

ОСНОВНЫЕ ПРИНЦИПЫ СОЗДАНИЯ ЕДИНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА КОРПОРАТИВНЫХ СИСТЕМ

Abstract: In the paper there are discussed the principles of creating unique information space of the corporative systems, i.e. the systems of great dimension, combining hetero geneour components, developed in different time and different designers. There are marked out main, more complicated aspects of creating unique information space, discussed problems, arising when creating planed the ways of their decision and possibilities of creating correspondent instruments.

Key words: unique information space, corporative systems, data representation, data synonymy, data semantix.

Анотація: У статті розглянуті принципи створення єдиного інформаційного простору корпоративних систем, тобто систем великої розмірності, що поєднують різномірні компоненти, розроблені в різний час і різними розробниками. Виділено основні, найбільш складні аспекти створення єдиного інформаційного простору, розглянуто проблеми, що виникають при створенні, намічені шляхи їхнього рішення і можливості створення відповідних інструментальних засобів.

Ключові слова: єдиний інформаційний простір, корпоративні системи, представлення даних, синонімія даних, семантика даних.

Аннотация: В статье рассмотрены принципы создания единого информационного пространства корпоративных систем, т.е. систем большой размерности, объединяющих разнородные компоненты, разработанные в разное время и различными разработчиками. Выделены основные, наиболее сложные аспекты создания единого информационного пространства, рассмотрены проблемы, возникающие при создании, намечены пути их решения и возможности создания соответствующих инструментальных средств.

Ключевые слова: единое информационное пространство, корпоративные системы, представление данных, синонимия данных, семантика данных.

1. Вступ

При создании мощных распределенных информационных, информационно-управляющих, учетных и многих других типов корпоративных систем, основной целью создания которых является повышение эффективности основной деятельности корпорации, разработчики сталкиваются с проблемой создания единого информационного пространства корпоративной системы (ЕИП КС).

Такое информационное пространство необходимо для создания систем поддержки принятия стратегических решений руководством корпорации при работе в ситуационных центрах [1].

Разновидностью корпоративных систем являются системы органов государственного управления. Проблемы создания ЕИП для системы Министерства обороны рассмотрены в [2].

Информационное пространство корпоративной системы, т.е. системы большой размерности, распределенной территориально и функционально, должно представлять собой систему, позволяющую надежно хранить большие объемы данных различного типа, оперативно их обновлять, вести эффективный поиск необходимых данных, их группирование и агрегирование. При этом пользователь должен иметь ясное и наглядное представление о содержании информации и ее структуризации. Пользовательское представление может не совпадать с физическим видом хранения данных, но должно давать возможность строить эффективно работающие запросы к данным.

Создание ЕИП КС предполагает «изотропность» информационного пространства, т.е. наличие единой спецификации средств (языка манипулирования данными) для обработки, подготовки и представления данных. Это означает, что любое приложение, разработанное для любого фрагмента КС, будет работать в любом другом фрагменте без перепрограммирования.

ЕИП КС должно удовлетворять условию «взаимопонимания», т.е. информация в любом фрагменте КС должна быть доступна для пользователя любого другого фрагмента без «переводчиков» в соответствии со своими метаданными.

Сформированные выше требования к ЕИП можно рассматривать как основные принципы, реализация которых при создании ЕИП КС может обеспечить необходимый уровень сервиса для разработчиков любых приложений, создаваемых для обеспечения эффективной деятельности корпорации.

Создание ЕИП позволяет снизить затраты на разработку любого типа корпоративной системы и позволяет повысить эффективность и отдачу от разработанной системы.

Ниже рассмотрим более подробно возможности реализации принципов создания ЕИП.

2. Единое представление данных

Первым составляющим элементом ЕИП должно быть единое представление данных о нем (о составе и структуре) у всех его пользователей. Это единое представление может базироваться на любой общепринятой модели данных: реляционной, иерархической, сетевой и т.п. Причем единое представление данных должно иметь, как минимум, две формы визуализации: логическую и физическую.

Логическая форма – это представление, которое отображает взгляд пользователя корпорации, т.е. любого отраслевого специалиста (финансиста, экономиста, специалиста в технических направлениях деятельности корпорации и т.п.). Эта форма базируется на документах из системы документооборота корпорации, технической и нормативной документации, документах других предприятий и организаций.

Физическая форма – это формализованное представление в виде, соответствующем языку описания данных используемой СУБД, и форме представления реляционной модели данных, как общепринятому стандарту *de facto*.

Физическая форма представления данных технически хорошо реализована во многих программных продуктах, предназначенных для проектирования информационных систем (программные изделия с названием “Designer” разных фирм). Практически любая из этих программ позволяет описать реквизитный состав информационных структур (таблиц, деревьев), их связи между собой, наборы идентификационных реквизитов (ключи) и т.п. Подобное описание обычно дает пользователю-профессионалу ИТ все необходимые данные для успешной работы в информационном пространстве системы.

Сложнее решается вопрос с логической формой представления информационного пространства для отраслевого специалиста корпорации, который должен получить ответ на вопрос в терминах своего представления о работе корпорации. Это может быть представление экономиста, финансиста, топ-менеджера и любого другого специалиста, включая технических: технологов, конструкторов, строителей. Эти специалисты должны получить данные о ЕИП в своем понятийном представлении.

Наиболее близко к такому представлению развернутое текстовое описание с иллюстрациями, накладывающими описание отдельных единиц хранения на структурную схему

организации (корпорации) с выделением точек сбора, концентрации данных. Однако создание такого представления требует большого труда по реализации и ведению описания и графического представления. Готовой разработанной системы пользовательского представления, удовлетворяющего всем требованиям функционального специалиста, не существует. Наиболее близко к необходимому уровню описания приближаются развитые системы управления проектом с подсистемами ведения репозитория метаданных. За счет допустимых в этих системах развернутых комментариев и форм графического представления структур проекта можно с достаточной степенью точности описать ЕИП.

Для дальнейшего эффективного использования идеологии ЕИП необходимо разработать или адаптировать программные средства логического представления содержимого и структур ЕИП для отраслевого пользователя корпорации.

3. «Изотропность» единого информационного пространства

«Изотропность», т.е. сохранение свойств в любом фрагменте корпоративной сети, должна быть характерной для инструментария разработчика прикладного программного обеспечения. Разработчик включает в программное обеспечение операторы доступа к данным (чтения, записи, обновления, установления прав доступа и т.п.), которые одинаково выполняются в любом фрагменте корпоративной сети независимо от места нахождения данных и от места их потребления.

Для обеспечения этого требования может быть предложено несколько подходов. Первый, наиболее простой и эффективный, заключается в использовании во всех фрагментах корпоративной сети одной СУБД, которая имеет опции создания и ведения распределенного БД. К таким СУБД относятся СУБД типа ORACLE, INFORMIX, MS SQL Server и т.п. Используя возможности подобной СУБД, создается единая схема БД, на основе которой любой разработчик приложений получает инструментарий манипулирования данными с помощью операторов языка SQL с расширениями, введенными фирмой-производителем СУБД. Обычно возможностей фирменного языка SQL и аппарата хранимых процедур СУБД с избытком хватает для реализации любого алгоритма обработки данных в приложении пользователя.

Другим подходом является реализация ЕИП на основе нескольких типов СУБД, работающих в различных фрагментах корпоративной сети. Как правило, все современные СУБД имеют SQL-интерфейс, расширенный различными полезными дополнениями, внесенными фирмами-производителями. Однако эти дополнения у различных фирм-производителей различаются между собой как видом (интерфейсом) операторов, так и результатами их применения. Поэтому без дополнительных доработок универсальным средством манипулирования данными является стандарт языка SQL, который поддерживается всеми коммерческими СУБД. Однако использование аппарата хранимых процедур в этой конфигурации ЕИП невозможно, поскольку каждая фирма-разработчик поддерживает свои синтаксис и семантику языков создания таких процедур.

Указанные ограничения сильно осложняют разработку универсальных (в рамках корпорации) приложений и вызывают дополнительные расходы на адаптацию приложений для различных фрагментов корпоративной сети.

И самым неудобным для эксплуатации и разработки приложений является ситуация, когда часть ЕИП ведется средствами, отличными от SQL СУБД. В таких условиях становится невозможным создание ЕИП без проведения предварительных доработок. Самой радикальной доработкой может быть замена средств ведения базы данных на одну из коммерческих SQL СУБД. Конечно, такая операция потребует вложения определенных средств, но это вложение одноразовое и окупится возможностью использовать без дополнительных затрат приложения, разработанные в других подразделениях корпорации.

Замена всех средств ведения данных на SQL СУБД приведет к созданию ЕИП со средствами манипулирования данными, как описано для первого варианта создания изотропной среды манипулирования данными.

Еще одним методом создания «изотропности» является разработка специальных программных средств «перевода» операторов SQL в операторы той системы ведения данных, которая используется на данном фрагменте корпоративной сети. Подобные системы разрабатывались для создания распределенных систем в авиастроении [3]. Подобная же система предлагалась для использования фирмой Taxido.

Однако, поскольку последние 10-15 лет SQL-интерфейс утвердился как стандарт de-facto, практически все коммерческие СУБД имеют этот интерфейс. Использование других языков манипулирования данными сохранилось в системах, разработанных в 70-80 годах прошлого века. Как показывает практика, в странах СНГ систем, разработанных до 91 года, в эксплуатации не сохранилось. Поэтому такой вариант создания «изотропной» среды для этого региона не актуален.

Из всего вышесказанного можно сделать вывод, что наиболее эффективным и практически реализуемым является использование единой или нескольких SQL СУБД, что вполне доступно при минимальных затратах.

4. Синонимия метаданных у семантически совпадающих данных

Если каждый фрагмент корпоративной сети представляет собой разработанную информационную систему со своим набором структур БД и если эти системы создавались в разное время, то одни и те же данные о работе подразделений могут дублироваться многократно и иметь различные источники поступления в БД.

Наименования одних и тех же данных (метаописание данных) будут различными в разных фрагментах, т.е. в различных фрагментах будут наличествовать синонимы в метаданных.

Структуры описания данных существенно отличаются друг от друга, особенно если они созданы различными разработчиками и в разное время. Семантически одинаковые показатели в различных фрагментах сети могут иметь различные единицы измерения, и некоторые из них могут быть производными друг друга. Одни и те же показатели могут иметь различные наименования и относиться к различным подсистемам, то есть они между собой не сравниваются и могут иметь несоординированные значения. Для разрешения этой задачи необходимо применение специальных методов, которые должны быть формализованы и автоматизированы. Только при таком условии могут быть получены однозначные и адекватные результаты.

Семантический анализ показателей корпоративной системы, имеющей множество фрагментов сети, в каждом из которых функционирует своя, автономно разработанная

информационная система, может быть проведен метод определения семантики реквизитов таблиц БД через создание для каждого реквизита набора (вектора) понятий, который однозначно определяет содержание реквизита. Если понятия для характеристики реквизита берутся из определенного тезауруса и есть формализованная процедура обработки векторов понятий, то определение совпадений семантического содержания реквизитов таблиц из БД различных фрагментов корпоративной системы может быть достигнуто. После нахождения пар (троек, четверок и т.д.) семантически совпадающих реквизитов необходимо оставить ведение этого реквизита в одном фрагменте, если он имеет общекорпоративное значение, а в остальных организовать ссылку на него. Для определения основного места ведения реквизита анализируется первичность появления значений реквизита (из первичных или вторичных документов берется значение), временной разрез, т.е. частота отображения значений в БД должна быть достаточной для всех применений, точность представления и, возможно, некоторые другие специфические требования.

Если же найденные семантические синонимы имеют местные (локальные) значения, то имена этих реквизитов заносятся в таблицу синонимов, и таблица может быть использована для построения обобщенных документов и создания хранилищ данных.

Проблема создания тезауруса понятий для описания экспертами семантики реквизитов имеет несколько возможных решений.

Может быть создан тезаурус на основе объединения понятий из соответствующих стандартов типа «Термины и определения». Может быть использован метод создания словаря предметной области [4]. Могут быть использованы различные интерпретации таких методов.

4. Выводы

Суммируя рассмотренные принципы создания ЕИП корпоративных систем можно констатировать, что для обеспечения реальных сроков создания ЕИП и достижения высококачественных результатов необходимо разработать:

- алгоритмические и программные средства для работы в отдельных направлениях создания ЕИП;
- создать комплекс инструментальных средств для создания ЕИП на основе существующих автономных систем;
- систему для тестирования созданных инструментальных средств (на примере создания ЕИП достаточно большой размерности для отработки всех аспектов).

СПИСОК ЛИТЕРАТУРЫ

1. Морозов А.А. Ситуационные центры – основа управления организационными системами больших размерностей // Математические машины и системы. – 1997. – № 2. – С. 7–10.
2. Морозов А.О., Яровий А.Д., Пилипенко Ю.Г. и др. Основні положення концепції розподіленого інтегрального банку даних у складі ЕАСУ ЗС // Математические машины и системы. – 1999. – № 2. – С.143–152.
3. Кузьменко Г.Е., Пилипенко Ю.Г. Особенности проектирования распределенного банка данных КАСУ и программные средства доступа //Управляющие системы и машины. – 1989. – № 2.
4. Кунгурцев А.Б., Тихан И.В. Формирование словаря предметной области // Искусственный интеллект. – 2006. – № 1. – С.166–172.