

# Информационно-аналитическая платформа



## Предпосылки

### Идет цифровизация ВУЗов.

Цифровизация в российских вузах – это широкий и динамичный процесс, охватывающий различные аспекты образовательной деятельности. Цифровизация направлена на создание эффективной и открытой образовательной среды с помощью перехода к электронным системам, а также интеграции ИИ и онлайн-платформ.

Этот процесс включает в себя множество элементов: от электронных зачетных книжек и умных студенческих карт до развития онлайн-обучения и применения искусственного интеллекта. Примеры включают использование системы LMS Canvas в НИТУ "МИСиС" и проект "Современная цифровая образовательная среда в Российской Федерации".

Цифровизация влечет за собой ряд проблем, таких как вопросы безопасности данных, неравенство доступа к технологиям и необходимость постоянного обновления знаний преподавателей.

Несмотря на вызовы, цифровизация позволяет создать гибкую, доступную и персонализированную образовательную среду, подготавливая студентов к успешной карьере в меняющемся мире. Это ключевой элемент в развитии современного образования. Таким образом, цифровизация образования не просто отвечает современным

требованиям, но и открывает новые горизонты для инновационных обучающих подходов, поддерживаемых технологиями.

### Вузы собирают данные.

Большие данные в контексте высших учебных заведений – это различные массивы информации, собираемые из многочисленных источников, таких как академические записи, научные исследования, студенческие опросы и множество других. Эти данные значительно отличаются от традиционных университетских баз данных. Несмотря на то что их объём обычно может достигать несколько гигабайт, их структура масштабна и сложна, если вообще присутствует. В связи с этим для их обработки, управления и анализа требуются специализированные программные инструменты и аналитические навыки.

Такие данные могут использоваться для самых разных целей: от анализа образовательных трендов и оценки эффективности учебных методик до исследования студенческого благополучия и прогнозирования тенденций в научных исследованиях. Они помогают вузам адаптироваться к меняющимся требованиям образовательной среды и эффективно реагировать на потребности студентов и академического сообщества.

## Вузы не понимают, как работать с данными

Многие вузы не обладают достаточными ресурсами и знаниями для эффективной работы с плохо структурированной информацией, однако это необходимо в том числе для анализа образовательных трендов, повышения качества обучения и персонализации учебного процесса. Это требует привлечения специалистов и разработки специальных программ для повышения компетенций в этой области.

Кроме того, фрагментация и дублирование контента в образовании создают дополнительные сложности, снижая эффективность учебного процесса и качество образования в целом. Однако применение data-driven подхода позволяет преодолеть эти препятствия, оптимизируя управление контентом и повышая качество образования.



## Что это за подход

Data-Driven представляет собой стратегию, основанную на сборе, анализе и использовании данных для принятия обоснованных управленческих решений. В отличие от традиционных методов, основанных на интуиции или ограниченном опыте, этот подход использует объективные данные для получения глубоких и точных выводов, что способствует точному определению потребностей студентов, оценке эффективности учебных программ и методов, а также предсказанию будущих тенденций и адаптации к ним. Благодаря анализу различной информации о студентах, процессах обучения и результатов исследований, вузы могут более эффективно управлять ресурсами, улучшать учебные планы и создавать индивидуализированные образовательные траектории.

## Актуальность data-driven подхода в образовании

### Индивидуализация обучения.

Сбор и анализ данных о студентах являются центральными элементами современного образовательного процесса. Эти данные, включающие информацию о личных предпочтениях, обучающем стиле и индивидуальных потребностях каждого студента, предоставляют уникальную возможность для глубокого понимания их образовательного опыта. Подробный анализ этих данных помогает образовательным учреждениям выявлять ключевые аспекты, которые

влияют на успеваемость и мотивацию студентов, позволяя им создавать более целенаправленные и эффективные учебные программы.

Таким образом, использование данных об учащихся приводит к значительному улучшению качества образования. Это достигается путем адаптации учебных материалов и методов к индивидуальным особенностям каждого студента, обеспечивая тем самым более глубокое и всестороннее усвоение знаний. Включение этих подходов в образовательный процесс не только способствует лучшему пониманию учебного материала, но и ведет к развитию у студентов необходимых навыков для их будущей профессиональной деятельности.

### Оценка эффективности учебных методов.

Анализ данных в образовательной сфере играет ключевую роль в оценке эффективности учебных методов и материалов. Это включает в себя подробное изучение того, как различные группы студентов реагируют на разные подходы в обучении. Собирая и анализируя эту информацию, образовательные учреждения могут точно определить, какие методы наилучшим образом соответствуют потребностям и предпочтениям разных категорий учащихся. Это позволяет преподавателям и разработчикам учебных программ находить наиболее подходящие способы представления материала, что, в свою очередь, ведет к более глубокому пониманию и усвоению информации студентами.



Кроме того, такой аналитический подход способствует повышению качества учебных программ в целом. Понимание того, какие учебные материалы и методы наиболее эффективны для конкретных групп студентов, позволяет учебным заведениям адаптировать и модифицировать свои курсы таким образом, чтобы максимально соответствовать разнообразным образовательным потребностям. Это обеспечивает не только повышение качества образования, но и улучшает удовлетворенность студентов от обучения, что важно для их академического успеха и мотивации.

## Принятие обоснованных управленческих решений.

В условиях цифровой трансформации образовательной среды, ректорам и академическим директорам программ открываются новые перспективы для принятия решений, основанных на объективных данных. Это означает, что они могут опираться не только на личный опыт и интуицию, но и на точную, количественно измеряемую информацию, получаемую в результате анализа больших данных. Такой подход позволяет не только повышать эффективность учебных программ и управленческих решений, но и обеспечивает более глубокое понимание потребностей и предпочтений студентов, что способствует созданию более адаптированной и реактивной образовательной среды.

## Прогнозирование и планирование.

Data-Driven подход в образовании открывает новые возможности для

анализа текущего состояния учебного процесса. Это означает, что учебные заведения теперь могут собирать и обрабатывать большие объемы данных о студентах, методах обучения и материальном обеспечении учебного процесса. Анализ этих данных дает ценные взгляды на эффективность учебных программ, степень вовлеченности студентов и общее качество образовательного процесса. Эта информация помогает учебным заведениям внедрять своевременные изменения и улучшения, делая обучение более персонализированным, удобным и отзывчивым к нуждам студентов.

Кроме того, применение Data-Driven подхода позволяет не только оценивать настоящее, но и прогнозировать будущие тенденции в образовании. Используя данные для анализа и предсказания изменений в образовательных трендах и предпочтениях студентов, учебные заведения могут лучше планировать и адаптироваться к будущим вызовам. Это обеспечивает их способность оставаться актуальными и конкурентоспособными в быстро меняющемся мире, где технологические и социальные изменения постоянно влияют на сферу образования.

## Улучшение качества образования.

В конечном итоге, центральной задачей использования аналитики данных в образовании является повышение его качества, что стоит во

главе угла любого учебного заведения. Аналитика данных предоставляет уникальную возможность понимать и улучшать все аспекты образовательного процесса, начиная от индивидуального взаимодействия с каждым студентом до глобальной стратегии развития учебного заведения. Это включает в себя не только оптимизацию учебных программ и методик, но и улучшение управленческих решений, развитие инфраструктуры и повышение эффективности взаимодействия между студентами и преподавателями. Таким образом, аналитика данных становится неотъемлемым инструментом для достижения высокого качества образования, адаптированного к потребностям современного общества и рынка труда.

Таким образом, data-driven подход в образовании открывает новые горизонты для развития и совершенствования образовательных процессов, делая их более гибкими, эффективными и ориентированными на конкретные нужды студентов и образовательной системы в целом. Данный подход не только повышает качество образования и удовлетворенность студентов, но и служит катализатором инноваций. В современном мире, где информационные технологии стремительно развиваются, data-driven подход становится необходимым инструментом для улучшения образовательных систем и формирования стратегических решений.

## Примеры анализируемых данных

В контексте вуза, data-driven подход включает в себя анализ разнообразных типов данных для улучшения учебного процесса и управления учебным заведением.

### Студенческая успеваемость.

Динамика оценок, результатов тестов и экзаменов, исследовательских проектов студентов. Анализ этих данных может выявить области, в которых студенты испытывают трудности, и помочь разработать стратегии для улучшения образовательных результатов. Также это помогает в адаптации учебных программ, чтобы они соответствовали потребностям государства и способностям студентов.

### Выбор курсов.

Информация о курсах, которые выбирают студенты, может указывать на популярность определенных предметов и программ. Анализ этих данных помогает учебным заведениям понять, какие курсы наиболее востребованы, что важно для планирования ресурсов и разработки новых курсов.

### Данные по посещаемости.

Отслеживание посещаемости лекций и семинаров предоставляет информацию о студенческом вовлечении и помогает выявить студентов, которым может потребоваться дополнительная поддержка.

## Обратная связь от студентов и преподавателей.

Опросы, анкеты и обзоры могут предоставлять ценные данные о качестве обучения, удовлетворенности студентов, а также эффективности преподавателей. Анализ этой обратной связи важен для постоянного улучшения учебного процесса и условий обучения.

## Нагрузка на аудиторный фонд.

Занятость классических, компьютерных и гибридных аудиторий позволяет качественно планировать требуемые изменения для обеспечения корректировок учебных программ.

## Вызовы и перспективы

Несмотря на преимущества data-driven подхода, исследования Skillbox Media и НИУ ВШЭ указывают на замедленное внедрение этой стратегии в российских вузах по сравнению с мировыми стандартами. Основными препятствиями являются жесткие организационные структуры, трудности с сбором и анализом данных, особенно в региональных ВУЗах, и недостаточность финансирования. Ключевые проблемы включают сложности с доступом к данным, отсутствие унифицированных систем хранения и проблемы качества данных.

Положительные примеры внедрения подхода в России встречаются в коммерческом и онлайн-образовании, а также среди государственных вузов.

МГПУ, например, разработал предикативную модель, которая с 71% точностью предсказывает успеваемость студентов, что помогло снизить отчисления. Университет ИТМО использует анкетирование и интервью с студентами и выпускниками для улучшения учебных программ и методик преподавания.

МГУ шагнул в направлении data-driven подхода, создав систему "Истина" для сбора и анализа данных научных и образовательных организаций. Это способствовало улучшению качества научной и образовательной деятельности. Однако подобные системы еще не достигли уровня, когда их можно было бы широко внедрять в другие вузы. Наша информационно-аналитическая платформа предлагает готовое решение для организаций образовательного сектора, позволяя полноценно использовать возможности данных с минимальными изменениями в организационной структуре.

## Сбор данных

Для успешного сбора данных в учебном заведении важно разработать систематический и организованный подход. Вот основные шаги и методы сбора данных:

- **Определение целей сбора данных.**  
Прежде всего, необходимо четко определить, какие данные требуются и для каких целей. Это может быть улучшение качества обучения, повышение вовлеченности студентов, оптимизация учебных программ или оценка эффективности преподавателей.
- **Использование LMS.**  
Внедрение и использование систем управления обучением (LMS) позволяет автоматически собирать данные о студенческой активности, такие как посещаемость, выполнение заданий, активность в обсуждениях и результаты тестов.
- **Анкетирование и опросы.**  
Регулярные анкеты и опросы студентов, преподавателей и административного персонала помогают собирать качественные данные об удовлетворенности, предложениях и замечаниях, связанных с учебным процессом и условиями обучения.
- **Анализ результатов успеваемости.**  
Систематический сбор и анализ оценок, результатов экзаменов и курсовых работ предоставляют ценную информацию о успеваемости студентов и эффективности учебных программ.
- **Отслеживание посещаемости.**  
Использование электронных систем для отслеживания посещаемости лекций и семинаров помогает понять степень вовлеченности студентов и выявить тех, кто нуждается в дополнительной поддержке.
- **Сбор обратной связи.**  
Регулярное собирание обратной связи от студентов и преподавателей через интервью, фокус-группы или онлайн-платформы для сбора отзывов.
- **Интеграция и анализ данных.**  
Собранные данные должны быть интегрированы в единую систему для их анализа и интерпретации. Использование инструментов для анализа больших данных и машинного обучения может помочь в выявлении закономерностей и трендов.
- **Соблюдение конфиденциальности и этики.**  
При сборе и обработке данных необходимо строго соблюдать политику конфиденциальности и этические нормы, обеспечивая защиту личной информации студентов и сотрудников.
- **Расписание и аудиторная нагрузка.**  
Данные хранящиеся в системах университета позволяют планировать расписание и изменения аудиторного фонда.

## Анализ данных, принятие решений, оценка эффективности

Процесс анализа данных, принятия решений и оценки их результативности в образовательных учреждениях разворачивается через ряд последовательных этапов.

### Анализ данных.

После сбора данных они анализируются с использованием статистических инструментов и методов. Это может включать описательную статистику, корреляционный и регрессионный анализ, а также более сложные методы, такие как машинное обучение. Цель здесь – выявить закономерности, тренды и потенциальные проблемы. Например, проведя статистический анализ, учебное заведение может обнаружить, что студенты из определенных географических регионов или с определенным социально-экономическим статусом демонстрируют уровень успеваемости выше среднего.

### Принятие решений.

На основе результатов анализа данных принимаются решения по улучшению качества обучения и управления вузом. Данные помогают сделать эти решения обоснованными и направленными на решение конкретных задач вуза. Например, если анализ показывает, что студенты лучше усваивают материал в интерактивных форматах, учебное заведение может внести изменения в

методику преподавания или обновить учебные программы.

### Действия и оценка.

После принятия решений следует этап реализации изменений. Затем важно оценить эффективность этих изменений. Для этого проводится анализ и сравнение данных, чтобы понять, привели ли внесенные изменения к улучшению результатов. Например, после внедрения новых методов преподавания можно ознакомиться с динамикой успеваемости и обратной связи от студентов по определенной дисциплине, чтобы оценить эффективность этих методов.

## Суть продукта

Мы разработали информационно-аналитическую платформу которая представляет собой комплексное решение для внедрения data-driven подхода в учебные заведения. Она оптимизирует процессы сбора, хранения, обработки и анализа данных, что способствует повышению эффективности управления ресурсами вуза. Платформа помогает проанализировать индивидуальный подход к обучению студентов, нуждающихся в поддержке, и помогает проанализировать административные процессы в вузах. Она также помогает в прогнозировании успеваемости и управлении учебными программами на основе анализа данных о студентах и преподавателях.

Наш продукт позволяет настроить процесс обработки данных от момента их извлечения до создания аналитического дашборда, что делает его идеальным для учебных заведений с уникальными требованиями. Он не только обеспечивает широкий анализ данных, но и позволяет проводить более специфические исследования, такие как оценка качества преподавания, модификация учебных программ и эффективное управление ресурсами.

Представленные в Белой книге знания и практические рекомендации помогут понять значение и важность платформ по работе с большими данными в контексте учебных заведений и осознать шаги для успешной интеграции и использования

платформы в рамках своей организации.

## Целевая аудитория

Целевая аудитория данной Белой Книги включает в себя руководителей образовательных учреждений, администраторов и специалистов в области аналитики данных, IT-специалистов, директоров по инновациям и изменениям, а также преподавательский состав, стремящийся к улучшению процессов обучения и управления качеством образования.

## Ценностное предложение

Ценность продукта заключается в его способности адаптироваться к уникальным потребностям каждого учебного заведения, предоставляя глубокий и многофункциональный анализ данных. Продукт позволяет вузам эффективно оценивать и улучшать качество преподавания, оптимизировать учебные программы и управлять ресурсами, гарантируя повышение образовательной эффективности. Это решение обеспечивает институты мощными инструментами для принятия обоснованных решений, способствуя улучшению качества образования и обеспечивая удовлетворение потребностей студентов и преподавателей.



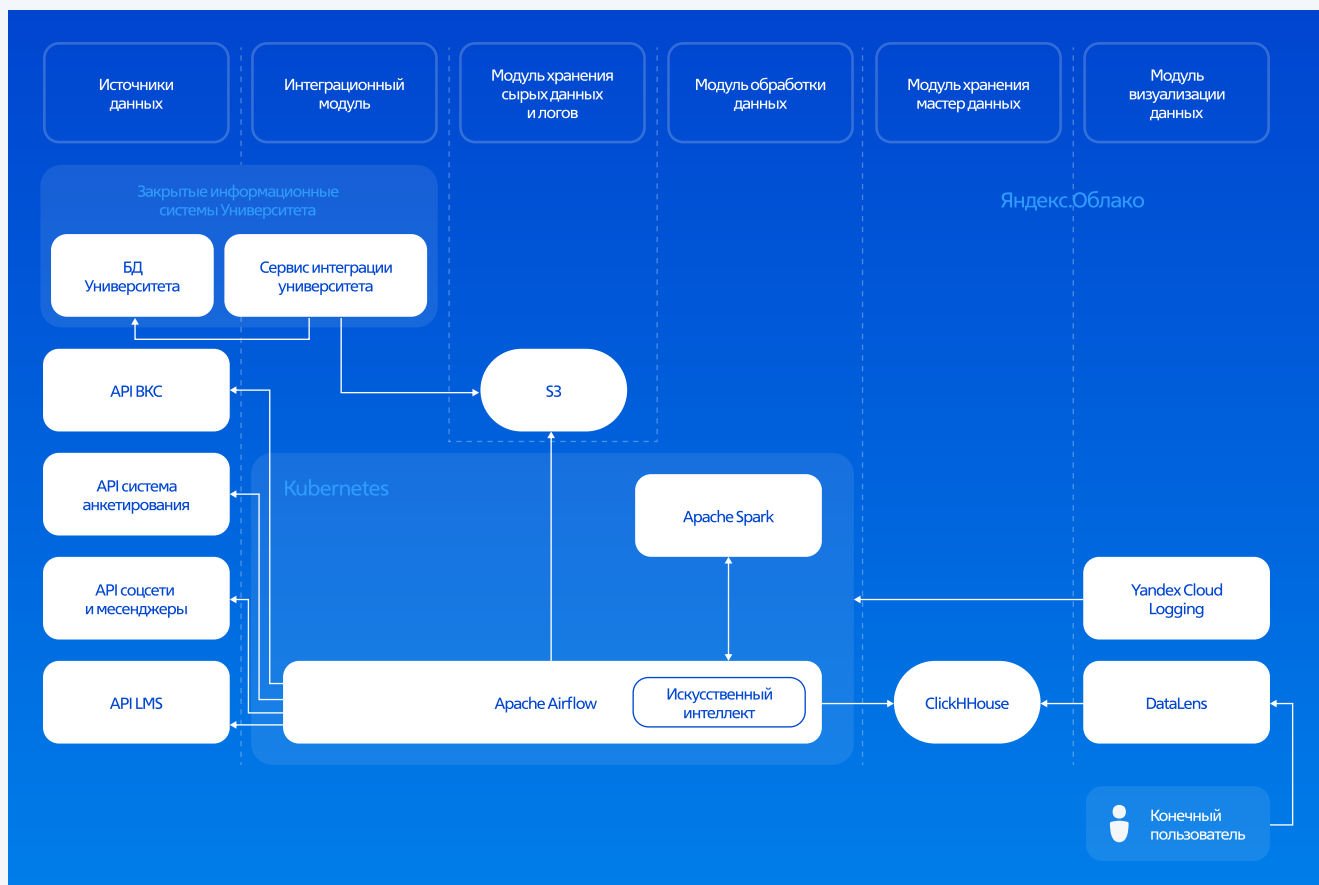
## Принципы, которые легли в основу продукта

Представляемая нами платформа для работы с большими данными - это интегрированное решение для учебных заведений, построенное на базе передовых (и тем не менее доступных большинству) технологий и инструментов. Она разработана для автоматизации сбора, обработки и анализа данных в вузах, что позволяет создавать аналитические дашборды, доступные пользователям в режиме 24/7. Ниже приведена схема, демонстрирующая структуру системы

и взаимодействие её модулей. Эта схема помогает визуализировать процесс обработки данных и показывает, как модули взаимодействуют друг с другом и как информация циркулирует внутри системы.

Принцип, лежащий в основе платформы, – это её модульное устройство, что позволяет легко адаптировать и расширять функционал в соответствии с уникальными задачами каждого учебного заведения.

## Из чего состоит продукт: модули



## Другие возможности платформы

- **Возможность обработки неструктурированных сырых данных.**

Модуль обработки и хранения мастер-данных позволяет проводить обработку данных, включая очистку, нормализацию и агрегацию, что улучшает качество данных для последующего анализа.

- **Интеграция.**

В системе предусмотрен специализированный Интеграционный модуль, отвечающий за связь с внешними источниками данных. Это подчеркивает важность взаимодействия с различными сервисами и предполагает использование API или других механизмов интеграции.

- **Централизованное хранение данных.**

Данные хранятся в централизованных хранилищах (сырые данные в хранилище S3 и обработанные данные в ClickHouse), что обеспечивает унифицированный доступ упрощает управление данными и их защиту.

## Лидеры отрасли

Большие данные — это не просто популярный термин, а ценный ресурс, который может улучшить бизнес-стратегии. Около 95% компаний сталкиваются с трудностями в управлении неструктурированными данными. Чтобы эффективно использовать большие данные, компаниям и учебным заведениям нужны простые в использовании инструменты для их обработки и анализа. Инструменты анализа данных помогают выявлять важные тенденции и предоставляют точную информацию для принятия решений.

Тем не менее, несмотря на широкий выбор инструментов, часто возникают проблемы с их настройкой под конкретные потребности учебных заведений. Продукт, о котором идет речь в этой Белой Книге, создан с учетом этих особенностей и предлагает гибкое решение, позволяющее детально настроить аналитические процессы в учебных заведениях.

Проанализируем рынок программ для обработки и анализа больших данных, сфокусировав внимание на ключевых игроках.

	Преимущества	Недостатки
SAS Analytics	Мощный аналитический функционал. Широкие возможности пользовательской настройки. Большое кол-во статистических библиотек.	Недоступен в России. Высокая стоимость. Высокий порог входа. Для использования наиболее сложных функций требуется знание R.
KNIME	Мощная No-Code система. Отличный функционал предобработки. Множество модулей для выстраивания пользовательских потоков обработки данных. Интеграция с ML-платформами (H2O) и библиотеками (Sklearn, Keras). Гибкий интерфейс. Open-source продукт.	На сложных данных система показывает меньшую эффективность, требуются навыки программирования.
Talend	Готов к использованию сразу после покупки, требует минимум времени для старта работ. Работает с большим кол-вом разных типов данных. Большое сообщество.	Недоступен в России. Слабая поддержка ML.
Klipfolio	Дашборды на уровне специализированных BI-систем. Использует Интернет для получения дополнительных данных из облака. Высокая масштабируемость.	Ограничения на тонкую настройку и предобработку данных. Невысокая скорость.
Looker	Глубокая интеграция с SQL. Интеграция с Google ML и Amazon SageMaker. Мощные API и средства разработки.	Для использования ML необходимы навыки программирования. Высокая стоимость.
Qlikview	Ассоциативная модель данных для исследования связей.	Для полноценной работы требуется большое кол-во дополнительных инструментов.
Canvas	Специализирован для образования. Поддержка различных образовательных инструментов.	Как инструмент аналитики заметно проигрывает специализированным решениям.
Blackboard Analytics	Глубокая интеграция с учебной средой Blackboard. Нацеленность на образовательные учреждения.	Ограниченная кастомизация. Функционал не дотягивает до специализированных аналитических решений.
Moodle Analytics	Интеграция с популярной системой управления обучением Moodle. Фокус на образовательной аналитике.	Для достижения высокой эффективности аналитических инструментов требуется ручная настройка.
Informatica	Мощный ETL функционал “из коробки”. Гибкие настройки.	Высокий порог входа. Высокая стоимость.
IBM SPSS Modeler	Обширные статистические инструменты. Развитая No Code система. Прогнозирование. Поддержка ML.	Недоступен в России. Высокая стоимость.
D2L Brightspace	В коробке с LMS. Мощная SaaS-аналитика.	Недоступен в России.

Часть представленных продуктов является аналитическими платформами широкого профиля, которые, тем не менее могут использоваться для образования. Характерным примером является KNIME, который, благодаря своей open-source природе и большому сообществу популярен в вузах. Однако многие из этих решений недоступны в Российской Федерации. Другая часть продуктов является дополнением к популярным системам управления обучением (LMS), такими как Moodle, Blackboard и Canvas.

Отдельно мы хотим выделить и детально рассмотреть следующие продукты – **Yandex Cloud** и **Yandex DataLens**. Эти решения важны для российского рынка образования. Yandex Cloud предлагает облачные решения для образовательной сферы, включая надежные инструменты для дистанционного обучения и масштабируемые вычислительные мощности. Yandex DataLens – система бизнес-аналитики, облегчающая работу с большими объемами данных. Продукты соблюдают требования 152-ФЗ, Yandex Cloud имеет сертификацию ФСТЭК по защите персональных данных и включены в Единый реестр российских программ для электронных вычислительных машин.

**Yandex Cloud** предлагает облачные решения для образовательной сферы, предназначенные для удовлетворения потребностей студентов, учителей и образовательных центров. Платформа поддерживает создание цифровой образовательной среды и концепцию цифрового университета. Она

включает надежные инструменты для дистанционного обучения, позволяя разрабатывать и запускать новые услуги удаленного обучения, и предлагает масштабируемые вычислительные ресурсы для обеспечения стабильности веб-сервисов при высоких нагрузках.

**Yandex DataLens** - это система бизнес-аналитики (BI), которая позволяет пользователям эффективно работать с данными. С ее помощью можно создавать интерактивные отчеты и панели мониторинга, анализировать большие объемы данных и демонстрировать результаты в наглядной форме. DataLens предоставляет инструменты для визуализации данных, позволяя легко превращать сложные данные в понятные диаграммы и графики. Этот продукт подходит для бизнес-пользователей, аналитиков и разработчиков, желающих оптимизировать процесс работы с данными.



Внедрение нашего продукта позволит собирать продвинутую статистику и осуществлять аналитическую работу в автоматическом режиме, результатом которой станет повышение качества принятия управленческих решений, оптимизация планирования и бюджетирования, повышение качества обучения, уменьшение количества отчисленных студентов по причине их слабой адаптации.

## Преимущества продукта:

### Образовательная специфика.

Платформа специально разработана для учебных заведений, в отличие от многих конкурентов, предлагающих более универсальные инструменты.

**Гибкость и масштабируемость.** Наш продукт дает учебным заведениям возможность настраивать и развивать собственную аналитику, адаптируя обработку данных и функционал в соответствии с их текущими потребностями. Кроме того, он легко интегрируется с уже существующими IT-решениями университетов, что делает его особенно удобным для использования в образовательной среде.

### Фокус на предобработке данных.

Настраиваемые процессы работы с данными обеспечивают эффективную подготовку, обработку и загрузку данных, что критично для точной аналитики.

**Импортозамещение.** Применение отечественного и open-source ПО снимает риски, связанные с возможным уходом из России зарубежных поставщиков.

## Ограничения продукта:

### Ограниченный набор визуализаций

**витрин данных** в базовой версии продукта, по сравнению с мировыми лидерами, чьи решения сейчас недоступны в России.

**Отсутствие полностью автоматизированной предобработки данных.** Это обеспечивает более гибкую и точную обработку сложных и неструктурированных образовательных данных.

## Под ключ

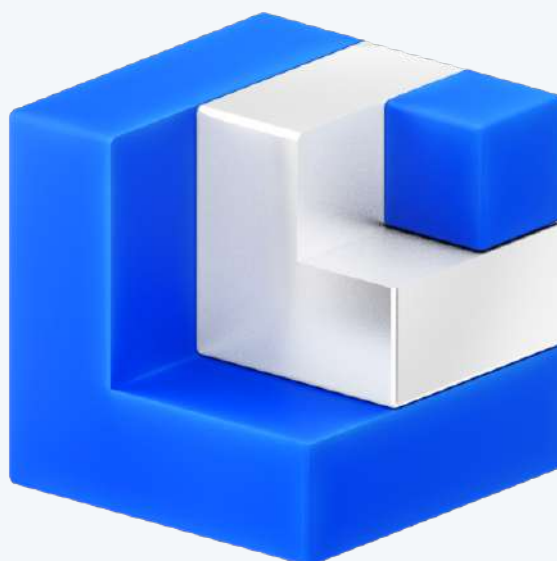
Наш продукт предоставляет полный комплекс услуг для внедрения data-driven подхода в учебных заведениях. Консультации и планирование внедрения осуществляется нашей командой, реализация и последующая поддержка выполняется командой технологических партнеров. Учебному заведению не требуется заботиться о технических аспектах интеграции; они просто оплачивают наши услуги. Мы полностью берем на себя ответственность за интеграцию нашего решения с существующими системами и базами данных вуза, обеспечивая гладкое внедрение и эффективное использование нашего продукта.

## Консалтинг

Мы предлагаем полный комплекс консультационных услуг и написание регламентов для внедрения нашей платформы в учебные заведения, обеспечивая организационную поддержку, включая методологическое встраивание data-driven подхода. Мы также активно сотрудничаем с вузами, имеющими опыт работы с нашей платформой, для перенимания лучших практик. Это обеспечивает более легкое и эффективное внедрение каждого нового решения, гарантируя гладкую интеграцию и оптимальное использование наших услуг, в то время как сам вуз обеспечивает интеграцию с сервисом и оплачивает его использование.

## Самостоятельное внедрение

Наше решение предусматривает возможность самостоятельного развертывания учебным заведением. В этом случае, вуз настраивает платформу самостоятельно на своих вычислительных мощностях или в облаке. Это подход обеспечивает полный контроль над процессом внедрения.







Пользователи имеют доступ к динамической статистике, основанной на опросах, оценках и других типах данных. Доступно множество фильтров, позволяющих оценить данных в различных разрезах. Витрины данных и дашборды могут выгружаться в форматах `xlsx`, `csv` или `pdf`.

**Модульность и взаимосвязи:** конвейер обработки данных, где каждый последующий модуль использует данные, обработанные предыдущим модулем, подчеркивает модульную структуру Платформы и важность управления данными между модулями.

**Визуализация данных:** Конечным результатом работы платформы являются аналитические дашборды, реализованные в Yandex DataLens.

**Функционал** определения тональности текста с помощью нейросети.

**Роли пользователей:** Реализованы различные уровни доступа, предусмотренные для администраторов и пользователей.

**Интеграционный модуль:** Позволяет подключаться к различным источникам данных от LMS типа Moodle до облачных хранилищ.

**Загрузка данных:** Данные загружаются из различных внешних источников, и по настраиваемому расписанию. - В случае закрытых систем используется локальный сервис для передачи данных в облачное хранилище (S3).

**Обработка данных:** Использование Apache Airflow для создания конвейеров обработки данных, автоматизации и планировании задач. Для оптимизации обработки используется Spark.

**Хранение данных:** Для хранения и работы с большими объемами данных используется СУБД ClickHouse от Yandex. Структура данных, учитывает все источники и обеспечивает их историчность.

**Технологические сервисы и инструменты:** Kubernetes используется для обеспечения масштабируемости системы и оркестрации сервисов и приложений. Платформы, как Yandex Managed Service обеспечивают хостинг и управление базами данных и кластерами.

Общая архитектура решения может быть представлена следующим образом:

**Уровень 1** - Интеграция с набором данных. Источниками данных могут выступать оценки, расписания, анкеты, опросы, логи, внутренняя статистика, тексты, чаты, видео, аудио, данные об административных процессах и прочие. Интеграция происходит через airflow REST API на кластере kubernetes.

**Уровень 2** – Хранение и обработка данных. Данные хранятся на S3, конвейер обработки включает в себя Spark, PostgreSQL, алгоритм NLP и нейросеть.

**Уровень 3** – Модуль хранения мастер-данных. Для этой цели используется хранилище S3 и реляционная СУБД PostgreSQL.

**Уровень 4** – Модуль визуализации данных. Обработанные данные с помощью ClickHouseL передаются в Yandex DataLens, где создаются аналитические дашборды.

Ниже приведена таблица, отражающая процесс обработки данных и дающая более подробное представление об архитектуре.

Подпроцесс	Описание
Чтение данных	Система читает сырые данные из объектного хранилища. Система читает последнюю версию справочника корректировки данных
Структурирование данных	Данные структурируются и сопоставляются с разработанной моделью данных в соответствии с документом «Отчет о структуре данных». В ходе сопоставления производится корректировка данных.
Обогащение данных	Ответы на опросы проходят процесс определения тональности при помощи NLP модели. Данные, полученные на предыдущем этапе, обогащаются результатами работы нейронной сети.
Агрегация данных	На данном этапе происходит агрегация и пред расчет некоторых данных для оптимизации последующего процесса визуализации данных в соответствии с документом «Отчет о структуре данных»
Сохранение данных	Обновленные и обогащённые данные записываются в базу данных ClickHouse.

# Технические и организационные требования

19

Первоначальная подготовка и настройка	Для управления доступом к облачным ресурсам потребуется регистрация аккаунта в Yandex Cloud. Описанный подход фокусируется на Yandex Cloud, однако платформа гибка и может быть развернута как в других облачных сервисах, так и на собственной инфраструктуре заказчика.
Настройка доступа и интеграционные процессы	<ul style="list-style-type: none"><li>— Подготовить API ключи для всех интегрируемых сторонних сервисов, включая Анкетолог, Moodle LMS и другие платформы</li><li>— Определить и сконфигурировать комплекты прав доступа и ролей для безопасного и эффективного управления данными в Яндекс Object Storage и прочих интегрируемых сервисах</li></ul>
Развертывание облачной инфраструктуры	<ul style="list-style-type: none"><li>— Описать последовательность шагов для развертывания Яндекс Object Storage, Яндекс Cloud Functions, управляемого ClickHouse сервиса от Яндекс, а также логирование в Яндекс.Облаке</li><li>— Настроить инфраструктуру безсерверных вычислений (Serverless) в Яндекс.Облаке</li><li>— Создать и настройте хранилища (buckets) для организации и хранения данных в соответствии с требованиями безопасности и логикой приложений</li></ul>
Конфигурирование серверного оборудования и интеграционных сервисов	<ul style="list-style-type: none"><li>— Запустить и настроить Kubernetes кластер, который будет основой для оркестрации контейнеризованных приложений</li><li>— Развернуть и сконфигурировать сервисы, обеспечивающие интеграцию с информационными системами студентов и учебными программами</li><li>— Установить Apache Airflow для автоматизации рабочих процессов, подготовьте соответствующие рабочие места и установите необходимые программные средства</li></ul>
Расширенные настройки	<ul style="list-style-type: none"><li>— Настроить группы узлов в Kubernetes для оптимизации процесса запуска приложений в контейнерах</li><li>— Установить и сконфигурировать оператор Apache SPARK</li><li>— Обеспечить доступ к Airflow извне с помощью конфигурирования ingress контроллера</li></ul>
Мониторинг и техническое обслуживание	<ul style="list-style-type: none"><li>— Настроить системы мониторинга для отслеживания состояния и производительности облачных сервисов.</li><li>— Разработать политики и процедуры для регулярного резервного копирования и оперативного восстановления данных в случае их потери или повреждения.</li></ul>

# Требуемые компетенции технических специалистов

20

- Понимание облачных сервисов Yandex и их компонентов (Yandex Cloud Functions, Yandex Message Queue, Yandex Object Storage, и других).
- Умение настраивать и управлять Serverless стеком.
- Навыки работы с Yandex Managed Service for Kubernetes для управления контейнеризованными приложениями.
- Опыт в использовании Yandex Managed Service for ClickHouse для больших данных.
- Навыки программирования на Python для создания сервисов интеграции и работы с API.
- Опыт интеграции с различными внешними сервисами, в том числе с SM платформами и сервисами для работы с историческими данными.
- Опыт работы с базами данных и системами управления базами данных (SQL, NoSQL).
- Понимание принципов ETL (Extract, Transform, Load) и опыт работы с соответствующими инструментами.
- Практические знания в области хранения, обработки и анализа больших объемов данных.
- Навыки конфигурирования и управления IAM (Identity and Access Management) для разграничения доступа в облачной среде и на уровне приложений.
- Знания в области информационной безопасности и применение практик для защиты данных.
- Владение Apache Airflow, включая разработку, планирование и отслеживание рабочих процессов данных.
- Опыт создания дашбордов и визуальных отчетов, желательно с использованием Yandex Datalens или аналогичных BI-систем.
- Системное администрирование.
- Опыт управления Linux/Unix-подобными системами.
- Навыки контейнеризации (Docker) и оркестрации (Kubernetes).
- Знание инструментов и методик логирования и мониторинга, в том числе Yandex Cloud Logging и APM (Application Performance Management) инструментов.

# Изменения в организационной структуре ВУЗа

Для успешного внедрения в университете должна быть выстроена модель управления данными по примеру той, что приведена ниже.



# Изменения в организационной структуре ВУЗа

22

Внедрение data driven подхода, в высших учебных заведениях требует как структурной, так и культурной изменений в организации. Основные организационные требования для внедрения data driven подхода в ВУЗе включают:

## **Поддержка руководства и персонала.**

Руководство передовых учебных заведений уже активно поддерживает внедрение управления данными, признавая его значимость в процессе принятия решений. Важно также обеспечить, чтобы понимание ценность были доступны и понятны всем участникам образовательного процесса, включая преподавателей.

**Управление данными.** Первый шаг к внедрению data driven подхода – это политика управления данными. В политику включают стандарты качества, целостности и доступности данных. Шаг номер два – назначение ответственных за управление данными. Ответственность ложится на таких специалистов, как Data Stewards и Data Managers, которые следят за точностью и актуальностью информации.

**Обучение и развитие.** Для развития подхода внутри вуза желательно создать системы обучения и профессиональной подготовки для сотрудников, чтобы они могли эффективно работать с данными. А для администрации и учебного персонала желательно внедрить лекции и вебинары для ознакомления с методами анализа данных.

## **Инструменты анализа данных.**

Необходимо внедрение или разработка аналитических инструментов и платформ для обработки больших объемов данных и выявления тенденций и закономерностей.

## **Лучшие практики работы с данными.**

Необходимо регулярно улучшать процессы управления, в которых принятие решений основывается на анализе данных и использовать опыт ВУЗов, успешно реализовавших подобный подход.

**Контроль эффективности.** После внедрения следует установить систему KPI и регулярный мониторинг для оценки эффективности и корректировки подходов.



Наша Аналитическая платформа была успешно внедрена и опробована в Высшей Школе Менеджмента СПбГУ. Для обработки и анализа данных ВШМ были разработаны специальные процессы обработки данных. Эти процессы позволили организовать и систематизировать данные, что облегчило создание интуитивно понятных аналитических дашбордов. В рамках проекта было реализовано:

- Автоматизированная система оценки качества образования по опросам студентов, доступного 24/7 для руководства ВШМ СПбГУ, директоров, заведующих кафедр и преподавателей.
- Автоматизация всего процесса оценки качества образования.
- Построение единой модели данных из различных источников и форматов.

Аналитические дашборды используются на ученых советах для принятия взвешенных решений. Во время совещания результаты анкетирования студентов выводятся онлайн на большом экране всем присутствующим, которые могут видеть, как студенты оценивают качество преподавания, а также тональность комментариев, классифицированных нейросетью. Личные встречи Директора Школы с преподавательским и управленческим составом сопровождаются выводом, как личных достижений преподавателя по оценке студентов (включая рейтинг относительно других), так и информацию в разрезе кафедр и уровней образования для ответственных директоров и

заведующих кафедр.

Другой пример – использование дашбордов руководителями направлений для контроля применения новых методологий в образовании. При принятии решений учитывается отраженная в дашбордах статистика: динамика оценок, количество неявок, количество пересдач, результаты опросов студентов. Дашборд позволяет оценить данные показатели в разном разрезе: на курсе, у конкретного преподавателя или дисциплины в целом.

Еще один пример использования – статистика загруженности аудиторий. Данный дашборд используется для сравнения нагрузки гибридных и обычных аудиторий. Представленная в нем статистика позволяет определить: занятые, недогруженные аудитории, распределение нагрузки по дням и по часам, а также нагрузку на учебный день и вечер.

Всего после внедрения нашей аналитической платформы:

- Было обработано более **1700 опросов**
- Более **5700 комментариев** было оценено нейросетью по тональности.
- Преподаватели получили прямой доступ к аналитическим данным.
- Произведена оценка результатов внедрения новой методологии учебной траектории “Цифровые инструменты для менеджеров”.

Выше было представлено описание реализованной нами информационно-аналитической платформы, включая принципы ее работы, ключевые преимущества для российских вузов, а также успешных примеров ее внедрения.

Данное руководство позволяет оценить общую картину возможностей и потенциала платформы. Структурирование и интеграция больших данных на всех уровнях - от начального сбора информации до ее визуализации и использования в управленческих решениях - являются ключевыми элементами для их успешного управления и анализа.

Пример успешного внедрения демонстрируют, как разнообразный набор инструментов и интеграция данных открывают перед руководством вузов и преподавательским составом новые перспективы. С их помощью учебные заведения могут не только существенно улучшить качество образования, но и эффективно выстроить процесс мониторинга успеваемости студентов, а также оптимизировать использование ресурсов. Это делает нашу платформу не просто инструментом, а надежным партнером в процессе повышения эффективности и качества образования.

Термин	Пояснения
Data-driven подход	Подход к принятию управленческих решений, основанный на доверии к данным и результатам их анализа.
ETL-пайплайн	Набор задач, которые выполняются в определенной последовательности. К примеру, батчевый процесс в Apache Airflow (данные собираются частями, после чего запускается процесс по расписанию). Если работа ведется на распределённых системах, то разработка ETL-процесса может быть осуществляться таким образом, чтобы задачи выполнялись параллельно.
Администратор	Лицо, ответственное за работоспособность и надлежащее функционирование всех элементов Информационной Платформы, определяемое поручениями Функционального заказчика.
Витрина данных	Подмножество (срез) хранилища данных, представляющее собой массив тематической, узконаправленной информации, ориентированной на пользователей Информационной Платформы.
Пользователь	Человек, использующий функции Системы с определенным набором прав, определяемый поручениями Функционального заказчика.
Информационная система	Совокупность содержащейся в базах данных информации и информационных технологий, а также программных и технических средств, обеспечивающих ее обработку.
Яндекс.Облако	Публичная облачная платформа от российской интернет-компании «Яндекс», используемая в качестве базовой в работах по созданию опытного образца.
NLP модель	Модель для анализа тональности текстов обеспечивает определение эмоциональной окраски текста
RAW данные (сырые данные)	Формат данных, содержащий необработанные, или обработанные в минимальной степени данные.
Большие данные (Big Data)	Огромные объёмы структурированных и неструктурированных данных, часто производимые в реальном времени, которые превышают способности традиционных баз данных в плане обработки и хранения.
Аналитические инструменты (Analytics Tools)	Программное обеспечение и приложения, используемые для анализа больших наборов данных, чтобы выявлять закономерности, тренды и получать инсайты, способствующие бизнес-принятию решений.
ETL (Extract, Transform, Load)	Процесс, который использует для извлечения данных из различных источников, их преобразования в нужный формат и загрузки в целевое хранилище.
No-Code система	Платформа, позволяющая пользователям создавать приложения и процессы без необходимости написания кода, через графический пользовательский интерфейс.

Термин	Пояснения
Машинное обучение (ML)	Область искусственного интеллекта, которая использует статистические методы, чтобы дать компьютерам способность "учиться" с данными и совершать действия, не являющиеся явно запрограммированными.
Кастомизация (Customization)	Настройка программного обеспечения или систем для соответствия конкретным потребностям пользователя или организации.
API (Application Programming Interface)	Набор правил и протоколов для создания и взаимодействия программного обеспечения. API позволяют разным программам коммуницировать друг с другом.
Кластер (Cluster)	Группа связанных серверов, работающих вместе таким образом, что во многом они могут быть восприняты как один системный узел.
Дашборд (Dashboard)	Визуальная панель для отображения самых важных информационных показателей, которые необходимы для достижения одной или нескольких целей; они собраны на одном экране для легкости мониторинга.
Airflow	Платформа для оркестрации рабочих процессов, используемая для автоматизации процедур интеграции данных.
Kubernetes (K8s)	Система для автоматизации развёртывания, масштабирования и управления контейнеризированными приложениями.
S3 (Simple Storage Service)	Объектное хранилище данных, предлагаемое компаниями облачных сервисов (например, AWS)
Модуль	Отдельный блок образовательного курса или программы, посвященный определенной теме или набору умений, и имеющий определенные образовательные цели и результаты.
Spark	Распределённая вычислительная система, используемая для обработки больших объёмов данных.
PostgreSQL	Распространённая реляционная база данных с открытым исходным кодом.
Yandex DataLens	Сервис от Яндекс для создания аналитических дашбордов и визуализации данных.
API-ключи	Уникальные идентификаторы, используемые для авторизации и взаимодействия с внешними сервисами через API.
Serverless	Парадигма облачных вычислений, позволяющая пользователям запускать код без необходимости управления серверами
ClickHouse	Высокопроизводительная OLAP (онлайн-аналитическая обработка) база данных, предназначенная для обработки больших объёмов данных на лету.
Анкетолог и Survey Monkey	Сервисы для проведения опросов и анкетирования.



## Соловьев Анатолий

Директор по управлению проектами Банка ВТБ «ПАО»

С 2004 работал в банках с государственным участием и руководил проектными подразделениями. Реализовано более 20 крупных проектов и систем. С 2019 отвечал за направление по сбору биометрических данных в государственный сервис «Единая биометрическая система». С 2021 года отвечает за развитие цифровых навыков студентов бакалавриата ВШМ СПбГУ и цифровых образовательных платформ в вузах, где банк является попечителем.



## Гаршин Василий

Управляющий директор ПАО Банк ВТБ

Разработчик платформы SimBA. Имеет опыт реализации проектов по цифровой трансформации (направление Big Data, ML, AI). Занимается развитием магистерской программы ВШМ СПбГУ «Бизнес-аналитика и большие данные». Преподаватель ВШМ СПбГУ.

