SQS Test Data Management

Data protection compliant, assure security, reduce costs and improve quality

sqs.com

Introduction

Security breaches are everywhere in the news. We read about personal information being traded, the unwanted disclosure of data and even extortions using such data. Much more important than the penalties associated with the offence of revealing protected data, are the obvious damage to reputation. The frequency of these occurrences is increasing.

It is obviously far too easy to obtain sensitive data.

- Why is it this the case?
- How can I protect myself?

You can find out how SQS organises test data management for protection and security, whilst lowering the test costs and simultaneously improving the quality of your software.

Authors: Dr. Kai-Uwe Gawlik
Global Head of Service Management
SQS Software Quality Systems AG
kai-uwe.gawlik@sqs.com

Rüdiger Louis
Delivery Manager/Management Consultant
SQS Software Quality Systems AG
ruediger.louis@sqs.com

Published: March 2010

Using production data favours data misuse

The following scenarios facilitate the misuse of sensitive data:

- In test and development environments, production data are used for standard and individual development

- In test and development environments, the access to this data is inadequately restricted

- Data and files can easily be moved to storage media or sent to any address

Traceability of data movements and formal commitments are of little use if the damage has already occurred. Actions can be taken which may hinder the misuse of data or even make it impossible. SQS Test Data Management is designed to prevent the security breach of critical data from the beginning.

Prevention is the best protection

The possibilities of a security breach can be averted as long as production data is not used in the test and development environment. Therefore the use of synthetic test data is obvious:

- Synthetic data are fictitious and do not need to be protected.

- Required test data can be adapted to the requirements of the tests (test cases) and do not necessarily need to reflect a real-world situation.

In addition, synthetic data have the following advantages:

- The desired testing coverage is achieved.

- Dates are set by parameterisation as needed. Sets of data can be generated dependent on a particular day's date.

- The volume of test data is low.

- Run time for test data availability preparation is low.

- Data can be loaded in testing into each existing database version.

BUT

- The cost increases with the complexity and the multitude of systems which need to be supplied consistently by data.

A mix of synthetic and anonymous real data is the best way

From the point of view of testing, there are some good reasons to use production data, because production data show the following characteristics, which synthetic data do not necessarily have:

- Large and complex systems can easily be supplied through the implicit consistency of production data.

- Productive data usually contain unknown data defects, which can lead to crashes and errors particularly on processing deadlines. These can be identified and eliminated through testing. The software can be designed defensively against these defects.

- Run-time and performance measurements are only meaningful with the corresponding data volume

BUT

- Production data comply with the control processes and cover the test, for example at the limits, at the most up to 60 %.

- Production data age. Some data may mature and therefore becomes unsuitable for use within a test case.

- The volume of test data is extensive.

- Durations of the test data provision are high.

- Production data do not fit easily into the more advanced database models in the test.

- Production data must be protected from unauthorized access.

- The use of production data in the test must meet the requirements of data protection.

In summary, only the appropriate mix of synthetic and anonymous/pseudonymous production data meets the requirements of both the data security and the requirements of the test.

To answer the question of the appropriate mix of data, usually one has to consider other requirements on the test data management:

- Test teams and operating departments must have confidence on the quality of test data.

- Operation departments and respective application administrators in projects or maintenance must be in a position to specify data in detail.

- The charge of the project budgets through test data organization and management must either generate a sustainable added value or be amortized over the project period and the product cycle.

- Test data and test environment management must be synchronised to create synergy.

SQS PractiQ®: Test Data Management

The SQS Test Data Management is the result of our more than 30 years experience in over 7,000 projects. Based on our sustainable approach, we would like to present you the following offer.

General approach

The SQS approach is integral and combines aspects of a systematic test approach together with the requirement of test data supply. This way, both data are aligned to the requirements of the test and, at the same time, a highly efficient form of testing is chosen.

The SQS Test Data Management is aligned with the actual need for test data and follows strict cost optimisation. With our approach for test data management, we lay the foundation of successful test automation. Test automation brings a major contribution to the relief of operation departments. Through our test automation specialists, we are able to create a prerequisite for this in the form of test data management.

As an independent service company, we are not tied to a fixed tool configuration when implementing the SQS Test Data Management. Instead, we can rely on existing tool landscapes in the customer environment alternatively assist in the selection of necessary tools. The goal is to implement an optimum solution in terms of application and architecture.

Our operations will be conducted under the compliance of:

- International standards, such as for example ISTQB,

- National legal framework,

- Company-specific regulations, such as for example IT security and data security,

- Accountability for audit and auditors.

The various required standards will be agreed upon according to each customer environment.

Test data management

The SQS Test Data Management affects the test process through the optimisation of needs-based approach. Costs are reduced and simultaneously the required quality is assured. Through the structured approach, the test process wins transparency.

The need for test data is derived from:

- Test objects which are based on test levels up to the developers' test and quality parameters (ISO9126).

- Test sets: which are executable representatives (test combination) of test cases. In this context test cases stand as order forms for test data which are derived from the requirements.

The requirements are therefore classified according to the following questions:

- Project view: Which documented requirements (technical approaches) are available in regard to an application and must therefore be tested using appropriate test data?

- Application view: What data are representative for the application and business professional and are therefore particularly relevant for regression testing?

- Data Quality: What data exist in the application and thus can lead to data quality problems (e.g. pollution effects, and legacy data)?

The data quality is of great relevance, because it can lead to unwanted interference particularly in the processing of bulk data lead. The result can be unwanted incidents such as delays in the process, job failure or even system crashes.

Providing test data

For the supply of test data, SQS Test Data Management deliberately schedules a division of the process. The extraction of test data is separated from the supply of test environment landscape with test data:

- The task of data extraction is to provide actual test data (databases, files, scripts), which are ready to be supplied upon demand for the test data.

- The task of supplying test data is to optimise the test data on request to the demands of the test environments.

Thus, a human, time and technical content separation of the test deployment process is achieved. This provides independent test environments, which can be supplied with test data.

The SQS approach makes the test data a manageable and controllable process: projects will be supported in an efficient and flexible way.

During the implementation of test runs, separation of extraction and provision of test data management guarantees following:

- Shorter lead times to generate and deliver test data

- A faster data supply through parameterised provision instructions (application, test level, test object)

An essential part of the SQS Test Data Management is to classify the data according to their stability, because a set of rules must be derived for the generation of test data and implications ensue for the deployment processes of this data.

- Configuration data, e.g. for proper control of processes to test environment provision

- Master data, e.g. customer data

- Inventory data, e.g. accounts and deposits

- Transaction data, e.g. transactions (sales, deposits/ withdrawals, securities purchase or sale, debit)

The system of rules together with the test levels of the test process specifies in which form a mix of the following processes should occur and is an essential element of an efficient and demand-oriented test data provision:

- Production data (only exceptionally allowed and anonymised)

- Anonymised tool based test data

- Test data (automatically) generated through the launch of the application

- Synthetic tool based test data

Key challenges in the test data are

- Test data which become unusable because of aging of data

- Creation of timely consistency, particularly in transaction data

The SQS Test Data Management meets both challenges. Depending on the procedure it ensures that aging of data is prevented through placing (synthetic) test data within set variables or data which are subjected to tool-based, time dependant data.

Implementing test data management

The technical implementation of a test data management is similar to a software development process as it runs through the following phases:

- Analysis

- Design

- Implementation

- Test

In the context of synthetic data and test automation, we have successfully used the "block" model as a design so that maintenance of test scripts remains manageable. When using production data, anonymisation and data reduction are implemented on a representative portfolio during these phases. In this case, cooperation of industry specialists and technicians will be the key to success. Both will take up the important task to identify dependencies in data processing which are not stored in the referenced in the database. Additional requirements are that Business Specialists are able to describe requirements (for example, in the form of use cases), to identify test objects and to formulate test cases in the sense of request forms for test data. Test case methodology covers up errors early in the conception phase and helps preventing subsequent costs and damages (Early Error Detection).

SQS forges links between requirements and testing.

Under these premises, an organised test data management forms a nucleus for sustainable optimisation of testing and provides cost savings and quality improvement. The introduction of the SQS Test Data Management can be carried out iteratively and involves applications which can be incrementally prioritised dependant on test levels in test data management.

This approach spreads the required investment over several phases.

Initial savings can already be done during the implementation of the overall project. Milestones and feedback loops and related practices are interwoven. Test data management is an ongoing task and should be centrally located. SQS Test Data Management is focused on recurrent testing activities (release-exchange, platform-changing, regression) and on securing of business processes (process quality).

Our global approach together with a combination of test data requirements and test deployment unburdens project budgets whilst ensuring quality.

Test data management is a rolling process which includes a current test data production for the release at the end of the test. For a follow up release new features have to be added or changes have to be adapted. An inventory at the end of the project ensures that unnecessary data are deleted.

Besides the purely technical implementation, SQS also provides for necessary changes and integration in business organizations.

Typical project phases for the introduction of test data management

Phase 1: Creation of a solution concept

To create a solution concept, an upstream analysis phase and a rough design phase must be performed in order to grasp the actual situation and the requirements and also to analyse the experiences based on hitherto test data management and test requirements. In the concept, the best practices are processed with the customer's situation (reality) so that a tailor made approach can be developed regarding

- Requirements and deployment process

- Provision of test data

- Reduction, anonymisation and synthetisation

- Procedural conditions

Phase 2–4: Implementation per application

The project is conducted iteratively in the following phases:

- Phase 2: Creation and optimization of test cases, test data, order forms, especially for regression

- Phase 3: Enrichment of the test system with synthetic data and test automation

- Phase 4: anonymisation and reduction of production data for the integration test

Each application goes through the phases 2 to 4. The activities of the respective phases can be made in parallel. This means that an average of 3 to 4 applications can be implemented in a period of 4 months. We recommend starting with 2 applications. Once the cooperation and the processes have stabilized, the frequency can be increased to 4 applications per phase until the end of the project.

Scalability: You decide which phases apply for you.

For the implementation following skills are involved:

- Test manager with deep industry knowledge for analysis, planning test level, test object definition and assessment

- Test designer for the creation of test cases (test order forms) especially for the regression

- Automation technician for the provision of test data

- Reduction / alienation technician for the provision of test data

If necessary, a specialist for synchronisation between the configuration management plan of testing and applications could be assigned for the duration of the project.

Tool product range

The testing process is a chain of individual steps which are executed cross-platform. With the Test Process Automation (TPA) and Distributed Test Environment (DTE) components of the SQS-TEST®/Professional tool, which are the preferred core components of the SQS Test Data Management, these individual steps can be triggered, synchronised and controlled. SQS-TEST®/Professional allow the integration of various tools, which are used in individual steps. It is based on a script-generating approach, which reduces maintenance costs in a sustainable way and makes them controllable. In this way, for example,

- SQL Loading and unloading routines

- Manipulation scripts

- Shell scripts

- Capture-replay scripts e.g. Quality Center

- Data alienation scripts

- Comparison scripts

- Test data from requirements (also from SQS-TEST®/Professional)

can be generated and integrated and controlled in a synchronized sequence. At the same time, tool platforms which already exist at the customer can be integrated.

The SQS Test Data Management process with the assistance of SQS-TEST®/Professional has been certified in interaction with SAP HR and eCATT of Data Protection of Schleswig-Holstein (Germany). Simultaneously the approach is integrated into the in-house standard SQS PractiQ® test automation. SQS-TEST®/Professional is targeted at a general test process to be controlled (driver of many other tools) in contrast to the local tool solutions for special themes, such as Capture Replay, SQL or comparison tools.

To reduce and anonymise production data we seek the best possible support for customers through marketable tools and we subject these tools to a proof of concept and pilot operation. In this context we also examine the customer's existing developments and evaluate them for further use in an integrated approach (on-board resources).

Summary

The SQS Test Data Management focuses primarily on test data, but has lasting side effects, which are based on the many years of experience of SQS in quality assurance. The solution is comprehensive and future-proof. The life cycle of test data is considered as a whole, and takes into account aspects such as data movement and data aging.

This leads to:

- Reduction of cost, time and complexity in the test through appropriate test organization and increase of flexibility

- Increased software quality through improvements in the testing process and test methodology as well as accompanying automation of test processes

- Shorter delivery times and procedures in the testing process (working vs. scenario days)

- Automated regression testing

- Clear Quality Gates in the testing process on all test levels

- Application of a common test methodology

- Centralised expertise to provide test data

- Centralised management and deployment of test environments

... and avoidance of the use of production data wherever possible and economically feasible.