

УДК 004.8

КОНЦЕПЦИИ И МОДЕЛИ АВТОМАТИЗИРОВАННОГО АНАЛИЗА В ХРАНИЛИЩАХ И ПОТОКАХ ДАННЫХ

Нгуен Дык Тхань, Чернов Э.А.

В настоящее время основой разработки систем автоматизированного анализа данных являются технологии Business Intelligence – BI. Контент-анализ российских и зарубежных источников, посвящённых концепциям, алгоритмам и ПО СУБД в формате BI выявил модели и тенденции развития современных систем автоматизированного анализа в хранилищах и потоках данных и позволил определить Business Intelligence как технологии поддержки принятия оптимальных бизнес-решений с помощью автоматизированных аналитических систем, связанных с ними инструментов и методологий, обеспечивающих оперативное влияние на результаты бизнес-процессов разного уровня. Отсюда, совокупность аналитических приложений BI для автоматизации интеллектуальной поддержки процессов принятия управленческих решений можно определить как автоматизированные аналитические системы (ААС), в которых программные продукты и аппаратные решения представляют подсистемы ААС.

Концепция корпоративного Хранилища данных (Enterprise Data Warehouse – EDW) начала формироваться в середине 1980 гг., а в 1992 г. W. Inman дал развёрнутое описание этой концепции как предметно-ориентированного, привязанного ко времени и неизменяемого собрания данных для поддержки принятия управляющих решений. Сегодня EDW считается лучшим решением для ААС. EDW обеспечивает всестороннюю и своевременную информацию, отвечающую требованиям всех уровней управления и аналитиков в рамках всей организации, которые используют информацию для принятия решений. EDW содержит интегрированные, стандартизированные, подробные, всеобъемлющие, текущие и исторические данные, обеспечивая единый источник бизнес-аналитики для поддержки принятия решений на стратегическом, тактическом и оперативном уровнях предприятия.

Достижение этих преимуществ требует, чтобы EDW было текущим, или активным. «Активный» означает то, что данные загружаются из всех систем-источников по крайней мере ежедневно, а в случае данных о рабочих процессах (Workflow) – практически непрерывно. Текущая информация обеспечивает своевременный тактический и оперативный анализ и принятие решений. «Активный» также означает, что EDW предоставляет регулярный или непрерывный мониторинг фактического состояния бизнеса и результатов в отношении определенных целей, с информированием ЛПР или обратную связь данных для операционных систем (workflow automation).

Выделяют два основных подхода к архитектуре хранилищ данных. К первому, предложенному W. Inman, относится нормализованное Хранилище данных с

пространственными витринами итоговых данных, которое также называют корпоративной информационной фабрикой (Corporate Information Factory – CIF). Второй подход – пространственное Хранилище данных с архитектурой шины (Data Warehouse Bus – DW BUS) – разработал R. Kimbal.

Инструменты для извлечения, преобразования и загрузки данных (ETL, Extraction-Transformation-Load tools) несут основную ответственность за передачу данных из транзакционной системы и интернета в хранилище данных. Инструменты ETL также проверяют качество данных, сосредоточивая внимание на тех элементах данных, которые будут наиболее часто использоваться в отчётах, информационных панелях, таблицах и других выходных материалах.

ETL-процесс для решения многих задач, является узким местом концепции хранилищ данных и при построении хранилища данных наибольшие затраты, как правило, приходится именно на этап ETL. Правильный подход в реализации процессов ETL позволят существенно оптимизировать затраты при построении современного аналитического информационного комплекса и повысить его эффективность. Следовательно, сложность логических и физических процессов, определяющих функционирование EDW, сопровождается многовариантностью их организации, что в свою очередь создаёт условия для формирования задач по оптимизации организации процессов EDW, направленных на снижение затрат времени на обработку данных и выполнение запросов, а также постоянных и переменных издержек, связанных с приобретением и эксплуатацией EDW. Ограничениями в этих оптимизационных задачах выступают требования к конечным результатам, получаемым с помощью EDW.

Автоматизированные системы управления Workflow Management Systems (WMSs) являются технологией автоматизированного управления потоком работ и через него бизнес-процессом, то есть технологией, основанной на процессном подходе к управлению организацией. WFMSs совершенствуют бизнес-процессы за счёт их автоматизации, получения нужной информации в нужном месте для выполнения конкретных работ, и интегрируя информацию на предприятии. WMSs поддерживают выполнение бизнес-процессов, как требует определение процессов, автоматизируют разработку и осуществление технологических операций, руководствуясь бизнес-правилами и логикой исполнения, и, наконец, документируя выполнения всех этапов бизнес-процесса.

В процессе выполнения Workflow изменения состояния распределителя работ (workflow engine) отражаются в контрольном журнале – журнале аудита (workflow log) или базе данных, в виде так называемого audit trail – следа контроля (аудита) – контрольная запись, отражающая, в частности, кто и какие операции в заданный период времени выполнял. Workflow logs содержат очень ценную информацию о фактическом исполнении бизнес-процессов (в отличие от только заданного или желаемого описания бизнес-процессов). Они могут быть очень ценным ресурсом для совершенствования бизнес-процессов и реинжиниринга и предоставлять информацию для процессов контроля и мониторинга. Вместе с тем Workflow log является источником для аналитической технологии поиска информации в данных (workflow

mining) в формате модели Workflow Audit Trail Data – WATD, которая основана на разработке процедуры анализа структуры бизнес-процессов с целью обнаружения их аномального выполнения.

Для выявления нарушений в структуре и ходе выполнения бизнес-процесса могут быть использованы различные подходы: основанные на использовании аппарата сетей Петри, распознавания структуры процесса по файлу статистики, анализа дерева достижимости, применении шаблонов (pattern analysis) и др.

Логической моделью workflow mining может служить модель WATD, основанная на сетях Петри – WF-net (рисунок 1). Предположим, что мы имеем журнал регистрации (workflow log), сформированный на базе множества исполнений бизнес-процесса, описываемых сетью WF_1 . Основываясь на этом журнале регистрации потока операций и используя соответствующий алгоритм поиска информации (mining algorithm) мы создаём сеть WF_2 . Задача автоматизированного анализа аудита данных потока операций в этом случае заключается в проверке равенства $WF_1 = WF_2$. Случай невыполнения этого равенства является сигналом о нарушении правил выполнения и/или результатов данной операции и необходимости принятия соответствующего решения по исправлению ситуации в автоматическом режиме или ЛППР.

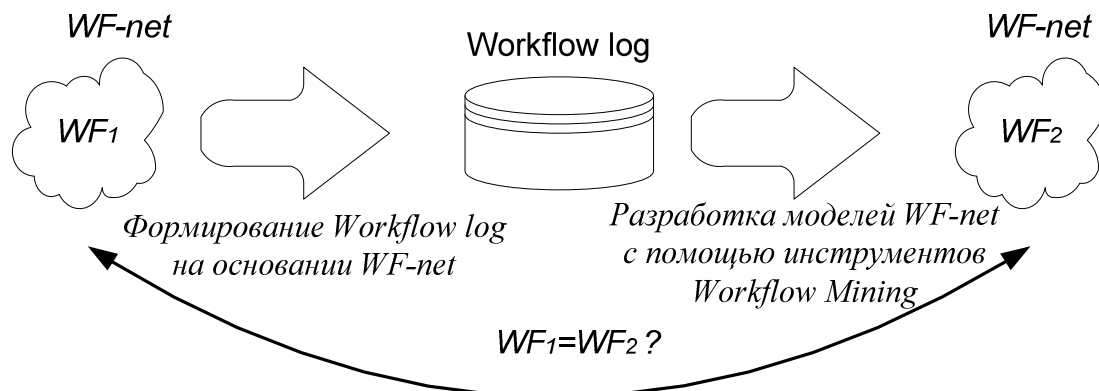


Рисунок 1 – Логическая модель аналитической технологии WATD оценки выполнения бизнес-операции

В диссертации рассмотрена концептуальная модель процесс-ориентированного Хранилища данных, в которой объединены данные аудита потока операций (WATD) с бизнес-данными объекта. ETL-процессы соответствующих уровней, импортируют WATD так же как и бизнес-данные объекта в общий архив. Данные отформатированы согласно соответствующим наборам метаданных, которые описывают формат и семантику основных структур исходных («сырых») данных.

Полученные концептуальные предпосылки и модели автоматизированного анализа данных в хранилищах и потоках данных позволили перейти к анализу возможностей применения методов аналитических технологий в автомобильной промышленности.

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.