

УДК 004.6

**МЕТОДЫ И ПОДХОДЫ К ИНТЕГРАЦИИ ДАННЫХ, ПРОБЛЕМЫ
ИНТЕГРАЦИИ ИНФОРМАЦИОННЫХ СИСТЕМ****Янаева Марина Викторовна,**

кандидат технических наук, доцент,
Кубанский государственный технологический университет,
Россия, г. Краснодар.
yanaevam@mail.ru

Керопова Анна Сергеевна,

студент 3 курса,
Кубанский государственный технологический университет,
Россия, г. Краснодар.
keropova.anna@gmail.com

Харченко Анатолий Сергеевич,

студент 3 курса,
Кубанский государственный технологический университет,
Россия, г. Краснодар.
anatoliy.h90@mail.ru

Аннотация

В статье рассматривается понятие интеграции данных, её роль в разработке систем и бизнес-аналитике. Определены цели и задачи интеграции. Приведены основные методы интеграции систем, их достоинства, недостатки и различия. Перечислены проблемы интеграции, проведен обзор существующих решений.

Ключевые слова: интеграция данных, интеграция информационных систем, разработка, бизнес-процесс, данные, проблемы интеграции, интерфейс.

**METHODS AND APPROACHES TO DATA INTEGRATION, PROBLEMS OF
INFORMATION SYSTEM INTEGRATION****Yanaeva Marina Viktorovna,**

candidate of technical sciences, associate professor,
Kuban State Technological University,
Russia, Krasnodar.
yanaevam@mail.ru

Keropova Anna Sergeevna,

3rd year student,

Kuban State Technological University,
Russia, Krasnodar.
keropova.anna@gmail.com

Kharchenko Anatoly Sergeevich,

3rd year student,
Kuban State Technological University,
Russia, Krasnodar.
anatoliy.h90@mail.ru

ABSTRACT

The article discusses the concept of data integration, its role in system development and business intelligence. The goals and objectives of integration are defined. The main methods of systems integration, their advantages, disadvantages and differences are given. The problems of integration are listed, a review of existing solutions is carried out.

Keywords: data integration, information systems integration, development, business process, data, integration problems, interface.

Понятие интеграции данных, постановка проблемы

Создание единой информационной структуры предполагает интеграцию, как приложений, так и данных, это помогает автоматизировать управление процессами между ними и облегчить использование системы.

При паре информационных систем (далее – ИС), особых проблем не должно возникнуть, но со временем информационная структура может стать более комплексной, может понадобиться интеграция трех и более автоматизированных систем, и выбор способа интеграции, проведение интеграции и их совместная работа могут вызвать сложности.

При осуществлении интеграции разработчики чаще всего сталкиваются со следующими проблемами:

- ненадежность передачи данных по сети;
- низкая скорость передачи данных;
- технологические различия между интегрируемыми системами;
- неизбежность изменений в интеграционных потоках между системами после внесенных в них изменений.

Теперь, когда мы очертили назначение интеграции ИС на предприятии и перечислили основные проблемы, возникающие в ее процессе, сформулируем определение в общем смысле: интеграция данных включает в себя объединение данных, которые находятся в разных источниках, позволяя представить их пользователю в стандартизированном виде. Иначе говоря, «речь идет о поддержке представления совокупности данных из множества независимых источников в терминах единой модели данных.» [5]

Как мы указали выше, этот процесс используется в решении коммерческих задач, например, в случае, когда двум схожим компаниям нужно объединить базы данных, но

также интеграция данных может быть использована и в научных целях. С увеличением объема и необходимости совместного использования данных, повышается и роль интеграции.

Интеграция данных и бизнес-аналитика

Если переформулировать определение, интеграция – это автоматизированное взаимодействие между системами или сервисами, подразумевающая под собой общий бизнес-процесс для разных систем. Вопросами интеграции занимается бизнес-аналитика, которая ставит следующие вопросы:

1. В чем ценность интеграции? Для чего она проводится?
2. Какой целевой бизнес-процесс? Что нужно получить в итоге?
3. Что происходит в каждой системе?
4. Какими данными системы должны обмениваться?
5. Каковы условия обмена данными?
6. Какие функции должны выполняться при обмене данными?

Когда ответы на поставленные вопросы определены, возникает следующий – как обучить две разные системы общаться друг с другом? Каждая из них написана на своем языке программирования, имеет собственный стек и прочие особенности. Эта ситуация не нова и уже давно имеет решение, которое сводится к взаимодействию двух компьютеров. Для того, чтобы это взаимодействие реализовать, были созданы протоколы и форматы данных. По сути, это инструменты взаимодействия между компьютерами.

Протокол – это некое соглашение о том, в каком формате, порядке, с какими инструментами будут взаимодействовать две и более системы. Протоколов существует великое множество, самые распространенные и используемые: HTTP, TCP/IP, FTP. Обычно протокол подразумевает, что используется определенный формат данных. Протокол может быть строго завязан на одном формате или поддерживать несколько.

Формат данных – это некая конкретная форма представления данных, с предустановленным ограничением типа данных, например, JSON, XML, CSV.

Для того, чтобы произвести интеграцию между системами, нужно договориться об общем протоколе и формате данных. Иными словами, протокол – это связующее звено между интегрирующимися сервисами. Все это выливается в выбор способа интеграции.

Выделяют несколько основных способов интеграции:

- удаленный вызов процедур, самый распространенный на данный момент;
- обмен сообщениями, следующий по популярности;
- передача файлов, достаточно прост по сравнению с предыдущими;
- общая база данных.

1. Как же выбрать способ интеграции? При сборе требований полезно рассмотреть следующие вопросы:
2. Как быстро клиент хочет получать ответ?
3. Как долго сервер будет формировать ответ?
4. Какая доступность у систем?
5. Какой объем данных нужно формировать?
6. Как часто нужно передавать данные?
7. Нужно ли беспокоиться о безопасности?
8. Какие в компаниях уже есть способы интеграции?

9. В какие сроки нужно произвести интеграцию?

Не все протоколы подходят под определенные способы интеграции и нужды бизнес-процесса, поэтому подходить к этому нужно комплексно, исходя из целей интеграции и способностей разработчиков в компании.

Уровни интеграции данных

Существует несколько уровней интеграции данных, а именно: физический, логический и семантический. Рассмотрим каждый из них:

1. Интеграция на физическом уровне ставит как цель то, чтобы конвертировать данные из разрозненных источников в общий, единый формат физического представления, который требуется.
2. Интеграция на логическом уровне предполагает обеспечение доступа к данным, что содержатся в разных источниках, в терминах единой схемы, глобальной, которая предоставляет описание их совместного представления, учитывая структурные и поведенческие свойства данных.
3. Интеграция на семантическом уровне обеспечивает поддержку унифицированного представления данных с учетом их семантических свойств в контексте единой онтологии предметной области. [1]

На любом из уровней интеграции основная проблема – неоднородность источников данных. При интеграции данных на физическом уровне это различные форматы файлов, но логическом – неоднородность моделей данных или отличные друг от друга схемы данных, у которых модель данных одна. Источники также могут быть разными, одни веб-сайты, а другие – объектные базы данных. На семантическом уровне это различные онтологии, соответствующие различным источникам данных.

Исходя из требований к интеграционной системе, формируется целый ряд задач, к которым относятся:

- составление архитектуры системы интеграции данных;
- создание интегрирующей модели данных, которая станет основой унифицированного пользовательского интерфейса системы;
- разработка методов отображения моделей данных и их построение в интегрирующую модель для конкретных моделей, которые поддерживаются отдельными источниками данных;
- интеграция метаданных, которые используются в системе источников данных;
- преодоление неоднородности источников данных;
- разработка механизмов семантической интеграции источников данных.

Методы интеграции данных

Интеграция систем направлена на повышение эффективности бизнес-процессов компании, где используются ИС, чаще всего, это вынужденная, но необходимая мера, так как развитие компании невозможно без совершенствования систем, которые она использует, потому как представить даже слияние двух изначально различавшихся отделов без интеграции нельзя. Соответственно, и подходов к интеграции существует несколько. Мы перечислим основные методы:

1. Отсутствие интеграции между системами. Этот подход может привести к потраченному времени на выполнение повторяющихся процедур в разных системах, из чего может получиться ощутимая доля издержек фирмы.
2. Вертикальная интеграция. Здесь системы интегрируются по принципу функциональных экспертиз, где одна из экспертиз находится выше другой. Этот метод позволяет значительно сократить трудозатраты на повторяющиеся операции, но подобную систему крайне трудно подвергнуть функциональному расширению. [3]
3. Интеграция «многие ко многим». При таком методе интеграции, каждая из подсистем имеет возможность обратиться к функционалу любой другой подсистемы, также они могут использоваться друг другом. В этом случае возможности интеграции почти не ограничены, но поддерживать такую схему затратно при увеличении интегрированных подсистем. [4]
4. Горизонтальная интеграция. Суть данного подхода в использовании специального «промежуточного» (middleware) ПО. Преимущество его в том, что сами системы могут заменяться в пределах существующей спецификации опубликованных функций, при этом никаких изменений в других системах не требуется, а подключить новую систему достаточно просто. [3]
5. Отсутствие необходимости в интеграции. Случай, когда система изначально единая и обеспечивает связность между модулями гораздо лучше, чем при любом из вышеперечисленных подходов.

Так как информационная система представляет из себя совокупность компонентов, мы можем переформулировать определение интеграции ИС. Иными словами, интеграция ИС заключается в интеграции одного или нескольких компонентов интегрируемых объектов интеграции, которая делится на:

- интеграцию платформ;
- интеграцию данных;
- интеграцию приложений;
- интеграцию бизнес-процессов.

Проблемы интеграции данных

Теперь, когда мы дали определение интеграции систем, перечислили ее виды и разобрали методы и подходы к ней, поговорим о проблемах и их решениях.

Главной особенностью информационных систем является то, что они, как правило, работают не автономно, требуя надежного комплексирования со множеством сопряженных систем. [1] Создавая интегрированный комплекс, нужно обеспечить целостность и полноту информации, что используется в процессе их работы, а это не всегда представляется возможным по целому ряду причин.

Обзор существующих решений

Интеграция ИС сводится к обеспечению единого унифицированного интерфейса, предоставляющего доступ к определенной совокупности неоднородных и независимых источников данных. Важно отметить, что этот процесс «освобождает пользователей от необходимости знания, данные из каких источников, кроме интегрированного, они используют, каковы свойства этих источников и как осуществить доступ к ним.» [1]

Проблемы, возникающие при разработке и внедрении ИС можно разделить на категории: системные, методологические и организационные. Мы рассмотрели несколько видов проблем интеграции данных и перечислили варианты для решения или избежания этих проблем.

Проблемы интеграции системного характера:

1. Недостаточная корректность структурирования бизнес-процессов. Для решения проблемы недостатка глубины проработки сформированной структуры бизнес-процессов, которая в последствии может привести к рассогласованности производственных и функциональных процессов, наилучшим вариантом, по нашему мнению, будет формирование и ведение классификатора бизнес-процессов. Его задача будет заключаться в актуализации во время выполнения проектов создания информационных систем. При использовании такого решения в процессе разработки комплекса ИС, который обеспечивает требуемый уровень информатизации, параллельно будет создаваться база знаний. Она в свою очередь поспособствует последующему развитию и улучшению реализуемых процессов.
2. Излишнее усложнение решения по интеграции. Очень важным этапом является выбор платформы для интеграции данных, но он может повлечь за собой проблемы, так как вариантов платформ существует множество и не всегда представляется возможным ясно определиться, какая платформа подходит под определенные цели. Наш совет в таком случае: не покупать дорогое решение, если стоит цель только перенести данные между системами. Существуют бесплатные продукты, удовлетворяющие эту потребность, а если же она достаточно специфическая, можно написать собственный программный продукт. Самое главное, это помнить о том, что организация сложных процессов передачи больших объемов данных и их очистка требует действительно качественной интеграционной платформы.
3. Неправильное использование интеграционной платформы. Данная проблема очень схожа с предыдущей, поэтому совет общий и очень простой – нужно искать золотую середину. Если вам нужно преобразовать данные с помощью языка SQL, не имеет смысла приобретать серьезную интеграционную платформу вроде Informatica PowerCenter. Излишнее стремление к оптимизации и желание применить все существующие новейшие технологии и форматы только усложняет процесс интеграции, здесь нужна программа минимум. Если, например, при организации обмена данными через web-сервисы нет необходимости в протоколах с шифрованием, гарантированной доставкой, не стоит усложнять и без того тяжелый процесс передачи и обработки больших данных.

Проблемы интеграции методологического характера:

1. Недостаточная корректность выделения видов деятельности: Мы считаем, что для большинства организаций и предприятий, использующих

классификацию видов деятельности в Стратегии и Концепции информатизации, изначально требуется актуализация базового классификатора, а в дальнейшем и его детализация. Данное решение будет наилучшим среди прочих, так как классификатор используется в качестве отправной точки в процессе планирования процессов информатизации и при разработке новых проектов, а также для интеграции ранее созданных ИС.

2. Недостаточная корректность разделения бизнес-процессов жизненного цикла продуктов и активов. В качестве решения данной проблемы мы предлагаем ведение учета изменения состояния, конфигурации и иных характеристик активов и продуктов предприятия для возможности рассмотрения их взаимосвязи в любой момент времени этапа их жизненного цикла. Такое решение не даст снизить эффективность оптимизации.
3. Отсутствие информационных моделей бизнес-процессов. Чтобы избежать этой проблемы, мы советуем уделить достойное внимание информационным моделям при описании бизнес-процессов, так как в последующем возможно возникновение трудностей с вопросами организации хранения данных, размещения необходимых сведений, а также автоматизации процесса формирования документов. Часто данный фактор упускается в угоду структурирования самих процессов и создания схем обмена информацией на уровне классов документов. Необходимо помнить про каждый из вышеперечисленных пунктов.

Проблемы интеграции организационного характера:

1. Недостаточность координации исполнителей при разработке информационных систем. Исходя из накопленного опыта, в решении этой проблемы помогает следующее правило: сторона, заинтересованная в данных, должна выполнять всю основную работу по организации интеграции и ее дальнейшему сопровождению. Иными словами, при интеграции двух систем, ответственность за работоспособность несет либо большая, сложная и более развитая система, подгоняя под свои требования более маленькую систему, либо та система, у которой больше запросов и выгоды из интеграции. Вопросы об обеспечении качества нужно решать на берегу. Естественно, что залог хорошего результата – коллективная работа, но возникают ситуации, когда заинтересованных в интеграции нет, в таком случае к двум сторонам назначается ответственный от административного ресурса. Бизнес-аналитики и прочие бизнес-специалисты все также будут нести ответственность за качество данных и их передачи.
2. Недостаточность ресурсов для решения интеграционных задач комплекса информационных систем. Данную проблему целесообразно решать следующим образом – выносить интеграцию комплекса информационных систем в отдельный проект. Большинство организаций считают, что такой вопрос должен решаться исполнителями проекта самостоятельно, но на

практике выходит иначе. Опыт наглядно демонстрирует, что увязка систем представляет собой не частную задачу, но требует во многих случаях системных решений.

3. Закрытость служб сопровождения и разработчиков информационных систем компании. Избежать такой проблемы, по нашему мнению, возможно правильно организовав взаимодействие между заказчиком и разработчиком. Иначе бизнес-аналитик не сможет правильно сформулировать требования заказчика к интеграции систем, сложится неполная картина о существующих данных компании, что в результате приведет к неполному хранилищу этих самых данных.

Рекомендации и обобщения

Подводя итог нашего обзора на проблемы интеграции систем, мы хотим сформулировать общий вывод и рекомендации по этому немаловажному в работе любой компании процессу.

Первым делом, приступая к интеграции систем, нужно максимально четко выделить те системы, которые действительно нуждаются в обмене данными друг с другом, определить все возможные протоколы взаимодействия, сформулировать требования по объему данных, и уже после приступать к выбору программных средств. Нельзя забывать и о необходимости участия человека во взаимодействии ИС, ведь от этого зависит необходимость внедрения решений, часто не очевидных и не относящихся к интеграции в целом.

Из множества систем, подлежащих интеграции, нужно сразу определить те, которые войдут в связку. Если вначале произвести интеграцию между двумя системами с помощью определенного инструмента, а через время, осознав, что этого не хватает, попытаться внедрить третью – можно наткнуться на множество проблем, связанных с невозможностью применить предыдущий подход к интеграции. Будут потеряны время и большие средства, так как интеграция будет полностью реализована, а интеграционное программное обеспечение изучено, и лицензия на него приобретена.

Продолжая мысль, уже после сбора и документирования информации о будущем виде взаимодействий, нужно отобрать тот набор приложений, который будет покрывать все возможные варианты взаимодействия и протоколов, требования по виду и объему данных, по стилям интеграции, но при этом будет являться минимальным. При наличии такой минимальной репрезентативной выборки можно выбрать нужный набор продуктов и запускать пилотный проект.

Немаловажный совет: нет смысла покупать интеграционное программное обеспечение еще до готовности пилотного проекта, попробуйте получить доступ к временной лицензии и протестировать продукт. Это убережет вас от лишних затрат, и кроме того, вы будете иметь возможность получить техническую поддержку еще до покупки полной версии продукта.

В заключении наших рекомендаций следует сказать, что интеграция систем – сложная, но необходимая процедура для любого предприятия. С течением времени и развитием технологий во всех областях нашей жизни, компании все чаще нуждаются в усовершенствовании своей деятельности, расширении своих возможностей. От объединения двух отделов в один до соединения двух компаний в одну – везде понадобится интеграция. В зависимости от целей интеграции, нужно подобрать соответствующий подход, чтобы улучшить производительность бизнес-процессов и свести к минимуму

трудозатраты работников предприятия. Самое главное – не перестараться, ведь конечный продукт должен быть не только эффективным, но и понятным.

Список литературы:

1. Когаловский М. Р. Методы интеграции данных в информационных системах // Институт проблем рынка РАН. – 2010. – Т. 74.
2. Думченков, И. А. Обзор методов интеграции информационных систем, их преимуществ и недостатков / И. А. Думченков. – Текст: непосредственный // Молодой ученый. – 2018. – № 23 (209). – С. 176-177. – URL: <https://moluch.ru/archive/209/51296/>
3. Красиков Виктор Евгеньевич Обзор основных характеристик и проблем при интеграции информационных систем // Вопросы науки и образования. 2020. №15 (99). URL: <https://cyberleninka.ru/article/n/obzor-osnovnyh-harakteristik-i-problem-pri-integratsii-informatsionnyh-sistem>
4. Кусов А. А. Проблемы интеграции корпоративных информационных систем // Управление экономическими системами: электронный научный журнал. – 2011. – №. 28. – С. 103-109.
5. Когаловский М. Р. Интеграция данных в информационных системах // Сб. трудов Третьей Всероссийской конференции “Стандарты в проектах современных информационных систем”, Москва. – 2003. – С. 23-24.

References:

6. Kogalovsky M. R. Methods of data integration in information systems // Institute of Market Problems of the Russian Academy of Sciences. - 2010. - T. 74.
7. Dumchenkov, I. A. Overview of information system integration methods, their advantages and disadvantages / I. A. Dumchenkov. – Text: direct // Young scientist. - 2018. - No. 23 (209). – S. 176-177. – URL: <https://moluch.ru/archive/209/51296/>
8. Krasikov Victor Evgenievich Overview of the main characteristics and problems in the integration of information systems // Problems of science and education. 2020. No. 15 (99). URL: <https://cyberleninka.ru/article/n/obzor-osnovnyh-harakteristik-i-problem-pri-integratsii-informatsionnyh-sistem>
9. Kusov A. A. Problems of integration of corporate information systems // Management of economic systems: electronic scientific journal. – 2011. – no. 28. - S. 103-109.
10. Kogalovsky M. R. Data integration in information systems // Sat. Proceedings of the Third All-Russian Conference "Standards in the projects of modern information systems", Moscow. - 2003. - S. 23-24.