

Название образовательной организации
Факультет / институт
Кафедра

Пояснительная записка к курсовой работе
по дисциплине «Базы данных»

Тема: Информационная система для управления ML-экспериментами и
моделями

Выполнил: _____

Группа: _____

Руководитель: _____

Город, 2025 г.

Содержание

Введение

Объект исследования — информационная система для управления жизненным циклом ML-экспериментов и моделей. Предмет исследования — проектирование реляционной базы данных, поддерживающей аудит, аналитические запросы и интеграцию с backend-API. Цель работы — разработать систему, демонстрирующую проектирование БД, ограничения целостности, функции, триггеры, оптимизацию запросов и batch-import.

1. Аналитическая часть

Системы типа MLflow/Weights&Biases фиксируют конфигурации запусков, метрики и артефакты. В курсовой работе важно обеспечить целостность данных и воспроизводимость, а также поддерживать аналитические отчёты. Ключевые сценарии: регистрация пользователей, создание проектов и экспериментов, запуск runs, логирование метрик, хранение ссылок на артефакты и сравнение результатов.

2. Проектная часть

2.1. Архитектура системы

Архитектура включает PostgreSQL 16, backend на FastAPI + SQLAlchemy + Alembic и frontend для демонстрации. Компоненты запускаются через Docker Compose.

2.2. Проектирование структуры базы данных

Схема включает 18 таблиц, обеспечивающих связи 1:1, 1:N и N:M:

- users, organizations, org_members
- ml_projects, project_members
- datasets, dataset_versions
- experiments, runs, run_configs

- metric_definitions, run_metric_values
- artifacts, run_artifacts
- audit_log
- batch_import_jobs, batch_import_errors
- project_metric_summary

Ключевые связи:

- 1:1 — runs → run_configs.
- 1:N — projects → datasets; experiments → runs.
- N:M — runs ↔ artifacts (через run_artifacts).

Ограничения целостности реализованы через PK, FK, UNIQUE, CHECK и NOT NULL, с каскадным удалением/обновлением для зависимых сущностей. Поля времени: created_at / updated_at (где применимо).

2.3. Представления

Реализованы три VIEW:

- v_runs_with_final_metrics — финальные метрики по runs;
- v_best_runs_per_experiment — лучший run по ключевой метрике;
- v_project_quality_dashboard — агрегаты по проекту (success rate, медиана времени, best metric).

2.4. Функции и триггеры

Функции:

- fn_best_run_id — скалярная функция, возвращающая лучший run по метрике и цели (min/max/last).
- fn_experiment_leaderboard — табличная функция для top-N runs по метрике.