

УДК 004.658.6

А.Ю.Якимайнен (6 курс, каф. ИУС), И.В.Бутенко (асп., каф. ИУС),  
Д.Ф.Дробинцев, ст. преп.

## МЕХАНИЗМ ФОРМИРОВАНИЯ МЕТАДАННЫХ НА БАЗЕ СТАНДАРТА CWM

Ключевым фактором рыночного успеха в сегодняшних условиях высокой конкуренции становится оперативное принятие эффективных деловых решений. Однако многие организации сталкиваются с труднопреодолимым препятствием – огромный объем и высокая сложность исходных данных. Решением этой проблемы может являться построение информационно-аналитической системы, которая позволяет манипулировать не только данными, но и их отношениями, зафиксированными в рамках метаданных.

Наиболее трудной задачей в технологиях хранилищ данных и аналитических систем является задача обмена метаданными между различными компонентами, входящими в состав любой системы. Такими компонентами могут быть информационные источники, хранилища и витрины, ETL-средства, аналитические средства и т.д. Каждый из этих компонентов имеет свои метаданные, хранящиеся в специальных форматах. Проблема состоит в том, что все эти разнородные по структуре и синтаксису метаданные семантически не взаимосвязаны, поэтому для согласованной и корректной работы системы в целом их необходимо передавать от одних средств другим, совместно использовать, устранять несоответствия и противоречия. Для решения этой проблемы и представления всевозможных метаданных, корпорацией OMG (Object Management Group) был разработан стандарт Common Warehouse Metamodel (CWM).

CWM основан на открытых объектно-ориентированных технологиях и стандартах, используя UML в качестве языка моделирования, XMI и XML – для обмена метаданными и язык программирования JAVA – для реализации моделей и спецификаций.

В основе CWM лежит модельно-ориентированный подход к обмену метаданными, в соответствии с которым объектные метамодели строятся в соответствии с синтаксическими и семантическими спецификациями некоторого общего стандарта. Каждая метамодель является отдельным модулем, что позволяет минимизировать зависимости между различными компонентами, уменьшить сложность и повысить наглядность модели. Метамодель (или средство моделирования) предназначена для представления определенного типа метаданных хранилища. Фактически CWM состоит из ряда метамodelей, которого достаточно для формирования метаданных обо всем хранилище данных.

В данной работе предлагается следующий механизм формирования метаданных.

Сначала формируются метаданные об исходных источниках информации и целевых базах данных. Для этого используются метамодели, позволяющие работать с существующими и новыми источниками данных, в том числе, с реляционными базами данных, ориентированными на запись базами данных (record oriented databases), а также XML- и основанными на объектах (object-based) источниками данных. Например, метамодель Relational (реляционная метамодель) позволяет собрать метаданные, включающие сведения о базе данных: ее название, месторасположение, хозяина, названия таблиц, столбцов, а также связи между ними, ключевые поля и т.д.

Далее используются метамодели, отвечающие за моделирование сервисов информационного анализа, которые обычно используются в хранилище данных. Эти метамодели определяют метаданные для преобразования данных, аналитических служб, визуализации информации/репортирования (business nomenclature) и data mining. Например, метамодель Transformation (ETL-средства) позволяет сформировать информацию обо всех

преобразованиях над данными, которые происходили при их извлечении из исходных источников и загрузки в целевые базы данных. Кроме того, эта метамодель включает сведения о том, когда и кто производил преобразования.

В любом хранилище необходимы средства, позволяющие моделировать процедуры по управлению хранилищем, устанавливать регламент их выполнения, специфицировать процессы контроля и протоколирования для загрузки информации и произведенных корректировок данных хранилища. В стандарте CWM предусмотрены такие метамодели. Они отвечают за управление хранилищем данных, представляют стандартные процессы и операции, журнализации (activity tracking) и планирования работ (scheduling). Например, метамодель Warehouse Operation (операции хранилища) содержит метаданные обо всех ежедневных загрузках и выгрузках информации.

Создание информационно-аналитических систем, основанных на технологиях хранилищ данных, невозможно без применения стандарта, подобного CWM. В качестве альтернативы пришлось бы разрабатывать собственные элементы стандарта, реализуемые в виде уникальных программных шлюзов для каждой пары интегрируемых систем. Использование CWM позволяет решать проблему обмена как экземплярами метаданных, так и метамоделями между гетерогенным программным обеспечением, поставляемым различными производителями. При этом CWM позволяет расширять имеющиеся метамодели, а так же создавать новые.