

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Порохня Андрей Александрович
Должность: И.о. директора института перспективной инженерии
Дата подписания: 19.06.2026 16:00:03
Уникальный программный ключ:
990bea3aeec48c23bd8699e48288f8d1960c600

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**
**Федеральное государственное автономное образовательное учреждение
высшего образования**
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

И.о. директора института
перспективной инженерии,
канд. техн. наук, доцент
А.А. Порохня

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения текущего контроля успеваемости и промежуточной аттестации по
дисциплине

Инжиниринг данных

Направление подготовки	<u>09.03.04 Программная инженерия</u>
Направленность (профиль)	<u>Разработка и сопровождение программного обеспечения</u>
Форма обучения	очная
Год начала обучения	2026
Реализуется в	6 семестре

Введение

1. Назначение: Оценочные материалы (оценочные средства) предназначены для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся, и представляют собой совокупность контрольно-измерительных материалов и методов их использования для измерения уровня достижения обучающимися установленных результатов обучения.

2. ФОС является приложением к программе дисциплины «Инжиниринг данных».

3. Разработчик Потапов И.Р., ассистент межинститутской базовой кафедры департамента цифровых, робототехнических систем и электроники

4. Проведена экспертиза ФОС.

Члены экспертной группы:

Председатель: Новикова Елена Николаевна, зав. межинститутской базовой кафедрой

(ФИО, должность)

Члены комиссии:

Николаев Евгений Иванович, доцент департамента цифровых и робототехнических систем и электроники.

(ФИО, должность)

Винокурский Д. Л., доцент департамента цифровых, робототехнических систем и электроники

(ФИО, должность)

Представитель организации-работодателя: А. В. Новиков, начальник отдела АСУТП ООО «ЕНДС-Ставрополь»

(ФИО, должность)

Экспертное заключение: Оценочные материалы (оценочные средства) по дисциплине «Паттерны проектирования» составлены в соответствии с требованиями самостоятельно устанавливаемого Федерального государственного образовательного стандарта высшего образования по направлению 09.03.04 «Программная инженерия» и позволяют оценить уровень сформированности компетенции ПК-3, ПК-5, ПК-7.

5. Срок действия ФОС определяется сроком реализации образовательной программы.

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

Уровни сформированности компетенци(ий), индикатора (ов)	Дескрипторы			
	Минимальный уровень не достигнут (Неудовлетворительно) 2 балла	Минимальный уровень (удовлетворительно) 3 балла	Средний уровень (хорошо) 4 балла	Высокий уровень (отлично) 5 баллов
Компетенция: ПК-3 – Владеет навыками разработки программного обеспечения: способен реализовывать алгоритмы, писать код на современных языках программирования, создавать веб- и desktop-приложения, работать с базами данных				
Результаты обучения по дисциплине (модулю): <i>Индикатор:</i> ПК-3ид-1 – Разрабатывает алгоритмы и реализует их на современных языках программирования (C++, C#, Python, JavaScript/TypeScript, Go) с использованием эффективных структур данных и стандартных библиотек	Не владеет Python на уровне реализации ETL. Не может написать код для извлечения данных из API или файлов. Не знает Pandas, SQLAlchemy, Requests.	Пишет простые скрипты на Python для извлечения данных из одного источника (например, CSV → Pandas DataFrame). Использует Requests для GET-запросов без обработки ошибок. Знает основы SQLAlchemy (подключение, выполнение простого SELECT). Код неструктурирован, отсутствует обработка исключений.	Реализует полноценный ETL-пайплайн: извлечение из REST API (с пагинацией, повторными попытками), трансформацию с помощью Pandas (фильтрация, группировка, merge), загрузку в базу данных через SQLAlchemy (insert, upsert). Использует параметризацию запросов, логгирование. Понимает эффективность векторизованных операций Pandas.	Разрабатывает модульные ETL/ELT-процессы с учётом инкрементальной загрузки, идиоматичности. Оптимизирует производительность: использует типы данных с низким потреблением памяти в Pandas, применяет chunked-загрузку, async/await для Requests. Пишет тесты для каждого этапа. Настраивает пулы соединений SQLAlchemy. Демонстрирует знание альтернатив (Dask, Polars) для больших объёмов.
Компетенция: ПК-5 – Способен выполнять проектирование и разработку архитектуры программных систем среднего и крупного масштаба с применением современных паттернов и методов проектирования				
Результаты обучения по дисциплине (модулю): <i>Индикатор:</i> ПК-5ид-1 – Контейнеризует приложения с помощью Docker	Не знаком с Docker. Не может написать Dockerfile. Не запускал контейнеры.	Создаёт простой Dockerfile (FROM python, COPY . . , RUN pip install). Собирает образ и запускает контейнер вручную. Может	Упаковывает ETL-сервис с зависимостями, использует .dockerignore. Применяет multi-stage build для разделения сборки и	Создаёт production-готовые образы: минимизирует количество слоёв, использует alpine/slim, настраивает non-root

(Dockerfile, multi-stage builds) и оркестрирует многоконтейнерные стеки через Docker Compose		написать docker-compose.yml с одним сервисом. Не использует multi-stage builds и не оптимизирует размер образа.	выполнения. Пишет docker-compose для стека: ETL-сервис, PostgreSQL, Redis (например, для кэширования). Настраивает зависимости (depends_on) и переменные окружения.	пользователя. В docker-compose описывает несколько ETL-сервисов (извлечение, трансформация, загрузка) с общими сетями и томами. Настраивает healthcheck, restart policies, логирование. Интегрирует docker-compose в CI/CD.
Результаты обучения по дисциплине (модулю): <i>Индикатор:</i> ПК-5ид-2 – Разворачивает и настраивает инфраструктуру на облачных платформах (VM, Managed DB, Object Storage, K8s) с использованием IaC (Terraform, Ansible)	Не имеет опыта работы с облачными платформами и IaC. Не знает, что такое S3-совместимое хранилище.	Вручную создаёт виртуальную машину в облачной консоли. Знает про S3 (объектное хранилище), но не использовал для пайплайнов. Может написать простой Terraform-файл для создания одного ресурса (например, VM) по готовому шаблону.	Разворачивает инфраструктуру под ETL-пайплайн в Yandex Cloud с помощью Terraform: VM, Managed PostgreSQL, bucket. Настраивает доступ к S3-хранилищу из ETL-сервиса (чтение/запись parquet). Использует ansible для установки Docker и запуска контейнеров на VM.	Полностью автоматизирует развёртывание через Terraform (модули, remote state, переменные окружения). Настраивает serverless-опции (например, Object Storage + Cloud Functions для триггеров). Обеспечивает безопасность: IAM-роли, secrets manager. Реализует отказоустойчивость: резервное копирование bucket, multi-AZ для БД. Пишет документацию по воспроизведению инфраструктуры.
Компетенция: ПК-7 – Способен применять современные информационные технологии, инструменты и платформы (включая отечественные), использовать DevOps-практики, облачные сервисы и средства автоматизации				

Результаты обучения по дисциплине (модулю): <i>Индикатор:</i> ПК-7_{ид-1} – Использует современные IDE, линтеры, формatters и инструменты статического анализа (SonarQube) для обеспечения качества кода	Не использует линтеры. Не слышал о SonarQube. Не знает Apache Airflow.	Устанавливает и запускает линтер (например, flake8) вручную. Пишет простой DAG с одним PythonOperator. Не использует сенсоры или SQLExecuteQueryOperator. Не настраивает статический анализ.	Настраивает pre-commit хуки с линтерами (black, isort, pylint) и форматтерами. Интегрирует SonarQube для анализа кода DAG-файлов (качество, дублирование, сложность). Создает DAG с сенсорами (FileSensor, S3KeySensor) и SQLExecuteQueryOperator для выполнения запросов в БД.	Автоматизирует проверку качества DAG в CI: запуск линтеров, статического анализа SonarQube с порогом качества. Проектирует DAG с динамическими задачами, XCom, пользовательскими операторами. Использует сенсоры с тайм-аутами и повторными попытками. Настраивает параллелизм, пулы, обратные вызовы для алертов. Документирует DAG в соответствии с best practices Airflow.
Результаты обучения по дисциплине (модулю): <i>Индикатор:</i> ПК-7_{ид-2} – Настраивает мониторинг, логирование и алертинг (Prometheus, Grafana, ELK/Loki, Sentry) для отслеживания состояния ПО	Не знаком с Prometheus/Grafana. Не использует CI/CD.	Запускает готовый стек Prometheus + Grafana для наблюдения за одним сервисом. Смотрит логи через консоль. Знает GitHub Actions в теории, но не настраивал. Тесты пайплайнов запускает вручную.	Настраивает экспорт метрик для ETL-сервиса (количество обработанных строк, ошибки, длительность) в Prometheus, строит дашборды в Grafana. Собирает логи через Loki. Пишет простой GitHub Actions workflow: на push запуск pytest для DAG-файлов и линтеров.	Развёртывает полноценный стек мониторинга: Prometheus с алертами (Alertmanager), Grafana с готовыми дашбордами для Airflow и БД, Loki для агрегации логов, Sentry для отслеживания исключений в DAG. Настраивает уведомления в Telegram/Slack. В GitHub Actions реализует многоступенчатый CI:

				проверка кода, сборка образа, запуск тестов с testcontainers, деплой на staging при успехе, автома- тическое тести- рование ETL- пайплайна с проверкой мет- рик.
--	--	--	--	---

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕРКИ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Номер задания	Правильный ответ	Содержание вопроса	Компетенция
1.	1	Какой механизм Docker позволяет контейнеру читать и записывать файлы в директорию хоста? 1) монтирование томов (bind mount) 2) копирование файлов через docker cp 3) использование EXPOSE 4) наследование образа	ПК-5
2.	2,4	Какие два преимущества даёт использование Docker для ETL-скриптов? 1) ускорение выполнения скрипта 2) воспроизводимость окружения 3) уменьшение размера образа 4) изоляция от хостовой системы	ПК-5
3.	3	Какая SQL-конструкция в PostgreSQL используется для удаления дубликатов с сохранением только самой свежей записи? 1) GROUP BY ... HAVING COUNT(*) = 1 2) DISTINCT ON (id) ORDER BY update_date DESC 3) ROW_NUMBER() OVER (PARTITION BY id ORDER BY update_date DESC) 4) LAG() OVER (PARTITION BY id ORDER BY update_date)	ПК-3
4.	2	Какой тип пагинации в REST API считается предпочтительным для больших наборов данных? 1) offset/limit 2) cursor-based (токен следующей страницы) 3) page number 4) URL следующей страницы в теле ответа	ПК-7
5.	1,3	Какие два метода в Pandas чаще всего используются для обработки пропущенных значений? 1) fillna() 2) drop_duplicates() 3) dropna()	ПК-3

		4) astype()	
6.	2	Какой формат хранения данных в Pandas обеспечивает сжатие, встроенную схему и колоночную структуру? 1) CSV 2) Parquet 3) JSON 4) Excel	ПК-7
7.	1	В схеме «звезда» таблица фактов обычно содержит: 1) внешние ключи на измерения и числовые меры 2) только атрибуты измерений 3) только транзакционные данные без агрегатов 4) денормализованные текстовые описания	ПК-3
8.	3,4	Какие два поля обязательно присутствуют в таблице измерения при реализации SCD Type 2? 1) is_deleted 2) change_date 3) valid_from и valid_to 4) is_current	ПК-3
9.	2	Что такое DAG в Apache Airflow? 1) формат хранения метаданных 2) направленный ациклический граф задач 3) тип оператора для баз данных 4) механизм повторных попыток	ПК-7
10.	1,4	Какие два оператора Airflow чаще всего используются для загрузки CSV в PostgreSQL? 1) PostgresOperator с командой COPY 2) BashOperator с psql 3) FileSensor 4) PythonOperator с библиотекой psycopg2	ПК-7
11.	2	Какой механизм Kafka гарантирует сохранение порядка сообщений внутри одной партии? 1) группировка consumer'ов 2) offset (смещение) сообщения	ПК-7

		3) репликация 4) сжатие сообщений	
12.	3	Какой инструмент предназначен для валидации качества данных с созданием HTML-отчётов? 1) dbt 2) Apache Superset 3) Great Expectations 4) Prometheus	ПК-7
13.	2	Какой подход реализует dbt: извлечение → загрузка → трансформация? 1) ETL 2) ELT 3) ETLT 4) LT	ПК-7
14.	1,3	Какие два метода инкрементальной загрузки наиболее распространены на практике? 1) на основе timestamp 2) на основе триггеров удаления 3) на основе hash-столбца 4) на основе WAL-логов	ПК-3
15.	2	Какой тип метрики Prometheus используется для измерения распределения значений (например, времени выполнения)? 1) Counter 2) Histogram 3) Gauge 4) Summary	ПК-7
16.	1	Какой сервис в Yandex Cloud является аналогом AWS Lambda? 1) Cloud Functions 2) Object Storage 3) Managed Service for Kubernetes 4) Data Proc	ПК-5
17.	2,4	Какие два события в GitHub Actions чаще всего запускают CI-пайплайн для ETL-проекта?	ПК-7

		1) schedule 2) push в ветку main 3) fork 4) pull request	
18.	3	Какой компонент Feast отвечает за низколатентное хранение признаков для инференса? 1) offline store 2) entity 3) online store 4) feature view	ПК-7
19.	1	Какой BI-инструмент используется в лабораторных работах для визуализации данных из витрины? 1) Apache Superset 2) Tableau 3) Power BI 4) Metabase	ПК-7
20.	2	Какая команда dbt используется для выполнения всех моделей и их зависимостей? 1) dbt compile 2) dbt run 3) dbt build 4) dbt seed	ПК-7
21.	4	Что такое «водяной знак» (watermark) в контексте инкрементальной загрузки? 1) максимальное количество строк в таблице 2) флаг успешной загрузки 3) контрольная сумма данных 4) значение (время или ID), до которого данные уже загружены	ПК-3
22.	2	Как в PostgreSQL выполнить upsert (обновление существующей или вставка новой записи)? 1) MERGE INTO 2) INSERT ... ON CONFLICT DO UPDATE 3) REPLACE INTO 4) UPDATE OR INSERT	ПК-3

23.	1,3	Какие два инструмента используются в курсе для мониторинга ETL-пайплайнов? 1) Prometheus 2) Zabbix 3) Grafana 4) Nagios	ПК-7
24.	3	Какой метод в Pandas используется для удаления дубликатов строк? 1) df.dropna() 2) df.unique() 3) df.drop_duplicates() 4) df.deduplicate()	ПК-3
25.	2	Какой тип SCD используется, когда нужно хранить полную историю изменений атрибутов измерения? 1) SCD Type 1 2) SCD Type 2 3) SCD Type 3 4) SCD Type 0	ПК-3
26.	1,4	Какие два утверждения о слое staging (промежуточном слое) в хранилище данных верны? 1) данные хранятся в максимально близком к исходному виде 2) слой предназначен для конечных отчётов 3) в staging обязательно проводят агрегацию данных 4) допускаются дубликаты и «грязные» данные	ПК-3
27.	3	Какая команда Docker используется для запуска контейнера с монтированием текущей директории в папку /data внутри контейнера? 1) docker run -v /data:(pwd)image < br > 2)dockerrun --mounttype = bind, source =/data, target =(pwd) image 3) docker run -v \$(pwd):/data image 4) docker run --volume /data image	ПК-5
28.	2	Что произойдёт с контейнером после выполнения команды docker run --rm? 1) контейнер останется остановленным 2) контейнер будет автоматически удалён после остановки 3) контейнер будет перезапущен 4) контейнер получит доступ ко всем томам	ПК-5

29.	1,3	Какие два типа оконных функций в PostgreSQL чаще всего используются для дубликации? 1) ROW_NUMBER() 2) SUM() 3) RANK() 4) AVG()	ПК-3
30.	4	Какой HTTP-статус означает превышение ограничения частоты запросов (rate limiting) в REST API? 1) 200 2) 404 3) 500 4) 429	ПК-7
31.	2	Что такое экспоненциальная задержка (exponential backoff) при повторных попытках? 1) задержка, которая уменьшается с каждой попыткой 2) задержка, которая увеличивается в геометрической прогрессии 3) фиксированная задержка между попытками 4) случайная задержка	ПК-7
32.	3	Какой метод Pandas позволяет преобразовать столбец в тип datetime с заменой ошибок на NaT? 1) df['col'].astype('datetime64') 2) df['col'].convert_datetime() 3) pd.to_datetime(df['col'], errors='coerce') 4) df['col'].to_datetime(errors='ignore')	ПК-3
33.	1,4	Какие два преимущества даёт формат Parquet перед CSV для хранения очищенных данных? 1) сжатие данных без потери информации 2) поддержка макросов 3) возможность редактирования в блокноте 4) встроенная схема данных	ПК-7
34.	2	В схеме «звезда» таблица измерений dim_date обычно содержит поля year, month, day вместо хранения только full_date. Зачем это делается? 1) для соблюдения нормализации	ПК-3

		2) для ускорения группировок по уровням иерархии 3) для экономии места 4) для поддержки SCD Type 2	
35.	3	Какой тип SCD используется, когда при изменении атрибута старое значение просто перезаписывается новым без сохранения истории? 1) SCD Type 0 2) SCD Type 2 3) SCD Type 1 4) SCD Type 4	ПК-3
36.	1,2	Какие два значения обычно используются для valid_to активной записи в SCD Type 2? 1) '9999-12-31' 2) NULL в некоторых реализациях 3) текущая дата 4) дата следующего изменения	ПК-3
37.	4	Какой параметр в DAG Airflow отвечает за то, чтобы пропущенные интервалы (backfill) не выполнялись автоматически? 1) schedule_interval=None 2) start_date 3) retries 4) catchup=False	ПК-7
38.	2	Как в Airflow передать данные из одной задачи в другую (например, строку или число)? 1) через переменные окружения 2) через XCom 3) через глобальные переменные Python 4) через аргументы командной строки	ПК-7
39.	1	Какой оператор Airflow следует использовать для выполнения SQL-запроса в PostgreSQL? 1) PostgresOperator 2) SqlExecutorOperator 3) JdbcOperator 4) DatabaseOperator	ПК-7

40.	3	Что произойдёт, если в Kafka consumer указать <code>auto_offset_reset='earliest'</code>? 1) consumer начнёт читать только новые сообщения 2) consumer упадёт с ошибкой 3) consumer начнёт читать с самого старого доступного сообщения в партиции 4) consumer будет игнорировать смещения	ПК-7
41.	1,4	Какие две стратегии используются для обеспечения <code>exactly-once</code> семантики в Kafka? 1) идемпотентный producer (<code>enable.idempotence=true</code>) 2) увеличение числа партиций 3) отключение сжатия 4) транзакции (<code>transactional API</code>)	ПК-7
42.	2	Как в Great Expectations называется набор ожиданий (проверок) для конкретного набора данных? 1) Validation 2) Expectation Suite 3) Data Docs 4) Checkpoint	ПК-7
43.	4	Какая команда dbt используется для генерации HTML-документации по проекту? 1) <code>dbt compile</code> 2) <code>dbt seed</code> 3) <code>dbt run</code> 4) <code>dbt docs generate</code>	ПК-7
44.	1	В dbt функция <code>{{ ref('model_name') }}</code> используется для: 1) создания зависимости и ссылки на другую модель 2) импорта макроса 3) загрузки seed-файла 4) определения источника данных	ПК-7
45.	3	Какой метод инкрементальной загрузки не требует наличия поля <code>updated_at</code> в источнике? 1) <code>timestamp-based</code> 2) <code>watermark</code> по ID 3) <code>hash-based</code> (сравнение контрольной суммы строки) 4) логическая репликация	ПК-3

46.	2	Как в PostgreSQL выполнить ON CONFLICT только для вставки, но не обновления (игнорировать конфликт)? 1) ON CONFLICT DO UPDATE NOTHING 2) ON CONFLICT DO NOTHING 3) ON CONFLICT IGNORE 4) ON CONFLICT (id) DO NOTHING	ПК-3
47.	1	Что такое Prometheus exporter? 1) процесс, который собирает метрики с приложения и отдаёт их в формате Prometheus 2) инструмент для визуализации метрик 3) база данных для хранения метрик 4) язык запросов для метрик	ПК-7
48.	2,3	Какие два типа метрик Prometheus подходят для подсчёта количества обработанных строк в ETL? 1) Gauge 2) Counter 3) можно использовать Histogram с наблюдениями 4) Summary	ПК-7
49.	4	Какой протокол используется для работы с Yandex Object Storage из Python-скрипта (совместимость с AWS)? 1) FTP 2) WebDAV 3) NFS 4) S3	ПК-5
50.	1	Какой ресурс Yandex Cloud позволяет автоматически запускать функцию при загрузке файла в бакет? 1) триггер Object Storage 2) сервисный аккаунт 3) API Gateway 4) Yandex IoT Core	ПК-5
51.	2	Что означает зелёная галочка в GitHub Actions при выполнении workflow? 1) workflow запущен 2) все задания (jobs) завершились успешно	ПК-7

		3) есть предупреждения 4) тесты не запускались	
52.	1,3	Какие два действия обычно входят в CI-пайплайн для ETL-проекта? 1) установка зависимостей и запуск линтеров 2) деплой в production 3) запуск pytest с unit-тестами 4) ручное одобрение	ПК-7
53.	3	Что такое feature store? 1) хранилище для моделей машинного обучения 2) база данных для метаданных 3) централизованное хранилище для управления признаками (features) 4) инструмент для визуализации данных	ПК-7
54.	4	Как в Feast получить признаки для модели в режиме реального времени (инференс)? 1) через offline store и SQL-запрос 2) через materialize сначала в CSV 3) через feast materialize только для обучения 4) через get_online_features	ПК-7
55.	2	Какой тип визуализации в Apache Superset лучше всего подходит для отображения тренда продаж по дням? 1) круговая диаграмма 2) линейный график (Time-series Line Chart) 3) тепловая карта 4) столбчатая диаграмма	ПК-7
56.	1	Как в Superset создать глобальный фильтр, влияющий на все диаграммы дашборда? 1) добавить фильтр в режиме редактирования дашборда 2) прописать условие в каждой диаграмме отдельно 3) использовать SQL Lab 4) настроить параметры подключения к БД	ПК-7
57.	2,3	Какие два компонента обязательны для работы Apache Airflow в production? 1) Redis 2) метабазы (PostgreSQL/MySQL)	ПК-5

		3) Scheduler и Web Server 4) Kafka	
58.	3	Какой инструмент используется для создания таблиц и витрин данных через SQL с поддержкой версионности и тестов? 1) Apache Airflow 2) Great Expectations 3) dbt 4) Pandas	ПК-7
59.	1	Какой метод в Pandas позволяет заменить все пропуски в столбце на среднее значение этого столбца? 1) <code>df['col'].fillna(df['col'].mean())</code> 2) <code>df['col'].replace(np.nan, df['col'].mean())</code> 3) <code>df['col'].interpolate()</code> 4) <code>df['col'].dropna()</code>	ПК-3
60.	4	Что означает параметр <code>errors='coerce'</code> в <code>pd.to_numeric()</code> ? 1) выбрасывать исключение при ошибке 2) игнорировать ошибки и оставлять исходные значения 3) заменять ошибки на 0 4) заменять ошибки на NaN	ПК-3
61.	2	Какая SQL-функция в PostgreSQL объединяет несколько столбцов в одну строку с разделителем? 1) <code>CONCAT()</code> 2) <code>CONCAT_WS()</code> или 3) <code>ARRAY_AGG()</code> 4) <code>STRING_AGG()</code>	ПК-3
62.	1,4	Какие два типа индексов в PostgreSQL наиболее полезны для ускорения запросов к таблице фактов в схеме «звезда»? 1) индекс на внешние ключи (<code>date_key</code> , <code>customer_key</code>) 2) полнотекстовый индекс 3) хэш-индекс 4) составной индекс на часто фильтруемые колонки	ПК-3
63.	3	Как в Airflow реализовать ожидание появления файла в директории перед запуском основной задачи?	ПК-7

		1) использовать PythonOperator с циклом while 2) использовать BashOperator с командой test -f 3) использовать FileSensor 4) использовать HttpSensor	
64.	2	Какой тип материализации в dbt создаёт представление (VIEW), а не таблицу? 1) table 2) view 3) incremental 4) ephemeral	ПК-7
65.	1	Что такое макрос (macro) в dbt? 1) многоразовый SQL-код с использованием Jinja 2) встроенная функция тестирования 3) способ загрузки seed-данных 4) тип материализации	ПК-7
66.	3	Какой метод инкрементальной загрузки гарантирует обнаружение изменений, даже если поле updated_at не обновлялось? 1) timestamp-based с большим запасом 2) watermark по ID 3) hash-based (сравнение хеша строки) 4) триггерная таблица	ПК-3
67.	2	Как в PromQL получить количество метрик за последние 5 минут? 1) sum(metric_name) 2) rate(metric_name[5m]) 3) increase(metric_name[5m]) 4) avg(metric_name)	ПК-7
68.	1,3	Какие два способа позволяют обеспечить безопасность статических ключей доступа к Yandex Object Storage в Cloud Function? 1) хранение в переменных окружения функции 2) запись прямо в код 3) использование сервисного аккаунта с ролью storage.editor 4) передача через URL	ПК-5
69.	4	Какой шаг в GitHub Actions workflow отвечает за клонирование репозитория? 1) setup-python	ПК-7

		2) pip install 3) run: git clone 4) actions/checkout@v4	
70.	2	Что такое entity в Feast? 1) название набора признаков 2) ключ, по которому объединяются признаки (например, customer_id) 3) источник данных 4) онлайн-хранилище	ПК-7
71.	1	Как в Apache Superset создать вычисляемое поле (например, profit = sales - cost) внутри Dataset? 1) используя SQL Lab или опцию «Add calculated column» 2) только через SQL-запрос при создании датасета 3) в настройках дашборда 4) через плагин	ПК-7
72.	3	Какой тип визуализации в Superset лучше всего показывает структуру целого (доли) по категориям? 1) линейный график 2) гистограмма 3) круговая диаграмма (Pie Chart) 4) тепловая карта	ПК-7
73.	2	Какая команда Airflow CLI используется для запуска DAG вручную? 1) airflow start 2) airflow dags trigger <dag_id> 3) airflow run 4) airflow task test	ПК-7
74.	1	Что означает параметр depends_on_past в DAG Airflow? 1) задача не запустится, если предыдущий запуск этой же задачи не завершился успешно 2) задача зависит от всех предыдущих DAG-запусков 3) задача будет ждать выполнения всех upstream-задач 4) включает backfill	ПК-7
75.	3	Как в PostgreSQL выполнить кастинг строки к целому числу с возможностью обработки ошибок?	ПК-3

		1) CAST(value AS INTEGER) 2) value::INTEGER 3) value::INTEGER с обработкой через блок BEGIN EXCEPTION или NULLIF 4) CONVERT(value, INTEGER)	
76.	2,4	Какие два инструмента можно использовать для автоматической проверки стиля кода Python в CI? 1) pytest 2) flake8 3) bandit 4) black --check	ПК-7
77.	1	Что такое seed в dbt? 1) CSV-файлы, загружаемые в БД как таблицы-справочники 2) случайное начальное значение для макросов 3) тип материализации 4) тест на уникальность	ПК-7
78.	3	Какой аргумент Pandas read_csv позволяет указать, что первая строка файла содержит имена столбцов? 1) index_col=0 2) sep=',' 3) header=0 4) names=['col1','col2']	ПК-3
79.	1	Какая команда Docker используется для просмотра запущенных контейнеров? 1) docker ps 2) docker images 3) docker container ls -a 4) docker stats	ПК-5
80.	2,4	Какие два варианта использования Kafka в ETL-пайплайнах наиболее типичны? 1) постоянное хранение данных на года 2) буфер между producer и consumer для асинхронной обработки 3) реляционная база данных для витрин 4) потоковая передача данных между микросервисами	ПК-7
81.	3	Как в Great Expectations запустить проверку набора данных? 1) через GE CLI командой expect run	ПК-7

		2) через Python, создав ExpectationSuite и вызвав run_checkpoint() 3) через Checkpoint (validations) 4) автоматически при изменении данных	
82.	1	Что такое Data Docs в Great Expectations? 1) HTML-отчёт с результатами валидации данных 2) документация по коду 3) схема базы данных 4) набор тестов	ПК-7
83.	4	Какая функция PostgreSQL используется для генерации ряда дат? 1) GENERATE_SERIES() 2) DATE_RANGE() 3) SEQUENCE() 4) generate_series() для дат	ПК-3
84.	2	Как в dbt указать, что модель должна быть материализована как таблица, а не представление? 1) в YAML-файле 2) через {{ config(materialized='table') }} внутри SQL-файла 3) через аргумент командной строки 4) через макрос	ПК-7
85.	1,3	Какие два типа соединений (JOIN) чаще всего используются при заполнении таблицы фактов из нормализованных OLTP-таблиц? 1) INNER JOIN 2) FULL OUTER JOIN 3) LEFT JOIN для проверки целостности 4) CROSS JOIN	ПК-3
86.	4	Какой параметр в Airflow DAG задаёт cron-подобное расписание? 1) start_date 2) catchup 3) default_args 4) schedule_interval	ПК-7
87.	2	Как в Pandas выполнить группировку с несколькими агрегатными функциями для разных колонок? 1) df.groupby().aggregate('sum', 'mean')	ПК-3

		2) <code>df.groupby().agg({'col1':'sum', 'col2':'mean'})</code> 3) <code>df.groupby().apply()</code> 4) <code>df.pivot_table()</code>	
88.	3	Какой тип метрики Prometheus (Gauge) подходит для отображения текущего количества активных worker'ов? 1) Counter 2) Histogram 3) Gauge 4) Summary	ПК-7
89.	2	Что произойдёт, если в Cloud Function Yandex Cloud не обработать исключение, а просто вызвать raise? 1) функция будет перезапущена автоматически 2) выполнение функции завершится с ошибкой, и в логах будет traceback 3) функция продолжит работу с последней точки 4) будет отправлено уведомление в Telegram	ПК-5
90.	1	Как в Feast материализовать признаки за определённый интервал времени? 1) <code>feast materialize-incremental <end_date></code> 2) <code>feast apply</code> 3) <code>feast materialize offline</code> 4) <code>feast retrieve</code>	ПК-7
91.	3,4	Какие два способа позволяют в Airflow обрабатывать ошибки в задаче и выполнять повторные попытки? 1) использовать TaskGroup 2) установить <code>depends_on_past=True</code> 3) задать параметры <code>retries</code> и <code>retry_delay</code> в <code>default_args</code> 4) использовать <code>on_failure_callback</code>	ПК-7
92.	1	Какой метод в Pandas удаляет строки, содержащие хотя бы один пропуск? 1) <code>df.dropna()</code> 2) <code>df.fillna()</code> 3) <code>df.isnull()</code> 4) <code>df.drop_duplicates()</code>	ПК-3
93.	2	Как в PostgreSQL создать индекс для ускорения поиска по полю <code>customer_id</code> в таблице фактов?	ПК-3

		1) CREATE INDEX ON fact_sales USING HASH (customer_id) 2) CREATE INDEX idx_fact_customer ON fact_sales (customer_id) 3) CREATE INDEX ON fact_sales (customer_id, date_key) 4) ALTER TABLE fact_sales ADD PRIMARY KEY (customer_id)	
94.	3	Какая команда Airflow используется для проверки корректности DAG-файла без его выполнения? 1) airflow test 2) airflow dags list 3) airflow dags show <dag_id> 4) airflow dags validate	ПК-7
95.	1	Как в dbt настроить пользовательский тест (custom test) в виде SQL-запроса? 1) создать .sql файл в папке tests/, который возвращает строки-нарушения 2) написать макрос с именем test_ 3) добавить тест в schema.yml с параметром expression 4) использовать dbt test --custom	ПК-7
96.	2	Какой тип визуализации в Superset лучше всего подходит для отображения распределения числовой переменной (например, суммы продаж)? 1) линейный график 2) гистограмма (Histogram) 3) круговая диаграмма 4) таблица	ПК-7
97.	3	Как в GitHub Actions закешировать зависимости pip для ускорения сборки? 1) использовать actions/checkout 2) использовать actions/setup-python без кэша 3) использовать actions/cache с ключом для pip 4) положить зависимости в репозиторий	ПК-7
98.	1,4	Какие два формата данных чаще всего используются для обмена сообщениями в Kafka? 1) JSON 2) XML 3) YAML 4) Avro	ПК-7
99.	2	Что означает is_current = TRUE в таблице измерения с SCD Type 2?	ПК-3

		1) запись была удалена 2) эта версия записи является актуальной на данный момент 3) запись требует обновления 4) запись используется в фактах	
100.	3	Какая команда dbt обновляет документацию и запускает локальный веб-сервер для её просмотра? 1) dbt docs generate --serve 2) dbt docs serve --generate 3) dbt docs serve после dbt docs generate 4) dbt serve	ПК-7

2. Описание шкалы оценивания

Оценка успеваемости студентов по дисциплине осуществляется в ходе текущего контроля и промежуточной аттестации. Для проверки уровня сформированности компетенций используется совокупность оценочных средств, позволяющих оценить знания, умения и навыки, приобретённые студентами в процессе изучения дисциплины.

Основными формами текущего контроля являются:

- выполнение и защита лабораторных работ;
- устные опросы (собеседования) по темам занятий;
- проверка индивидуальных заданий;
- тестирование по отдельным разделам дисциплины.

Промежуточная аттестация проводится в форме зачета с оценкой и включает выполнение практического задания и/или ответы на теоретические вопросы.

3. Критерии оценивания компетенций заданий открытого типа

Оценка «отлично» выставляется студенту, если в ходе выполнения лабораторных работ и итогового проекта продемонстрировано умение самостоятельно разрабатывать полноценные ETL/ELT-пайплайны на Python с учётом инкрементальной загрузки, идемпотентности и обработки ошибок. Студент использует эффективные структуры данных и стандартные библиотеки (Pandas, Requests, SQLAlchemy), оптимизирует производительность (векторизованные операции, chunked-загрузка, async/await). Реализованы модульные тесты для каждого этапа, код структурирован и документирован. В проекте применяются продвинутые SQL-конструкции (оконные функции, CTE, SCD Type 2). Студент демонстрирует знание альтернативных библиотек (Dask, Polars) для работы с большими объёмами данных. Компетенция ПК-3 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если студент реализует базовый ETL-пайплайн: извлечение из одного источника (CSV или API без обработки пагинации), трансформацию с помощью Pandas (фильтрация, группировка), загрузку в базу данных через SQLAlchemy. Код содержит обработку основных исключений, но не всегда учитывает инкрементальность или идемпотентность. Используются стандартные SQL-запросы, но продвинутые конструкции (оконные функции, CTE) применяются фрагментарно. Тесты написаны только для критических функций. Компетенция ПК-3 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если студент пишет простые скрипты на Python (чтение CSV → вывод в консоль) без использования специализированных библиотек для ETL. Отсутствует обработка ошибок, пагинации при работе с API, не применяются транзакции при загрузке в БД. SQL-запросы имеют базовый уровень (SELECT, INSERT), но не используются JOIN или агрегации. Код неструктурирован (один файл, нет функций). Тесты отсутствуют. Компетенция ПК-3 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, если студент не владеет Python на уровне, достаточном для реализации ETL-задач. Не может написать код для извлечения данных из API или файлов. Не знает Pandas, Requests, SQLAlchemy. Работы с базами данных не продемонстрировано. Компетенция ПК-3 не сформирована.

Оценка «отлично» выставляется студенту, если в ходе выполнения лабораторных работ и итогового проекта продемонстрировано умение самостоятельно создавать production-готовые Docker-образы (multi-stage builds, минимальный размер, non-root пользователь) и описывать многоконтейнерные стеки в docker-compose с healthcheck, политиками перезапуска и логированием. Студент полностью автоматизирует развёртывание инфраструктуры в облаке (Yandex Cloud) с помощью Terraform (модули, remote state, переменные окружения) и настраивает serverless-компоненты (триггеры Object Storage + Cloud Functions). Обеспечивается безопасность (IAM-роли, secrets manager) и отказоустойчивость (резервное копирование, multi-AZ для БД). В отчёте представлены диаграммы архитектуры

и обоснование выбора ресурсов (CPU, память, сеть). Компетенция ПК-5 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если студент создаёт Dockerfile для упаковки ETL-скрипта и docker-compose для стека (ETL-сервис + PostgreSQL). Используются переменные окружения, volumes, depends_on. Студент вручную разворачивает инфраструктуру в облачной консоли (VM, Managed DB, bucket) и настраивает доступ к объектному хранилищу из ETL-сервиса. Применяет базовые конструкции Terraform (один ресурс) или Ansible для установки Docker на VM. Отсутствует полная автоматизация, часть шагов выполняется вручную. Компетенция ПК-5 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если студент создаёт простой Dockerfile (без оптимизации) и запускает контейнер вручную. docker-compose используется для одного сервиса. Облачные сервисы не применяются либо используются только для хранения файлов (bucket) без автоматизации развёртывания. IaC (Terraform/Ansible) не используется. Отчёт не содержит архитектурных схем или обоснования выбора ресурсов. Компетенция ПК-5 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, если студент не знаком с Docker, не может написать Dockerfile, не запускал контейнеры. Не имеет опыта работы с облачными платформами. Инфраструктура не развёрнута либо все компоненты установлены на локальной машине без контейнеризации. Компетенция ПК-5 не сформирована.

Оценка «отлично» выставляется студенту, если в ходе выполнения лабораторных работ и итогового проекта продемонстрировано умение самостоятельно проектировать сложные DAG в Apache Airflow с динамическими задачами, XCom, пользовательскими операторами и сенсорами (с тайм-аутами и повторными попытками). Настроены параллелизм, пулы, обратные вызовы для алертов. Реализован полноценный стек мониторинга: Prometheus с экспортом метрик ETL-пайплайна (время выполнения, количество строк, ошибки), Alertmanager, Grafana с дашбордами для Airflow и БД, Loki для агрегации логов, Sentry для отслеживания исключений. Настроены уведомления в Telegram/Slack. В GitHub Actions реализован многоступенчатый CI/CD: линтеры (flake8, black), статический анализ SonarQube с порогом качества, сборка образов, запуск тестов с testcontainers, автоматический деплой на staging. Компетенция ПК-7 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если студент создаёт простой DAG в Airflow (несколько последовательных задач) с использованием FileSensor и SQLExecuteQueryOperator. Настроен pre-commit с линтерами (black, flake8). Prometheus собирает базовые метрики (например, количество запусков), в Grafana построены простые графики. Логи просматриваются через консоль (без агрегации). GitHub Actions запускает pytest и линтеры при push. Отсутствуют алертинг, продвинутые сенсоры и динамические задачи. Компетенция ПК-7 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если студент запускает готовый DAG из примера (без собственной модификации). Линтеры не используются либо запускаются вручную. Prometheus и Grafana не развёрнуты либо используются для мониторинга не ETL-пайплайна, а стороннего сервиса. CI/CD не настроен; тесты запускаются локально. Студент не применяет современные DevOps-практики. Компетенция ПК-7 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, если студент не использовал Apache Airflow, не знаком с Prometheus/Grafana, не слышал о GitHub Actions. Лабораторные работы выполнены без применения оркестрации, мониторинга и CI/CD. Код не проверяется линтерами. Компетенция ПК-7 не сформирована.

4. Критерии оценивания компетенций заданий закрытого типа

Оценка «отлично» выставляется студенту, если он правильно ответил на 88-100% от общего количества заданий. Компетенция ПК-3 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если он правильно ответил на 72-87% от общего количества заданий. Компетенция ПК-3 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если он правильно ответил на 53-71% от общего количества заданий. Компетенция ПК-3 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, выставляется студенту, если он правильно ответил на менее 53% от общего количества заданий. Компетенция ПК-3 не сформирована.

Оценка «отлично» выставляется студенту, если он правильно ответил на 88-100% от общего количества заданий. Компетенция ПК-5 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если он правильно ответил на 72-87% от общего количества заданий. Компетенция ПК-5 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если он правильно ответил на 53-71% от общего количества заданий. Компетенция ПК-5 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, выставляется студенту, если он правильно ответил на менее 53% от общего количества заданий. Компетенция ПК-5 не сформирована.

Оценка «отлично» выставляется студенту, если он правильно ответил на 88-100% от общего количества заданий. Компетенция ПК-7 освоена на высоком уровне.

Оценка «хорошо» выставляется студенту, если он правильно ответил на 72-87% от общего количества заданий. Компетенция ПК-7 освоена на среднем уровне.

Оценка «удовлетворительно» выставляется студенту, если он правильно ответил на 53-71% от общего количества заданий. Компетенция ПК-7 освоена на минимальном уровне.

Оценка «неудовлетворительно» выставляется студенту, выставляется студенту, если он правильно ответил на менее 53% от общего количества заданий. Компетенция ПК-7 не сформирована.