

**Создание DataOps-конвейера для работы с данными в государственном секторе
Шенцов Д.И. (ИТМО)**

**Научный руководитель – доктор технических наук, профессор Марусина М.Я.
(ИТМО)**

Введение. В условиях активной цифровизации государственного управления обработка значительных объёмов данных приобретает особую значимость. Эффективное принятие решений во многом определяется скоростью и качеством подготовки информации, что предполагает внедрение автоматизированных инструментов и регламентов по работе с данными. В данной статье освещаются принципы построения DataOps-конвейера в государственном секторе, а также ключевые технологические и методологические аспекты, которые необходимо учитывать при его интеграции в существующие информационные системы. Отдельное внимание уделяется вызовам при масштабировании данной модели и преимуществам, которые она предоставляет при взаимодействии с разнородными источниками данных.

Основная часть. В ходе проведённого исследования был рассмотрен подход к автоматизации работы с государственными данными в формате DataOps-конвейера, охватывающего ETL-процессы, процедуры контроля качества и последующую аналитику в режиме непрерывного обновления. Выявлено, что в традиционных государственных информационных системах часто встречаются низкая адаптивность и высокая задержка обновления, что подтверждается данными ряда работ [1, 2].

Для оценки эффективности DataOps-подхода были выполнены экспериментальные запуски с набором данных из различных открытых обезличенных баз данных государственных органов. По результатам апробации удалось добиться значительного ускорения и универсализации процесса подготовки данных для дальнейшей аналитики и построения ML/AI-моделей. Достижение подобных показателей стало возможным благодаря системной автоматизации рутинных операций, разработки единого алгоритма работы с данными.

Внедрение комплексного контроля качества позволило снизить процент несоответствий в итоговых наборах данных, что подтверждает выводы о надёжности подобных решений в публичном секторе. Подобные механизмы особенно актуальны при необходимости оперативного формирования отчётности для управленческих целей [3].

Важной частью работы стал анализ операционной эффективности государственных учреждений при переходе к DataOps-модели. Согласно результатам внедрения пилотных проектов, автоматизация обработки данных способна уменьшить затраты на инфраструктуру благодаря более гибкому распределению вычислительных ресурсов. Подобные практики оказывают положительный эффект на бюджетные параметры и позволяют высвободить средства для дальнейшего развития информационных сервисов, что подтверждается в том числе результатами других исследований [4, 5].

Выводы. В работе рассмотрено внедрение DataOps-конвейера в государственном секторе и его влияние на оптимизацию управления данными. DataOps позволяет ускорить обработку данных, повысить их точность и снизить затраты на аналитику за счет автоматизации и интеграции с DevOps-подходами. Однако процесс внедрения сталкивается с организационными и технологическими барьерами, такими как сопротивление изменениям и нехватка специалистов. Для успешной реализации необходимы стандартизация процессов, развитие культуры работы с данными и усиление контроля качества. Перспективы развития DataOps связаны с интеграцией облачных технологий, ИИ

и AutoML, что позволит создать гибкие и интеллектуальные системы управления данными в госуправлении. Перспективы дальнейших исследований связаны с масштабированием технологии в разных ветвях государственного управления и разработкой дополнительных модулей для глубокой проверки качества и усиления безопасности данных.

Список использованных источников:

1. Петров А.Ю., Саифуллина Л.Д. Информационные технологии в государственном управлении // Евразийский Союз Ученых. – 2017. – № 11-2 (56). – С. 36–37
2. Чанов С.В. Государственные информационные системы: риски и угрозы использования // Взаимодействие власти, бизнеса и общества в развитии цифровой экономики. – Саратов, 2018. – С. 163–166.
3. Борисяк, М.А. Методы машинного обучения для контроля качества данных в научных экспериментах: дис. ... канд. комп. наук / Нац. исслед. ун-т «Высшая школа экономики». – Москва, 2020. – 116 с.
4. Боллини, С. DataOps: применение принципов DevOps для управления данными / S. Bollineni // Journal of Artificial Intelligence, Machine Learning & Data Science. – 2023. – Т. 14, № 4. – С. 1285-1288. – DOI: 10.51219/JAIMLD/satyadeepak-bollineni/293.
5. Yin Z., Zhou S., Zhou J., Tian M., Lin M., Liu S. Research on DataOps Capability - Practice and Development // Proceedings of the 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). – Exeter, United Kingdom, 2023. – С. 2170-2174. – DOI: 10.1109/TrustCom60117.2023.00303.