

Анисимов Галина Борисовна, кандидат физико-математических наук,
ст.н.с., доцент Донской государственной технической университет,

anisimovagalina@mail.ru

Валуйский Максим Сергеевич, магистрант, Донской государственной
технической университет, pochta_my_live@mail.ru

Минаева Наталья Евгеньевна, магистрант, Донской государственной
технической университет, minaeva.natali2012@yandex.ru

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ МИГРАЦИИ ДАННЫХ МЕЖДУ ОБЛАКАМИ И ЛОКАЛЬНЫМИ СИСТЕМАМИ

Аннотация: В данной статье рассматривается проблема автоматизации процессов миграции данных между облаками и локальными системами, анализируются существующие подходы, их преимущества и недостатки, предлагается концептуальная модель автоматизированного процесса миграции данных, основанная на применении инструментов ETL и оркестрации, описываются особенности тестирования облачных баз, данных при миграции. В заключении обсуждаются перспективы развития автоматизации процессов миграции данных в контексте растущей сложности и разнородности ИТ-инфраструктур.

Ключевые слова: автоматизация процессов, миграция данных, стратегия миграции, облачные вычисления, облачная база данных, локальная система, оркестрация, ETL, Kubernetes, DataOps.

Abstract: This article examines the problem of automating data migration processes between clouds and local systems, analyzes existing approaches, their advantages and disadvantages, suggests a conceptual model of the automated data migration process based on the use of ETL and orchestration tools, describes the features of testing cloud databases during migration. In conclusion, the prospects for the development of automation of data migration processes in the context of the growing complexity and heterogeneity of IT infrastructures are discussed.

Keywords: process automation, data migration, migration strategy, cloud computing, cloud database, on-premises system, orchestration, ETL, Kubernetes, DataOps.

Введение

В современном мире, характеризующемся экспоненциальным ростом объёмов данных и возрастающей потребностью в гибкости и масштабируемости ИТ-инфраструктуры, миграция данных становится критически важной задачей для многих организаций. Миграция данных может быть обусловлена различными факторами, такими как переход на облачные платформы, консолидация центров обработки данных (ЦОД), обновление устаревших систем, внедрение новых бизнес-приложений и обеспечение соответствия нормативным требованиям (например, Федеральному закону №152-ФЗ «О персональных данных»).

Традиционные подходы к миграции данных, выполняемые вручную или с использованием скриптов, часто оказываются неэффективными, трудоёмкими, подверженными ошибкам и дорогостоящими. Это связано с необходимостью выполнения большого количества рутинных операций, сложностью преобразования данных между различными форматами и структурами, а также риском потери данных или нарушения их целостности. Автоматизация процессов миграции данных позволяет значительно сократить время и затраты на выполнение миграции, повысить качество данных, снизить риски ошибок и обеспечить большую гибкость и управляемость процесса.

1. Подходы и технологии автоматизации миграции данных

В сфере автоматизации миграции данных существуют четыре основных подхода:

1. Использование ETL-инструментов, обеспечивающих широкий функционал при высокой стоимости [1].

2. Применение облачных сервисов миграции, отличающихся простотой и масштабируемостью, но ограниченной кастомизацией [2, 6].

3. Разработка собственных скриптов, предоставляющих полный контроль ценой высокой трудоёмкости [3].

4. Использование инструментов оркестрации для централизованного управления процессами миграции [4].

При миграции в облачные среды Kubernetes требуется автоматизация доступа к базам данных, что решается запуском временных контейнеров для получения данных из Vault и подготовки баз данных [5]. Для безопасного хранения данных в облаке применяются решения с локальным шифрованием по алгоритму AES, включающие установление соединения с хранилищем и выбор способа взаимодействия [2].

Стратегии миграции в облако включают: Lift-and-Shift (минимальные изменения архитектуры), Re-platforming (частичное использование облачных функций) и Application Rewrite (полная переработка приложений) [6]. Для оптимизации вычислений разработана структура PMCO, позволяющая экономить до 44% времени выполнения и 84% энергии за счёт прозрачного переноса процессов с мобильных устройств на вычислительную инфраструктуру [7].

2. Концептуальная модель автоматизированного процесса миграции данных, основанная на совместном применении решений

В данном разделе к рассмотрению предлагается концептуальная модель автоматизированного процесса миграции данных, основанная на совместном применении инструментов ETL и системы оркестрации. Такой подход, сочетающий специализированные инструменты ETL и централизованную систему оркестрации, позволяет достичь значительных преимуществ в гибкости, управляемости, масштабируемости и отказоустойчивости процесса миграции.

IDEF0 как часть методологии структурного подхода является эффективным инструментом для анализа. Декомпозиция IDEF0-модели даёт возможность подробно рассмотреть процессы функционирования, а также наглядно описать логику и взаимодействие процессов [8, 9].

На рисунках 1 и 2 представлены контекстная диаграмма процесса «Миграция данных» и диаграмма декомпозиции первого уровня, соответственно. Диаграммы концептуальной модели разработаны в соответствии с методологией IDEF0.

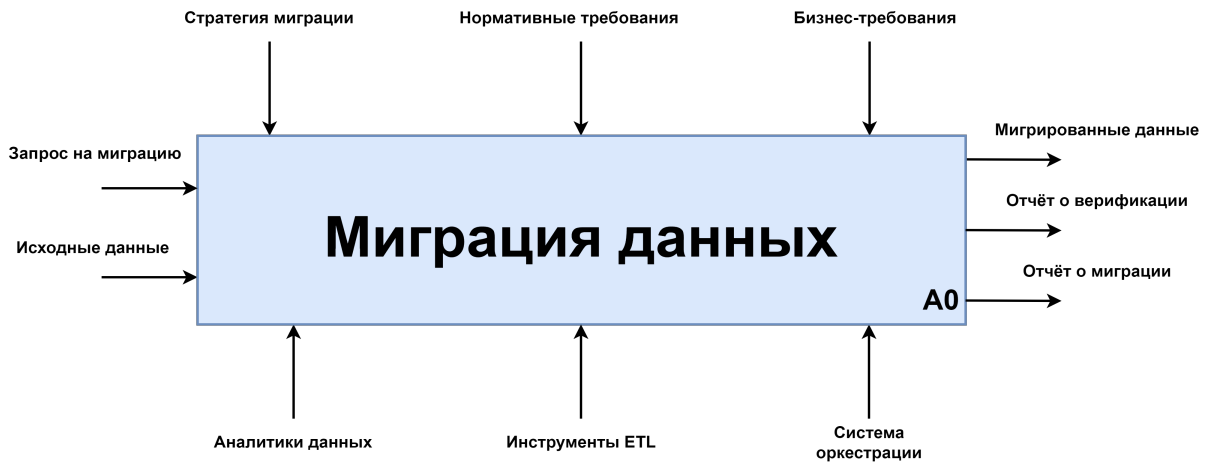


Рис. 1. — Контекстная диаграмма (A-0)

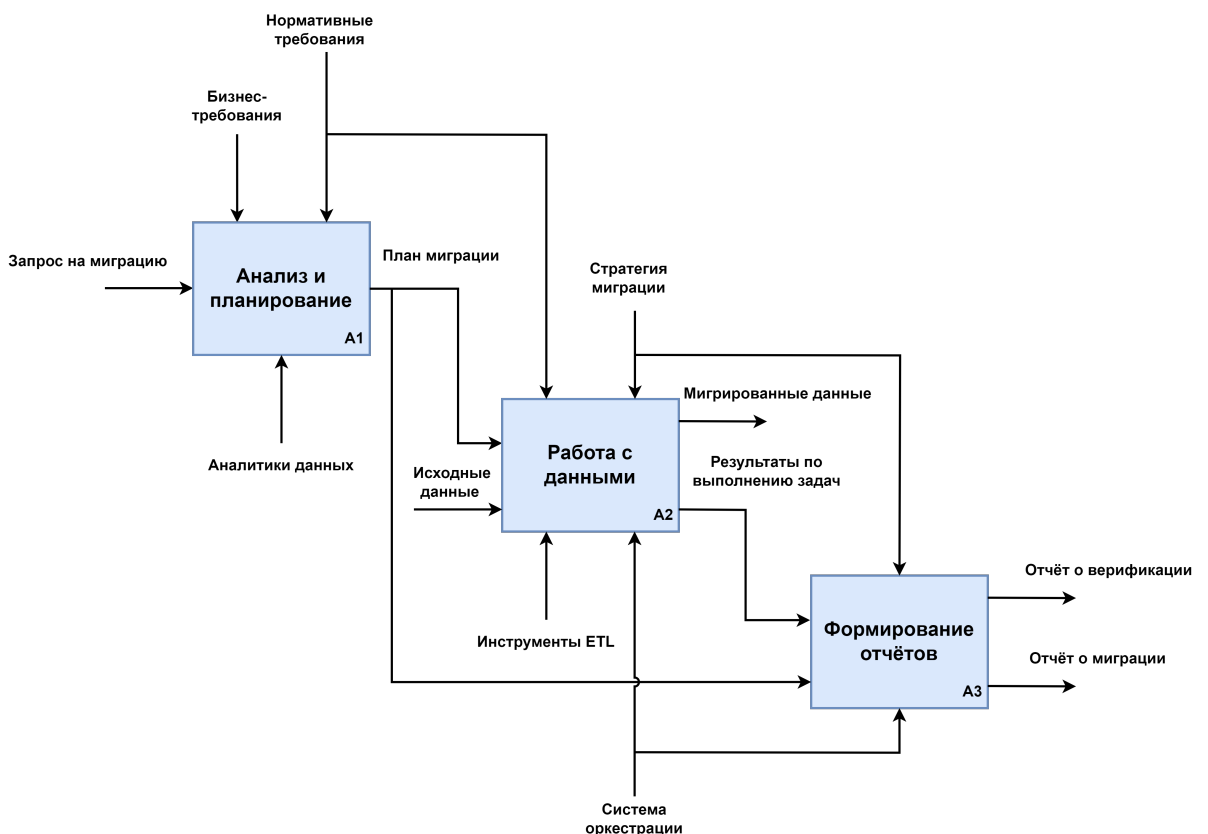


Рис. 2. — Диаграмма декомпозиции первого уровня (A0)

Основные функциональные блоки:

1. Анализ и планирование. На этом этапе проводится анализ исходных и целевых систем, определение требований к данным, разработка стратегии миграции, определение правил преобразования данных и планирование этапов миграции. Важным аспектом является оценка влияния миграции на бизнес-процессы и определение допустимого времени простоя системы.

2. Работа с данным. Включает функциональные блоки по извлечению данных, преобразованию данных, загрузки данных, верификации данных. Данные извлекаются из исходных систем с использованием различных методов, таких как прямое подключение к базам данных, чтение файлов, использование. Затем данные преобразуются в формат, совместимый с целевой системой (при этом автоматизированное преобразование данных выполняется с использованием специализированных инструментов ETL). Преобразованные данные загружаются в целевую систему. Важно обеспечить целостность данных, скорость загрузки и минимизировать влияние на работу целевой системы. Наконец, после загрузки данных проводится верификация данных для обеспечения их качества и соответствия требованиям.

3. Формирование отчётов. Отчётность предоставляет информацию о ходе и результатах выполнения миграции и качестве данных.

3. Особенности тестирования облачных баз данных при миграции

При миграции данных в облачные системы, такие как DBaaS Postgres Pro, необходимо проводить тщательное тестирование и контроль качества. Тестирование DBaaS является объёмным и сложным процессом из-за особенностей облачных баз данных и бизнес-логики. Авторами статьи также рассматривается пирамида автоматизации тестирования, включающая unit-тесты, компонентные тесты, интеграционные тесты и end-to-end тесты [10].

«Выделяется три основных подхода к автоматизации тестирования:

1. Тестирование на уровне кода (unit-тестирование).
2. Тестирование пользовательского интерфейса.
3. Тестирование API-интерфейса.» [10].

«При тестировании DBaaS Postgres Pro, рекомендуется использовать нижеследующие инструменты:

1. Playwright для тестирования UI.
2. Go и библиотеки Go для тестирования backend.
3. Allure TestOps для хранения тестовой документации и формирования отчетности.
4. Gitlab для CI/CD.» [10].

4. Перспективы развития автоматизации процессов миграции данных

Автоматизация процессов миграции данных является динамично развивающейся областью, в которой постоянно появляются новые инструменты и технологии. Перспективы развития автоматизации процессов миграции данных главным образом связаны с:

1. Облачные платформы предоставляют все более широкий спектр сервисов для миграции данных, которые позволяют автоматизировать процесс переноса данных между локальными системами и облаком, а также между различными облачными сервисами. В будущем ожидается дальнейшее развитие этих сервисов, включая поддержку новых форматов данных, более гибкие возможности настройки и кастомизации, и интеграцию с другими облачными сервисами.

2. Использование искусственного интеллекта и машинного обучения может значительно повысить эффективность автоматизации процессов миграции данных. В перспективе машинное обучение может быть использовано для автоматического определения правил преобразования данных, выявления аномалий в данных и прогнозирования времени выполнения миграции.

3. Инструменты оркестрации позволяют автоматизировать и координировать выполнение различных задач в процессе миграции данных, обеспечивая централизованное управление процессом миграции, параллельное выполнение задач, обработку ошибок и мониторинг выполнения

заданий. В будущем ожидается дальнейшее развитие этих инструментов, включая поддержку новых технологий и интеграцию с другими инструментами разработки и управления ИТ-инфраструктурой.

4. DataOps — это методология, которая применяет принципы DevOps к управлению данными. DataOps стремится к автоматизации, сотрудничеству и непрерывной интеграции и доставке в процессах, связанных с данными, включая миграцию. Внедрение DataOps позволяет ускорить и упростить процессы миграции, повысить качество и надёжность данных [11, 12].

Заключение

Автоматизация процессов миграции данных является критически важной задачей для многих организаций, стремящихся к повышению гибкости и масштабируемости ИТ-инфраструктуры. Существующие подходы к автоматизации, такие как использование инструментов ETL, облачных сервисов миграции данных и разработка собственных скриптов и приложений, имеют свои преимущества и недостатки. Предлагаемая в этой статье концептуальная модель автоматизированного процесса миграции данных позволяет значительно сократить время и затраты на выполнение миграции, повысить качество данных, снизить риски ошибок и обеспечить большую гибкость и управляемость процесса за счёт совместного использования инструментов ETL и оркестрации.

Стратегические векторы развития автоматизации процессов миграции данных определяются эволюцией гибридных и мультиоблачных систем, интеграцией когнитивных технологий машинного обучения, а также совершенствованием платформ оркестрации. Ключевым драйвером выступает концепция DataOps, которая трансформирует традиционные подходы за счёт внедрения принципов сквозной автоматизации, CI/CD-практик для конвейеров данных и коллаборации между командами разработки, аналитики и эксплуатации. Это формирует основу для самообучающихся систем миграции, способных адаптироваться к динамике инфраструктур и предиктивно устранять аномалии.

Литература

1. Баева В. Р., Дроздов А. Ю. ETL: Актуальность и применение. Преимущества и недостатки ETL инструментов // Международный научный журнал «ВЕСТНИК НАУКИ» № 5 (14) Т.4. URL: cyberleninka.ru/article/n/etl-aktualnost-i-primenenie-preimuschestva-i-nedostatki-etl-instrumentov/viewer.
2. Бусенков А. А., Багажков Д.И., Чернов В. В., Панов А. И., Башмуров Н. А. Разработка алгоритма и программная реализация средства защиты персональных данных в облачных хранилищах // «ИННОВАЦИИ И ИНВЕСТИЦИИ». 2021. №12. URL: cyberleninka.ru/article/n/razrabotka-algoritma-i-programmnaya-realizatsiya-sredstva-zaschity-personalnyh-dannyh-v-oblachnyh-hranilischah/viewer.
3. Харенслак Б., Руйтер дэ Д. Apache Airflow и конвейеры обработки данных: практическое руководство. Москва: ДМК Пресс, 2022. 502 с.
4. Сайфан Джиджи. Осваиваем Kubernetes. Оркестрация контейнерных архитектур. Санкт-Петербург: Питер, 2019. 400 с.
5. Лазарева Н. Б., Ловцова Н. Н. Автоматизация получения доступа к базам данных для компонентов в среде Kubernetes // Инженерный вестник Дона. 2021. №3. URL: ivdon.ru/ru/magazine/archive/n3y2021/6844.
6. Abbas Zafer. Cloud Migration Strategies: Moving Applications and Workloads to the Cloud. Department of Computer Science, University of South Coria (2022). URL: www.researchgate.net/publication/372826166_Cloud_Migration_Strategies_Moving_Applications_and_Workloads_to_the_Cloud.
7. Yousafzai Abdullah, Ibrar Yaqoob, Muhammad Imran, Abdullah Gani, and Rafidah Md Noor. Process migration-based computational offloading framework for IoT-supported mobile edge/cloud computing. IEEE Internet of Things Journal (2019). URL: ieeexplore.ieee.org/document/8847424.
8. Анисимова Г. Б., Фролов Р. А. Разработка информационной системы автоматизации деятельности автосалона // Инженерный вестник Дона. 2022. №5. URL: ivdon.ru/ru/magazine/archive/n5y2022/7647.

9. Панкратов А. А., Анисимова Г. Б. Проектирование информационной системы оптимизации работы автостоянки // Инженерный вестник Дона. 2018. №3. URL: ivdon.ru/ru/magazine/archive/n3y2018/5091.

10. Мересий А. В., Полевщиков И. С. Автоматизация тестирования систем управления облачными базами данных на примере DBaaS Postgres Pro // Инженерный вестник Дона. 2024. №8. URL: ivdon.ru/ru/magazine/archive/n8y2024/9406.

11. John G. Schmidt, Kirit Basu. DataOps. The Authoritative Edition. Panther Publishing, 3102 Rexford Dr., Austin, TX 78723, USA. 2019. 135 p.

12. Simon Trewin. The DataOps Revolution: Delivering the Data-Driven Enterprise. CRC Press 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742. 2022. 167 p.

References

1. Bayeva V. R., Drozdov A. YU. ETL: Aktual'nost' i primeneniye. Preimushchestva i nedostatki ETL instrumentov // Mezhdunarodnyy nauchnyy zhurnal «VESTNIK NAUKI» №5 (14) T.4. URL: cyberleninka.ru/article/n/etl-aktualnost-i-primenenie-preimushchestva-i-nyedostatki-etl-instrumentov/viewer.

2. Busenkov A. A., Bagazhkov D.I., Chernov V. V., Panov A. I., Bashmurov N. A. Razrabotka algoritma i programmaya realizatsiya sredstva zashchity personal'nykh dannykh v oblachnykh khranilishchakh // «INNOVATSII I INVESTITSII». 2021. №12. URL: cyberleninka.ru/article/n/razrabotka-algoritma-i-programmnaya-realizatsiya-sredstva-zaschity-personalnyh-dannyh-v-oblachnyh-hranilischah/viewer.

3. Kharenslak B., Ruyter de D. Apache Airflow i konveyyery obrabotki dannykh: prakticheskoye rukovodstvo. Moskva: DMK Press, 2022. 502 p.

4. Sayfan Dzhidzhi. Osvaivayem Kubernetes. Orkestratsiya konteynerykh arkhitektur. Sankt-Peterburg: Piter, 2019. 400 p.

5. Lazareva N. B., Lovtsova N. N. Avtomatizatsiya polucheniya dostupa k bazam dannykh dlya komponentov v srede Kubernetes // Inzhenernyy vestnik Dona. 2021. №3. URL: ivdon.ru/ru/magazine/archive/n3y2021/6844.

6. Abbas Zafer. Cloud Migration Strategies: Moving Applications and Workloads to the Cloud. Department of Computer Science, University of South Coria (2022). URL: www.researchgate.net/publication/372826166_Cloud_Migration_Strategies_Moving_Applications_and_Workloads_to_the_Cloud.
7. Yousafzai Abdullah, Ibrar Yaqoob, Muhammad Imran, Abdullah Gani, and Rafidah Md Noor. Process migration-based computational offloading framework for IoT-supported mobile edge/cloud computing. IEEE Internet of Things Journal (2019). URL: ieeexplore.ieee.org/document/8847424.
8. Anisimova G. B., Frolov R. A. Razrabotka informatsionnoy sistemy avtomatizatsii deyatel'nosti avtosalona // Inzhenernyy vestnik Dona. 2022. №5. URL: ivdon.ru/ru/magazine/archive/n5y2022/7647.
9. Pankratov A. A., Anisimova G. B. Proektirovanie informatsionnoy sistemy optimizatsii raboty avtostoyanki // Inzhenernyy vestnik Dona. 2018. №3. URL: ivdon.ru/ru/magazine/archive/n3y2018/5091.
10. Meresiy A. V., Polevshchikov I. S. Avtomatizatsiya testirovaniya sistem upravleniya oblachnymi bazami dannykh na primere DBaaS Postgres Pro // Inzhenernyy vestnik Dona. 2024. №8. URL: ivdon.ru/ru/magazine/archive/n8y2024/9406.
11. John G. Schmidt, Kirit Basu. DataOps. The Authoritative Edition. Panther Publishing, 3102 Rexford Dr., Austin, TX 78723, USA. 2019. 135 p.
12. Simon Trewin. The DataOps Revolution: Delivering the Data-Driven Enterprise. CRC Press 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742. 2022. 167 p.