

УДК 004.658.6

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

## ПРИНЦИПЫ ПОСТРОЕНИЯ КЛАССИЧЕСКОГО ХРАНИЛИЩА ДАННЫХ НА ОСНОВАНИИ РАЗВИВАЕМЫХ ВИТРИН ДАННЫХ

Хранилище данных (ХД) – широко распространённый и уникальный репозиторий информации предприятия. Среда Хранилища предназначена только для чтения и состоит из детальных и агрегированных данных, которые полностью очищены и интегрированы; кроме того, в нем хранится обширная и детальная история данных на уровне транзакций.

Достоинствами архитектуры классического Хранилища данных являются: непротиворечивость информации, один набор процессов извлечения и бизнес-правил, общая семантика, централизованная, управляемая среда, легко создаваемые и наполняемые витрины данных, единый репозиторий метаданных.

Недостатки такого архитектурного решения: реализация требует больших затрат, высокая ресурсоемкость, потребность в системах и ресурсах в масштабе всего предприятия, рискованный сценарий ("все поставлено на карту").

С точки зрения архитектурного решения хранилище данных реализует свои функции, прежде всего, через подмножество зависимых витрин данных (ВД). Для реализации огромного потенциала, присущего ХД, с одновременным устранением таких недостатков, как большие затраты на реализацию, высокая ресурсоемкость, длительное время разработки до появления системы на рынке используется архитектура постепенно развиваемых витрин данных.

Такая архитектура обеспечивает быстрое развертывание, запуск в промышленную эксплуатацию и возврат инвестиций, но она нацелена на решение конкретного спектра задач, и при появлении новых необходимо частично перепроектировать систему. Факт перепроектирования влечет за собой значительные временные и финансовые затраты по доработке и тестированию ХД. Для уменьшения трудозатрат при доработке ХД в случае изменения требований необходимо определять их таким образом, чтобы заложить возможные тенденции дальнейшего развития системы. В результате возникает задача формирования структуры развиваемых витрин данных таким образом, чтобы при появлении новых требований система претерпевала минимум изменений даже на этапе разработки.

При построении развиваемых витрин данных разработчик должен составить иерархию функциональных требований (ИФТ), которая в дальнейшем определяет модель данных высокого уровня ХД. При этом модель данных строится на основании базовых понятий, выявленных из ИФТ.

Составляя дерево требований, мы постепенно переходим от общих понятий к частным понятиям и сущностям предметной области. Эти понятия являются определяющими информационное наполнение системы, поскольку ссылаются на объекты транзакционных систем и других внешних источников. Зная информационные сущности или исходные показатели, мы сможем сформировать базовые объекты хранилища данных. Базовые понятия определяют не специфику конкретной аналитической задачи, а специфику отрасли в целом. Считается, что базовые понятия по своей сути статичны и практически не меняются со

временем. Такой подход позволяет максимально переиспользовать спроектированное ХД для решения других задач. В результате, при таком подходе мы получаем максимальное переиспользование процедур загрузки, очистки, агрегации и конвертации данных, поступающих в хранилище из внешних источников.