

ВВЕДЕНИЕ

В настоящее время для управления бизнес-процессами каждого современного промышленного производства (ПП) используются информационные системы (ИС) различных уровней, том числе [1]:

- автоматизированные системы управления технологическим процессом (АСУ ТП), предназначенные для управления технологическим оборудованием, используемого для производства продукции;
- автоматизированные системы управления производственными процессами предприятия в режиме реального времени (*Manufacturing Execution System* — MES-системы);
- автоматизированные системы планирования, учета, контроля и анализа каждого бизнес-процесса промышленного предприятия (*Prise Resource Planning* — ERP-системы).

Данные ИС, помимо решения перечисленных выше задач, обеспечивают сбор (в том числе и в режиме реального времени) и хранение в различных автономных базах данных (БД) или хранилищах данных (ХД) различной информации (данных) о производственных процессах предприятия, зачастую разнородной по своей структуре [2]. Использование автономных БД и ХД обусловлена тем, что каждая из данных типов ИС имеет собственные, уникальные, источники информации, а также скорость генерации информации и структуру информации. (Отметим, что отмеченная особенность данных о производственных процессах промышленного предприятия позволяет называть эти данные гетерогенными данными (ГТ), соответственно хранилища этих данных — хранилищами гетерогенных данных (ХГД).)

В результате информация о жизненном цикле одной и той же единицы продукции (ЕП) ПП, которая требуется для решения задач управления, оптимизации бизнес-процессов ПП, выявления и устранения причин брака, оказывается размещенной в нескольких автономных БД и/или ХД. При этом закономерно возникает проблема поиска информации, релевантной решаемой

задаче, которая хранится одновременно в различных автономных ХД. При этом дополнительные трудности возникают в связи с тем, что в большинстве случаев информация, хранящаяся в ХД одной ИС, оказывается несинхронизованной во времени с информацией, хранящейся в ХД другой ИС. Перечисленные выше проблемы, связанные с вынужденным распределением технологической информации по нескольким автономным БД и/или ХД и необходимостью организации в этой связи информационного взаимодействия различных ИС, в полной мере присущи современному металлургическому производству (МП), на котором реализуется полный жизненный цикл продукции МП: сырье–чугун–сталь–металлопрокат, а также, при необходимости, утилизация продукции МП. Отмеченные обстоятельства закономерно приводят к динамическому изменению структуры технологической информации, соответствующей ее текущему состоянию, на каждом из этапов ЖЦ продукции МП.

Способ решения описанных выше проблем достаточно очевиден — необходимо разработать единое ХД, пригодное для хранения всей технологической информации, создающейся в процессе производства продукции ПП и привязанной к единой временной шкале, что позволит хранить, извлекать и анализировать информацию, структура которой принципиально разнородна и может изменяться с течением времени. При этом очевидно, что для создания единого ХД ПП необходимо решить многочисленные проблемы интеграции разнородных и зачастую несогласованных друг с другом, создаваемых разнородными источниками данных, работающими на различных физических принципах и обеспечивающими различные темпы генерации информации.

Анализ работ, посвященных различным аспектам разработки ХД, обеспечивающих интеграцию разнородных данных, показал, что в настоящее время для создания подобных ХД разработана теоретическая база, в том числе:

- реляционные модели данных [3–8];
- объектно-ориентированные модели данных [9–13];
- темпоральные модели данных [14–15],

существуют технологии, предназначенные для создания различных ХД, в том числе:

- технология транзакционной организации записи *OLTP* (*Online Transaction Processing*); предназначенная для ХД с заранее выбранной и далее не модифицируемой структурой хранения данных [16];

- технология *OLAP* (*Online Analytical Processing*), ориентированная на хранение большого объема структурированных данных, относящихся к определенной предметной области [17];
- технология *Oracle Streams* [18], предназначенная для интеграции данных, обмена данными и сообщениями с помощью механизма *Advanced Queuing* [19] в однородной среде и гетерогенных средах;

а также разработан ряд методов интеграции разнородных данных, в том числе:

- интеграции на уровне брокеров [20];
- интеграции на уровне интерфейсов (физических, программных и/или пользовательских) [21–23];
- интеграции на функционально-прикладном и организационном уровнях [24];
- интеграции на функционально-прикладном и организационном уровнях [24];
- интеграции на уровне корпоративных программных приложений [26,27]
- интеграции при помощи Web-сервисов [24–27];
- интеграции на уровне данных [21];
- интеграции на уровне сервисов [26];
- интеграции на уровне пользователя [27],

которые сегодня представлены в соответствующем сегменте рынка программного обеспечения в виде тех или иных программных инструментов, встроенных в том числе в такие популярные сегодня системы управления базами данных (СУБД), как Oracle, MS SQL и др.

Однако, несмотря на существенный прогресс в области разработки ХД в Российской Федерации и методов интеграции разнородных данных, как показали результаты анализа ИС и ХД, используемых на МП, на момент начала исследования не было создано единого ХГД МП. Данная ситуация, с нашей точки зрения, обусловлена в первую очередь отсутствием методологии структурного синтеза ХГД, под которой мы понимаем совокупность моделей (структурные модели информационных потоков промышленного производства; информационные и онтологические модели источников информации) и технологии разработки ХГД, выбор которых должен базироваться на результатах системного анализа выбранного ПП и его информационных потоков.

Анализ состояния предметной области, краткие результаты которого изложены выше, позволили сформулировать основную цель исследования: разработка методологии структурного синтеза ХГД и ее апробация на примере МП.

Для достижения цели исследования были решены следующие задачи.

1. Анализ современного состояния методов интеграции разнородных данных, технологий хранения и доступа к разнородным данным, программных инструментов, предназначенных для сознания хранилищ разнородных данных.

2. Системный анализ производственных процессов многоэтапного ПП на примере МП с целью обоснования гетерогенной структуры данных МП и выбора базовых информационных сущностей, достаточных для разработки онтологической модели данных МП и логической структуры единого ХГД МП.

3. Разработка на примере МП методологии структурного синтеза единого ХГД и обоснование возможности ее использования для разработки ХГД каждого многоэтапного ПП.

4. Подтверждение работоспособности методологии структурного синтеза единого ХГД на примере разработки ХГД МП как неотъемлемой части Автоматизированной системы выпуска металлургической продукции (АС ВМП).

В первой главе: изложены результаты анализа состояния предметной области, в том числе методы интеграции разнородных данных, модели разнородных данных и технологий хранения данных; определены следующие базовые понятия, используемые в работе: гетерогенные данные как результат интеграции данных, описываемых различными информационными моделями; источники гетерогенных данных; ХГД, предназначенное для сбора, хранения и доступа к гетерогенным данным; гетерогенная информационная система, обеспечивающая сбор, анализ и обработку гетерогенных данных; гетерогенный запрос, обеспечивающий извлечение релевантной информации из гетерогенного хранилища данных. Сформулированы цель и задачи исследования.

Во второй главе: изложены результаты системного анализа информационных потоков (ИП); определен состав множества источников информации (ИИ) и разработана их структурные схемы; предложена классификация данных МП, в соответствии с которой данные МП классифицированы на три группы: статистические, квазидинамические и динамические данные, структура которых может меняться в процессе ПП. Принимая во внима-

ние описанное выше свойство данных МП, сделан обоснованный вывод о том, что они имеют гетерогенную природу, в связи с чем для решения системных задач управления МП и повышения качества выпускаемой продукции необходимо перейти от использования автономных БД ERP, БД MES и АСУ ТП к хранению информации в едином ХГД МП.

Третья глава посвящена созданию методологии создания единого ХГД МП, предусматривающая последовательную разработку: информационных моделей введенных авторами информационных сущностей «план производства», «готовая продукция», «единица продукции» выбранного производства; онтологии гетерогенных данных выбранного производства; инфологической модели ХГД МП, далее выбор СУБД, разработка единого ХГД и обоснование возможности ее применения на любом многоэтапном промышленном производстве.

В четвертой главе описаны: разработанные на основании предложенной методологии структурного синтеза единого ХГД МП программные инструменты, обеспечивающие функционирование модуля ХГД МП в составе автоматизированной системы выпуска металлургической продукции (АС ВМП); способ реализации запросов к ХГД МП, обеспечивающий приемлемое с точки зрения практики время его выполнения, основанный на совместном использовании MD-индекса (многомерного индекса на основе R-дерева, обеспечивающего поиск по множеству параметров ЕП, наиболее подходящих для структуры хранения данных в ХГД МП), ET-индекса (битового индекса выполнения условия, обеспечивающего поиск точного соответствия условию), FE-индекса (индекса, кэширующего пересчитанные значения выражений *value* в хэш-таблице), а также результаты опытного тестирования ХГД МП, подтвердившие адекватность предложенных научно обоснованных решений и работоспособность соответствующих программных инструментов.

В заключении подведены итоги проведенного исследования.