

ТЕРН Профилирование данных

Версия 2.0

Обзор функциональных возможностей

Листов 24

2024 год

АННОТАЦИЯ

В настоящем документе представлены общие сведения о программном продукте «Терн Профилирование данных» (далее – Система), описаны функциональные возможности, приводится информация по их использованию.

Оглавление

АННОТАЦИЯ	2
ГЛОССАРИЙ	4
1. Общие сведения	6
2. Назначение системы	6
2.1. Полное наименование продукта и его условное обозначение	6
2.2. Назначение и область применения	6
2.3. Задачи, решаемые системой	7
3. Функциональные возможности системы	8
3.1. Авторизация в системе	8
3.2. Соединения (коннекторы)	8
3.3. Построение профиля источника данных	9
3.4. Профилирование столбцов таблиц	11
3.5. Сравнение	13
3.6. Аналитика данных	14
3.7. Конструктор	17
3.8. Поиск	18
3.9. Планирование	22
3.10. Мониторинг	23

ГЛОССАРИЙ

№	Термин	Определение
1.	Базовое состояние Системы, Базовый вид Системы	Экранный вид Системы с единственным открытым модальным окном проводника
2.	БД, служебная БД Системы	Внутренняя база данных Системы под управлением СУБД Постгрес
3.	Документ	Многомерное понятие (или бизнес-термин), отождествляемое Системой по определенным правилам с некоторым набором столбцов в конкретной таблице источника данных
4.	Источник данных, Информационный источник, Источник	Реляционная СУБД, доступная для подключения к Системе посредством ODBC
5.	Конкурентные сценарии	Из множества доступных сценариев организации событий в один момент времени может быть выбран только один. При выполнении дополнительных условий возможна однократная (многократная) смена сценария
6.	Поле документа	Многомерная структура, описывающая тип данных, размерность, правила заполнения, ограничения по содержанию определенных символов, отождествляемая Системой по определенным правилам с конкретным столбцом из набора, формирующего Документ. Может быть одинарным или составным
7.	Составное поле документа	Объект учета в Системе, описывающий семантически связанный, сгруппированный набор полей (пример – адрес, ФИО и т.д.)
8.	Пользователь	Абстрактное физическое лицо, обремененное полномочиями и компетенцией для исполнения осознанных, целевых действий в Системе
9.	Профилирование	Исследование источника данных с целью дальнейшей аналитической обработки полученной информации
10	Профиль	Объект учета в Системе, содержащий информацию, собранную особыми способами с внешних источников информации. Могут быть двух типов – А и С с разным назначением и функциональными возможностями
11	Профиль (Тип А)	Объект учета Системы, содержащий метаданные, описывающие состояние источника данных, собранные в конкретный момент времени. Фактические данные из источника в Системе не хранятся и не обрабатываются. Полный перечень метаданных см. Приложение 1

12	Профиль (Тип С)	Объект учета Системы, содержащий метаданные, описывающие состояние столбцов заданной таблицы. Полный перечень метаданных см. Приложение 2
13	Связанность профилей	Профили Тип С могут быть построены на основе профилей Типа А с использованием зафиксированных в родительском профиле метаданных. При таком способе формирования профилей Тип С, у обоих профилей появляется значок связанности. При обработке связанных профилей действуют определенные ограничения на копирование, вырезание и удаление
14	Система	Программное обеспечение Терн Профилирование данных
15	Стандартные функции проводника, стандартные функции	<p>Функции проводника Системы, доступные при нажатии правой кнопки мыши на любом из профилей:</p> <ul style="list-style-type: none"> • Открыть; • Переименовать; • Удалить; • Вырезать; • Копировать; • Вставить; • Сравнить; • Запустить; • Запланировать.
16	СУБД	Система управления базами данных

ВАЖНО: Так будут выделены важные замечания.

1. Общие сведения

Терн Профилирование данных – решение в области изучения и исследования данных, разработанное компанией ТЕРН (tern.ru).

Использование современных методов и подходов к обработке больших объемов данных, в том числе с использованием искусственного интеллекта, машинного обучения, механизмов самообучения нейронных сетей, должно позволить создать решение, обладающее уникальными преимуществами перед уже существующими на рынке аналогами.

2. Назначение системы

2.1. Полное наименование продукта и его условное обозначение

Полное наименование системы: программный продукт «ТЕРН Профилирование данных».

Условное обозначение продукта: Система.

2.2. Назначение и область применения

«Терн Профилирование данных» (далее – Система) – решение в области выявления объектной структуры данных в системах-источниках и проверки качества данных при формировании корпоративного хранилища данных и разработки ETL/ELT в рамках их подготовки к интеллектуальному анализу, интеграции данных в корпоративных хранилищах данных (Data Warehouse) или информационных системах.

Система предназначена для подразделений, отвечающих за

- обеспечение качества данных (Data Quality);
- построение и автоматизацию процессов ETL;
- формирование и сопровождение корпоративных хранилищ данных, построение корпоративной модели данных, работу с большими данными (Big Data): аналитиков и инженеров данных (Data Scientist, Data Engeneer).

Система применяется для широкого круга задач в области построения и использования корпоративных хранилищ данных, работе с большими данными, обеспечения качества данных.

2.3. Задачи, решаемые системой

Задачами, решаемыми Системой, являются:

- Маркировка источников данных информацией объектах корпоративной модели, содержащихся в этих источниках со сбором дополнительной статистики при помощи механизмов машинного обучения;
- Верификация данных до и после ETL – обработки на предмет их качества и соответствия бизнес-требованиям;
- Поиск аномалий и представление общей картины о качестве и структуре данных в источнике;
- Автоматизированная аналитическая обработка собранных статистических данных;
- Расширение возможностей Системы в части разметки источников за счет автоматизируемых алгоритмов машинного обучения (AutoML);
- Оптимизация и ускорение процессов подготовки к загрузке и трансформации данных, а также смежных с ними задач, за счет использования готовых отраслевых объектных моделей («золотых записей» предметных областей)

К возможностям системы относятся:

- Исследование и разметка типов данных в подключаемых информационных источниках.
- Исследование, поиск аномалий и формирование аналитических метрик для таблиц баз данных;
- Формирование, хранение и ведение массива метаданных, описывающих характеристики источников и приемников данных.

- Аналитическая обработка массива метаданных в целях обеспечения решения указанных выше задач.

3. Функциональные возможности системы

ВАЖНО: Представленная версия системы использует наборы данных, профили типов А и С, документы и из поля, состоящие из тестовых и демонстрационных данных.

Работа в системе «ТЕРН Профилирование данных» осуществляется через любой современный интернет-браузер (предпочтительным является Яндекс.Браузер).

3.1. Авторизация в системе

Авторизация в Системе происходит путем ввода логина и пароля, после чего пользователю становится доступен функционал Системы.

ВАЖНО: В представленном экземпляре Системы, механизмы авторизации отключены.

3.2. Соединения (коннекторы)

Работа с источниками данных начинается с создания для них соединений (коннекторов) – именованных множеств параметров, предназначенных для определения методов доступа к источникам данных.

С соединениями доступны следующие операции:

- Создание соединения (определение названия соединения, типа соединения, источника, типа соединения).
- Редактирование соединения (доступно редактирование всех ранее введенных полей).
- Удаление соединения (для удаления соединения необходимо дополнительно подтвердить данное действие).

ВАЖНО: Удаление соединений, в представленной версии ПО невозможно.

В связи со спецификой подключения к различным СУБД, в представленной версии ПО поддерживаются соединения с реляционными базами данных Postgres и Oracle.

3.3. Построение профиля источника данных

Основной особенностью работы Системы, является использование заранее обученного набор нейронных сетей, предназначенных для исследования (профилирования) внешних по отношению к Системе, источников данных. Исследование производится с использованием механизмов искусственного интеллекта и сигнатурного анализа. Суть процесса профилирования строится на следующих фактах:

- Используются многомерные понятия (или бизнес-термины), отождествляемые Системой по определенным правилам с некоторым набором столбцов в конкретной таблице источника данных. Эти понятия называются – Документы.

ВАЖНО: Документом может являться электронный образ юридически значимого документа или любой иной объект, которому можно дать формальное описание с использованием заранее сформированных правил, а также для которого можно создать соответствующую среду машинного обучения.

- Система содержит в себе пополняемый набор Документов, наличие которых способна распознать в источниках данных.

ВАЖНО: Документом в Системе считается объект, который имеет формальное описание и для которого сформирована обученная нейросеть (одна или несколько), способная его распознать в источниках данных.

- Профилирование источников данных (построение профиля) заключается в формировании для каждой из выбранных для профилирования таблицы в источнике контрольной выборки определенного размера (обычно несколько десятков тысяч строк) и

последующего анализа этих выборок с использованием обученных нейросетей на предмет поиска в них Документов.

ВАЖНО: Система имеет в своем составе некоторый набор объектных моделей предметных областей (их иногда называют – «золотые записи») – наборов Документов, наиболее характерных (часто используемых) в данных областях.

ВАЖНО: В соответствии с ФЗ № 152 от 27.07.2006 в базовой поставке Системы отсутствуют любые объекты, которые можно было бы отождествить с персональными данными.

ВАЖНО: Система не занимается сбором, хранением и накоплением каких-либо данных или сведений, содержащихся во внешних источниках.

ВАЖНО: Модуль, обеспечивающий пополнение набора Документов в представленной версии ПО не предоставляется.

- Результатом профилирования источника данных является объект (профиль), содержащий следующие метаданные:
 - Название:
 - Профиля;
 - Источника/Приемника;
 - Схемы;
 - Базы данных;
 - Локальные данные:
 - Тип СУБД;
 - Строка подключения;
 - Количество таблиц;
 - Перечень Документов, участвующих в профилировании;
 - Наличие/отсутствие признака приемника данных
 - Длительность профилирования;
 - Дата начала профилирования;
 - Дата окончания профилирования;
 - Виртуальный путь к профилю;
 - Наличие/отсутствие связанности с другими профилями;
 - Наличие/отсутствие ошибок в процессе профилирования;

- Таблица:
 - Название;
 - Количество столбцов;
 - Количество строк;
 - Распознанный тип данных (документ);
 - Вероятность, с которой был распознан тип данных (качество данных);
 - Количество аномалий;
 - Запрет на профилирование (при наличии);
 - Размер контрольной выборки при профилировании;
- Столбец:
 - Название;
 - Порядковый номер;
 - Тип данных;
 - Размерность;
 - Пустые значения (NULL);
 - Значения по умолчанию (если есть);
 - Комментарии (если есть);
 - Соответствие полю документа (если есть);

3.4.Профилирование столбцов таблиц

Для оценки качества данных в источниках и приёмниках данных, при проектировании и верификации данных при ETL/ELT обработке, и ряде других задач, в Системе реализован инструмент, позволяющий проводить исследование (профилирование) столбцов таблиц во внешних, по отношению к Системе, источниках данных. Целью исследования является поиск аномалий, а также сбор статистических данных об их структуре, качестве и т.д. У данного инструмента существует два режима работы:

- Полное профилирование. При полном профилировании в оценке данных участвуют все строки выбранных таблиц. В зависимости от размера таблиц, данный режим может занимать много времени, выдавая при этом абсолютно точный результат. При полном профилировании используются оптимизированные аналитические алгоритмы.

- Быстрое профилирование. Быстрое профилирование производится на основании контрольных выбор небольшого размера (несколько десятков тысяч строк), что дает серьезный выигрыш по времени, однако итоговый результат имеет вероятностную природу. При быстром профилировании используются заранее обученные нейронные сети.

При профилировании столбцов таблиц, Система производит замер по следующему не редактируемому набору метрик:

1. Имя столбца;
2. Тип данных столбца;
3. Длина строки (максимальная разрядность):
 - i. Максимальное значение в столбце;
 - ii. Минимальное значение в столбце;
 - iii. Медиана;
 - iv. Среднее;
4. Наличие пустых значений (NULL):
 - i. Общее количество;
 - ii. Процентное содержание по отношению к общему количеству строк в столбце;
5. Количество нулевых значений:
 - i. Общее количество;
 - ii. Процентное содержание по отношению к общему количеству строк в столбце;
6. Уникальные значения:
 - i. Общее количество;
 - ii. Процентное содержание по отношению к общему количеству строк в столбце;
7. Повторы (одинаковые значения):
 - i. Общее количество;
 - ii. Процентное содержание по отношению к общему количеству строк в столбце;
8. Аномалии по типу:
 - i. Общее количество;
 - ii. Процентное содержание по отношению к общему количеству строк в столбце;

Для столбцов строкового типа должны быть доступны следующие метрики:

1. Имя столбца;
2. Тип данных столбца;
3. Длина строки;
4. Количество в строке:
 - i. заглавных букв;
 - ii. прописных букв;
 - iii. пробелов;
 - iv. точек;
 - v. точек запятой;
 - vi. двоеточий;
 - vii. запятых;
 - viii. тире;
 - ix. дефисов;
 - x. кириллических символов;
 - xi. латинских символов;
 - xii. ASCII символов;
5. Количество повторяющихся (одинаковых) строк.

Результаты профилирования сохраняются в БД Системы и доступны для последующего использования.

3.5.Сравнение

В сравнении могут участвовать профили одного того же источника, сделанные в разное время. Количество одновременно сравниваемых профилей не имеет значения и определяется читаемостью итоговой таблицы (условиями решаемой задачи). Один из профилей должен быть выбран основным (мастер) – все остальные будут сравниваться с ним.

Использование данной функции предполагается при решении следующего круга задач:

- Верификация данных до и после ETL – обработки на предмет их качества и соответствия бизнес-требованиям;
- Поиск аномалий и представление общей картины о качестве и структуре данных в источнике;

- Автоматизированная аналитическая обработка собранных статистических данных.

ВАЖНО: Смена основного профиля возможна в любой момент времени не ограниченное количество раз.

Результаты сравнения представляются в виде таблицы, где столбцами являются использованные профили, а строками – сравниваемые объекты. Сравнение происходит по-объектно, в объеме метаданных, указанных в п.3.3.

ВАЖНО: В сравнении участвуют только метаданные, хранимые в Системе. Фактические данные в источниках не задействуются.

При сравнении, для облегчения восприятия представленных сравнений, имеется возможность использования следующих опций:

- Скрывать 100% совпадения (исключение столбцов). Данная опция позволяет исключать из результирующей таблицы профили, совпадающие по критическим метаданным.

ВАЖНО: при этом, часть метаданных (количество строк, длительность профилирования и ряд других), не являющихся критическими и в которых всегда есть отличия, при выборе рассматриваемой опции игнорируются для профилей, остальные метаданные в которых, совпадают с основным.

- Показывать только отличия (исключение строк). При сравнении больших профилей (много таблиц и много столбцов в каждой из них), представленные результаты, среди которых большая часть полностью совпадает, затрудняют восприятие. Рассматриваемая опция, позволяет заметно оптимизировать представленные в итоговой таблице результаты.

3.6. Аналитика данных

В Системе есть возможность сравнить и сопоставить профили разных источников данных (профили типа А).

Использование данной функции предполагается при решении следующего круга задач:

1. Верификация данных до и после ETL – обработки на предмет их качества и соответствия бизнес-требованиям;
2. Поиск аномалий и представление общей картины о качестве и структуре данных в источнике;
3. Автоматизированная аналитическая обработка собранных статистических данных.

Использование данной возможности, позволят проверять ряд гипотез (метрик), представленных ниже. Система предоставляет гибкие механизмы по определению необходимых метрик, а также формированию наборов исходных данных, требуемых для их построения.

Метрики могут быть выбраны из следующего набора:

1. Наиболее часто встречающиеся Документы в наборе профилей;
2. Анализ групп дубликатов Документов. Документ А содержится в одной группе профилей, Документ Б содержится в другой группе профилей и т.д.;
3. Карта уникальных Документов (Документ входит только в один профиль);
4. Карта уникальных по Документам профилей (не содержатся пересечения с другими профилями);
5. Не встречающиеся Документы (Документы, которые заведены в систему, но не встречаются в выбранном наборе профилей);
6. В каких профилях наиболее полно встречается заданный набор Документов (при сравнении используется один из вариантов: полное или частичное совпадение). Определяется максимальное количество Документов из выбранных в рамках одного профиля;

7. В каких профилях наименее часто встречаются заданные Документы (Поиск минимума – в идеале поиск профиля, в котором ни один из представленных Документов не встречается);
8. Поиск пересечений профиля по Документам с другими выбранными профилями (разница с предыдущими метриками состоит в том, что набор Документов не задается отдельно, а берется из указанного профиля);
9. Есть ли у профиля пересечение по заданному набору документов (один или более документ) с другими профилями из набора (Разница с п.8 – поиск случайного пересечения и проверка отсутствия/наличия пересечений по заданным условиям).
10. В каких профилях содержится Документ (если документов задано несколько, ответ строится индивидуально для каждого из документов)?
11. В каких профилях Документ не содержится (если документов задано несколько, ответ строится индивидуально для каждого из документов)?
12. Показать качество данных для выбранного документа (если документов задано несколько, ответ строится индивидуально для каждого из документов) для всех профилей, куда он входит (с возможностью сортировки);
13. Сравнение профилей, содержащих одинаковые документы (если совпадений по документам несколько, то ответ строится индивидуально для каждого документа) по дополнительным показателям – количеству строк в таблицах, частоте обновления, количеству столбцов, их типам и т.д. Необходимо также предусмотреть возможность сортировки.
14. Построение Карты пересечения профилей (разница с п.8 - строится карта по принципу – каждый с каждым) по данным

Для формирования набора исходных данных, необходимых для расчета указанных выше правил, используются следующие правила:

1. Набор Документов Системы (можно выбрать из доступного списка, либо выбрать любой профиль источника, указав качестве входного параметра, весь список всех распознанных в нем документов).
2. Набор профилей. Должны соблюдаться следующие ограничения:
 - a. Один источник может быть представлен только одним профилем в указанном наборе.
 - b. Если исходный набор документов взят из какого-либо профиля, все профили этого источника в наборе выбираемых профилей, участвовать не могут.

3.7.Конструктор

Конструктор позволяет осуществлять функции по формированию, ведению и модификации объектов Системы. В частности, к возможностям Конструктора относятся:

- Добавление и редактирование Документов и их полей. При этом реализованы возможности формирования сложных семантических конструкций, учитывающих размерность, виды допустимых и не допустимых символов, участвующих в описании поля или документа, способы задания исключений, использование внешних справочников и т.д. Также представлены возможности по формированию необходимых и достаточных условий для обучения нейронных сетей, распознавать новые виды объектов.
- Добавление и редактирование подключений к источникам данных. Создан механизм, позволяющий формировать наборы метаданных (совокупность сведений, необходимых для

физического подключения к внешнему источнику) без какой-либо генерации кода.

- Подключение внешних справочников. Правила формирования и заполнения определенной части полей ряда документов, регламентированы внешними организациями, ведущими общефедеральные справочники. Реализована возможность импорта этих справочников в Систему.
- Неотъемлемой частью современных бизнес-процессов является использование производственных календарей, учитывающих все рабочие и не рабочие дни. В Системе существуют механизмы, позволяющие как создавать, так и импортировать производственные календари, для дальнейшего их использования.
- Использование механизма событий – триггеров, приходящих из внешних систем и используемых для активации некоторых из задач профилирования, является одной из значимых особенностей системы. Формирование, описание и настройка данного механизма, также осуществляется в Конструкторе.

ВАЖНО: В представленном экземпляре ПО, функции Конструктора отключены.

3.8. Поиск

В Системе реализована функция, позволяющая отображать или исключать результаты поиска по разнообразным, заранее задаваемым критериям, а также фильтровать эти результаты.

К критериям выбора для осуществления поиска относятся:

1. Имя профиля. Для поиска должно быть указано имя профиля. Вместо имени допускается использование маски, где символ «?», указывает на необходимость подстановки одной позиции любого

допустимого символа, символ «*» - неограниченного количества допустимых символов. «?» может использоваться неограниченное количество раз в любых сочетаниях с другими допустимыми символами.

2. **Виртуальный путь хранения.** При отсутствии явного указания данного параметра, поиск производится по всей базе данных. При задании виртуального пути хранения, поиск производится только в виртуальных папках более низкого уровня вложения на всю глубину вложенности. Должна быть реализована возможность указания нескольких различных виртуальных путей хранения. В этом случае, поиск должен производиться последовательно, по всем из них, на всю глубину вложенности, если она там есть.
3. **Выбор типа профиля.** Должна быть реализована возможность фильтрации результатов поиска по типу профилей (А или С). Критерий является не обязательным, при отсутствии явной спецификации выбора, выводятся результаты, содержащие оба типа профилей (при наличии в итоговой выборке результатов).
4. **Источник.** Должна присутствовать возможность выбора источника, среди профилей которого осуществляется поиск. В выборе должны быть использованы источники, которые зарегистрированы в Системе. Допускается выбор только одного источника. Редакция этого критерия Пользователем не допускается. Критерий не является обязательным, при его отсутствии, в поиске участвуют все источники, зарегистрированные в системе, с учетом ограничений, выставленных предыдущими критериями.
5. **Схема.** Должна присутствовать возможность выбора схемы в рамках источника, среди профилей которого осуществляется поиск. Допускается выбор только одной схемы. В выборе должны быть использованы схемы, которые существуют в выбранном

источнике. Редакция этого критерия Пользователем не допускается. Критерий не является обязательным, при его отсутствии, в поиске участвуют все схемы, существующие в источнике,

6. БД. Должна присутствовать возможность выбора имени БД в рамках источника, среди профилей которого осуществляется поиск. Допускается выбор только одного имени БД. Ввод данного поля возможен Пользователем только в ручном режиме. Проверки корректности ввода не производятся. Критерий не является обязательным, при его отсутствии, в поиске участвуют все имена БД, существующие в источнике,
7. Документы. При задании этого критерия должны быть указаны Документы, наличие которых в профиле является обязательным для попадания в итоговую выборку. Выбор Документов должен осуществляться из вызываемого модального окна, в котором отображается весь перечень Документов, существующих в Системе. При этом, должна быть доступна возможность выбора как всех документов, так и любого их сочетания. Осуществленный выбор должен быть подтверждаем, также должна быть предусмотрена возможность отмены ранее осуществленного выбора. Критерий не является обязательным.
8. Время создания. Должен быть реализован критерий, описывающий точные или граничные значения времени создания профилей. Используется для сужения результатов поисковой выборки или в аналитических целях. Критерий не является обязательным. Временные параметры должны задаваться с использованием отдельного модельного окна, содержащего графические элементы часов, с необходимостью подтверждения выбора или его отмены. Критерий не является обязательным.

9. Дата создания. Должен быть реализован критерий, описывающий точные или граничные значения даты создания профилей. Используется для сужения результатов поисковой выборки или в аналитических целях. Критерий не является обязательным. Временные параметры должны задаваться с использованием отдельного модельного окна, содержащего графические элементы календаря, с необходимостью подтверждения выбора или его отмены. Критерий не является обязательным.
10. Время профилирования. Должен быть реализован критерий, позволяющий описать граничные условия длительности хода процесса профилирования для сужения результатов поисковой выборки или в аналитических целях. Критерий не является обязательным.
11. Количество таблиц. В зависимости от задач, количество таблиц в профилях одного источника, сделанных в различное время, может отличаться. В этой связи, должен быть реализован критерий, позволяющий задать соответствующий диапазон значений. Критерий не является обязательным.

Фильтрация результатов поиска возможна по следующим ключам:

- Имя профиля;
- Связанность;
- Тип;
- Источник;
- Схема;
- Имя БД;
- Дата создания;
- Время создания.

ВАЖНО: При использовании двух и более ключей для фильтрации результатов поиска, последовательность их указания, значения не имеет.

3.9. Планирование

Для повышения гибкости функционала Системы, были реализованы функциональные возможности формирования, ведения и использования механизмов внутреннего расписания, регулирующего процессы профилирования, а также обеспечивающая возможность инициации процессов профилирования за счет приема и обработки соответствующих сообщений от внешних источников.

Механизмы внутреннего расписания реализованы максимально гибким и легко настраиваемым образом, что позволяет Пользователям решать широкий круг задач. Имеются возможности:

1. Настройки однократных, многократных и непрерывных запусков;
2. Формирование рабочего календаря, а также запуск процессов профилирования в зависимости от дня недели, числа в месяце, номера недели и т.д.
3. Формирования табло запланированных к запуску процессов в одном из режимов монитора.

ВАЖНО: Фактическое исполнение задач профилирования при использовании функциональных возможностей внутреннего расписания, происходит с использованием Apache AirFlow.

ВАЖНО: Запуск групповых сценариев в Системе не предусмотрен.

ВАЖНО: В Системе реализованы различные триггеры для запуска процессов профилирования. Кроме временных, существуют триггеры типа «Событие» - инициализация процесса профилирования по сигналу, пришедшему из внешней системы (является настраиваемым, количество внешних систем, и

ТЕРН Профилирование данных. Описание функциональных возможностей. Версия 2.0
соответственно сигналов не ограничено), а также триггеры типа «Действие», возможности которых завязаны на интеграцию с ПО «Терн ETL».

3.10. Мониторинг.

Каждый процесс профилирования представляет собой последовательность выполняемых операций – таких как, чтение из базы данных, подготовительные мероприятия, аналитическая обработка, формирование метаданных.

Система предоставляет возможность осуществлять контроль над процессами профилирования и давать оценку качества их работы (мониторинга процессов профилирования).

В рамках мониторинга предоставляются способы оценки текущих и законченных процессов профилирования, а также оценка с возможностью внесения модификаций в запланированные процессы.

Мониторинг позволяет получить информацию по следующим параметрам процессов профилирования:

- **Имя профиля.** Наименование объекта, который будет сохранен во внутренней БД Системы по завершению конкретного процесса профилирования.
- **Статус.** Индикатор, который может изменяться в следующих значениях – в процессе, готово, ошибка.
- **Тип.** Тип конкретного профиля – источник, приёмник, столбцы таблицы.
- **Источник, Схема, имя БД** – сведения, уточняющие метаданные конкретного профиля.
- **Дата старта.** Формат: время запуска процесса профилирования в формате год-месяц-день часы:минуты:секунды.
- **Прогресс.** Динамический коэффициент, дающий представление о ходе процесса профилирования.

- Время в процессе. Информация о времени, прошедшем с начала конкретного процесса профилирования.
- Путь. Виртуальный адрес хранения данного профиля.

ВАЖНО: Информация в окне мониторинга обновляется раз в 10 секунд.

Параметр не может быть изменен в текущей версии Системы.