

Программа курса

«BDAM: Аналитика больших данных для руководителей»

О курсе: Все, что нужно знать для успешной работы с большими данными: методы аналитики и машинного обучения, принципы работы и функциональные возможности компонентов экосистемы Hadoop, объектных S3-like хранилищ, безопасность озера данных, цифровизация бизнеса.

Аудитория: Руководство государственных предприятий и частных компаний, менеджеры и специалисты.

Уровень подготовки: предварительный опыт не требуется.

Соотношение теории к практике 50/50

Продолжительность курса: 24 академических часа, 5 дней

Содержание программы

1. Введение в Big Data (Большие данные)

- Большие данные и цифровизация бизнеса.
- Характеристики data-driven организации. Data-driven и data-informed организации: основные отличия.
- Принятие data-driven решений и путь к ценности данных.
- Базовые принципы и отличия от классических подходов к работе с данными.
- Сравнительный анализ фреймворков управления проектами анализа больших данных CRISP DM, Domino, TDSP, SEMMA, Enterprise Big Data Framework.
- Отраслевая специфика аналитики больших данных. Общий разбор отраслевых сценариев (cases).
- Тенденции в подходах Big Data и что актуально на сегодняшний день.
- Технологии Big Data в условиях импортозамещения.

2. Правовые аспекты организации защиты персональных данных

- Правовое регулирование в области защиты персональных данных
- Права субъекта и обязанности оператора при обработке ПД
- Нарушители безопасности ПД
- Государственные регуляторы и их нормативно-правовая документация в области защиты ПД
- Осуществление проверки соблюдения правил в области защиты ПД и ответственность за их нарушение
- GDPR
 - **Практическое задание** — постановка целей для функции Big Data в вашей организации, выявление зон и способов развития организации на пути к data-driven.

3. Понимание Бизнеса (Business Understanding)

- Влияние Big Data на бизнес. Правильные вопросы бизнесу.
- Определение бизнес целей для проекта Big Data.
- Этапы жизненного цикла проекта анализа больших данных, модель DSCPLC.
- Инициация проекта — критические факторы успеха. Основные проблемы.
- Оценка ситуации: риски, ROI, IRR, доступные ресурсы, оценка зрелости компании.
- Приоритизация задач: Что делаем, а что нет.
- Высокоуровневый план проекта.
 - **Практическое задание** — выявление проблем, с которыми может столкнуться организация при реализации Big Data проектов, постановка вопросов бизнесу.

4. Понимание данных (Data Understanding)

- Проблемы при работе с большими данными, модель 4V's
- Определение источников данных.

- Специфика работы с потоковыми и пакетными данными в Big Data.
- Принципы формирования Data Lake: выбор платформы.
- Первичный сбор и анализ данных: инструментарий и доступные методы.
- Описание данных и сбор метаданных.
- Data management и Data Governance.
- Оценка качества данных Data Quality.
- Основные трудности и проблемы, критические факторы успеха, роли и навыки.
- Разбор сценариев (cases) для фазы Data Understanding.

5. Подготовка данных (Data Preparation)

- Подготовка данных — подходы Data Science: нормализация, очистка, выборки, enrichment, форматирование данных.
- Подготовка данных — как процесс формирования Data Pipeline:
- Процессы ETL и ELT,
- Зонирование Data Lake и сегментирование данных,
- Назначение и сравнительная характеристика компонент экосистемы Apache Hadoop, NoSQL, DWH, платформ потоковой обработки для хранения и обработки Big Data на примерах (отраслевые сценарии),
- Инструменты оркестрирования и автоматизации (Data Flow).
- И снова о Метаданных: Data Lineage, Data Provenance, Data Governance, ...
- Безопасность больших данных.
- Основные трудности и проблемы, критические факторы успеха.
- Специалисты и их компетенции на данной стадии.
 - **Практическое задание** — выявление факторов и рисков реализации Big Data проектов в вашей организации; идентификация датасетов, необходимых для исследования в рамках Big Data проектов; анализ стратегий управления данными с учетом челенджей больших данных в соответствии с моделью 4V's; выявление зон развития функции управления большими данными.

6. Выбор и построение моделей (Modeling)

- Классы аналитических задач и подходы к их решению.
- Обзор техник моделирования.
- Построение моделей и оценка моделей.
- Что нужно для успешного моделирования.
- Инструментарий для решения аналитических задач этапа моделирования.
- Оценка моделей и среда тестирования моделей: технические метрики оценки качества проведенного моделирования.
- А также песочницы, Machine & Deep learning, AI, Нейронные сети и многое другое.
- Основные трудности и проблемы фазы моделирования, критические факторы успеха.
- Облачные платформы для быстрой разработки.
- Место DevOps, MLOps для организации процессов разработки.
- Рассмотрение фазы моделирования на сквозных сценариях (cases): место, участники, взаимодействие с другими фазами и процессами.

7. Построение команды Data Science

- Таксономия Data Science. Система OSEMN.
- Команда Data Science: специалисты и их компетенции, роли. Таксономия Data Science. Система OSEMN.
- Техники найма команды Data Science. Удержание и развитие талантов в команде Data Science
 - **Практическое задание** — построение команды Data Science в вашей организации, определение ключевых ролей и навыков команды, разработка стратегии развития и удержания команды.

8. Оценка результатов (Evaluation)

- Бизнес-метрики оценки качества моделирования.
- Отличие Data-driven метрик от традиционных.
- Операционные метрики. Vaniti метрики. Фрэймворк AARRR.
- Оценка качества моделирования.

- Что делать если все плохо? – возвращаемся на предыдущие фазы.
- Рассматриваем данную стадию в рамках наших сценариев (cases).
- Отличия среды разработки и эксплуатации.
- Особенности этапа оценки.

9. Развертывание (Deployment)

- Планирование развертывания модели.
- Мониторинг и обслуживание модели.
- Методологии автоматизации и вывода продуктов в промышленную эксплуатацию – DevOps и MLOps.
- Задачи финального обеспечения фаз жизненного цикла Data Science: цикличность reviews, обновления и вывод из эксплуатации.
 - **Практическое задание** — выявление качественных и количественных критериев успеха для Big Data-проектов.

10. Экономическая оценка проекта

- Метрики экономической эффективности проекта
- Принципы разработки бизнес-кейсов для проектов анализа больших данных. Типовые ошибки при разработке бизнес-кейсов. Разбор примеров бизнес-кейсов.
- Защита бизнес-кейсов и бюджета проекта

11. Цифровая трансформация организации

- Факторы цифровой трансформации организации
- Этапы цифровой трансформации. Уровни зрелости организации
- Как не повторить «ловушки стартапов». Челенджи на пути к цифровой трансформации
- Критерии оценки зрелости аналитики данных в организации
 - **Практическое задание** — разработка бизнес-кейсов для Big Data-проектов, разработка плана цифровой трансформации вашей организации.