

УДК 004.65

Д.И. Осенняя, И.И. Закирова, П.Д. Дробинцев

РАЗРАБОТКА ПРОГРАММНОГО ПРОДУКТА ДЛЯ ПОДБОРА КОНФИГУРАЦИЙ ИНФОРМАЦИОННЫХ ЦЕНТРОВ

D.I. Osennaya, I.I. Zakirova, P.D. Drobintsev

DEVELOPING SOFTWARE FOR DATA CENTER CONFIGURATION

Изучена разработка программного продукта для автоматизированного подбора конфигураций информационных центров. Рассмотрены разработанная методика для подбора конфигураций информационных центров и подход к анализу характеристик полученной системы.

ИНФОРМАЦИОННЫЙ ЦЕНТР; EMC; ECLIPSE; EMF; GMF.

The paper describes an approach to the development of software products aimed at automating configuration of data centers. Also, the paper presents methods to select configurations for information centers and an approach to analyse their properties.

DATA CENTER; EMC; ECLIPSE; EMF; GMF.

Хранение информации – важнейшая составляющая информационных технологий. Объем информации, производимой компаниями и частными лицами, растет с экспоненциальной скоростью. И вместе с этим растет и сложность систем, обеспечивающих хранение таких объемов данных.

Успех в бизнесе зависит от быстрого и надежного доступа к соответствующей информации. Значимость информации для бизнеса постоянно увеличивается, вследствие чего усиливается потребность в защите и управлении данными. Информацию необходимо эффективно хранить, защищать, оптимизировать и управлять ею.

Для централизованной обработки информации всего предприятия организации содержат информационные центры. Информационные центры хранят огромное количество данных, необходимых для выполнения различных задач, и управляют этими данными [1].

Деятельность многих организаций предъявляет чрезвычайно высокие требования к доступности и производительности приложений, работающих с данными. Таковыми являются, например, финансовые организации, для которых важны как производительность (поскольку задержки в

функционировании основных систем могут стоить миллионы долларов), так и надежность (поскольку потеря важной финансовой информации может повлечь за собой еще более тяжелые последствия) [7].

Корпорация EMC [3] помогает предприятиям и поставщикам услуг преобразовать свои бизнес-процессы и предоставлять информационные технологии как услугу. В основе этой трансформации лежит облачная инфраструктура. Благодаря инновационным продуктам и услугам EMC ускоряет переход к облачным технологиям, позволяя ИТ-подразделениям более гибко, надежно и экономично хранить и защищать информацию, а также управлять ею.

Каждому поставщику ИТ-услуг требуется определенная конфигурация информационного центра, которая отвечает его запросам, удовлетворяет специфике поставленной задачи, а также обеспечивает оптимальное соотношение производительности, надежности и совокупной стоимости [8]. Поэтому процесс подбора конфигурации каждой системы трудоемок, требует индивидуального подхода и становится особенно актуальной задачей с ростом объема информации, подлежащей хранению.

Часто за основу информационного цен-

тра берется приблизительная конфигурация системы, которая в дальнейшем расширяется с помощью рекомендаций систем мониторинга. Эти системы определяют, какие элементы в информационном центре не справляются с нагрузкой и где требуется добавление дополнительных ресурсов.

Существует большое количество таких систем мониторинга. В арсенале компании EMC есть продукт ViPR SRM, позволяющий визуализировать взаимосвязи в информационном центре, анализировать тенденции нагрузок и конфигурацию системы, получать графики зависимостей приложений от информационного центра.

Компания IBM предлагает продукт Tivoli Storage Productivity Centre, обеспечивающий схожие сервисы: мониторинг информационного центра, формирование отчетов и управление емкостью, сетями, системами и сервисами репликации.

Данный подход нельзя назвать эффективным, т. к. конфигурация информационного центра, взятая за основу, часто бывает очень далека от оптимальной. Поэтому первоначальную конфигурацию системы необходимо выбирать, опираясь на результаты предварительного моделирования системы, в ходе которого рассчитываются ожидаемые характеристики надежности и производительности.

В настоящей статье изучаются методы для предварительного моделирования системы, которые дают возможность проводить ее анализ без непосредственной сборки системы, что позволяет существенно сократить затраты на ее построение.

Моделирование информационных систем

В качестве основы для моделирования информационных центров используется предоставление бизнес-процесса предприятия в разрезе информационных потоков и, в конечном счете, приложений, которые ими оперируют. Однако в данном подходе наблюдается ряд сложностей, основной из которых является разрыв между представлением менеджмента о бизнес-процессах, выражающего требования к информационному центру в терминах приложений и количества пользователей, и ИТ специалиста,

мыслящего категориями объемов памяти хранилищ, количества операций ввода/вывода в секунду и типов элементов, из которых строится система. Так, например, менеджер может указать в качестве требования необходимость размещения на системе хранения приложения Microsoft Exchange с общим количеством пользователей 50 человек и средней интенсивностью входящих и исходящих писем равной 20. Но данное требование абсолютно непонятно ИТ инженеру, которого в конечном итоге интересует количество оперативной памяти или частота и количество процессоров в составе серверов. Решение данной проблемы лежит в области совместной работы менеджмента, представляющего интересы бизнеса, и инженеров, являющихся специалистами в области построения информационных центров.

Для решения проблемы предлагается разработать программный продукт, позволяющий подбирать конфигурацию информационного центра в зависимости от требований заказчика к степени защищенности данных, с которыми оперируют его приложения, быстродействию приложений, а также к стоимости. Кроме того, приложение должно позволять рассчитывать различные характеристики будущей системы, определять ее «узкие места» и отказоустойчивость при выходе из строя различных элементов.

Предлагается собирать данные о приложениях в формате xls таблиц, которыми удобно оперировать менеджеру при составлении требований и строить графическое представление модели информационного центра, на основе которого обеспечить коммуникации менеджмента и инженеров. В результате подобный программный продукт позволит визуализировать предложенное решение для информационного центра и проводить ручное редактирование как набора составных элементов, так и их свойств.

Методика подбора конфигураций информационных центров

В процессе изучения предметной области была разработана методика подбора конфигураций информационных центров (рис. 1). Основой методики является созда-

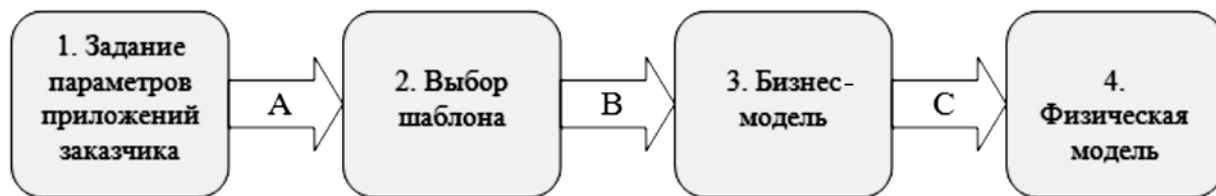


Рис. 1. Методика построения информационных центров

ние моделей различных уровней, которые могут использоваться как при формировании требований бизнеса к будущей системе, так и при разработке детального представления информационного центра на уровне физических элементов (серверов, коммутаторов и т. д.).

Рассмотрим более детально процесс подбора конфигураций информационных центров по предлагаемой методике.

На этапе 1 происходит инициализация исходных данных. В диалоговом окне пользователь указывает xls-файлы, которые содержат характеристики и требования приложений. Основными характеристиками является количество операций ввода-вывода в секунду (IOPS), средний размер одного блока информации, процент операций чтения и необходимый объем памяти. Требования приложений определяют уровень защиты от сбоев серверов и массивов данных. Таблицы составляются менеджерами, исходя из знания и понимания бизнес-процессов. После окончания этапа загруженная информация отображается в диалоговом окне и может быть при необходимости отредактирована.

На этапе 2 с помощью набора правил А, на основе загруженных на предыдущем

шаге данных, происходит выбор списка шаблонов, т. е. часто используемых конфигураций бизнес-моделей, которые подходят в данном случае. Бизнес-модель (БМ) – это графическое представление структуры информационного центра в виде логических сущностей (таких как кластер серверов, уровень виртуализации, система хранения данных [1]) и связей между ними. Здесь пользователю предоставляется выбор: отдать предпочтение системе с большей надежностью, но и с большей стоимостью, или же более дешевой и менее надежной системе.

Набор правил А, который применяется в прототипе программного продукта и позволяет задавать три уровня защиты от сбоев и использовать пять шаблонов, может быть представлен в виде таблицы (см. табл.).

На этапе 3 элементы шаблона с помощью набора правил В, заполняются свойствами на основе загруженных на первом этапе данных, и полученная модель визуализируется. На этом этапе пользователю предоставляется возможность редактирования модели для более тонкой настройки.

Одно из правил набора В, которое позволяет рассчитать количество серверов на каждом хосте, представлено ниже:

Выбор шаблонов на основе требований приложений

Уровень защиты от сбоев серверов	Уровень защиты от сбоев массивов данных	Шаблоны, отвечающие требованиям
Высокий	*(любой)	Gold, Diamond
*(любой)	Высокий	Gold, Diamond
Средний	Средний	Silver, Gold, Diamond
Средний	Низкий	Bronze, Silver, Gold, Diamond
Низкий	Средний	Bronze, Silver, Gold, Diamond
Низкий	Низкий	Wood, Bronze, Silver, Gold, Diamond

$$N_{\text{Server}} = \frac{\sum_{i=1}^k (\text{size } W[i] \cdot W\%[i] + \text{size } R[i] \cdot R\%[i]) \cdot \text{IOPS}[i]}{p},$$

где k – количество приложений в системе; p – пропускная способность портов хоста; $\text{size } W$, $\text{size } R$ – размеры блоков информации на запись и чтение соответственно; $W\%$, $R\%$ – процент операций записи и чтения.

На этапе 4 с использованием набора правил преобразования С, происходит преобразование схемы БМ в схему ФМ и ее визуализация. Физическая модель (ФМ) – подробное графическое представление конфигурации системы с помощью реальных устройств (таких как сервер, switch, VPLEX, RPA [2]). Как и для БМ, на этом

этапе существует возможность редактирования схемы. Также появляется возможность моделирования отказов элементов системы и определения «узких мест».

Описание одного из правил набора С приведено ниже. Элемент БМ Replication преобразуется в набор элементов ФМ – Recover Point Appliances (RPA) – устройств, обеспечивающих репликацию данных. Их количество рассчитывается исходя из пропускной способности элемента БМ Storage, с которым связан данный элемент Replication:

$$N_{\text{RPA}} = \left[\frac{\text{Storage.CommonTrafficCapacity_GbPerSec}}{P_{\text{portRPA}}} \right] + 1,$$

где $P_{\text{portRPA}} = \text{TrafficCapacity_GbitPerSec}$ – пропускная способность порта RPA.

В дальнейшем на основе ФМ проводится анализ системы на соответствие ее характеристик требованиям заказчика. В ходе работы рассмотрено две области характеристик информационных центров: надежность и производительность [8].

Надежность в данном случае характеризуется следующими параметрами:

RPO (Recovery Point Objective, Директивный срок восстановления);

RTO (Recovery Time Objective, Директивное время восстановления);

СМВО (Среднее время между отказами), СВВ (Среднее время восстановления) [1].

Параметры RPO и RTO являются очень удобными для задания требований к системе, т. к. они понятны заказчику и в то же время очень много определяют для исполнителя.

RPO – это момент времени, к которому система должна быть восстановлена после простоя. Он характеризует тот объем данных, который предприятие может позволить себе потерять. Чем больше RPO, тем выше устойчивость бизнеса заказчика к информационным потерям. Основным определяющим фактором RPO является период

создания резервной копии.

RTO – промежуток времени, за который система, приложения и функции восстанавливаются после сбоя. Этот параметр определяет время, которое организация может позволить себя находиться в состоянии простоя и при этом избежать в дальнейшем серьезных последствий.

Рассчитанные в автоматическом режиме характеристики производительности позволяют показать, насколько оптимальна подобранная конфигурация информационного центра и какова рабочая эффективность всех компонент.

В рамках анализа производительности системы оцениваются такие характеристики, как время отклика, утилизация, емкость.

Время отклика – это обратное время запроса на операцию ввода/вывода, т. е. общее время от прибытия заявки до отправки из системы. Среднее время отклика зависит как от времени обслуживания и времени, затрачиваемого на обработку запроса, так и от загрузки системы. Несмотря на то что основным критерием работоспособности системы является доступность информации, время отклика может быть настолько большим, что работу системы нельзя будет назвать приемлемой.

Оценка уровня утилизации применима как к отдельным компонентам системы, так и к группам однотипных элементов и показывает, на сколько процентов используются доступные ресурсы устройств. Элементы с высоким уровнем утилизации могут стать причиной снижения производительности системы, поэтому они должны быть выявлены и, по возможности, либо продублированы, либо оснащены дополнительными ресурсами. Очень низкий уровень утилизации может указывать на то, что устройства «простаивают» и, возможно, для поставленных требований может быть разработана более оптимальная конфигурация системы.

Емкость характеризует сумму доступных ресурсов. Например, это может быть количество доступных портов на коммутаторе, или объем дискового пространства в хранилище. Управление емкостью лежит в основе обеспечения требований заказчика на работу приложений и уровень обслуживания.

Входными данными для расчета описанных характеристик информационного центра является ФМ системы, где для каж-

дого элемента известен набор характеристик, таких как средняя наработка до отказа, среднее время восстановления, скорость обработки информации и т. д. Также необходимо задать условия работоспособности системы в целом.

Описанная методика была использована при создании продукта DataCenter, предназначенного для построения, конфигурации и анализа информационных центров. Разработка продукта производилась в среде Eclipse с помощью инструментов Eclipse Modelling Framework (EMF) [4] и Graphical Modelling Framework (GMF) [5]. Для поддержки работы продукта с моделями различных уровней на первом этапе создаются EMF [4] метамодели классов и их связей друг с другом для БМ и ФМ, и по ним генерируется Java-код этих классов. На втором этапе, с помощью инструмента GMF, на основе сгенерированного кода классов моделей генерируется код для графических редакторов. Далее, оба полученных редактора объединяются в одно RCP (Rich Client Platform) [6] приложение, в которое по мере необходимости разработчик вручную вносит требуемые изменения.

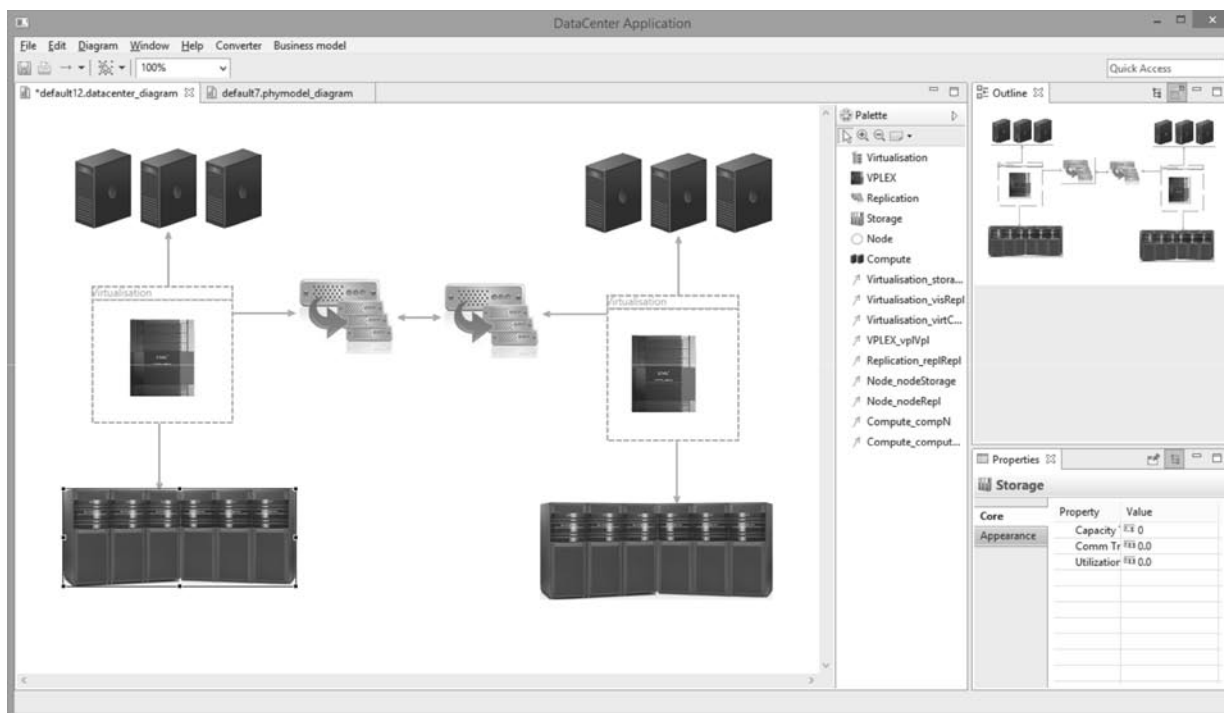


Рис. 2. Приложение DataCenter. Бизнес-модель

К настоящему моменту разработан прототип программного продукта, позволяющий:

1) создавать, редактировать, сохранять (в формате xml) БМ и ФМ;

2) читать информацию о приложениях из xls-файлов, и на их основе выбирать список подходящих шаблонов БМ;

3) заполнять свойствами и визуализировать выбранный шаблон;

4) генерировать на основе БМ ФМ с использованием временных правил преобразования;

5) проводить расчет и анализ характеристик надежности и производительности создаваемого информационного центра.

На рис. 2 изображено главное окно приложения DataCenter. Слева область редактора бизнес-модели, в нем визуализируется конфигурация информационного центра с двумя кластерами серверов, двумя уровнями виртуализации, содержащими по одному VPLEX [1], метро репликацией и двумя системами хранения данных. Правее модели находится палитра, а в правом нижнем углу параметры выделенного элемента (в данный момент это кластер массивов хра-

нения данных [1]).

Результаты. Четырехэтапная методика, описанная выше, лежит в основе проекта DataCenter и на данной стадии разработки показывает неплохие результаты для решения задачи конфигурирования информационных центров. Используя даже тестовые алгоритмы, она позволяет получить прототип результирующей модели. Несмотря на стремление к автоматизации процесса, на каждом этапе пользователю доступно ручное редактирование параметров системы, в будущем это должно позволить моделировать самые разнообразные индивидуальные решения. Усовершенствуя правила преобразования и расширяя набор встроенных шаблонов, можно добиться более точной работы программы.

Разрабатываемый программный продукт позволяет автоматизировать процесс подбора оптимальной конфигурации информационных центров для конкретных заказчиков, а также оценивать характеристики и отказоустойчивость будущей системы что существенно снижает затраты на построение информационных центров.

СПИСОК ЛИТЕРАТУРЫ

1. От хранения данных к управлению информацией. EMC. СПб.: Питер, 2010. 544 с.

2. **Dharma R., Jiang L.** *Networking for Storage Virtualization and EMC RecoverPoint*. EMC, 2013. 102 с.

3. EMC – ведущий поставщик проверенных ИТ-решений, технологий для облачных вычислений и работы с большими данными [электронный ресурс] / URL: <http://russia.emc.com/index.htm> (дата обращения 10.11.2013).

4. Eclipse Modelling Framework Project (EMF): [электронный ресурс] / URL: <http://www.eclipse.org/modeling/emf/> (дата обращения 11.11.2013).

5. Model Driven Architecture approach to

domain of graphical editors [электронный ресурс] / URL: <http://eclipse.org/gmf-tooling/> (дата обращения 11.11.2013).

6. Rich Client Platform: [электронный ресурс] / URL: http://wiki.eclipse.org/index.php/Rich_Client_Platform (дата обращения 11.11.2013).

7. **Горбунов-Посадов М.М., Корягин А.Н.** *Современные системы хранения данных старшего класса*. М.: Препринты ИПМ им. М.В. Келдыша, 2008.

8. **Лобанов А.К.** *Методы построения систем хранения данных* [электронный ресурс] / URL: <http://citforum.ru/hardware/data/db/> (дата обращения 24.05.2014).

REFERENCES

1. *От khraneniya dannyh k upravleniyu informatsiyey*. EMC, St. Petersburg: Piter Publ., 2010, 544 p. (rus)

2. **Dharma R., Jiang L.** *Networking for Storage Virtualization and EMC RecoverPoint*. EMC, 2013, 102 p.

3. *EMC – vedushchiy postavshchik proverennykh IT-resheniy, tekhnologiy dlya oblachnykh vychisleniy*

i raboty s bolshimi dannymi Available: <http://russia.emc.com/index.htm> (Accessed 10.11.2013). (rus)

4. *Eclipse Modelling Framework Project (EMF)*. Available: <http://www.eclipse.org/modeling/emf/> (Accessed 11.11.2013).

5. *Model Driven Architecture approach to domain of graphical editors*. Available: <http://eclipse.org/gmf-tooling/> (Accessed 11.11.2013).

6. *Rich Client Platform*. Available: http://wiki.eclipse.org/index.php/Rich_Client_Platform (Accessed 11.11.2013).

7. **Gorbunov-Posadov M.M., Koryagin A.N.** *Sovremennyye sistemy khraneniya dannykh starshego*

klassa. Moscow: Preprinty IPM im. M.V. Keldysha Publ., 2008. (rus)

8. **Lobanov A.K.** *Metody postroyeniye sistem khraneniya dannykh*. Available: <http://citforum.ru/hardware/data/db/> (Accessed 24.05.2014). (rus)

ОСЕННЯЯ Дарья Игоревна – студентка кафедры информационных и управляющих систем Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: osen1313@gmail.com

OSENNYA, Daria I. *St. Petersburg State Polytechnical University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: osen1313@gmail.com

ЗАКИРОВА Ирина Игоревна – студентка кафедры информационных и управляющих систем Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: from.mumla@gmail.com

ZAKIROVA, Irina I. *St. Petersburg State Polytechnical University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: from.mumla@gmail.com

ДРОБИНЦЕВ Павел Дмитриевич – доцент кафедры информационных и управляющих систем Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: drob@ics2.ecd.spbstu.ru

DROBINTSEV, Pavel D. *St. Petersburg State Polytechnical University.*

195251, Politekhnikeskaya Str. 29, St. Petersburg, Russia.

E-mail: drob@ics2.ecd.spbstu.ru