

УДК 658.5:681.3.016

А.М. Валуев, А.С. Панкратов

**СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ИНТЕГРАЦИИ
ИНФОРМАЦИИ ИЗ НЕЗАВИСИМЫХ БАЗ ДАННЫХ
И ВОЗМОЖНОСТИ ИХ ПРИМЕНЕНИЯ В ЗАДАЧАХ
ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ**

Рассмотрена проблема совместного использования независимых баз данных в задачах планирования и управления. Представлены две схемы такого информационного взаимодействия, первая для случая использования баз данных подразделений одной организации, вторая — для задач координации деятельности бизнес-партнеров при условиях строгой конфиденциальности нерелевантной информации.

Ключевые слова: управление, планирование, база данных, информационная система, запрос, конфиденциальность.

Среди многочисленных тенденций, наблюдаемых в современной жизни нашей страны, можно выделить две, на первый взгляд взаимно противоречащие, тем не менее, органично дополняющие друг друга. С одной стороны, развитие частного бизнеса и отход государства от прямого управления производственными и другими бизнес-процессами внутри страны ведёт к повышению уровня самостоятельности и автономности отдельно взятых отраслей, предприятий, корпораций. Иными словами, наблюдается тенденция децентрализации различных сфер общественной жизни. С другой стороны, именно эта децентрализация ведёт к необходимости формирования устойчивых межведомственных и межкорпоративных связей, объединения на информационном и иных уровнях в рамках одной предметной области. Иными словами, в настоящее время возникает потребность в разного рода интеграционных процессах, актуальность которых ощущается в настоящее время во всём мире.

Одним из важнейших аспектов процесса интеграции является проблема консолидации информации из различных и независимых источников, в частности, объединение в единое целое разнородных баз данных, без ущерба их независимости. Можно назвать ряд предметных областей, где подобная задача интеграции может возникнуть естественным образом:

- медицинская сфера: объединение информации, содержащейся в электронных медицинских картах пациентов;
- торговая сфера: поддержка объединённой информации о наличии товаров и ценах на них в магазинах различных торговых сетей;
- транспортно-туристическая сфера: поддержка постоянно обновляемого электронного расписания различных видов транспорта, причём как на магистральных, так и на местных рейсах;
- горно-геологическая сфера: объединение в единое целое баз данных различных ведомств, содержащих информацию о месторождениях, к примеру, каменного угля.

Данной проблематике посвящён ряд работ российских и зарубежных авторов. Среди российских учёных, занимающихся проблемами интеграции можно назвать работы Л. В. Масель, А. В. Черноусова, А. Кудинова. Однако следует отметить, что в настоящее время лишь намечаются основные контуры обобщённого подхода к решению задачи интеграции данных из разных компьютерных систем (см., например, [1]), большинство публикаций на эту тему затрагивает лишь отдельные аспекты проблемы либо частные предметные области (например, учёт объектов недвижимости) со своими специфическими особенностями [2, 3]. Настоящая работа посвящена некоторым задачам управления горным производством и также не претендует на общность, однако представляется, что описанный ниже подход может иметь и более широкое применение.

В горном производстве имеются многочисленные примеры интеграции и взаимозависимости бизнес-процессов собственного горного производства и смежных сфер, например:

1. взаимная зависимость карьера и железной дороги в отношении поставки порожних составов и отправки груженных, особенно в случае прямой отгрузки полезного ископаемого из забоев;

2. желательность координации деятельности угольной шахты (карьера) и электростанции, работающей на угле, в следующих аспектах:

- уточнение информации о качестве угля по данным о процессе его сжигания и данным лабораторных исследований на электростанции;

- регулирование качественного состава отгружаемого угля;

- регулирование объемов поставок в случае работы «с колес», т.е. отсутствия запасов угля на электростанции. В по-

следнем случае процесс отгрузки нужно координировать также с железнодорожной организацией;

3. аналогичные зависимости в деятельности нескольких горных предприятий, работающих на одну обогатительную фабрику, и самой обогатительной фабрики.

Все перечисленные аспекты взаимозависимости деятельности деловых партнеров, требующие постоянного информационного обмена, актуальны, прежде всего, для оперативного планирования и управления. Для среднесрочного планирования конкретная информация из баз данных смежников едва ли может быть доступна; во всяком случае, объем релевантной информации невелик и проблемы ее передачи не возникает. Для этого уровня управления важнейшее значение приобретает интеграция информации о залежи из разных источников. Обязательным ее источником служат данные детальной и эксплуатационной разведки, причем последние регулярно пополняются. Анализ проб позволяет определять состав пород и содержания компонентов в породах, а методами геостатистики [4, 5] эти данные распространяются на смежные объемы. Но и границы пластов или рудных тел также получаются по данным геологической разведки лишь путем интерполяции, что зачастую в силу сложной структуры залежи и нарушений залегания не может дать вполне достоверную информацию с требуемой точностью. Наоборот, геофизические методы позволяют выявить поверхности раздела пород как единое целое, но из важных свойств устанавливают только значение плотности, что не позволяет ни однозначно идентифицировать породы, ни тем более определить значения содержания компонентов.

Зато совместное использование двух источников резко повышает объем и достоверность информации о месторождении. Во-первых, по пересечениям границ раздела пород с колонками скважин можно установить, какие именно поверхности раздела служат границами пластов или рудных тел. Само по себе измеренное в пробах значение плотности породных комплексов, например, кондиционного угля или углистого аргиллита, дает основание для их идентификации исключительно на основе средних значений плотности. Корреляционная зависимость между значениями плотности и основным показателем качества — содержанием основного металла для руды, зольности или теплотворной способности для угля или сланца — также позволяет повысить достоверность оценки запасов. Это может быть выполнено путем коррекции значения показателя, рассчитанного методами геостатистики, для точек, в которых плотность определяется геофизическими методами, по измеренному значению плотности на основе корреляционных зависимостей.

В данном случае мы имеем дело с двумя информационными массивами, пополняемыми двумя отдельными организациями или подразделениями одной организации, работающими на общий конечный результат. Поэтому их базы данных взаимно доступны для любых запросов без изменения содержания баз. Для сведения этих данных в единое целое может быть применена технология так называемых медиаторов. Медиаторами называются программные компоненты, принимающие запросы от пользователей и затем направляющие их к соответствующим источникам, возможно, с предварительной трансляцией; полученные ответы от источников приводятся в соответствие со структурой ме-

диатора, объединяются и выдаются пользователю. Это может быть предусмотрено следующим образом.

Взаимосвязь элементов данных из различных источников (необходимую при исполнении запросов к медиатору) описывается, как правило, с помощью таблиц соответствия. Пусть данные геологической разведки о пробах угля в колонках скважин хранятся в базе данных DB1 в таблице GEOSK, а данные о высотных отметках точек поверхностей раздела и о плотности пород хранятся в базе данных геофизической разведки DB2 в таблице SLOI. Для совместного их использования при интеграции данных с помощью медиатора, представляющего, по сути, виртуальную базу данных (представление для пользователя), эта виртуальная база имеет свои характеристики (реквизиты), которые также заносятся в общую таблицу соответствия.

Для остальных источников, равно как и для подключения новых источников, в таблицах TABLE_CORR_GEO и TABLE_CORR_POINTS заводятся новые колонки — в простейшем случае с названиями соответствующих атрибутов, в более сложных случаях с функциями преобразования.

Совершенно иную ситуацию представляет собой использование горным предприятием данных железной дороги, электростанции или металлургического предприятия. Эти последние, хотя и являются деловыми партнерами горного предприятия, преследуют собственные цели. Сферой их взаимных интересов являются: своевременная подготовка горного предприятия к приму очередной партии вагонов под погрузку, что приводит к более быстрой оборачиваемости вагонов; обеспечение требуемого качества шихты и своевременная поставка продукции

Таблица 1

Таблица TABLE_CORR_GEO, описывающая простые соответствия атрибутов

Медиатор. таблица	Медиатор. таблица. поле	DB1. таблица	DB1. таблица. поле	DB2. таблица	DB2. таблица. поле
GEO	KROVLA	GEOSK	KROVLA	SLOI	GL
GEO	POCHVA	GEOSK	POCHVA	SLOI	NULL
GEO	ZOLA	GEOSK	ZOLA	SLOI	NULL
GEO	UDV	GEOSK	UDV	SLOI	PLOTN
GEO	PORODA	GEOSK	PORODA	SLOI	NULL

Таблица 2

Таблица TABLE_CORR_POINTS, содержащая функции приведения координат точки к формату медиатора (географическая широта и долгота) F1X, F1Y, F2X, F2Y:

Медиатор. таблица	Медиатор. таблица. поле	DB1. значение	DB2. значение
POINTS	POINT_X	F1X(MR.ORIGIN_X, MR.ORIGIN_Y, SKVG.X, SKVG.Y)	F2X(PROFIL.PSTART_X, PRO- FIL.PSTART_Y, PROFIL.PEND_X, PROFIL.PEND_Y, TOCHKA.DIST)
POINTS	POINT_Y	F1Y(MR.ORIGIN_X, MR.ORIGIN_Y, SKVG.X, SKVG.Y)	F2Y(PROFIL.PSTART_X, PRO- FIL.PSTART_Y, PROFIL.PEND_X, PROFIL.PEND_Y, TOCHKA.DIST)

горного предприятия или, наоборот, запланированная задержка поставки при переполнении склада. Естественно предполагать, что партнеры во избежание утечки конфиденциальной информации будут максимально ограничивать доступ к своей БД той выборкой, которая совершенно необходима для осуществления делового взаимодействия.

Рассмотрим эту возможность на примере. Представим себе, что некоторая подсистема базы данных железнодорожной организации включает таблицы Составы, Вагоны и График (движения). Допустим, что параметр текущего состояния вагона следующие: № вагона, Тип вагона, Состояние (порожний, груженный), Объем, Грузоподъемность, Назначение (станция), Адресат (организация), № состава. Другие атрибуты, как-то: Владелец, Дата изготов-

ления, Проверить до и т.п. не должны интересоваться адресата и поэтому не сообщаются. В результате о рейсах, включающих вагоны, поступающих под погрузку в адрес горного предприятия «Шахта Новая» на его грузополучательную станцию «Шахтная», можно сформировать представление на основе следующего SQL-запроса

```
SELECT Вагоны. № состава,
COUNT(*) AS Количество_вагонов,
SUM (Вагоны.Грузоподъемность) AS
Сум_Грузоподъемность, SUM (Ваго-
ны.Объем) AS Сум_Объем FROM Ваго-
ны GROUP BY Вагоны.№ состава
WHERE Вагоны.Адресат = «Шахта Но-
вая» AND Вагоны. Назначение =
«Шахтная»
```

Здесь не хватает, однако, информации о расчетном времени прибытия. Сообщать весь маршрут состава, доставляющего порожние вагоны, не нуж-

но и нежелательно. Полная информация о времени прибытия требуемых составов в пункт назначения будет получена путем естественного соединения по полю № состава вышеописанного представления с другим представлением

SELECT График. № состава, График. Дата_прибытия, График. Время_прибытия FROM График WHERE График. Станция = «Шахтная» ORDER BY График.Дата_прибытия, График. Время_прибытия

Последнее представление содержит информацию обо всех составах, делающих остановку на станции «Шахтная», независимо от наличия в нем вагонов, адресованных шахте. Однако при использовании его не самого по себе, а во внутреннем соединении с предыдущим запросом информационной системе шахты не будут доступны сведения о составах, не имеющих к ней отношения.

В свою очередь, база данных оперативно-диспетчерского управления горного предприятия в части движения вагонов на территории предприятия

имеет отношение к формированию составов, отправляемых с грузоотправительской станции, из конкретных груженых вагонов, что интересует железнодорожную организацию. Эта информация может предоставляться на тех же условиях конфиденциальности всей нерелевантной информации для запросов со стороны железнодорожников. Таким образом, доступной должна быть только информация о плановых моментах поступления груженых вагонов на грузоотправительскую станцию. На ее основании диспетчеры железной дороги должны принимать решения о необходимости остановки составов на грузоотправительской станции шахты, времени ожидания еще не поступивших груженых вагонов и конкретном наборе вагонов, которые будут присоединены к прибывающему составу. Разумеется, все представления, применяемые в задачах управления сторонней организации, не могут быть обновляемыми во избежание нежелательных воздействий на процессе управления в организации — источнике данных.

СПИСОК ЛИТЕРАТУРЫ

1. Юмагузин Н.В. Классификация взаимосвязей в схемах данных // Программные продукты и системы. — 2007 — № 3. — С. 204—212.

2. Торшин Д.В. Организация единого интегрированного пространства на основе универсального формата обмена данными // Научно-технические ведомости СПбГПУ, серия «Информатика. Телекоммуникации. Управление». — 2009. — № 2(71). — С. 26—32.

3. Воробьева М.С. Построение модели и интеграции данных в информационно-

управляющих системах // Модернизация образования в условиях глобализации: Круглый стол «Образование через науку и инновации», 14—15 сентября 2005 г. / Под ред. В. Н. Кутрунова. Тюмень: Изд-во ТюмГУ, 2005. — С. 26—28.

4. Матерон Ж. Основы прикладной геостатистики. — М.: Мир, 1968. — 408 с.

5. Давид М. Геостатистические методы при оценке запасов руд. — Л.: Недра. — 1980. — 360 с. **ПЛАБ**

КОРОТКО ОБ АВТОРАХ

Валуев А.М. — профессор, доктор физико-математических наук, Московский государственный горный университет, amvaluev@online.ru

Панкратов А.С. — кандидат физико-математических наук, доцент, Российский университет дружбы народов, sasha.pankratov@gmail.com