

На правах рукописи

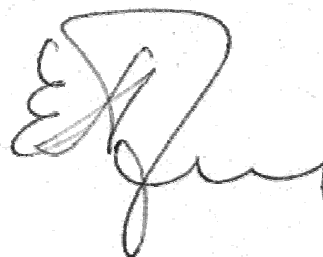
Гришенков Евгений Анатольевич

**ПОСТРОЕНИЕ КОНЦЕПТУАЛЬНОЙ МОДЕЛИ БАЗ ДАННЫХ
ПРИ ПОМОЩИ ШАБЛОНА МОДЕЛИ ЕДИНОГО
КОРПОРАТИВНОГО ПРОСТРАНСТВА ДАННЫХ**

Специальность: 05.13.01. – «Системный анализ, управление и обработка информации (в науке и промышленности) по техническим наукам»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук



Нижний Новгород,
2010

Работа выполнена на кафедре «Вычислительные системы и технологии»
Института радиоэлектроники и информационных технологий
Государственного образовательного учреждения высшего
профессионального образования
Нижегородский государственный технический университет
им. Р.Е. Алексеева

Научный руководитель: - доктор технических наук, доцент
Мисевич Павел Валерьевич

Официальные оппоненты: - доктор технических наук, доцент
Хранилов Валерий Павлович
- кандидат физико-математических
наук,
Шапошников Дмитрий Евгеньевич

Ведущая организация: Вычислительный центр
им. А.А. Дородницына
Российской академии наук

Защита состоится 23 декабря 2010 г. в 11 часов на заседании
диссертационного совета Д.212.165.05 в Нижегородском государственном
техническом университете им. Р.Е. Алексеева по адресу: 603950, г.
Нижний Новгород, ГСП, ул. К. Минина 24

С диссертацией можно ознакомиться в библиотеке НГТУ им. Р.Е.
Алексеева.

Отзыв на автореферат в двух экземплярах, заверенный печатью
организации, просим направлять по вышеуказанному адресу на имя
ученого секретаря диссертационного совета.

Автореферат разослан « ____ » ноября 2010 г.

Ученый секретарь
диссертационного совета



А.С.Суркова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Совершенствование методов проектирования интегрированных корпоративных баз данных (БД) актуально сегодня по причине качественного повторения проблемной ситуации 60-70 годов с разрозненностью данных применительно к разнообразным СУБД - источникам интегрирования данных по технологии Data Warehouse процессами ETL (извлечение, преобразование, загрузка).

В 1975 году трехуровневая модель абстракции данных стандарта ANSI/SPARC определила путь преодоления недостатков файловых систем и интеграции данных коллективного использования. Явное выделение концептуального уровня абстракции данных придало данным смысл целостного корпоративного ресурса, отделенного от программ их обработки и способов физического хранения. Он объединяет разнообразия представлений о данных пользователей и прикладных программистов (внешний уровень), а также решений внутреннего (физического) уровня.

Однако, несмотря на достижения моделирования сущность-связь (ER) и теории нормализации баз данных, за 20 последующих лет никому не удалось построить единую нормализованную корпоративную базу данных, хотя цель такая ставилась, как отмечал идеолог Microsoft Дэвид Васкевич.

На практике для решения проблемы интеграции данных было развито два технологических подхода: киоски данных (Data Marts или DM) и хранилище данных (Data Warehouse или DW), их сравнение широко обсуждалось. DW основано на идее проектирования корпоративной аналитической БД, физическом разделении учетных и аналитических систем. Обработка данных также разделена на транзакционную (OLTP) и аналитическую (OLAP) части, различающиеся: объемами транзакций, оптимизацией на ввод или чтение, типом пользователей и др. Заполнение структуры DW основано на обширной ETL обработке, массивованных пересылках данных предельно большого объема.

При DM подходе выборочные подключения к СУБД источникам позволяют «на лету» формировать слой предметно ориентированных данных (Universe). Подход DM легок в реализации, но с ростом числа киосков данных быстро нарастает множество связей между киосками и источниками вместе с разногласиями в пользовательских представлениях.

Подходы DM и DW характеризуются как нисходящий и восходящий способы проектирования. Однако результаты их не стыкуются - вместо единого DW удается строить только разрозненные «склады» данных Detail Data Store (DDS) для каждой предметной области с ограниченным кругом задач. Модель данных DDS может содержать сотни таблиц и тысячи показателей, объединение моделей очень сложная задача.

Такое развитие свидетельствует о качественном воспроизводстве сложностей и недостатков времен файловых систем, обусловленным теоретическим вакуумом в сфере концептуального моделирования данных.

Эдсгер Дейкстра также считал главной причиной несовершенства существующих информационных систем - невозможность преодоления сложности описания «картины мира» предметной области. Он поставил задачу, ставшую побудительным мотивом настоящей работы: «научиться отличать необходимую сложность системы от нагромождений, возникающих в силу случайных причин, определить какую часть архитектуры всей системы нужно представлять в интерфейсах».

Данная работа посвящена проблеме концептуального моделирования баз данных посредством шаблона модели единого корпоративного пространства данных (ЕКПД). Шаблон основан на представлении фактов хозяйственной деятельности в координатах многомерного пространства предметно-ориентированных измерений с универсальным интерфейсом сводного отчета для ввода-вывода. В основу положен практический опыт разработки ряда корпоративных аналитических систем. Концептуальную схему диссертационной работы представляет Рисунок 1.

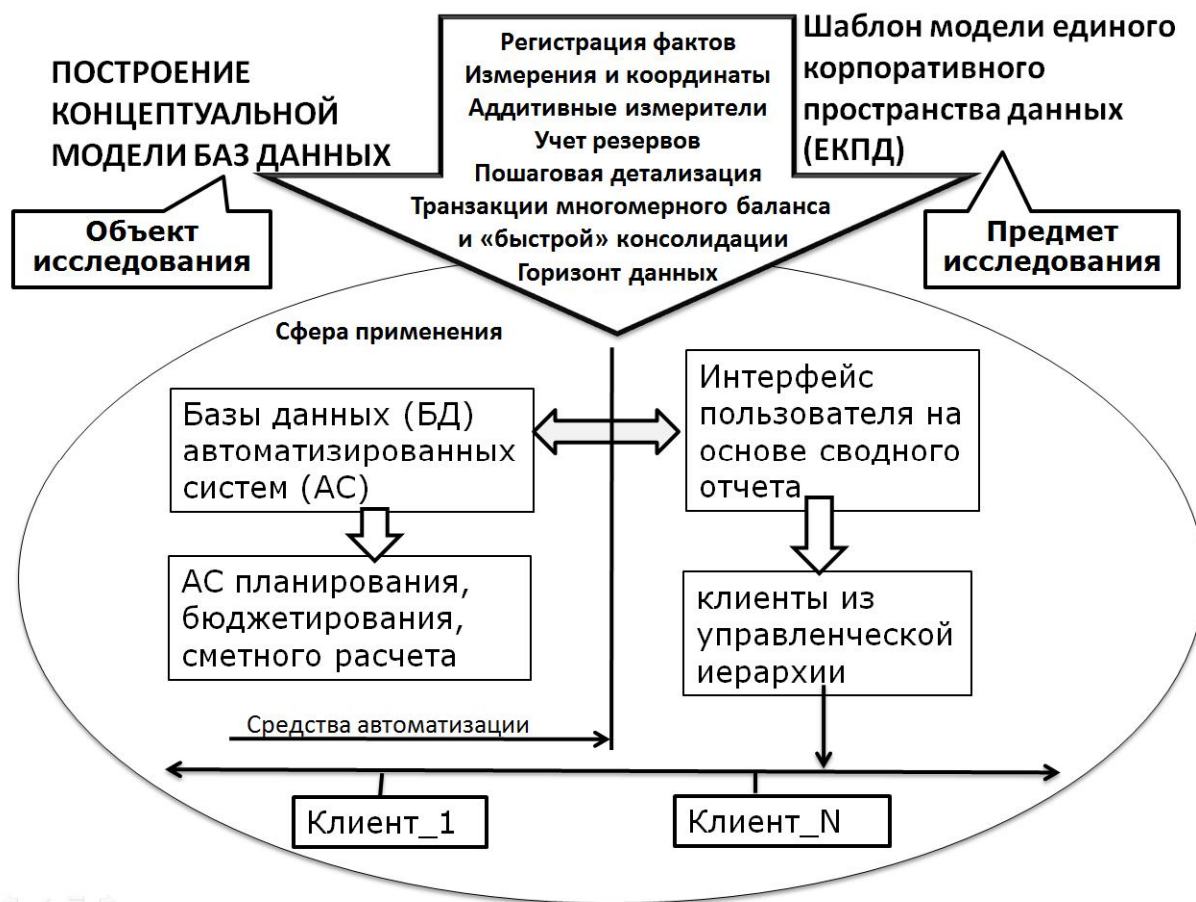


Рисунок 1 – Концептуальная схема диссертационной работы

Цель работы заключается в разработке концептуального шаблона построения баз данных, описывающих потоки корпоративных ресурсов, подчиняющиеся действию законов сохранения ресурса, на основе многомерной модели единого корпоративного пространства данных, а

также разработка типовых операций с данными посредством интерфейса сводного отчета.

Для достижения сформулированной цели решены следующие **задачи**:

1. Исследование методов концептуального проектирования БД, выявление причин, препятствующих построению неограниченно больших БД.
2. Разработка концептуального шаблона модели корпоративной БД на основе ЕКПД с ограничениями, отражающими специфику корпоративных данных: аддитивные показатели, «горизонт» данных, закон сохранения ресурса.
3. Определение места интерфейса в общей архитектуре системы как проекционной гиперплоскости шаблона модели ЕКПД, разработка многомерной модели перераспределения аддитивного ресурса на основе баланса с учетом членов с остатками (невязки или резервы), разработка основных операций с БД.
4. Проверка применения шаблона ЕКПД и основных операций на тестовом макете и на реальной задаче сметного планирования.

Объектом исследования стали проблемные и системные методы АС, методы концептуального проектирования корпоративной БД, трехуровневая модель абстракций данных стандарта ANSI/SPARC.

Предметом исследования является шаблон многомерной модели единого корпоративного пространства данных с ограничениями, отражающими специфику корпоративной БД: аддитивные показатели, «горизонт» данных, закон сохранения ресурса.

При выполнении работы использовались следующие **методы проведения исследований**: многомерный анализ данных, линейная алгебра и геометрия, теория графов, теория множеств, методы и модели проектирования систем банков и баз данных, моделирование сущность-связь, теория нормализации БД, шаблоны проектирования.

Достоверность и обоснованность полученных в работе результатов обеспечивается преемственностью с апробированными теоретико-практическими подходами к вопросам построения АС, теоретические результаты прошли практическую проверку на макете системы, на решении реальной актуальной задачи сметного планирования.

Научная новизна представленной работы заключается:

1. Рассмотрение частных моделей баз данных (представлений пользователей) как типовых элементов шаблона модели ЕКПД, подчиняющихся структурным ограничениям шаблона. Такой подход отличается от рассмотрения частных моделей как эмпирической данности с произвольным графическим представлением. Предложенная

стандартизация позволяет реализовать функционирование и поддержку неограниченно больших корпоративных моделей на основе массового самообслуживания.

2. Использование шаблона модели ЕКПД в форме инструментария построения и формализации концептуальной модели баз данных в предметной области проектирования БД высокого уровня сложности. Подход отличается от известных сведениями части процессов проектирования и формализации концептуальной модели БД к описанию объекта исследования при помощи типовых информационных структур. Это позволяет создать программные инструментарию проектирования концептуальной модели, которые обеспечивают построение стандартных подсхем данных и их объединение в схему.
3. Предложен способ использования интерфейса сводного (OLAP) отчета не только для просмотра и навигации по данным, а для редактирования изменений и области определения БД. Для обоснования способа была сформулирована многомерная модель баланса аддитивных данных с учетом остатков (резервов) и исследованы её инвариантные свойства. Это позволяет использовать транзакции многомерного баланса и быстрой консолидации для редактирования детальных и агрегатных значений данных, а также измерений и их значений с использованием зарекомендовавшего себя инструментария сводного отчета OLAP.

Практическая ценность и внедрение результатов заключается в снятии ограничений на формирование и поддержку, в перспективе, неограниченно больших корпоративных моделей на основе массового самообслуживания. Результаты работы использовались при реализации проектов по заказу службы экономики и развития Куйбышевской ж.д. в 1999-2002 гг., при реализации Адресной программы формирования специализированного жилого фонда ОАО «РЖД» в 2006-2008 и 2009-2010 гг., при решении задач реализации жилищной политики ОАО «РЖД» в 2005-2010 гг.

Область применения результатов не ограничивается железнодорожной отраслью. Изложенные в работе методы и средства позволяют обеспечить совместное оперативное взаимодействие сотрудников службы экономики, например, большой территориально распределенной корпорации.

Апробация работы и публикации. Результаты работы были представлены на международных отраслевых конференциях в Санкт-Петербурге: “Инфотранс-98”, “Инфотранс-99”, “Инфотранс-2000”, а также за рубежом: SAS User Group Conference (SUGI26), США, 2001 г. По результатам диссертации опубликовано 8 печатных работ, из них 2 в журналах списка ВАК.

Структура и объем работы. Текст диссертации состоит из введения, четырех глав, заключения и списка литературы, включающего 100 наименований. Диссертация содержит 98 страниц машинописного текста, 31 иллюстрацию, 1 таблицу.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность работы, определяется цель и задачи исследования, формулируется научная новизна, приводятся сведения о структуре и краткое описание основных положений диссертационной работы. Она ориентирована на разработку возможностей технологической ниши, создаваемой опережающим развитием внешней памяти. Это позволяет широко использовать внешнюю память для хранения агрегатных значений, что кардинально меняет ландшафт оперативной обработки (Рисунок 2), нивелируя различия между видами обработки OLAP и OLTP, за счет использования учетных транзакций для корректировки агрегатов.

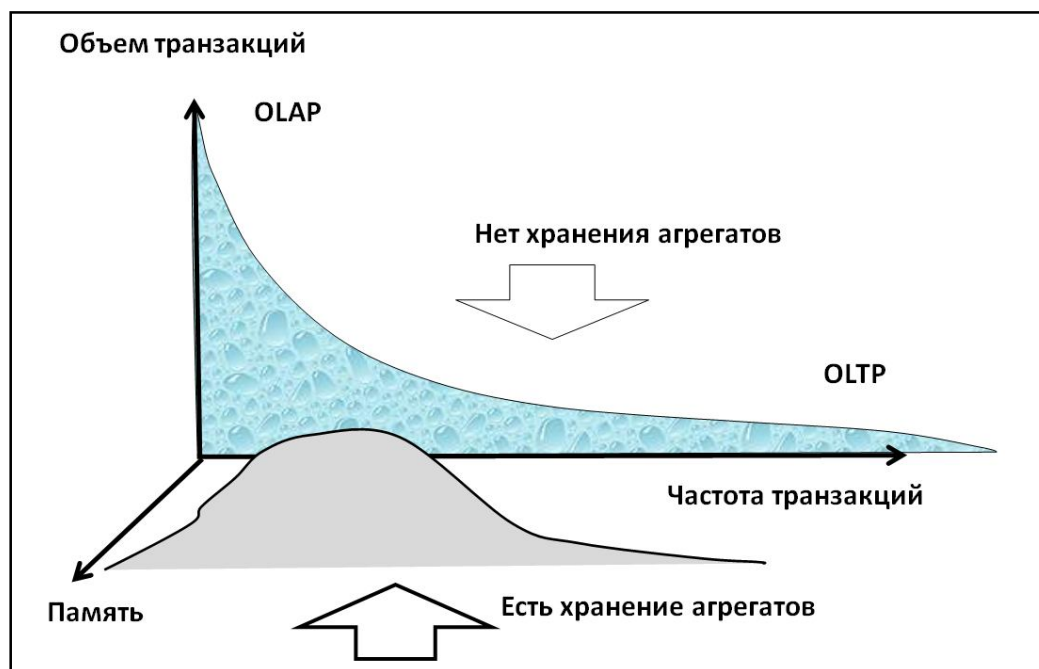


Рисунок 2 – Выравнивание ландшафта оперативной обработки включением в рассмотрение внешнюю память

В первой главе приводится обзор современных технологий концептуального проектирования баз данных, дается историческая справка о развитии технологий баз данных в качестве основы создания современных информационно-аналитических систем. Приводится описание технологий хранилищ данных (Data Warehouse) и киосков данных (Data Mart) как практических подходов на пути концептуального проектирования единой корпоративной базы данных. Рассматриваются тенденции и перспективы развития отрасли OLAP продуктов и возможности их использования на отечественном рынке информационных

систем. Описываются проблемы получения неограниченно большой корпоративной базы данных на основе объединения моделей смежных предметных областей, реализованных по технологии хранилищ данных моделями «складов» данных Detail Data Store (DDS) на примере отраслевых решений Системы SAS.

Вторая глава посвящена теоретическому обоснованию предлагаемого метода. Формулируется суть проблемы концептуального моделирования корпоративных данных. Она заключается в объединении всего спектра пользовательских перспектив на данные, абстрагируясь от уровня физической реализации, например, экономии памяти, быстродействия алгоритмов, дублирования данных. Качество модели заключается в её способности адекватно отражать «картину мира» на каждом уровне детализации, достигаемой добавлением измерений. Мощность множества возможных представлений данных, допускаемых моделью, всегда должна превышать мощность множества пользовательских представлений. Для всего этого модель должна обладать инвариантными свойствами, независимыми от числа измерений, быть проста в понимании и обладать интуитивно понятным интерфейсом.

Для решения проблемы используем многомерную модель данных (ММД) общего вида, дополненную уровнями резерва на измерениях, и интерфейсом из гиперплоскости с ограничением на число одновременно используемых измерений. Исходим из того, что корпоративные данные описывают материальные и финансовые потоки в системе координат, обеспечивающей представления данных с разной степенью детализации. Потоки подчиняются законам сохранения, а пространство обеспечивает многомерную сетку, в узлах которой регистрируются значения потоков.

Основой ММД служат записи фактов из значения измерителя и его координат по измерениям. Измеритель – числовой показатель. Считаем его аддитивным, а измерения – оси со шкалой наименований, среди которых есть точка 0 и Резерв. Координаты отвечают на основные вопросы факта регистрации хозяйственной деятельности - Что? Кто? Где? Когда? Как? и т.п., а измеритель - Сколько?

Декартово произведение осей измерений, пересекающихся в точках 0, образуют многомерное координатное пространство. Измерение - аналог домена реляционной модели данных (РМД), а измеритель - атрибут, присутствующий во всех сущностях вычислительно согласованными значениями, отражая закон сохранения описываемого ресурса.

Область определения данных задается в ММД бинарной функцией в пространстве, вместе со всеми проекциями в подпространствах, это геометрическая фигура, инвариант перестановок измерений и их значений.

Аналогом сущности в ММД выступает подпространство, как комбинация ключевых атрибутов кортежа РМД. Для N-мерного пространства максимальное число подпространств составляет $2^N - 1$, а вместе пространством максимальное число порождаемых сущностей РМД достигает 2^N .

Концептуально ММД представляется ER-моделью в виде бинарного гиперкуба, ребра его выражают отношение один ко многим и ориентированы от начала координат в глубину пространства к детальной таблице фактов. Каждой вершине гиперкуба соответствует агрегатное подпространство, представленное таблицей его непустых точек. Наборы параллельных ребер отражают операцию свертки-развертки пространства по измерениям (на основе суммирования).

Проследим изменения и усложнения ER-модели при добавлении измерений, начиная с нулевого, одного, двух, трех измерений и т.д., пока абстрагируясь от измерения Время (Рисунок 3, а, б, в, г).

Модель нулевого числа измерений содержит одно число - суммарное значение измерителя по корпорации и нулевые значения координат по всем измерениям. Модель одного измерения, представляет раскрытие общей суммы по значениям измерения, например, по филиалам корпорации, тип сущности меняется с 00 на 10, представляя элементарную операцию по свёртке-развёртке значений. Модель содержит две сущности, соединенные отношением один ко многим, то есть одномерный бинарный гиперкуб.

При добавлении каждого нового измерения модель усложняется удвоением гиперкуба и разнесением удвоенных сущностей по новому измерению с добавлением новых связей-ребер. Для двумерной модели, (Рисунок 3-г, и Таблица 1), имеем ромб с нуль сущностью наверху, два верхних ребра представляют развертку по двум измерениям, нижняя сущность представляет детальную таблицу в пространстве двух измерений, например, численность персонала по подразделениям. Другие сущности представляют итоговые строку и столбец, а также общую сумму.

Ромб получается объединением двух иерархических моделей (Рисунок 3-в), нижние сущности которых содержат ключевые атрибуты Филиал и Служба. Эти модели образуют левую и правую ветви единой модели, склеенную по общим сущностям: Корпорация и Подразделение. Аналогично склеиваются иерархические модели для трех и четырех измерений (Рисунок 4), формируя схему бинарного гиперкуба. Модель N-мерного пространства составлена из $N!$ иерархических моделей, путей по ребрам гиперкуба из начала координат в детальное пространство (вершину на главной диагонали).

При увеличении числа измерений виртуальный объем формируемого пространства начинает заведомо превышать объем области определения БД, отражающей реальность. Это будет проявляться в изменении ориентации ребер гиперкуба модели. Например, добавление измерения, функционально связанного с уже имеющимся измерением, приведет к ребрам с обратной ориентацией или ребрам типа 1:1. Общим признаком такой ситуации является отсутствие новой информации, прекращение роста числа точек области определения, при добавлении новых измерений.

Для РМД такое измерение превращается в атрибут сущности, что существенно упрощает модель (но не работу с данными). Однако для

ММД можно оставлять не нормализованное представление, это приводит к дублированию данных, что допустимо при автоматическом поддержании.

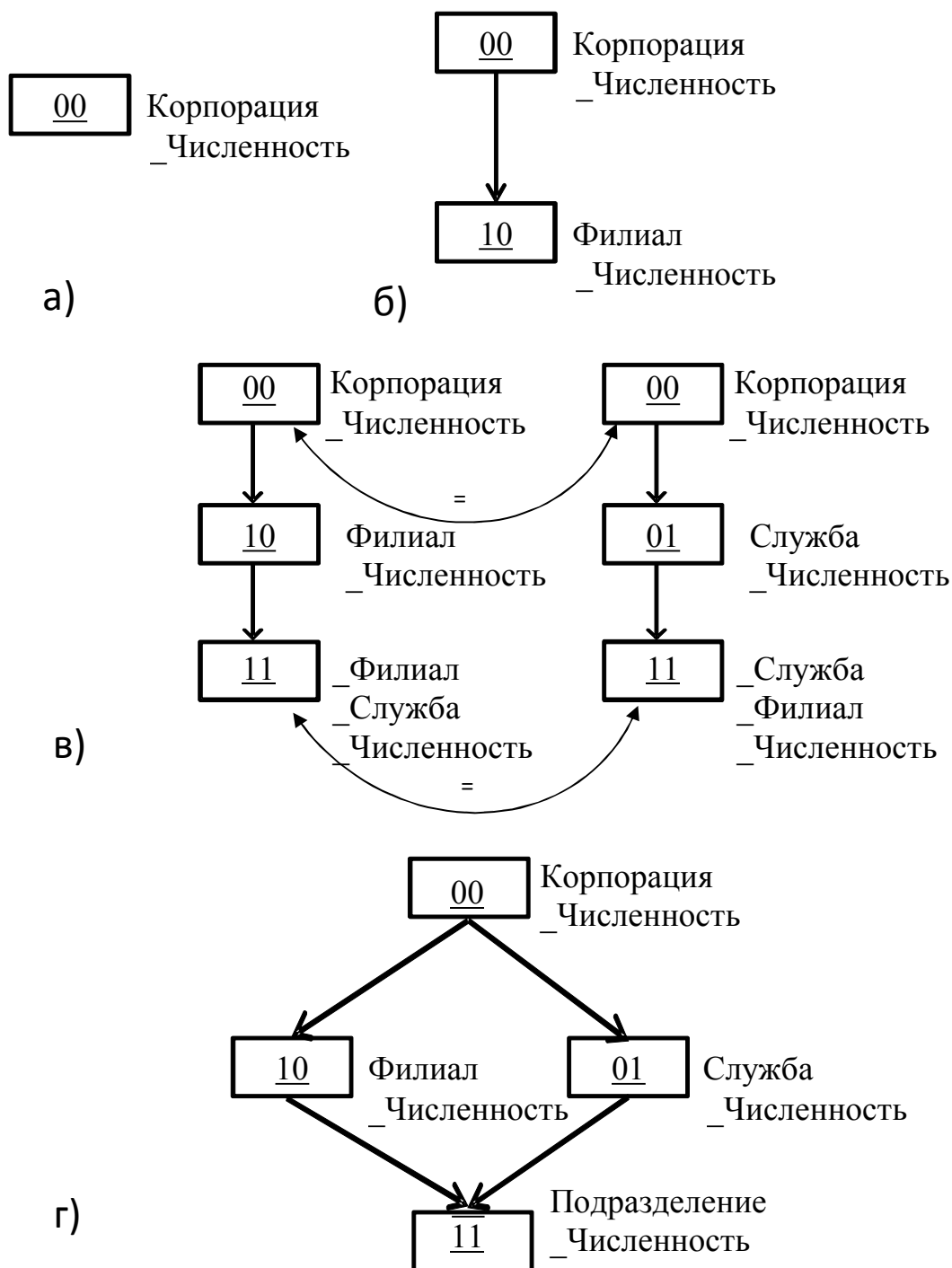


Рисунок 3 - Концептуальная модель для: а) нуля, б) одного, в) двух иерархических моделей, частей модели двух измерений (г) с измерителем Численность персонала

Другой причиной прекращения этого роста служит достижение предела дискретности данных или точности счета. В любом случае, при достаточно большом числе измерений, имеется предел роста числа точек области определения, назовем его «горизонтом» данных. Использование

этого явления позволяет существенно упростить корпоративную модель и работу с её данными. Вводим ограничение на число измерений в интерфейсе, вне пределов которого, до самого «дна», модель основана на связях типа 1:1, не требующих обработки и логически реализуемых только резервными точками.

Инвариантные свойства и редактирование агрегатов. Если в координатах факта нет нулей, то это детальная точка пространства, если есть нули, то агрегатная точка, если есть Резерв, то это резервная точка, она может быть агрегатной или детальной.

Множество резервных детальных точек взаимно-однозначно с множеством агрегатных точек. Доказательство этого утверждения просто. Заменой всех нулевых координат для любой агрегатной точки на координаты резерв получаем единственным образом детальную резервную точку. Обратной заменой получаем для каждой детальной резервной точки единственную агрегатную точку. Важно то, что это свойство инвариантно, то есть не зависит от числа измерений модели.

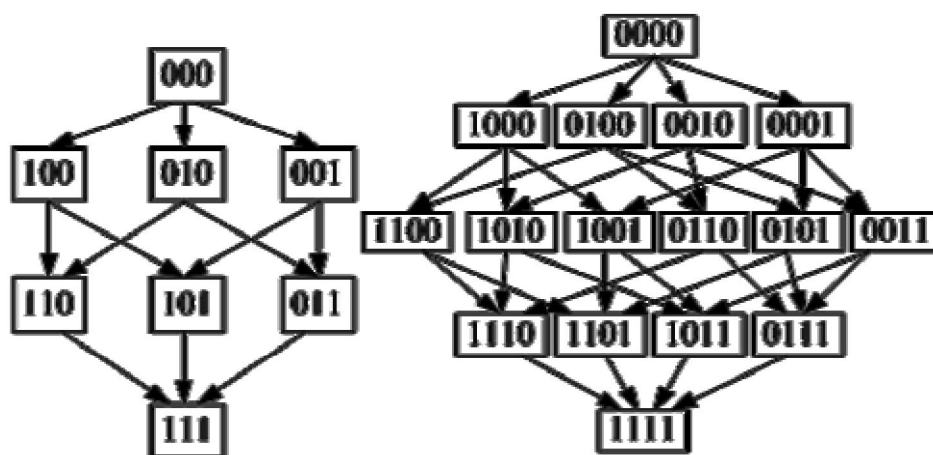


Рисунок 4 - Концептуальная модель для 3 и 4 измерений, раскрытые измерения помечены 1 в именах таблиц

Вторым инвариантным свойством является декомпозиция модели с аддитивным измерителем на элементы данных (кубы данных), каждый построен на своей детальной точке и несет её значение на своих вершинах. В каждой агрегатной точке суммируются значения проходящих через неё элементарных кубов. Таким образом, каждое агрегатное значение всегда декомпозируется на сумму некоторых детальных значений. Элементарный куб инвариантен относительно перестановок измерений и прост для консолидации, поскольку имеет на вершинах постоянное значение.

На инвариантных свойствах построим правило редактирования агрегатных значений (агрегатов). Добавление в систему элементарного куба данных резервной детальной точки со значением необходимого приращения, приносит его в соответствующую агрегатную точку.

Интерфейс как гиперплоскость. Проекционная гиперплоскость – это гиперплоскость, содержащая все ячейки (точки) многомерного пространства с фиксированным значением одной или нескольких координат. В том числе нулевые значения координат позволяют представлять в интерфейсе агрегатные значения. Похожая конструкция известна как сводный отчет, например, в MS Excel. Такой интерфейс обеспечивает непосредственный доступ к детальным и агрегатным точкам пространства и редактирование их значений с соблюдением балансных соотношений за счет учета вносимых расхождений (невязок) в пространстве точек Резерв.

Многомерная модель как контейнер. Детальное рассмотрение элементов ММД показывает, что она содержит определенные множества иерархических, реляционных, сетевых моделей данных. В этом смысле ММД является контейнером моделей остальных типов, поэтому на концептуальном уровне абстракции данных достаточно ограничиться ММД моделями, детализация которых относится к физической реализации - внутреннему уровню абстракции данных.

Любой путь погружения в N-мерное пространство, максимум из N шагов, обеспечивает доступ к детальным данным и представляет иерархическую подсхему. Две и более иерархии дают реляционную подсхему. Они пересекаются по общим сущностям (вершинам гиперкуба), а начала и концы иерархий можно не использовать. Если взять нижнюю сущность (детальную таблицу данных) и выходящие из неё иерархии путей сверток, вплоть до начала координат, то получим подсхему данных типа звезда. Подсхема снежинка, состоит из двух звезд, построенных на разных сущностях, рассматриваемых как детальные таблицы. Оказывается, что подсхемы звезда и снежинка всегда являются частью некоторой ММД, это же относится и к моделям данных других типов.

Не каждая реальная иерархическая или реляционная модель явно представит подсхему общей схемы многомерной модели. Однако верно обратное утверждение. Схема многомерной модели порождает подсхемы иерархического, реляционного, сетевого типа, которых вполне достаточно для описания интегрированных данных. Такого вывода о роли многомерной базы как «контейнера» моделей других типов не было замечено в известных автору работах по сопоставлению многомерного и реляционного моделирования данных.

На проведенном рассмотрении формулируется основная рекомендация для проектирования баз данных. Нужно представлять реляционные подсхемы, принадлежащие многомерной модели, как совокупность иерархий по вершинам бинарного гиперкуба. Такой подход облегчит объединение разных подсхем в одну. Если же рисовать сущности в произвольном порядке, то одну и ту же схему можно преобразовать неузнаваемым образом, даже сохраняя связи между сущностями (разновидность известной задачи изоморфизма графов). Поэтому еще более сложно объединять подсхемы смежных предметных областей.

Чтобы объединить две подсхемы, следует классифицировать их сущности по количеству входящих и выходящих ссылок, чтобы отобразить на шаблон многомерной схемы и идентифицировать сущности с вершинами бинарного гиперкуба, а связи – с его ребрами. После такой классификации объединение подсхем представляет чисто техническую, автоматизируемую задачу. В настоящее время такое объединение подсхем из сотен сущностей и тысяч атрибутов представляет большую проблему.

В шаблоне ЕКПД перспективы пользователей на данные отличаются порядком перечисления сущностей, то есть путями по схеме из начала координат в детальное пространство. Таких путей всего $N!$ для N -мерной модели. Поскольку никакой пользователь не пользуется одновременно всеми измерениями, у каждого подразделения корпорации имеется присущий ему «горизонт» данных. Для смежных по производственным процессам подразделений общие данные находятся на пересечении подпространств.

Этот факт позволяет нам сделать основное предположение шаблона ЕКПД, которое сводится к ограничению доступа в интерфейсе только к любым $0,1,2,\dots,N$ измерениям из всех 2^N корпоративных измерений. Все возможные подсхемы данных расположены в N -мерных гранях, а их объединения, как интегрированная модель, сосредоточены в тонком N -мерном слое бинарного гиперкуба ЕКПД из 2^N измерений.

Третья глава посвящена методике применения шаблона ЕКПД, которое должно основываться на следующих принципах: 1)пошаговая детализация; 2)интерактивность; 3)параллелизм; 3)самообслуживание. Любой пользователь может формировать свою БД из измерений, представлений (кубов данных) и области определения. На этом проектирование заканчивается, следует стадия заполнения БД данными, если же данные выверены, то они сами формируют область определения.

Каждый пользователь может формировать свое представление данных, начиная буквально с одной нулевой точки пространства, предпочитая использовать уже готовые измерения, если они подходят для решаемой задачи. Он может создать полностью свои измерения, однако получит автономно работающую БД, что не является необычным в сегодняшней практике. Если же он будет использовать имеющиеся измерения смежных с ним по производственному процессу пользователей, то это обеспечит связь с ними по данным без потребности в ETL процессе.

Основное требование шаблона ЕКПД к физическому уровню, заключается в реализации транзакций «многомерного» баланса и «быстрой» консолидации при редактировании агрегатных и детальных значений в интерфейсе «сводного отчета», которые следуют из следующего рассмотрения операций над данными.

Перераспределение ресурсов между точками многомерного пространства с сохранением общего баланса ресурса представляется основной операцией пользователей. При регистрации фактов,

соответствующие значения должны вычитаться из планов. Рассмотрим вычислительную модель для одного, двух и более измерений.

Для одного измерения имеем сумму по измерению в нулевой точке X_0 , значение X_R в точке Резерв и отдельные значения по n филиалам X_i , где $i=1, 2, \dots, n$, как представляет формула (1).

$$X_0 = X_R + \sum_{i=1}^n X_i \quad (1)$$

где, например, X_0 – численность работников компании, X_i – численность филиала i , X_R – невязка.

Есть два режима работы с данными: сбор заявок и планирование. При сборе заявок происходит обычное суммирование X_0 , резервы не участвуют. При планировании используется операция пересылки ресурса из одного места пространства в другое, не нарушая закон сохранения ресурса. Операция задаётся количеством ресурса, координатами ячеек источника и цели.

В режиме планирования источником ресурса для филиалов является резерв X_R и общая сумма X_0 при распределениях по филиалам не изменяется. Будем считать, что она принадлежит области ответственности другого пользователя, который изменяет X_0 , чтобы установить лимит ресурса. Операция изменения X_0 является редактированием агрегатного значения по правилу добавления в систему элементарного гиперкуба, основанного на соответствующей детальной точке. В данном случае это одномерный куб с одинаковыми значениями на вершинах X_0 и X_R .

Таким образом, X_R представляет невязку двух пользовательских взглядов на текущее состояние данных (Рисунок 5) и хранит заявки на ресурсы (отрицательные значения) или лимиты для распределения (положительные значения). Пользователи добиваются согласования планов сведением невязки к нулю или достаточно малым значениям.

Обобщение на многомерный случай заключается в многократном повторении одномерной модели, что проявляется в появлении индексов у её членов. Каждый новый индекс соответствует еще одному измерению. Обозрение на экране ограничено двумя физическими измерениями, поэтому в интерфейсе представлена плоскостная проекция многомерной структуры данных, в которой желтым цветом выделены клетки с невязкой.

Покажем, что режим сбора заявок является частным случаем планирования. Можно представить, что в режиме сбора заявок источником ресурса назначается некая «внесистемная» точка и введенные детальные значения влияют на суммарное значение X_0 , например, $XR0$ для распределения по строке $XR20$ (Таблица 1).

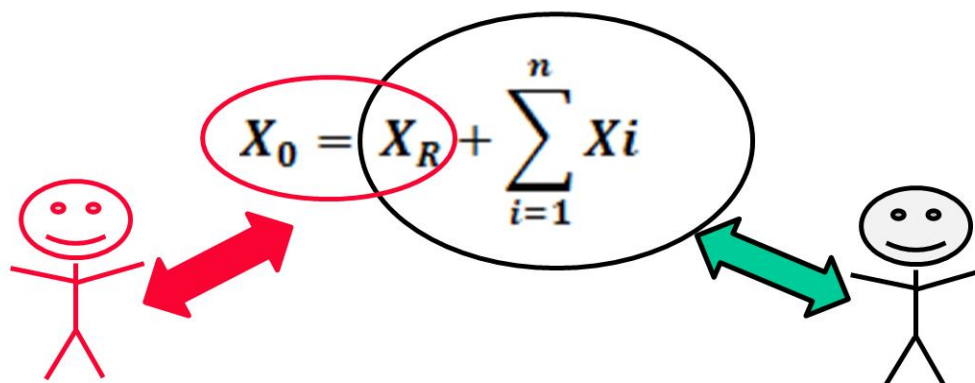


Рисунок 5 – Интеграция перспектив пользователей на примере одномерной модели

Двумерная модель формируется декартовым произведением одномерных моделей, порождая более сложные формулы, нагляднее их представляет таблица сопряженности (Таблица 1), помимо строк и столбцов с итогами в ней присутствуют строки и столбцы с резервами. Для аддитивного измерителя X_{ij} , например, численность персонала, строки образованы измерением Филиал, а столбцы - измерением Служба. Модель сущность связь этого объекта была рассмотрена выше (Рисунок 3-г).

Таблица 1. Многомерная модель как декартово произведение двух измерений Филиал и Служба

Служба Филиал	Итого	Резерв	Служба 1	Служба 2	Служба 3	Служба 4
Итого	X00	X0R	X01	X02	X03	X04
Резерв	XR0	XRR	XR1	XR2	XR3	XR4
Филиал 1	X10	X1R	X11	X12	X13	X14
Филиал 2	X20	X2R	X21	X22	X23	X24
Филиал 3	X30	X3R	X31	X32	X33	X34
Филиал 4	X40	X4R	X41	X42	X43	X44

Здесь присутствуют четыре сущности, три агрегатных таблицы: 1) сумма по строкам и столбцам X00; 2) суммы по строкам X0j; 3) суммы по столбцам Xi0; и одна детальная таблица Xij, где i, j=R, 1, 2, 3, 4.

При изменении детального значения, например, X23, нужно откорректировать ещё три точки: X20, X03 и X00 для сохранения «многомерного баланса». При пересылке ресурса между двумя любыми детальными точками, общая сумма X00 не меняется, к ней добавляется и вычитается одно и то же значение. Таким образом, пересылки ресурса вызывают в пространстве БД локальные возмущения, не выходящие за подпространство, задаваемое источником и целью. Этот эффект верхней границы консолидации позволяет обеспечить независимую работу пользователей и подразделений с коммуникацией по данным посредством полей резерва в интерфейсе.

Для трех и большего числа измерений описанная картина принципиально не меняется. Каждая операция пересылки ресурса сама задает **верхнюю границу консолидации** через координаты точек источника и цели назначения ресурса. Это позволяет, через доступ пользователей к определенным измерениям, обеспечить нисходящие и восходящие процессы планирования.

Нижняя граница консолидации определяется текущим горизонтом данных: при редактировании агрегатного значения, изменяется значение соответствующей детальной резервной точки, доступной в интерфейсе, а не во всем 2^N -мерном пространстве. При добавлении в интерфейс измерения, текущими детальными резервными точками становятся другие, значения в них автоматически корректируются, сохраняя вычислительную целостность. При накоплении в интерфейсе N любых измерений оказывается, что гиперплоскость, задаваемая каждой детальной точкой, есть бинарный 2^N - N мерный гиперкуб с постоянным значением на вершинах. Ведь, в подпространствах с числом измерений от $N+1$ до 2^N только резервные точки могут нести отличные от нуля значения. Таким образом, процессы консолидации всегда ограничены N измерениями.

Структурная схема применения шаблона модели ЕКПД содержит по 3 элемента на стороне клиента и сервера (Рисунок 6):

1. База данных многомерной структуры по шаблону ЕКПД.
2. Метабаза содержит текущее описание данных БД.
3. Подсистема реализации запроса к базе данных.
4. Данные запроса к БД на чтение или на редактирование.
5. Запрос к БД.
6. Интерфейс пользователя на основе сводного отчета.

Построение концептуальных моделей на основе шаблона ЕКПД не требует рассмотрения на физическом уровне реализации и задания типа СУБД. Однако ожидается наиболее естественная стыковка шаблона модели ЕКПД с СУБД UMS-FAD отечественной компании «Х-Технология». Она реализует быстрый доступ (Fast Accessible Data или FAD) к многомерным данным на основе сетевой организации ссылок между записями с одинаковыми координатами, оптимизацией запроса перестройкой дерева поиска на основании ведения частотных статистик в метаданных.

Функциональная схема применения шаблона модели ЕКПД (Рисунок 7) охватывает внешний и концептуальный уровни абстракции данных. Далее при описании структуры шаблона модели кратко отражен подход к реализации структуры хранения данных методом многомерно-сетевым способом доступа.

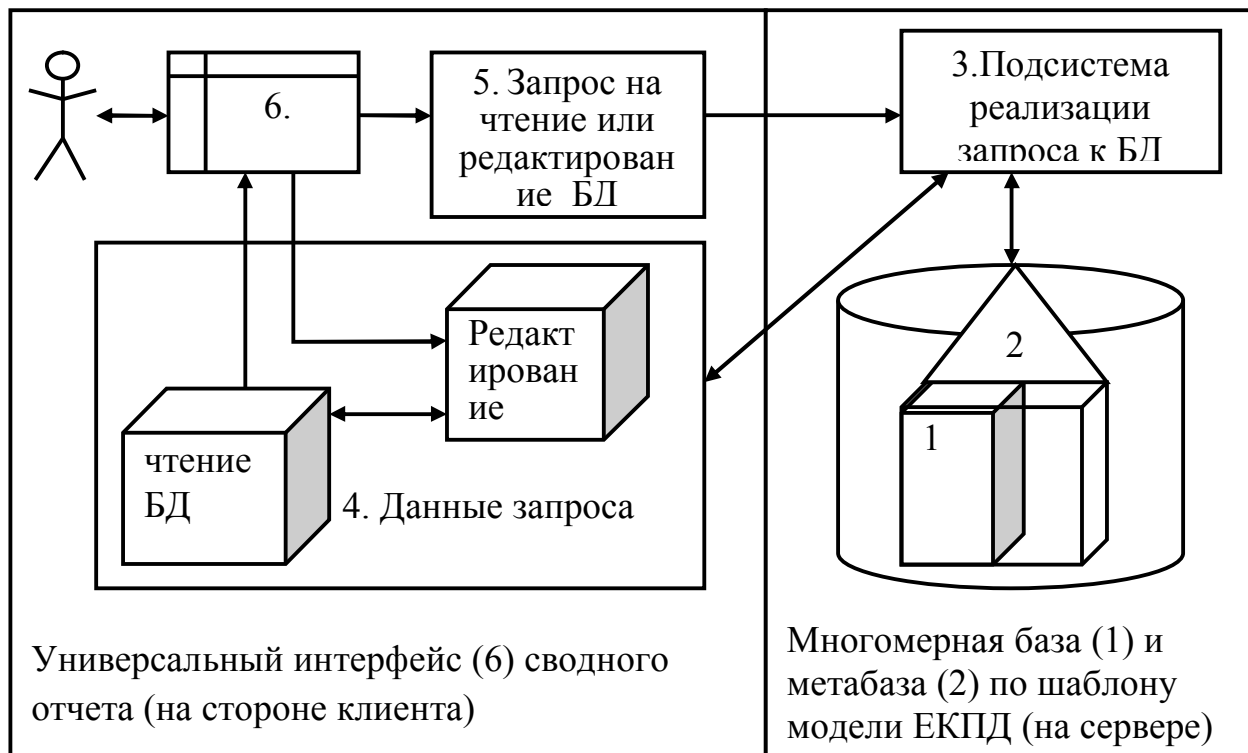


Рисунок 6 - Структурная схема применения шаблона ЕКПД состоит из 6 основных элементов, размещенных на стороне клиента и сервера

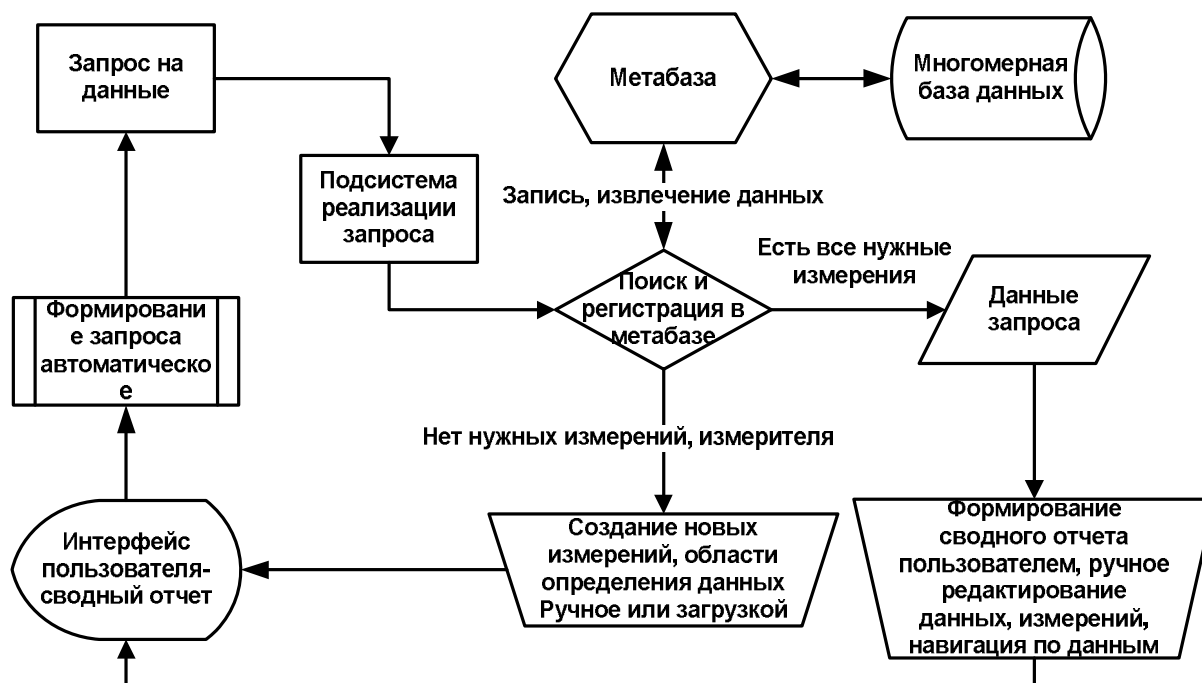


Рисунок 7 - Функциональная схема применения шаблона модели ЕКПД

Рассмотрим подробнее шесть элементов и реализуемые ими функции структурной схемы применения шаблона модели ЕКПД.

1. **База данных** многомерно-сетевой структуры обеспечивает хранение многомерной структуры данных на устройстве прямого доступа в виде записей переменной длины. Эти записи представляют собой описания регистрации фактов, содержат ID измерителя и его числового значения, а также список значений координат, каждая задается парой из ID измерения и ID значения измерения. Для каждого значения координаты в записи хранится указатель на адрес следующей записи с таким же значением координаты. Список координат в записи ограничен «горизонтом» данных из N измерений, это свойство шаблона ограничивает размер каждой перспективы пользователя на данные, обычно N=16. При этом количество всех измерений БД неограниченно.
2. **Метабаза** представляет собой описание данных БД и является частью многомерной структуры в виде справочников измерений, содержит связь между ID и именем для измерения и для каждого его значения. Для каждого значения каждого измерения справочник также хранит адрес последней записи с таким же значением измерения и общее число таких записей.
3. **Подсистема реализации запроса** к базе данных обеспечивает поиск и извлечение из БД всех записей запроса, то есть задаваемой в интерфейсе гиперплоскости, консолидацию данных до нужного уровня детализации. Обеспечивает оценку объема запрошенного количества данных по статистике в метабазе и оптимизацию дерева поиска записей запроса, перестройкой уровней, запись в базу данных новых значений измерителя по месту или добавлением записей, помечает удаленные записи. Работает на основе принципов транзакций «многомерного» баланса и «быстрой» консолидации.
4. **Данные запроса** к БД представляют собой выборку подмножества записей БД по созданному в интерфейсе запросу к БД на чтение или запись. Данные запроса передаются на компьютер клиента или обратно после правок и дополнений после обработки их посредством интерфейса пользователя.
5. **Запрос к БД** формируется инструментальными средствами интерфейса и может описывать данные для чтения из БД или данные для записи в БД. Запрос аналогичен команде SQL, содержит измеритель и его значения со списком измерений и значений измерений, которые однозначно задают координаты ячеек многомерной структуры для чтения или записи значений измерителя.
6. **Интерфейс пользователя (ИП)** является основным инструментальным средством применения шаблона ЕКПД посредством автоматического формирования запросов к БД на основе действий пользователя. Реализует стандартные функции представления данных на экране в виде сводного отчета и навигации по данным. Реализует дополнительные функции по редактированию детальных и агрегатных

значений данных, значений измерений, заданию направлений планирования как набор измерений, задающий текущую ячейку-источник распределяемого ресурса. В интерфейс можно включить любые N измерений из неограниченного числа измерений БД. ИП позволяет задавать область определения БД посредством связывания значений измерений через 0 значение в ячейке, оно не влияет на сумму, но само присутствие его в ячейке разрешает задавать в ней значений измерителя. ИП организован на базе сводного отчета, пример реализации приведен при описании содержания главы 4.

В четвертой главе описаны результаты практического применения шаблона ЕКПД, проверку реализацией в Системе SAS и MS Excel 2007 при решении задачи оценки потребности в средствах для Адресной программы формирования специализированного (технологического) жилого фонда ОАО «РЖД» в разрезах по профессиям, станциям, годам, филиалам, другим измерениям, с учетом региональных нормативов.

Установлена граница оперативной работы (до 1 млн. записей) через интерфейс сводного отчета с базой данных в виде детальной таблицы, при этом все агрегаты считаются «на лету». Серия рисунков, копий экранов работающего макета системы, показывает функциональность интерфейса как инструментального средства шаблона модели (Рисунок 8) и (Рисунок 9).

Интерфейс на основе сводного отчета позволяет редактировать значения в ячейках и распределять данные посредством множественной пересылки. Описаны операции интерфейса для формирования, буквально с одной ячейки, архитектуры БД из новых или имеющихся в метаданных измерений, задания области определения БД.

Интерфейс тестового макета в среде MS Excel2007, функциональность объекта Pivot Table расширена 6 следующими командами всплывающего контекстного меню:

1. «Добавить новое значение измерения/В строку/В столбец» позволяет ввести вручную имя, для больших справочников, например, для железнодорожных станций, организован выбор из справочника.
2. «Синонимы» позволяет объединить разные имена значений измерения под одним именем, для непосредственной правки неверных наименования.
3. «Режим планирования/Режим сбора заявок» включает функцию выбора ячейки источника планирования, задавая верхнюю границу консолидации.
4. «Определить ячейку как источник» позволяет указать ячейку-источник в режиме планирования вместо принятого по умолчанию значения.
5. «Распределить/Равномерно/Пропорционально/Обнулить» - добавление или удаление ресурса в ячейки из множественного выбора областей.
6. «Опустошить ресурс» позволяет разорвать ошибочные связи между измерениями через значения (даже нулевые).

The screenshot displays a software application window titled "AddShortcutMenuItem_v18_19_5_DEMO". It features a pivot table and a configuration panel on the right.

Pivot Table Data:

	Вакансия	Пенсия	Прочее	Ротация	Резерв	Общий итог
ВОСТОЧНО-СИБИРСКАЯ	22	58	68	26		174
ГВЦ				2		2
ГОРЬКОВСКАЯ	15	13	50	3		81
ДАЛЬНЕВОСТОЧНАЯ	101	100	132	11		344
ДЖВ	1			14		15
ЗАБАЙКАЛЬСКАЯ	181	70	94	35		380
ЗАПАДНО-СИБИРСКАЯ	126	229	59	75		489
КАЛИНИНГРАДСКАЯ	16	14	15	5		50
КРАСНОЯРСКАЯ	40	65	23	14		142
КУЙБЫШЕВСКАЯ	5	88	61	31		185
МОСКОВСКАЯ	55	100	113	26		294
ОКтябрьская	71	10	131	31		243
ПРИВОЛЖСКАЯ	196	19		21		236
САХАЛИНСКАЯ	4		27			31
СВЕРДЛОВСКАЯ	54	10	34	35		133
СЕВЕРНАЯ	72	78	50	30		230
СЕВЕРО-КАВКАЗСКАЯ	8	120	37	3		168
ФПД	8	3	4	6		21
ЦДРВ		4		19		23
ЦДРП	22	33	112	6		173
ЦСС	11	19	13	70		113
ЮГО-ВОСТОЧНАЯ	25	27	99	39		190
ЮЖНО-УРАЛЬСКАЯ	33	60	106	49		248
Резерв					9	9
Общий итог	1 066	1 120	1 228	551	9	3 974

Configuration Panel (Список полей сводной таблицы):

- Выберите поля для добавления в отчет:
 - Филиал ОАО РЖД
 - Отделение дороги как территория
 - Основание потребности
 - Потребность в помещениях ЖФК шт
 - Профессия должность
 - Оценка площади ЖФК м2
 - Оценка площади ЖФК 25м2
 - Оценка средств без инфляции тыс руб
 - Оценка средств млн руб
 - Должностной коэффициент площади
 - Должностной коэффициент стоимости
 - Дополнительная площадь м2
 - Хозяйство
 - Дирекция
 - Региональная дирекция
 - Комментарий
 - Железная дорога как территория
 - Станция
 - Регион
 - Региональная стоимость 1м2
- Перетащите поля между указанными ниже областями:
 - Фильтр отчета:
 - Профессия должк...
 - Отделение доро...
 - Год
 - Станция
 - Хозяйство
 - Названия столбцов:
 - Основание потр...
 - Названия строк:
 - Филиал ОАО РЖД
 - Значения:
 - Сумма по полю П...

Рисунок 8 - Инструментальное средство шаблона ЕКПД на основе сводного отчета с расширенной функциональностью, ячейка источник выделена рамкой, ресурс можно назначать в любые ячейки данных

В заключении приводятся основные результаты и выводы по работе, обсуждаются перспективы дальнейшего развития исследований в области концептуального проектирования баз данных.

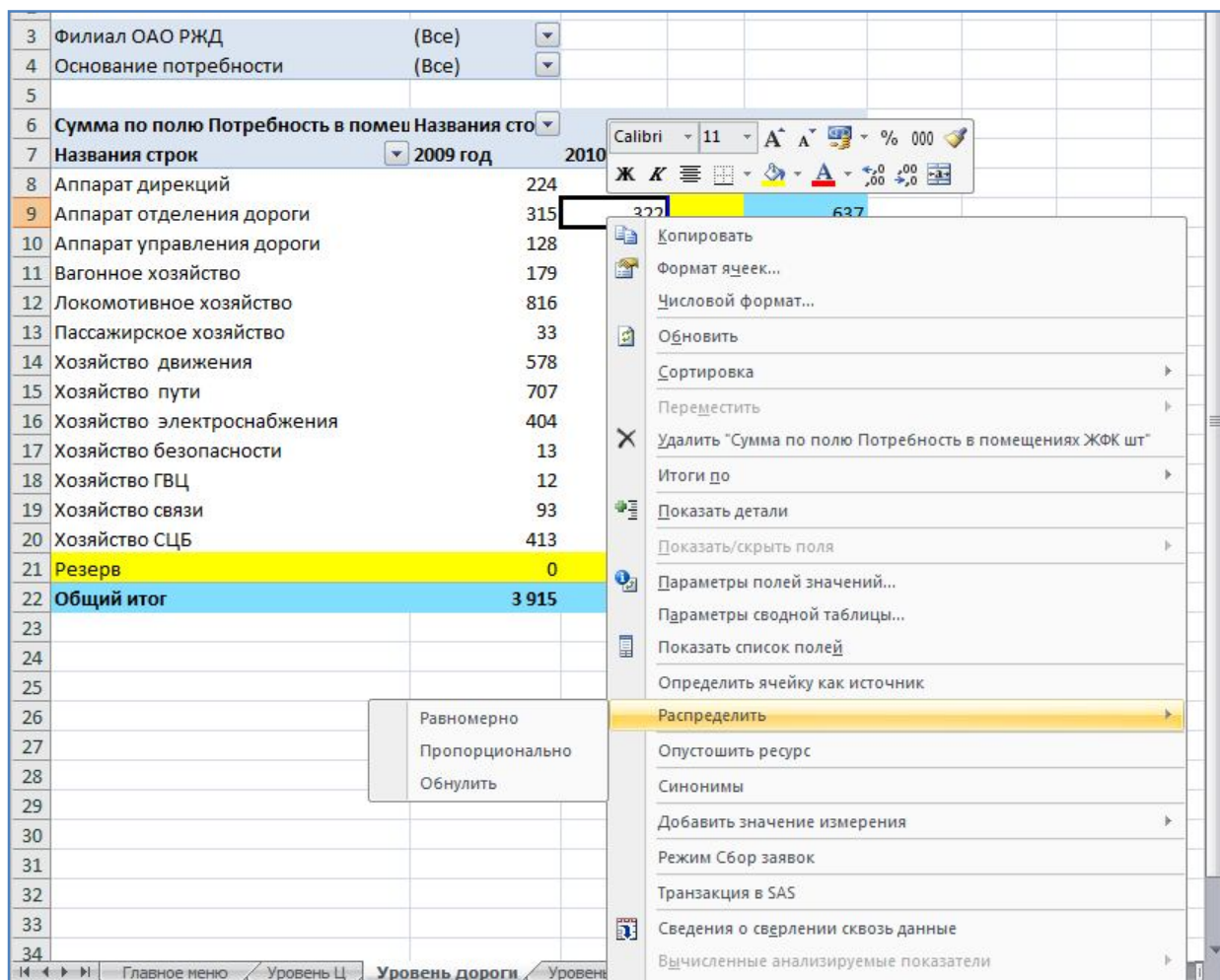


Рисунок 9 - Всплывающее контекстное меню содержит команды по редактированию сводного отчета

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Главным результатом работы является решение актуальной научно-технической задачи, состоящей в разработке комплекса методов и средств концептуального проектирования баз данных на основе шаблона единого корпоративного пространства данных. При этом:

1. Исследованы возможности современных подходов концептуального проектирования баз данных, выявлены причины, препятствующие объединению моделей и построению неограниченно больших БД.
2. Предложен способ проектирования и работы с базами данных на основе многомерного подхода – шаблона ЕКПД, который обеспечивает построение стандартных подсхем данных и объединения их в корпоративную схему.
3. Предложен метод использования интерфейса сводного отчета не только для просмотра и навигации по данным, но и для редактирования измерений и области определения БД, а также детальных и агрегатных значений данных на основе транзакций «многомерного» баланса и «быстрой» консолидации.
4. Показано, что шаблон ЕКПД представляет собой «контейнер» различных типов БД иерархической, реляционной, сетевой структур, который наилучшим образом подходит для задач концептуального проектирования БД. При этом получаем архитектуру БД с естественно заданным интерфейсом сводного отчета (гиперплоскости), с доступом к любому элементу данных с возможностью его редактирования.
5. Предметная область применения - системы корпоративного управления верхнего уровня, поддерживающие процессы планирования, бюджетирования, управленческой отчетности на основе оперативного взаимодействия. Метод позволяет системно увязать одновременную работу иерархии многих пользователей, например, службы экономики крупной корпорации.
6. Показано, что структура шаблона ЕКПД и инструментальные средства интерфейса позволяют ускорить процессы принятия решений экономистами. В каждом документе интерфейса есть строка и столбец с координатами Резерв, представляющие текущие невязки ресурсов. Они обеспечивают коммуникации пользователей при одновременной работе над общими данными. Ориентируясь на значения невязки при принятии решения, экономисты могут самостоятельно оценивать её значимость, и не обязаны дожидаться сбора данных до последнего процента.
7. Показана возможность достигать лучшего позиционирования информационной службы, снимая с неё ответственность за качество данных и сроки их сбора (аналогия с электронной почтой), поскольку интуитивно понятный интерфейс, дополненный журналом регистрации

событий, позволит управлять обработкой данных в режиме самообслуживания.

СПИСОК ПУБЛИКАЦИЙ

Статьи в рецензируемых изданиях, рекомендованных ВАК

1. Гришенков Е.А. Построение концептуальных моделей баз данных при помощи шаблона модели единого корпоративного пространства данных// **Системы управления и информационные технологии**, 3(41), 2010. - С. 81-87
2. Гришенков Е.А. Планирование и консолидация многомерной базы // **"Открытые системы/СУБД"**, апрель, 2001, С. 65-72
<http://www.osp.ru/os/2001/04/180095/p3.html>, авторский вариант:
<http://www.olap.ru/basic/consolidation.pdf>

Материалы международных конференций

3. Eugene Grishenkoff, Informational System for Planning and Consolidation // **SUGI26 Proceedings, April 2001, USA**, С.116-121,
<http://www2.sas.com/proceedings/sugi26/p116-26.pdf>
4. Гришенков Е.А. Информационная геометрия // Тезисы докладов 2-й Московской конференции “Декомпозиционные методы в математическом моделировании и информатике”, ВЦ РАН, Москва, 2004, С. 50-51, <http://www.ccas.ru/Pavlovsky/confer/works.pdf>
5. Гришенков Е.А., Гельфанд Е. В. Информационно-аналитическая система анализа, оценки, планирования контингента производительности труда и ФОТ на Куйбышевской железной дороге // Информационные технологии на железнодорожном транспорте : доклады IV Междунар. науч.-практич. конф. ИНФОТРАНС-99 (4 ;1999; С.-Петербург) / ПГУПС и др.-СПб;ПГУПС, 1999,303с.ил.-ISBN 5-7641-0036-4 , С.125
6. Гришенков Е. А., Информационно-аналитическая система сметного планирования // Информационные технологии на железнодорожном транспорте: Аннот. докл. междунар. науч.-практ. конф. Инфотранс-2000 (5; 2000; СПб) / ПГУПС и др.-СПб,2000.-80с.,С.34,
http://library.pgups.ru/jirbis/index.php?option=com_irbis&Itemid=300
7. Гришенков Е. А., Информационно-аналитическая система сметного планирования//Информационные технологии на железнодорожном транспорте: Докл. и аннот. докл. междунар. науч.-практ. конф. Инфотранс-2000 (5; 2000; СПб): сборник/ ПГУПС, Октябрьская ж.д.. Информац.-вычислит. центр, МПС РФ, Междунар. акад. наук высш. шк.. Петербург. отд-ние, Академия транспорта РФ, Ассоциация вузов

железнодорожного транспорта, Digital Design. Инфотранс-2000 (5; 4-7 окт. 2000; СПб). - СПб, 2000. - 304 с.: ил., С.127

Статьи

8. Грищенко Е.А. Железная дорога в фокусе информационных технологий // "READ.ME", № 07, 2007, С. 23-26,
<http://www.sas.com/offices/europe/russia/articles/1999/rm07.html>