

П Р И К Л А Д Н А Я

ИНФОРМАТИК@

научно-практический
журнал

№ 4 (52) 2014

ISSN 1993-8314



СИИЕРГИЯ ПРИИИТ

П Р И К Л А Д Н А Я

ИНФОРМАТИК@

научно-практический журнал

№ 4(52) 2014

Июль-август

ISSN 1993-8314

Московский финансово-промышленный университет «Синергия»

С 19 февраля 2010 года журнал включен в Перечень ведущих периодических изданий, рекомендованных ВАК для публикации результатов диссертационных исследований.

РЕДАКЦИОННЫЙ СОВЕТ

Главный редактор

Емельянов А. А., докт. экон. н., проф., Национальный исследовательский университет «МЭИ»; Национальное общество имитационного моделирования, Санкт-Петербург

Сопредседатели редакционного совета

Рубин Ю. Б., докт. экон. н., проф., чл.-корр. РАО, ректор МФПУ «Синергия», зав. кафедрой Теории и практики конкуренции

Мешалкин В. П., докт. техн. н., проф., чл.-корр. РАН, Директор Института логистики ресурсосбережения и технологической инноватики, РХТУ им. Д. И. Менделеева

Члены редакционного совета

Амбросов Н. В., докт. экон. н., проф., зав. кафедрой Информатики и кибернетики, БГУЭФ (Иркутск)

Багриновский К. А., докт. экон. н., проф., зав. лабораторией Имитационного моделирования, ЦЭМИ РАН

Бендиков М. А., докт. экон. н., проф., зав. кафедрой ИУИМ МФПУ «Синергия», вед. научн. сотр. ЦЭМИ РАН

Брекис Эд., Dr. Oes., доцент, зав. кафедрой Эконометрики и бизнес-информатики, факультет Экономики и Управления, Латвийский Университет, Рига, Латвия

Бугорский В. Н., канд. экон. н., проф., кафедра ВТиП, СПбГЭУ

Волкова В. Н., докт. экон. н., проф., кафедра Информационных систем в экономике и менеджменте, СПбГПУ

Дик В. В., докт. экон. н., проф., зав. кафедрой Информационного менеджмента и электронной коммерции МФПУ «Синергия»

Диго С. М., канд. экон. н., проф., отв. за работу с авторизованными учебными центрами и образовательными учреждениями, Компания «1С», Москва

Дли М. И., докт. техн. н., проф., зав. кафедрой МИТЭ, зам. директора Филиала НИУ «МЭИ» в Смоленске

Дови'В., Dr of Physics, Полный профессор теории развития процессов, Университет Генуи, Италия

Клемеш Йржи, Dr, проф., Факультет IT Исследовательского института химии и инженерии процессов, Паннония Университет, Веспрем, Венгрия

Козлов В. Н., докт. техн. н., проф., зав. кафедрой Системного анализа и управления, СПбГПУ

Коршунов С. В., канд. техн. н., проф., проректор МГТУ им. Н. Э. Баумана

Милош М., PhD, проф., зам. директора Института компьютерных наук, Люблинский Технологический Университет, Люблин, Польша

Павловский Ю. Н., докт. физ.-мат. н., проф., чл.-корр. РАН, зав. отделом Имитационных систем, ВЦ им. А. А. Дородницына РАН

Потёмкин А. И., докт. техн. н., зав. кафедрой Корпоративного управления и электронного бизнеса, РГУТиС, Московская обл. (п. Черкизово)

Праузелло Фр., PhD, Полный профессор международной экономики, директор Департамента Экономики и Финансов, Университет Генуи, Италия

Пузанков Д. В., докт. техн. н., проф., зав. кафедрой Вычислительной техники, СПбГЭУ «ЛЭТИ» им. В. И. Ульянова

Пуйджанер Л., Dr, проф., директор Центра процессов и технологий воздействия на окружающую среду, Политехнический Университет Каталонии, Барселона, Испания

Росс Г. В., докт. экон. н., докт. техн. н., проф., зам. директора ВНИИ ПВТИ

Сухомлин В. А., докт. техн. н., проф., зав. лабораторией Открытых информационных технологий, факультет ВМК, МГУ им. М. В. Ломоносова

Халин В. Г., докт. экон. н., проф., зав. кафедрой Информационных систем в экономике, Экономический факультет СПбГУ

Хубаев Г. Н., докт. экон. н., проф., зав. кафедрой Экономической информатики и автоматизации управления РГЭУ (РИНХ, Ростов-на-Дону)

Чистов Д. В., докт. экон. н., проф., зав. кафедрой Информационных технологий, Финансовый университет при Правительстве РФ

Шорилов А. Ф., докт. физ.-мат. н., проф., главн. научн. сотр. Центра экономической безопасности Института экономики Уральского отделения РАН (Екатеринбург)

Штельцер Д., Dr, проф., Департамент информации и управления знаниями, Факультет Экономики, Технологический Университет Ильменау, Тюрингия, Германия

Заместители главного редактора

Власова Е. А., научная редакция МФПУ «Синергия»

Прокимов Н. Н., канд. техн. н., доцент, кафедра Информационных систем, МФПУ «Синергия»

IT-менеджмент

Управление ресурсами

О. А. Жданович

Система управления комплексом научного оборудования коллективного пользования на основе облачных технологий 5

Управление эффективностью

С. А. Мещерин, Д. Г. Какауридзе, А. И. Волков, Е. Ю. Пустозеров,

Применение онтологий для создания и эксплуатации хранилищ финансовых данных . . . 16

IT в естественных науках

Химико-физические технологии

А. В. Пыжов

Разработка программного обеспечения подсистемы реестра РАО и кадастра приповерхностных хранилищ в системе СГУК РВ и РАО 24

Н. С. Устинова, Е. В. Бурляева

Функциональное моделирование производства модулей анодного заземления для электрохимической защиты от коррозии. 29

В. А. Дударев, В. В. Масютин

Интеграция автономных источников данных для исследования свойств полупроводниковых материалов 34

IT и образование

Образовательное пространство

Пресс-релиз

Всероссийская студенческая олимпиада по направлениям «Прикладная информатика» и «Математическое обеспечение и администрирование информационных систем» . . 39

Технологии обучения

В. С. Лаврентьев, Н. Н. Прохимнов, К. М. Сергеев

Ресурсосберегающее тестирование знаний на основе облачных технологий 41

Инструментальные средства

Эффективные алгоритмы

И. С. Калинин

Алгоритм поиска приближенной композиционной модели Липшиц-ограниченной сюръективной функции. 49

М. А. Баранов

Параллельная версия жадного алгоритма кластеризации. 58

Сетевые технологии

В. А. Бобков, А. С. Черкашин

Обработка и визуализация пространственных данных на гибридном вычислительном кластере 70

Simulation

Теория и практика

В. В. Девятков

Некоторые вопросы развития методологии имитационных исследований 81

Лаборатория

Исследование процессов и систем

Ю. Г. Древис, М. С. Сергеев

Обнаружение скрытых периодичностей методом вейвлет-анализа 89

Системы поддержки принятия решений

Е. Д. Стрельцова, А. И. Бородин, С. В. Фурсов

Инструментарий стратегического управления промышленным предприятием 95

В. И. Карпов, С. А. Корзунов, С. Е. Грачева

Оптимизация календарного плана выполнения комплекса взаимосвязанных работ в системе поддержки принятия решений 101

Точка зрения

Методология науки

Л. С. Болотова, А. П. Новиков, А. А. Никишина

Архитектура оболочек программных систем с порождающими механизмами (ч. 2) 109

История специальности

Становление информатики

К 85-летию Р. С. Гиляревского

Информатика вчера, сегодня, завтра 131

К 45-летию кафедры №29

Национального исследовательского ядерного университета «МИФИ»

Управляющие интеллектуальные системы — знаменательная дата. 140

О. А. Жданович, канд. техн. наук, доцент, заместитель директора по развитию Научно-исследовательского института химических реактивов и особо чистых химических веществ, г. Москва, oljan@inbox.ru

Система управления комплексом научного оборудования коллективного пользования на основе облачных технологий

Поставлена задача формирования системы управления коллективным использованием оборудования научно-производственного предприятия. Предложена теоретико-множественная модель элементов производственной системы. На основе частного облака типа *IaaS* с облачным брокером предложена архитектура информационной системы, обеспечивающая поддержку управления в условиях информационных воздействий и позволяющая принимать управленческие решения в динамичных условиях осуществления деятельности.

Ключевые слова: система управления, облачный брокер, научное оборудование, коллективное использование.

Введение

Научно-технический прогресс, интеграция России с высокоразвитыми промышленными странами привели к формированию в нашей стране производства и рынка наукоемкой высокотехнологичной продукции. Значительная часть этого рынка принадлежит *научно-исследовательским предприятиям* (НПП) различного профиля. НПП возникали, как правило, на основе профильных научно-исследовательских институтов или по кооперации с ними. Многие НПП относятся к сфере малого и среднего бизнеса, их *основная деятельность* (ОД) как по масштабам, так и по содержанию характеризуется как *единичное или мелкосерийное производство* (ЕМП), что привносит соответствующую специфику в задачи управления [4].

Следует подчеркнуть, что в современных условиях как государственным, так и негосударственным (в особенности) НПП приходится самим решать на основе самокупаемости все вопросы, связанные с ОД, а также разнообразные задачи ее обеспечения.

При этом нужно иметь в виду, что основные *бизнес-процессы* (БП), определяемые профилем деятельности, — научные исследования и высокотехнологичное производство, как правило, жестко регламентированы, результаты исследований и производимая продукция подлежат сертификации. Кроме того, деятельность контролируется соответствующими отраслевыми надзорными структурами и им подотчетна.

Выполнение научных исследований в интересах высокотехнологичного производства — неотъемлемая часть деятельности НПП; соответствующие БП обеспечиваются комплексом специального и универсального оборудования, которое может быть ресурсом двойного назначения: оно может применяться как в научных исследованиях, так и в производстве продукции. В связи с этим управление всеми БП использования таких ресурсов является достаточно специфическим. Важно отметить, что формирование и использование парка оборудования являются практически в полном объеме прерогативой менеджмента предприятия, что требует соответствующей методологической основы [1, 3].

Теоретико-множественная модель ресурсов

Каждая единица научно-производственного оборудования (НПО) может выполнять определенный набор операций O_j и использоваться в различных производственных режимах Rg_k ; различные сочетания операций и режимов определяют номенклатуру услуг N , которые может оказать данное НПО, причем процесс выполнения каждой услуги имеет соответствующую длительность производственного цикла (ДПЦ), определяемую сочетанием используемых операций и режимов. Таким образом, множество N_i для каждого i -го НПО является отображением кортежа операций и режимов

$$\left\{ O_j^M, Rg_k^M \mid j = \overline{1, I^O}; k = \overline{1, I^{Rg}} \right\}_i \xrightarrow{I^N} N_i = \left\{ N_n, ДПЦ_n, D^N \mid n = \overline{1, I^N} \right\}_i, i = \overline{1, I^M}, \quad (1)$$

где I^M — количество единиц НПО; I^O — количество операций; I^{Rg} — количество производственных режимов его использования; I^N — количество видов (номенклатура) услуг, производимых на i -м НПО; D^N — множество атрибутов услуг (наименование, описание, порядок выполнения и т. д.).

Кроме того, каждая единица НПО характеризуется совокупной стоимостью владения (ССВ) или TCO (*Total Cost of Ownership*, англ.); TCO_i представляет собой множество

$$TCO_i = \left\{ C_j, D^C \mid j = \overline{1, I^C} \right\}, i = \overline{1, I^M}, \quad (2)$$

где I^C — количество составляющих; D^C — множество атрибутов составляющих TCO_i (наименование, описание и т. д.).

Состав TCO_i определяется принятой моделью; так, по аналогии с [5] может использоваться детальная модель, позволяющая учесть особенности жизненного цикла (ЖЦ) НПО как изделия (рис. 1).

В этой модели содержание составляющих достаточно очевидно; можно только отметить, что затраты, непосредственно связанные с использованием НПО при оказании услуг (зарплата персонала, амортизация, содержание производственных помещений, технологические затраты энергии и материалов и т. д.), входят в состав элемента «Использование».

Величина ССВ — это сумма всех затрат, связанных с использованием каждого НПО в течение всего его ЖЦ, т. е.

$$C_{TCO_i} = \sum_{i=1}^{I^C} C_{ij}, i = \overline{1, I^M}. \quad (3)$$



Рис. 1. Модель совокупной стоимости владения

При определении величины суммарных затрат $C_{TCO_i}, i = \overline{1, I^M}$ необходимо учитывать, что составляющие $C_{ij}, i = \overline{1, I^M}, j = \overline{1, J^C}$ изменяются в течение ЖЦ и зависят от режима использования НПО. Так, в модели, изображенной на рис. 1, видно, что затраты в стадиях консалтинг, проектирование, изготовление, внедрение и освоение имеют место в основном в начале ЖЦ; затраты по функциям поддержка, обслуживание, сопровождение, испытания и информационная поддержка по этапам жизненного цикла могут возникать случайно в течение ЖЦ. Затраты, связанные с использованием НПО, также являются случайными и зависят от режима использования НПО. В соответствии с этим для составляющих справедливо выражение

$$C_i = \int_{T_n}^{T_k} C_i(t) dt, i = \overline{1, I^C}, \quad (4)$$

где T_n, T_k — начало и конец ЖЦ соответственно.

Тогда суммарные затраты

$$C_{TCO_i} = \sum_{l=1}^{j^c} \int_{T_n}^{T_k} C_l(t) dt = \int_{T_n}^{T_k} \sum_{l=1}^{j^c} C_l(t) dt = \int_{T_n}^{T_k} \delta C_{TCO_i}(t) dt, i = \overline{1, I^M}, \quad (5)$$

где $\delta C_{TCO_i}(t)$ — суммарные затраты по всем составляющим ССВ в текущий момент времени, их величина ограничена доступными в данный момент средствами, т. е. $\delta C_{TCO_i}(t) \leq \delta C_{TCO_i}(t)_{\text{доп}}$.

В этих условиях модель НПО представляет собой множество

$$M_i = \{N_i, TCO_i, D^M\}, i = \overline{1, I^M}, \quad (6)$$

где в дополнение к ранее принятым обозначениям D^M — множество атрибутов НПО (наименование, место расположения, условия применения и т. д.).

В соответствии с принятой моделью каждая единица НПО может оказать/произвести некоторый максимальный, или потенциальный, объем услуг/продукции $P_{\text{пот}i}$, представляющий собой производственный ресурс соответствующей единицы НПО, и получить при его реализации соответствующий доход; величина $P_{\text{пот}i}$ является отображением множества M_i :

$$M_i \xrightarrow{f^P} P_{\text{пот}i}, \quad (7)$$

где f^P — функция отображения. Таким образом, на оказание услуг в объеме $P_{\text{пот}i}$ затрачиваются средства в объеме C_{TCO_i} . Соотношение $P_{\text{пот}i}$ и C_{TCO_i} определяет потенциальную экономическую эффективность применения данного НПО в деятельности НПП, т. е.

$$\Theta_{\text{пот}i} = \frac{P_{\text{пот}i}}{C_{TCO_i}}. \quad (8)$$

Комплекс НПО (КНПО), имеющийся в распоряжении НПП, — множество

$$KM = \{M_i | i = \overline{1, I^M}\} = \bigcup_{i=1}^{I^M} M_i, \quad (9)$$

объединяет все НПО во всех отношениях — и по операциям, и по режимам, и по номенклатуре услуг [6]. Так, автономными НПО в комплексе выполняется множество NM услуг N_i , т. е.

$$N^M = \bigcup_{i=1}^{I^M} N_i, i = \overline{1, I^M}, \quad (10)$$

где учитывается, что разные НПО могут оказывать одни и те же услуги. Кроме того, в составе комплекса, кроме множества NM , могут быть оказаны также некоторые комплексные услуги N_q , для выполнения которых требуется совместное использование нескольких различных НПО, т. е. рабочих центров (РЦ); множество таких услуг

$$N^Q = \{N_q, ДПЦ_q, D^Q | q = \overline{1, I^Q}\}, \quad (11)$$

где I^Q — количество комплексных услуг; D^Q — множество атрибутов комплексных услуг (наименование, описание, порядок

выполнения и т. д.). Таким образом, номенклатура услуг КНПО является объединением как услуг отдельных НПО, так и комплексных услуг, т. е.

$$N_{KM} = N^M \cup N^Q. \quad (12)$$

Суммарный ресурс КНПО $P_{\text{пот}\Sigma}$ — сумма ресурсов отдельных НПО, т. е.

$$P_{\text{пот}\Sigma} = \sum_{i=1}^M P_{\text{пот}i} \quad (13)$$

это максимальный, или потенциальный, объем услуг/продукции, который может быть выполнен на этом комплексе; он является ресурсом предприятия.

Совокупная стоимость владения КНПО $C_{TCO\Sigma}$ отражает суммарные затраты предприятия на создание и использование комплекса, т. е.

$$C_{TCO\Sigma} = \sum_{i=1}^M C_{TCOi}, \quad (14)$$

соотношение $P_{\text{пот}\Sigma}$ и $C_{TCO\Sigma}$ определяет потенциальные экономические показатели применения КНПО в деятельности НПП, в частности его эффективность

$$\Theta_{\text{пот}\Sigma} = \frac{P_{\text{пот}\Sigma}}{C_{TCO\Sigma}}. \quad (15)$$

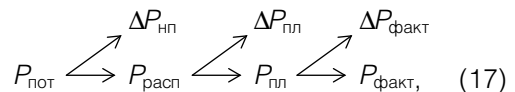
Постановка задачи управления ресурсами

В практике применения НПО нужно управлять ресурсами в полном соответствии с принципами производственного менеджмента. Потенциальный объем ресурсов $P_{\text{пот}\Sigma}$, в конце концов — парк НПО, или КНПО, определяется при стратегическом планировании деятельности (эта задача, а также распределение ресурсов по структуре НПП выходят за рамки настоящей работы) [7, 8]. При этом важно учитывать, что невозможно постоянно полностью использовать потенциальные ресурсы. Обычно в распоряжении менеджмента бывает только какая-то их часть — *располагаемые ресурсы* $P_{\text{расп}} < P_{\text{пот}}$, что определяется степенью освоения НПО. Это означает, что использо-

вание НПО нужно планировать и оценивать их использование в пределах имеющихся ресурсов. При планировании использования НПО на основе регламентов выполнения работ и ДПЦ формируются графики использования выделенных ресурсов и определяются потребности в ресурсах $P_{\text{пл}}$, причем $P_{\text{пл}} < P_{\text{расп}}$. Это означает, что для выполнения работ должны быть такие ресурсы, чтобы все работы выполнялись по графику в заданные сроки. Величина фактического потребления ресурсов при выполнении работ $P_{\text{факт}}$ может быть как больше, так и меньше плановой, что определяется соотношением качества планирования и производственного менеджмента. Таким образом, при организации управления ресурсами в качестве ее основы строится следующая цепь ключевых показателей:

$$P_{\text{пот}} \rightarrow P_{\text{расп}} \rightarrow P_{\text{пл}} \rightarrow P_{\text{факт}}. \quad (16)$$

На каждом из переходов по этой цепи имеют место потери ресурсов. На этапе освоения — это *недоступные потери* $\Delta P_{\text{нп}} = P_{\text{пот}} - P_{\text{расп}}$, которыми менеджмент не может воспользоваться; *потери при планировании* $\Delta P_{\text{пл}} = P_{\text{расп}} - P_{\text{пл}}$ характеризуют качество плана и по существу являются резервом в управлении; величина $\Delta P_{\text{факт}} = P_{\text{пл}} - P_{\text{факт}}$ при $\Delta P_{\text{факт}} < 0$ отражает перерасход, при $\Delta P_{\text{факт}} > 0$ — экономию ресурсов по отношению к плановым показателям при выполнении работ. Таким образом, сводная картина в задаче управления ресурсами имеет следующий вид:



характеристики использования ресурсов представлены в табл. 1.

Показатели эффективности управления ресурсами могут быть различными, но в соответствии с неизбежными потерями всегда будет $\Theta_{\text{пот}} > \Theta_{\text{расп}} > \Theta_{\text{пл}}$. Эффективность — ключевой показатель качества управления. Для ее повышения нужно увеличивать объем используемых ресурсов, т. е. объем ока-

Таблица 1

Характеристики использования ресурсов

Наименование показателя	Показатель	
	плана	менеджмента
Объем	$P_{пл}$	$P_{факт}$
Резерв	$\Delta P_{пл} = P_{расп} - P_{пл}$	$\Delta P_{факт} = P_{пл} - P_{факт}$
Загрузка	$Z_{пл} = \frac{P_{пл}}{P_{расп}}$	$Z_{факт} = \frac{P_{факт}}{P_{расп}}$
Качество	$K_{пл} = \frac{\Delta P_{пл}}{P_{расп}}$	$K_{мен} = \frac{\Delta P_{факт}}{P_{пл}}$
Напряженность	$H_{пл} = \frac{\Delta P_{пл}}{P_{пл}}$	$H_{мен} = \frac{\Delta P_{факт}}{P_{факт}}$
Использование резерва	$I_{рез} = \frac{\Delta P_{факт}}{\Delta P_{пл}}$	

занных услуг / выпущенной продукции, или снижать уровень потерь; суммарные потери равны [7]

$$\Delta P_{\Sigma} = \Delta P_{нп} + \Delta P_{пл} + \Delta P_{факт} = P_{пот} - P_{факт} \quad (18)$$

откуда следуют пути их уменьшения, т. е.

$$\Delta P_{\Sigma} \xrightarrow{\Delta P_{нп}, \Delta P_{пл}, \Delta P_{факт}} \min. \quad (19)$$

Для повышения эффективности также нужно снижать ССВ, т. е.

$$C_{ТСО\Sigma} \xrightarrow{C_{ТСОi}, i=1, M} \min, \quad (20)$$

где, в свою очередь,

$$C_{ТСОi} \xrightarrow{C_{ij}, j=1, J^C} \min. \quad (21)$$

Общая характеристика условий задачи управления

По существу, задачей управления ресурсами является удовлетворение потребностей НПП в ресурсах в заданные сроки по их номенклатуре и объему. Сроки определяются программой выполнения исследований и планом производства продукции,

номенклатура и объем по позициям задаются характером выполняемых БП. При всей очевидности и типовом характере такой постановки задачи на практике ее решение в условиях НПП оказывается сложным.

Во-первых, в современных условиях деятельности НПП целесообразно формировать по возможности плоскую иерархию системы управления для повышения ее эффективности — это одно из основных направлений совершенствования менеджмента [8]. При этом подразделениям вменяются в качестве ОД определенные БП, для выполнения которых они наделяются ресурсами P_c . Правда, в связи с высокой стоимостью НПО и необходимостью его загрузки обеспечить им полностью БП всех подразделений не удастся, и подразделения кооперируются в вопросах использования тех или иных НПО. Для этого центру приходится делегировать полномочия и предоставлять возможно более широкие права в области принятия решений производственным подразделениям, в том числе и в вопросах управления ресурсами. Это означает, что подразделения сами планируют работы, формируют график подачи ресурсов в производство, заказывают ресурсы $P_{вк}$ по внут-

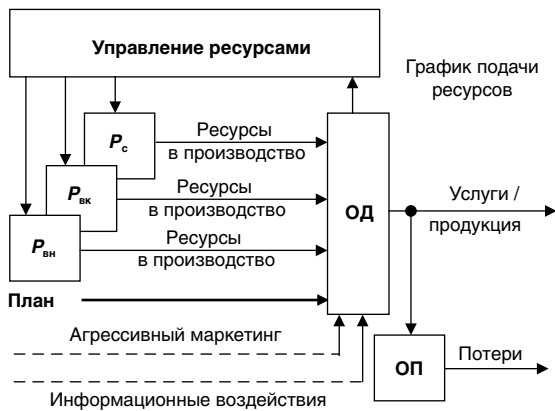


Рис. 2. Схема управления ресурсами

ренней кооперации другим подразделениям — владельцам ресурсов, а также ресурсы $P_{вн}$ у сторонних партнеров и получают их в свою ОД, контролируют их потребление и потери в исследованиях и производстве. Как следствие, подразделения принимают на себя все связанные с этими процессами риски (рис. 2).

Во-вторых, потребление различных ресурсов и соответственно потребности в них характеризуются высоким уровнем неопределенности, которая вытекает из особенностей выполнения научных исследований в условиях единичного или мелкосерийного производства, что существенно осложняет управление ресурсами и снижает их эффективность.

В-третьих, реальные возможности производственных подразделений в вопросах управления за пределами своей бизнес-сферы ограничены, обеспечить высокое качество процессов получения ресурсов, как правило, не удастся; снижение рисков срыва графика выполнения работ и плана производства достигается за счет снижения показателей напряженности и увеличения нормы страховых запасов ресурсов. В связи с этим возникают потери в виде невыполненных исследований и невыпущенной продукции [4].

В-четвертых, в настоящее время деятельность НПП осуществляется в условиях существенной нестабильности рынков, что

приводит к повышению риска срыва получения ресурсов по внешним для НПП причинам. Внешние факторы — *информационные воздействия*, в частности *агрессивный маркетинг* поставщика ресурсов, — могут побуждать подразделения НПП увеличивать страховой объем ресурсов по сравнению с обоснованным объемом [2].

При этом проявляется действие так называемого «*принципа дефицита*» из теории активных систем: *ценность чего-либо позитивного в наших глазах существенно увеличивается, если оно становится недоступным* [2, с. 163]. В этих условиях стремление снизить риск остановки работ из-за возникновения дефицита по внешним причинам приводит к стремлению увеличить ресурс P_c парка НПО, находящегося в распоряжении подразделения, и к завышению объема заявок на поставки ресурсов по внутренней $P_{вк}$ и по внешней $P_{вн}$ кооперации по сравнению с нормативными объемами. Как следствие, при сверхнормативном увеличении объема ресурсов могут существенно снижаться показатели использования ресурсов НПО, в частности их эффективность, поскольку возрастут соответствующие составляющие затрат в составе ССВ.

Блок-схема алгоритма управления ресурсами представлена на рис. 3.

В соответствии с рис. 2 и 3 целевая функция управления ресурсами в форме минимизации потерь может быть описана следующим выражением:

$$\begin{aligned} \Delta P_{оп} &= \Delta P_{пл} + \Delta P_{факт} = \\ &= \Phi \{Z_c; Z_{вк}; Z_{вн}; Q_c; Q_{вк}; Q_{вн}; \Delta G_c; \\ &\Delta G_{вк}; \Delta G_{вн}; C_{TCO}; K_{нп}; K_{п}\} \rightarrow \min, \end{aligned} \quad (22)$$

где введены следующие обозначения:

$\Delta P_{оп}$ — потери ресурсов в оперативном управлении;

$Z_c; Z_{вк}; Z_{вн}$ — заявки на ресурсы собственного НПО, а также по внутренней и внешней кооперации соответственно; заявки подают все подразделения независимо друг от друга в соответствии с планом и графиком работ;

С. А. Мещерин, аспирант Московского физико-технического института, sergey.metcherin@gmail.com

Д. Г. Какауридзе, аспирант Московского физико-технического института, kakauridze.d@gmail.com

*А. И. Волков, генеральный директор ЗАО «РДТех Разумные информационные технологии»,
аспирант Московского физико-технического института, Anatoli.Volkov@rdtex.ru*

*Е. Ю. Пустозеров, руководитель направления ЗАО «РДТех Разумные информационные технологии»,
Москва, Evgeny.Pustozеров@rdtex.ru*

Применение онтологий для создания и эксплуатации хранилищ финансовых данных¹

В статье представлен аналитический обзор существующих подходов к описанию хранилищ данных и технологии загрузки гетерогенных данных из внешних источников. В качестве инструмента для работы с гетерогенными данными используется OWL-онтология. Описано использование OWL-онтологии на этапе загрузки внешних данных и на этапе получения аналитических данных из хранилища данных.

Ключевые слова: хранилища данных, СУБД, онтологии.

Введение

Хранилище данных (ХД) на протяжении многих лет признается надежной технологией для хранения и анализа больших объемов структурированных данных. Особенно удобно использование хранилищ данных в сферах деятельности с повышенными требованиями к отказоустойчивости, защите и средствам интеллектуального анализа данных. Например, в банковском деле, биллинговых системах, страховых компаниях.

Однако в последнее время в связи с лавинообразным ростом объемов открытых данных [1] в сети Интернет (социальные сети, блоги, мультимедийный контент, семантически размеченные биржевые данные) все большую актуальность приобретает использование гетерогенных внешних данных в хранилищах данных. При этом в силу разнородности данных проводить полную интеграцию внешних данных в хранилище (пол-

ноценный ETL-процесс) становится нецелесообразным. Можно показать связи таких гетерогенных данных друг с другом и с существующими данными хранилища с помощью языка семантической разметки, а сами внешние данные — хранить без изменений и создания классических реляционных отношений. Далее будет дан обзор технологии хранилищ данных и описано использование семантической разметки совместно с хранилищами данных.

Хранилища данных и витрины данных

Хранилище данных — это база данных, предоставляющая возможность интегрировать данные из различных источников и предназначенная для аналитической обработки и составления отчетов. При решении задач организационного управления ХД позволяет собирать исторические данные и проводить бизнес-анализ, способствующий увеличению эффективности бизнес-процессов в компании [2]. ХД включает в себя базу данных (БД), средства обработки и клиентскую аналитику.

¹ Работа выполнена при поддержке РФФИ, проект № 14-07-31288.

Хранилища данных обладают следующими свойствами:

- предметная ориентации — данные в ХД объединяются по предметным категориям;
- интегрированность — данные в ХД поступают из различных источников, при этом устраняются конфликты имен, обеспечивается целостность данных;

- неизменность данных — данные в ХД не корректируются, они сохраняются там единожды и в дальнейшем доступны только для чтения, это соответствует назначению ХД — возможности проведения анализа данных, актуальных на любой заданный промежуток времени;

- временная зависимость — специфика бизнес-анализа — изменение данных во времени, поэтому в ХД могут храниться большие объемы исторических данных.

Далее рассмотрим характерные особенности хранилищ данных, которые отличают их от OLTP-систем.

- Производительность. В OLTP осуществляются только predetermined операции, в ХД — произвольные, поэтому производительность предсказать сложно.

- В ХД загрузка данных производится регулярно с заданной периодичностью (ETL-процесс) и в больших объемах, в OLTP — в виде относительно небольших по объему согласованных транзакций. Как правило, ХД обеспечивает высокую актуальность данных, но с некоторой задержкой после их возникновения в автоматизированной системе.

- Для ХД характерна схема данных «звезда» (рис. 1), или «снежинка», способствующая оптимизации производительности запросов по нескольким «измерениям». В OLTP-системах обычно используется нормализованная схема, рассчитанная на обеспечение оптимизации согласованных DML-операций (update, insert, delete).

- Типичный запрос. В ХД запрос подразумевает обработку $\approx 10^5$ – 10^6 строк. Пример запроса: «Найти суммарные продажи за месяц по всем клиентам». OLTP-запросы, как правило, более специфичны и подразумевают обработку нескольких строк из связан-

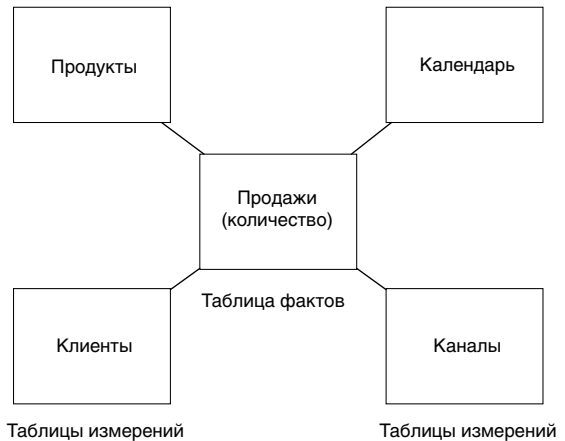


Рис. 1. Схема «звезда», типичная для ХД

ных таблиц. Пример запроса: «Извлечь текущий номер заказа для данного клиента».

- Исторические данные. OLTP-системы очень часто хранят данные за несколько дней или недель для выполнения текущих транзакций, в ХД — сведения для бизнес-анализа, полученные за месяцы и годы.

Архитектура ХД изображена на рис. 2. Рассмотрим назначение представленных на схеме блоков.

Предварительные данные — предварительно вычисленные значения (например, агрегации за месяц), увеличивающие производительность. *Данные из таблиц* — данные, например, из тех же OLTP БД.

Метаданные — формат данных для согласования источников данных, периодичности пополнения (облегчение стандартизации источников данных). Часто для подготовки данных перед загрузкой используется промежуточная область — *область предварительной обработки*. Такой подход реализован в ХД Oracle Data Warehouse (DWH).

Витрины данных — множество тематических баз данных, содержащих информацию, относящуюся к отдельным аспектам деятельности организации. Архитектура ХД с витринами данных изображена на рис. 3.

Применение витрин данных имеет следующие преимущества:

- аналитики видят и работают только с теми данными, которые им реально нужны;

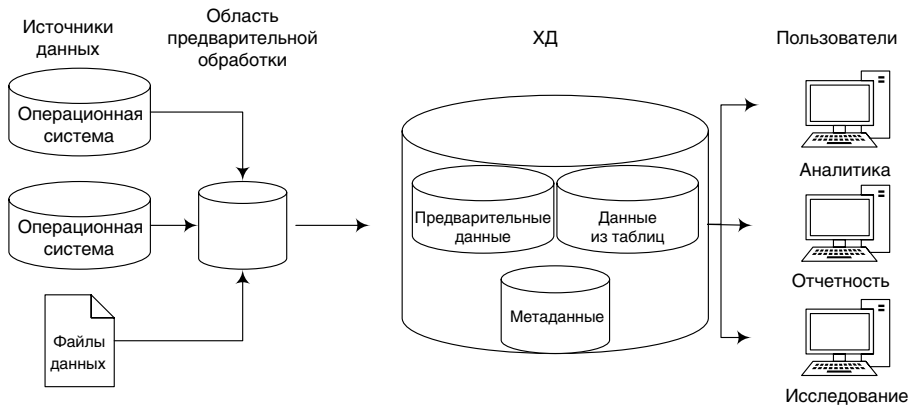


Рис. 2. Базовая архитектура хранилища данных

- целевая БД максимально приближена к конечному пользователю;
- витрины данных обычно содержат тематические подмножества заранее агрегированных данных, их проще проектировать и настраивать;
- для реализации витрин данных не требуется высокопроизводительная вычислительная техника.

Виды хранилищ данных

В отношении архитектуры существует два основных подхода к построению хранилищ данных. Это так называемая корпоративная информационная фабрика (Corporate Information Factory, сокращенно CIF) Билла Инмона (Bill Inmon), описанная в [3, 4], и ХД с архитектурой шины (Data Warehouse Bus, сокр. BUS) Ральфа Кимболла (Ralph Kimball), описанная в [5]. На рисунке 4 представлен

подход, используемый в хранилищах данных с архитектурой CIF.

Работа такого хранилища начинается с процесса извлечения данных из источников, например транзакционных систем банка. После этого загружается реляционная база данных с третьей нормальной формой, содержащая атомарные данные. Получившееся нормализованное хранилище используется для того, чтобы наполнить информацией дополнительные репозитории данных, подготовленных для анализа: специализированные хранилища для изучения и «добычи» данных (Data Mining), а также витрины данных.

В качестве отличительных характеристик подхода Билла Инмона к архитектуре хранилищ данных можно назвать следующие:

- использование реляционной модели организации атомарных данных и многомерной — для организации суммарных данных;

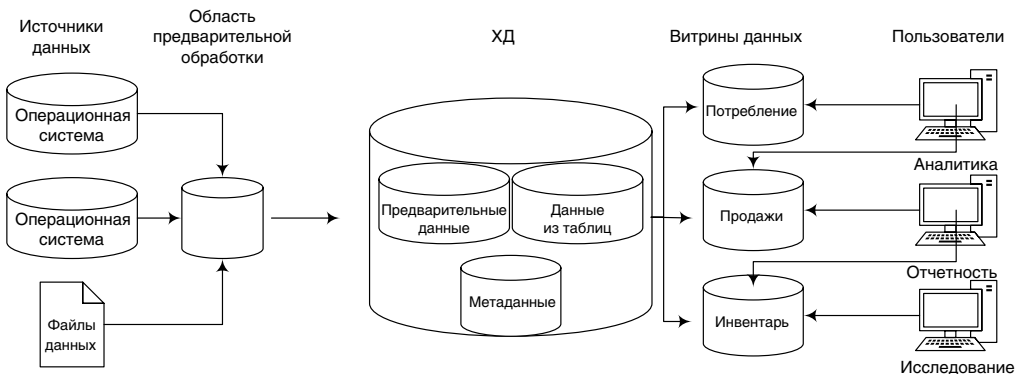


Рис. 3. Архитектура ХД с витринами



Рис. 4. Нормализованное ХД с многомерными витринами итоговых данных (CIF)

- использование итеративного или «спирального» подхода при создании больших хранилищ данных, т. е. «строительство» хранилища не сразу, а по частям. Это позволяет при необходимости вносить изменения в небольшие блоки данных или программных кодов и избавляет от необходимости перепрограммировать значительные объемы данных в хранилище. То же самое можно сказать и о потенциальных ошибках: они также будут локализованы в пределах сравнительно небольшого массива без риска испортить все хранилище;

- использование третьей нормальной формы для организации атомарных данных, что обеспечивает высокую степень детальности интегрированных данных и соответственно предоставляет корпорациям широкие возможности для манипулирования ими и изменения формата и способа представления данных по мере необходимости;

- ХД — это проект корпоративного масштаба, охватывающий все отделы и обслуживающий нужды всех пользователей корпорации;

- ХД — это не механическая коллекция витрин данных, а физически целостный объект.

В модели Ральфа Кимболла (рис. 5) первичные данные преобразуются в информацию, пригодную для использования, на этапе подготовки данных. При этом обязательно

принимается во внимание требования к скорости обработки информации и качеству данных. Как и в модели Билла Инмона, подготовка данных начинается во время скоординированного извлечения данных из источников. Ряд операций совершается централизованно, например поддержание и хранение общих справочных данных, другие действия могут быть распределенными.

Область представления структурирована с использованием нескольких измерений, при этом она может быть централизованной или распределенной. Многомерная модель ХД содержит ту же атомарную информацию, что и нормализованная модель (см. подход Билла Инмона), но информация структурирована по-другому для облегчения ее использования и выполнения запросов. Эта модель включает как атомарные данные, так и обобщающую информацию (агрегаты в связанных таблицах или многомерных кубах) в соответствии с требованиями производительности или многомерной организации данных. Запросы в процессе выполнения обращаются ко все более низкому уровню детализации без дополнительного перепрограммирования со стороны пользователей или разработчиков приложения.

В отличие от подхода Билла Инмона, многомерные модели строятся для обслуживания бизнес-процессов (которые, в свою оче-

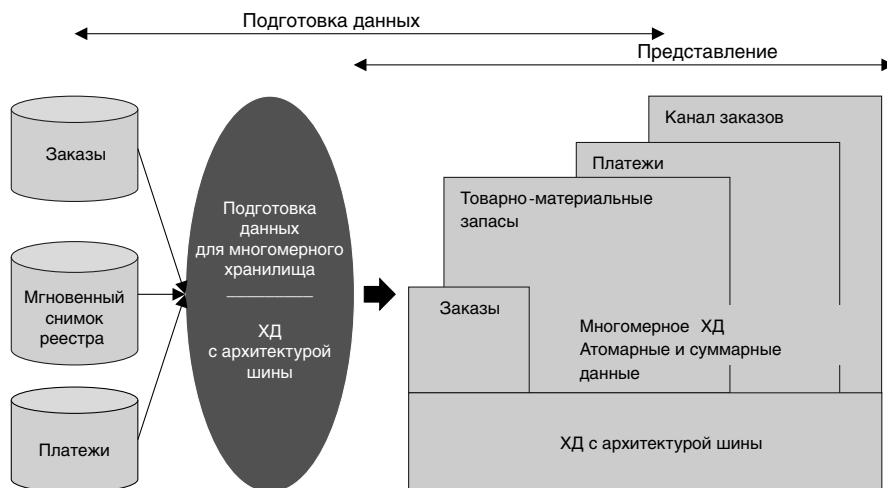


Рис. 5. Подход Кимбалла, многомерное ХД

редь, связаны с бизнес-показателями или бизнес-событиями), а не бизнес-отделов. Например, данные о заказах, которые должны быть доступны для общекорпоративного использования, вносятся в многомерное ХД только один раз, в отличие от CIF-подхода, в котором их пришлось бы трижды копировать в витрины данных отделов — маркетинга, продаж и финансов. После того как в хранилище появляется информация об основных бизнес-процессах, консолидированные многомерные модели могут выдавать их перекрестные характеристики. Матрица корпоративного хранилища данных с архитектурой шины выявляет и усиливает связи между показателями бизнес-процессов (фактами) и описательными атрибутами (измерениями).

Первое существенное отличие между этими архитектурами — различные подходы к построению баз данных, составляющих основу хранилища. Если Ральф Кимболл использует многомерную организацию баз данных (dimensional data bases) с так называемой архитектурой «звезда» как на стадии подготовки, так и презентации данных, то Билл Инмон комбинирует два подхода. В его модели атомарные данные организованы в реляционные базы и содержатся в нормализованном хранилище данных, причем суммарные данные доступны для

использования через специализированные хранилища, средства data mining и OLAP. Только зависимые витрины данных организованы с помощью многомерных моделей, как и у Ральфа Кимболла.

Таким образом, архитектуры отличаются лишь способами обращения с атомарными данными: их многомерной организацией у Кимболла и нормализованной у Инмона [6].

Второе принципиальное отличие этих двух подходов, отчасти вытекающее из первого, — физическая организация хранилища. Если у Инмона ХД — это физически целостный реально существующий объект, то хранилище Кимболла — скорее, «виртуальный» объект. Это коллекция витрин данных, которые могут быть разобращены по нескольким измерениям.

Этими двумя основными отличиями в целом и исчерпывается принципиальная разница между моделями.

Онтологии и хранилища данных

Современное информационное пространство характеризуется в первую очередь большими объемами данных. Данные (новости, банковские сводки, мультимедиаданные) обновляются в реальном времени и имеют гетерогенную природу, т. е. данные могут нахо-