

МЕТОДИКА ПРОЕКТИРОВАНИЯ КОРПОРАТИВНОГО ХРАНИЛИЩА ДАННЫХ НА БАЗЕ ПЛАТФОРМЫ SAP NET WEAVER BUSINESS WAREHOUSE

С.А. Тоноян

tonoyansl@mail.ru

В.А. Высочанский

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Приведен сравнительный анализ существующих методов построения классических хранилищ данных. Показано, что при проектировании корпоративного хранилища данных на базе ключевых показателей эффективности такие методы используют лишь частично. Предложена методология, описывающая конкретные шаги по созданию корпоративного хранилища данных, которое имело бы свойства архитектуры Layered Scalable Architecture. Реализован новый практический подход для проектирования корпоративного хранилища данных на базе платформы SAP Net Weaver Business Warehouse, использующий стандарт Layered Scalable Architecture, предлагаемый компанией SAP. Приведена практическая реализация разработки корпоративного хранилища данных отдела таможенного оформления подразделения нефтегазовой компании, подтверждающая адекватность и работоспособность описываемой методики

Ключевые слова

Хранилище данных, SAP, многомерная модель данных, витрина данных, схема-звезда, OLAP, OLTP, инфокуб, агрегирование, LSA

Поступила в редакцию 26.05.2015

© МГТУ им. Н.Э. Баумана, 2016

Актуальность работы. Дочерние компании крупных предприятий используют изолированные ERP-системы SAP R/3, а подразделения — небольшие витрины данных на базе SAP Net Weaver Business Warehouse (SAP BW). При необходимости многокритериального анализа бизнеса возникает потребность в показателях эффективности работы дочерних компаний и подразделений. Организовать единую отчетность On-Line Analytical Processing (OLAP) на совокупности разнородных систем On-Line Transaction Processing (OLTP) не представляется возможным из-за сложности запросов и недопустимо длительного времени их исполнения [1, 2].

Следовательно, требуется создать систему ключевых показателей эффективности (КПЭ), которая консолидирует данные из разрозненных исходных OLTP-систем для анализа. Если прибегнуть к созданию классического хранилища данных, то возникнут трудности в поддержке большого числа источников данных

(экстракторов). Поэтому возникает потребность в подходе, описывающем способ проектирования унифицированных хранилищ, витрин данных, совокупность которых объединяется в корпоративное хранилище данных (КХД), обладающее гибкой структурой, и предоставляет пользователям единую OLAP-отчетность.

Корпоративное хранилище данных представляет собой систему обработки и многомерного анализа оперативных, исторических и прогнозных данных предприятия на основе системы КПЭ.

Наиболее популярным инструментальным программным средством для реализации хранилищ данных крупных предприятий является система SAP BW. В методической литературе, предлагаемой компанией SAP и сторонними авторами, чаще всего описываются принципы построения классических хранилищ данных, которые частично используются при проектировании КХД на базе КПЭ. Таким образом, возникает необходимость в методике, учитывающей специфику КХД и возможности SAP BW при проектировании хранилища данных.

Классическая структура хранилища данных на базе SAP BW. Неотъемлемой чертой любого хранилища данных является наличие средств аналитической обработки в реальном времени OLAP, иногда называемых средствами многомерного анализа. Данные инструменты основаны на концепции многомерной модели базы данных, позволяющей исключить недостатки использования реляционной базы данных с высокой степенью нормализации, которые задействованы в OLTP-системах, ориентированных на обработку транзакций в реальном времени [2, 3]. Платформа SAP BW предоставляет собой широкий набор инструментов OLAP, в основе которых лежит идея построения OLAP-кубов (инфокубов в терминологии SAP).

Связь между OLTP- и OLAP-системами реализуется в виде ETL-процессов (Extract, Transform, Load — извлечение, преобразование, загрузка), схематическое изображение которых показано на рис. 1.

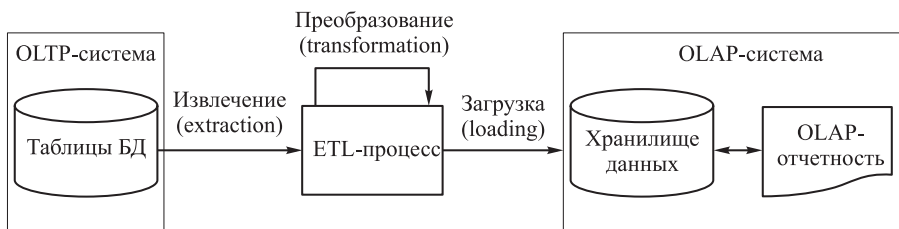


Рис. 1. ETL-процесс

Хранилище данных в SAP BW строится из набора инфокубов. Каждый состоит из одной таблицы фактов и нескольких таблиц измерений. Подобная структура денормализована и нередко избыточна в целях повышения скорости выполнения запросов к инфокубам, что достигается при отсутствии необходимости в соединении (JOIN в терминологии SQL) множества таблиц и выполнении специализированных запросов [1, 3]. Непосредственно на OLAP-кубах в SAP BW строятся аналитические отчеты, являющиеся конечной целью разработки хранилища данных.

Основной таблицей инфокуба является таблица фактов, в которой отражаются некоторые события, значимые для дальнейшего анализа [4]. Очевидно, что для описания любого факта требуется набор параметров, уникально идентифицирующих данный факт, и набор числовых значений, характеризующих его. В SAP BW параметры фактов содержатся в таблицах измерений, а числовые характеристики называются показателями и находятся непосредственно в таблице фактов [5]. Таким образом, значение показателя соответствует уникальной комбинации из ключевых полей измерений.

Структура классического OLAP-куба, часто называемая схемой-звездой, приведена на рис. 2.

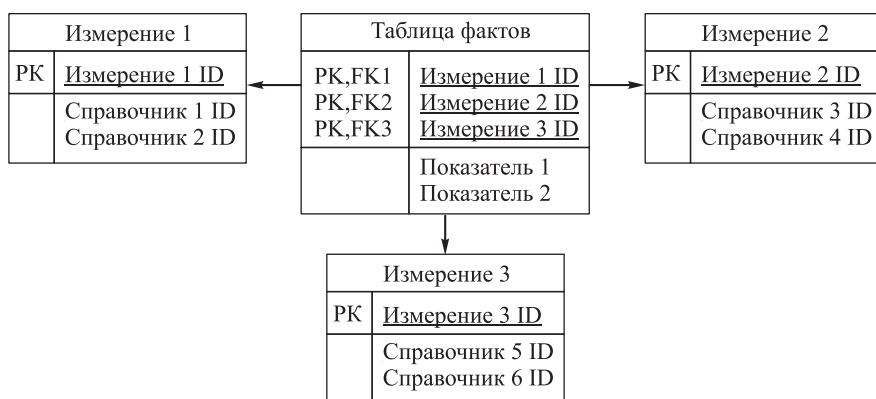


Рис. 2. Классическая многомерная модель OLAP-куба

В SAP BW используется усовершенствованная схема-звезда, которая отличается от классической схемы тем, что значения справочников не хранятся в таблицах измерений. В SAP-схеме для каждого справочника генерируется так называемый «суррогатный ключ» (SID), который заменяет ключ справочника в таблице измерения. Таким образом, несколько инфокубов могут независимо использовать одни и те же справочники в таблицах измерений [2]. Расширенная схема-звезда, используемая в SAP BW, приведена на рис. 3.

Как было отмечено ранее, структура OLAP-куба оптимизирована для запросов чтения благодаря денормализации [4]. Однако системы, из которых данные поступают в хранилище (т.е. в инфокубы) очень часто построены по совершенно иным принципам: к примеру, ERP-системы базируются на реляционных базах данных в третьей нормальной форме. Это означает, что между хранилищем данных и исходной системой должен существовать промежуточный уровень, преобразующий информацию без потерь и агрегирования [4].

Основным источником данных для любого хранилища, как было отмечено ранее, является OLTP-система, которая в семействе программных продуктов SAP представлена платформой R/3. Для создания потока данных между системами SAP R/3 и SAP BW необходимо задействовать процессы ETL [4], содержащие следующие стадии (рис. 4).

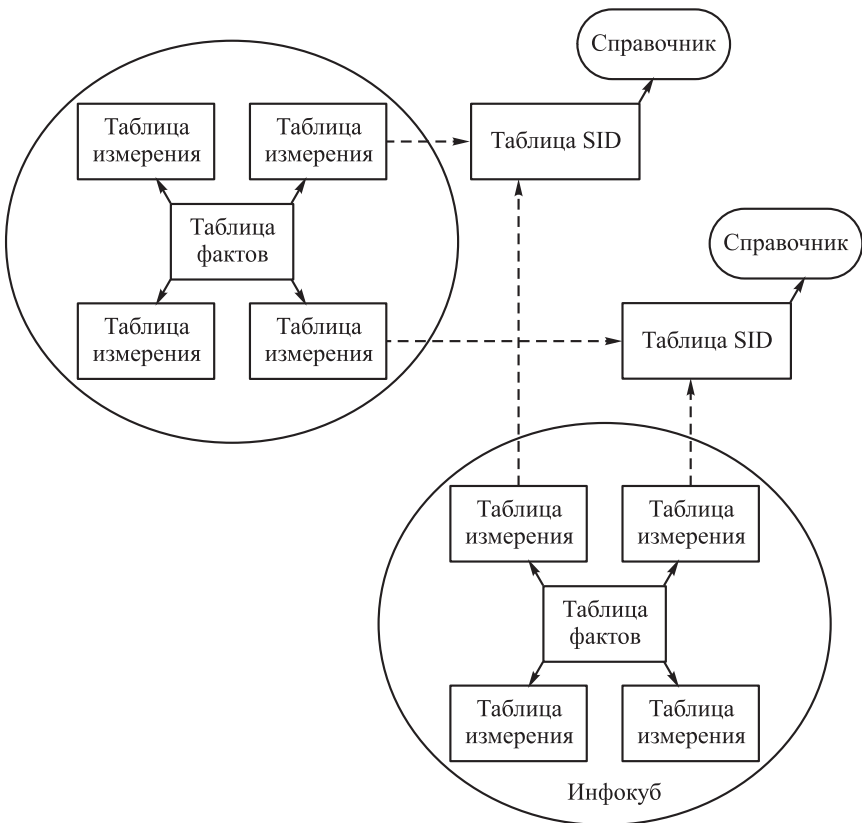


Рис. 3. Расширенная схема-звезда SAP BW

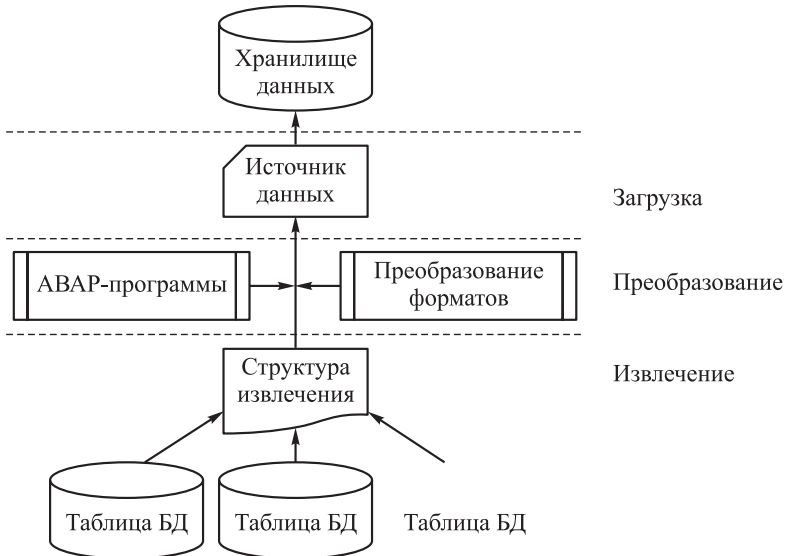


Рис. 4. ETL-процесс в SAP BW

1. Извлечение данных из таблиц OLTP-системы с помощью структур извлечения — особых виртуальных таблиц, являющихся проекциями реально существующих таблиц базы данных.

2. Преобразование извлеченных данных посредством стандартных операций (перевод в другой формат) или программ, написанных на встроенном объектно-ориентированном языке АВАР.

3. Загрузка в хранилище данных.

Требования к проектированию КХД. Корпоративное хранилище данных, в отличие от классического, должно содержать данные разной степени агрегирования. Это означает, что одна и та же запись, извлеченная из исходной системы, хранится в КХД на нескольких уровнях:

а) изначальное представление — наиболее детализированная информация;

б) промежуточное представление — данные агрегированы по нескольким признакам, например: выручка в рублях из сбытовых контрактов суммируется по странам, т. е. исключается признак «Сбытовой контракт», но остается признак «Страна»;

в) окончательное представление — данные находятся в наиболее обобщенном виде, в котором набор ограничен пятью–десятью измерениями.

Отсюда следует, что для построения КХД недостаточно использовать систему инфокубов, на каждом из которых базируется OLAP-отчет, так как подобная схема позволяет хранить данные только на одном уровне детализации.

Следующая особенность КХД — наличие ключевых показателей эффективности. В отличие от показателей классического хранилища данных, повторяющих показатели исходной системы (суммы, объемы, выручки, расходы, прибыль и т. п.), КПЭ наиболее точно отражают состояние бизнеса. Очевидно, что концепция КПЭ подразумевает сложную систему расчета и извлечения данных (иногда для расчета одного КПЭ требуются данные из разных несвязанных систем), не предоставляемую классическим хранилищем.

Как правило, КХД внедряют достаточно крупные предприятия–учредители, желающие проводить анализ эффективности работы своих дочерних предприятий. Отсюда следует, что единое КХД разделяется на «витрины данных» — небольшие хранилища отдельных дочерних обществ, формирующие КПЭ в рамках своей деятельности [4, 5]. Часто возникает необходимость во взаимной интеграции витрин в связи с получением обобщенных КПЭ.

Обязательное наличие оптимизированной структуры хранения архивных (исторических) данных также отличает КХД от обычного хранилища [6, 7]. Как правило, в каждой витрине данных создается архивный OLAP-куб, в который передаются наиболее агрегированные записи (для уменьшения объема хранимой информации) по закрытым отчетным периодам. В отличие от обычных инфокубов, архивный куб редко используется для построения OLAP-отчетов — его основной целью является сохранение исторических данных, которые в перспективе можно использовать для расчета, например, прогнозных значений [2].

Таким образом, для реализации КХД требуется организовать особый подход, учитывающий его особенности и предполагающий его использование в рамках крупного предприятия.

Классификация подходов к проектированию КХД. Хранилище данных как система, использующая технологии OLAP, может быть реализована посредством одного из следующих принципов, касающихся способов представления данных.

Реляционная OLAP-система (ROLAP) — данные хранятся в реляционной базе данных, как и таблицы фактов и измерений OLAP-кубов. Для организации агрегатов необходимо создание отдельных таблиц.

Многомерная OLAP-система (MOLAP) — информация содержится непосредственно в упорядоченных многомерных массивах (OLAP-кубах) в виде значений показателей, отраженных относительно фиксированного набора измерений. Классическими схемами MOLAP являются модели «звезда» и «снежинка».

Гибридная OLAP-система (HOLAP) — сочетание технологий ROLAP и MOLAP, наиболее детальные данные хранятся в реляционной базе данных, а агрегированные данные для отчетов — в многомерных структурах.

Следует отдельно упомянуть виртуальное хранилище данных (ВХД) — это система, эмулирующая работу хранилища данных и внедряемая в существующую OLTP-систему в качестве отдельного слоя в виде набора представлений (view). Преимуществом ВХД является возможность формирования аналитической отчетности без организации новых структур хранения и процессов преобразования данных в новый формат. Тем не менее, к очевидным недостаткам ВХД относятся увеличение нагрузки на OLTP-систему при менее производительной (по сравнению с самостоятельными OLAP-системами) системой аналитических запросов, а также невозможность работы с историческими данными.

Платформа SAP BW позволяет создавать хранилища данных по принципам гибридных OLAP-систем. В настоящее время в области построения корпоративных хранилищ данных можно выделить несколько общепризнанных подходов, применяемых в тех или иных ситуациях. Наиболее обобщенными являются подходы «сверху–вниз» и «снизу–вверх».

При более детальной классификации выделяют следующие подходы к проектированию КХД.

Объединенные хранилища данных / витрины данных (Federated Data Warehouse/Federated Data Mart) — это подход, основанный на методе «снизу–вверх», но не предполагающий построение хранилища данных уровня предприятия. В соответствии с FDW/FDM на предприятии используются несколько самостоятельных хранилищ данных, которые связаны между собой наличием общих точек соприкосновения: некоторых источников данных, объектов хранения данных (DSO в терминах SAP BW) и т. д. Данный подход позволяет выполнить параллельную разработку, обеспечивающую непротиворечивость данных, но порождает очень сложную информационную среду предприятия.

Двухуровневое хранилище данных подразумевает использование принципов подхода «сверху–вниз», т. е. создается единое хранилище данных — источник агрегированных и детальных данных для построения аналитической отчетности в каждом подразделении. Первый уровень представляется источниками данных, в которые загружаются записи исходных систем с помощью ETL-процессов. На втором уровне реализуется непосредственно централизованное хранилище данных, состоящее из одного или нескольких OLAP-кубов, связанных с источниками данных через систему объектов промежуточного хранения. Таким образом, все аналитические отчеты пользователей формируются на основе единственного источника данных. Благодаря следованию принципам «сверху–вниз» в данном подходе реализуется строгое соответствие данных документированным стандартам и бизнес-правилам, однако такое представление не всегда оказывается удобным для конечных пользователей.

Трехуровневое хранилище данных также относится к разновидностям подхода «сверху–вниз», но при этом устраняется недостаток КХД, состоящего из двух уровней — появляется поддержка потребностей отдельных подразделений компании. Первые два уровня структурно повторяют двухуровневое хранилище, однако, появляется третий уровень — витрины данных, которые снабжают аналитические отчеты отделов предприятия специализированными детальными данными. При этом источником данных для витрин по-прежнему остается единое хранилище данных, что гарантирует целостность и непротиворечивость информации. В этом случае удовлетворяются потребности пользователей, но возникает избыточность в хранении данных. Следовательно, принципиальные отличия данной архитектуры от двухуровневой заключаются в том, что появляются возможности гибкой настройки аналитической отчетности подразделений, а также распределения нагрузки на систему между витринами данных, что ранее было невозможно из-за обращения всех запросов к единому источнику.

Подход SAP LSA к проектированию КХД. Многоуровневая масштабируемая архитектура SAP Layered Scalable Architecture (LSA) — это стандарт проектирования глобальных масштабируемых хранилищ данных, разработанный компанией SAP для применения в КХД [3, 9]. За основу был взят подход SAP Enterprise Data Warehouse (EDW — корпоративное хранилище данных), недостатки которого — плохая масштабируемость и низкая производительность при обработке интенсивного потока данных.

Корпоративное хранилище данных, построенное по принципам LSA, состоит из семи уровней (слоев), приведенных на рис. 5. Следует отметить, что в LSA активно используются объекты хранения Operational Data Store (ODS) — таблицы базы данных, объединяющие ключевые поля нескольких справочников и предназначенные для промежуточного хранения данных.

1. Уровень сбора данных (DAL — Data Acquisition Layer), как и в классическом хранилище данных, используется для передачи информации из системы-источника.

2. Уровень качества и гармонизации данных (QHL — Quality & Harmonization Layer) является необязательным и нужен в тех случаях, когда требуется приведение данных к определенному формату: например, если данные поступают из разнородных систем, то требуется их преобразовать к общему виду.

3. Уровень корпоративной памяти (CML — Corporate Memory Layer) требуется для резервного копирования и восстановления. Данные на этом уровне хранятся не больше двух лет.

4. Уровень распространения данных (DPL — Data Propagation Layer) используется для хранения логически разделенных записей.

5. Уровень бизнес-трансформации данных (BTL — Business Transformation Layer) служит для реализации бизнес-логики с помощью трансформаций данных между объектами хранения.

6. Уровень отчетности и витрин данных (FRL — Flexible Reporting Layer; DRL — Data mart Reporting Layer) предназначен для хранения детализированных логически разделенных данных в инфокубах или объектах хранения.

7. Уровень виртуализации (VRL — Virtualization Reporting Layer) содержит мультипровайдеры — структуры, объединяющие несколько инфокубов и объектов хранения в единый источник информации.

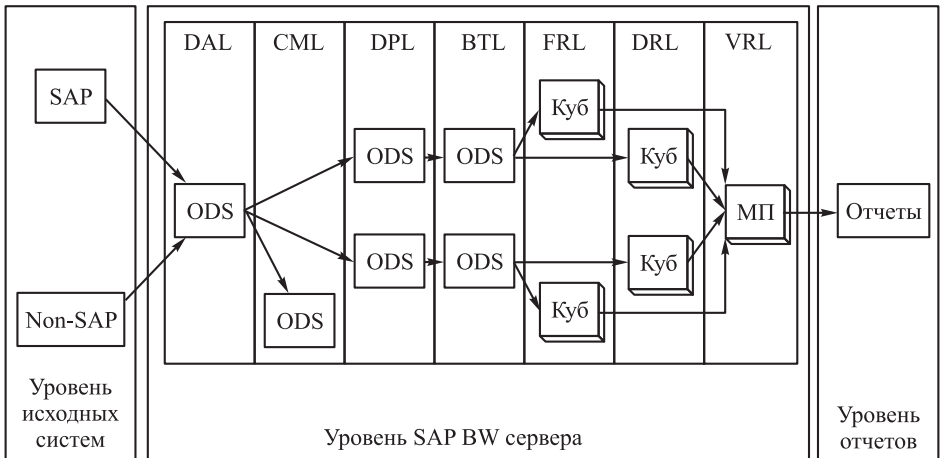


Рис. 5. Уровни архитектуры SAP LSA

Применение подобной многослойной структуры хранилища позволяет выполнить первое и основное требование к проектированию КХД — хранить данные на разных уровнях детализации. Более того, наличие нескольких последовательно связанных уровней позволяет реализовать сложные преобразования показателей — необходимое условие для внедрения КПЭ. Использование витрин данных также становится возможным в SAP LSA, так как значительно упрощается процесс их интеграции. Это связано с тем, что каждая витрина строится по одной и той же методологии, обеспечивая тем самым однородность показателей.

Добавление архивного инфокуба в концепции SAP LSA может проходить сразу на двух уровнях: FRL и DRL. Дело в том, что в некоторых случаях требуется сохранять в архив детализированные данные (FRL), представленные в десяти и более измерениях. Обосновать такое решение можно в том случае, если архивирование происходит в витрине, в которой интенсивность появления новых записей не слишком велика [1, 8]. В противном случае рекомендуется передавать в исторический куб агрегированные данные с уровня DRL.

Платформа SAP BW предусматривает создание специализированных инфокубов реального времени, способных работать в двух режимах: загрузка данных из предыдущего уровня (загрузка) и ввод новых записей из отчетов (планирование) [9, 10]. Режим загрузки, фактически, повторяет функциональность классического инфокуба, а режим планирования предназначен для ввода в инфокуб плановых показателей и расчета прогнозных показателей на основе существующих данных. В подходе LSA такие действия возможны на уровне VRL, так как все пользовательские формы ввода и механизмы расчета прогнозных КПЭ строятся на базе мультиспровайдеров [6, 11, 12].

Методика проектирования КХД с использованием подхода SAP LSA. Для успешной реализации КХД, в соответствии с принципами подхода SAP LSA, необходимо придерживаться следующих шагов, применяемых как для создания отдельных витрин данных, так и для хранилища данных уровня предприятия (рис. 6).

1. Определение перечня ключевых показателей эффективности, выработка алгоритмов их расчета и составление списка исходных систем. Данный этап предполагает подробное документирование, при этом особенно важным является проверка на соответствие разработанных алгоритмов бизнес-процессам.

2. Реализация уровня сбора данных (DAL) путем создания источников данных в виде экстракторов. Основная логика извлечения при этом формируется на стороне ERP-системы, а в системе SAP BW возникает потребность в преобразовании полей таблиц SAP R/3 в инфо-объекты.

3. Создание инфо-объектов в системе SAP BW в соответствии с принятой конвенцией технических наименований и требованиями проектного решения. Получившиеся признаки и показатели играют роль «строительного материала» для создания объектов хранения.

4. Создание уровней качества и корпоративной памяти в виде DSO с оптимизацией записи. Эти DSO должны содержать инфо-объекты, соответствующие всем полям экстракторов, что позволит хранить исходные данные в наибольшей степени детализации.

5. Создать стандартный DSO на уровне синхронизации с набором инфо-объектов, минимально требуемых бизнес-отчетностью соответствующей группы показателей. На этом этапе происходит разделение на потоки данных по наличию общих признаков в КПЭ. Например, из общего сбытового источника данных ERP-системы могут формироваться различные группы показателей: одни из них отражают объемы проданных материалов, а другие — выручку от продажи.

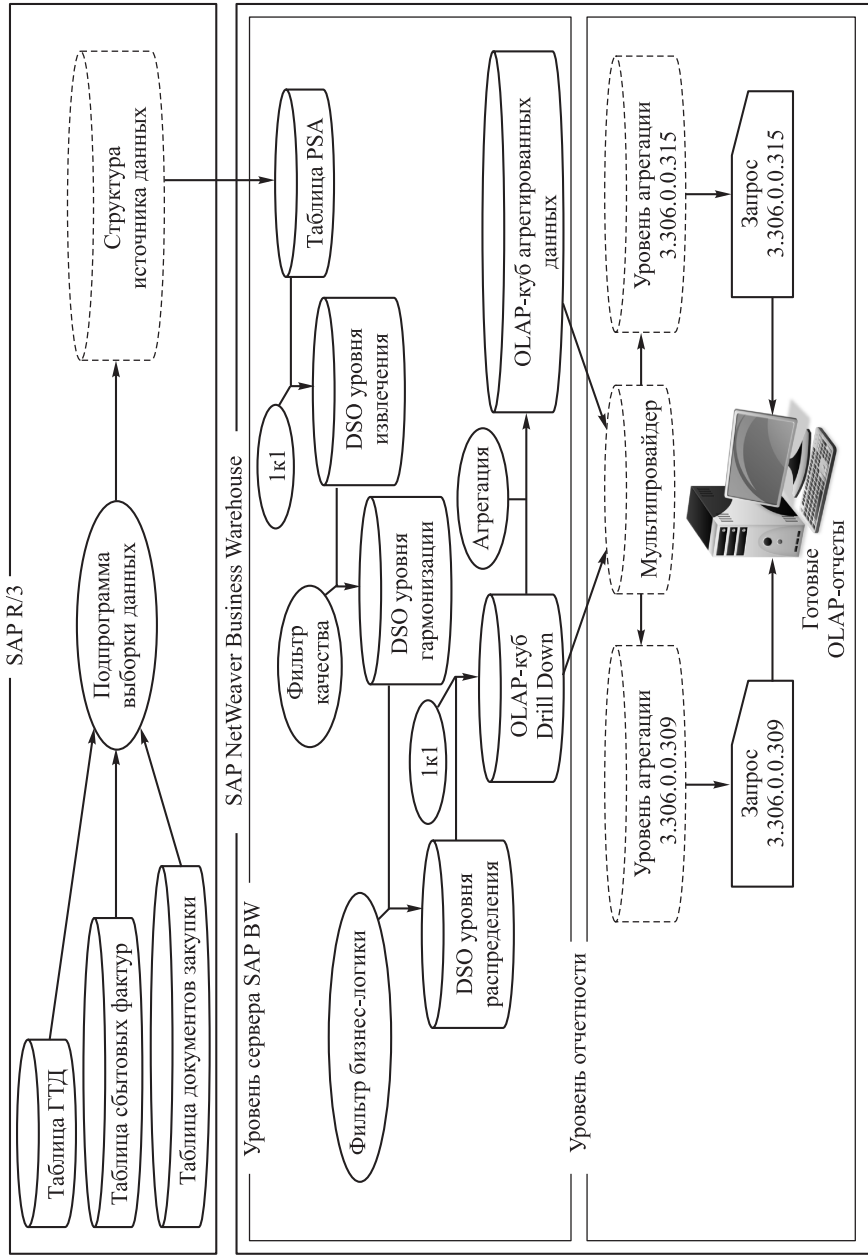


Рис. 6. Архитектура КХД

6. Для потока из одного–трех показателей рекомендуется проводить расчет КПЭ в трансформации перед стандартным DSO уровня распределения. При наличии большого числа показателей на уровне бизнес-трансформации требуется разместить стандартные DSO (по одному на каждую группу показателей). В трансформации перед этими DSO следует разместить необходимые расчеты, связанные с формированием КПЭ — это требуется в том случае, когда программы экстрактора недостаточно для реализации алгоритма подсчета показателя.

7. Реализация уровня отчетности с двумя подуровнями: детальные и агрегированные данные. Для детальных данных создаются инфокубы, причем их число не обязательно должно соответствовать группам показателей, определенным ранее. Рекомендуется создать группы показателей более высокого уровня (к примеру, по отношению показателей к тому или иному блоку ERP-системы). В таком случае сокращается число инфокубов с минимальной избыточностью в их измерениях. За агрегированные данные отвечает один инфокуб, соединенный трансформациями со всеми инфокубами детальных данных. В нем остаются только те инфо-объекты, которые необходимы для основной отчетности предприятия.

8. Создать мультипровайдеры (виртуальные структуры) на уровне виртуализации по одному на каждую укрупненную группу показателей (см. п. 7), в каждом из которых были бы включены инфокубы агрегированных и детальных данных. Это необходимо для обеспечения перехода в отчеты детальных данных (drill-down). Допускается совмещение уровня отчетности и виртуализации.

Примером использования данной методики является реализация корпоративного хранилища данных отдела таможенного оформления подразделения нефтегазовой компании. В соответствии с предложенным подходом были проведены следующие работы.

1. Сформированы паспорта ключевых показателей эффективности в соответствии с форматом бизнес-объектов предприятия: «Таможенные пошлины за реализованный газ» и «Таможенные пошлины за реализованные продукты переработки». Создано проектное решение с подробным описанием алгоритмов извлечения и расчета данных показателей.

2. В соответствии с требованиями проектного решения в OLTP-системе SAP R/3 создан экстрактор «Таможенные декларации», объединяющий поля из трех таблиц базы данных: ГТД, сбытовые фактуры и документы закупки. Для извлечения записей из таблиц создана программа на языке ABAP/4.

3. С использованием конвенции наименований, описанной в проектном решении, созданы признаки (справочники) и показатели в системе SAP BW.

4. Создан DSO с оптимизацией записи в качестве уровня извлечения, назначением которого является хранение записей исходной системы в формате хранилища.

5. Уровень синхронизации представлен стандартным DSO с набором признаков и показателей исходной системы, необходимых для расчета КПЭ. Фильтр качества создан для выделения только тех данных, которые оказывают прямое влияние на формирование ключевых показателей.

6. На основе записей, прошедших через фильтр бизнес-логики, выполняется расчет и КПЭ размещается в стандартном DSO уровня распределения.

7. Уровень отчетности совмещен с уровнем виртуализации. OLAP-куб Drill Down содержит детальные данные по КПЭ, т. е. имеет расширенный набор измерений по сравнению с OLAP-кубом агрегированных данных.

8. Создан мультипровайдер, построенный на обоих OLAP-кубах. Для формирования запросов предназначены уровни агрегации, каждый содержит только те признаки и показатели, которые необходимы для конкретных отчетов.

В результате было создано КХД, архитектура которого, приведенная на рис. 6, полностью соответствует принципам концепции LSA.

В качестве примера приведена даталогическая модель OLAP-куба Drill Down на рис. 7.

Созданное КХД может выступать как в роли обособленного хранилища, так и войти в состав более крупного ХД в качестве витрины данных — это возможно за счет унификации потоков данных — независимо от характера источника (таможенные декларации, сбытовые фактуры, финансовые документы и любые другие), назначение и формат уровней хранилища остаются одинаковыми. Следовательно, значительно упрощается процесс внедрения новых ключевых показателей (в рамках существующих потоков) и новых витрин данных. Более того, становится возможным хранение КПЭ из всех потоков в едином OLAP-кубе агрегированных данных.

Таким образом, приведенная методика проектирования КХД может быть успешно и практически реализована в среде SAP BW, при этом выполнение её этапов гарантирует соответствие разрабатываемой архитектуры общепринятой практике SAP LSA.

Выводы. Многие крупные предприятия, использующие на протяжении нескольких лет ERP-системы SAP R/3, активно внедряют КХД на базе SAP BW для эффективной поддержки принятия стратегических решений при работе с большими массивами накапливаемых данных. Тем не менее, не всегда уделяется должное внимание проектированию архитектуры КХД — часто отдается предпочтение классической структуре хранилища данных, которую проще реализовать и быстрее внедрить.

Предложен подход, позволяющий создавать КХД любых масштаба и сложности, при этом сохраняя прозрачность структуры и возможности интеграции с другими системами.

В настоящее время успешное использование методологии SAP LSA для проектирования КХД доказывает, насколько мощным средством является данный подход. Однако каждый отдельный проект КХД не обходится без составления индивидуальной архитектуры на основе LSA, отвечающей требованиям бизнес-процессов конкретного предприятия.

Предложенная методика имеет ряд недостатков. Во-первых, появляется избыточность в хранении данных из-за наличия одних и тех же КПЭ, представленных в OLAP-кубах разной детализации. Во-вторых, при наличии нескольких

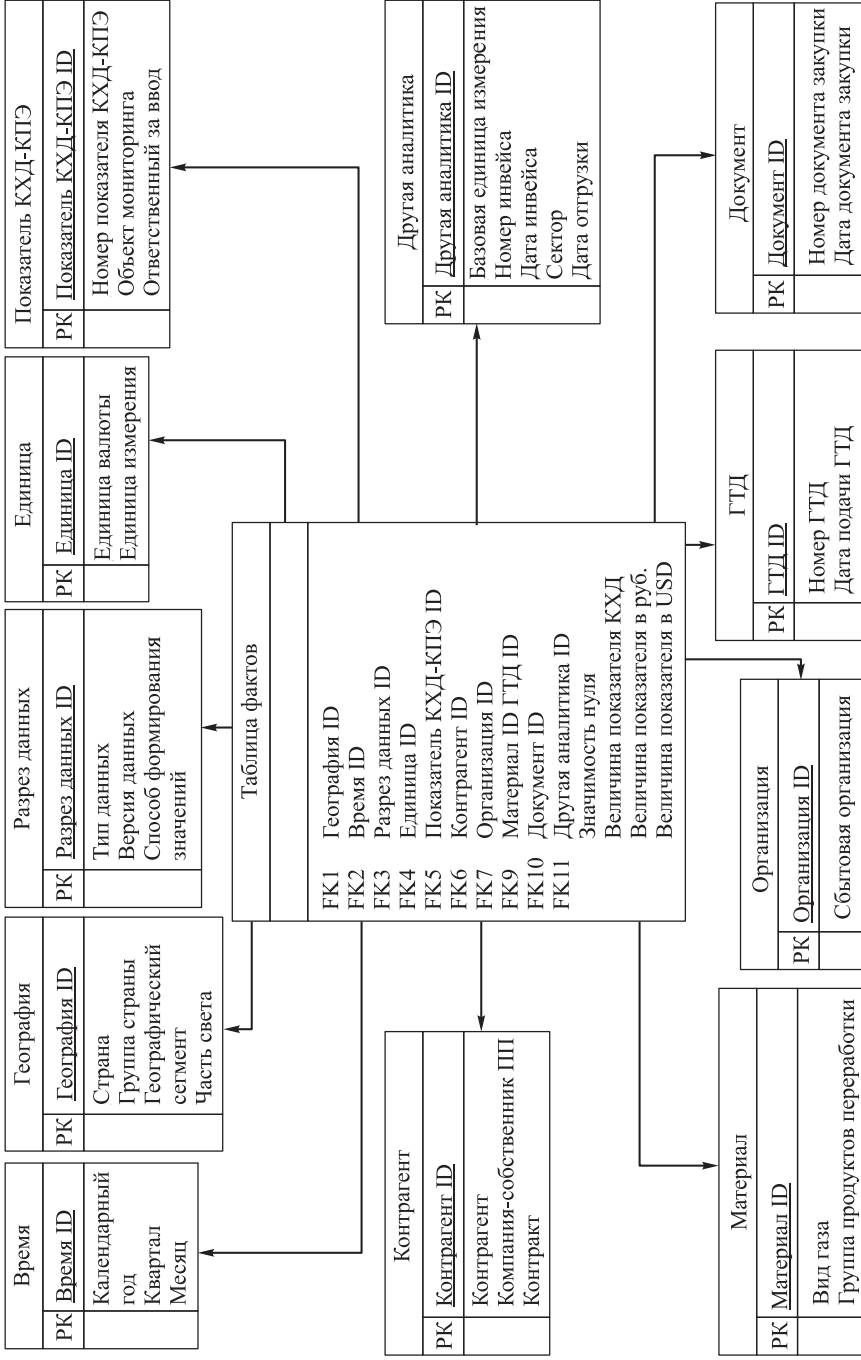


Рис. 7. Даталогическая модель OLAP-куба Drill Down

источников данных необходимо синхронизировать отдельные потоки на уровне отчетности. Перечисленные недостатки свойственны всем КХД и обусловлены структурой агрегации данных на разных уровнях иерархии для обеспечения оперативного анализа данных, формировать аналитические отчеты.

ЛИТЕРАТУРА

1. Балдин А.В., Тоноян С.А., Елисеев Д.В. Анализ избыточности хранения темпоральных данных средствами реляционных СУБД // Инженерный журнал: наука и инновации. 2014. № 4. DOI: 10.18698/2308-6033-2014-4-1273 URL: <http://engjournal.ru/catalog/it/hidden/1273.html>
2. Тоноян С.А., Сараев Д.В. Темпоральные модели базы данных и их свойства // Инженерный журнал: наука и инновации. 2014. № 12. DOI: 10.18698/2308-6033-2014-12-1333 URL: <http://engjournal.ru/catalog/it/hidden/1333.html>
3. Хаунт Ж. Многоуровневая масштабируемая архитектура (LSA). SAP NetWeaver, RIG VI EMEA, 2009. 43 с.
4. Технологии моделирования данных для хранилищ данных / Ч. Баллард, Д. Хэрреман, Д. Шау, Р. Белл, Е. Ким, А. Валенчик. IBM Redbooks, 1998. 216 с.
5. Балдин А.В., Тоноян С.А., Елисеев Д.В. Язык запросов к миварному представлению реляционных баз данных, содержащих архив информации из предыдущих кадровых систем // Инженерный журнал: наука и инновации. 2013. № 11. URL: <http://engjournal.ru/catalog/it/hidden/1053.html>
6. Пател Б., Палекар А., Ширалкар Ш. Практический курс SAP Net Weaver Business Warehouse (BW) 7. Бонн: Galileo Press, 2008. 689 с.
7. Паклин Н.Б., Орешков В.И. Бизнес аналитика: от данных к знаниям. СПб.: Питер, 2013. 704 с.
8. Барсегян А.А. Анализ данных и процессов. СПб.: БХВ–Петербург, 2009. 512 с.
9. Черненький В.М. Толочко С.И. Анализ информационных систем и определение понятий информационная система поддержки оперативных решений // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2011. Спецвып. № 5. С. 69–80.
10. Виноградова М.В., Игушев Э.Г. Принципы организации структуры данных с произвольным доступом и быстрой операции вставки // Инженерный журнал: наука и инновации. 2012. Вып. 3. DOI: 10.18698/2308-6033-2012-3-101 Available at: <http://engjournal.ru/eng/catalog/it/asu/101.html>
11. Сарка Д., Лах М., Йеркич Г. Microsoft SQL — Server 2012. Реализация хранилищ данных. Учебный курс Microsoft / пер. с англ. М.: Русская редакция, 2014. 816 с.
12. Спирли Э. Корпоративные хранилища данных. Планирование, разработка и реализация. М.: ИД «Вильямс», 2010. 400 с.

Тоноян Славик Анушаванович — канд. техн. наук, доцент кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Высочанский Владимир Александрович — студент кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Пробьба ссылаться на эту статью следующим образом:

Тоноян С.А., Высочанский В.А. Методика проектирования корпоративного хранилища данных на базе платформы SAP NetWeaver Business Warehouse // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2016. № 4. С. 33–48.

DOI: 10.18698/0236-3933-2016-4-33-48

ENTERPRISE DATA WAREHOUSE DESIGN METHOD USING SAP NET WEAVER BUSINESS WAREHOUSE

S.A. Tonoyan

tonoyansl@mail.ru

V.A. Vysochanskiy

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

Data warehouse is an informational system optimized to store large amounts of data and create analytical reports for decision support. Principles of classic data warehouse architecture cannot be applied for enterprise data warehouses (EDW) based on key performance indicators (KPI). In this article we offer a methodology that takes into account the KPI, we review and justify the enterprise data warehouse design technique using SAP NetWeaver Business Warehouse. Our step by step guide to data warehouse design features SAP LSA (Layered Scalable Architecture) standard. It consists of 7 layers, each serving the particular purpose of EDW functionality. LSA is flexible; it can be easily modified to fit various kinds of business processes. We provide a practical example of an EDW design for customs department of an oil and gas company

Keyword

Data warehouse, SAP, multidimensional data model, star-schema, OLAP, OLTP, info-cube, aggregation, LSA

REFERENCES

[1] Baldin A.V., Tonoyan S.A., Eliseev D.V. Analysis of temporal data storage redundancy by means of RDBMS. *Jelektr. nauchno-tekh. izd. "Inzhenernyy zhurnal: nauka i innovacii"* [El. Sc.-Tech. Publ. "Eng. J.: Science and Innovation"], 2014, iss. 4.

DOI: 10.18698/2308-6033-2014-4-1273 Available at: <http://engjournal.ru/eng/catalog/it/hidden/1273.html>

[2] Tonoyan S.A., Saraev D.V. Temporal database models and their properties. *Jelektr. nauchno-tekh. izd. "Inzhenernyy zhurnal: nauka i innovacii"* [El. Sc.-Tech. Publ. "Eng. J.: Science and Innovation"], 2014, iss. 12. DOI: 10.18698/2308-6033-2014-12-1333

Available at: <http://engjournal.ru/eng/catalog/it/hidden/1333.html>

[3] Haupt Juergen. LSA. SAP NetWeaver, RIG BI EMEA, 2009. 43 p.

[4] Ballard C., Herreman D., Schau D., Bell R., Kim E., Valencic A. Data modeling techniques for data warehousing. International Technical Support Organization, IBM, 1998. 216 p.

[5] Baldin A.V., Tonoyan S.A., Eliseev D.V. Query language to mivar representation of relational databases, containing information archive from previous human resources systems. *Jelektr. nauchno-tekh. izd. "Inzhenernyy zhurnal: nauka i innovacii"* [El. Sc.-Tech. Publ. "Eng.

J.: Science and Innovation”], 2013, iss. 11. DOI: 10.18698/2308-6033-2013-11-1053 Available at: <http://engjournal.ru/eng/catalog/it/hidden/1053.html>

[6] Patel B., Palekar A., Shiralkar Sh. SAP Net Weaver Business Warehouse (BW) 7. Bonn: Galileo Press, 2008. 689 p.

[7] Paklin N.B., Oreshkov V.I. Biznes analitika: ot dannykh k znaniyam [Business intelligence: From data to knowledge]. St. Petersburg, Piter Publ., 2013. 704 p.

[8] Barsegyan A.A. Analiz dannykh i protsessov [Analysis of data and processes]. St. Petersburg, BKhV–Peterburg Publ., 2009. 512 p.

[9] Chernen'kiy V.M., Tolochko S.I. Information system analysis and the definition of a notion of information system of prompt decision support. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Priborostr., Spetsvyp no. 5* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng., Spec. Issue no. 5], 2011, pp. 69–80 (in Russ.).

[10] Vinogradova M.V., Igushev E.G. Principles of data structure organization with random access and fast insert and delete operation. *Jelektr. nauchno-tekh. izd. “Inzhenernyy zhurnal: nauka i innovacii”* [El. Sc.-Tech. Publ. “Eng. J.: Science and Innovation”], 2012, iss. 3.

DOI: 10.18698/2308-6033-2012-3-101

Available at: <http://engjournal.ru/eng/catalog/it/asu/101.html>

[11] Sarka D., Lah M., Jerkic G. Exam 70-463: Implementing a data warehouse with Microsoft SQL server 2012 Microsoft. Microsoft Press, 2013.

[12] Sperley E. Planning, building, and implementation (Enterprise data warehouse). Prentice Hall, 1999.

Tonoyan S.A. — Cand. Sci. (Eng.), Professor of Information Processing and Control Systems Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Vysochanskiy V.A. — student, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Tonoyan S.A., Vysochanskiy V.A. Enterprise Data Warehouse Design Method using Sap Net Weaver Business Warehouse. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2016, no. 4, pp. 33–48. DOI: 10.18698/0236-3933-2016-4-33-48