

ИНФОРМАЦИОННОЕ ХРАНИЛИЩЕ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ ИСПЫТАНИЙ И ВЕРИФИКАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Ю. М. ШЕРСТЮК¹, А. А. МУСАЕВ²

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

¹<yusher@incomsys.ru>, ²<musayev@yandex.ru>

УДК 621.4

Шерстюк Ю. М., Мусаев А. А. Информационное хранилище в автоматизированной системе испытаний и верификации математических моделей // Труды СПИИРАН. Вып. 4. — СПб.: Наука, 2007.

Аннотация. Описывается подход к осуществлению автоматизированного хранения и использования опыта испытаний и отладки математических моделей как совокупности ретроспективных значений параметров, характеризующих функционирование изучаемой системы на определенном временном отрезке. Особое внимание уделено применению для этих целей информационных хранилищ и их отличий от традиционных операционных систем учета. Кратко изложены основные характеристики систем испытаний и отладки, создаваемых на основе технологий хранилищ данных (Data Warehouse), оперативной аналитической обработки данных (OLAP) и интеллектуального анализа данных (Data Mining). — Библиограф. 10 назв.

UDC 621.4

Sherstyuk Yu. M., Musayev A. A. Data Warehouses Application in the Automated Testing and Verification System for Mathematical Models // SPIIRAS Proceedings. Issue 4. — SPb.: Nauka, 2007.

Abstract. This article outlines an approach to the automated storing and using the results of testing and debugging of mathematical models which defined as a set of historical parameters values describing the state of the testing system in time periods. Particular attention is given to the data warehouses application for these purposes and to their difference from the traditional logging systems. There are also briefly explanation of the core features of the testing and debugging systems based on data warehouses, OLAP and data mining. — Bibl.10 items.

1. Введение

Одним из центральных этапов построения современных систем управления является разработка математических моделей (ММ), отражающих взаимосвязь входных и выходных параметров, а также параметров состояния объекта управления (ОУ). Формирование ММ может осуществляться априори, исходя из общих усредненных представлений о структуре и свойствах ОУ, или апостериори, базируясь на результатах мониторинга его функционирования. Однако во всех случаях ММ управления нуждаются в испытаниях, отладке и верификации, предусмотренных государственными стандартами на разработку автоматизированных систем (АС).

Практический опыт испытаний и отладки математических моделей как совокупность знаний и практически усвоенных навыков и умений играет важнейшую роль в процессах создания современных систем управления. При этом корпоративный опыт проведения испытаний и отладки ММ не сводим к сумме опыта отдельных разработчиков (хотя и определяется им), т.е. обладает свойством эмергентности, характерным для сложных систем [1]. Изначально носителем любого, в т. ч. экспериментального, опыта является человеческая память со всеми присущими ей достоинствами и недостатками. Стремление к ав-

томатизации накопления и хранения опыта как составной части автоматизации процессов управления является естественным следствием появления вычислительной техники и развития соответствующих информационных технологий. Однако хранение любой информации в виде совокупности данных в ЭВМ предполагает обязательную формализацию этой информации (пусть даже в виде, например, текстов на естественном языке, которые рассматриваются как последовательности символов и по отношению к которым можно выполнять ряд формальных операций поиска, редактирования и т. д.). В этом плане опыт по испытанию и отладке ММ в его изначальном понимании объектом автоматизированного хранения быть не может, ибо он потенциально не формализуем в частности его составляющие, тесно связанные с интуицией, эмоциональной окраской оценки событий и т. д. С учетом вышеизложенного хранение опыта (здесь и далее рассматривается только автоматизированное хранение, т. е. хранение с применением средств вычислительной техники) рассматривается далее в более узком смысле, который можно определить исходя из методологии системного анализа применительно к задачам управления.

2. Задача мониторинга объектов моделирования

Объект моделирования (ОМ) может рассматриваться как динамичная совокупность компонентов, в качестве которых выступают различные подсистемы и элементы, выделяемые по различным признакам (организационным, функциональным, техническим) со множественными связями различной природы. Каждый компонент ОМ характеризуется рядом параметров, значения которых изменяются в результате происходящих событий (поскольку наблюдение непрерывных процессов также носит дискретный характер, измерение величины параметра, характеризующего непрерывный процесс, также можно рассматривать в качестве события). Таким образом, в любой произвольный момент времени состояние ОМ S характеризуется совокупностью $P_S = \langle X(t), Z(t) \rangle$, где $X(t)$ — множество в общем случае взаимосвязанных параметров, причем от времени зависят как значения этих параметров, так и их состав, а $Z(t)$ — множество значений этих параметров. Изменение как $Z(t)$, так и $X(t)$ происходит событийно, то есть

$$(X(t_{i+1}), Z(t_{i+1})) = F((X(t_i), Z(t_i)), A(t_{i+1})),$$

где $X(t_{i+1}), Z(t_{i+1})$ — значения $X(t_i), Z(t_i)$ после свершения $(i+1)$ -го события; t_i, t_{i+1} — времена свершения двух последовательных $(i$ - и $(i+1)$ -го) событий; $A(t_{i+1})$ — атрибуты $(i+1)$ -го события; F — функциональная зависимость характеристик ПСП от происходящих событий.

Поскольку ОМ, как правило, представляет собой систему с управлением, одна часть параметров $X(t)$ (соответственно и значений $Z(t)$) является параметрами управления ($X_C(t)$), а другая — параметрами состояния ($X_S(t)$).

При наличии автоматизированных систем мониторинга, осуществляющих сбор (автоматический или через «ручной ввод») и хранение данных, характеризующих ОМ, часть $P_S (P_S' = \langle X'(y), Z'(t) \rangle \subset P_S = \langle X(y), Z(t) \rangle)$ имеет автоматизированное хранение — как правило, в базах данных (БД) автоматизированных

систем (АС). Тем самым на ОМ существуют информационные отображения $R: S \rightarrow P_S$ и $R': R: S \rightarrow P_S'$.

Оставив за рамками рассмотрения целый ряд вопросов реализации этих отображений (шум квантования по значениям и по времени, дублирование хранения значений, целостность и непротиворечивость данных и т. д.), с некоторыми допущениями можно считать, что базы данных автоматизированных систем ОМ содержат информационную модель P_S' , отображающую его состояние со степенью полноты отображения P_S' / P_S .

Следует отметить, что абсолютное большинство баз данных АС ориентировано на информационное отображение текущего состояния ОМ, т.е. $P_S'(t)$ на момент времени t . Изменение содержимого этих БД происходит событийно, благодаря чему в мировой практике подобные БД и использующие их АС получили название транзакционных (OLTP — On-Line Transaction Processing) [2]. В то же время в зависимости от реализации и характера решаемых задач в БД OLTP-систем в момент времени t могут храниться не только $Z(t)$, но и значения в предшествующие моменты времени за некий интервал хранения T : для j -й БД можно записать, что ее содержимое (здесь и далее — только в части характеристик ПСП) на момент t есть $B_j = (X_j'(t), Z_j'(t))[t - T_j, t]$.

С учетом вышеизложенного опыт, хранимый информационными системами с использованием средств вычислительной техники, — это $O = B_1 \cup B_2 \cup \dots \cup B_n$, где n — количество БД, т.е. это хранимая история поведения ОМ. Использование опыта, хранимого подобным образом, может осуществляться двумя способами:

- при решении задач проверки адекватности ММ, по своей сути требующих обязательного учета функционирования ОУ за некоторый период, — например составление разного рода отчетных документов;
- в рамках поддержки принятия решений по испытаниям и собственно управлению на основе аналитических технологий.

Первый способ является традиционным, он реализуется *информационными системами (ИС)*, предназначенными для решения соответствующих задач. Второй способ получает свое распространение в основном в последнее десятилетие, что связано с развитием аналитических технологий — информационных технологий, ориентированных на задачи прогнозирования состояния сложных динамических систем в нестационарных и неоднородных средах, разработку сценариев развития ситуаций в условиях комплексной динамической неопределенности, ситуационный анализ текущей обстановки и т. п. [2–5].

3. Применение аналитических технологий

Постоянно возрастающий интерес к аналитическим технологиям, рассматриваемым в контексте систем управления, обусловлен прежде всего тем, что важнейшей составной частью любого процесса управления являются действия по сбору и анализу информации, содержание которых составляет оценка текущей ситуации и прогноз ее развития.

Под ситуацией в данном случае понимается совокупность характеристик ОУ и среды его функционирования в некоторый момент времени. Оценка си-

туации подразумевает вычисление значений этих количественных и качественных характеристик (как правило, с их объединением и обобщением) и их критериальное соотнесение с требованиями.

Аналитические технологии ориентированы на решение задач поддержки принятия решений. К таким задачам относятся:

- оценка текущего и прогнозируемого состояния ОУ и (или) среды его функционирования;
- обнаружение и исследование скрытых закономерностей, факторов, тенденций и взаимосвязей;
- обобщение информации как агрегация и интеграция сведений различного характера;
- формирование альтернативных решений и выбор «оптимального» в соответствии с заданным критерием, а также результатами анализа сценариев развития ситуаций;
- моделирование процесса эволюции состояния объекта в нестационарной неоднородной среде и т. д.

В свою очередь повышение качества процессов обобщения и анализа информации на основе аналитических технологий и их реализующих современных средств достигается автоматизированной реализацией процессов выявления скрытых факторов влияния, их идентификации, использования опыта на основе автоматизированного формирования и поиска прецедентов анализируемых ситуаций в массивах ретроспективных данных, высокодостоверного прогнозирования эволюции состояния объекта исследования, в т. ч. выявления предпосылок к скачкообразному изменению этого состояния.

Перечисленные ключевые возможности аналитических технологий, реализуемых посредством *аналитических информационных систем (АИС)*, позволяют рассматривать их как аналитическую и информационную основу не только процессов обобщения и анализа информации о текущем и прогнозируемом состоянии объекта управления и среды его функционирования, но и процессов испытаний и верификации ММ, используемых для построения соответствующих систем управления.

Данный аспект применения аналитических технологий осуществляется посредством использования:

- результатов прогнозирования развития процессов и явлений для оценки эффективности возможных (в т. ч. альтернативных) вариантов решений, планов и программ испытаний и верификации;
- способов ретроспективного анализа содержимого хранилищ данных для поиска типовых ситуаций, требующих управления, и отладки вариантов решений на основе аналогий и накопленного опыта управления;
- специально разработанного комплекса испытаний и верификации управленческих решений, основанных на формировании виртуальных сценариев развития частично управляемых ситуаций.

Таким образом, в основе инструментария аналитических технологий лежит применение методов анализа и прогнозирования на основе ретроспективных данных, т. е. истории функционирования ОМ. Тем самым наличие накопленного опыта превращается в необходимое условие применения аналитических технологий и реализующих их АИС, причем от объема и качества этого опыта во многом зависит успех указанного применения.

Данное положение заставляет по-новому оценить ситуацию с хранением опыта в транзакционных базах данных. С позиций накопления опыта функцио-

нирования ОМ в интересах его последующего использования в интересах поддержки принятия решений в АИС его хранение в только в транзакционных системах имеет следующие очевидные недостатки:

- распределенное хранение в нескольких, как правило, разнородных БД существенно затрудняет использование этих данных из-за сложности их одновременной выборки из нескольких источников, возможных пересечений перечней хранимых данных и (или) нарушений их согласованности (целостности);
- запросы на выборку ретроспективных данных снижают реактивность транзакционных систем;
- глубина ретроспективного анализа определяется алгоритмами работы транзакционных систем и в общем случае равна $\min\{T_i\}$.

4. Концепция информационных хранилищ

В силу указанных недостатков, а также принципиальных различий между системами OLTP и АИС (табл. 1) данные транзакционных систем практически бесполезны для непосредственного применения в АИС. Таким образом, складывается парадоксальная ситуация — «отсутствие информации при ее наличии или даже избытке». В результате этого с начала 90-х годов стала складываться концепция информационных хранилищ и соответствующая ей технология хранилищ данных (Data Warehouse, DW), в настоящее время подкрепленная довольно большим множеством средств, как правило, входящих в состав систем управления базами данных. В числе основных производителей средств построения DW выступают такие известные фирмы, как Oracle, SAS Institute, Microsoft и IBM (Informix) [6, 7].

Таблица 1

Сравнительный анализ характеристик данных в транзакционных и аналитических системах

Признак	Транзакционные ИС	Аналитические ИС
Источники данных	Внутренние	Внутренние и внешние, гетерогенные
Уровень агрегации данных	Детализированные данные	В основном агрегированные
Возраст данных	Оперативные и исторические за небольшой период времени	Исторические, за длительные временные периоды
Частота обновления и объем обновляемых данных	Высокие; добавление данных небольшими порциями	Низкая, загрузка данных большими объемами после предварительной обработки. Ранее загруженные данные не изменяются

Информационное хранилище (ИХ) представляет собой базу данных определенной структуры, содержащую информацию о производственном процессе компании в историческом контексте. Главное назначение хранилища — обеспечивать быстрое выполнение произвольных аналитических запросов.

К числу главных преимуществ хранилищ данных относят следующие [8].

Единый источник информации: компания получает выверенную единую информационную среду, на которой будут строиться все справочно-аналитические приложения в той предметной области, по которой построено хранилище. Эта среда будет обладать единым интерфейсом, унифицирован-

ными структурами хранения, общими справочниками и другими корпоративными стандартами, что облегчает создание и поддержку аналитических систем.

Производительность: физические структуры хранилища специальным образом оптимизированы для выполнения абсолютно произвольных выборок, что позволяет строить действительно быстрые системы запросов.

Интегрированность: интеграция данных из разных источников уже сделана, поэтому не надо каждый раз производить соединение данных для запросов, требующих информацию из нескольких источников.

Историчность и стабильность: информационное хранилище данных нацелено на долговременное хранение информации в течении 10–15 лет с адаптацией хранимой информации к изменениям структуры и параметров, происходящих в отображаемом объекте. Благодаря этому появляется возможность осуществлять исторический анализ информации.

Независимость: разделение информационного хранилища и OLTP-систем существенно снижает нагрузку на последние со стороны аналитических приложений; тем самым производительность существующих систем не ухудшается, а на практике происходит уменьшение времени отклика и улучшение доступности систем.

На рис. 1 приведена типовая обобщенная схема компонентов ИХ как совокупности информационных массивов и средств, обеспечивающих заполнение (ведение) и применение (использование) ИХ совместно со средствами OLAP (On-Line Analyzis Processing) и DM (Data Mining — интеллектуальный анализ данных) [9].

Можно выделить следующие три ключевых момента приведенной схемы:

1. Первичные данные из транзакционных систем с некоторой периодичностью (например, раз в сутки) загружаются в ИХ. При загрузке происходит проверка данных на непротиворечивость и логическую целостность, по результатам которой может проводиться ручная или автоматическая корректировка данных, а также фильтрация данных.

2. В общем случае хранению в ИХ подлежат:

исторические данные — важнейшими свойствами данных в аналитических задачах является их исторический характер и обязательная спецификация времени, которому эти данные соответствуют;

агрегированные данные — чем выше должностной уровень руководителя, управляющего, аналитика, тем выше уровень агрегации данных, используемых им для принятия решения;

прогнозируемые данные — наряду с данными, отражающими свершившиеся события, в аналитических системах предусматриваются структуры, ориентированные на получение прогнозов поведения различных параметров системы.

3. Данные, находящиеся в ИХ, используются средствами OLAP и DM.

Таким образом, информационное хранилище рассматривается в настоящее время как обязательный компонент АИС, с создания которого и начинается внедрение аналитических технологий как в систему испытаний ММ, так и в саму систему управления.

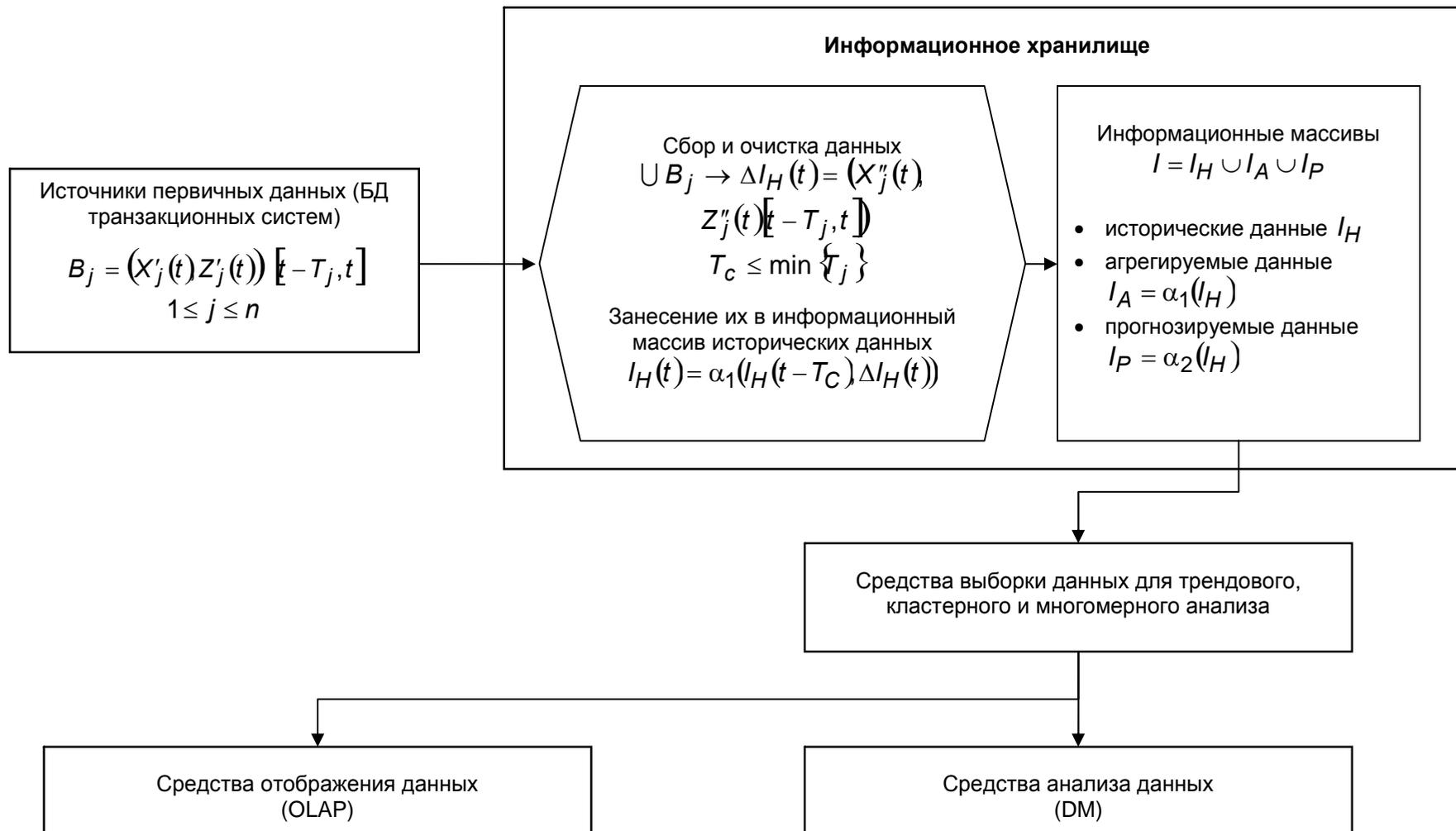


Рис. 1. Типовая обобщенная схема компонентов информационного хранилища.

5. Заключение

В заключение следует отметить, что создание ИХ, отражающих опыта функционирования ОУ, позволяет последовательно совершенствовать степень адекватности соответствующих ММ, и, как следствие, повысить качество процессов автоматизированного управления. Накопление опыта в форме развивающихся знаний создает реальные предпосылки для более глубокой степени автоматизации процессов отладки и верификации ММ, позволяющей перейти к новому поколению интеллектуальных СУ.

Литература

- [1] *Морозов В. П., Дымарский Я. С.* Элементы теории управления ГАП: Математическое обеспечение. Л.: Машиностроение, 1984. 384 с.
- [2] *Кречетов Н., Иванов П.* Продукты для интеллектуального анализа данных // *Computer Week*. 1997. № 14–15. С. 32–39.
- [3] *Бирюков А.* Системы принятия решений и хранилища данных // *СУБД*. 1997. № 4. С. 37–41.
- [4] *Львов В.* Создание систем поддержки принятия решений на основе хранилищ данных // *СУБД*. 1997. № 3. С. 30–40.
- [5] *Карпов Е. А., Мусаев А. А., Шерстюк Ю. М.* Многоцелевая аналитическая информационная система. Методология создания и основные проектные решения. СПб.: ВУС, 2000. 143 с.
- [6] *Коровкин С. Д., Левенец И. А., Ратманова И. Д., Старых В. А., Щавелев Л. В.* Решение проблемы комплексного оперативного анализа информации хранилищ данных // *СУБД*. 1997. № 5–6. С. 47–51.
- [7] *Бриттов П. А., Липчинский Е. А.* Практика построения хранилищ данных: Система SAS // *СУБД*. 1998. № 4–5. С. 14–20.
- [8] ТехноСерв А/С. Хранилище данных [Электронный ресурс] // <<http://technoserv.ru>> (по состоянию на 10.03.2007)
- [9] *DataMining with Microsoft SQL Server 2000. Technical Reference.* Washington. Microsoft Press, 2001. 368 p.