

ИНФОРМАЦИОННОЕ ХРАНИЛИЩЕ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ: КОНЦЕПЦИЯ СИСТЕМЫ ХРАНЕНИЯ ПРОИЗВОДСТВЕННОГО ОПЫТА

А. А. Мусаев, Ю. М. Шерстюк

Описывается современный подход к осуществлению автоматизированного хранения и использования производственного опыта как совокупности ретроспективных значений параметров, характеризующих функционирование производственной системы предприятия на определенном временном отрезке. Особое внимание уделено отличиям информационных хранилищ от традиционных операционных систем учета. Кратко изложены основные характеристики аналитических информационных систем, создаваемых на основе технологий *хранилищ данных* (Data Warehouse), *оперативной аналитической обработки данных* (OLAP) и интеллектуального анализа данных (Data Mining).

Являясь компонентом производительных сил (наряду с технологией, оборудованием и т. д.), производственный опыт как совокупность знаний и практически усвоенных навыков, умений производственной деятельности играет важнейшую роль в функционировании промышленного предприятия. Существенно, что корпоративный производственный опыт не сводится к сумме опыта отдельных индивидуумов (хотя и определяется им), т. е. обладает свойством эмергентности, характерным для сложных систем [1].

Носителем любого, в т. ч. производственного, опыта изначально является человеческая память – со всеми присущими ей достоинствами и недостатками. Стремление к автоматизации накопления и хранения производственного опыта как составной части автоматизации производства является естественным следствием появления вычислительной техники и развития соответствующих информационных технологий. Однако хранение любой информации в виде совокупности данных в памяти ЭВМ предполагает обязательную формализацию соответствующей информации (например, в виде текстов на естественном языке, которые рассматриваются как последовательности символов, по отношению к которым можно выполнять ряд формальных операций поиска, редактирования и т. д.). В этом плане производственный опыт в его изначальном понимании объектом автоматизированного хранения быть не может, поскольку он потенциально не формализуем. В частности, его составляющие, оказываются тесно связанными с интуицией, эмоциональной окраской оценки событий и т. д.

С учетом вышеизложенного, хранение производственного опыта (здесь и далее рассматривается только автоматизированное хранение, т. е. хранение с применением

средств вычислительной техники) далее рассматривается в более узком смысле, который можно определить исходя из методологии системного анализа применительно к управлению производственной системой предприятия.

Производственная система предприятия (ПСП) может рассматриваться как динамичная совокупность компонентов, в качестве которых выступают различные подсистемы и элементы, выделяемые по различным признакам (организационным, функциональным, техническим) с множественными связями различной природы (материальной, информационной, энергетической) как между собой, так и с элементами внешней по отношению к ПСП среды (среды погружения ПСП). Каждый компонент ПСП 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 ПСП 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 $Z(t)$ и $X(t)$ происходит событийно, т. е.

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

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) -го и $(i+1)$ -го событий;

$A(t_{i+1})$ – атрибуты $(i+1)$ -го события;

F – функциональная зависимость характеристик ПСП от происходящих событий.

Поскольку ПСП 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 ($X_S(t)$).

Благодаря наличию на предприятии разного рода автоматизированных систем, осуществляющих сбор (автоматический или через "ручной ввод") и хранение данных, характеризующих ПСП. Часть P_S ($P_S' = \langle X'(t), Z'(t) \rangle \subset P_S = \langle X(t), Z(t) \rangle$) имеет автоматизированное хранение – как правило, в *базах данных (БД) автоматизированных систем (АС)*. Тем самым на предприятии существуют информационные отображения $R: S \rightarrow P_S$ и $R': S \rightarrow P_S'$.

Оставив за рамками рассмотрения целый ряд вопросов реализации этих отображений (шум квантования по значениям и по времени, дублирование хранения значений, це-

лостность и непротиворечивость данных и т. д.), с некоторыми допущениями можно считать, что базы данных автоматизированных систем предприятия содержат информационную модель P_S' , отображающую ПСП со степенью полноты отображения P_S'/P_S .

Следует отметить, что абсолютное большинство баз данных АС предприятия независимо от принадлежности этих АС к *АСУ предприятием* (АСУП) или *АСУ технологическими процессами* (АСУ ТП) ориентированы на информационное отображение текущего состояния ПСП, 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 $V_j=(X'_j(t), Z'_j(t))[t-T_j, t]$.

С учетом вышеизложенного производственный опыт, хранимый на предприятии с использованием средств вычислительной техники – это $O = V_1 \cup V_2 \cup \dots \cup V_n$, где n – количество БД, т. е. это хранимая история поведения ПСП.

Таким образом, хранение производственного опыта имеет место на любом предприятии, на котором имеется хотя бы одна БД, для которой период хранения $T \neq 0$.

Использование производственного опыта, хранимого подобным образом, может осуществляться двумя способами:

при решении задач контроля и управления, по своей сути требующих обязательного учета функционирования объекта управления (наблюдения) за некоторый период (например, составление разного рода отчетных документов);

в рамках поддержки принятия решений по управлению ПСП и ее компонентами на основе аналитических технологий.

Первый способ является традиционным, он реализуется *автоматизированными системами* (АС), предназначенными для решения соответствующих задач.

Второй способ получил распространение в последнее десятилетие и связан с развитием *аналитических информационных технологий* (АИТ). К ним относится класс информационных технологий, ориентированных на задачи прогнозирования состояния сложных динамических систем в нестационарных и неоднородных средах, разработки сценариев развития ситуаций в условиях комплексной динамической неопределенности, ситуационного анализа текущей обстановки и т. п. [2-5].

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 составляют оценка текущей ситуации и прогноз ее развития.

Под ситуацией в данном случае понимается совокупность характеристик объекта управления и среды его функционирования в некоторый момент времени. Оценка ситуации подразумевает вычисление значений этих количественных и качественных характеристик (как правило, с их объединением и обобщением) и их критериальное соотнесение с требованиями.

Главным назначением АИТ является автоматизированное решение задач поддержки принятия управленческих решений. К таким задачам относятся:

- оценка текущего и прогнозируемого состояния объекта управления 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 закономерностей и факторов;

- количественной оценки факторов влияния и угроз в сложившейся ситуации;

- 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 ана-

лиза информации о текущем и прогнозируемом состояниях объекта управления и среды его функционирования, но и процессов подготовки (выработки и оценки) вариантов решений.

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 действий, а также управляющих воздействий иного рода;

способов ретроспективного анализа содержимого хранилищ данных для поиска типовых ситуаций, требующих управления, и выработки вариантов решений на основе аналогий и накопленного опыта управления;

специально разработанного комплекса анализа эффективности управленческих решений, основанных на формировании виртуальных сценариев развития частично управляемых ситуаций.

Таким образом, в основе инструментария АИТ лежит применение методов анализа и прогнозирования на основе ретроспективных данных, т. е. истории, опыта производства. Тем самым наличие производственного 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 производственного опыта в транзакционных базах данных. С позиций накопления производственного опыта 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 $\min \{T_i\}$.

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".

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	Высокая, добавление данных небольшими порциями	Низкая, загрузка данных большими объемами после предварительной обработки. Ранее загруженные данные не изменяются

В данной ситуации с начала 90-х годов стала складываться концепция информационных хранилищ и соответствующая ей технология хранилищ данных (DW – *Data Warehouse*), 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].

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 [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, унифицированными 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 приведен 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 – интеллектуального анализа данных) [9].

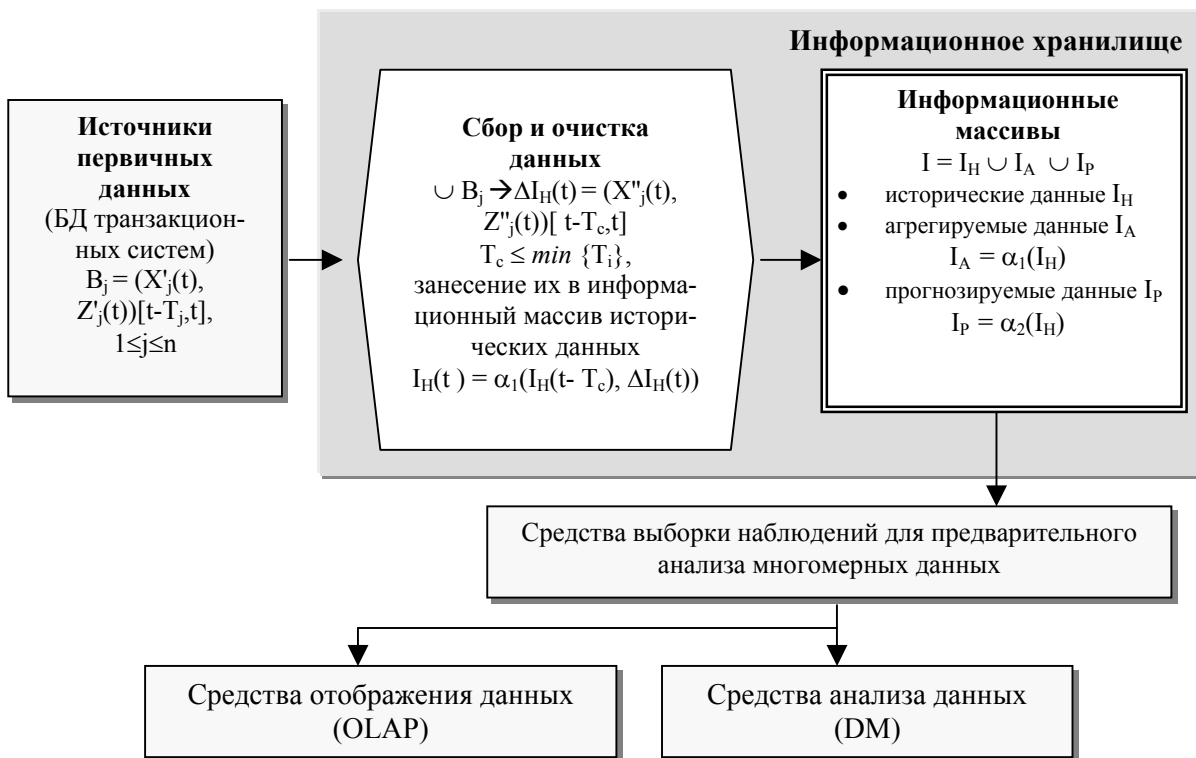


Рис. 1. Основные информационного хранилища

Можно выделить следующие три ключевые момента приведенной схемы:

1) Первичные данные из транзакционных систем с некоторой периодичностью (например, раз в сутки) загружаются в ИХ. При загрузке происходят:

проверка данных на непротиворечивость и логическую целостность, по результатам которой может проводиться (ручная или автоматическая корректировка данных);
фильтрация данных.

2) В общем случае хранению в ИХ подлежат:

исторические данные - важнейшим свойством данных в аналитических задачах является их исторический характер и обязательная спецификация времени, которому эти данные соответствуют;

агрегированные данные - чем выше должностной уровень руководителя, менеджера, 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 АИС на предприятии.

По данным [8] согласно исследованию META Group, 90 - 95% компаний списка Fortune 2000 активно применяют информационные хранилища, чтобы добиться преимущества в конкурентной борьбе и получить значительно большую отдачу от своих инвестиций. Трехлетнее изучение опыта 62 организаций, проведенное International Data Corporation (IDC) показало, что эти организации в среднем получили 400-процентный возврат своих инвестиций в АИС. В качестве примера внедрения ИХ на промышленных предприятиях России можно привести, в частности, заявление Магнитогорского металлургического комбината о создании полнофункционального ИХ сведений о ПСП (предполагается, что внедрение АИС на комбинате затронет более 800 рабочих мест) [10].

В заключение следует отметить, что создание ИХ производственного опыта для промышленного предприятия имеет целый ряд особенностей, связанных со спецификой функционирования ПСП с непрерывным производством, необходимостью одновременного учета событий применительно к непрерывным и дискретным процессам и т. д., что соз-

дает определенные трудности прежде всего в формировании структуры ИХ и выборе процедур его заполнения.

Литература:

1. Морозов В.П., Дымарский Я.С. Элементы теории управления ГАП: Математическое обеспечение. – Л.: Машиностроение, 1984. – 384с.
2. Кречетов Н., Иванов П. Продукты для интеллектуального анализа данных. – Computer Week-Москва. - 1997. - N 14-15. – С.32-39.
3. Бирюков А. Системы принятия решений и хранилища данных. – СУБД, 1997, N4, с. 37-41.
4. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных. - СУБД, 1997, N3, 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. – 143с.
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. Бритов П. А., Липчинский Е. А. Практика построения Хранилищ Данных: Система SAS . - СУБД, 1998, №4-5, с.14-20.
8. ТехноСерв А/С Хранилище данных. – Материалы Интернет, сайт <http://win.technoserv.ru/resh/sppr/dw.html>.
9. Data Mining with Microsoft SQL Server 2000. Technical Reference. – Microsoft Press, Washington, 2001. –368 p.
10. Проект OLAP DSS на Магнитогорском металлургическом комбинате. - Материалы Интернет, сайт <http://www.olap.ru/best/news/n001015151.asp>