ВТБ

2023

Платформа для моделирования задач бизнес-аналитики и анализа данных

Simulation platform for Business Analytics and big data (SimBA)

Введение в проблему

Работа над обновлением магистерской программы «Бизнес-аналитика и большие данные» (Master in Business Analytics and Big Data или сокращенно MiBA) в Высшей школе менеджмента СПбГУ (далее – Школа) потребовала актуализации не только дисциплин технологического трека, но и соответствующих им инструментов.

Необходимо отметить, что данная актуализация пришлась на период COVID-19, что выдвигало дополнительные требования в части возможности удаленного доступа к инструментам преподавания таких дисциплин, как “Основы разработки” или “Машинное обучение и большие данные”. Одновременно с этим, в рамках программы требовалось разработать новые курсы по изучению технологий работы с большими данными.

Существующий к тому моменту подход к преподаванию технологических дисциплин предполагал разворачивание необходимых инструментов и сред на компьютерах в компьютерном классе Школы силами самих обучающихся при помощи преподавателей соответствующих курсов.

Одновременно с этим в Школе проводились мастер-классы и работа в лабораториях, предусматривающая анализ данных силами проектных команд, что, в свою очередь, требовало доступ к вычислительным ресурсам, а также к ресурсам для хранения данных для проектов. При этом в целях экономии ресурсов требовалось обеспечить гибкость в части выделения ресурсов, их эффективное масштабирование как в сторону быстрого увеличения при одновременном старте работ многих команд, так и быстрого сворачивания после завершения основной части разработки и вычислений.

Таким образом, возникла потребность в доступном, гибким с точки зрения настроек и легко масштабируемом инструменте для обеспечения преподавателей, обучающихся и участников проектных групп возможностями по изучению технологических курсов (в первую очередь курсов по программированию), а также работе с данными. Отсутствие такого инструмента привело бы к сложностям в преподавании курсов, возрастанию временных и финансовых затрат на организацию соответствующей инфраструктуры.

Для кого написан отчет

Настоящий документ предлагает описание варианта реализации платформы для эффективного решения подобных задах, апробированного в Школе не только в рамках магистерской программы MiBA, но и на других направлениях. Документ описывает основные элементы решения в виде набора шагов по разворачиванию облачной ИТ платформы и рецептом по ее настройке и интеграции в учебные и исследовательские проекты.

В документе изложены:

- Принципы функционирования платформы
- Подходы к организации работы на платформе
- Основные требования к ресурсам для разворачивания платформы
- Элементы технического решения для платформы
- Примеры технической конфигурации под различные образовательные и исследовательские задачи

Документ предназначен для преподавателей дисциплин по программированию и работе с большими данными, руководителей исследовательских направлений или проектов по анализу данных, руководителей и директоров ИТ служб учебных заведений, в сферу ответственности которых входит обеспечение учебных процессов необходимой инфраструктурой.

Обзор существующих решений

Современные требования к анализу данных включают в себя необходимость работы в том числе с инструментами по обработке больших массивов как структурированных, так и слабоструктурированных данных. Аналитики данных в современном мире должны быть знакомы с основными концепциями и фреймворками обработки больших данных на различных уровнях – от базовой архитектуры решения до точечного владения выбранными инструментами сбора, обработки, анализа и визуализации.

В настоящее время в большинстве образовательных курсов или внутри специализированных программ, предлагаемых учебными заведениями, при изучении основ программирования, а также основ аналитики данных используется язык программирования Python и множество связанных с ним фреймворков. Для его изучения, а также для работы с широким спектром модулей (библиотек) для анализа данных, учащимся в большинстве случаев предлагается либо установить на свои персональные компьютеры различного рода программы, в основе своей содержащие интерпретатор языка, а также IDE (Integrated Development Environment, интегрированная среда разработки), либо использовать облачные платформы, предназначенные в первую очередь для ML (machine learning) задач. Именно к классу последних и относится Simulation platform for Business Analytics and Big Data (или сокращенно SimBA).

Перечень основных представителей облачных платформ, предназначенных в той или иной мере для решения задач по анализу данных, приведен в ПРИЛОЖЕНИИ. Далее остановимся на преимуществах и недостатках этих решений.

Существующие альтернативы для учебных заведений, в которых осуществляется преподавание дисциплин, требующих доступ к ИТ продуктам или решениям для образовательных задач, варьируются от собственных компьютерных классов до использования внешних облачных платформ или продуктов. Сравнение различных альтернативных вариантов с учетом преимуществ и недостатков каждого из решений приведено ниже в таблице.

	Преимущества	Недостатки
Компьютерный класс	Прямой контакт с преподавателем Возможность настройки единой среды для всех пользователей	Невозможность быстро масштабировать вычислительные мощности Ограничения по доступу в класс, невозможность провести занятия в другой аудитории. Издержки на содержание специализированной аудитории
Персональные компьютеры обучающихся	Удобство для пользователей, работа на знакомом оборудовании	Естественные ограничения вычислительных мощностей Невозможно управлять рабочей средой, издержки на ее настройку для каждого пользователя. Сложно организовать работу в команде и авторизованный доступ к данным
Google Colab	Широкая доступность. Низкий порог входа. Возможность бесплатного доступа к ресурсам облака, Linux-среда, бесплатные GPU. Интеграция с другими сервисами Google	Ограничения по времени использования. Настройка специализированной среды требует от пользователя дополнительных навыков Нет гарантированной доступности вычислительных ресурсов (GPU). Доступные ресурсы ограничены по мощности. Риски доступности в силу санкционных ограничений
Kaggle	Доступность при регистрации на самой платформе. Интеграция рабочей среды в комьюнити, доступ к данным и обмен опытом. Возможность командной работы. Доступ к вычислительным ресурсам (GPU)	Ограничения по времени использования. Доступные ресурсы ограничены по мощности. Специализированное решение под задачи платформы
Google Cloud ML	Привычная Linux-среда для разработки. Доступ к неограниченным вычислительным ресурсам (в том числе GPU). Поддержка облачных решений. Интеграция с продуктами Google	Нет бесплатной версии. Относительно высокий порог входа для неопытных пользователей. Риски доступности в силу санкционных ограничений, проблемы с оплатой
Microsoft Azure	Широкий набор инструментов для разметки и предобработки данных Динамическое масштабирования вычислительных ресурсов (GPU)	Нет бесплатной версии Относительно высокий порог входа для неопытных пользователей Риски доступности в силу санкционных ограничений, проблемы с оплатой
Yandex DataSphere	Эффективное решение от российской компании. Доступ к экосистеме Yandex Cloud. Доступ к вычислительным ресурсам различной конфигурации. Удобная тарификация. Удобная интеграция	Высокий порог входа для неопытных пользователей Настройка специализированной среды требует от пользователя дополнительных навыков. Отсутствие бесплатного режима при сравнительно высокой стоимости вычислительных ресурсов

	Преимущества	Недостатки
Amazon SageMaker	Наличие готовых блок-схем для обучения и визуализации данных	Небольшое сообщество Недостаточно мануалов. Недоступен в России
IBM Watson	Специализация на задачах компьютерного зрения и NLP Наличие утилит по предобработке данных	Узкая специализация. Полностью недоступен для России
Oracle Cloud	Мощный GPU ресурс на рынке со сравнительно низкой стоимостью. Удобная среда Oracle SQL. Высокая скорость обучения моделей	Непрозрачная тарификация. Недоступен в России
Alibaba Cloud	Доступен из России при соблюдении ряда условий	Сложная регистрация. Небольшое сообщество. Документация частично на китайском языке
Deepnote	Доступен в России. Есть бесплатная версия. Удобная SQL-среда. Возможность гибкого масштабирования	Крайне ограниченный вычислительные ресурсы в бесплатной версии. Риски доступности в силу санкционных ограничений, проблемы с оплатой
Datalore	Доступен в России. Есть бесплатная версия. Целостная интегрированная среда с SQL и Mongo клиентами	Крайне ограниченные вычислительные ресурсы в бесплатной версии. Риски доступности в силу санкционных ограничений, проблемы с оплатой
H2O.ai	Высокая скорость обучения Открытый исходный код Интеграция с ChatGPT Наличие удобных API для разработки	Сложный интерфейс, высокий порог входа для неопытных пользователей. Риски доступности в силу санкционных ограничений, проблемы с оплатой

В настоящее время позиционирование платформы SimBA относительно основных конкурентов выглядит следующим образом:

Обзор существующих решений

К классу облачных ML-платформ относится и Simulation platform for Business Analytics and Big Data (или сокращенно SimBA).
Опыт Школы в области образовательных курсов показал, что в большинстве случаев либо используются личные ноутбуки обучающихся и вопросы настройки инструментария перекладываются на их плечи, либо студентам предлагается использовать распространенные и доступные облачные сервисы. Чаще всего это Google Colab или Kaggle, и это во многом тот же способ решать вопросы настройки и доступа к нужным инструментам за счет усилий самих студентов. Еще одним недостатком подобного подхода является невозможность воспроизвести нужную рабочую среду и процесс получения данных.

Отдельно необходимо рассмотреть вопрос организации исследовательских проектов и лабораторий по работе с данным индустриальных партнеров Школы, где вопрос доступа к вычислительным ресурсам, а также вопросы управления доступом к данным, уже невозможно решить за счет самих участников, которые не имеют в своем распоряжении нужный объем ресурсов и не всегда обладают необходимыми компетенциями по настройке рабочего окружения.

Преимущества иностранных платформ очевидны – опыт эксплуатации, разнообразие инструментов, удобство получения данных через встроенные инструменты интеграции внутри облачных платформ, ресурсы для масштабирования вычислительных мощностей, наличие большого сообщества (примеры – Kaggle, Google Cloud, Microsoft).
Однако риски также очевидны – это проблемы с оплатой, зависимость от политической конъюнктуры, риски отключения.

Реализация образовательных программ и курсов, а также организация проектов в области работы с данными в настоящее время требуют платформенных решений для создания обучающимся и участникам проектных команд удобной и производительной среды.
Вместе с тем, рассмотренные выше и представленные на рынке решения в той или иной мере ограничивают возможности гибкого преподавания широкого спектра дисциплин по программированию, работе с данными и анализу данных и требуют либо значимых финансовых затрат, либо специфических знаний и навыков как обучающихся, преподавателей и поддерживающих процессы ИТ-специалистов.

Платформа

Рассматриваемое в данном документе решение опирается на следующие основные принципы:

Использование современных фреймворков

являющихся де-факто индустриальными стандартами для работы с данными и одновременно State-of-the-Art (SotA) подходами в отрасли

Низкий порог входа для пользователей

в том числе для тех, кто не имеет глубокого технического бэкграунда

Эффективное управление затратами

на разворачивание и поддержание платформы с учетом необходимости быстрого масштабирования как в терминах количества пользователей, так и добавления новых фреймворков

Широкий спектр современных фреймворков

Для доступа пользователей к широкому спектру современных фреймворков применяется контейнеризация, таким образом достигается возможность использования в образовательных и исследовательских задачах целого ряда решений. Как пример, на платформе можно реализовать изучение:

- Основ работы с Unix операционными системами
- Использования Git, bash/shell в разработке
- Основ работы с реляционными базами данных (БД), изучение SQL (PostgreSQL)
- Нереляционных NoSQL БД (MongoDB, HBase, ClickHouse)
- Фреймворков для массово-параллельной обработки больших данных Hadoop, (Map-Reduce, Apache Spark)
- Специализированных инструментов управления процессами моделирования и обработки данных (MLflow, Apache Airflow)
- Машинного и глубокого обучения, анализа и визуализации данных (с использованием языков программирования Python, R, Julia и соответствующих модулей)

Модульный подход на основе контейнеризации позволяет гибко дополнять или менять перечень доступных на платформе инструментов, дополнять платформу новыми фреймворками под определенные курсы или проекты.

Платформа

Низкий порог входа

для пользователей на платформе стал возможен за счет использования в качестве IDE интерактивных ноутбуков Jupyter / JupyterLab с возможностью доступа к ним через сеть Интернет с использованием обычного браузера без установки на персональный компьютер каких-либо дополнительных приложений. При этом интерфейс ноутбуков максимально прост, не содержит избыточного набора элементов управления и надстроек, обладая при этом достаточной функциональностью для решения задач по программированию, анализа и обработки данных.

Интерактивность ноутбуков позволяет пользователям видеть результат выполнения кода непосредственно на рабочем экране сразу за выполняемыми строками кода, что существенно упрощает пользовательский опыт для начинающих пользователей.

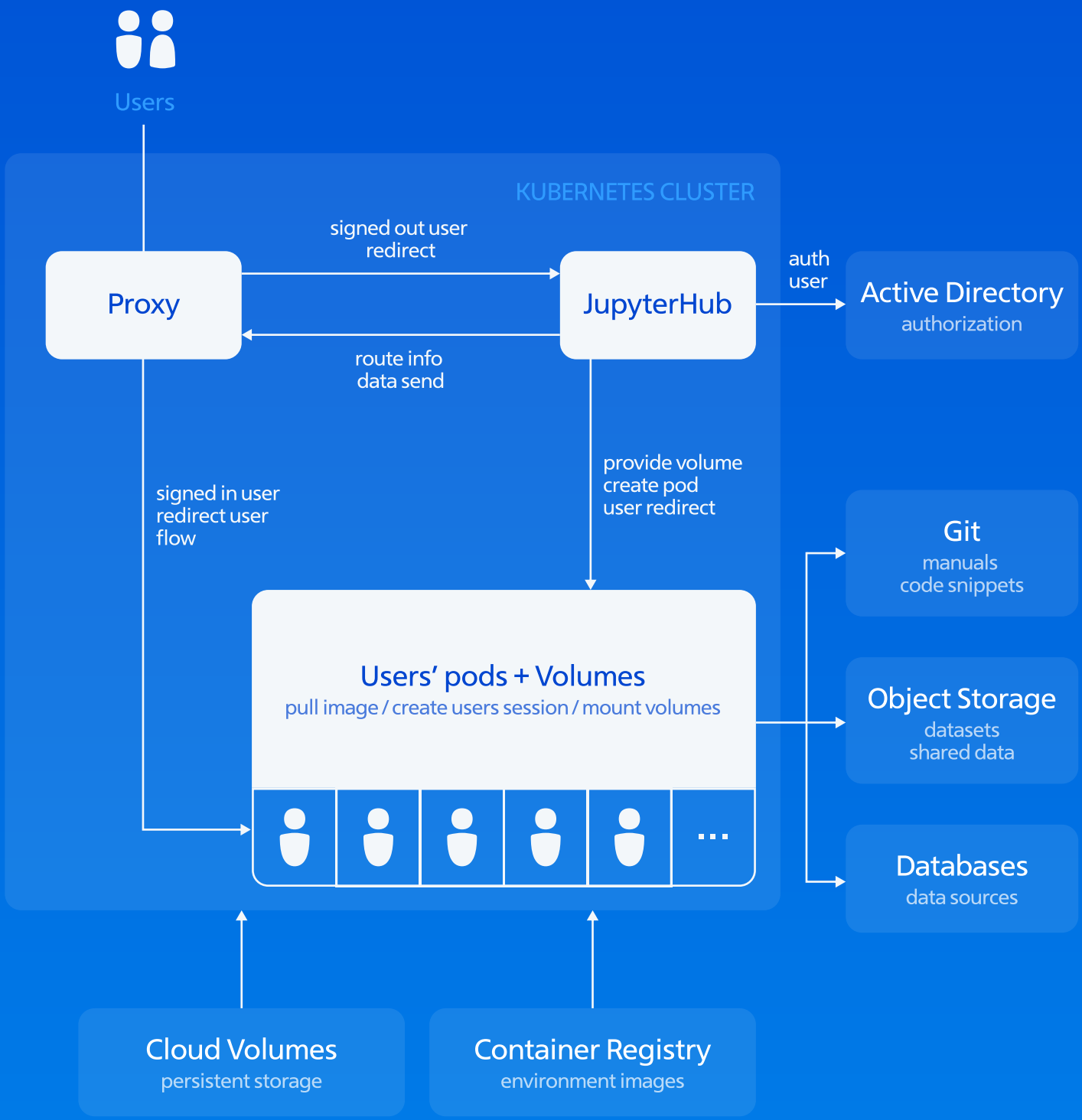
Простой вход и настройка прав на платформе могут быть достигнуты за счет интеграции с уже существующими в образовательном учреждении средствами авторизации или через взаимодействие со сторонними сервисами. Пользователям в зависимости от настроек могут быть выданы доступы к данным или дополнительным окружениям внутри платформы, а также более мощным вычислительным конфигурациям.

Эффективное управление затратами

на работу платформы становится возможным за счет ее разворачивания на облачных ресурсах (например, Yandex Cloud или VK Cloud) с использованием Kubernetes. Такое решение делает возможным гибкое управление вычислительными ресурсами – например, быстрое добавление вычислительных мощностей при наплыве пользователей и выключение ненужных ресурсов при их оттоке. Это свойство наиболее важно для образовательных учреждений, для которых вычислительные ресурсы нужны только на время проведения занятий или при выполнении студентами домашнего задания или исследовательских проектов.

Разворачивание платформы на Kubernetes кластере позволяет также гибко подключать различные варианты вычислительных ресурсов в различных конфигурациях, под специализированные задачи. Например, для подключения пользователей к внешним источникам данных, в том числе с реализацией разграниченного доступа к чувствительным данным.

Архитектура решения строится на использовании Managed Kubernetes кластера, с развернутым на нем JupyterHub. Пользовательские среды при запуске используют Docker контейнеры, собранные под определенные задачи, пользовательские данные хранятся в объектном хранилище. Авторизация пользователей основана на интеграции с Active Directory учебного заведения. Принципиальная архитектурная схема решения представлена на диаграмме ниже.



Платформа

Реализация решения в соответствии с выбранной архитектурной схемой выдвигает определенные технические и организационные требования.

Организационные требования

Предусматривают наличие специалистов, с компетенциями в области:

— Разворачивания и управления платформой (DevOps) с учетом указанного в технических требованиях стека технологий (Kubernetes, Docker, базы данных и фреймворки обработки данных), а также CI/CD инструментария (Helm, kubectl)

— Администрирования решения на уровне управления пользователями и пользовательскими ресурсами с использованием встроенных инструментов JupyterHub, инструментария для настройки прав доступа (OAuth, Active Directory)

— Поддержки пользователей с учетом специфики решаемых на платформе задач и соответствующего пользовательского опыта

В зависимости от масштабов платформы, количества пользователей и численности ИТ подразделения, обеспечивающего разворачивание и сопровождение решения, компетенции могут быть совмещены в пределах одного-двух специалистов.

Технические требования

В целом ограничиваются возможностью доступа к облачным средам, предоставляющих для использования:

— Managed Kubernetes сервисы для разворачивания JupyterHub

— Container Registry для хранения docker-образов, которые используется при запуске непобедимых сред

— Object Storage для создания пространства хранения данных

— Репозиторий для хранения конфигурационных файлов (возможно использование GitHub для этих целей)

Примерами таких облачных сред могут служить **VK Cloud** или **Yandex.Cloud**. Для реализации доступа через OAuth необходимо наличие на стороне организации решения, с которым можно осуществить интеграцию для авторизации пользователей.

Платформа

Пример реализации в ВШМ СПбГУ

Ниже приведен пример реализации платформы для моделирования задач бизнес-аналитики и анализа данных на облачной инфраструктуре в Школе с учетом особенностей образовательного процесса в бакалавриате и магистратуре. Другое название платформы – Simulation platform for Business Analytics and Big Data (SimBA).

Для пользователей платформы доступны следующие варианты рабочих сред или окружений:

Data Science environment – основные библиотеки для задач обработки данных и машинного обучения на языке Python.

Minimal Python environment – окружение для тренировки навыков разработки на Python, с возможностями дополнительного использования Unix среды.

Spark environment – для работы с большими данными и использования высокопроизводительного фреймфорка Apache Spark.

PostgreSQL environment – для работы с базой данных PostgreSQL.

MongoDB – для работы с базой данных MongoDB.

Airflow environment – для изучения фреймворка Apache Airflow.

Hadoop (with YARN) and Spark environment – для изучения возможностей платформы Apache Hadoop, Map-Reduce операций и Apache Spark.

R environment – изучение экосистемы языка R.

GPU environment – окружение с доступными графическими ускорителями для высокопроизводительных задач.

Платформа развернута в Яндекс.Облаке с использованием сервисов:

Managed Service for Kubernetes – для разворачивания кластера вычислительных ресурсов с неограниченным количеством узлов. Один рабочий узел (нода) имеет 16 vCPU, 64 ГБ RAM. 96 ГБ SSD и в зависимости от настрое позволяет разместить от 3 до 10 пользователей.

Object Storage – для хранения пользовательских данных и датасетов для образовательных курсов или совместной проектной работы.

Container Registry – для хранения образов, на основе которых запускается требуемое пользовательское окружение.

Compute cloud – для разрешения виртуальных серверов под отдельные задачи по работе с данными в рамках отдельных образовательных активностей. На серверах могут быть установлены различные решения, такие как PostgreSQL, ClickHouse, Apache Atlas и др.

Cloud Logging – для мониторинга активности пользователей на платформе и оценки ее загруженности.

Платформа

Пример реализации в ВШМ СПбГУ

В зависимости от окружения пользователям доступны вычислительные ресурсы в конфигурациях 3-5 vCPU, 10-16 ГБ RAM, 12 ГБ SSD. Сверх этого могут быть выделены дополнительные вычислительные ресурсы под вычислительно сложные проекты, дополнительное дисковое пространство для хранения большого объема данных (от 1 ТБ).

Предусмотрена интеграция с Active Directory Школы, что с одной стороны позволяет пользователям входить на платформу под своей учетной записью в Школе, а с другой – идентифицировать пользователей и назначать им права и доступы к чувствительным данным.

На платформе реализуются следующие курсы:

- Data-driven Decision Making for Product Managers
- Machine Learning and Big Data Analysis
- Deep Learning for Business Applications
- Introduction to Big Data Modern Technologies
- Development Essentials
- Applied mini-project on Industrial Datasets
- Mastering Applied Skills in Management, Analytics and Entrepreneurship
- Цифровые инструменты для менеджеров

Также платформа используется студентами и преподавателями Школы в различных проектах по анализу данных, машинному и глубокому обучению.

Платформа

Если остановиться на подходах к внедрению платформы SimBA в образовательный процесс, то следует отметить такие ключевые аспекты как:

- Наличие в учебном заведении профильных образовательных программ по аналитике данных, программированию, машинному обучению или направлений, где платформа могла бы быть востребованной и где проще всего начать ее внедрение
- Квалификация преподавателей, их опыт в разработке ИТ решений, использовании современных инструментов в разработке или анализе данных
- Готовность преподавателей к использованию элементов платформы в своих курсах и проектах, участие преподавателей в развитии платформы
- Формализация использования платформы через указание платформы как инструмента в рабочих программах курсов
- Создание доступных пользовательских инструкций и руководств, интерактивных ноутбуков (образцов кода) для демонстрации возможностей платформы
- Развитие сообщества разработчиков и пользователей платформы внутри образовательного учреждения для обмена опытом и выработки предложений по улучшению платформы

В завершение следует отметить, что в зависимости от практических задач в области образования, анализа данных, совместной проектной работы, платформа может быть дополнена следующими элементами:

- Модуль автоматической проверки заданий по программированию, настраиваемый под тот или иной образовательный курс
- Инструменты low-code и no-code для дальнейшего снижения порога вхождения пользователей и разработки новых образовательных курсов
- Модули интеграции с базами данных на основе приложений-клиентов с дружественным пользовательским интерфейсом для изучения работы с современными базами данных
- BI клиенты с расширенными возможностями по визуализации данных

Заключение

Настоящая книга содержит описание платформы для моделирования задач бизнес-аналитики и анализа данных (платформа SimBA) от положенных в основу ее разработки принципов до примера практической реализации.

Платформа SimBA является эффективным и удобным инструментом для решения образовательных задач, может быть встроена в учебных процесс и использована для курсов по разработке и программированию, статистике и визуализаций данных, продвинутой аналитике (эконометрика, машинное обучение), курсы по работе с базами данных и инструментами обработки больших данных. Платформу также можно использовать в качестве инструмента командной работы над проектами в области анализа данных.

Несомненными преимуществами платформы SimBA являются:

— Низкий порог входа для слушателей без технологических навыков

— Возможность настройки необходимого технологического стека как для продвинутых пользователей, так и для курсов начального уровня

— Изучение SoTA (state-of-the-art) технологий анализа и хранения данных

— Управление доступными пользователям ресурсами и настройка доступа пользователей к размещаемым на платформе данным

Данный материал позволит читателю ознакомиться с платформой SimBA, оценить новые возможности и перспективы, которые могут стать доступными при использовании этого решения. Мы надеемся, что этой информации было достаточно для понимания функционала SimBA и возможности ее применения для ваших уникальных задач.

Терминология

Термин	Пояснения
IDE	Integrated Development Environment или интегрированная среда разработки, комплекс программных средств, используемый пользователями для работы над программным кодом
Школа	Высшая школа менеджмента СПбГУ
MiBA	Магистерская программа Высшей школы менеджмента СПбГУ «Бизнес-аналитика и большие данные» (Master in Business Analytics and Big Data)
Интерактивный ноутбук	Веб-приложение для создания и совместного использования документов с программным кодом. Гибкий интерфейс позволяет пользователям создавать код, запускать его ноутбука и получать результаты исполнения кода непосредственно в интерфейсе ноутбука. Эти особенности сделали данный класс приложений популярными в задачах data science, машинного обучения и анализа данных. Jupyter Notebook является в настоящее время самым распространенным вариантом такого рода приложений
SotA	State of the Art термин относится к наивысшему в данный момент уровню развития устройства, технологии или научной области
БД	База данных, структурированная коллекция информации, которую можно легко обрабатывать, управлять и обновлять
GPU	Graphic Processor Unit или видеокарта. Класс устройств, использующих высокопроизводительные графические процессоры и применяемые в настоящее время не только для решения задач в области графики, но и для осуществления высокопроизводительных вычислений, основанных на матричных операциях
SQL (Structured Query Language)	Стандартизированный язык баз данных для управления и обработки данных
Big Data	Огромные объемы информации, обрабатываемые и анализируемые для получения ценных знаний и информации
Машинное обучение (Machine Learning)	Методология и подход в области искусственного интеллекта, которая позволяет системе автоматически обучаться и улучшать свою работу на основе выявления закономерностей в наборе данных
NLP	Natural Language Processing, область машинного обучения, относящаяся к задачам обработки естественного языка, работы с текстом
No-code, low -code	Подход к разработке информационных систем, когда отсутствие необходимость в написании кода, вместо этого система может быть усилиями аналитика или продвинутого пользователя через взаимодействие со специализированным графическим интерфейсом
Открытый исходный код (Open source)	Тип программного обеспечения, исходный код которого доступен для общественного использования и модификаций

Термин	Пояснения
Облачные вычисления (Cloud computing)	Использование удаленных серверов в интернете для хранения, управления и обработки данных, вместо локального сервера или персонального компьютера
API (Application Programming Interface)	Набор процедур, протоколов и инструментов для обмена данными между различными информационными система и приложениями
Распределенная система	Система, в которой компоненты находятся на сетевых компьютерах, которые взаимодействуют и координируют свои действия только путем передачи сообщений
Кластер (Cluster)	Группа связанных компьютеров, работающих вместе, чтобы улучшить производительность или надежность.
MongoDB	Кроссплатформенная документо-ориентированная база данных с открытым исходным кодом. Она отвечает требованиям NoSQL и использует формат JSON-подобных документов и схемы
NoSQL	Подход к проектированию баз данных, который обеспечивает гибкость, масштабируемость и производительность, которых может не хватать в традиционных базах данных
Hadoop	Открытый фреймворк для обработки большого количества данных на кластерах компьютеров с простыми моделями программирования
Docker	Программное обеспечение с открытым исходным кодом, которое автоматизирует развертывание, масштабирование и управление приложениями
Контейнеризация	Метод виртуализации, при котором приложения и их зависимости объединяются в один пакет, что упрощает развертывание и управление
Kubernetes	Открытая платформа для автоматизации развертывания, масштабирования и управления приложениями в контейнерах
BI (Business Intelligence)	Технологии, приложения и практики для сбора, интеграции, анализа и представления бизнес-информации для поддержания более эффективного принятия решений
Spark	Программный продукт для массово-параллельной обработки больших данных, предназначенный для эффективной, удобной и быстрой аналитики
MLflow	Открытая платформа для управления полным циклом жизни машинного обучения. Она обеспечивает работу с моделями, включая их экспериментирование, воспроизводимость и развертывание

Дополнительные материалы

Ниже приведен перечень источников, которые были использованы при разработке платформы и написании данной книги:

— Экосистема Jupyter Ссылка

— JupyterHub как многопользовательский подпроект Jupyter Ссылка

— JupyterHub на кластере Kubernetes Ссылка

— Репозиторий JupyterHub на кластере Kubernetes Ссылка

— Репозиторий платформы SimBA (инструкция пользователя) Ссылка

— Репозиторий платформы SimBA (инструкция администратора) Ссылка

— Высшая школа менеджмента СПбГУ Ссылка

— Магистерская программа Высшей школы менеджмента СПбГУ «Бизнес-аналитика и большие данные» Ссылка

— JupyterHub, или как управлять сотнями пользователей Python. Лекция Яндекса Ссылка

— Teaching and Learning with Jupyter Ссылка

— Managing a 1,000+ Student JupyterHub without Losing Your Sanity Ссылка

— Stanford Seminar - Jupyter Notebooks and Academic Publication Ссылка

Эксперты

Гаршин Василий

Управляющий директор

ПАО Банк ВТБ

Идеолог и разработчик платформы SimBA

Опыт реализации проектов по цифровой трансформации (направление Big Data, ML, AI)

Развитие магистерской программы ВШМ СПбГУ «Бизнес-аналитика и большие данные»

Преподаватель ВШМ СПбГУ

Авторы

Поликарпов Владислав Аркадьевич


ВТБ