

Интеграция информационных систем в цифровые платформы для оказания социальных сервисов населению

С.Б. Попов^{1,2}, П.В. Хрипунов²

¹Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

²Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Рассматриваются методологические аспекты трансформации существующих информационных систем оказания социальных сервисов населению в цифровые платформы, которые обеспечивают использование новых инновационных технологий обработки и аналитики больших данных, машинного обучения и искусственного интеллекта. Выделяются три информационных технологии, на которых базируется такая трансформация: технологии облачных вычислений; технологии больших данных; технологии разработки программного обеспечения на базе контейнеров в соответствии с микросервисной архитектурой. Предлагается организация управления данными цифрового профиля человека на основе введения нового уровня абстракции – информационного обеспечения – и использования микросервисной архитектуры, реализующей концепцию Data-as-a-Service.

1. Введение

Совершенствование оказания социальных сервисов населению в настоящее время связывается с реализацией концепции «Государство-как-Платформа» [1], важнейшей частью которой является цифровая трансформация процессов взаимодействия органов государственной и муниципальной власти и населения.

В государственном секторе нет конкурентного давления, заставляющего компании трансформировать свои бизнес-процессы, поэтому в разных странах уровень цифровизации органов государственного управления и степень развития цифровых платформ значительно отличаются. Тем не менее, получив опыт получения широкого спектра услуг в различных областях посредством бизнес-платформ, граждане ожидают, что государство также обеспечит предоставление базовых социальных услуг в максимально удобном для пользователей цифровом формате в режиме онлайн.

В настоящее время предоставление социальных сервисов населению в электронном виде (через интернет) осуществляется через интернет-порталы федерального и регионального уровня, информационные системы различных ведомств. Флагманом здесь выступает Единый портал государственных услуг и функций (ЕПГУ), который постоянно расширяет спектр предоставляемых федеральных и региональных услуг и наращивает пользовательскую базу,

совершенствует способы взаимодействия с населением через мобильные приложения портала «Госуслуги» [2].

2. Базовые архитектуры информационных систем в сфере оказания социальных сервисов населению

Существующие решения в сфере информационных систем государственных ведомств и органов власти различного уровня, которые часто обобщённо именуют как электронное правительство (ЭП), представляют собой промежуточный этап на пути к созданию целевой архитектуры для «Государства как платформы». Обобщая мировой опыт создания таких систем, можно отметить, что, в том или ином виде, электронное правительство включает в себя следующие элементы:

- Информационная система предоставления государственных услуг, реализованная, как правило, в виде единого портала госуслуг или как комплекс порталов государственных ведомств и органов власти различного уровня с единым обликом (пользовательский интерфейс и внешний вид).
- Технологическая инфраструктура электронного взаимодействия государственных органов и иных государственных организаций или система межведомственного электронного документооборота.
- Система идентификации и аутентификации.

Эти три составляющие в различных странах, в разные периоды времени реализовывались по-разному.

На начальном этапе развития базовые составляющие цифрового правительства создавались в виде информационной системы уровня предприятия, условно говоря, в виде отдельного приложения с удалённым доступом пользователей по локальной сети или через интернет. Информационные системы обслуживали внутренние потребности правительственных ведомств и других органов государственной власти (ОГВ). Технологическая инфраструктура электронного взаимодействия ОГВ также реализовывалась в виде информационных систем на основе типовых баз данных, которые выступали как точки обмена информацией между другими участниками электронного документооборота. В настоящий момент подобные системы являются внутренним слоем правительственных информационных систем и напрямую не взаимодействуют с внешними потребителями государственных услуг (госуслуг) – отдельными гражданами, организациями, предприятиями, ассоциациями и т.д.

Возникновение концепции электронного правительства привело к необходимости обеспечения доступа потребителей госуслуг к информационным системам ОГВ. При этом объединённая архитектура приложений электронного правительства была ориентирована на порталы, а порталные приложения были расширениями некоторых внутренних информационных систем государственных ведомств.

Развитие технологий проектирования сложных распределённых информационных систем [3], внедрение методологий, основанных на концепции сервис-ориентированной архитектуры [4], привело к созданию отдельного слоя электронного правительства на основе программно-реализованной магистральной шины приложений и сервисов, которая координирует и интегрирует не только специально разработанные для этого слоя сервисы, но и унаследованные приложения.

На этом этапе развития ЭП наибольший импульс развития получила технологическая инфраструктура электронного взаимодействия, которая реализовывалась в виде магистральной шины приложений и набора специализированных сервисов. Элементы системы идентификации и аутентификации также оформлялись в виде сервисов.

На современном этапе существующие приложения постепенно модернизируются, монолитные приложения преобразуются в распределённые программные системы на основе технологий и сервисов, основанных на процессах обработки данных, создавая, таким образом, архитектуру приложений переходного этапа.

После преобразования всех ранее монолитных приложений в процессы и сервисы, архитектура ЭП формируется в виде облака государственных услуг (сервисов). Приложения электронного правительства преобразуются в простые коннекторы к этому облаку. Данная архитектура является хорошей основой для оформления ЭП в цифровую платформу. На данном этапе все элементы ЭП, в том числе и технологическая инфраструктура электронного взаимодействия, представлены в виде сервисов, что иллюстрирует пример платформы X-tee (X-Road до 2018 г.) в Эстонии.

Австралия внедряет в ЭП облачную платформу cloud.gov.au, которая значительно облегчает развертывание и обслуживание цифровых услуг на основе сервисов. Платформа cloud.gov.au обеспечивает прозрачно для пользователей быстрый, безопасный, стандартный способ изменения веб-приложений. Это облегчает правительству выпуск, мониторинг и расширение пользовательских цифровых услуг, а в сочетании с онлайн-платформой «Цифровой рынок» (Digital Marketplace) австралийское правительство полностью трансформирует процесс государственных закупок цифровых услуг и упрощает доступ к государственным контрактам для предприятий любого размера.

Развитие электронного правительства в направлении действительно электронной социальной системы превратит архитектуру ЭП в социальное облако, взаимодействующее с облаком государственных сервисов (услуг). Последнее выступает в качестве платформы для социальных, профессиональных, частных, добровольных и других услуг, которые должны быть интегрированы в электронно-социальную систему.

Та часть электронного правительства, которая взаимодействует с партнерами, является коллективным экстранетом, который похож на популярные социальные сетевые инструменты (Facebook, LinkedIn и т.д.) и электронный банкинг. Его основным функционалом являются следующие составляющие:

- безопасное хранилище коротких сообщений, документов и видео;
- специализированная (включая ролевую) информация и функциональность;
- разнообразные услуги, как небольшие модульные приложения, подключаемые по мере необходимости
- прямой канал к государственным бизнес-процессам;
- единое представление (пользовательский интерфейс и внешний вид) центральных, региональных и местных органов власти.

Правительственная часть электронного правительства (набор приложений) является интеграционным и координационным потенциалом, необходимым для удовлетворения потребностей партнеров.

Важно, чтобы слой электронного правительства был отделён от слоя существующей государственной информационной среды. Это разделение означает оперативную и эволюционную независимость.

3. Архитектурные решения и технологии современных цифровых платформ

В условиях цифровой трансформации цифровые платформы должны обеспечивать базовой организации использование новых инновационных технологий обработки и аналитики больших данных, машинного обучения и искусственного интеллекта, поддержку интернета вещей.

Традиционные или существующие системы, архитектура которых сложилась в конце 1980-х годов, просто не могут поддерживать эти технологии эффективным способом, если вообще способны работать с ними.

Современное обновление инфраструктуры информационных технологий, в первую очередь, связывают со следующими инновационными технологиями:

- технологии, основанные на модели облачных вычислений [5];
- технологии распределённого отказоустойчивого хранения, обработки и аналитики больших данных [6];

- технологии разработки программного обеспечения на базе контейнеров в соответствии с микросервисной архитектурой.

3.1. Облачные технологии

Ключом к появлению облачных вычислений и платформ, на которых они развертываются, является концепция виртуализации, в которой обособленные вычислительные ресурсы отделяются (или «абстрагируются») от их оригинального оборудования и объединяются в «виртуальный» пул ресурсов. Это значительно облегчает облачным поставщикам использование программного обеспечения для автоматизации подготовки и настройки систем, для которых традиционно требуется конфигурация аппаратного уровня.

Первоначальная идея виртуализации ИТ-ресурсов предприятия или приобретения этих ресурсов у поставщиков общедоступного облака эволюционирует, такие понятия как контейнеры, микросервисы, бессерверные архитектуры приложений и другие типы программного обеспечения для управления ресурсами начинают создавать более сложную картину со множеством нюансов.

Из всего разнообразия моделей обслуживания облачных технологий в целях разработки цифровых платформ следует обратить внимание на следующие важные разновидности базовой модели Платформа-как-Услуга:

- integration Platform as a Service (iPaaS) – платформа интеграции как услуга;
- Containers-as-a-Service (CaaS) – платформа управления контейнерами как услуга.

Платформа интеграции как услуга (iPaaS) – это набор облачных сервисов, обеспечивающих разработку, выполнение и управление интеграционными потоками, соединяющими любые комбинации локальных и облачных процессов, сервисов, приложений и данных внутри одной или нескольких организаций. iPaaS можно интерпретировать как систему метапрограммирования программных потоков вычислений на основе сервисов, которая формирует программную сеть из независимых сервисов, связывая их между собой, согласовывая их интерфейсы специализированными коннекторами.

Платформы iPaaS предоставляют централизованную консоль для управления, регулирования и интеграции облачных приложений. Эти средства работают путем подключения облачных приложений/сервисов и управления потоками интеграции (потоков передачи данных между сервисами). Они могут ускорить разработку продукта за счет интеграции существующих инструментов и увеличения объема данных за счет использования внешних источников. Компании используют эти инструменты для масштабирования производительности, добавления функциональности продукта и структурирования интегрируемых приложений. Функции или данные можно быстро добавлять или удалять, сокращая обработку отказа, время простоя и длительность разработки.

Существуют определённые пересечения между инструментами ESB (Enterprise Service Bus – архитектурное решение в виде корпоративной сервисной шины) и решениями iPaaS, но применяются они для разных систем. iPaaS интегрирует облачные приложения, в то время как инструменты ESB обычно используются в составе корпоративных информационных систем [7]. Containers-as-a-Service (CaaS) – это комплекс сервисов управления рабочими нагрузками на основе контейнеров. CaaS предлагает платформу для развертывания и управления кластерами приложений и контейнеров путем предоставления пользователям средств исполнения контейнеров, служб их оркестровки и базовых ресурсов. Платформы управления контейнерами облегчают организацию и виртуализацию программных контейнеров, что иногда называется виртуализацией на уровне операционной системы. Разработчики используют контейнеры для запуска, тестирования и защиты приложений в независимых от ресурсов средах. Контейнеры инкапсулируют компоненты приложений, необходимые им библиотеки или интерактивные программные модули. Платформы управления помогают пользователям распределять ресурсы для оптимизации производительности и балансировки рабочих нагрузок системы. Контейнеры обеспечивают гибкую, портативную платформу для организации, автоматизации и распространения приложений. Компании используют платформы управления контейнерами

для упрощения доставки контейнеров, чтобы избежать сложностей взаимозависимых архитектур систем. Эти инструменты являются масштабируемыми и могут значительно повысить производительность распределённых приложений.

3.2. Технологии больших данных

Термин «Большие данные» (Big Data) может рассматриваться как характеристика наборов данных сверхбольшого объёма и как характеристика инновационных технологий сбора, систематизации и обработки данных в целях извлечения знаний о явлениях и объектах, которые связаны с этими наборами.

Распространённым подходом в первом случае является использование Gartner'овской триады, определяющей данные сверхбольшого объёма как совокупность трех характеристик 3V: Volume (объём), Velocity (скорость поступления/наращивания), Variety (разнообразие). Несмотря на то, что в оригинале [8] речь шла о проблемах технологий обработки данных, и термин «Большие данные» не упоминался, был сформулирован тренд выхода за рамки традиционных подходов к управлению данными методом прямого перебора, запрос на использование новых разнообразных архитектурных решений, которые существенно повлияют на состав используемых бизнесом приложений и стратегий обработки данных.

Собственно второй вариант определения как раз и говорит о том, что технологии больших данных – это технологии обработки информации, которые применяются тогда, когда традиционные технологии обработки на базе реляционных баз данных не применимы для решения стоящих задач.

С тех пор многочисленные предложения других авторов сформировали 5V больших данных – добавились ещё два аспекта: Veracity (достоверность) и Value (ценность).

Последний аспект (Value) имеет ещё и значение «Смысл». И это приводит к замечательной краткой формулировке консалтинговой компании Forrester: «Большие данные объединяют техники и технологии, которые извлекают смысл из данных на экстремальном пределе практичности».

Обобщая многочисленные дискуссии в этой области, можно дать следующее определение: «Технологии Big Data – это технологии преобразования информации в знание». Здесь следует отметить, что знание – это не просто информация, это информация со связями, зависимостями, контекстом, историей её изменения (во времени), моделью (ее получения и использования). Главное отличие знаний от данных состоит в их структурности и активности, появление в базе данных новых фактов или установление новых связей может стать источником формирования новых знаний, а, следовательно, изменений в принятии решений. Для принятия решения необходимы знания, а не информация.

Почему в области обработки данных сверхбольшого объёма не сработали традиционные технологии обработки на основе реляционных баз данных?

Потому, что они не создают знания. Традиционные технологии, опираясь на уже имеющиеся знания, создают основу для накопления и интерпретации информации в рамках существующего знания. Традиционный подход предполагает, что мы заранее знаем способы использования всех помещаемых в хранилище данных и технология их обработки в большинстве случаев уже определена, возможны некоторые вариации, но в рамках заранее продуманных процедур. Выход за эти рамки невозможен или сопряжён со значительными доработками программного обеспечения, большими временными затратами на конвертацию или переформатирование данных. Основная проблема – этап структурирования данных при внесении их в базу данных, необходимость заранее сформировать формат записи, в соответствии с которым будет храниться информация. В случае больших данных разнородность этой информации может быть настолько велика, что создание единого шаблона её хранения невозможно, а количество форматов записи единичной порции информации сопоставимо с количеством этих порций.

Концепция Big Data оформилась в тот момент, когда пришло осознание, что огромный пласт данных, порождаемых организацией и связанных с ней людей, машин и других разнообразных источников информации [9], не фиксируется вообще или неосознанно (мертвым грузом)

хранится на компьютерах пользователей, серверах подразделений. Причем на текущий момент невозможно предсказать какого рода информация, с какими связями и временными отметками может быть востребована позднее, какие методы будут использоваться при её комплексировании, обработке и анализе.

Поэтому на начальном этапе становления технологий больших данных появилась потребность сохранять *все* данные на всякий случай, понять существующие источники данных, сохранить *все* (возможно неструктурированные, «как есть») данные в некое единое «большое» хранилище.

Поскольку информация не структурирована, то в соответствии с концепцией она сохраняется со всеми сопутствующими связями, временными и пространственными метками. Однако без соответствующих методов обработки подобная неструктурированная информация мертва. Собственно концепция Big Data и появилась в связи с созданием нового инструментария для хранения, обработки и анализа такой информации в виде связки нескольких информационных технологий – распределённой файловой системы Google File System (GFS), парадигмы масштабируемой распределённой обработки данных MapReduce и распределённой нереляционной базы данных Bigtable. Эти технологии были разработаны компанией Google и представлены в виде докладов на нескольких конференциях [10-12].

Бурное развитие технологии больших данных получили с появлением программной инфраструктуры распределённых вычислений Hadoop, реализованной на Java в виде проекта с открытым исходным кодом. Используя идеи исследователей из компании Google, разработчики Hadoop создали программную платформу, которая обеспечивает распределённое хранение и параллельную обработку неструктурированных данных на кластерных вычислительных системах.

До недавнего времени проект Apache Hadoop содержал четыре базовых компонента:

- Hadoop Common – набор базовых модулей, которые обеспечивают программную поддержку других компонент и фреймворков всей инфраструктуры Hadoop.
- Hadoop Distributed File System (HDFS) – распределённая файловая система, которая обеспечивает высокопроизводительный доступ к данным распределённого приложения, запускаемого на платформе Hadoop.
- Hadoop YARN – система планирования заданий и управления ресурсами вычислительного кластера.
- Hadoop MapReduce – фреймворк параллельной обработки больших наборов данных с использованием системы YARN.

В октябре 2018 года в состав базовых модулей был включён проект Apache Hadoop Ozone, который предоставляет семантику хранилища объектов для Hadoop, использует распределённое хранилище данных Hadoop Distributed Data Storage (HDDS) для уровня хранения данных. HDDS – ещё один новый подпроект Apache Hadoop.

Apache Hadoop можно рассматривать и как систему обработки данных с MapReduce в качестве программного механизма обработки по умолчанию, и как инфраструктуру (экосистему) для целого ряда технологий обработки больших данных. Используемые в рамках инфраструктуры Hadoop фреймворки взаимозаменяемы и/или могут использоваться в tandem. Например, Apache Spark, фреймворк пакетной и потоковой [13-15] обработки данных, может быть подключен к Hadoop, заменяя MapReduce. Разнообразные платформы обработки и программные компоненты обработки интегрируются в Hadoop для использования HDFS и менеджера ресурсов YARN. Эта функциональная совместимость между компонентами является одной из причин того, что Apache Hadoop в настоящее время является базовой платформой в области обработки больших данных.

Работа с большими данными не похожа на обычный процесс бизнес-аналитики, где простое сложение известных значений приносит результат: например, итог сложения данных об оплаченных счетах становится объемом продаж за год. При работе с большими данными результат получается в процессе их очистки путём последовательного моделирования: сначала выдвигается гипотеза (или даже набор гипотез), строится статистическая, визуальная или

семантическая модель (набор моделей), на ее основании проверяется верность выдвинутой гипотезы и затем (но скорее параллельно) выдвигается (или анализируется) следующая. Этот процесс требует от исследователя либо интерпретации визуальных значений [16-17] или составления интерактивных запросов на основе знаний, либо разработки адаптивных алгоритмов машинного обучения, способных получить искомым результат.

Традиционные подходы при проведении исследований предполагают первоначальное формирование некоторой модели, затем накопление и обработку данных с последующей оценкой параметров такой заранее сформированной модели. Технологии больших данных в процессе формирования новых знаний интегрируют моделирование, вычислительный эксперимент и информационные технологии [18]. В процессе исследований в соответствии с методологией больших данных решается задача нахождения модели, максимально соответствующей некоторым экспериментальным данным [16-17]. Причём данная методология направлена на поиск наиболее интересных (т.е. неожиданных, действенных) и робастных (предсказывающих будущее поведение) моделей [19].

На современном этапе развития технологий больших данных наблюдается тенденция отхода от идеологии хранения *всех* данных. Больше внимания уделяется аспектам Veracity (достоверность) и Value (ценность). Развивается концепция жизненного цикла данных, в рамках которой разрабатываются процессы управления данными, формируется понимание как данные появились, кто за них отвечает, и кто в них заинтересован, какой вклад они обеспечивают в формирование целевого продукта – знания.

Процессы управления данными приобретают самостоятельное значение, они формируют достоверные данные, отвечают за поддержание их достоверности (Veracity) [9], контролируют востребованность конкретных данных, оценивают их вклад в создание прибавочной стоимости (Value), прогнозируют целесообразность последующего хранения конкретных данных.

3.3. Микросервисная архитектура и контейнеры

Архитектура микросервисов [20] – архитектурный подход, который ставит во главу декомпозицию приложений на взаимодействующие, функционально простые, слабосвязанные сервисы, что обеспечивает поставку и поддержку сложных программных систем со скоростью и качеством, необходимыми для современного цифрового бизнеса. Микросервисы можно использовать для разбиения большого монолитного приложения на более мелкие и простые сервисы. Каждый микросервис реализует одну функцию, и делает это хорошо, так что "микро" в микросервисах относится к сфере функциональности сервисов, а не к количеству строк кода.

Микросервисы характеризуются независимостью от языка программирования, платформы и операционной системы. Основным инструментом, обеспечивающим эту независимость, в настоящее время является технология контейнеров [21].

Микросервисный подход часто неэффективен при создании простых приложений и лучше подходит для сложных приложений, которые разрабатываются в течение определенного периода времени, с большими базами кода и командами. Микросервисы идеально подходят для реализации множества бизнес-функций, при этом каждый микросервис фокусируется на реализации одной бизнес-функции или подфункции. Команды могут независимо модернизировать код в соответствии со своим собственным планом.

Возможность быстрого обновления приложения и значительное сокращение времени, необходимого для подготовки нового релиза, которые обеспечивает микросервисный подход к проектированию, являются ключевыми факторами, продвигающими этот стиль архитектуры.

Концепция микросервисов схожа с сервис-ориентированной архитектурой (SOA). При сравнении концепций SOA (сервис-ориентированной архитектуры) и архитектуры микросервисов следует отметить, что архитектура микросервисов представляет собой шаблон архитектуры «разделяй-на-минимально-возможные-части», который делает упор на концепцию разграничения контекста, тогда как SOA – это архитектурный шаблон «разделяй-на-максимально-возможные-части», который делает упор на повторное использование абстракций и бизнес-функций.

4. Методологические аспекты трансформации информационных систем в цифровые платформы

Модернизация информационной системы – сложный и долговременный процесс. Отправной его точкой чаще всего является решение не переписывать имеющиеся системы заново, а трансформировать их. Такое решение обычно принимается из-за того, что в организации скапливается огромное количество прежних наработок, многие из которых написаны на устаревших языках и с трудом допускают дальнейшее развитие и поддержку.

Существующие информационные системы на основе сервис-ориентированной архитектуры (наиболее распространённый вариант в настоящее время) могут быть достаточно технологично перенесены на облачные инструментальные платформы, которые обеспечат необходимый уровень масштабируемости, адаптируемости по модели IaaS. Наличие сервисной модели позволяет использовать модули существующих информационных систем при реализации расширенной функциональности цифровой платформы либо через непосредственной использование существующего программного интерфейса модулей, либо через разработку специализированных адаптеров к ним.

Дальнейшее направление модернизации связано с уменьшением связности системы за счёт выделения функциональных модулей информационной системы в отдельные независимые сервисы, разработанные и объединённые в единую среду на выбранной платформе, предоставляющей услуги по модели PaaS или, в случае перехода к микросервисной архитектуре ПО, по модели SaaS.

Реорганизация существующего информационного слоя систем, т.е. баз данных, как правило, не требуется, либо незначительна. Однако, это не означает, что механизмы хранения и организации доступа к данным при переходе на модель цифровых платформ не изменяются. Наоборот, создание цифровой платформы предполагает, что будут использованы новейшие технологии распределённого хранения и анализа больших данных, которые обеспечат высокий уровень масштабирования, прозрачное подключение современных методов анализа данных, принятие решений с использованием элементов искусственного интеллекта [19, 22-23].

При этом существующие базы данных могут быть использованы как один из источников при формировании общего «озера» данных целевой платформы. В процессе развития целевой платформы может возникнуть потребность формирования реплик существующих баз данных, разбиения их на необходимые домены или использования комплексных схем масштабирования для обеспечения расширенного спектра потребностей всех составляющих платформы. Дальнейшие сценарии интеграции унаследованных баз данных, как правило, предполагают перенос всех данных в хранилища, реализованные на основе современных распределённых платформ обработки данных.

4.1. Слой информационного обеспечения

В настоящее время формируется понимание необходимости введения нового слоя в стек программного обеспечения предприятия – dataware или информационного обеспечения.

Информационное обеспечение – это уровень абстракции, который позволяет управлять данными как важнейшим корпоративным ресурсом, отделённым от любых других зависимостей. Введение нового программного слоя обеспечивает гораздо меньшую совокупную стоимость владения при управлении данными в разных хранилищах, в процессе использовании широкого спектра технологий и инструментов обработки данных и, что еще важнее, возможность быстрее извлекать конкурентные знания из данных [7].

Преыдушие поколения систем управления данными были точечными продуктами или инструментами, решающими конкретную проблему. Концептуально это похоже на сетевые устройства или сервисы, рассматриваемые как управляемый ресурс. Информационное обеспечение эффективно оперирует разнообразными типами данных, способами доступа к данным и системными инструментами, необходимыми для управления данными как корпоративным ресурсом, независимо от базовой инфраструктуры и местоположения.

Информационное обеспечение предоставляет предприятиям согласованный подход к управлению и защите данных. Набор стандартных API-интерфейсов позволяет использовать

данные широким спектром приложений и инструментов. Разнообразие и большой объём данных преодолеваются за счёт возможности обработки данных в разных местах с использованием как локальной, так и облачной аппаратной инфраструктуры, а также контейнеров.

Информационное обеспечение оптимизирует весь жизненный цикл данных – от приема до обработки – для поддержки приложений, которые одновременно требуют аналитики в реальном времени, машинного обучения и искусственного интеллекта [17, 20-21]. Информационное обеспечение обеспечивает полную гибкость в использовании базовой инфраструктуры (локальная, облачная или контейнерная инфраструктура) и шаблонов развертывания (гибридная или мульти облачная).

Таким образом, устаревшие монолитные информационные системы требуют коренной трансформации для включения их элементов в целевую цифровую платформу. Более современные информационные системы, построенные с использованием сервис-ориентированной архитектуры, могут быть сначала незначительно модернизированы за счёт развёртывания их в инфраструктуре облачных технологий (IaaS), с последующей адаптацией для использования большинства их отдельных составляющих на платформенной архитектуре (PaaS).

Базы данных, используемые в информационных системах, подключаются к целевым платформам в качестве одного из источников данных, служат для первоначального наполнения базового информационного слоя в платформе, но в процессе функционирования и развития платформы данные из унаследованных баз данных постепенно будут перемещаться в выбранную для целевой платформы распределённую платформу обработки данных.

5. Проектные решения реализации цифрового профиля человека на платформе социальных сервисов

Трансформация информационных систем в цифровые платформы для оказания социальных сервисов населению на основе введения нового слоя – информационного обеспечения – и использования микросервисов формирует новый подход к организации управления данными на основе концепции цифрового профиля человека.

Именно цифровой профиль гражданина целесообразно выбрать в качестве базовой единицы учёта при создании Государства-как платформы [1].

Для реализации этого подхода авторы предлагают использовать концепцию Data-as-a-Service [24]. Основные проектные решения при этом развивают и адаптируют принципы технологии Big Data применительно к задачам информационного обеспечения оказания социальных сервисов населению: (1) все компоненты системы управления данными реализуются как сервисы, размещенные в облачной вычислительной среде; (2) в качестве интеграционного решения используется децентрализованное объединение равноправных взаимодействующих друг с другом сервисов, т.е. федерация сервисов; (3) основной структурной единицей хранения в системе является цифровой профиль человека, который доступен пользователю как отдельный сервис хранения со своим уникальным именем в федерации сервисов; (4) сервисы цифрового профиля используют принцип обработки и хранения данных в оперативной памяти; (5) сервисы регистрируются в распределенном репозитории, который обеспечивает группировку всех сервисов по передаваемой от них информации; (6) организация вычислений с использованием информации из цифрового профиля человека использует ориентированный на данные подход, при котором в процессе обработки на узлы распределенной системы вместо данных рассылаются процедуры их обработки; (7) в процессе обработки данные цифрового профиля человека не изменяются, а добавляется новая сущность в его зависимости, которая верифицируется с использованием отдельного процесса обработки данных, запускаемого автономно в фоновом режиме, предоставляя пользователю необходимые ему функции (чаще всего это визуализация) и данные непосредственно сразу после их создания/обновления.

Сервис цифрового профиля человека самостоятельно обеспечивает формирование достоверных данных для сторонних сервисов обработки данных, отвечает за поддержание их достоверности, контролирует востребованность конкретных сущностей профиля, управляет жизненным

циклом хранения составляющих его данных. Цифровой профиль человека формирует граф информационных связей, описывающий все цифровые следы, оставляемые им при взаимодействии с цифровым пространством [25]. Сервис цифрового профиля конкретного человека обеспечивает защищённое изолированное соединение с системами идентификации, защиты данных и раскрытия информации. Данная концепция полностью совместима с планируемой к реализации в госуслугах системой биометрической идентификации на основе распознавания лица и голоса человека [26-28]. Дальнейшее развитие интернета вещей повысит цифровую прослеживаемость данных, связанных с конкретным человеком, поможет формировать медицинскую составляющую профиля человека, прогнозировать его продуктивные предпочтения, социальное поведение, персонализировать и прогнозировать жизненные траектории человека.

Цифровой профиль-как-сервис позволит перейти от оказания единичных цифровых услуг к проактивному предоставлению услуг с использованием технологий и инструментов формирования процессов взаимодействия человека и государства для оказания всего комплекса взаимосвязанных услуг в конкретной жизненной ситуации граждан.

Формирование программных процессов при этом выполняется в рамках интеграционной платформы (iPaaS), которая объединяет сервисы цифрового профиля и сервисы обработки в единый программный поток вычислений с использованием метапрограммирования на основе специализированных шаблонов, описывающих процесс исполнения услуг в соответствии с необходимым регламентом.

Предлагаемый подход обеспечивает для социальных цифровых платформ прозрачное использование новых инновационных технологий обработки и аналитики больших данных в реальном времени, машинного обучения и искусственного интеллекта, поддержку технологий интернета вещей.

6. Литература

- [1] Шклярчук, М.С. Государство как платформа: Люди и технологии. – М: РАНХиГС, 2019. – 111 с.
- [2] Подведены итоги работы Единого портала госуслуг в 2018 году / Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации, 2019 [Электронный ресурс]. – Режим доступа: <https://digital.gov.ru/ru/events/38738/> (03.03.2019).
- [3] Kazanskiy, N.L. The distributed vision system of the registration of the railway train / N.L. Kazanskiy, S.B. Popov // *Computer Optics*. – 2012. – Vol. 36(3). – P. 419-428.
- [4] Волотовский, С.Г. Оценка производительности приложений параллельной обработки изображений / С.Г. Волотовский, Н.Л. Казанский, С.Б. Попов, П.Г. Серафимович // *Компьютерная оптика*. – 2010. – Т. 34, № 4. – С. 567-572.
- [5] Kazanskiy, N.L. Cloud computing for nanophotonic simulations / N.L. Kazanskiy, P.G. Serafimovich // *Lecture Notes in Computer Science*. – 2013. – Vol. 7715. – P. 54-67. DOI: 10.1007/978-3-642-38250-5_7.
- [6] Kazanskiy, N.L. Distributed storage and parallel processing for large-size optical images / N.L. Kazanskiy, S.B. Popov // *Proceedings of SPIE*. – 2012. – Vol. 8410. – P. 84100I. DOI: 10.1117/12.928441.
- [7] Surnin, O.L. Big data incorporation based on open services provider for distributed enterprises / O.L. Surnin, P.V. Sitnikov, A.V. Ivaschenko, N.Yu. Ilyasova, S.B. Popov // *CEUR Workshop Proceedings*. – 2017. – Vol. 1903. – P. 42-47.
- [8] Laney, D. 3D Data Management: Controlling Data Volume, Velocity and Variety // *Application Delivery Strategies*, File 949. – META Group, 2001 – [Electronic resource]. – Access mode: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [9] Евсютин, О.О. Обзор методов встраивания информации в цифровые объекты для обеспечения безопасности в «интернете вещей» / О.О. Евсютин, А.С. Кокурина, Р.В. Мещеряков // *Компьютерная оптика*. – 2019. – Т. 43, № 1. – С. 137-154. – DOI: 10.18287/2412-6179-2019-43-1-137-154.

- [10] Ghemawat, S. The Google file system / Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung // Proceedings of the 19th ACM Symposium on Operating Systems Principles. – Bolton Landing, NY, 2003. – P. 29-43.
- [11] Dean, J. MapReduce: Simplified Data Processing on Large Clusters / J. Dean, S. Ghemawat // Proceedings of the Sixth Symposium on Operating System Design and Implementation, 2004. – P. 137-150.
- [12] Chang, F. Bigtable: A Distributed Storage System for Structured Data / F. Chang, J. Dean, S. Ghemawat, W.C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, Robert E. Gruber // Proceedings of the 7th Symposium on Operating Systems Design and Implementation, 2006. – P. 205-218.
- [13] Protsenko, V.I. Firmware and hardware infrastructure for data stream processing / V.I. Protsenko, P.G. Seraphimovich, S.B. Popov, N.L. Kazanskiy // CEUR Workshop Proceedings. – 2016. – Vol. 1638. – P. 782-787. DOI: 10.18287/1613-0073-2016-1638-782-787.
- [14] Казанский, Н.Л. Сравнение производительности систем потокового анализа данных в задаче обработки изображений скользящим окном / Н.Л. Казанский, В.И. Проценко, П.Г. Серафимович // Компьютерная оптика. – 2014. – Т. 38, № 4. – С. 804-810.
- [15] Проценко, В.И. Анализ параметров систем детектирования множественных визуальных объектов в режиме реального времени / В.И. Проценко, Н.Л. Казанский, П.Г. Серафимович // Компьютерная оптика. – 2015. – Т. 39, № 4. – С. 582-591. DOI: 10.18287/0134-2452-2015-39-4-582-591.
- [16] Смелкина, Н.А. Реконструкция анатомических структур на основе статистической модели формы / Н.А. Смелкина, Р.Н. Косарев, А.В. Никоноров, И.М. Байриков, К.Н. Рябов, А.В. Авдеев, Н.Л. Казанский // Компьютерная оптика. – 2017. – Т. 41, № 6. – С. 897-904. DOI: 10.18287/2412-6179-2017-41-6-897-904.
- [17] Никоноров, А.В. Реконструкция изображений в дифракционно-оптических системах на основе сверточных нейронных сетей и обратной свертки / А.В. Никоноров, М.В. Петров, С.А. Бибииков, В.В. Кутикова, А.А. Морозов, Н.Л. Казанский // Компьютерная оптика. – 2017. – Т. 41, № 6. – С. 875-887. DOI: 10.18287/2412-6179-2017-41-6-875-887.
- [18] Popov, S.B. The Big Data methodology in computer vision systems // CEUR Workshop Proceedings. – 2015. – Vol. 1490. – P. 420-425. DOI: 10.18287/1613-0073-2015-1490-420-425.
- [19] Кропотов, Ю.А. Метод прогнозирования изменений параметров временных рядов в цифровых информационно-управляющих системах / Ю.А. Кропотов, А.Ю. Проскуряков, А.А. Белов // Компьютерная оптика – 2018. – Т. 42, № 6. – С. 1093-1100. DOI: 10.18287/2412-6179-2018-42-6-1093-1100.
- [20] Lewis, J. Microservices / J. Lewis, M. Fowler, 2014 [Electronic resource]. –Access mode: <https://martinfowler.com/articles/microservices.html> (03.03.2019).
- [21] Kane, S. Docker: Up and Running: Shipping Reliable Containers in Production / S.P. Kane, K. Matthias – O'Reilly Media Inc., 2018. – 352 p.
- [22] Kazanskiy, N.L. Machine vision system for singularity detection in monitoring the long process / N.L. Kazanskiy, S.B. Popov // Optical Memory and Neural Networks (Information Optics). – 2010. – Vol. 19(1). – P. 23-30. DOI: 10.3103/S1060992X10010042.
- [23] Шаталин, Р.А. Обнаружение нехарактерного поведения в задачах видеонаблюдения / Р.А. Шаталин, В.Р. Фидельман, П.Е. Овчинников // Компьютерная оптика. – 2017. – Т. 41, № 1. – С. 37-45. DOI: 10.18287/2412-6179-2017-41-1-37-45.
- [24] Popov, S. Software threads of distributed image processing based on microservices // CEUR Workshop Proceedings. – 2017. – Vol. 1989. – P. 215-220.
- [25] Рыцарев, И.А. Кластеризация медиа-контента из социальных сетей с использованием технологии BigData / И.А. Рыцарев, Д.В. Кириш, А.В. Куприянов // Компьютерная оптика. – 2018. – Т. 42, № 5. – С. 921-927. DOI: 10.18287/2412-6179-2018-42-5-921-927.
- [26] Калиновский, И.А. Обзор и тестирование детекторов фронтальных лиц / И.А. Калиновский, В.Г. Спицын // Компьютерная оптика. – 2016. – Т. 40, № 1. – С. 99-111. DOI: 10.18287/2412-6179-2016-40-1-99-111.

- [27] Никитин, М.Ю. Нейросетевая модель распознавания человека по лицу в видеопоследовательности с оценкой полезности кадров / М.Ю. Никитин, В.С. Конушин, А.С. Конушин // Компьютерная оптика. – 2017. – Т. 41, № 5. – С. 732-742. DOI: 10.18287/2412-6179-2017-41-5-732-742.
- [28] Сорокин, В.Н. Распознавание личности по голосу: аналитический обзор / В.Н. Сорокин, В.В. Вьюгин, А.А. Тананыкин // Информационные процессы. – 2012. – Т. 12, № 1. – С. 1-30.

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН - соглашение № 007-ГЗ/ЧЗ363/26.

Digital Transformation Legacy Social Service Information System

S.B. Popov^{1,2}, P.V. Khripunov²

¹Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

²Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The article considers the methodological aspects of the digital transformation of legacy social service information systems. The result of this transformation is the social service digital platform that ensure the use of new innovative technologies for processing and analysing big data, machine learning and artificial intelligence. The base of this transformation is cloud computing; big data technology; microservice architecture with container-based software development technology. We propose a person digital profile as a basis of data management in the social service digital platform. The proposed one is based on a new abstraction level of software stack – dataware. The implementation uses a microservice architecture and a container-based software development technology in according to the Data-as-a-Service concept.