва 2023

Содержание

1. Наименование дисциплины.....	3
2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине.....	3
3. Место дисциплины в структуре образовательной программы.....	4
4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся.....	4
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	5
5.1. Содержание дисциплины.....	5
5.2. Учебно-тематический план.....	10
5.3. Содержание семинаров, практических занятий.....	11
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине.....	14
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	14
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю.....	17
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	20
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	23
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	24
10. Методические указания для обучающихся по освоению дисциплины.....	25
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	25
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	25

1. Наименование дисциплины

«Управление данными предприятия».

2. Перечень планируемых результатов освоения образовательной программы с указанием индикаторов их достижения, соотнесенных с планируемыми результатами обучения по дисциплине

В результате изучения дисциплины у студентов должны быть сформированы следующие компетенции:

Таблица 1

Компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПКН-1	Способность внедрять транзакционные системы и консультировать по вопросам систем сбора, накопления и хранения транзакционных данных	1. Проводит анализ рынка систем сбора, накопления и хранения транзакционных данных.	Знать: современное состояние рынка систем сбора, накопления и хранения транзакционных данных и перспектив их использования в организации. Уметь: использовать современные технологии анализа рынка систем сбора, накопления и хранения транзакционных данных, обобщения и систематизации информации.
		2. Внедряет системы сбора, накопления и хранения транзакционных данных.	Знать: современные приемы внедрения системы сбора, накопления и хранения транзакционных данных. Уметь: организовывать процесс внедрения системы сбора, накопления и хранения транзакционных данных, используя современные технические средства
		3. Консультирует по вопросам применения систем сбора, накопления и хранения транзакционных данных.	Знать: современное состояние основных транзакционных систем Уметь: проводить консультации по вопросам применения систем сбора, накопления и хранения транзакционных данных.
ПКН-2	Способность анализировать и проектировать информационные потоки организации	1. Анализирует информационные потоки организации.	Знать: понятие информационных потоков организации, критерии классификации информационных потоков хозяйствующих субъектов; Уметь: повышать качество управления, организовать эффективную работу внутрикорпоративных структур, целенаправленно распределять финансовые и материальные ресурсы с целью достижения конкурентного преимущества

Компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
		2. Создают модели «как есть» и «как должно быть» информационных потоков организации.	Знать: модели «как есть» и «как должно быть» информационных потоков организации Уметь: применять основные подходы к моделированию механизмов непосредственного влияния информации на бизнес-процессы организации

3. Место дисциплины в структуре образовательной программы

Дисциплина «Управление данными предприятия» относится к общефакультетскому (предпрофильному) циклу части, формируемой участниками образовательных отношений по направлению подготовки: 38.03.05 Бизнес-информатика, образовательная программа «Цифровая трансформация управления бизнесом», профили: «ИТ-менеджмент в бизнесе», «Технологии цифровых бизнес-моделей».

4. Объем дисциплины в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся (в семестре)

Таблица 2

Вид учебной работы по дисциплине	Всего (в зач. ед. и часах) 2021г./2022г.	Семестр 5*/5** (в часах)	Семестр 6*/6** (в часах)
Общая трудоемкость дисциплины	7 зач. ед. /252 ч.	108	144
Контактная работа - Аудиторные занятия	102	68/50	34/34
<i>Лекции</i>	50/32	34/16	16/16
<i>Практические и семинарские занятия</i>	52/52	34/34	18/18
Самостоятельная работа	150/168	40/58	110/110
Вид текущего контроля	РАР/ контрольная работа	Расчетно-аналитическая работа	Контрольная работа
Вид промежуточной аттестации	зачет, экзамен	зачет	экзамен

**набор 2021 г.; ** набор 2022, 2023 года*

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Характеристики больших данных

Понятие «Big Data». Источники и характеристики больших данных. Экосистема, значение больших данных. Принципы работы с большими данными: горизонтальная масштабируемость, отказоустойчивость, локальность данных. стек (экосистема) инструментов Hadoop. Программная модель MapReduce. Поток данных в MapReduce. Проект Apache Spark. Преимущества Spark и назначение модулей Spark. Облачные технологии обработки больших данных. Озера данных и современные хранилища данных.

Тема 2. Платформа управления данными Ростелекома

Потоковая и пакетная обработки данных (настройка и мониторинг etl/elt-процессов): RT.Streaming на базе распределенного брокера сообщения Apache Kafka и проекта для автоматизации потоков данных Apache NIFI; RT.DataLoader - набор инструментов автоматизации извлечения данных из систем-источников и доставки выгружаемых данных в «озеро».

Формирование озера «сырых» неструктурированных данных: RT.DataLake - фреймворк для обработки и хранения данных с набором утилит и программ из Hadoop-экосистемы (Hadoop, Hive, Spark, Zookeeper, Hbase, Hue, Ranger, Oozie, Tez, Livy, Phoenix, Zeppelin, Sqoop, Flume, Solr, Knox, Flink); RT.KeyValue - NoSQL база данных для высоко нагруженных систем.

Построение реляционной области классического хранилища данных для BI-отчетности и аналитики: RT.Warehouse - массивно-параллельная БД для хранения структурированных данных на базе ПО Greenplum Databases; RT.WideStore - аналитическая база данных для эффективной обработки больших набор данных с высокой скоростью выполнения аналитических функций, но урезанная версия языка запросов на базе ClickHouse.

Управление данными организации с построением единого глоссария и сквозного data lineage: RT.DataGovernance - управление корпоративным глоссарием, описание жизненного цикла данных организации, построение карты данных и единого реестра отчетов в качестве единой точки доступа к корпоративной отчетности; RT.MDM - Классический инструмент централизации и контроля целостности (согласованности) нормативно-справочной информации, обеспечения единых процессов гармонизации и ведения справочников.

Сервисное управление компонентами платформы в paas/on-premises модели: RT.DataSync - Синхронизация БД Greenplum между собой, архивирование и восстановления БД; RT.ClusterManager - Набор необходимых функций для развертывания, управления и мониторинга компонент кластера с помощью визуального интерфейса в режиме «единого окна».

Тема 3. Стек технологий Arenadata

Программные компоненты с открытым исходным кодом корпоративной платформы данных Arenadata Enterprise Data Platform (EDP):

Arenadata DB (ADB) — аналитическая, распределённая СУБД, построенная на MPP-системе с открытым исходным кодом Greenplum, предназначенная для хранения и обработки больших объёмов информации — для создания корпоративного хранилища данных.

Apache Hadoop Arenadata Hadoop (ADH) - полноценный дистрибутив на базе предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.

Arenadata QuickMarts (ADQM) - кластерная колоночная система управления базами данных, созданная на основе ClickHouse.

Arenadata Postgres (ADPG) - коммерческий дистрибутив популярной объектно-реляционной базы данных с открытым кодом PostgreSQL для эффективной работы с нагрузкой разного профиля (в первую очередь OLTP — Online Transaction Processing).

Arenadata Streaming (ADS) — масштабируемая отказоустойчивая система для потоковой обработки данных в режиме реального времени, адаптированная для корпоративного использования и построенная на базе Apache Kafka и Apache Nifi.

Arenadata Cluster Manager - — универсальный оркестратор гибридного ландшафта. Он позволяет быстро устанавливать, настраивать все data-сервисы компании и управлять ими.

Arenadata Catalog – ПО для внедрения практики Data Governance, позволяющее решать задачи по управлению информационными активами компании и ведению корпоративного бизнес-гlossария в едином интерфейсе.

Тема 4. Инструменты и технологии сбора, хранения, обработки и анализа больших данных VK Cloud Solutions

Вычисления: виртуальные серверы, Облако 152-ФЗ, Аренда VPS / VDS, Autoscaling, Cloud GPU

Хранилище: Cloud Storage S3, Cloud Disaster Recovery

Kubernetes: Cloud Containers, Enterprise Kubernetes, Autoscaling

Базы данных: Cloud Databases (MySQL, PostgreSQL, MongoDB, Redis, ClickHouse, Postgres Pro, Arenadata DB, Tarantool Cloud)

Машинное обучение в облаке: Cloud ML Platform, Cloud Voice, Vision

Cloud Big Data: S3 хранилище VK Cloud Storage, in-memory платформа Tarantool, Apache Hadoop, Apache Spark, ClickHouse, Storm, Kafka, Arenadata DB

Тема 5. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Cloud (Sber)

Аналитика данных: Cloud Search Service - распределенная поисковая система на основе Elasticsearch; Data Lake Insight (Apache Spark и Apache Flink) - сервис для анализа больших данных с поддержкой потоковой и пакетной обработки; Data Warehouse Service - надежное хранилище для данных большого объема; DAYU (Data As You Use) - универсальная платформа для работы с данными; Graph Engine Service - облачный сервис для построения запросов и анализа данных графовой структуры; Log Tank Service - облачный сервис для сбора, хранения и анализа логов с виртуальных машин и облачных сервисов; MapReduce Service - сервис для хранения, анализа и обработки больших данных на базе кластеров больших данных корпоративного уровня в облаке, в которых можно запускать компоненты больших данных, такие как Hadoop, Spark, HBase, Kafka и Storm; Document Database Service (MongoDB) - высокодоступный и легко масштабируемый сервис для управления документоориентированной базой данных, полностью совместимый с MongoDB; Database Security Service проводит аудит базы данных, обнаруживает атаки с помощью внедрения SQL-кода и выявляет операции с высоким риском.

Хранение и резервное копирование: Object Storage Service - масштабируемое объектное хранилище, полностью совместимое с S3; Data Warehouse Service - хранилище для данных большого объема; резервное копирование в облако; файловое хранилище NFS; Distributed Cache Service for Memcached; Distributed Cache Service for Redis; Elastic Volume Service - сервис блочного хранилища данных; Scalable File Service Turbo - сервис для управления файловой системой NFS; Cloud Backup and Recovery; резервное копирование VM; Аварийное восстановление DRaaS; Software Repository for Container.

Базы данных: Document Database Service (MongoDB), Data Admin Service - облачный сервис для управления базами данных, Relational Database Service (RDS) for SQL Server — высокодоступный сервис управления реляционными базами данных; RDS for MySQL - платформа управления базами данных MySQL; Distributed Database Middleware - сервис для распределенного хранения данных на нескольких серверах баз данных MySQL; RDS for PostgreSQL - платформа управления базами данных PostgreSQL

AI & ML: Cloud AI Services - сервисы на базе искусственного интеллекта для решения прикладных бизнес-задач; Christofari и Christofari Neo - суперкомпьютеры для любого бизнеса; DataHub ML Space - хаб предобученных моделей, датасетов и контейнеров; AI Services - AI-сервисы для решения бизнес задач, а также интеграции в приложения и веб-сервисы; ML Space - платформа для ML-разработки полного цикла и совместной работы DS-команд; ruGPT-3 & family - инструменты для обработки языка

Тема 6. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Yandex Cloud

Платформа данных: Managed Service for PostgreSQL; Managed Service for MySQL; Managed Service for ClickHouse; Managed Service for YDB - распределённая

отказоустойчивая Distributed SQL СУБД; Managed Service for Redis; Managed Service for MongoDB; Managed Service for Apache Kafka; Managed Service for Greenplum; Managed Service for Elasticsearch; Managed Service for OpenSearch New; Data Proc - управление кластерами Apache Hadoop; Data Transfer; Object Storage - масштабируемое хранилище данных; Message Queue; Data Streams; Yandex Query - бессерверная аналитика данных из Object Storage и потоковых данных; DataSphere - среда ML-разработки полного цикла; DataLens - визуализация и анализ данных; Monitoring - сбор и визуализация метрик

Контейнерная разработка: Managed Service for Kubernetes; Container Registry - управление Docker-образами; Serverless Containers - запуск контейнеров без Kubernetes.

Бессерверные вычисления: Cloud Functions - запуск кода в виде функции; Managed Service for YDB; API Gateway - интеграция с сервисами Yandex Cloud при помощи API; IoT Core - решения для интернета вещей.

Машинное обучение: DataSphere; SpeechKit; Translate; Vision.

Тема 7. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Selectel

ML и обработка данных: Data Science Virtual Machine - виртуальные машины с готовыми к работе приложениями для разработчиков ИИ и аналитиков данных; Data Analytics Virtual Machine - виртуальные серверы с предустановленными инструментами для дата-инженеров, специалистов по Data Science и аналитиков; суперкомпьютер для HPC - HPC-кластер для решения задач ML и AI.

Платформа обработки данных: Платформа для аналитики и работы с Big Data на вычислительных мощностях Selectel; ML-платформа - преднастроенная инфраструктура для ClearML и Kubeflow.

Тема 8. Машинное обучение для анализа больших данных

Машинное обучение для решения задач Data Mining. Градиентный спуск в машинном обучении. Деревья решений. Бэггинг, бустинг, стекинг. Фреймворки машинного обучения. Кластерный анализ, алгоритм k-means. Анализ временных рядов. Автоматическое машинное обучение (AutoML).

Машинное обучение и прогнозное моделирование с помощью BigQuery ML. Создание и применение моделей машинного обучения для структурированных данных большого объема непосредственно в BigQuery, используя простой SQL. Разработка в BigQuery ML следующие типов моделей: линейная регрессия для прогнозирования; бинарная и мультиклассовая логистическая регрессия для классификации; кластеризация K-средних для сегментации данных; прогнозирование временных рядов; Boosted Tree для создания моделей классификации и регрессии на основе XGBoost; feature preprocessing; Hyperparameter tuning; evaluation functions; inference functions; AI Explanation functions.

Создание рабочих процессов инженерной аналитики (трансформации и моделированию данных) с помощью фреймворка dbt; создание надежных наборов данных для отчетов, создания моделей машинного обучения и операционных рабочих процессов.

Тема 9. Распределенные вычисления Apache Spark

Big Data фреймворк Apache Spark с открытым исходным кодом для распределённой пакетной и потоковой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop. Анализ больших данных с использованием фреймворка Spark. Spark SQL модуль Apache Spark для работы со структурированными данными. Многофункциональный веб-блокнот Apache Zeppelin, обеспечивающий считывание, анализ и визуализацию данных, их обмен и взаимодействие с Hadoop и Spark. Эмуляция распределенных вычислений Spark в Google Colab.

Тема 10. Аналитика данных в SQLite

Библиотека на языке C SQLite, реализующая небольшой, быстрый, автономный, высоконадежный, полнофункциональный механизм базы данных SQL. Аналитика данных в Google Colab базы данных SQLite, подключение к базе данных SQLite из SQL-клиента (DBeaver).

Тема 11. OLAP in-memory database DuckDB

Аналитика данных в OLAP in-memory database DuckDB. Обработка и хранение наборов табличных данных, из файлов CSV или Parquet. Интерактивный анализ данных, объединение нескольких больших таблиц. Применение JupyterSQL и DuckDB для эффективной визуализации наборов данных, превышающих объем памяти.

Тема 12. Унифицированный интерфейс для распределенных вычислений Fugue

Использование FugueSQL для определения сквозных рабочих процессов поверх Pandas, Spark и Dask DataFrames. Аналитика больших объемов структурированных данных с использованием DuckDB и Fugue.

Тема 13. Apache Superset

Применение платформы Apache Superset для визуального анализа и исследования данных. Подключение Superset к источникам данных на основе SQL через SQLAlchemy, включая современные облачные корпоративные базы и хранилища данных. Интеграция данных и их исследование с использованием простого конструктора визуализаций без кода, либо интегрированной средой разработки SQL.

Современная корпоративная веб-ориентированная платформа для бизнес-аналитики RT.DataVision (на Apache Superset, Celery, Redis DB, PostgreSQL DB), позволяющая разрабатывать корпоративную отчетность различного уровня сложности, от простых таблиц до комплексных дашбордов и представлений данных с импортом в файлы различных форматов.

5.2. Учебно-тематический план

Таблица 4

№п/п	Наименование темы (раздела) дисциплины	Трудоемкость в часах (2021 г./2022 г., 2023)					Формы текущего контроля успеваемости
		Всего	Контактная работа-Аудиторная работа			Самостоятельная работа	
			Общая	Лекции	Практические и семинарские занятия		
5 семестр							
1	Тема 1. Характеристики больших данных	12/10	8/6	4/2	4/4	4/4	Выполнение индивидуальных заданий
2	Тема 2. Платформа управления данными Ростелекома	14/10	8/4	4/-	4/4	6/6	Выполнение индивидуальных заданий
3	Тема 3. Стек технологий Arenadata	14/12	8/6	4/2	4/4	6/6	Выполнение индивидуальных заданий
4	Тема 4. Инструменты и технологии сбора, хранения, обработки и анализа больших данных VK Cloud Solutions	10/10	6/4	4/2	2/2	4/6	Выполнение индивидуальных заданий
5	Тема 5. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Cloud (Sber)	10/10	6/4	4/2	2/2	4/6	Выполнение индивидуальных заданий
6	Тема 6. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Yandex Cloud.	8/12	4/4	2/2	2/2	4/8	Выполнение индивидуальных заданий
7	Тема 7. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Selectel.	8/12	4/4	2/2	2/2	4/8	Выполнение индивидуальных заданий
8	Тема 8. Машинное обучение для анализа больших данных	32/32	24/18	10/4	14/14	8/14	Выполнение индивидуальных заданий
	Итого семестр 5	108/108	68/50	34/16	34/34	40/58	
6 семестр							
9	Тема 9. Распределенные вычисления Apache Spark	28/28	8/8	4/4	4/4	20/20	Выполнение индивидуальных заданий

10	Тема 10. Аналитика данных в SQLite	28/28	4/4	2/2	2/2	24/24	Выполнение индивидуальных заданий
11	Тема 11. OLAP in-memory database DuckDB	32/32	8/8	4/4	4/4	24/24	Выполнение индивидуальных заданий
12	Тема 12. Унифицированный интерфейс для распределенных вычислений Fugue.	30/30	4/4	2/2	2/2	26/26	Выполнение индивидуальных заданий
13	Тема 13. Apache Superset	26/26	10/10	4/4	6/6	16/16	Выполнение индивидуальных заданий
	Итого семестр 6	144/144	34/34	16/16	18/18	110/110	
	В целом по дисциплине	252/252	102/84	50/32	52/52	150/168	РАР/контрольная работа
	Итого в %		40/33	49/38	51/62	60/67	

5.3. Содержание практических и семинарских занятий

Таблица 4

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Тема 1. Характеристики больших данных	1. Типы данных, функции и операторы SQL в Google BigQuery. 2. Создание проекта, импорт данных в Google BigQuery. 3. Интерфейс редактора запросов Google BigQuery 4. Добавление публичных датасетов Google в свой проект. 5. Простые запросы SQL. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 2. Платформа управления данными Ростелекома	1.Создание кластера RT.DataLake. 2. Добавление хостов в кластер. 3.Запуск установки. 4. Работа со статусами и логами установки. 5. Проверка состояния кластера. 6. Изменение репликации. 7.Операции обслуживания кластера. 8. Аналитика SQL: Агрегирование, оконные функции. 9. Классы оконных функций: агрегирующие (Aggregate), ранжирующие (Ranking), функции смещения (Value). Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 3. Стек технологий Arenadata	1. Агрегирование. Операции соединения таблиц, вложенные запросы в качестве таблиц, условий фильтрации, полей. 2. Рекурсивные запросы (конструкция WITH). Соединение запросов, как множеств (UNION, INTERSECT, EXCEPT). Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий.

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Тема 4. Инструменты и технологии сбора, хранения, обработки и анализа больших данных VK Cloud Solutions	1. Применение математических XYZ-анализа 2. Применение математических и оконных функций для ABC-анализа Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 5. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Cloud (Sber)	Многомерный ABC-XYZ анализ на SQL Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 6. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Yandex Cloud.	Клиентская аналитика SQL: RFM-анализ на больших объемах структурированных данных. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 7. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Selectel.	Клиентская аналитика SQL: когортный анализ на больших объемах структурированных данных. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 8. Машинное обучение для анализа больших данных	Создание, обучение, оценка, объяснение модели машинного обучения, создание прогнозов на новых данных: 1. Создание регрессионной модели машинного обучения средствами SQL в Google BigQuery 2. Разработка модели бинарной классификации средствами SQL в Google BigQuery 3. Применение ансамблей деревьев решений для решения задач классификации и регрессии в BigQuery (использование фреймворка XGBoost) 4. Разработка модели кластеризации средствами SQL в Google BigQuery 5. Разработка модели анализа временных рядов средствами SQL в Google BigQuery. Создание рабочих процессов инженерной аналитики (трансформации и моделированию данных) с помощью фреймворка dbt. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 9. Распределенные вычисления Apache Spark	Применение фреймворка Spark для анализа и машинного обучения на больших данных (модуль pyspark.sql): - Apache Zeppelin (интерактивный блокнот) для анализа и визуализации данных, а также совместной работы над данными с использованием средств Apache Spark, обеспечивающий считывание, анализ и визуализацию данных, их обмен и взаимодействие с Hadoop и Spark. - Эмуляция распределенных вычислений Spark в Google Colab.	Выполнение индивидуальных заданий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
	Основная литература: Дополнительная литература:	
Тема 10. Аналитика данных в SQLite	1. Аналитика данных в Google Colab базы данных SQLite. 2. Подключение к базе данных SQLite из SQL-клиента (DBeaver). Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 11. OLAP in-memory database DuckDB	1. Аналитика данных в OLAP in-memory database DuckDB. 2. Обработка и хранение наборов табличных данных, из файлов CSV или Parquet. 3. Интерактивный анализ данных, объединение нескольких больших таблиц. 3. Применение JupyterSQL и DuckDB для эффективной визуализации наборов данных, превышающих объем памяти. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 12. Унифицированный интерфейс для распределенных вычислений Fugue.	1. Использование FugueSQL для определения сквозных рабочих процессов поверх Pandas, Spark и Dask DataFrames. 2. Аналитика больших объемов структурированных данных с использованием DuckDB и Fugue. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий
Тема 13. Apache Superset	1. Применение платформы аналитики RT.DataVision (на Apache Superset, Celery, Redis DB, PostgreSQL DB) для визуального анализа и исследования данных. 2. Подключение Superset к источникам данных на основе SQL через SQLAlchemy, включая современные облачные корпоративные базы и хранилища данных. 3. Интеграция данных и их исследование с использованием конструктора визуализаций без кода, и с использованием интегрированной средой разработки SQL. 4. Разработка корпоративной отчетности различного уровня сложности, от простых таблиц до комплексных дашбордов и представлений данных с импортом в файлы различных форматов. Основная литература: Дополнительная литература:	Выполнение индивидуальных заданий

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Таблица 5

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Тема 1. Характеристики больших данных	Источники и характеристики больших данных. Экосистема, значение больших данных. Какие используются инструменты и компоненты решения для работы с большими данными.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 2. Платформа управления данными Ростелекома	Функциональная роль платформы управления данными в ИТ-архитектуре. Решения на базе платформы: - Аналитика данных и отчетность. - Клиентская аналитика. - Интернет вещей (IoT). - Security, Risk, & Compliance. Решения по отраслям экономики.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 3. Стек технологий Arenadata	Применение платформы Arenadata для различных секторов бизнеса: - Импортозамещение иностранных СУБД - Транспортные компании - Телеком - Промышленность - Интернет-компании - Добыча сырья - Государственным заказчикам - Банки и страхование - Ритейл и e-commerce	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 4. Инструменты и технологии сбора, хранения, обработки и анализа больших данных VK Cloud Solutions	Решения на базе платформы VK Cloud: - Образование - Медиахостинг - Медицина - Логистика - Разработка - E-commerce	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 5. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Cloud (Sber)	Решения по отраслям на базе платформы Cloud: - Ритейл - IT-разработка - Транспорт и логистика - Образование - Медиа и реклама	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников.

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
	- Энергетика - Здравоохранение - Государство Решения по задачам: - Чат-боты на Serverless - Сайт в облаке - 1С в облаке - Разработка и тестирование	Выполнение сквозного индивидуального задания
Тема 6. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Yandex Cloud.	Решения Yandex Cloud по отраслям: Розничная торговля; E-Commerce; Автомобильная промышленность; Цифровые отрасли; Финансы и страхование; Энергетика; Здравоохранение; Образование; Малый и средний бизнес; Фармацевтика; Цифровизация промышленных предприятий. По типу задач: - Сайт в облаке - Интернет-магазин - Хранение и обработка персональных данных - 1С в облаке - Сервисы Microsoft в облаке - Автоматизация колл-центров - Рекомендательная система для ритейла и e-commerce - Корпоративное хранилище данных - Бизнес-аналитика - Serverless - Чат-боты на Serverless - Микросервисная архитектура - Data Science в облаке - Yandex Distributed Cloud - Yandex Cloud Security Solution Library - Миграция в облако	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 7. Инструменты и технологии сбора, хранения, обработки и анализа больших данных Selectel.	Решения Selectel по бизнес-задачам: - Миграция в облако - Оптимизация IT-затрат - Соответствие 152-ФЗ - Отказоустойчивая IT-инфраструктура - Информационная безопасность - Анализ данных - Приватные решения в публичном облаке	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 8. Машинное обучение для анализа больших данных	Машинное обучение в BigQuery: - Формулировка задачи машинного обучения - Типы задач машинного обучения - Построение регрессионной модели - Выбор метки - Выбор признаков в наборе данных	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников.

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
	- Создание обучающего набора данных - Обучение и оценка модели - Получение прогнозов с помощью модели - Исследование весов модели - Создание модели классификации - Обучение - Оценка - Прогнозирование - Выбор порога - Настройка механизма машинного обучения в BigQuery - Управление разделением данных - Балансировка классов - Регуляризация	Выполнение сквозного индивидуального задания
Тема 9. Распределенные вычисления Apache Spark	Типичные сценарии обработки больших данных Apache Spark: - Извлечение, преобразование и загрузка (ETL). - Обработка потоков данных в реальном времени. - Пакетная обработка. - Машинное обучение с использованием MLlib. - Обработка графов с помощью GraphX. - Обработка SQL и структурированных данных с помощью Spark SQL.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 10. Аналитика данных в SQLite	- Преимущества SQLite. - Недостатки SQLite. - Применение SQLite. - консольная утилита SQLite3 для работы с SQLite.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 11. OLAP in-memory database DuckDB	- Преимущества DuckDB. - Примеры использования DuckDB. - DuckDB в действии с Python. - Ограничения DuckDB.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания
Тема 12. Унифицированный интерфейс для распределенных вычислений Fugue.	- Несовместимость между Pandas и Spark. - Снижение дорогостоящих ошибок. - Интерфейс SQL. - Fugue vs Koalas vs Modin.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Тема 13. Apache Superset	- Администрирование платформы. - Методы авторизации пользователей. Методы разграничения доступов. Управление ролями. - Настройка RLS. - Встраивание контента RT.DataVision во внешние сервисы.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников. Выполнение сквозного индивидуального задания

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные тестовые задания для текущего контроля:

1.

Какие из следующих технологий СУБД не используют принцип MapReduce

Hadoop
Cassandra
Redis
HDI Insight

2.

Какие вероятные разочарования тренда больших данных?
из-за угрозы безопасности личной жизни (privacy) граждан будут упрощены процедуры сбора данных, что приведёт к падению ценности больших данных
из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных
Нет

3.

Выберите технологию потоковой обработки событий в режиме реального времени
MapReduce
Apache Hadoop
Apache Kafka
Spark Streaming

4.

Для полнотекстового интеллектуального поиска и аналитики по полнотекстовым данным в формате JSON отлично подходит СУБД
HBase
Cassandra
Elasticsearch
Hive

5.

Формат Parquet считается
колоночным (столбцовым)
неструктурированным

полуструктурированным
строковым

6.

Для машинного обучения подходят данные

Бинарные

Числовые типа int

Любых форматов в цифровом виде

Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов

7.

Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит

Apache Kafka

Apache AirFlow

Apache Hadoop

Apache Spark

8.

Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью

Apache AirFlow

Apache Kafka

Apache Hive

Apache Hadoop

9.

Apache NiFi используется для
эффективного хранения больших данных
визуализации результатов аналитики
маршрутизации потоков Big Data и построения ETL-конвейеров
оптимизации SQL-запросов к DWH

10.

Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов

Можно

Нельзя

Задание на расчетно-аналитическую работу

Студенту в начале семестра предлагается выбрать кейс, на основе которого разрабатываться аналитический проект. Для этого можно использовать датасеты с Kaggle <https://www.kaggle.com/datasets?search=customer>, содержащие клиентские данные (всего 2466 различных Datasets).

Все практические занятия и расчетно-аналитическая работа выполняются для выбранного варианта.

Задание:

1. Импортировать данные в BigQuery. Обогащать датасет, проводя клиентскую аналитику: выполнить ABC-XYZ, RFM, когортный -анализ.

2. Создать, обучить, оценить по метрикам качества, объяснить модели машинного обучения используя SQL; создать интерактивный отчет (дашборд) в Looker Studio.

Примерные темы расчетно-аналитической работы:

1. Customer Personality Analysis.
2. Telco Customer Churn.
3. Shop Customer Data.
4. Customer Segmentation.
5. Mall Customer Segmentation Data.
6. Credit Card customers.
7. Starbucks Customer Survey.
8. Airlines Customer satisfaction.
9. Loan Prediction Based on Customer Behavior.
10. Brazilian E-Commerce Public Dataset by Olist.

Задание на контрольную работу

На данных кейса предыдущего семестра:

1. В Apache Zeppelin или Colab провести исследовательский анализ данных на Spark, создать и оценить модель машинного обучения.
2. Провести ABC-XYZ, RFM, когортный -анализ в OLAP in-memory database DuckDB.
3. Создать несколько дашбордов в RT.DataVision с использованием интегрированной средой разработки SQL, выявить инсайты, настроить встраивание контента RT.DataVision во внешние сервисы.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине содержится в разделе 2.

Типовые контрольные задания или иные материалы, необходимые для оценки умений, знаний

Таблица 6

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотношенные с индикаторами достижения компетенции	Типовые контрольные задания
ПKN-1	1. Проводит анализ рынка систем сбора, накопления и хранения транзакционных данных.	Знать: современное состояние рынка систем сбора, накопления и хранения транзакционных данных и перспектив их использования в организации. Уметь: использовать современные технологии анализа рынка систем сбора, накопления и хранения транзакционных данных, обобщения и систематизации информации.	Задание 1 Проведите анализ текущего состояния рынка систем сбора, накопления и хранения транзакционных данных, рассматривая основные компании на российском рынке. Охарактеризуйте тенденции и направления развития рынка: интеграция с BI-системами и с инструментами машинного обучения, разработка инструментов предиктивной аналитики. Задание 2 Проведите анализ текущего состояния мирового рынка систем сбора, накопления и хранения транзакционных данных. Охарактеризуйте тенденции и направления развития рынка: интеграция с BI-системами и с инструментами машинного обучения, разработка инструментов предиктивной аналитики.
	2. Внедряет системы сбора, накопления и хранения транзакционных данных.	Знать: современные приемы внедрения системы сбора, накопления и хранения транзакционных данных. Уметь: организовывать процесс внедрения системы сбора, накопления и хранения транзакционных данных, используя современные технические средства	Задание 1 Организуите облачное развертывание системы сбора, накопления и хранения транзакционных данных, используя современные технические средства стека технологий Arenadata Задание 2 Организуите облачное развертывание системы сбора, накопления и хранения транзакционных данных, используя современные технические средства Yandex Cloud
	3. Консультирует по вопро-	Знать: современное состояние основных транзакционных систем	Задание 1 Для выбранного предприятия опреде-

	сам применения систем сбора, накопления и хранения транзакционных данных.	Уметь: проводить консультации по вопросам применения систем сбора, накопления и хранения транзакционных данных.	лите области, в которых могут быть использованы систем сбора, накопления и хранения транзакционных данных. Укажите возможные направления их внедрения и эффект от их использования. Задание 2 Для выбранного предприятия проведите консультацию заказчика по вопросам применения систем сбора, накопления и хранения транзакционных данных.
ПКН-2	1. Анализирует информационные потоки организации.	Знать: понятие информационных потоков организации, критерии классификации информационных потоков хозяйствующих субъектов; Уметь: повышать качество управления, организовать эффективную работу внутрикорпоративных структур, целенаправленно распределять финансовые и материальные ресурсы с целью достижения конкурентного преимущества	Задание 1 На основе анализа информационных потоков организации, предложите различные варианты разработки хранилища данных «под ключ» и создания рекомендательных сервисов для оптимизации производственных процессов. Задание 2 В целях обеспечения импортозамещения технологий обработки и хранения данных, внедрения процесса управления данными в организации, обеспечения горизонтального масштабирования, повышения качества управления, организации эффективной работы внутрикорпоративных структур, предложите различные варианты развертывания платформы управления данными у ведущих облачных провайдеров, необходимых для построения, создания масштабных озер данных и реализации алгоритмов искусственного интеллекта на базе моделей Data Science.
	2. Создают модели «как есть» и «как должно быть» информационных потоков организации.	Знать: модели «как есть» и «как должно быть» информационных потоков организации Уметь: применять основные подходы к моделированию механизмов непосредственного влияния информации на бизнес-процессы организации	Задание 1 Для выбранного предприятия выявите необходимых изменений в информационных потоках организации, определите мероприятия адаптации ИТ-инфраструктуры для внедрения платформы управления данными. Разработайте предложения для заказчика по выбору направлений изменений ИТ-ландшафта предприятия. Задание 2 Для выбранного предприятия определите необходимый состав компонентов платформы управления данными, проведите проектирование целевой архитектуры информационных потоков организации. Разработайте предложения для заказчика по выбору направлений изменений ИТ-ландшафта предприятия

Примерные вопросы для подготовки к зачету:

1. Охарактеризуйте понятие «Big Data», источники и характеристики больших данных.
2. Охарактеризуйте принципы работы с большими данными: горизонтальная масштабируемость, отказоустойчивость, локальность данных.
3. Охарактеризуйте стек (экосистему) инструментов Hadoop.
4. Охарактеризуйте программную модель MapReduce, поток данных в MapReduce.
5. Охарактеризуйте проект Apache Spark, преимущества Spark и назначение модулей Spark.
6. Охарактеризуйте облачные технологии обработки больших данных.
7. Охарактеризуйте отличия озера данных и хранилища данных.
8. Охарактеризуйте платформа управления данными Ростелекома.
9. Охарактеризуйте стек технологий Arenadata.
10. Охарактеризуйте инструменты и технологии сбора, хранения, обработки и анализа больших данных VK Cloud Solutions.
11. Охарактеризуйте инструменты и технологии сбора, хранения, обработки и анализа больших данных Cloud (Sber).
12. Охарактеризуйте инструменты и технологии сбора, хранения, обработки и анализа больших данных Yandex Cloud.
13. Охарактеризуйте инструменты и технологии сбора, хранения, обработки и анализа больших данных Selectel.
14. Охарактеризуйте возможности машинного обучения для анализа больших данных в Google BigQuery.

Примерные вопросы для подготовки к экзамену:

1. Охарактеризуйте источники больших данных: веб-сайты и мобильные приложения, корпоративные приложения, IoT.
2. Охарактеризуйте экосистему Hadoop
3. Охарактеризуйте ключевые возможности HDFS
4. Охарактеризуйте технологию Map Reduce
5. Охарактеризуйте ключевые возможности Hive
6. Охарактеризуйте ключевые возможности Impala
7. Охарактеризуйте ключевые возможности Apache Spark
8. Охарактеризуйте Apache Spark MLlib
9. Охарактеризуйте ключевые возможности Apache Kafka
10. Охарактеризуйте ключевые возможности Apache Presto и HBase
11. Охарактеризуйте ключевые возможности Apache Flink
12. Охарактеризуйте ключевые возможности Apache Beam
13. Охарактеризуйте понятие и назначение Data Lakes
14. Охарактеризуйте ключевые возможности Amazon Web Services S3 (Simple Storage Service)
15. Охарактеризуйте ключевые возможности Google BigQuery
16. Охарактеризуйте ключевые возможности DataLens для аналитики Big Data

17. Охарактеризуйте ключевые возможности Google Data Studio для аналитики Big Data
18. Охарактеризуйте ключевые возможности продукты AutoML (на примере любого облачного провайдера)
19. Охарактеризуйте ключевые возможности аналитики Big Data в публичном облаке VK Cloud Solutions
20. Охарактеризуйте ключевые возможности ClickHouse Database
21. Охарактеризуйте ключевые возможности Greenplum Database
22. Охарактеризуйте возможности сервиса Яндекс.Облака Yandex Data Proc
23. Охарактеризуйте ключевые возможности Yandex Data Science Virtual Machine
24. Охарактеризуйте ключевые возможности MapReduce Service Sbercloud
25. Охарактеризуйте ключевые возможности Data Lake HDInsight Sbercloud
26. Охарактеризуйте ключевые возможности Dataproc Sbercloud
27. Охарактеризуйте ключевые возможности ML Spase Sbercloud
28. Data Mining и основные типы.
29. Охарактеризуйте метрики качества модели классификации.
30. Охарактеризуйте метрики качества задачи кластерного анализа.
31. Охарактеризуйте метрики качества задачи регрессии.
32. Оценка качества обучения модели.

Образец экзаменационного билета

Кейс 60 баллов

1. Охарактеризуйте ключевые возможности ML Spase Sbercloud.
По индивидуально заданию (кейс №...):
2. импортировать данные в BigQuery. Обогащать датасет, проводя клиентскую аналитику: выполнить ABC-XYZ, RFM, когортный -анализ.
3. Создать несколько дашбордов в RT.DataVision с использованием интегрированной средой разработки SQL, выявить инсайты, настроить встраивание контента RT.DataVision во внешние сервисы.

*Методические материалы, определяющие процедуры оценивания
знаний, умений*

Приказ от 23.03.2017 №0557/о «Об утверждении Положения о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по программам бакалавриата и магистратуры в Финансовом университете».

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Нормативно-правовые акты

1. Национальный проект «Цифровая экономика Российской Федерации»
<https://digital.ac.gov.ru/>
2. Федеральный закон от 27 июля 2006 г. N 149-ФЗ «Об информации, информационных технологиях и о защите информации».

3. Федеральный закон от 27.07.2006 №152-ФЗ «О персональных данных».
4. Государственная программа Российской Федерации "Информационное общество" (в ред. [Постановления](#) Правительства РФ от 29.03.2019 N 356-24).
5. Стратегия развития информационного общества в Российской Федерации, на 2017 – 2030 годы. Указ Президента РФ от 9 мая 2017 г. № 203.

а) основная:

1. Кондрашов, Ю. Н. Анализ данных и машинное обучение на платформе MS SQL Server: учебное пособие / Ю. Н. Кондрашов. — Москва : Русайнс, 2020. — 303 с. — Текст : непосредственный. — То же 2023. — ЭБС BOOK.ru. — URL: <https://book.ru/book/947076> (дата обращения: 12.05.2023). — Текст : электронный.
2. Макшанов, А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2022. — 188 с. - ЭБС Лань. — URL: <https://e.lanbook.com/book/198599> (дата обращения: 12.05.2023). — Текст : электронный.

б) дополнительная:

3. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения: Практическое пособие / С. Рашка; пер.с англ. А. В. Логунова. – Москва : ДМК Пресс, 2017. - 418 с. — ЭБС ZNANIUM.com. — URL: <http://znanium.com/go.php?id=1027758> (дата обращения: 12.05.2023). - Текст : электронный.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium <http://www.znanium.com>
5. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
6. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
7. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
8. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
9. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
11. Национальная электронная библиотека <http://нэб.рф/>

12. <https://cloud.google.com/bigquery/docs/sandbox> – страница BigQuery sandbox.
13. <https://www.anaconda.com/> – страница загрузки Anaconda.
14. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндекс.
15. Портал открытых данных Российской Федерации <https://data.gov.ru>
16. <https://cloud.yandex.ru/training/corpplatform> - практический курс «Построение корпоративной аналитической платформы»
17. <https://practicum.yandex.ru/ycloud/> - бесплатный курс «Инженер облачных сервисов»
18. [https://rise.articulate.com/share/BtQjK0gEy1lktRKR6q2hPZ5KnRDJhB8k#/?](https://rise.articulate.com/share/BtQjK0gEy1lktRKR6q2hPZ5KnRDJhB8k#/) - Ростелеком бизнес. Платформа управления данными

10. Методические указания для обучающихся по освоению дисциплины

Студентам необходимо руководствоваться «Методическими рекомендациями по планированию и организации внеаудиторной самостоятельной работы по образовательным программам бакалавриата и магистратуры в Финансовом университете» (Приказ ректора № 1040_о от 11.05.2021) и данной рабочей программой дисциплины.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект лицензионного программного обеспечения:

1. ОС Astra Linux
2. Компьютерные программы общего назначения Windows, Microsoft Office
3. Антивирус ESET Endpoint Security

11.2 Современные профессиональные демонстрационные и информационные справочные системы:

Не предусмотрены.

11.3 Сертифицированные программные и аппаратные средства защиты информации:

Не предусмотрены.

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

Помещения для проведения лекций, семинарских занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.