

Хранилище данных для стратегии дифференцирования

Хранилища Данных: что это и кому они нужны



Данные, информация, знания, решения. Как эти термины связывает хранилище данных? Как хранилище данных поможет дифференцироваться на рынке. Об этом рассказывает Максим Бодаев, директор по развитию бизнеса Citia BTC

Максим Бодаев,
директор по развитию бизнеса Citia BTC

Назначение ХД

Каких-то десять лет прошло с тех пор, как в нашей стране перестали ассоциировать Хранилища Данных (ХД) со шкафами, битком набитыми жесткими дисками: сейчас «шкафы» принято называть системами хранения данных (СХД), а представление о том, что такое «хранилище данных» у большинства ИТ-специалистов практически совпадает с его определением в Википедии – очень большая предельно-ориентированная информационная корпоративная база данных, специально разработанная и предназначенная для подготовки отчетов, анализа бизнес-процессов с целью поддержки принятия решений в организации.

Концепт ХД

Если «копнуть» еще глубже в историю, то сама концепция хранилища данных (применительно к бизнесу) появилась с подачи компании IBM в конце 80-х. В то время эта концепция представляла собой модель архитектуры для организации потока

данных из оперативных (учетных) систем к лицам, принимающим решения. С тех пор хранилище данных как система получило массу имен: система поддержки принятия решений, система бизнес-анализа, корпоративная информационно-аналитическая система.

Архитектура «слоев»

Концептуальная архитектура же подобной системы остается неизменной по сей день, включая в себя следующие «слои»:

Слой систем-источников данных: ими служат любого рода учетные системы, использующиеся в организации – ERP, CRM, MRP, автоматизированные банковские системы, биллинговые системы (в телекоммуникационных компаниях), разрозненные Excel-файлы в конце концов;

Слой доступа к данным: фактически интерфейс хранилища данных к системам-источникам, отвечающий за извлечение, преобразование и загрузку (ETL – extract, transform, load) данных в хранилище;

Слой хранения данных и метаданных: собственно, сердце хранилища – та самая «очень большая ... база данных»;

Слой доступа к информации: интерфейс конечных пользователей системы (потребителей информации, полученной из данных) для работы с хранилищем данных.

Единый знаменатель

Казалось бы, если все так просто и прозрачно – в какой из учетных систем нет модуля отчетности? – зачем было придумывать что-то новое? Ответ достаточно тривиален.

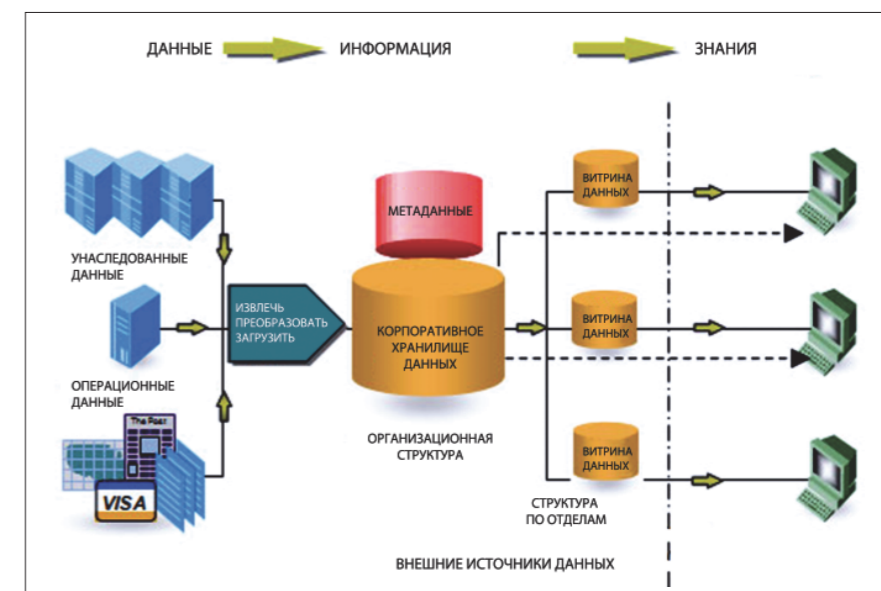
Начнем с того, что учетных систем в любой организации всегда несколько, и в каждой из них одни и те же данные учитываются по-разному: «М» и «Ж» для определения пола клиента в одной, и «0» и «1» для той же цели – в другой, например. Поэтому для ответа на простейший вопрос «каково соотношение мужчин и женщин в части приносимых организации доходов», например, уже требует как минимум одной «лишней» опера-

ции – приведения значений такого измерения анализа как «пол» к единому знаменателю.

А если все учетные системы от одного поставщика (фантастика, конечно, но допустим) и они настолько интегрированы друг с другом, что подобные казусы и разночтения исключены? Не так все просто: заранее редко можно «просчитать» какого рода информация потребуется тому или иному лицу, принимающему решения, а плодить отчеты сотнями на все случаи жизни – такого ни одному ИТ-директору не пожелаешь.

Учет vs. анализ

Но ведь можно взять удобное средство бизнес-анализа (термин появился также в конце 80-х) и «прикрутить» его прямо к базам данных учетных систем: теперь каждый пользователь может сам себя обеспечить информацией. Оказывается наиболее «одаренные» аналитики своими запросами запросто кладут на лопатки рано или поздно любую учетную систему: вместо того,



занных лиц например) – и этот перечень далеко не полный.

Принципы организации данных

Хранилища данных избавлены от таких проблем, так как данные в них (кстати, физически перемещен-

к некоторому промежутку или моменту времени.

От данных к решению

Если данный перечень дополнить еще и рядом требований к информации, которые удовлетворяет хранилище данных – правильность (за счет проверки и очистки данных), своевременность (за счет предрасчитанных показателей), доступность, релевантность, краткость и защищенность (за счет применения инструментов бизнес-анализа) – становится очевидным, что подобные системы служат эффективной поддержкой следующей важной цепочки: данные → информация → знания → решения.

А в любом конкурентном бизнесе именно информация и знания могут служить самым главным дифференциатором от конкурентов. Свидетельством этому служат некоторые из компаний, которым Citia BTC помогала и помогает внедрять Хранилища Данных для обеспечения их аналитической информацией, такие как Киевстар и Астелит (данные по трафику и начислениям, абонентской базе), Банк «Дельта» (данные по клиентам и их взаимодействиям с банком), АТБ Маркет (данные по продажам, остаткам и движению товаров) – они являются лидерами в своих сегментах бизнеса и на них равняются многие другие организации.

Системы [хранения данных] служат эффективной поддержкой следующей важной цепочки: данные → информация → знания → решения

чтобы учитывать, она начинает часами пыхтеть, чтобы вернуть пользователю выборку типа «все операции всех клиентов по дням за последние три года».

Проблемы трактовки

А если увеличить аппаратные мощности? Не тут-то было: оказывается аналитикам нужна историчность – а сколько нам был должен клиент позавчера, а в каком регионе он находился месяц назад, а к какому сегменту относился в прошлом году? В учетных системах историчности нет априори. Но даже если решить и эту проблему, остаются проблемы одинаковой трактовки информации, проверки, очистки и стандартизации данных (адресной например), восполнение недостающих данных (определение свя-

ные из учетных систем) организованы согласно следующих принципов:

Проблемно-предметная ориентация. Данные объединяются в категории и хранятся в соответствии с областями, которые они описывают, а не с приложениями, которые они используют.

Интегрированность. Данные объединены так, чтобы они удовлетворяли всем требованиям предприятия в целом, а не единственной функции бизнеса.

Некорректируемость. Данные в хранилище данных не создаются: т.е. поступают из внешних источников, не корректируются и не удаляются.

Зависимость от времени. Данные в хранилище точны и корректны только в том случае, когда они привязаны