Обзор платформы Dataplan: качественное решение аналитических задач ИБ

С каждым годом растет спрос на решения для обеспечения информационной безопасности. Важно комплексно подходить к анализу и выявлению инцидентов. Зачастую в организациях значительную часть ресурсов как финансовых, так и человеческих выделяют на построение активной системы защиты информации, которая редко учитывает налаженные в компании бизнес-процессы. Такая система становится помехой для комфортной и продуктивной работы сотрудников и в основном обеспечивает защиту информации только от внешних воздействий. Борьба с потенциальными внутренними нарушителями сводится к реализации ограничительных политик безопасности.

Dataplan — платформа для решения аналитических задач в ИБ. Применяется как инструмент анализа данных для определения текущего состояния, потребностей пользователей в доступе к информационным ресурсам и повышения защищенности информационных систем. Например, для оценки необходимости или перед внедрением DLP, DAM-систем. Выявлять наиболее востребованные ресурсы для разработки и внедрения механизмов повышения отказоустойчивости, а также составлять поведенческую картину использования ресурсов пользователями.

Основные функциональные возможности Dataplan:

- Сбор и обработка больших массивов данных из разных источников.
- Долговременное хранение данных.
- Расширенная поведенческая аналитика (с применением ML) действий пользователей (UBA/UEBA) и систем, с которыми они взаимодействуют (хосты, базы данных, таблицы, процессы, приложения и пр.).
- Формирование индивидуальных запросов на обработку хранимых данных.
- Графическое отображение результатов анализа.
- Ролевая модель разграничения доступа к данным платформы.
- Уведомление пользователей и ответственных лиц о результатах анализа.

Такие функции позволяют использовать Dataplan в системах организаций с разными числом сотрудников, масштабами инфраструктуры и составом средств защиты информации.

Из текущих инсталляций платформа используется:

- в ФОИВ, банках, лизинговых и страховых компаниях для анализа журналов событий доступа пользователей к критически важным базам данных для выявления инсайдерской деятельности. При этом, ряд баз данных функционирует под управлением СУБД заказных разработок, ряд из которых не имеет собственной системы логирования;

- в другом ФОИВ — для анализа журналов событий почтовой службы и выявления случаев компрометации учетных записей сотрудников, находящихся в командировках. Чтобы решить эту задачу используют открытые базы GeoIP;

- в банках и других коммерческих организациях для анализа отклонений в типовых действиях пользователей при доступе к сетевым информационным ресурсам, выявления нетиповой сетевой активности в инфраструктуре и пр. А также для обогащения данными при расследовании инцидентов информационной безопасности.

Платформой можно пользоваться специалистам с различной квалификацией – как офицерам информационной безопасности, не обладающих знаниями Data Science, так и аналитикам, создающим уникальные SQL-запросы и витрины данных.

Так, например, встроенные алгоритмы машинного обучения позволяют строить поведенческие профили (рис. 1) пользователей в несколько «кликов».

← Просмотр профиля

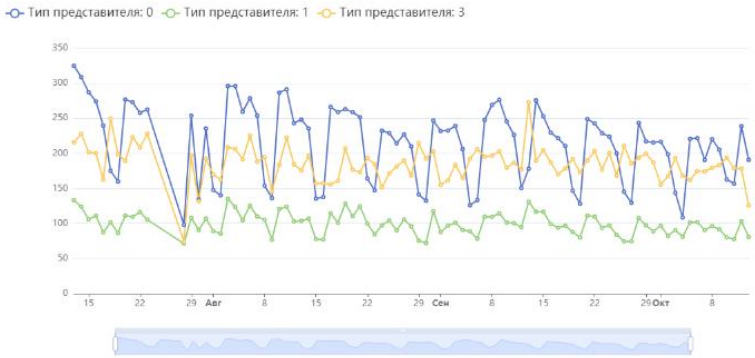
[redacted]
выполнен

Рассчитать

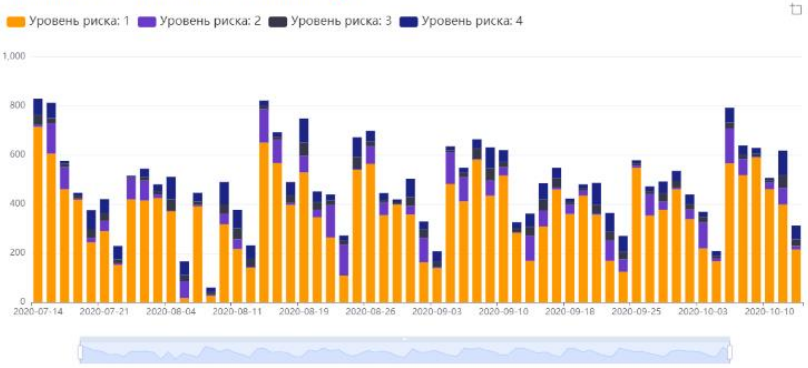
Прогнозирование загрузки - 3

Интервал 13/07/2020, 00:00 - 13/10/2020, 00:00

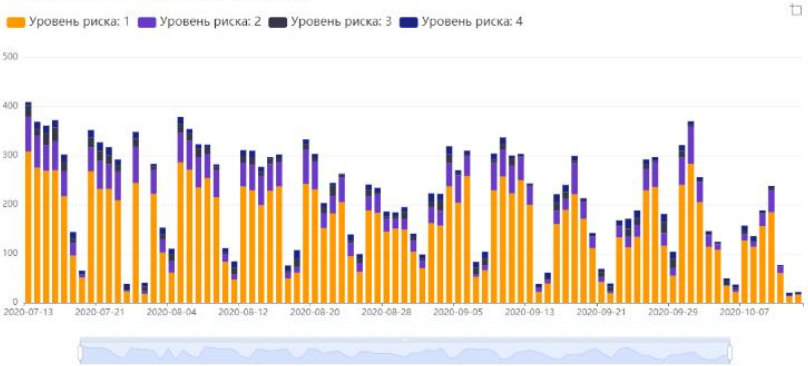
Усредненное поведение пользователя: среднее



Общее количество индивидуальных аномалий



Общее количество групповых аномалий



Уровень риска по индивидуальной и групповой статистике

Представитель пользователя	Аномалий по инд. статистике	Сумм. риск по инд. статистике	Аномалий по групп. статистике	Сумм. риск по групп. статистике
[redacted]	6	7	87	348
[redacted]	6	6	85	250
[redacted]	3	3	71	228
[redacted]	4	4	114	222
[redacted]	7	7	62	213
[redacted]	5	5	63	211
[redacted]	13	16	46	184

Рис. 1. Пример построения поведенческого профиля

Построенный профиль покажет усредненные значения, определяющие поведение пользователей (или другой исследуемой сущности) за заданный временной интервал, отобразит уровни индивидуального и группового рисков. Таким образом можно выявить отклонения в действиях пользователя по отношению к самому себе или ко всей группе пользователей (рис. 2).

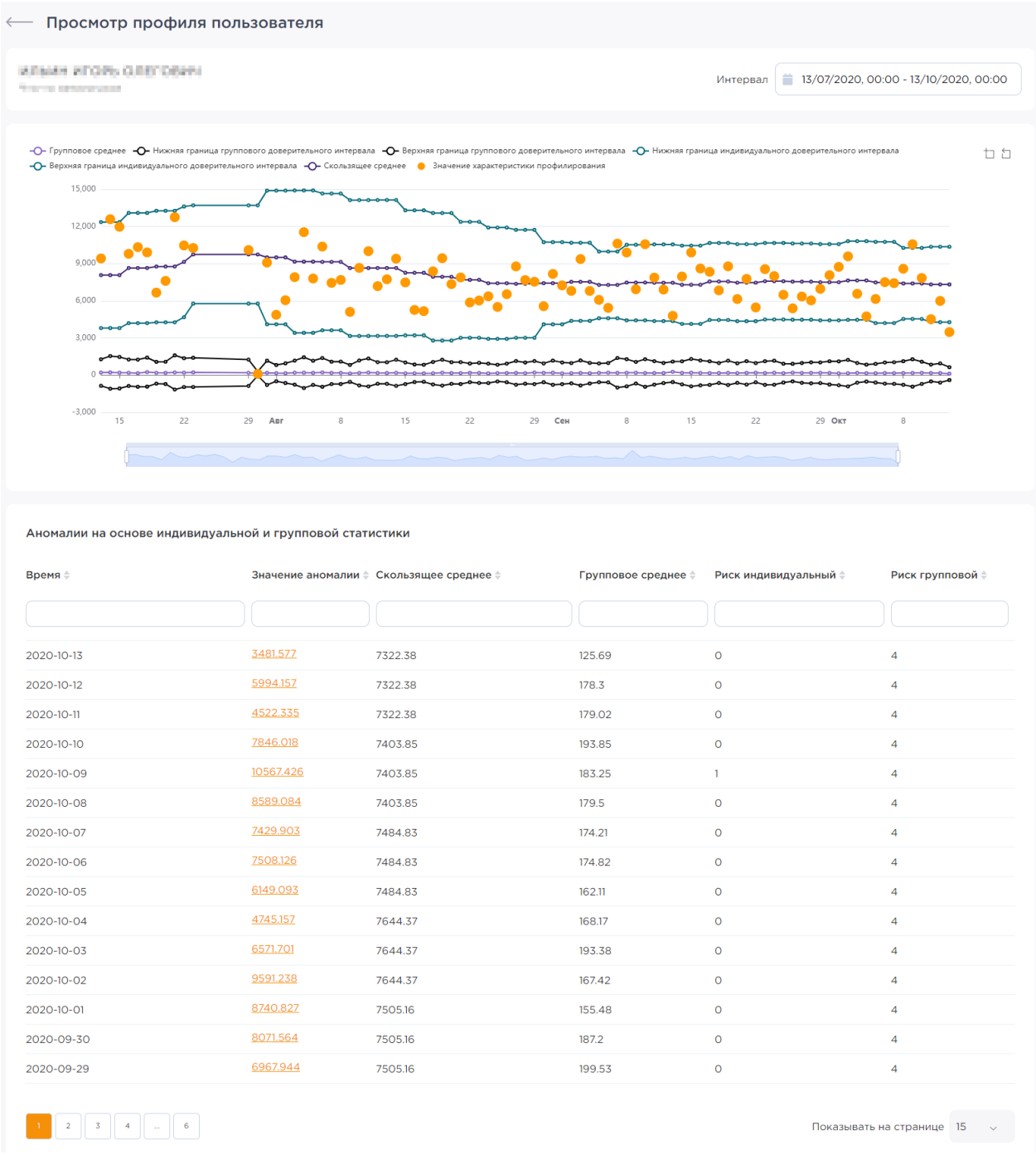


Рис. 2. Пример детализации результатов анализа по конкретному пользователю

При необходимости можно проанализировать данные, на базе которых было выявлено то или иное отклонение и выгрузить их для более детального анализа.

Поведенческие профили могут быть построены на основании данных из разных систем, в том числе представляющих сведения разного уровня важности. При этом быть направлены на решение одной задачи, например, выявление компрометации учетных данных. Для упрощения анализа такие

профили могут быть объединены в метапрофили (рис. 3), которые представляют сводную статистику по всем пользователям (или другим сущностям) этих профилей. Такие сведения позволят оценить суммарные уровни риска и аномалий, состояние системы защиты информации или бизнес-процессов, снизить количество ложных срабатываний по отдельным профилям.

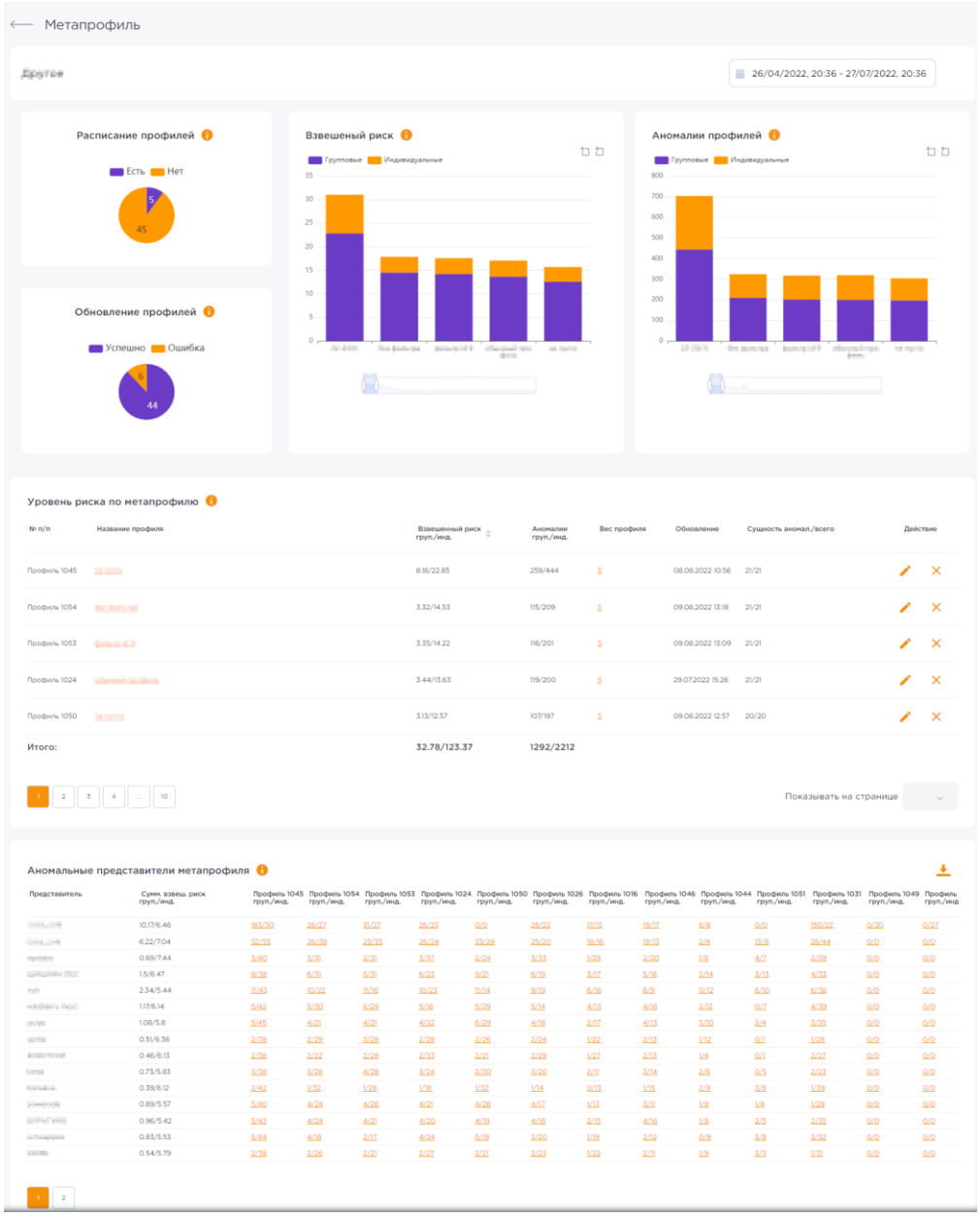


Рис. 3. Пример сводной статистики, приведенной в метапрофиле

Технические характеристики

В платформе Dataplan используется модульная архитектура, построенная с применением компонентов с открытым исходным кодом, что в том числе позволяет выполнять распределенную установку и масштабирование.

В качестве типовых форматов (протоколов) источников платформа поддерживает Syslog, ODBC/JDBC, BEATS, также можно загрузить отдельные файлы форматов CSV и JSON. Для нетиповых источников можно самостоятельно разработать коннектор, в том числе путем добавления Python-скрипта. Дополнительно разрабатывается и постоянно пополняется библиотека шаблонов коннекторов.

Долговременное хранение данных в Dataplan выполняется в столбчатой (колончатой) СУБД. Jyf позволяет применять в СХД низкоскоростные накопители информации, а встроенные механизмы сжатия позволяют использовать СХД меньших объемов.

Запросы к хранилищу данных платформы выполняются с использованием SQL-запросов или с применением соответствующих Python-скриптов. Ограничения на синтаксис запроса отсутствуют. Результаты выполнения запроса могут быть выведены непосредственно в интерфейсе платформы для ознакомления.

Графическое отображение результатов анализа выполняется в виде графиков, диаграмм, таблиц и других элементов визуализации с возможностью просмотра данных, на основе которых получены результаты (drill-down).

Разграничение доступа к хранимым в платформе данным выполняется на уровне баз данных и их таблиц, а к пользовательским данным — на уровне визуализаций, отчетов, запросов.

Платформа формирует уведомления о событиях, в том числе о состоянии потребления ресурсов, поступления новых данных, которые отображаются в виде push-уведомлений в Web-интерфейсе, а также могут быть отправлены в почтовых сообщениях и по syslog.

Сценарии применения Dataplan

Основной сценарий применения платформы – это инсталляция на высокопроизводительных вычислительных мощностях одного сервера, режим Standalone.

Рис. 4. Вариант 1

Вариант 1 подходит для небольших или средних организаций (финансовых, консалтинговых, проектных и пр.), в инфраструктуре которых используются пять-семь средств защиты информации. Ориентировочно до 10-15 информационных систем.

Для организаций с самодостаточной системой защиты информации предусмотрен второй вариант применения Dataplan – Integrated.

Рис. 5. Вариант 2

Вариант 2 подходит для средних и больших организаций (банков, лизинговых и страховых организаций, территориально-распределенные организации, корпорации), в инфраструктуре которых развернуто семь и более средств защиты информации, используется больше 15 информационных систем.

Во втором варианте Dataplan используется как средство анализа поведения пользователей по настраиваемым алгоритмам для конкретной организации или бизнес-процесса. При этом платформа взаимодействует с уже установленными системами мониторинга ИБ, которые выступают как источники событий, так и потребители результатов анализа.

Такой сценарий применения позволяет принимать более взвешенные решения по выявленным инцидентам, значительно увеличить скорость их расследования, облегчить работу сотрудников службы безопасности, исключить «человеческий фактор», повысить вероятность выявления действий внутренних нарушителей и т.д.

Состав и архитектура решения

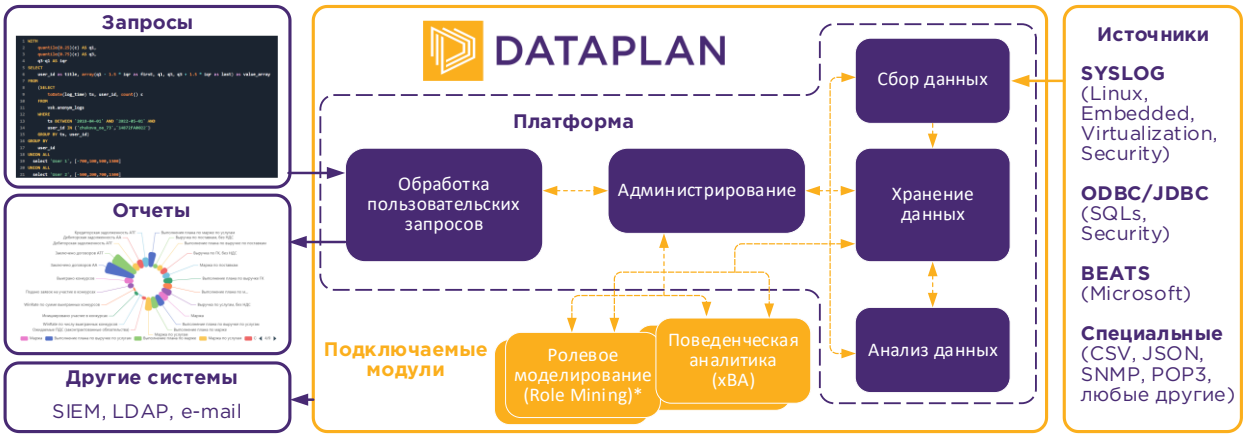


Рис. 6. Архитектура решения

Архитектура платформы, механизмы Machine Learning и анализа данных – собственная разработка NGR Softlab.

В составе платформы используются высокопроизводительные компоненты для:

- сбора данных – Logstash;
- хранения данных – ClickHouse;
- хранения служебных данных – PostgreSQL.

Системные требования и порядок внедрения

Минимальные системные требования отмечены на *рис. 7*:

Компонент	Требования					
	Характеристика	Одна нода*	Многонодовая конфигурация**			
			Сбор данных	Хранение данных	Анализ данных	Администри-рование
			Нода 1	Нода 2	Нода 3	Нода 4
			Нода 1, 2	Нода 6-8	Нода 3, 4	Нода 5
Центральный процессор	x64, не менее 2 ГГц	16 ядер	8 ядер	8 ядер	8 ядер	8 ядер
Оперативная память	от 2 ГГц	32 Гб	8 Гб	8 Гб	16 Гб	16 Гб
Хранилище (1 год хранения)	RAID 5, 6, 10	2 ТБ	500 ГБ	2 ТБ	500 ГБ	500 ГБ
Сетевой интерфейс	Ethernet	1 Гбит/с	1 Гбит/с	1 Гбит/с	1 Гбит/с	1 Гбит/с

Операционная система — Ubuntu версии не ниже 18.04 редакции LTS. Возможно развертывание в изолированном сегменте (обновление с помощью установки соответствующих патчей). Основные обновления выпускаются не чаще двух раз в год, неосновные – не чаще одного раза в квартал. Их установка выполняется при необходимости устранения критических уязвимостей или использования новых функций обновленной версии.

Интеграция с другими решениями

Для решения задач информационной безопасности в качестве источников данных могут выступать операционные системы и службы от Microsoft (контроллер домена, файловый сервер, Exchange и пр.), Unix-системы (в том числе встроенное программное обеспечение коммутаторов, маршрутизаторов и пр.), СУБД, вспомогательные системы (СКУД, видеонаблюдение и пр.), а также средства защиты информации разных классов (от антивирусных средств до межсетевых экранов, СКЗИ, SIEM, PAM и пр.).

В качестве потребителей данных могут выступать SIEM, SOAR и почтовые службы (e-mail-рассылки).

Dataplan позволяет выявлять инсайдерскую деятельность или нецелевое использование ресурсов, обнаруживать компрометацию учетных данных или правил разграничения доступа, выявлять начало атаки или подготовки к ней, а также определять различные отклонения в работе пользователей, информационных ресурсов, оборудования и др.

Платформа Dataplan может применяться в организациях с разным количеством сотрудников и информационных ресурсов. Использование решения, независимо от применяемых средств защиты информации и систем обработки данных, повышает прозрачность инфраструктуры, позволяет исследовать не только уровень защищенности информации, но и состояние бизнес-процессов компании.