

ПРОБЛЕМЫ ИНТЕГРАЦИИ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ — МЕТОДЫ И ТЕХНОЛОГИИ

PROBLEMS OF INTEGRATION OF CORPORATE INFORMATION SYSTEMS — METHODS AND TECHNOLOGIES

A. Lebedev

Summary. The article considers the problems of integration of corporate information systems (IS) in the conditions of metallurgical holdings. The structure of the levels of integration of corporate IS and the resource planning system of the enterprise, as a central module of the IT landscape, are presented. The structure of general integration approaches is proposed at different architectural levels of IS, providing a uniform, unified presentation of data from different sources and taking into account different types of heterogeneity of IS. Various integration approaches are analyzed in accordance with the level of abstraction at which IS integration is performed. It is shown that explicit and accurate semantics of integrable data is important for semantically correct and meaningful integration results.

Keywords: information systems, resource planning system, corporate information integration, application integration, data integration, platform integration, single access to data.

Лебедев Александр Сергеевич

*Старший преподаватель, Уральский федеральный
университет имени первого Президента
России Б. Н. Ельцина
aslebedev@urfu.ru*

Аннотация. В статье рассмотрены проблемы интеграции корпоративных информационных систем (ИС) в условиях металлургических холдингов. Приведены структура уровней интеграции корпоративных ИС и система планирования ресурсов предприятия, как центрального модуля ИТ-ландшафта. Предложена структура общих интеграционных подходов на разных архитектурных уровнях ИС, обеспечивающих однородное, унифицированное представление данных из разных источников и учитывающих различные виды гетерогенности ИС. Проанализированы различные интеграционные подходы в соответствии с уровнем абстракции, на которой выполняется интеграция ИС. Показано, что явная и точная семантика интегрируемых данных важна для семантически правильных и значимых результатов интеграции.

Ключевые слова: информационные системы, система планирования ресурсов, интеграция корпоративной информации, интеграция приложений, интеграция данных, интеграция платформ, единый доступ к данным.

Анализ структуры информационной подсистемы современных крупных организаций, корпораций, холдингов показывает характеристики, общие для всех организаций. К этим характеристикам относится большое число информационных систем (ИС), автоматизирующих отдельные процессы, сферы деятельности и организационные подразделения. Информационное пространство включает ИС разных производителей, с разными датами запуска в эксплуатацию, использующие различные аппаратные и программные платформы, имеющие различную информационную архитектуру и т.д. Типичную ситуацию для крупных и средних современных организаций принято называть гетерогенным ИТ-ландшафтом.

Центральным модулем ИТ-ландшафта можно считать систему планирования ресурсов предприятия (Enterprise Resource Planning, ERP), разработанную в качестве важнейшего вспомогательного инструмента для обеспечения поддержки принятия управленческих решений. Эффективность управления и принятия решений была значительно улучшена ERP-системами, что способствует развитию предприятия. В последнее время в свя-

зи с ростом и расширением бизнеса, формированием холдингов, растущее предприятие планирует или уже достигло обновления и модернизации своей ERP-системы. Следовательно, различные традиционные системы ERP (legacy information system [1]) появляются в разных организациях и отделах.

Большинство традиционных ERP-систем представляют собой гетерогенные системы, которые могут разрабатываться различными компаниями-разработчиками программного обеспечения в рамках разных сред разработки, поэтому для организаций это создаёт большую проблему для внедрения интегрированной системы бизнес-аналитики на основе различных информационных систем ERP и для достижения совместного принятия решений в процессе управления. Вследствие развития и значительного прогресса в технологии ИС (например, сервис-ориентированная системная архитектура, бизнес-аналитика и т.д.) существует тенденция, согласно которой многие предприятия планируют разработать и внедрить систему бизнес-аналитики на основе своих существующих ERP-систем для оперативного и эффективного реагирования в динамичной бизнес-среде. По-

этому проблемы эффективного и действенного согласования и интеграции этих разнородных ERP-систем стали неотложной задачей, которая должна быть исследована и решена.

В целом, интеграция нескольких информационных систем направлена на объединение выбранных систем таким образом, чтобы они образовывали единое новое целое и давали пользователям возможность взаимодействия с единой информационной системой. Причина интеграции двояка: во-первых, учитывая набор существующих информационных систем, можно создать интегрированное представление для облегчения доступа к информации и ее повторного использования через единую точку доступа к информации. Во-вторых, учитывая определенную потребность в информации, данные из различных дополняющих информационных систем объединяются, чтобы получить более полную основу для удовлетворения потребности.

В контексте предприятия проблема интеграции обычно упоминается как интеграция предприятия (enterprise integration, EI). Корпоративная интеграция означает способность интегрировать информацию и функциональные возможности из различных информационных систем на предприятии. Это включает интеграцию корпоративной информации (enterprise information integration, EII), которая касается интеграции на уровне данных и информации, и интеграцию корпоративных приложений (enterprise application integration, EAI), которая рассматривает интеграцию на уровне логики приложения.

EAI представляет сложную и многогранную технологию, охватывающую все уровни корпоративной системы: архитектуру, процессы, программное и аппаратное обеспечение. EAI обеспечивает интеграцию следующих уровней:

- ◆ Интеграция бизнес-процессов (Business Process Integration, BPI). Для функционирования BPI компании необходимо определять, управлять и проводить процессы обмена корпоративной информацией среди различных бизнес-систем. На основе этого организация имеет возможность сократить расходы, упростить операции и улучшить обработку запросов клиентов.
- ◆ Интеграция приложений (Enterprise Application Integration, EAI). Назначением EAI является консолидация данных или функций отдельных приложений и подсистем с другими, вследствие чего достигается интеграция, приближенная к реальному времени. При этом возможна потребность специальной интеграции, в первую очередь, когда необходима консолидация существующего

приложения с вновь внедряемым ERP-приложением.

- ◆ Интеграция данных (Data Integration). Для успешной интеграции бизнес-процессов и приложений необходима интеграция систем баз данных и данных. Задачи интеграции данных включают консолидацию и анализ данных из учетных систем, сбор и подготовка данных (Extract Transform Loading, ETL), управление основными данными (Master Data Management) и ряд других. Перед началом интеграции, следует построить модель данных, определить местонахождение (идентифицировать) и каталогизировать данные. После завершения этих этапов данные становятся пригодными для совместного использования и распространения в системах баз данных.
- ◆ Стандарты интеграции (Standards of Integration). Для проведения интеграции данных нужно обобщить стандартные форматы данных. Для интеграции стандартами являются форматы, поддерживающие распространение и использование бизнес данных и информации, т.е. стандарты становятся основой осуществления интеграции корпоративных приложений.
- ◆ Интеграция платформ (Platform Integration). Завершением интеграции систем, как базовой архитектуры, так и программного и аппаратного обеспечения является интеграция разнесенных частей гетерогенных сетей. Интеграция платформ также относится к процессам и инструментам, которыми эти системы осуществляют оптимальный и безопасный обмен информацией. Результатом является беспрепятственная передача данных различными приложениями.

Интеграция нескольких информационных систем, как правило, направлена на объединение выбранных систем таким образом, чтобы они образовывали единое новое целое. Пользователям предоставляется однородное логическое представление данных, которые физически распределены по разнородным источникам данных. Для этого все данные должны быть представлены с использованием одних и тех же принципов абстракции (унифицированная глобальная модель данных и унифицированная семантика). Эта задача включает в себя обнаружение и разрешение конфликтов схемы и данных, касающихся структуры и семантики.

По определению общепризнанного "отца" технологии хранилища данных В. Inmon, проблема с различными приложениями и огромным объемом избыточности данных в том, что не существует единой "системы записи". Проблема заключается в архитектуре, созданной для поддержки ранних приложений. Избыточность данных является просто побочным продуктом неправильной

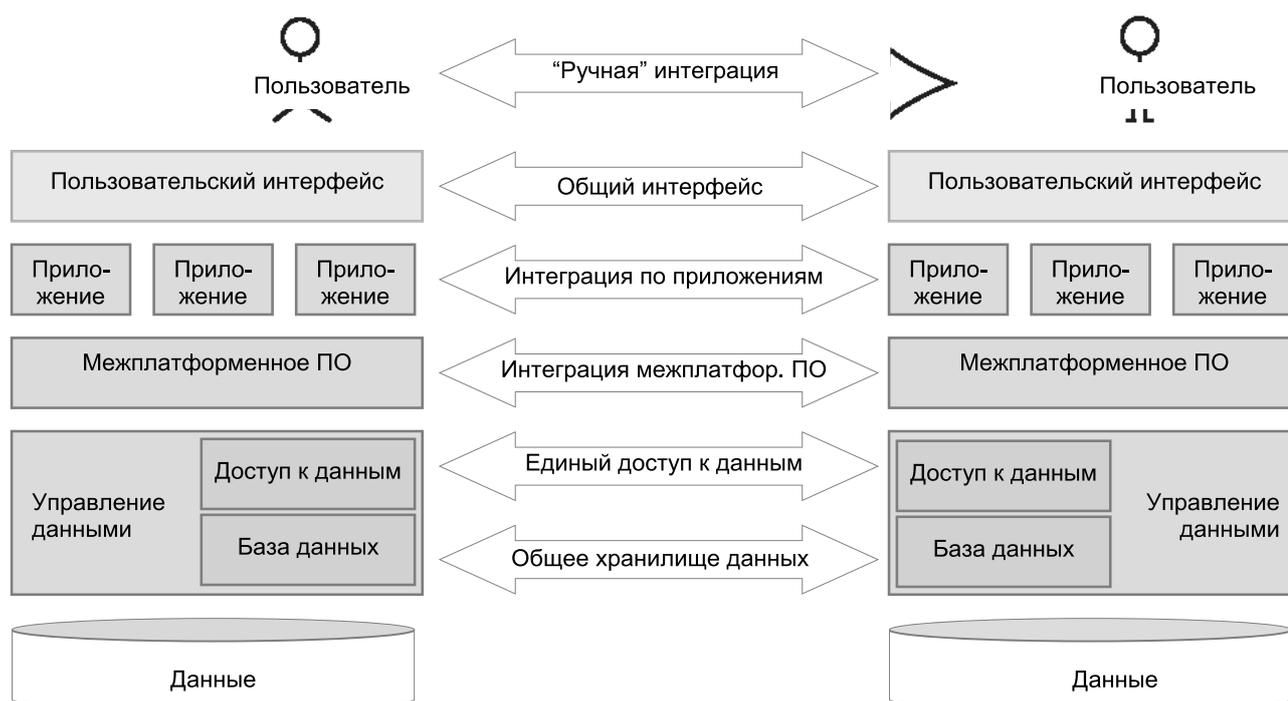


Рис. 1. Общие интеграционные подходы на разных архитектурных уровнях

архитектуры. Реальность такова, что огромные объемы избыточных данных находятся внутри корпорации [2].

В целом, информационные системы не ориентированы на интеграцию. Поэтому всякий раз, когда требуется интегрированный доступ к различным исходным системам, источники и их данные, которые не соответствуют друг другу, должны объединяться дополнительными функциями адаптации и согласования. Следует учитывать, что не существует единственной проблемы интеграции. Хотя цель всегда состоит в том, чтобы обеспечить однородное, унифицированное представление данных из разных источников, конкретная задача интеграции определяет:

- ◆ архитектурный вид информационной системы (рис. 1),
- ◆ содержание и функциональность компонентов системы,
- ◆ вид информации, которой управляют компонентные системы (буквенно-цифровые данные, мультимедийные данные; структурированные, полуструктурированные, неструктурированные данные),
- ◆ требования относительно автономности компонентных систем,
- ◆ предполагаемое использование интегрированной информационной системы (доступ только для чтения или запись),
- ◆ требования к производительности и

- ◆ доступные ресурсы (время, деньги, человеческие ресурсы и т.д.) [3].

Кроме того, обычно необходимо учитывать несколько видов гетерогенности. К ним относятся различия в:

- ◆ аппаратном обеспечении и операционных системах,
- ◆ программном обеспечении управления данными,
- ◆ моделях данных, схемах и семантике данных,
- ◆ промежуточном программном обеспечении,
- ◆ пользовательских интерфейсах и
- ◆ бизнес-правилах и ограничениях целостности.

Проанализируем различные интеграционные подходы в соответствии с уровнем абстракции, на которой выполняется интеграция.

Информационные системы могут быть описаны с использованием многоуровневой архитектуры, как показано на рис. 1.: на самом верхнем уровне пользователи получают доступ к данным и услугам через различные интерфейсы, которые работают поверх различных приложений. Приложения могут использовать промежуточное программное обеспечение — мониторы обработки транзакций (transaction processing, TP), промежуточное программное обеспечение, ориентированное на сообщения (message-oriented middleware, MOM), промежуточное программное обеспечение SQL и т.д., — для доступа к данным через уровень доступа к данным. Сами

данные управляются системой хранения данных. Обычно системы управления базами данных (СУБД) используются для объединения доступа к данным и уровня хранения.

В целом, проблема интеграции может быть решена на каждом из представленных уровней системы. Для этого доступны следующие основные подходы, как показано на рис. 1.

Ручная интеграция

Здесь пользователи напрямую взаимодействуют со всеми соответствующими информационными системами и вручную интегрируют выбранные данные. То есть пользователям приходится иметь дело с различными пользовательскими интерфейсами и языками запросов. Кроме того, пользователи должны иметь подробные знания о местоположении, логическом представлении данных и семантике данных.

Общий пользовательский интерфейс

В этом случае пользователю предоставляется общий пользовательский интерфейс (например, веб-браузер), который обеспечивает единый внешний вид. Данные из соответствующих информационных систем по-прежнему представляются отдельно, поэтому пользователям все еще необходимо обеспечить гомогенизацию и интеграцию данных (например, в поисковых системах).

Интеграция с помощью приложений

В этом подходе используются приложения интеграции, которые обращаются к различным источникам данных и возвращают интегрированные результаты пользователю. Это решение практично для небольшого числа компонентных систем. Тем не менее, приложения становятся все более объемными, так как число системных интерфейсов и форматов данных для гомогенизации и интеграции растет.

Интеграция с помощью межплатформенного ПО

Межплатформенное (промежуточное) программное обеспечение предоставляет повторно используемую функциональность, которая обычно используется для решения специализированных аспектов проблемы интеграции, например, как это делается в промежуточном программном обеспечении SQL. В то время как приложения освобождаются от реализации общих функций интеграции, в приложениях все еще требуются усилия по интеграции. Например, промежуточное программное обеспечение SQL предоставляет единую точку до-

ступа для отправки запросов SQL во все подключенные системы компонентов. Однако результаты запроса не интегрированы в один однородный набор результатов. Кроме того, для создания интегрированных систем обычно приходится объединять различные инструменты промежуточного программного обеспечения.

Единый доступ к данным

В этом случае логическая интеграция данных осуществляется на уровне доступа к данным. Глобальные приложения обеспечиваются единым глобальным представлением физически распределенных данных, хотя на этом уровне доступны только виртуальные данные. Локальные информационные системы сохраняют свою автономию и могут поддерживать дополнительные уровни доступа к данным для других приложений. Однако глобальное предоставление физически интегрированных данных может занимать много времени, поскольку доступ к данным, их гомогенизация и интеграция должны выполняться во время выполнения.

Общее хранилище данных

Здесь физическая интеграция данных выполняется путем передачи данных в новое хранилище данных; местные источники могут быть либо удалены, либо оставлены в рабочем состоянии. В целом, физическая интеграция данных обеспечивает быстрый доступ к данным. Однако, если локальные источники данных удалены, приложения, которые к ним обращаются, также должны быть перенесены в новое хранилище данных. Если локальные источники данных остаются работоспособными, необходимо периодически обновлять общее хранилище данных.

На практике конкретные интеграционные решения реализуются на основе представленных шести общих интеграционных подходов.

Вместе с тем интеграция представляется большим, чем просто структурная или техническая проблема. Технически довольно легко подключить разные реляционные СУБД. Более сложным является интеграция данных, описываемых различными моделями данных; еще сложнее проблемы, вызванные данными с разнородной семантикой. Следовательно, явная и точная семантика интегрируемых данных важна для семантически правильных и значимых результатов интеграции.

В области базы данных семантика может рассматриваться как интерпретация людьми данных и элементов схемы в соответствии с их пониманием мира в определенном контексте. При интеграции данных типом рассматриваемой семантики является, как правило, семантика

реального мира, связанная с отображением объектов в модельном или вычислительном мире в реальный мир [4]. В этом случае семантическая интеграция является задачей группировки, объединения или дополнения данных из разных источников с учётом явной и точной семантики данных во избежание структурного объединения семантически несовместимых данных. Следовательно, семантическая интеграция должна гарантировать, что объединяются только данные, относящиеся к одной и той же или достаточно сходной реальной сущности или концепции. Необходимым условием для этого является разрешение семантической неоднозначности в отношении интегрируемых данных с помощью явных метаданных, чтобы выявить все соответствующие неявные предположения и информацию, лежащую в основе контекста.

Онтологии, которые можно определить как явные, формальные описания концепций и их взаимосвязей, существующих в определенном пространстве дискурса, вместе с общим словарным запасом для ссылки на эти концепции — могут способствовать решению проблемы семантической гетерогенности. По сравнению с другими схемами классификации, такими как таксономии, тезаурусы или ключевые слова, онтологии позволяют создавать более полные и более точные модели предметной области [5].

Содействие гибкому сотрудничеству между предприятиями является еще одной областью, которая создает проблемы для интеграции данных. Например, для виртуальных организаций как наборов организационных единиц, которые работают для достижения общей цели, интеграция данных на лету чрезвычайно важна из-за их динамического характера [6]. Для эффективного своевременного предоставления необходимой информации всеми сотрудничающими партнерами, каждый из которых находится в различном реальном контексте, имея в виду собственную концептуальную модель мира, гибкая и адаптированная интеграция данных является необходимым условием. На основе адекватно интегрированных данных могут быть реализованы необходимые приложения, такие как управление цепочками поставок (SCM), планирование ресурсов предприятия (ERP) и управление взаимоотношениями с клиентами (CRM).

Поскольку одной из целей интеграции данных является предоставление унифицированного доступа к нескольким источникам данных, конфиденциальность и безопасность являются важными вопросами. Таким образом, необходимы гибкие, но эффективные средства контроля доступа в интегрированных системах [7]. Несмотря на то, что интеграция может обеспечить много преимуществ, интеграция данных и обмен данными часто затрудняются проблемами конфиденциальности [8].

Например, компании воздерживаются от обмена данными из-за боязни быть эксплуатируемой конкурентами или регулирующими органами. Аналогичным образом, интегрированный доступ к данным пациентов может способствовать развитию медицинских исследований, но может оказаться невозможным без проверенных мер защиты конфиденциальности и контроля доступа. Поэтому разработка методов, гарантирующих интеграцию данных и обмен данными без потери конфиденциальности, является крайне важной.

Качество данных, которое можно охарактеризовать с помощью точности, полноты, своевременности и согласованности данных, представляет большой интерес для удобства использования интегрированных данных. Однако в области интеграции данных часто необходимо принимать во внимание сложные потоки данных между производителями данных, интеграторами данных и потребителями интегрированных данных, чтобы обеспечить надлежащие решения для обеспечения качества данных. При этом, схемы с улучшенной онтологией, используемые в семантической интеграции данных, представляют собой важную предпосылку для высококачественных интегрированных данных и, таким образом, могут облегчить проблемы, связанные с качеством [9]. В частности, возможность для пользователей проверить, откуда данные берутся и как они были объединены и преобразованы в их текущую форму, имеют ключевое значение для предоставления пользователям возможности различать факты и убеждения и, как следствие, установить доверие к интегрированным данным [9]. Поэтому проблемы происхождения данных и прослеживаемости, вероятно, будут играть важную роль в будущих системах интеграции, особенно когда речь идет о сложных преобразованиях данных в широко распространенных источниках данных. Кроме того, глобальное обеспечение ограничений целостности может помочь пользователям доверять интегрированным данным из разных источников [10].

Таким образом, проведенный анализ проблем и основных подходов в области интеграции корпоративных информационных систем показал, что однозначного и эффективного решения для всех ситуаций пока нет. Наиболее сложные проблемы интеграции данных вызваны семантической неоднородностью.

Для обеспечения научного обоснования решения проблем интеграции корпоративных информационных систем, необходимо определить методы и модели управления основными данными, адекватные конкретной ситуации и задачам информационной системы предприятия, ориентируясь на формирование автоматизированных систем интеллектуального анализа данных.

ЛИТЕРАТУРА

1. Brodie M. L. The promise of distributed computing and the challenges of legacy information systems // *Interoperable database systems*. — 2013. — Vol. 7. — P. 1–31.
2. Inmon B. What is data redundancy? — <https://searchdatamanagement.techtarget.com>.
3. Dittrich K.R., Jonscher D. All Together Now — Towards Integrating the World's Information Systems // In: *Advances in Multimedia and Databases for the New Century*. — World Scientific Press. — Kyoto, Japan, November 30 — December 2, 2018. — P. 109–123.
4. Ouksel A.M., Sheth A. P. Semantic Interoperability in Global Information Systems: A Brief Introduction to the Research Area and the Special Section. / *SIGMOD Record*. — 2017. — Vol. 28(1). — P. 5–12.
5. Huhns M.N., Singh M. P. Agents on the Web: Ontologies for Agents // *IEEE Internet Computing*. — 2017. — Vol. 1(6). — P. 81–83.
6. Winslett M. Databases in Virtual Organizations: A Collective Interview and Call for Researchers // *SIGMOD Record*. — 2015. — Vol. 34(1). — P. 86–89.
7. Jonscher D., Dittrich K. R. An Approach for Building Secure Database Federations // In: *20th International Conference on Very Large Data Bases (VLDB2014)*. — Morgan Kaufmann. — Santiago de Chile, Chile, September 12–15, 2014. — P. 24–35.
8. Clifton C., et al. Privacy-Preserving Data Integration and Sharing // In: *9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD2014)*. — Paris, France, June 13, 2014. ACM. — P. 19–26.
9. Gertz M., et al. Report on the Dagstuhl Seminar "Data Quality on the Web" // *SIGMOD Record*. — 2014. — Vol. 33(1). — P. 127–132.
10. Conrad S., et al. Schema Integration with Integrity Constraints // In: *15th British National Conference on Databases (BNCOD2017)*. — London: Springer.UK. — July 7–9, 2017. — P. 200–214.

© Лебедев Александр Сергеевич (aslebedev@urfu.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»



Уральский федеральный университет им. первого Президента России Б.Н. Ельцина