

РЕШЕНИЯ ХРАНИЛИЩ ДАННЫХ В АВТОМАТИЗИРОВАННЫХ АНАЛИТИЧЕСКИХ СИСТЕМАХ АВТОМОБИЛЬНОЙ ПРОМЫШЛЕННОСТИ

Нгуен Дык Тхань, Чернов Э.А.

Автоматизированные системы управления автозаводов, разработанные за последнее время, решают в основном задачи в рамках одного информационного поля накопления данных. Способы доступа к накопленным данным, сводятся к простым механизмам поиска по прямому совпадению. Однако, любое хранилище и поток информации ценны не только и столько конкретными данными, образующими их, но связями между информационными объектами. Подобная метаинформация не может быть извлечена из баз данных, например, с помощью прямого использования реляционных запросов или технологии OLAP. При этом выявление связей и закономерностей в массивах данных традиционно остаётся работой аналитиков.

На современных автомобильных заводах используются гетерогенные программные средства для различных типов задач по заводскому планированию и производственным операциям. При этом системы программного обеспечения на операционном уровне в большинстве случаев ещё не интегрированы и поддерживают отдельные задачи, такие как управление производственными заказами, мониторинг производственных процессов, идентификация автомобилей, управление качеством, управление обслуживанием, контроль материалов и другие.

Вместе с тем, как показывает проведенный анализ, повышение эффективности автоматизированных аналитических систем (ААС) является многоуровневой задачей, решение которой необходимо обеспечивать как на уровне разработки, так и использования ААС для поддержки принятия решений. Решения на уровне разработки ААС в первую очередь связаны с повышением эффективности хранилища – снижением затрат времени и других ресурсов на обработку данных.

Следовательно, повышение эффективности хранилищ данных является многоуровневой задачей, решение которой необходимо обеспечивать как на уровне эффективных решений организации хранилища данных, так и в отношении форматов использования автоматизированных аналитических систем для поддержки принятия решений. При этом повышение эффективности технологии ETL должно обеспечивать снижение затрат времени и других ресурсов на обработку данных. Реализация вопросов повышения эффективности ААС на уровне форматов их использования определяется содержанием производственных задач и рассмотрены в 4-й главе. При решении задачи повышения эффективности хранилища данных также следует выделить два самостоятельных уровня, в соответствии с подсистемами хранилища – подсистема ETL сбора, накопления и использования данных и подсистема OLAP оперативной аналитической обработки данных.

Вопросы поддержки принятия решений также имеют характерные особенности и, в частности, обычно они являются довольно сложными. Особенности как баз данных систем поддержки принятия решений, так и выполняемых в них запросов являются причиной того, что акцент в этом случае переносится на проектирование с позиции повышения производительности, особенно – производительности процессов пакетного добавления данных и произвольного извлечения данных. Но, по мнению К. Дейта, такое состояние дел должно учитываться на этапе не логического, а физического проектирования.

В соответствии с задачами по обеспечению эффективности ААС в производственных процессах, в диссертации рассматриваются проблемы и их возможные решения по организации ETL-процессов в хранилище реального времени (Real Time Data Warehousing) или Активном хранилище данных (Active Data Warehousing, ADW).

Концепция ADW ориентирована на поддержку принятия оперативных решений на основании данных, полученных в системе управления потоком работ (Workflow Management System, WMS) в процессе мониторинга производственных операций. WMS-системы обеспечивают всестороннюю поддержку ранним стадиям жизненного цикла бизнес-процесса, но часто не имеют достаточных возможностей к установлению обратной связи и прозрачности выполнения процесса. Хотя WMS-системы часто регистрируют подробную информацию о ходе выполнения процесса, вместе с тем многие WMS-системы испытывают затруднения в использовании этой информации в целях контроля и анализа. Одной из причин ограниченных возможностей WMS-систем называют отсутствие широкой поддержки промышленного стандарта для данных аудиторского контроля технологического процесса (Workflow Audit Trail Data, WATD) и журнала аудита (workflow log), которые осуществляются основными WMS-системами. Поэтому, до последнего времени аналитикам было сложно использовать информацию аудита технологического процесса для получения чёткой картины состояния и выполнения производственных процессов.

В качестве решения проблем, связанным с применением ААС в режиме реального времени, в диссертации рассматривается организация ADW на основе архитектуры корпоративной информационной фабрики (CIF). В этом решении основной функцией CIF является предоставление базы данных для процесс-управляемой поддержки принятия оперативных решений в условиях постоянного контроля и улучшения производственных процессов. В терминах системного подхода CIF выполняет функции процессора, входом которого является серверная часть приложений, включающая комплекты ПО для обработки сообщений в корпоративных ИТ-системах, а выходом – ПО пользовательского интерфейса, содержащее инструменты мониторинга, подготовки отчётов, OLAP, Data Mining, инструментальные панели и т.п.

CIF состоит из трёх основных компонентов:

- 1) ETL-контейнер, который используется для передачи и трансформации WATD;

2) хранилище данных процесса, которое является частью системы хранилища корпоративных данных (EDW) и используется для хранения большой группы исторических данных workflow для поддержки принятия стратегических решений

3) банк данных процесса, который содержит очень подробные за последний день данные workflow, запущенных в настоящее время процессов и может быть использован для поддержки принятия тактических и оперативных решений.

На рисунке 1 представлена схема ETL-процесса для WATD, который поддерживается ETL-контейнером. На схеме показаны этапы извлечения, преобразования и загрузки. Кроме того, здесь включен этап оценки метрик workflow, что позволяет их оценить сразу же после того как они были рассчитаны. На основании результатов оценки, ETL-контейнер может либо посылать уведомления или инициировать любой вид производственных операций.

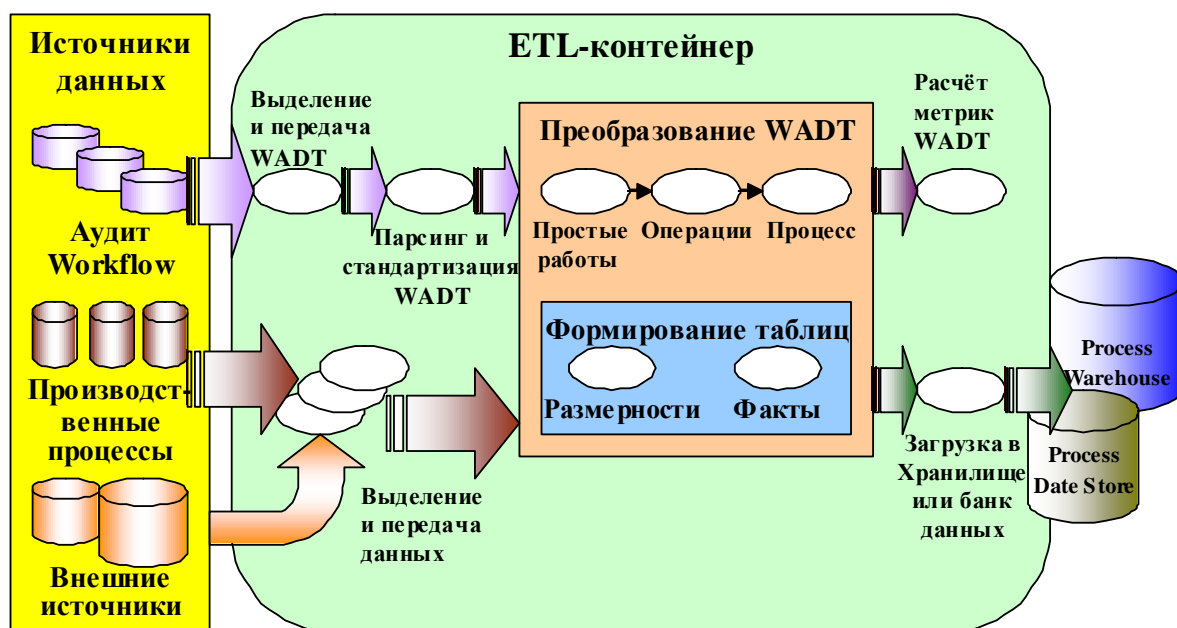


Рисунок 1 - ETL-процесс для данных аудита технологического процесса (WATD)

Как только данные (WATD) разобраны и стандартизированы, они проходят через ряд компонентов, поддерживающих всю необходимую информацию о простых (элементарных) работах, отдельных видах операций и процессах, расчёта и оценки показателей workflow и загружаются в таблицы размерностей и фактов системы хранилища данных. Для расчёта метрик ETL-контейнер может также использовать информацию из других источников данных.

На рисунке 3 показаны этапы обработки данных (WATD) для заполнения различных таблиц фактов метриками workflow. Некоторые из шагов ETL-обработки могут обеспечить получение данных из предыдущих этапов обработки, чтобы иметь возможность сформировать определённые агрегаты этих данных. Существенно то, что окружающая среда ETL может эффективно управлять взаимозависимостью этих этапов обработки WATD.

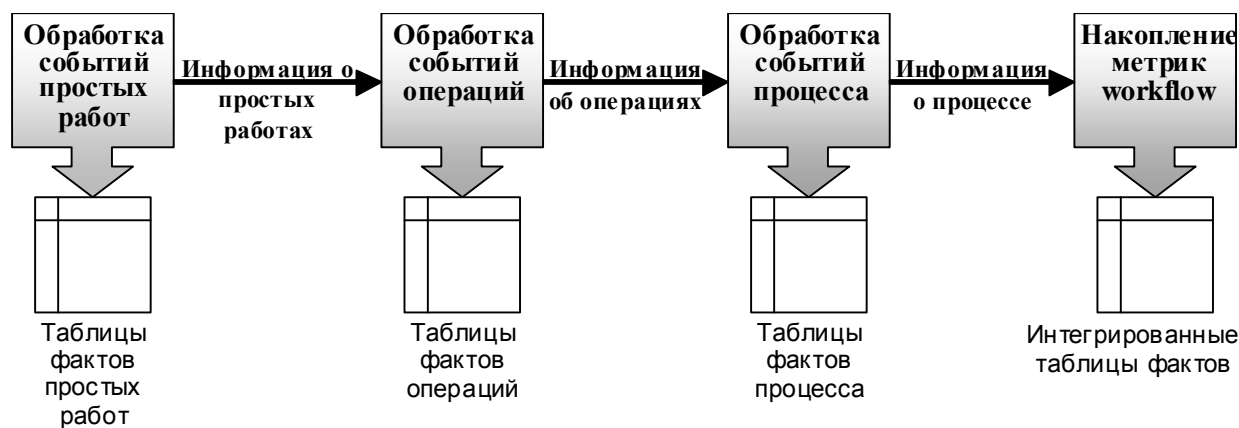


Рисунок 2 - Заполнение таблиц фактов

Во время обработки WATD, данные размерностей также будут преобразованы или заполнены. Все таблицы фактов на рисунке 3 обращаются к ряду таблиц размерностей. На рисунке 4 показаны различные типы таблиц размерностей, заполняемых в процессе обработки WATD. Следует учитывать, что различные типы ПО преобразователей могут включать сложную логику для обеспечения корректных ссылок или заполнения данных размерностей и поддержки истории этих данных.

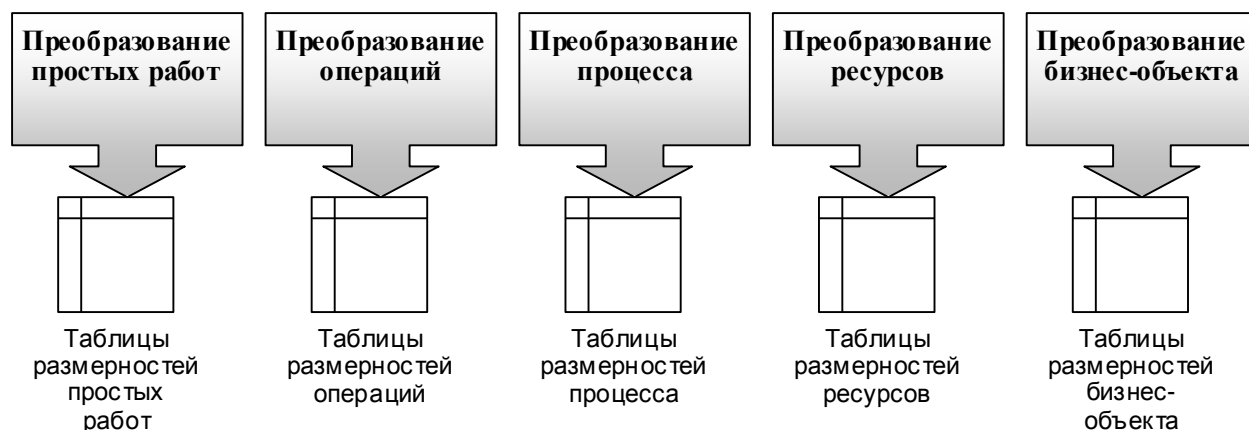


Рисунок 3 - Преобразования и заполнение размерностей данных

В диссертации представлены решения по организации ETL-контейнера для обработки событий Workflow и многопоточности в ETL-контейнере.

Эффективное использование подсистемы хранилища для оперативной аналитической обработки данных (OLAP) и многомерных технологических решений в традиционном формате для управления оперативными системами связано с целым рядом ограничений. В диссертации рассматривается решение этой задачи в условиях, характерных для автомобильной промышленности. Исходя из сформулированных требований к архитектуре, модель хранилища данных документооборота производственных процессов автозавода представлена в виде гиперкуба. Конструкцию

гиперкуба формируют шесть размерностей: производственные процессы (workflows), участники, организация, серверы, время и меры.

Список информационных источников

- [1] Нгуен Дык Тхань. Инструменты Data Mining в информационных системах автомобильной промышленности //Сборник научных трудов кафедры АСУ. 2011
- [2] Нгуен Дык Тхань. Технология автоматизированного анализа данных Data Mining // Международное научное издание «Современные фундаментальные и прикладные исследования». № 3. 2011
- [3] Нгуен Дык Тхань. Концепции и модели автоматизированного анализа в хранилищах и потоках данных // Сборник Перспективные вопросы мировой науки. № 28. 2011
- [4] Нгуен Дык Тхань. Business Intelligence в автоматизации поддержки управленческих решений //Сборник Перспективные разработки науки и техники. № 57. 2011.
- [5] Нгуен Дык Тхань. Автоматизация анализа данных для принятия решения в режиме реального времени в автомобильной промышленности // Вестник МАДИ. № 1. 2011
- [6] Нгуен Дык Тхань. Применение ассоциативных правил Data Mining в анализе гарантийных данных по гарантийным обязательствам автотранспорта // Грузовик. № 3. 2012.