

ОБ ИСПОЛЬЗОВАНИИ КОНТРОЛЬНЫХ ДАННЫХ В СИСТЕМЕ ОБЕСПЕЧЕНИЯ КАЧЕСТВА СЛОЖНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

В. М. Гребенюк

Московский государственный технический университет
радиотехники, электроники и автоматики

Получена 30 августа 2013 г.

Аннотация: Рассмотрены современные методы управления и контроля тестовых данных, выявлены преимущества и недостатки каждого метода, а также сформулированы проблемы и особенности управления тестовыми данными в сложных системах.

Ключевые слова: сложные системы, информационные системы, контрольные данные, обеспечение качества.

Abstract. The article is outlining best practices of using various reference data management methods and technologies, highlighting pros and cons of each method. It is also highlighting the key issues of managing reference data in complex systems.

Keywords: complex systems, information systems, test data, quality assurance.

Введение.

Зачастую, контрольным (тестовым) данным, используемым для верификации правильности работы сложных информационных систем, не уделяется должного внимания, однако, они могут иметь определяющее значение в поведении тестируемой системы в процессе испытаний, препятствовать решению проблем интероперабельности (то есть способности систем и компонентов к взаимодействию), а также могут повлиять на итоговые результаты испытаний. Задача подбора контрольных данных позволяющих получить максимально реалистичные результаты тестов, в ряде случаев, является нетривиальной. Методы, рассмотренные в данной статье, могут помочь в выборе наиболее оптимального набора контрольных данных.

Анализ существующих методов контроля тестовых данных.

Анализ публикаций подтверждает особую роль контрольных (тестовых) данных в процессе обеспечения качества сложных систем: будь то сложная система военного или гражданского назначения, для проведения интеграционного тестирования необходим набор реалистичных контрольных данных, присутствующий во всех подсистемах [1]. Проблема усугубляется тенденциями к укрупнению систем, росту сложности и объёмов хранимых данных. Если в «живой» среде большинство организаций стремится к поддержанию справочных данных о соответствии между подсистемами, то тестовая среда, зачастую, представляет собой не связанные и не синхронизированные между собой наборы тестовых сред.

На данный момент существует большое количество сервисов и программных решений для управления контрольными данными. Принципиально, такие решения можно разделить на несколько групп [2]:

1. Каталоги тестовых данных,
2. Средства реплицирования и обезличивания данных из «живой» среды,
3. Средства доступа и контроля источников данных.

Автоматическое или ручное создание каталога тестовых данных является наиболее распространённым средством управления тестовыми данными. В самом простом случае, такой каталог представляет собой список данных, которые предлагается использовать для тестирования. Основными недостатками такого способа являются необходимость постоянной поддержки каталога в актуальном состоянии и необходимость детального учёта всех изменений, так как каталог представляет собой «срез» данных на определённый момент времени без возможности автоматического обновления при изменении данных.

Другим распространённым средством управления и создания тестовых данных является копирование и обезличивание данных из «живой» среды [3]. Обезличивание необходимо для предотвращения утечки персональных и других данных, представляющих ценность или ограниченных к распространению. Существует несколько недостатков данного метода:

- некоторые методы обезличивания информации способны исказить логику построения контрольных данных,
- объёмы данных, полученных из «живой» среды могут быть слишком большими для тестовой среды.

Самым современным средством управления тестовыми данными являются средства доступа и контроля источников данных [4]. Такие средства позволяют устанавливать соединения непосредственно с источниками данных (например, базами данных), настраивать соответствия и ссылки между данными из разных источников, а также следить за обновлениями данным в автоматическом режиме. Недостатками таких средств являются:

- первоначальная настройка может потребовать определённых усилий и навыков,
- необходимость в стандартизации формата и набора данных между источниками,
- наличие непосредственного доступа к источникам данных каждой подсистемы.

С точки зрения взаимодействия с подсистемами, в основном выделяются 3 типа решений [5]:

- централизованное;
- интегрированное;
- комбинированное.

При централизованном подходе, выделяется ответственная организация или структура, которая отвечает за сбор, хранение и поддержку информации в актуальном состоянии. Этот способ является самым точным и позволяет избавиться от дублицирующей информации, однако, он требует чёткой координации и большое количество сил, из-за чего является самым дорогим.

Интегрированный подход подразумевает наличие централизованного узла, содержащего ограниченное количество информации и перекрёстных ссылок, при этом подсистемы сами ответственны за хранение и поддержание

информации в актуальном состоянии. Данный подход является одним из самых простых, но он не исключает риск изменения необходимой информации внутри отдельной подсистемы.

Комбинированный подход подразумевает, что отдельная часть (отдельные записи) информации из всех систем храниться и поддерживается в актуальном состоянии централизованно, а остальную информацию использует каждая из подсистем индивидуально в соответствии со своими планами. Данный подход является наиболее распространенным, но, как и первый, требует строгой координации между подсистемами.

Применение методов контроля тестовых данных при тестировании сложных информационных систем.

Стоит отметить, что не все из существующих решений применимы для сложных программно-аппаратных комплексов, в силу наличия подсистем, каждая из которых является отдельной системой с собственной «дорожной картой», бюджетом, организационной структурой и набором тестовых стендов. Наиболее распространенными трудностями, возникающими при проведении интеграционного тестирования сложных систем, являются:

1. разница в версиях приложений в разных тестовых средах и их возможная несовместимость,
2. разные расписания работы тестовых стендов различных подсистем,
3. возможное отсутствие тестового стенда для одной или нескольких подсистем.

Для обеспечения корректного использования контрольных данных в сложных информационных системах и обеспечения интероперабельности, в дополнение к методам управления контрольными данными, рассмотренными в предыдущем разделе, предлагается использовать следующие меры:

1. Централизованная координация использования тестовых сред на время проведения интеграционного тестирования во избежание непреднамеренного изменения контрольных данным в подсистемах.

2. Использование различных, независимых друг от друга, наборов тестовых данных для разных видов тестов в рамках интеграционного тестирования.

3. Внедрение ручных или автоматических методов определения целостности данных непосредственно до начала интеграционного тестирования.

4. Подготовка резервных наборов контрольных данных.

В отдельных случаях, решением проблемы подбора тестовых данных для верификации корректности функционирования сложных систем может стать тестирование в «живой» среде, однако, такой метод не является универсальным и на это есть несколько веских причин. Во-первых, если речь идёт о тестировании сложных систем на этапе разработки, то никакой «живой» среды не существует. Во-вторых, появляется риск нарушить нормальное функционирование системы из-за инъекций тестовых данных и/или проведения тестовых активностей. В-третьих, существует опасность утечки данных из реальной («живой») среды, включая персональные данные, которые могут представлять коммерческую или иную ценность, а также попадать под действие существующего законодательства.

Выводы.

Проблема выбора и контроля тестовых данных при обеспечении качества сложных систем и, в частности, проведения интеграционного тестирования, становится всё более актуальной из-за тенденций к укрупнению систем, росту сложности и объёмов хранимых данных. Данное исследование показало, что несмотря на наличие существующих решений, предназначенных для управления контрольными данными, существует необходимость в принятии дополнительных мер для обеспечения корректного использования контрольных данных и интероперабельности компонентов, в силу специфики сложных информационных систем.

Литература

1. Bock T., Easley T., Gibson J. Managing large distributed data sets for testing in a joint environment // EbookBrowse.net search engine, 2010. URL: <http://ebookbrowse.com/data-management-managing-large-distributed-data-sets-for-testing-in-a-joint-environment-aiaa-pdf-d329701848>
2. Haller K. Test Data Management in Practice // Software Quality Days 2013 Conference journal, 2013. URL: http://www.swisscom.com/content/dam/its/documents/whitepapers/wp_test-data-management_conference-journal_de.pdf
3. Hewlett Packard business white paper. How to address top problems in test data management // Hewlett-Packard Development Company, 2010. URL: http://www.hp.com/hpinfo/newsroom/press_kits/2010/HPSoftwareUniverseBarcelona2010/How_to_address_top_problems_WP.pdf
4. IBM Software Solution Brief. IBM InfoSphere Optim Test Data Management // IBM Corporation, 2012. URL: <ftp://public.dhe.ibm.com/common/ssi/ecm/en/ims14064usen/IMS14064USEN.PDF>
5. Reddy A.S. Reference Data Management: The Case for a Utility Model // Cognizant, 2012. URL: <http://www.cognizant.com/InsightsWhitepapers/Reference-Data-Management-The-Case-for-a-Utility-Model.pdf>