

Концепция управляющего хранилища

И.А. Степановская, Э.Г. Прохорова, Л.А. Сырых

Аннотация. Актуальной проблемой построения интеллектуальных систем управления является разработка аппарата многоуровневого иерархического принятия решений на основе нечетких множеств. Настоящая работа направлена на решение этой проблемы применительно к задачам системного интеграционного глобального мониторинга и анализа (СИГМА) крупномасштабных слабоформализуемых процессов. Примерами таких приложений являются мониторинги регионального, ведомственного и отраслевого масштаба. Предлагаются принципы и методы построения интеллектуальных СИГМА-приложений, ориентированных на задачи упреждающего мониторинга, сформулированные в виде концепции управляющего хранилища.

1. Постановка задачи

Неустанно расширяющаяся потребность в системном подходе к управлению в различных прикладных областях породила концепцию хранилища данных (ХД). Перспективность технологий ХД в системном управлении определяется их ориентацией на комплексное использование методов и средств, традиционно развиваемых в рамках независимых научно-практических направлений, включая интеграцию информации, управление базами данных, добычу знаний, многомерный анализ, нечеткую логику.

Высокий потенциал технологий ХД для информатизации обеспечил им использование в задачах контроля, анализа, прогнозирования и управления отраслевого, регионального и ведомственного масштаба. Примерами могут служить системы мониторинга здравоохранения, регионального социально-экономического состояния, градостроительной деятельности, природных ресурсов, экологического состояния, программно-целевого инвестирования, материально-технического состояния производственных мощностей и др.

Несмотря на значительные успехи, анализ показывает недостаточную эффективность практического внедрения технологии ХД. По сведениям специалистов около 60% проектов не дают ожидаемой результативности в части системного управления. Современное понима-

ние проблем системного управления связано с дальнейшим развитием и расширением концептуальных моделей ХД. Одним из перспективных направлений является ориентация на динамические стратегии выбора приоритетных объектов, задач, уровня и алгоритмов контроля, анализа и оперативного управления.

В рамках этой проблематики в работе предлагается концепция «управляющего хранилища» (controlling warehouse).

2. Анализ эффективности концепции хранилищ данных

Первоначальная концепция обработки информации в задачах системного управления представлена классической схемой сопровождения централизованного хранилища данных (Рис.1).

Она предполагает интеграцию данных из независимых источников информации, загрузку интегрированной информации в хранилище и проблемно-ориентированные витрины и массовую организацию взаимодополняющих контуров анализа и управления по разнообразным разрезам и уровням системного управления.

Активные разработки программного обеспечения классической схемы сопровождения хранилища привели к тому, что она приобрела черты комплексной «аналитической машины», реализующей мощные алгоритмы семантической

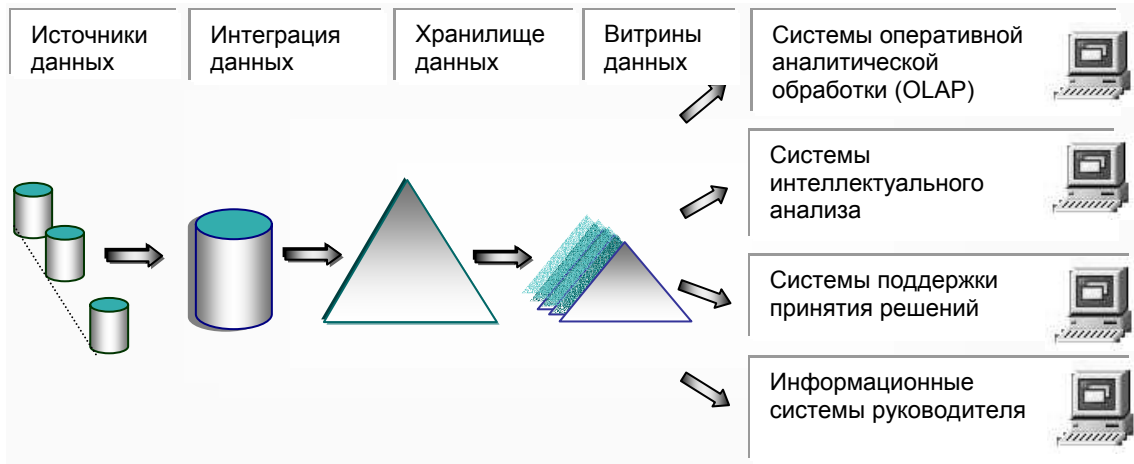


Рис. 1

обработки огромных массивов информации с использованием методов интеллектуальной обработки.

Примерами высокоразвитых программных комплексов сопровождения хранилищ данных могут служить разработки фирмы Sybase и компании Cognos. Система PowerDesigner фирмы Sybase обеспечивает исследование потоков данных (модуль Process Analyst), разработку концептуальной и физической моделей данных (модуль DataArchitect), создание объектов приложений (модуль AppModeler), исследование групповой работы, совместного доступа к информации и управления моделью (модуль MetaWorks), проектирование хранилищ данных и графического просмотра информации о модели (модули WarehouseArchitect и Viewer).

Системы Hupregion и PowerPlay компании Cognos представляют ставший типовым набор инструментальных средств аналитической обработки хранилища, используемый в задачах системного управления. Этот набор сочетает сервис гибкого доступа к информации с важнейшими инструментами факторного анализа (модуль Scenario) и нейросетевой технологии прогнозирования (модуль 4Thought).

Однако общепризнанным недостатком классической схемы является несоответствие высокого аналитического потенциала сравнительно низкому потенциалу в области интеграции данных. Суть проблемы в том, что классическая схема предусматривает простейшую стратегию интеграции, ориентированную на источники

однородной информации и чисто техническую технологию ETL (extraction, transformation, loading) извлечения данных, их приведения к одной измерительной шкале и точности, устранения избыточности, очистки, заполнения пропущенных значений, удаления ошибочных данных, фильтрации и преобразования к одному формату.

Стремление расширить возможности интеграции данных привело к развитию новой концепции сопровождения хранилища, предусматривающей интеграцию источников неоднородной информации (например, документальной, сигнальной и статистической), отвечающей информационным потребностям прикладных задач управления. Практические подходы к ее реализации связаны с отказом от принципа централизованного хранилища и развитием принципа федерализации источников, развитием новых моделей накопителей интегрированной информации – оперативного склада данных и виртуального хранилища. Модель оперативного склада данных ориентирована на управляющие приложения, в которых задачи оперативной обработки информации доминируют над задачами ее хранения. Модель виртуального хранилища направлена на включение интегрируемых источников в «единое информационное пространство». В обобщенном виде современная схема сопровождения хранилища представлена на Рис.2.

Примером практической реализации новых концепций может служить система управления холдингом. Она обеспечивает интеграцию

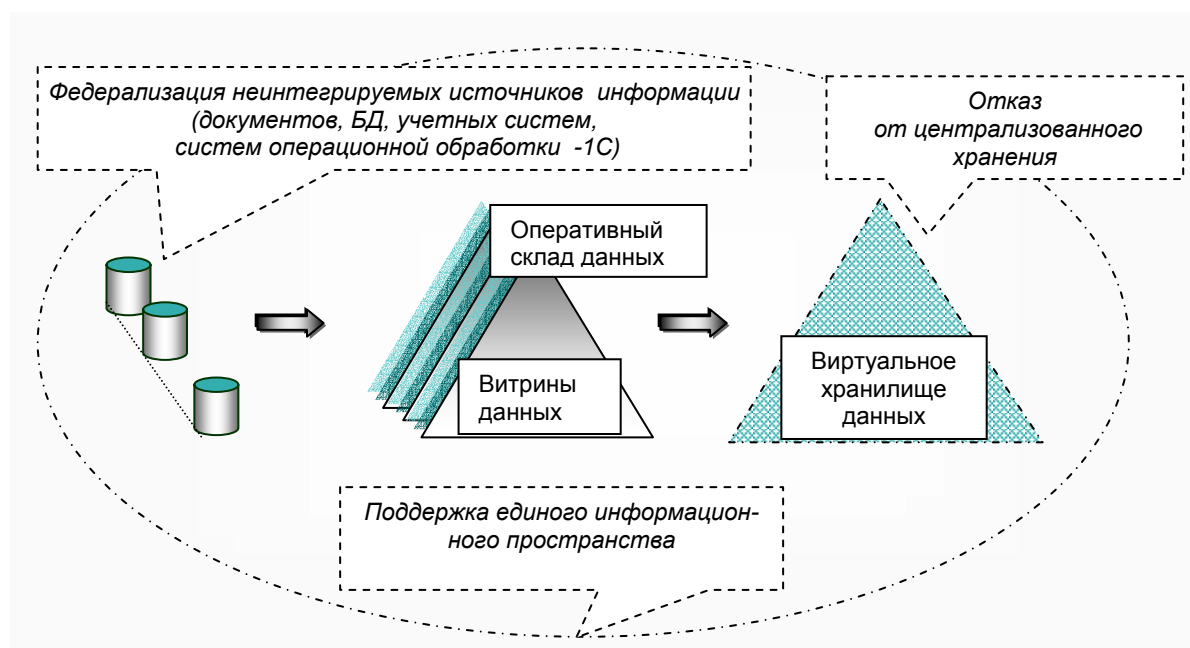


Рис.2

разнородных источников, включая учетные системы (1С, Ахapta, Navision, Парус, Галактика и др.), СУБД (Oracle, Sybase, MS SQL Server) и файловые системы формата DBF (Excel, XML), распределенные по филиалам компании. Принцип федерализации всех источников статистической и документальной отчетности позволяет организовать комплексную информационную систему, сохраняя имеющиеся в каждом из подразделений компании автономные системы обработки и хранения информации. Охват всех информационных источников в единый комплекс создает возможности для системного контроля, анализа, прогнозирования и управления бюджетом и финансовой политикой.

Приобретение аналитической машиной расширенных способностей к интеграции информации делают ее привлекательной для использования в автоматизированных контурах системного управления. Основным сдерживающим фактором является изначальное планирование хранилища «от источников данных», а не «от задач», при котором практически невозможно добиться адекватности информационной инфраструктуры требованиям замкнутых контуров управления.

В связи с этим актуальной проблемой является дальнейшее развитие концептуальных моделей ХД. С этой целью предлагается новая концепция управляющего ХД, настраиваемого на многоуровневую многообъектную модель контроля, анализа и управления множествами разнообразных объектов (регионы, отрасли, предприятия, здания, имущество, организационные системы и др.).

3. Основные положения концепции «управляющего хранилища»

Фундаментальное отличие предлагаемой концепции заключается в использовании методологического подхода системной интеграции, глобального мониторинга и анализа (СИГМА) данных о предметной области. СИГМА-приложения строятся на основе формального аппарата кластерно-ориентированного моделирования предметной области.

Данный аппарат поддерживает унифицированный подход как к разнообразным объектам прикладной области управления, так и по отношению к разным этапам решения задач системного управления, включая

- контролируемость,
- формирование концепции управляемости,

- реактивное управление,
- оценку качества управления и его модификацию.

Процесс кластерного моделирования предметной области включает построение лингвистической, структурной и функциональной моделей СИГМА-приложений.

3.1. Лингвистическая модель

Как известно, в отличие от привычной однородной шкалы измерительных осей, OLAP-куб использует иерархическую шкалу, предусматривающую применение иерархической системы группирующих множеств (Рис.3), задающих уровни обобщения группируемого множества.

Введем следующие типы кластеризаторов.

- Номенклатура как предметный кластеризатор потока элементов, т.е. бесконечного потока элементов, например, документов, сигналов и т.п.

- Предметный кластеризатор, представляющий этапы жизненного цикла объектов или процессов. Примерами могут служить этапы жизненного цикла образцов продукции (проектирование, производство, реализация), эксплуатируемых изделий (поступление, хранение, поверка, ремонт, снятие с эксплуатации, утилизация), процесса финансирования (предварительная заявка, план, выделение, реализация).

- Структурные кластеризаторы, отражающие такие категории отношений между объектами, как состав, декомпозиция, подчинение, принадлежность и др.

- Качественная кластеризация, представляющая динамические группировки объектов в зависимости от категории их состояния.

В процессе кластерного моделирования предметной области произвольное множество объектов Ω , включаемое в контур мониторинга, рассматривается в наборе

$$(\Omega, \{\Psi\}, K, X, Y),$$

где Ω – множество объектов,

$\{\Psi\}$ – множество кластеризаторов множества Ω (структурных, параметрических, пове-

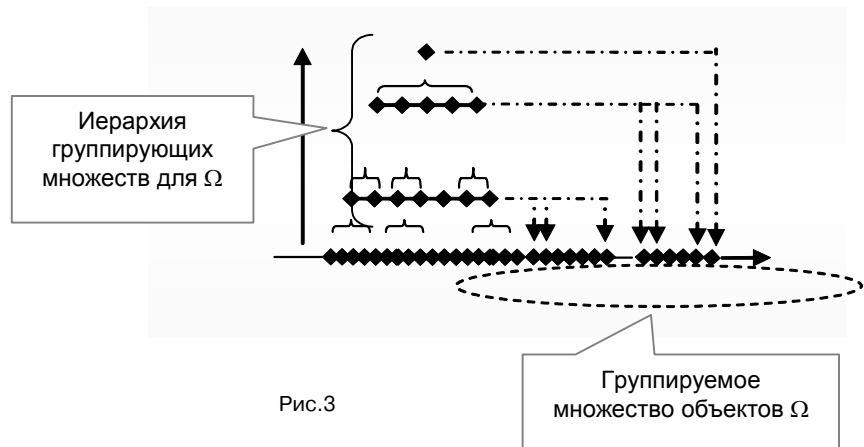


Рис.3

денческих, по связям и отношениям, по прогнозным или ретроспективным характеристикам и др.),

K – комплексный кластеризатор, построенный на множестве $\{\Psi\}$, определяющий семантику задачи контроля, анализа и управления элементами множества Ω ,

$X_i, i=1, n$ – многоуровневая система кластеров, определяемых кластеризатором K (имена всех подмножеств как элементов иерархического разбиения множества Ω),

$Y_i, i=1, n$ – планируемая многоуровневая управляющая система, представленная множеством управляющих воздействий, соответствующих каждому кластеру $X_i, I=1, n$.

Рассмотренный набор представляет стратегию управления применительно ко всем элементам множества Ω . По отношению к каждому элементу стратегия управления является нечеткой, поскольку зависит от того, к какому кластеру в какой момент времени он будет принадлежать.

Кластерно-ориентированная стратегия управления предполагает выбор управляемых множеств Ω и определение для каждого из них множества контролируемых кластеров. В результате в единый контур контроля и управления могут быть включены объекты разнообразной природы (регионы, отрасли, предприятия, здания, имущество, организационные системы и др.).

Преимуществами кластерно-ориентированной стратегии контроля, анализа и управления являются следующие.

- Отсутствие ограничений на тематическую направленность задач контроля, анализа и про-

гнозирования, практически неограниченное число объектов, включаемых в контур контроля.

- Гибкость, допускающая возможность смены представлений и поколений, перепрофилирования, переориентации, расширения задач.

- Открытость по отношению к привлекаемым источникам информации и методам ее анализа, определяющим возможность интеграции разноаспектной информации о наблюдаемой прикладной области.

- Наличие принципиальной возможности «целостного» контроля сильно связанных объектов и процессов с непредсказуемыми «системными» эффектами фрагментарных управляющих решений, для которых неадекватны методы декомпозиции на независимые подзадачи управления.

- Универсальность и возможность использования на разных этапах задач системного управления, включая моделирование, контроль, анализ, прогнозирование и оценку качества управления.

3.2. Структурная модель

Базовой единицей структурного моделирования является **OLAP-куб**.

Он определяется как декартово произведение мультикластеризуемых множеств, пересечение которых пусто $\{ \times \Omega_i, \cap \Omega_i = \emptyset, i \in I \}$. Произвольное множество Ω_i называется образующим i -ую ось, а множество меток этой оси представлено объединением образующего множества с одним из его кластеризаторов $\Omega_i \cup K, K \in \{ \Psi \}$.

В рамках понятийной модели **OLAP-кубы** настраиваются на систему обобщения и конфигурируются в OLAP-суперкубы соответственно семантике задач. В целом OLAP-суперкубы обеспечивают разноаспектное представление объектов и процессов.

3.3. Функциональная модель

OLAP-суперкуб допускает реализацию **итерационных запросов**, отражающих потребности системы аргументации и разработки сценариев принятия решений, когда при получении результатов запросов или отчетов возникает потребность в дополнительных вопросах. В рамках единой понятийной модели формирова-

ние итерационных запросов производится в терминологии, представленной лингвистической моделью информационно-управляющего процесса. Итерационные запросы составляют модель проектируемой автоматизированной обработки информации в интеллектуальных контурах управления. Предполагается, что итерационные запросы поддерживают задачи управления, сформулированные на обобщенном семантическом уровне, и обеспечивают глубокую детализацию ответов на последовательность уточняющих вопросов о том, какие объекты, явления или процессы являются приоритетными для контроля и каковы подходы к управлению ими.

Представленная понятийная модель является специализированной системой OLAP-технологии, а потому обладает возможностями ROLAP- или MOLAP-реализации [1].

Для создания, отладки, испытаний, сравнительного анализа качества и построения пилотных образцов проектируемых информационно-управляющих процессов предпочтительными являются ROLAP- или дешевые MOLAP-реализации. В режиме непосредственного контроля, анализа и управления допустимы дорогие MOLAP-реализации, учитывающие требования к аппаратно-программной платформе эксплуатируемых СИГМА-приложений.

Заключение

Преимущества и новые возможности концепции «управляющего хранилища»

Перспективы «управляющего хранилища» в значительной мере связаны с решением актуальных проблем развития нового направления интеллектуальных СИГМА-приложений, поддерживающих решение плохо формализованных информационно-управляющих задач. Основные положения и преимущества концепции управляющего ХД были продемонстрированы в ходе реализации ряда пилотных проектов СИГМА-приложений. Одним из наиболее показательных является мониторинг финансовых потоков, возникающих в ходе выполнения Федеральных целевых программ. Примерами могут служить задачи типа: «В чем суть проблемы недофинансирования», «Как можно оценить

качество структуры бюджета, каковы направления его совершенствования» в мониторингах финансовых потоков регионального, отраслевого и ведомственного уровней, или: «Какие могут быть даны рекомендации по развитию предприятия или группы предприятий в контексте отраслевой политики на реструктуризацию и организацию ассоциаций по производству перспективных образцов», или: «Как можно оценить организационную структуру предприятия, каковы проблемы, узкие места» в системе мониторинга ведомственного уровня.

Принципиально новыми являются также возможности для организации внедрения понятийных знаний на уровне настраиваемых на модель сервисов имитационного и анимационного моделирования, а также для использования методов проблемно-ориентированной добычи знаний.

Преимуществом концепции «управляющего хранилища» является поддержка развития новых концепций внедрения ХД и их использования наряду с классическими.

Наконец, концепция «управляющего хранилища» открывает новые возможности для разработки программного обеспечения СИГМА-приложений в рамках единой программной инженерии, направленной на удовлетворение современным требованиям модульности, унификации, стандартизации, сертифицируемости.

Литература

1. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. Том.1: Пер. с англ. М.: Вильямс, 2001.

Степановская Ираида Александровна. Окончила МГПИ им. В.И.Ленина в 1969 г., МГУ им. М.В.Ломоносова в 1983 г. Кандидат технических наук. Имеет более 60 публикаций. Область научных интересов: математическое моделирование, проблемно-ориентированные системы, основанные на знаниях, и экспертные системы, системы и технологии создания и поддержки проблемно-ориентированных баз данных, инженерно-технические и информационные автоматизированные системы мониторинга биоресурсов, биосферы и технических систем, программное обеспечение параллельных и распределенных вычислительных систем, надежность и безопасность технических систем, искусственный интеллект и принятие решений. Ведущий научный сотрудник ИПУ РАН.

Прохорова Элла Григорьевна. Окончила ХПИ им. В.И.Ленина в 1959 г. Имеет 38 публикаций. Область научных интересов: моделирование, информационно-аналитические системы, диспетчеризация вычислительных ресурсов, экологический мониторинг, мониторинг финансовых потоков. Научный сотрудник ИПУ РАН.

Сырых Лора Алексеевна. Окончила МИРЭА в 1985 г. Область научных интересов: моделирование, экологический мониторинг, мониторинг финансовых потоков. Научный сотрудник ИПУ РАН.