

УДК 004.6

С.В. Липатова, А.А. Смагин, А.М. Фирулин, П.И. Смикун, В.Г. Захаров

МЕТОДИКА ПОСТРОЕНИЯ СИСТЕМЫ КОМПАКТНОГО ХРАНЕНИЯ ДАННЫХ, ПРЕДНАЗНАЧЕННЫХ ДЛЯ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ

Липатова Светлана Валерьевна, кандидат технических наук, доцент, окончила факультет информационных телекоммуникационных технологий и сетей Ульяновского государственного университета. Доцент кафедры «Телекоммуникационные технологии и сети» УлГУ. Имеет статьи в области разработки информационных систем различного назначения. [e-mail: dassegel@mail.ru].

Смагин Алексей Аркадьевич, доктор технических наук, профессор, окончил радиотехнический факультет Ульяновского политехнического института. Заведующий кафедрой «Телекоммуникационные технологии и сети» УлГУ. Имеет статьи, изобретения, монографии в области разработки информационных систем различного назначения. [e-mail: smagina1@mail.ru].

Фирулин Александр Михайлович, аспирант, окончил факультет математики и информационных технологий УлГУ по специальности «Информационные системы и технологии». Младший консультант-разработчик регионального центра IBS «Ульяновск». Имеет статьи в области информационных систем и технологий. [e-mail: afirulin@mail.ru].

Смикун Петр Иванович, кандидат технических наук, окончил факультет автоматики и вычислительной техники Таганрогского радиотехнического института. Главный конструктор, начальник научно-исследовательского отделения ФНПЦ АО «НПО «Марс». Имеет ряд статей в области разработки программного обеспечения для АСУ. [e-mail: smikun@mail.ru].

Захаров Вадим Геннадьевич, заместитель главного конструктора ФНПЦ АО «НПО «Марс», окончил факультет информационных технологий и телекоммуникаций УлГУ, специализируется в области разработки ПО АСУ специального назначения. Имеет научные публикации в области разработки информационных систем различного назначения. [e-mail: vg.zaharov@gmail.com].

Аннотация

В данной статье предлагается методика построения системы компактного хранения данных, предназначенных для аналитической обработки. Здесь описываются все этапы построения системы: от анализа задачи в конкретной предметной области до внедрения. В частности, рассматриваются вопросы организации данных, их загрузки из различных источников, очистки, сжатия, восстановления и аналитической обработки. Для наглядности методика изображена, в том числе, в виде схемы бизнес-процессов.

Предлагаемая методика может использоваться в различных предметных областях для решения разнообразных задач. Авторами ведутся работы по применению методики к задаче оценки надежности программно-аппаратных средств. Статья рассчитана на людей, специализирующихся в области хранения и аналитической обработки данных.

Ключевые слова: компактное хранение, временной срез, анализ данных.

THE METHOD FOR CONSTRUCTING A SYSTEM OF COMPACT STORAGE OF DATA USED FOR ANALYTICAL PROCESSING

Svetlana Valerevna Lipatova, Candidate of Engineering, Associate Professor; graduated from the Faculty of Information Technologies and Telecommunications of Ulyanovsk State University; Lecturer at the Department of Telecommunications Technologies and Networks of Ulyanovsk State University; an author of articles in the field of multipurpose information system development. e-mail: dassegel@mail.ru.

Aleksei Arkadevich Smagin, Doctor of Engineering, Professor, Head of the Department of Telecommunications Technologies and Networks at Ulyanovsk State University; graduated from the Radioengineering Faculty of Ulyanovsk Polytechnic Institute; an author of articles, inventions, and monographs in the field of different-purpose information systems development. e-mail: smaginaa1@mail.ru.

Aleksandr Mikhailovich Firulin, Postgraduate Student; graduated from the Faculty of Mathematics and Information Technology of Ulyanovsk State University by the specialty in Information Systems and Technologies; Junior Consultant-Developer of the Regional Center of IBS "Ulyanovsk"; an author of articles in the field of information systems and technologies. e-mail: afirulin@mail.ru.

Peter Ivanovich Smikun, Candidate of Engineering; graduated from the Faculty of Automation and Computer Engineering of Taganrog Radioengineering Institute; Chief Designer, Head of the Research Department at FRPC JSC 'RPA 'Mars'; an author of a number of articles in the field of automated control system software development. e-mail: smikun@mail.ru.

Vadim Gennadevich Zakharov, Deputy Chief Designer at FRPC JSC 'RPA 'Mars', graduated from the Faculty of Information Technologies and Telecommunications at Ulyanovsk State University; specializes in the field of special-purpose automated control system software development; an author of scientific papers in the field of different-purpose information systems development. e-mail: vg.zaharov@gmail.com.

Abstract

The article considers a method for constructing a compact storage system of data used for analytical processing. The authors describe all stages of constructing system: from analysis of task in a particular subject area to implementation. Particularly, the question of organizing data, downloading them from various sources, purification, compression, restoration, and analytical processing are discussed in this article. In order to clarify some points, the method is shown as business process diagram.

The proposed method can be used in different subject areas for solving a wide range of tasks. The authors investigate the possibility of using the method for estimating the reliability of hardware-software equipment.

This article is intended for specialists in data storage and analytical data processing.

Key words: compact storage, time slice, data analysis.

ВВЕДЕНИЕ

Для аналитической обработки данных возникает необходимость их загрузки из различных источников, преобразования к единой структуре и дальнейшего хранения. При этом требуется хранить не только оперативные, «свежие» данные, но и исторические.

Объемы хранимых данных постоянно увеличиваются, растет и потребность в их анализе. В связи с этим строятся специализированные центры обработки данных [1], разрабатываются новые и совершенствуются существующие методы хранения данных [2–5].

Согласно данным портала TAdviser (<http://www.tadviser.ru/>), в России в третьем квартале 2014 года наблюдался рост емкости систем хранения данных на 22,3%, а стоимость систем, поставленных на российском рынке, превысила \$114,38 млн. Кроме того, все больше крупных российских компаний используют системы хранения данных («Разгуляй», «Банка Хоум Кредит», «Райффайзенбанка», телекоммуникационной компании «Мегафон» и других).

Задача хранения архивных данных может рассматриваться, во-первых, с точки зрения организации физической памяти (оптические диски, флеш-память, жесткие диски, технология Redundant array of independent disks), во-вторых – логической организации данных (хранилища данных, технологии Business intelligence (BI) и Extract, Transform, Load (ETL)), и в-третьих – сжатия данных (методы сжатия и архивации).

В данной статье рассматриваются вопросы разработки моделей представления данных для различных временных периодов их накопления, алгоритмов преобразования из одной модели в другую, включающих эффективное хранение актуальных и архивных данных, необходимых для аналитической обработки.

Целью статьи является разработка методики построения системы компактного хранения данных и проекта информационной системы, реализующей данную методику.

Для достижения этой цели были последовательно рассмотрены методы:

- представления данных с учетом периода их хране-

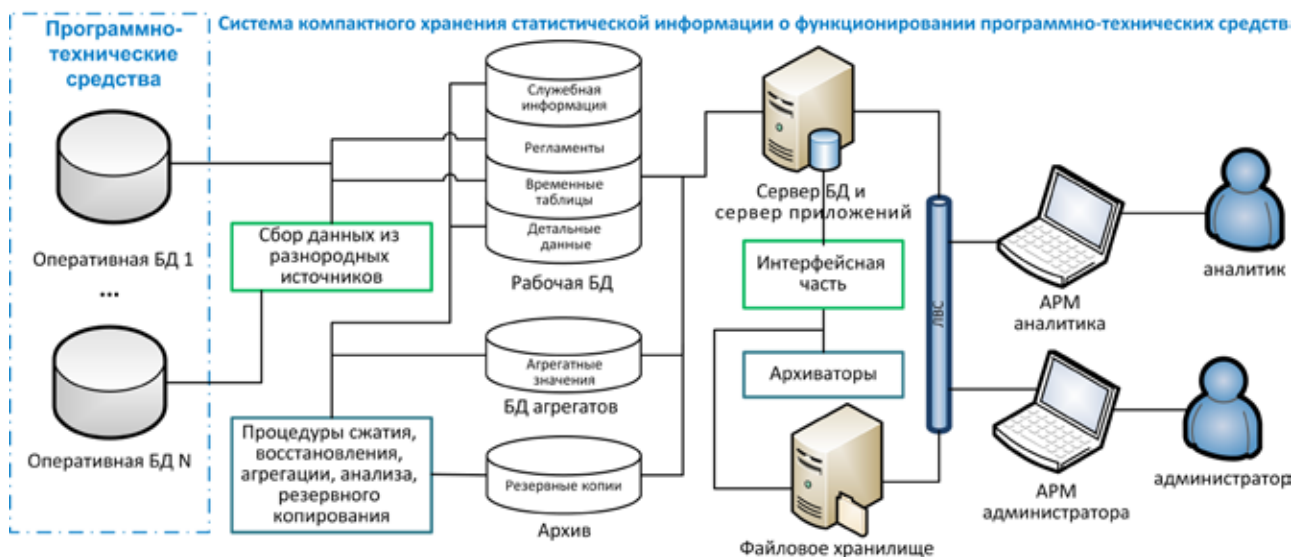


Рис. 1. Общая схема системы компактного хранения статистических данных

ния (текущие данные, квартальные данные, годовые, за последние 3–5 лет, архивные данные);

- синхронизации и очистки данных, поступающих из разных источников;
- сжатия информации для более компактного представления;
- агрегации данных (определение агрегированных показателей, способов их вычисления и регламента их получения);
- аналитической обработки для разработанных моделей представления данных.

ОПИСАНИЕ СИСТЕМЫ КОМПАКТНОГО ХРАНЕНИЯ ДАННЫХ

Входные оперативные данные, сохраняемые в хранилище, поступают из различных источников данных. Источник данных не обязательно является реляционной базой данных (БД), это может быть файл, многомерная БД и т. д.

Для разрабатываемой методики введем понятие «временной срез». В многомерной модели представления данных, наиболее часто используемой для аналитической обработки, срез представляет собой набор данных, полученных в результате фиксации одного или нескольких измерений.

В задаче компактного хранения под временным срезом понимается вся совокупность данных, накопленных за определенный период времени. Примерами периодов для временных срезов являются день, неделя, квартал, год и т. д. Количество срезов определяется для каждой конкретной задачи.

Система компактного хранения должна выполнять ряд функций:

- 1) сбор и согласование данных из гетерогенных источников данных;
- 2) предварительный анализ данных (выявление штатных и внештатных ситуаций);
- 3) получение агрегированных значений и статистических оценок за указанные периоды (на каждом временном срезе);

4) удаление лишней информации (не имеющей значения для задачи анализа);

5) сжатие и восстановление данных (отдельных полей таблиц, таблиц, БД);

6) ведения регламентов, описывающих расписание и схему работы процедур сжатия, агрегации и предварительного анализа;

7) представления данных в виде, удобном для аналитика (отображение детальных данных и агрегированных данных с возможностью выбора периода, объекта, изделия, сравнение данных за разные периоды).

Система хранения данных использует архитектуру «клиент-сервер» и для выполнения перечисленных функций предлагается следующая структура (рис. 1). К уровню хранения данных в системе относятся оперативные БД, являющиеся источниками данных, БД собственно хранилища (рабочая, агрегатов, резервных копий, физически это может быть одна база) и файловое хранилище сжатых резервных копий (каталог в файловой системе).

Уровень представления реализуется интерфейсной частью, которая позволяет обеспечивать администрирование процессов хранения и сжатия данных и просмотр сохраненных данных. На уровне бизнес-логики находятся процедуры обработки и сжатия данных. Пользователями системы являются администратор, который позволяет настроить соединение с источниками данных и контролирует работоспособность системы, и аналитик, который настраивает регламенты функционирования системы и использует данные для решения своих задач.

Базы хранилища содержат разнотипные данные (табл. 1). Исходные данные условно разделим на 2 типа: справочные и статистические. Справочные данные – это декларативные данные, которые редко изменяются в рамках одного источника (например, перечень оборудования на объекте, режимы работы). Статистические данные описываются временными рядами и хранят значения изменяющегося во времени параметра (например, текущая температура процессора, количество отказов за месяц).

В зависимости от типа данных выполняется различная обработка. В случае редко изменяющихся данных и содержащих много текстовой информации нет необходимости их сохранять на разных срезах. Статистическому анализу подвергаются данные, представленные временными рядами, и они контролируются на уровне задачи, обслуживаемой хранилищем.

В настоящей статье под внештатной ситуацией подразумевается отклонение контролируемых параметров (например, повышение температуры процессора до критической температуры) от заранее определенных интервальных значений. Данные о внештатных ситуациях сохраняются полностью и сжимаются без потерь, а в штатных – сохраняются агрегаты. Детальные данные хранятся на первом срезе, а последующие срезы накапливают агрегатные значения (например, среднее значение за час, среднее значение за день, среднее значение за неделю и т. д.).

Как известно, эффективного сжатия можно достичь, если его производить с потерями, и с этой целью необходимо определять на уровне описания штатных и внештатных ситуаций, какие именно данные можно отбросить, а что оставить в виде агрегатных показателей.

Сообщения небольшой длины имеют малый коэффициент сжатия. Поэтому сжатие данных на уровне атрибутов малоэффективно. С учетом необходимости получения среза данных для задачи анализа более эффективным будет сжатие всего среза, т. е. резервной копии данных за определенный период времени. Такие сжатые резервные копии можно хранить как в файловом хранилище, так и БД.

Для построения такой системы могут быть использованы следующие технологии хранилища данных: Online Transaction Processing (модели для хранения оперативной информации), Online Analytical Processing (модель представления данных для аналитической обработки),

ETL (алгоритмы преобразования, очистки и синхронизации данных), Data Mining (методы и алгоритмы аналитической обработки данных).

Предлагаемая в статье методика опирается на существующие подходы к обработке данных: методы сжатия, методы выявления внештатных ситуаций предметной области (на основе отклонений от заданных параметров значений атрибутов), а также на модель представления данных, модель временных срезов. Для построения хранилища для заданной задачи данные должны быть описаны аналитиком, с указанием того, что они являются входными данными для процессов хранения и обработки данных (анализ, сжатие, хранение).

Тогда распределенные и сохраненные по моделям представления для разных временных срезов массивы данных будут являться выходными наборами данных. На рисунке 2 представлена взаимосвязь данных, подлежащих процессам обработки и хранения результатов в хранилище.

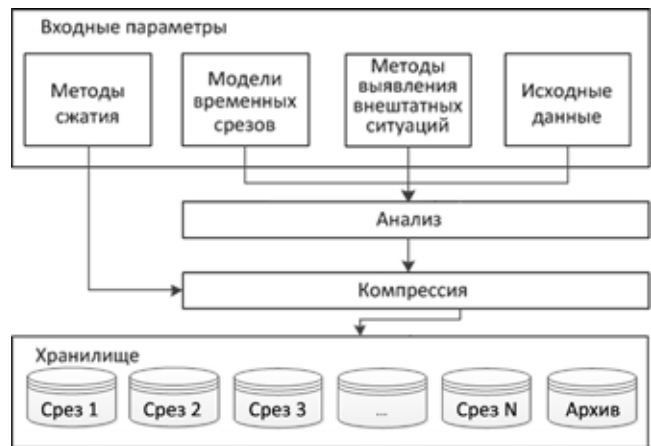


Рис. 2. Схема процессов обработки данных и хранения

Таблица 1

Организация структуры хранения системы

Система хранения	Область хранения	Вид информации	Временные срезы
Внешние БД (источники)	Оперативные БД	Оперативные данные	-
Рабочая БД системы	Область временного хранения	Данные из нескольких источников данных (несогласованные)	-
	Область служебной информации	Регламенты функционирования	-
	Область постоянного хранения детальных данных	Детальные согласованные данные	Справочные данные
Статистические данные			Срез 1
Рабочая БД системы и дисковое пространство (файловое хранилище) или отдельная архивная БД с VLOB	Архив	Резервные копии данных	-
БД агрегатов (при больших объемах данных и длительном периоде хранения на каждый срез своя база)	Область постоянного хранения агрегатов	Агрегатные значения	Срез 1–Срез N

Данные предварительно анализируются по временному фактору, подвергаются проверке на наличие отклонений от заданных значений сохраняемых атрибутов, затем происходит их сжатие и размещение на определенный уровень хранилища.

ОПИСАНИЕ МЕТОДИКИ ПОСТРОЕНИЯ СИСТЕМЫ КОМПАКТНОГО ХРАНЕНИЯ ДАННЫХ

На рисунке 3 представлена предлагаемая методика применения моделей и алгоритмов компактного хранения и аналитической обработки данных. Рассмотрим более подробно каждый из этапов методики.

Анализ задачи в конкретной предметной и проблемной области

На данном этапе определяется, требуется ли построение специального хранилища данных для конкретной задачи (возможно объем данных в задаче не является достаточно большим или в данной задаче недопустимо сжатие с потерями и не требуется получение агрегатных значений).

Проектирование системы компактного хранения

На этом этапе происходит определение перечня источников данных. Ими могут быть оперативные БД, файлы разных форматов, потоки сообщений и др. Перечень зависит от конкретной задачи.

Важной частью этапа является выбор модели представления данных, в качестве которой могут быть применены реляционная, постреляционная, многомерная, объектно-ориентированная, объектно-реляционная модели. Здесь необходимо учесть вид модели представления данных в самом хранилище, если оно уже существует или если уже выбрана система управления БД (СУБД).

Построение модели представления данных для конкретной задачи целесообразно проводить на концептуальном уровне (причем следует учитывать допустимость связей «много ко многому», использование для описания объектов предметной области естественно-языковых конструкций). Модель может быть представлена в виде ERD (Entity-Relationship Diagram). Не следует забывать, что разработанная модель должна содержать перечень всех рассматриваемых данных.

Объектом для хранения является информация, накапливаемая на протяжении долгого времени, и задача хранения – представление их для аналитической обработки по запросу пользователя этих данных в требуемый момент времени. Как правило, нет необходимости сохранения абсолютно всех данных, так как со временем часть информации теряет свою значимость и актуальность, становятся важны лишь агрегированные значения (среднее, максимальное, минимальное, математическое отклонения и т. д.)

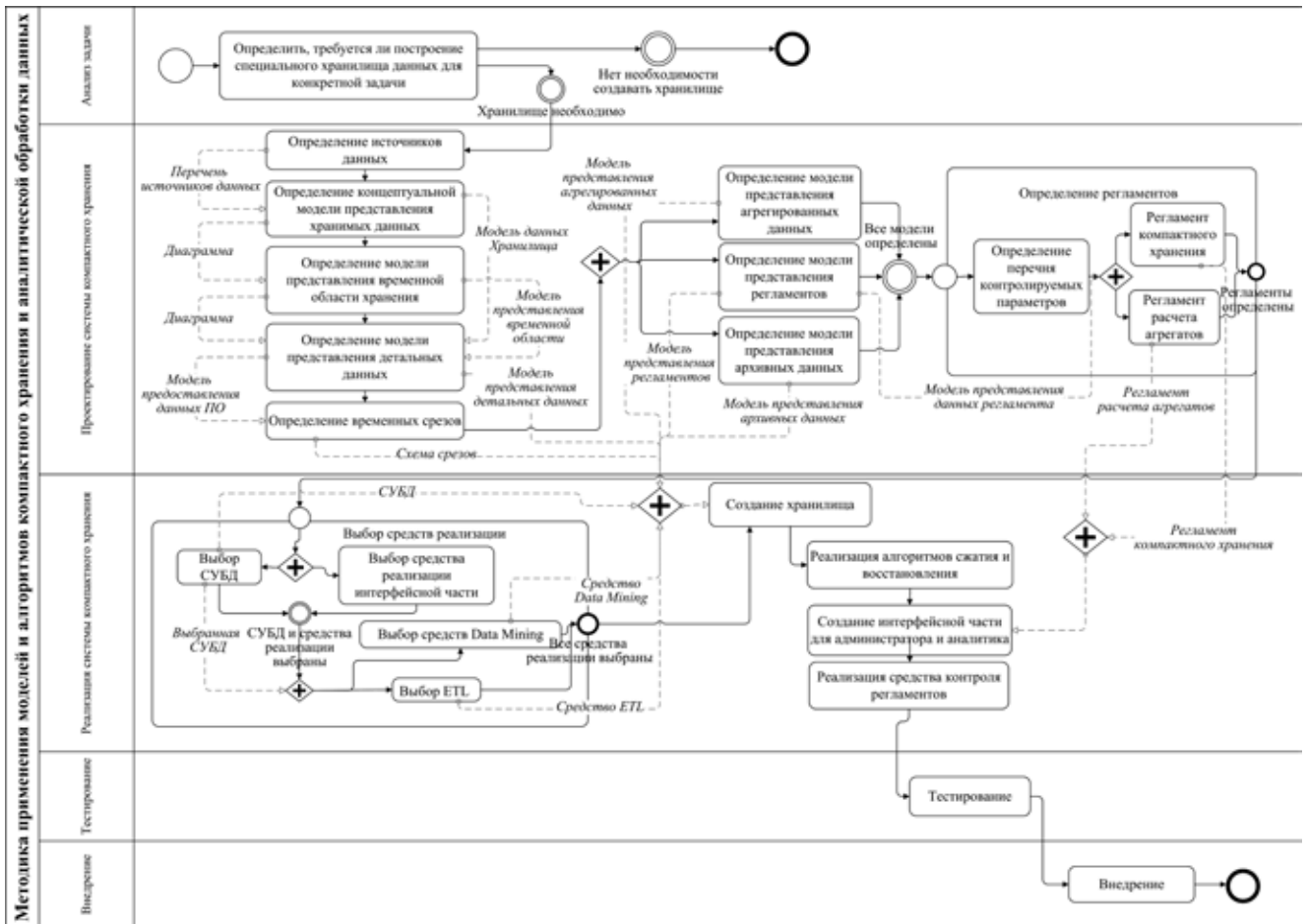


Рис. 3. Схема методики построения системы компактного хранения данных

для штатных ситуаций и отклонения от них – во внештатных ситуациях.

Учитывая эти особенности, на рассматриваемом этапе необходимо определить значимые для контроля и анализа временные интервалы (час, день, неделя, месяц и т. д.).

Здесь же производится построение логической модели представления данных для конкретной задачи (для описания объектов предметной области вводятся понятия и сокращения, если используется реляционная модель, то в ней проводится нормализация). Одним из вариантов модели может быть модель ERD.

Важным требованием к логической модели является необходимость сохранения в ней всех данных от источников; если одни и те же данные имеются в разных источниках, то во временной области все они сохраняются для последующего сравнения и выбора из них только одного.

Следует отметить, что для анализа данных часто не требуются оперативные данные, а нужны их интегральные (агрегатные) оценки, при этом учитывается, что агрегатная модель описывает все возможные агрегаты, требуемые для текущей задачи. Кроме собственно агрегатного значения, необходимо хранить время получения параметра и количество полученных за временной промежуток данных.

Отсюда возникает необходимость построения модели архивных данных, в которой указывается, с какой частотой будет формироваться архивная версия данных и каким способом будет храниться сжатая архивная версия.

Например, архивная версия может создаваться один раз в год. Значит, в одну архивную часть будут выбираться данные за текущей год и в модели должны быть параметры, описывающие архивную составляющую за год (год, количество записей, размер, контрольная сумма и т. д.). Для хранения архивной составляющей может быть выбран BLOB-формат поля или сохраняться ссылка на архивный файл. В этом случае должна быть еще определена структура файлового хранилища. Общая модель представления данных для хранения выглядит следующим образом (рис. 4).

Для выполнения компактного хранения требуются регламенты (порядок выполнения алгоритмов компактного хранения), которые также должны сохраняться в системе хранения и обработки. Поэтому возникает необходимость построения модели представления данных о регламентах. В такой модели должны присутствовать атрибуты, представленные в таблицах 2 и 3.

В построенной модели представления данных требуется заполнение таблиц регламентов для всех данных. Справочную информацию, которая редко изменяется, целесообразно сохранять один раз, так как такая информация требуется для всех уровней хранения (временных срезов). Она не содержит статистических данных и значит не требует агрегации (кроме подсчета количества записей). Поэтому она будет отправляться с оперативного уровня в архив на последнем временном слое, для которого выполняется сжатие всей резервной копии.

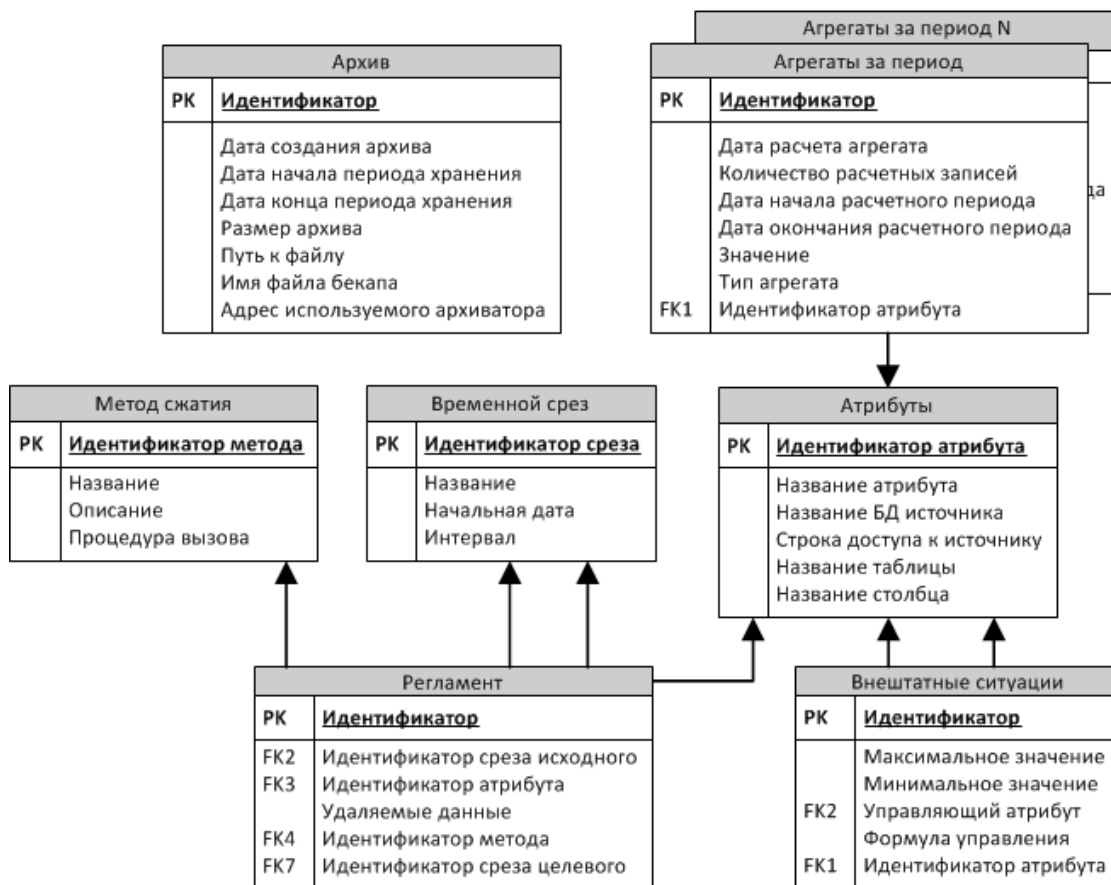


Рис. 4. Модель хранения данных о регламентах и архиве

Таблица 2

Регламент компактного хранения

Оперативная БД	Отношение	Атрибут	Временной срез исходный	Временной срез целевой	Класс данных	Удаляемые (истина /ложь)	Метод
----------------	-----------	---------	-------------------------	------------------------	--------------	--------------------------	-------

Таблица 3

Регламент расчета агрегатов

Таблица	Атрибут	Временной срез	Агрегатная функция
---------	---------	----------------	--------------------

Статистическая информация предполагает накопление агрегатов на промежуточных временных срезах: неделя, месяц, квартал, год.

Таким образом, можно условно выделить два класса данных: справочник и статистика.

Для компактного хранения данных можно использовать комбинацию подходов к представлению и сжатию данных (табл. 4).

Для описания методики эффективного хранения введем понятие класс данных, который будет обозначать, какой из подходов требуется применять к конкретным данным.

Для организации эффективного хранения данных необходимо:

- определить количество временных срезов данных (минута, час, день, сутки, неделя, месяц, квартал, год, 3 года, 5 лет, 10 и т. д.);

- определить для всех атрибутов отношений оперативной БД атрибут-класс-срез-конкретный метод (сжатия, аппроксимации, формулу и т. д.). Для этого необходимо заполнить таблицу – регламент компактного хранения (табл. 2).

Так как данные могут поступать из разных БД, то требуется указать БД, отношение и атрибут. Эти три параметра будут идентифицировать хранимый параметр в оперативных БД.

Возможно применение разных подходов для различных временных срезов. Поэтому требуется указание исходного среза данных и целевого, причем преобразование может выполняться не последовательно (например, с первого среза в третий). Для каждого подхода существует множество методов, которые реализованы в виде готовых программных средств и библиотек [6–11].

В процессе преобразования данных на новый срез, с текущего среза эти данные удаляются. Может возникнуть

ситуация, когда при сжатии возникают потери, которые не допустимы. Целесообразно не удалять сжатые данные на текущем слое, чтобы иметь возможность их восстановить при необходимости в полном объеме. Однако этот прием может привести к увеличению объема требуемой памяти, и его лучше использовать для небольшого процента данных, описывающих внештатные ситуации, при этом для каждого атрибута необходимо установить значение флага «Удаляемые».

Как указывалось ранее, для анализа данных часто не требуются оперативные данные, нужны их интегральные оценки. Для того чтобы иметь возможность настроить соответствие агрегированных функций временным срезам и атрибутам таблиц, необходим регламент, описывающий этот процесс. В качестве примера приведем регламент расчета агрегатов для задачи мониторинга программно-технических средств (табл. 3).

Реализация хранилища данных

В зависимости от имеющегося опыта, используемых на предприятии систем, требований к аналитическим функциям, платформе, выбранной модели представления данных и имеющихся ресурсов требуется выбрать СУБД, на основе которой будет строиться хранилище.

Технология ETL (от англ. Extract, Transform, Load) обеспечивает выполнение одного из основных процессов в управлении хранилищами данных, который включает в себя:

- извлечение данных из внешних источников;
- трансформацию и очистку данных для нужд бизнес-модели;
- загрузку в хранилище данных.

Подсистема, реализующая хранилище, использует об-

Таблица 4

Подходы к представлению и сжатию данных

Подход	Класс данных	Рекомендовано к применению
Сжатие без потерь	Класс 1	Применять к критичным данным, чье искажение недопустимо
Сжатие с потерями	Класс 2	Применять к данным, для которых допустим определенный процент потерь и искажения
Агрегирование значений	Класс 3	Применять при переходе от одного временного среза хранения к другому; к данным, требующим приблизительной оценки
Аппроксимация рядов значений одного параметра	Класс 4	Применять к данным, в значениях которых существует зависимость от временного параметра
Вычисление	Класс 5	Применять к параметрам, которые зависят от других параметров, и данная зависимость известна

ласть временного хранения данных для извлечения, преобразования и оперативного хранения данных, а область постоянного хранения – для загрузки и агрегации данных в хранилище.

Требуется оценить, насколько ценным является использование ETL. Часть СУБД поддерживают данные функции (BI), в некоторых задачах источники гомогенные, и объединение данных, их согласование не является сложной задачей. В этом случае не рекомендуется применять дополнительно ETL-систему. Однако для случая гетерогенных источников и разнообразных моделей представления данных получение и согласование данных может оказаться нетривиальной задачей, требующей специального инструмента. Здесь применение ETL оправдано и требуется выбрать одну из существующих систем, например: Warehouse Builder от Oracle, JasperETL от JasperSoft, Data Transformation Service от Microsoft, Open Studio от Talent, Pentaho ETL от The Pentaho BI Project, учитывая ресурсы, опыт разработчиков, платформу и функциональные возможности.

Возвращаясь к задачам, решаемым средствами DataMining, можно отметить, что они сводятся к распределению объектов по классам, установлению зависимостей непрерывных переменных от входных (регрессия), выявлению закономерностей между связанными объектами (ассоциация), установлению отношений порядка между исследуемыми наборами данных (сиквенциальный анализ), выявлению кластеров и распределению по ним объектов (кластеризация), установлению закономерностей между связанными во времени событиями (прогнозирование), выявлению нехарактерных шаблонов (анализ отклонений).

Для ряда задач может быть достаточно аналитических средств, доступных в СУБД, но для более сложных процедур по автоматизации анализа данных следует использовать инструменты DataMining.

При создании системы анализа данных необходимо разработать удобный интерфейс, который помог бы ответить на следующие основные вопросы:

- Какое именно приложение будет реализовано: desktop или web?
- Какой язык программирования использовать?
- Какую лицензию использовать?
- Какая платформа?

На базе выбранных инструментов (СУБД, ETL, DataMining) осуществляется реализация хранилища и построенных в нем моделей представления данных (на базе технологий OLAP или OLTP).

Что касается реализации алгоритмов сжатия и восстановления, то в этих случаях могут использоваться готовые программные продукты (WinRAR, 7-Zip, FreeArc, WinZip и многие другие). Выбор средств должен определяться эффективностью сжатия и удобством интеграции средства в хранилище.

На рисунке 5 показана схема процесса эффективного хранения анализируемых данных.

Интерфейсы для клиентской части, предоставляющие доступ к функциям хранилища администратору и аналитику, должны реализовывать следующие функции:

- ведение регламентов;
- восстановление данных за указанные периоды;
- сжатие данных;
- просмотр данных и агрегатов за указанные периоды;

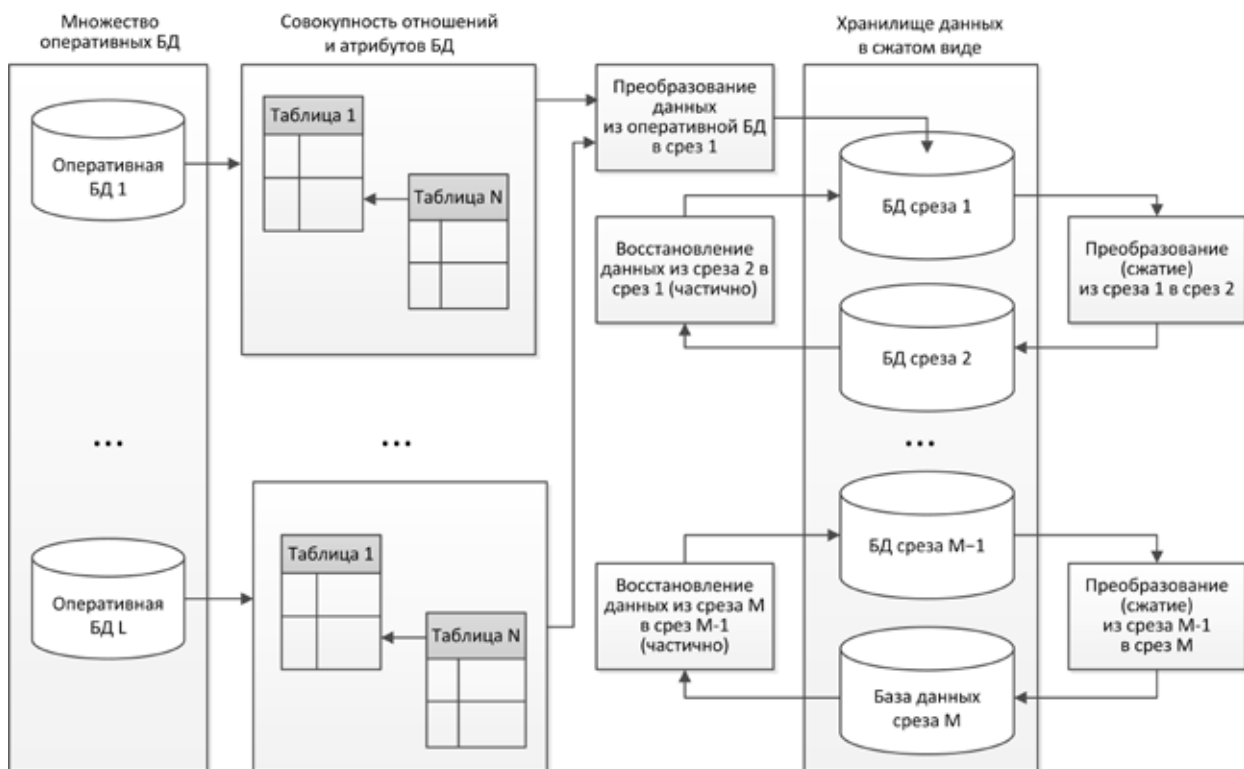


Рис. 5. Схема организации процесса эффективного хранения данных

- сравнение данных за разные периоды;
- определение критериев нештатных ситуаций.

На рисунках 6 и 7 показаны примеры компонентов интерфейсной части, выполненные в виде страниц web-приложения.

В заключительной части предложенной методики выполняются этапы тестирования и внедрения, структура и содержание которых определяются совместно разработчиком и пользователем и прогонкой необходимого минимального объема всех разнообразных данных, к которым планируется применить построенную систему.

ЗАКЛЮЧЕНИЕ

В данной статье была в общем виде рассмотрена методика применения моделей и алгоритмов компактного хранения и аналитической обработки данных, на основании которой можно построить достаточно эффективное программное средство для хранения и аналитической обработки данных. Описанная методика может быть применена в первую очередь для построения информационных систем, предназначенных для сбора, хранения и анализа статистических данных. Например, это могут быть контрольные данные о работе программных и аппаратных

средств, разные статистические сведения о населении, данные о производстве продукции и т. д.

Предложенные средства использованы в решении практической задачи построения хранилища данных для их анализа надежности программных и аппаратных средств.

Централизованный сбор в едином хранилище таких данных, полученных в различные периоды времени, позволяет не только компактно хранить их, но и использовать для разностороннего анализа.

В качестве средств реализации выбрана СУ БД PostgreSQL, средства ETL (решено не применять средства DataMining), программные коды компонентов системы реализуются на языках программирования Java и JavaScript.

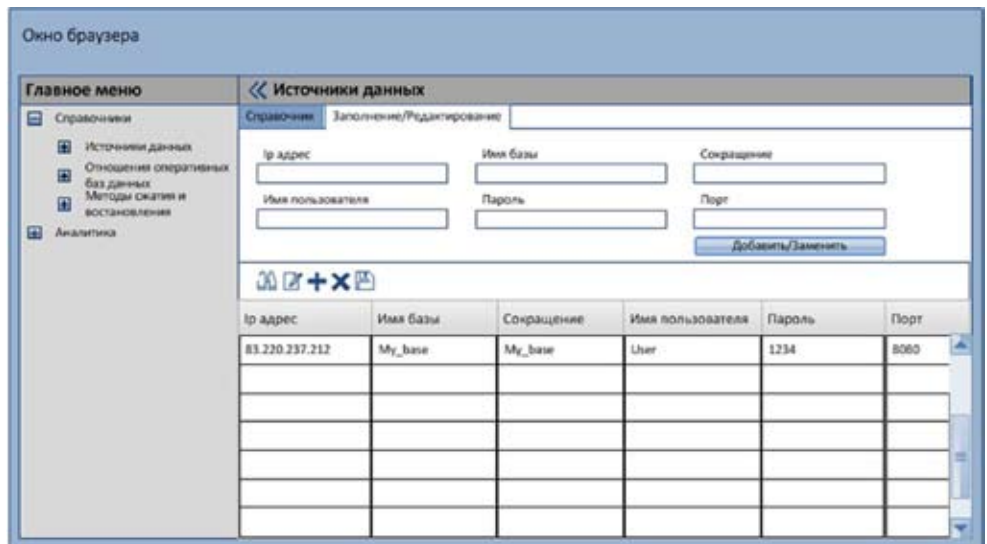


Рис. 6. Поля заполнения данных о временных срезах

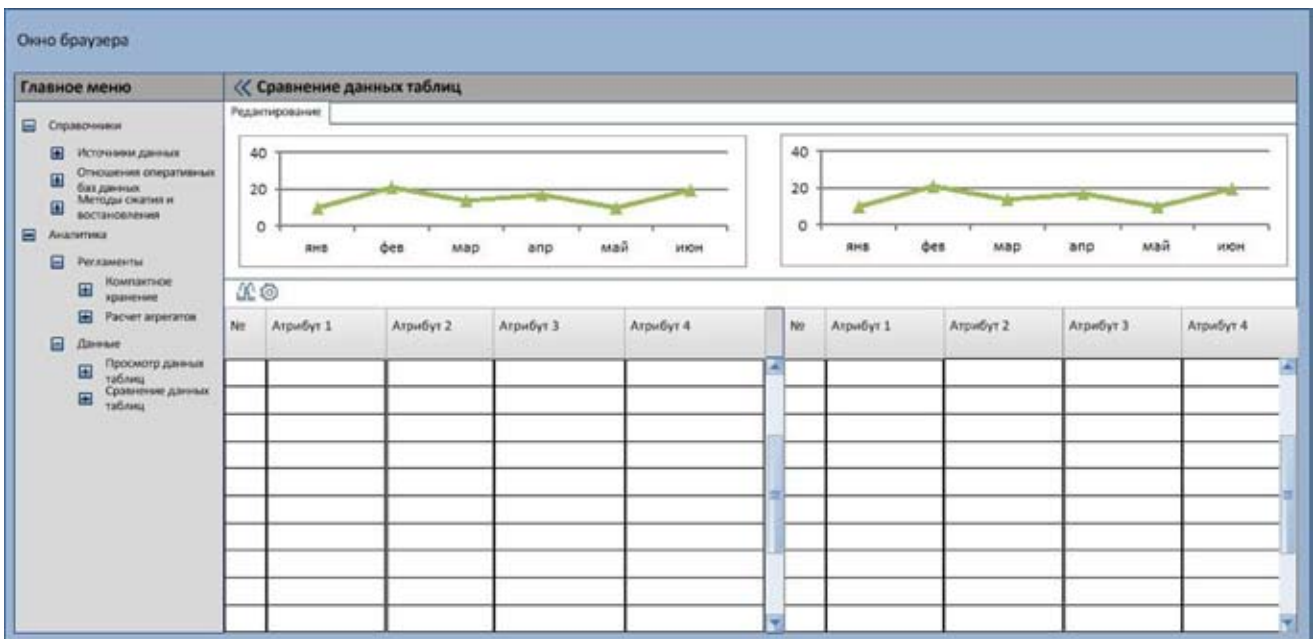


Рис. 7. Форма сравнения данных

СПИСОК ЛИТЕРАТУРЫ

1. Центр обработки данных (дата-центр). – URL: [http://tadviser.ru/index.php/Статья:Центр_обработки_данных_\(дата-центр\)#.D0.97.D0.B0.D0.B4.D0.B0.D1.87.D0.B8/](http://tadviser.ru/index.php/Статья:Центр_обработки_данных_(дата-центр)#.D0.97.D0.B0.D0.B4.D0.B0.D1.87.D0.B8/) (дата обращения: 10.07.2016).
2. Стоцкий М.В. Подсистема хранения данных мониторинга высокопроизводительных вычислительных систем // Информатизация и связь. – 2012. – № 9. – С. 42–46.
3. Тютляева Е.О., Московский А.А. Анализ основных тенденций в области хранения данных // Информационные технологии и вычислительные системы. – 2012. – № 2. – С. 64–75.
4. Фишер А.В., Дьяченко Р.А., Лоба И.С. Организация хранения хронологических данных в базах данных систем мониторинга и прогнозирования // Политематический сетевой научный журнал Кубанского государственного аграрного университета. – 2012. – № 79(5). – С. 271–280.
5. Шестюк Ю.М., Воронков К.Л. Обработка и хранение первичных данных технологического мониторинга // Вопросы радиоэлектроники. – 2012. – № 2. – С. 77–84.
6. Агрегирование // Академик. – URL: http://technical_translator_dictionary.academic.ru/3331/агрегирование (дата обращения: 30.08.2016).
7. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин – М. : Диалог-МИФИ, 2002. – 384 с.
8. Вычисление // Академик. – URL: http://polytechnic_dictionary.academic.ru/375 (дата обращения: 30.08.2016).
9. Сэлмон Д. Сжатие данных, изображений и звука. – М. : Техносфера, 2004. – 368 с.
10. Лоран П.Ж. Аппроксимация и оптимизация. – М. : Мир, 1975. – 496 с.
11. Преобразование данных в ETL. – URL: <http://bourabai.ru/tpoi/olap01-12.htm> (дата обращения: 30.08.2016).

REFERENCES

1. *Tsentr obrabotki dannykh (data-tsentr)* [Data Processing Center (Data-Center)]. Available at: http://tadviser.ru/index.php/Статья:Центр_обработки_данных_

(дата-центр)#.D0.97.D0.B0.D0.B4.D0.B0.D1.87.D0.B8/ (accessed: 10.07.2016).

2. Stotsky M.V. Podsystemakhraneniia dannykh monitoring vysokoproizvoditelnykh vychislitelnykh system [Subsystem of the Monitoring Data Storage from High Performance Computers]. *Informatizatsiia i sviaz* [Informatization and Communication], 2012, no. 9, pp. 42–46.

3. Tiutliaev E.O., Moskovskii A.A. Analiz osnovnykh tendentsii v oblasti khraneniia dannykh [Challenges and Opportunities in Data Storage]. *Informatsionnye tekhnologii i vychislitelnye sistemy* [Information Technologies and Computer Systems], 2012, no. 2, pp. 64–75.

4. Fisher A.V., Diachenko R.A. Loba I.S. Organizatsiia khraneniia khronologicheskikh dannykh v bazakh dannykh system monitoringa i prognozirovaniia [Storage Organization of Chronological Data in Databases of Monitoring and Forecasting Systems]. *Politematicheskii setevoi nauchnyi zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta* [Polythematic Online Scientific Journal of Kuban State Agrarian University], 2012, no. 79(5), pp. 271–280.

5. Shestiuk Iu.M., Voronkov K.L. Obrabotka i khraneniie pervichnykh dannykh tekhnologicheskogo monitoringa [Processing and Storage of Primary Data Technology Monitoring]. *Voprosy radioelektroniki* [Questions of Radio-Electronics], 2012, no. 2, pp. 77–84.

6. Agregirovanie [Aggregation]. *Akademik* [Academician]. Available at: http://technical_translator_dictionary.academic.ru/3331/агрегирование (accessed: 30.08.2016).

7. Vatoлин D., Ratushniak A., Smirnov M., Iukin V. *Metody szhatiia dannykh. Ustroistvo arkhivatorov, szhatie izobrazhenii i video* [Data Compression Methods. Archive Program Configuration, Compression of Images and Videos]. Moscow, Dialog MIFI Publ. 2002. 384 p.

8. Vychislenie [Computing]. *Akademik* [Academician]. Available at: http://polytechnic_dictionary.academic.ru/375 (accessed: 30.08.2016).

9. Salomon D. *Szhatie dannykh, izobrazhenii i zvuka* [Data Compression]. Moscow, Tekhnosfera Publ., 2004. 368 p.

10. Loran P. Sh *Approksimatsiia i optimizatsiia* [Approximation and Optimization]. Moscow, Mir Publ., 1975. 496 p.

11. *Preobrazovanie dannykh v ETL* [Data Transformation in ETL]. Available at: <http://bourabai.ru/tpoi/olap01-12.htm> (accessed: 30.08.2016).