

DataForge

Функциональные характеристики

Версия 1.1

Оглавление

1. Назначение и основные цели ПО.....	3
2. Основные функциональные блоки	3
2.1 Управление доступом и безопасность (Аутентификация)	3
2.2 Управление проектами и версиями	3
2.3 Реестр показателей и измерений (РПИ)	4
2.4 Таблицы фактов и Модель данных	4
2.5 Витрины данных	4
2.6 Генерация SQL и управление представлениями	5
2.7 Подключения к базам данных (БД).....	5
2.8 Администрирование.....	5
2.9 Интеграционное взаимодействие (API).....	6
3. Характеристики интерфейса пользователя.....	6
4. Нефункциональные требования (Ключевые)	6
5. Требования к хранению и обработке данных	6

1. Назначение и основные цели ПО

DataForge — это единая платформа для сквозного управления корпоративными аналитическими данными, обеспечивающую полный цикл работы с корпоративной аналитикой: от формализации бизнес-требований до физической реализации витрин данных в целевых СУБД.

Основные цели:

- **Стандартизация метаданных** — создание единого глоссария бизнес-показателей с контролем формул и зависимостей.
- **Визуальное моделирование хранилищ данных** — построение логических схем таблиц фактов с автоматическим разрешением зависимостей показателей.
- **Управление витринами данных** — конфигурирование аналитических структур с поддержкой различных типов объединения и уровней агрегации.
- **Версионирование и контроль изменений** — отслеживание эволюции моделей данных с возможностью сравнения и отката версий.
- **Физическая реализация** — автоматическая генерация SQL-кода и создание представлений непосредственно в подключенных базах данных.
- **Обеспечение качества** — встроенные механизмы проверки формул, тестирования данных через фильтры и просмотр происхождения показателей.

2. Основные функциональные блоки

2.1 Управление доступом и безопасность (Аутентификация)

Вход в систему:

- Используется форма с логином (email) и паролем.
- При успешной аутентификации пользователь перенаправляется на список проектов.

Управление сессиями:

- Принцип одной активной сессии (новый вход завершает старый).
- Время жизни сессии — не более 12 часов (для всех ролей).

Восстановление доступа:

- Функционал сброса пароля по email (ссылка с ограниченным сроком действия).

Защита от подбора:

- Вывод универсального сообщения об ошибке («Неверный логин или пароль») для предотвращения перебора.

IP-фильтрация (Безопасность):

- Режимы "Белый список" (доступ только разрешенным) и "Черный список" (запрет указанным).
- Возможность автоматической подстановки IP-адреса пользователя.

2.2 Управление проектами и версиями

Проекты:

- Контейнеры для изоляции данных (РПИ, таблицы, витрины) в рамках компании.
- Поддержка CRUD операций (создание, просмотр, редактирование, удаление) с валидацией названия (1-128 символов, уникальность в рамках компании).
- Клонирование проекта (копирование всех версий и настроек, кроме физических представлений в БД).

Версии:

- Срез состояния всех элементов проекта на определенный момент времени.
- Переключение версий через выпадающий список в интерфейсе.

2.3 Реестр показателей и измерений (РПИ)

Центральный семантический слой системы, состоящий из 4-х вкладок.

Показатели:

- Управление показателями: группировка, наименование, тип (Базовый/Расчетный), формула.
- **Формулы:** Поддержка арифметических операторов, ссылок на другие показатели (в квадратных скобках), агрегатных функций (SUM, AVG и др.).
- Валидация формул с подсветкой ошибок.

Измерения:

- Описание аналитических срезов (атрибутов).
- Типы: "Первичные" (прямая ссылка на поле БД) и "Производные" (вычисляются по формуле).

Факты:

- Числовые поля-источники данных. Типы: Первичный, Производный, Константа.

Справочники:

- Логическая группировка измерений с поддержкой иерархии (drag-and-drop).
- Настройка первичных и внешних ключей для связи с таблицами фактов.

Импорт/Экспорт:

- Загрузка данных из Excel (.xlsx, .xls) и CSV (макс. 10 МБ) с мастером сопоставления колонок и вариантов значений.
- Экспорт в XLSX.

Ручное заполнение:

- Табличный интерфейс с поддержкой Excel-подобной навигации (Tab, Enter, Ctrl+C/V), сортировкой и фильтрацией.

2.4 Таблицы фактов и Модель данных

Конфигурация таблицы фактов:

- Выбор элементов из РПИ (показателей, измерений, фактов, справочников).
- **Визуальная модель данных:** Интерактивная диаграмма, показывающая центральную таблицу фактов, присоединенные таблицы справочников и линии связей.
- Автоматическое разрешение зависимостей показателей (при добавлении показателя подтягиваются все необходимые для его расчета факты/измерения).

Модель данных проекта:

- Объединяет все таблицы фактов и справочники проекта в единую схему.
- Отображение связей и зависимостей показателей.

Работа с показателями в таблице фактов:

- Просмотр списка показателей, привязанных к таблице.
- **Фильтры для проверки:** Настройка сценариев проверки показателей.
- **Проверка показателя:** Генерация SQL-запроса по формуле для выбранного сценария проверки, выполнение его в БД и просмотр результата.
- **Происхождение:** Визуализация рекурсивных зависимостей показателя вплоть до полей исходных БД.

2.5 Витрины данных

Конфигурация витрин:

- Выбор элементов (показателей, фактов, измерений) из одной или нескольких таблиц фактов.

- **Тип витрины:** "С группировкой" (агрегированная) и "Без группировки" (детальная).
- **Тип слияния (для нескольких таблиц):**
 - UNION: Вертикальное объединение (добавляется служебное поле fact).
 - JOIN: Горизонтальное объединение (доступно только для агрегированных витрин, требует общих справочников).

Настройки агрегации:

- Возможность задавать индивидуальные уровни группировки для конкретных показателей (переопределение настроек витрины).

Визуализация:

- Отображение структуры витрины с указанием источников (таблиц фактов).

Проверка витрины:

- Генерация SQL SELECT-скрипта на основе конфигурации (с учетом формул из РПИ).
- Выполнение скрипта в БД и предпросмотр результата.

2.6 Генерация SQL и управление представлениями

Создание объектов:

- На основе витрины система генерирует SQL-скрипт на целевой СУБД (PostgreSQL, ClickHouse).
- Поддержка типов объектов: **Обычное представление (VIEW)**, **Материализованное представление (MATERIALIZED VIEW)**, **Таблица (TABLE)**.

Выполнение:

- Возможность выполнения скрипта непосредственно из интерфейса.
- Выбор целевой схемы/БД для развертывания.

Управление:

- Действия с созданными объектами: Показать (SELECT), Обновить (DROP/CREATE), Удалить (DROP).

2.7 Подключения к базам данных (БД)

Настройка:

- Хранение параметров подключения: тип СУБД, хост, порт, БД, схема, пользователь, пароль.
- **Оффлайн-режим:** После успешного подключения система делает "снимок" структуры БД (схема, таблицы, колонки) и работает с ним в оффлайн-режиме.

Управление:

- Проверка доступности, редактирование, временное отключение, удаление.
- Функция "Обновить все подключения" для массовой проверки.

2.8 Администрирование

Ролевая модель:

- **Наблюдатель:** Только чтение.
- **Аналитик:** Работа с РПИ, таблицами фактов, создание проектов (своих).
- **Разработчик:** Управление витринами, генерация SQL, редактирование чужих проектов.
- **Администратор:** Управление пользователями, лицензиями, просмотр журнала.
- **Суперадминистратор:** Управление компаниями (создание/удаление организаций).

Управление пользователями:

- Добавление/редактирование/удаление пользователей в рамках компании.
- Назначение ролей.

Лицензирование:

- Загрузка файла лицензии (.lic).

- Просмотр параметров лицензии (тип, сроки, лимиты по пользователям/проектам).

Системный журнал:

- Аудит всех значимых действий (входы, создания, удаления, изменения конфигураций).
- Фиксация параметров "До" и "После" для операций редактирования.

2.9 Интеграционное взаимодействие (API)

Аутентификация: API-ключи, привязанные к учетной записи пользователя (управляются администратором).

Функционал (Read-only):

- Получение списка проектов и версий.
- Получение данных РПИ (показатели, измерения, факты) в формате JSON или Excel (XLSX).
- Поддержка пагинации и локализации.

3. Характеристики интерфейса пользователя

Навигация: Единое меню (Проекты, РПИ, Таблицы фактов, Витрины, Подключения, Настройки). Отображение текущего проекта и версии в шапке.

Работа с данными:

- Табличный интерфейс с inline-редактированием (как в Excel)
- Поддержка Drag-and-Drop (особенно в конфигуляторах таблиц фактов и витрин).
- Визуальные диаграммы (происхождение, модель данных, таблицы фактов, витрины) с возможностью масштабирования, панорамирования и просмотра деталей (клик по элементу).

Локализация: Русский и Английский язык интерфейса (переключатель в профиле).

4. Нефункциональные требования (Ключевые)

Производительность:

- Отклик интерфейса: 2-4 секунды для базовых операций.
- Генерация SQL для сложной витрины: до 5 секунд.
- Поддержка нагрузки: 50+ активных пользователей, пик до 100.

Надежность:

- Время планового простоя: не более 4 часов в месяц.
- Восстановление из бэкапа: не более 4 часов.

Технологии развертывания:

- Поставляется в Docker-контейнерах.
- On-premise развертывание.
- Минимальные требования: 4 vCPU, 8 GB RAM, 100 GB SSD.
- Поддерживаемые браузеры: Chrome, Yandex Browser.

5. Требования к хранению и обработке данных

База данных метаданных: PostgreSQL (в составе Docker-образа).

Целевые СУБД для витрин: PostgreSQL (включая Postgres Pro, Jatoba), ClickHouse.

Безопасность данных:

- Передача данных только по HTTPS (TLS 1.2+).
- Хранение паролей пользователей в хэшированном виде (bcrypt/PBKDF2 с солью).
- Изоляция данных между компаниями (мультитенантность).
- Фиксация всех действий в системном журнале без возможности редактирования.