

РАЗРАБОТКА МЕТОДИКИ АДАПТИВНОГО УПРАВЛЕНИЯ ПРОЦЕССОМ ПРОЕКТИРОВАНИЯ И РАЗВИТИЯ СПЕЦИАЛИЗИРОВАННЫХ ХРАНИЛИЩ ДАННЫХ*

Предложена методика адаптивного управления процессом проектирования и развития хранилищ данных, обеспечивающая оригинальную формализацию процессов проектирования и развития модели данных, учитывая условия эксплуатации и специализированную информацию о предметной области, что позволяет повысить производительность хранилищ данных. Выполнено описание логики процесса проектирования и развития специализированного хранилища данных с использованием управляющей модели в нотации eEPC ARIS-методологии.

Ключевые слова: специализированное хранилище данных, адаптивное управление, проектирование, материализация представлений.

Адаптивное управление разработкой хранилища данных (ХД) подразумевает итерационные действия, направленные на достижение приемлемой производительности систем поддержки принятия решений. В качестве критерия оптимальности, относительно которого принимается решение, для модели адаптивного управления используется минимизация стоимости обслуживания и времени отклика, а в качестве управляемых переменных – характеристики хранилища в виде совокупности ограничений (максимальное время ожидания ответа, доступное дисковое пространство). Параметры объектов управления не могут быть однозначно определены и меняются в ходе реализации процессов проектирования и эксплуатации, происходящих в современных условиях развития.

Для повышения производительности систем поддержки принятия решений необходимо включать в модель хранилища данных итоговые агрегатные таблицы. Из-за ограничений дискового пространства и объема трудозатрат, связанных с необходимостью поддержания агрегатов в актуальном состоянии, нельзя вычислить и сохранить все возможные представления. Поэтому материализуются только избранные комбинации агрегатов, которые затем используются для более эффективного вычисления других агрегатов. Задача выбора агрегатов для материализации NP – трудная [1], число агрегатов экспоненциально зависит от количества измерений. При эксплуатации хранилища выбор представлений для материализации может основываться на информации о частоте и времени выполнения запросов. Такая информация автоматически накапливается СУБД в виде статистики по полям таблиц хранилища данных и отсутствует на этапе проектирования. При проектировании хранилища требуется разработать способы определения релевантных элементов в пространстве представлений кандидатов на материализацию, используя доступную информацию о предметной области.

Методика адаптивного управления процессом проектирования и развития специализированных хранилищ данных. Специализированное хранилище данных ориентировано на решение определенного класса аналитических

задач. Решение задач обеспечивается наличием четких описаний совокупностей методов расчета показателей предметной области. Предлагается определять релевантные элементы для включения в модель данных хранилища на основе анализа методик предметной области, переходя от содержательного описания к количественным оценкам. Такой подход позволит повысить производительность первоначального проекта специализированного хранилища, используя информацию о предметной области в отсутствие статистической. Развитие модели хранилища данных для повышения скорости доступа к информации должно соответствовать требованиям пользователей. Учет потребностей пользователей предлагается выполнять за счет включения семантических оценок методик предметной области в расчет релевантных элементов.

Предлагаемая методика построения специализированных хранилищ данных реализуется в три основных этапа.

Этап 1. Исследование предметной области: выделение измерений, показателей и методик расчета. Определение источников данных, частоты их обновления. Изучение структуры исходных данных для последующей разработки методов унификации и верификации данных. Определение релевантных данных с помощью методик расчета показателей предметной области.

Этап 2. Формирование модели хранилища данных: создание таблиц фактов и таблиц измерений. Включение в модель хранилища данных дополнительных структур, используя алгоритм выбора релевантных представлений. Формирование репозитория метаданных.

Этап 3. Эксплуатация хранилища данных и накопление статистики для последующего его развития с использованием алгоритма выбора релевантных представлений и рекомендаций автоматических средств системы управления базами данных (СУБД).

Проектирование и развитие хранилища данных может производиться с помощью инструментария одного производителя или нескольких, при этом возможные проблемы интеграции различных компонентов в работе не рассматриваются. Инструментарий хранилища обеспе-

* Работа выполнена при финансовой поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (ГК № 02.740.11.0621) и гранта ККФПН и НТД (Доп. сог. № 01/10 от 17 мая 2010 г.).

чивает загрузку, верификацию и актуализацию данных и использует определенный сервер СУБД. Применение принципа адаптации предполагает, что часть функций по получению, обработке и анализу недостающей информации о работе системы осуществляется автоматическими средствами СУБД в процессе нормальной эксплуатации хранилища данных, а не проектировщиком на предварительной стадии разработки.

Различные оптимизаторы запросов СУБД обладают специфическими свойствами, которые накладывают дополнительные требования на результат. Конкретный физический проект зависит от свойств выбранного оптимизатора. Необходимо, чтобы разрабатываемая методика проектирования и развития хранилищ данных не предполагала детального изучения работы и особенностей оптимизатора СУБД. Предлагается обеспечение методикой построения первоначальной модели хранилища данных, которая в процессе эксплуатации будет корректироваться автоматическими средствами СУБД в соответствии с внутренней реализацией конкретного сервера.

Методика адаптивного управления процессом проектирования и развития специализированных хранилищ данных основана на использовании алгоритма выбора релевантных представлений (алгоритм релевантности). Предложенный алгоритм включает определение релевантных элементов множества представлений и выбор представлений для материализации [2; 3]. Для описания логики процесса проектирования и развития специализированных хранилищ данных используется управляющая модель в нотации eEPC ARIS-методологии [4] (рис. 1). Нотация ARIS eEPC (extended Event Driven Process Chain) – расширенная нотация описания цепочки процесса, управляемого событиями, являющаяся расширением нотации IDEF3.

Управляющая модель объединяет организационную и функциональную модели, включая этапы проектирования и эффективного использования хранилища данных с возможностью его адаптивного развития. Модель отражает роли, которые играют автоматические средства оптимизации, инструментарий ХД, проектировщик и эксперт предметной области в процессе настройки и поддержки хранилища данных в актуальном и работоспособном состоянии. «Проектировщик» объединяет роли программиста и администратора хранилища данных и специалиста, проводящего исследование предметной области. «Эксперт» выступает в роли специалиста предметной области и (или) пользователя хранилища данных.

В процессе эксплуатации хранилища данных автоматическими средствами СУБД по полям таблиц хранилища собираются и обрабатываются статистические сведения, на основе которых формируются рекомендации по внесению изменений в модель данных. При поступлении новой информации предметной области или дополнительного дискового пространства, в случае отсутствия рекомендаций СУБД, решение об изменении модели хранилища данных принимается на основе выполнения алгоритма релевантности.

Действия по повышению производительности ХД выполняются до тех пор, пока структура хранилища данных не достигнет состояния, когда внесение в нее измене-

ний невозможно или не требуется (события управляющей модели «ХД работает удовлетворительно» или «ХД работает неудовлетворительно»). Такое состояние поддерживается, пока не инициализируется одно из событий, приводящее цепочку последовательного выполнения функций в действие, например, поступление дополнительного дискового пространства или появление новых требований пользователя.

Наиболее эффективная структура хранилища данных обеспечивается интеграцией автоматических средств оптимизации, действий проектировщика и пользователя в настройке и поддержке хранилища в актуальном работоспособном состоянии с учетом имеющейся информации на конкретный момент времени. Отличие от существующих подходов к проектированию хранилищ данных заключается в том, что особое внимание уделено проблеме организации данных при отсутствии статистики о работе хранилища данных.

Особенность предложенной методики отражена блоками управляющей модели, определяющими проектирование, развитие и изменение модели ХД на основе алгоритма релевантности [2; 3].

Состав и описание модели проектирования. Модель процесса проектирования и развития специализированного хранилища содержит следующие функциональные блоки.

1. Проектирование модели ХД (алгоритм релевантности).
2. Разработка методов автоматизированной загрузки.
3. Изменение модели ХД (алгоритм релевантности).
4. Получение и обработка дополнительной информации.
5. Проверка работы ХД.
6. Анализ причин неудачи.
7. Изменение настроек сбора статистики.
8. Сбор статистики.
9. Обработка статистики.

Каждый блок включает набор алгоритмов и методов для достижения целевого назначения.

Проектирование модели хранилища данных происходит общесистемными методами [5–7] и др., включая последовательное выполнение концептуального и логического этапов проектирования. Концептуальное и логическое проектирование может быть выполнено с использованием различных моделей данных [8–11], которые поддерживают реляционные структуры данных и обеспечивают простоту трансляции концептуальной модели данных в логическую [12] и др.

Предложенный в работе алгоритм релевантности, используемый для проектирования модели хранилища данных, состоит из алгоритма определения релевантности [2] и алгоритма выбора представлений для последующей материализации в информационном хранилище [1; 3; 13–15].

Выполнение функции «Разработка методов автоматизированной загрузки», вызванное наступлением события «Физическая модель данных реализована», производится разово. Последующий вызов функции не происходит до тех пор, пока не появится событие «Новые данные поступили». Иными словами, новые методы загрузки информации не требуются до тех пор, пока не поступят неучтенные ранее данные предметной области.

Хранилище данных является развивающейся средой, поэтому требуется изменение и развитие модели хранения в нем информации. Изменения, как правило, вызываются следующими событиями: первичным наполнением хранилища, периодическим поступлением новых данных, появлением новых расчетных методик предмет-

ной области, поступлением рекомендаций от автоматических средств СУБД, изменением дискового пространства для области агрегатов. После реализации модели хранилища данных и разработки средств автоматизированной загрузки выполняется наполнение хранилища. Дальнейшее развитие хранилища влечет за собой периоди-

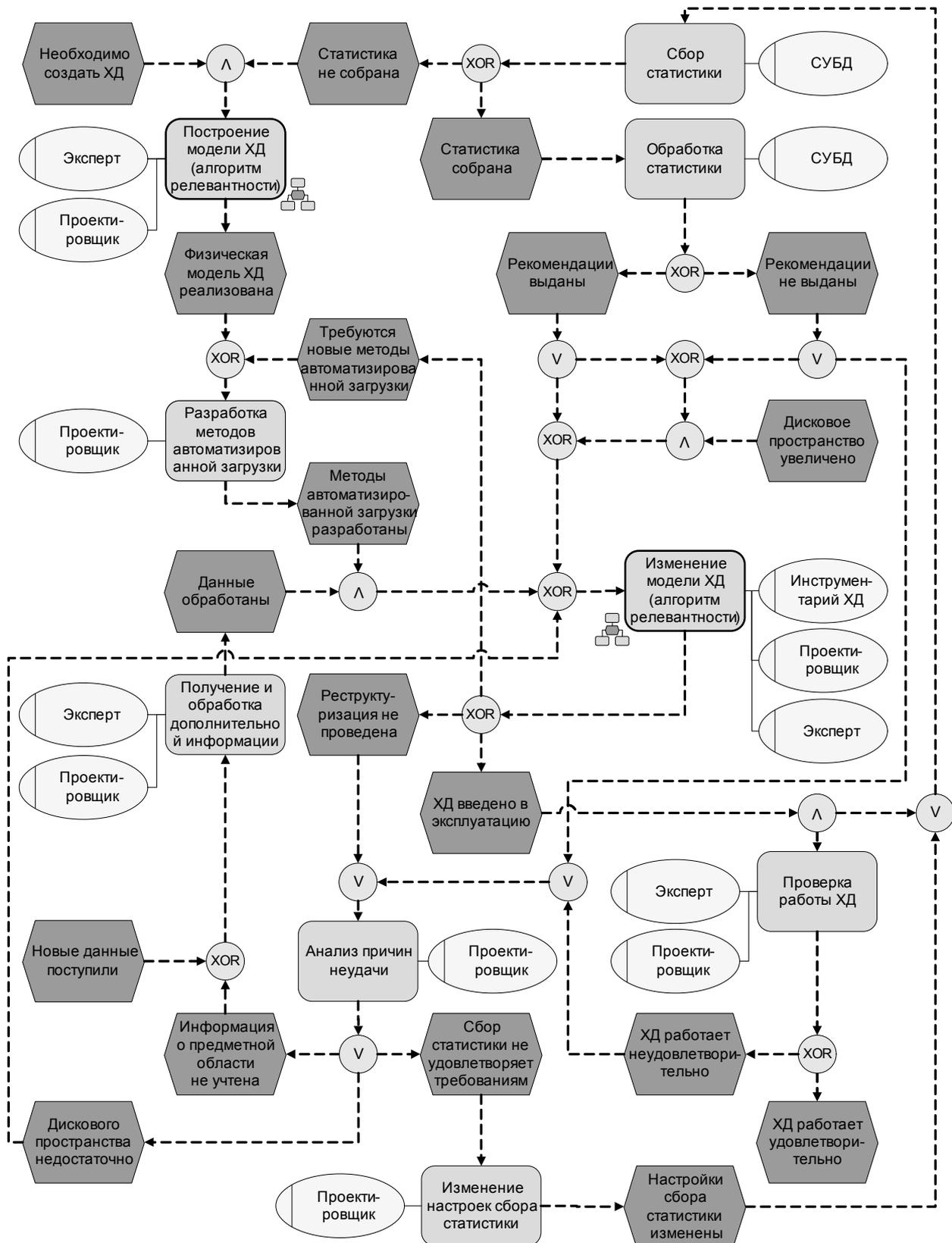


Рис. 1. Модель процесса проектирования и развития специализированного хранилища данных

ления для добавления в хранилище. Необходимо учитывать, что периодически объемы хранилища данных увеличиваются. Если для проведения актуализации данных недостаточно дискового пространства, то необходимо удалить часть таблиц с агрегатами, руководствуясь при этом статистической информацией или, при ее отсутствии, – алгоритмом релевантности [2; 3].

Декомпозиция функции «Изменение модели ХД (алгоритм релевантности)» модели проектирования и развития специализированных хранилищ данных показана на рис. 2. В случае успешного применения действий, вызванных событиями, требующими развития модели хранилища данных, результатом является измененная модель хранилища данных, готовая к дальнейшей эксплуатации. После внесения корректировок в модель хранилища данных выполняется проверка его работы.

Запланированная реструктуризация модели хранилища данных может быть не проведена по причине нехватки дискового пространства, которого недостаточно даже при удалении агрегированных данных из хранилища. В этом случае требуется дополнительное исследование ситуации, решение которой зависит от возможности поступления дискового пространства.

По результатам проверки работоспособности инициализируется функция анализа возможных причин неудачи или хранилище данных переходит в стационарное состояние удовлетворительной работы, сопровождаемое только сбором статистики. В стационарном состоянии хранилище находится до тех пор, пока не инициализируется одно из событий, требующих изменения модели или объемов данных.

Методика адаптивного управления процессом проектирования и развития специализированных хранилищ данных предполагает регулярный сбор статистики. Сбор и обработка статистики выполняются автоматически СУБД. Результаты сбора статистики используются для изменения модели хранилища данных и для изменения настроек автоматических средств анализа и сбора статистики.

Работа большого числа утилит по проведению анализа, настройки и сбору информации о базе данных автоматическими средствами СУБД может сильно понизить производительность системы, в этом случае следует уменьшить их количество или изменить способы их работы. Изменение настроек статистики может быть вызвано недостатком информации о частоте выполнения запросов. При наличии свободных системных ресурсов может быть выполнено изменение способов работы мастера настройки индексов или подключение и использование дополнительных утилит СУБД.

Существуют различные сочетания причин, в связи с которыми становится невозможным проведение изменения модели хранилища данных и обеспечивается неудовлетворительная работа. Функция «Анализ причин неудачи» выполняется проектировщиком, принимая во внимание общую картину функционирования хранилища данных. После проведения анализа разрабатываются способы устранения причин неудачи реструктуризации или неудовлетворительного функционирования ХД.

В процессе эксплуатации хранилища данных меняются требования пользователей к составу информации и

расчетным методикам, на основе которых выполнено проектирование и оптимизация модели данных. Кроме этого, могут быть выявлены особенности предметной области, не учтенные на первоначальном этапе разработки хранилища данных. Предложенная методика адаптивного управления предполагает выполнение обработки дополнительной информации о предметной области для корректировки модели хранилища данных. К множеству новых данных и расчетных методик применяется специально разработанный алгоритм, учитывающий поступление новой информации предметной области. При наличии соответствующей статистики о работе хранилища модель данных корректируется с учетом рекомендаций СУБД. Для новых данных и материализованных представлений разрабатываются средства автоматической актуализации.

Выполнено описание процесса проектирования и развития специализированных хранилищ данных на основе методики адаптивного управления процессом проектирования и развития специализированных хранилищ данных. Для описания логики процесса проектирования и развития специализированного хранилища данных используется управляющая модель в нотации eEPC ARIS-методологии. Функциональные блоки модели, определяющие проектирование и развитие модели хранилища данных на основе алгоритма выбора релевантных представлений для материализации, отражают особенность предложенной методики. Описаны функции и связи между элементами, представленные в модели. Модель обеспечивает наглядное представление процесса функционирования специализированного хранилища данных в рамках методики адаптивного управления, начиная с принятия решения о необходимости проектирования. Методика адаптивного управления процессом проектирования и развития позволяет повысить производительность специализированного хранилища за счет материализованных представлений, выбранных с использованием информации о предметной области в отсутствии статистики о работе хранилища данных. Предложенная методика не предполагает детального изучения работы и особенностей оптимизатора СУБД, обеспечивая построение первоначальной модели хранилища данных, которая в процессе эксплуатации корректируется автоматическими средствами в соответствии с внутренней реализацией конкретного сервера баз данных.

Библиографические ссылки

1. Gupta H., Afrati F. N., Kolaitis P. G. Selection of views to materialize in a data warehouse // Proc. of the 6th Intern. Conf. on Database theory // Eds. Lecture Notes In Computer Science. London : Springer-Verlag, 1997. P. 98–112.
2. Бадмаева К. В. Алгоритм оценки релевантности представлений для материализации в специализированном хранилище данных // Вестник СибГАУ. 2009. Вып. 1(22). В 2 ч. Ч. 2. С. 60–64.
3. Badmaeva K. The performance of specialized data warehouses increasing // Proc. of the IASTED Intern. Conf. on Automaton, Control and Information Technology. Novosibirsk, 2010. P. 206–210.

4. Моделирование бизнеса. Методология ARIS / М. Каменова, А. Громов, М. Ферапонтов, А. Шматалюк. М. : Метатехнология, 2001.
5. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. М. : Вильямс, 2001.
6. Kimball R., Ross M. The data warehouse toolkit: the complete guide to dimensional modeling. N. Y. : John Wiley & Sons, Inc., 2002.
7. Inmon W. H. Building the data warehouse. N. Y. : John Wiley & Sons, 1992.
8. Buzydlowski J. W., Song I., Hassell L. A framework for object-oriented on-line analytic processing // Proc. of the 1st ACM Intern. Workshop on Data Warehousing and OLAP. N. Y. 1998. P. 10–15.
9. Golfarelli M., Rizzi S. A methodological framework for data warehouse design // Proc. of the 1st Intern. Workshop on Data Warehousing and OLAP. Maryland, 1998. P. 3–9.
10. Extending the E/R model for the multidimensional paradigm / C. Sapia, M. Blaschka, G. Hööfling, B. Dinter // Proc. ER Workshop on Data Warehousing and Data Mining. Singapore, 1998. P. 105–116.
11. Tryfona N., Busborg F., Christiansen J. StarER: A conceptual model for data warehouse design // Proc. of the ACM 2nd Intern. Workshop on Data Warehousing and OLAP/Kansas City. 1999. P. 3–8.
12. Kamble A. S. A conceptual model for multidimensional data // Proc. of the Fifth on Asia-Pacific Conf. on Conceptual Modelling ; Australian Computer Society, Darlinghurst, Australia, 2008. Vol. 79. P. 29–38.
13. Baralis E., Paraboschi S., Teniente E. Materialized views selection in a multidimensional database // Proc. of the 23rd Intern. Conf. on Very Large Data Bases, Eds. Very Large Data Bases. San Francisco : Morgan Kaufmann Publishers, 1997. P. 156–165.
14. Harinarayan V., Rajaraman A., Ullman J. D. Implementing data cubes efficiently // Proc. of the 1996 ACM SIGMOD Intern. Conf. on Management of Data. Quebec, 1996. P. 205–216.
15. Efficient approaches for materialized views selection in a data warehouse / M. Hung, M. Huang, D. Yang, N. Hsueh // Information Sciences. 2007. № 177. P. 1333–1348.

K. V. Badmaeva

THE TECHNIQUE OF ADAPTIVE CONTROL IN DESIGN AND DEVELOPMENT PROCESS OF SPECIALIZED DATA WAREHOUSE

The technique of adaptive control in process of designing and developing data warehouses is proposed. The technique provides original formalization of the design and development of the data warehouse model and it takes into account the operating conditions and specific information about data domain. The description of the design and development of specialized data warehouse using the management model in the notation eEPC ARIS-methodology is executed.

Keywords: specialized data warehouse, adaptive control, designing, materialization of views.

© Бадмаева К. В., 2010

УДК 65.012.123

S. N. Masaev, M. G. Dorrer

COMPANY MANAGEMENT SYSTEM ESTIMATION ON THE BASIS OF ADAPTIVE CORRELATION TO THE ENVIRONMENT

The method of the structure and indicators analysis of company business processes based on the calculation of simple correlation between historic series of expenses is offered.

Keywords: correlation, adaptation, process, system analysis, management.

After the global crisis outbreak companies have changed a lot: stronger integration of companies, lead to numerous consolidations, acquisitions and mergers.

The Russian companies are not an exception. Holdings (hereinafter referred as to production systems) will be a general moving force, which diversified scope of the activity and the work was in hand inside. It is a gathering moment, but the other process has not completed yet – creating the model of decision-making system. If it is, Russian companies

can be considered well-functioning systems and not just a collection of heterodeneous assets [1].

A holding company manager will find the following difficulties, when managing the holding companies the decision-making system of holding (further production system):

- intercommunication between the companies is inaccurate measurement, subjectivity;
- intercommunication between accounting system and decision-making system;