How to address top problems in test data management

Data reuse, sub-setting and masking

Business white paper

Table of contents

Why you need test data management

Around the world, businesses are focussed on three critical business goals: to improve business agility, to reduce costs and to mitigate risks. These goals are among the key elements to corporate success.

Business agility allows an organisation to bring products to market in less time. Cost reduction helps a business stay profitable and maintain its competitive edge. And risk mitigation, including compliance with regulations, is one of the keys to corporate survival in a regulated world.

To help the business achieve these goals, IT application teams need to deliver quality software on time and on budget. In the software development and testing effort, application teams need to protect sensitive information to help the business meet its compliance requirements. And they need to leverage tools and processes that help reduce infrastructure and operations costs.

While these goals are clear, the reality is something else. Today's quality assurance (QA) teams face formidable challenges in the testing process. These challenges stand in the way of making your QA team a contributor to business agility, cost reduction and risk mitigation.

The challenges of preparing and managing test data

Realistic data is hard to collect and sort
With today's business applications, data is typically spread across multiple databases, including both internal and external ones. Customer records might be stored in one place and purchase orders in another. This can make data extraction a time-consuming process. Many test teams know that accessing and extracting test data can consume a large portion of the time spent in testing efforts.

In addition, testing organisations typically have limited access to the production systems or production backup systems where real customer information resides, and they have limited skills for dealing with the range of databases and schemas. This means they have to depend on other stakeholders – such as subject matter experts and database administrators – to provide the data they need. It all adds up to a lot of lost time in the testing process.

Unmasked sensitive data puts the business at risk
Social security numbers, credit card numbers and other personal and business information are an attractive target to hackers, data thieves and others.

That's why today's business applications are typically protected by multiple layers of data security. However, that is not the case with most test and development environments. When production data is used for QA tests, sensitive data can become a soft target for unscrupulous people.

What's more, the risk of data security breaches grows when QA organisations outsource testing to third parties, which might be anywhere in the world. When third parties have access to test data, how do you meet your compliance requirements for the protection of confidential customer information? These teams might not even be aware of the sensitivity of the data they are using in their tests.

Storage maintenance costs too much
As the number of business applications rises and the amount of data they handle explodes, storage maintenance costs are becoming a significant drain on IT budgets. While you might pay US$100 to acquire one terabyte of storage, you might end up spending almost US$700 to maintain that storage capacity. Part of this cost stems from the need for database administrators who are well versed in SQL script.

Given the high cost of storage maintenance, your QA team needs to reduce the amount of data it stores and manages. It is not cost effective to clone and maintain an entire production database when you actually need just a relevant subset of the data for testing. It is hard to maintain the referential integrity of data when you pull it out of a production environment. For example, if a customer name changes in one place, you need to make sure it gets changed in other places where it appears.

Data is not readily available for reuse
After all the work and expense of extracting data from a production database, the data used by QA specialists is often put to work in just one testing task or phase. Your QA organisation can get more out of the time and effort it puts into collecting and preparing data if that data is usable by all the testing tools.

And since data extraction can be required with every testing iteration or agile sprint, the use of time-consuming manual extraction methods can drive up costs and jeopardise schedules. In addition, human error is common when IT staff attempt to handle large quantities of data manually without a structured automation solution.
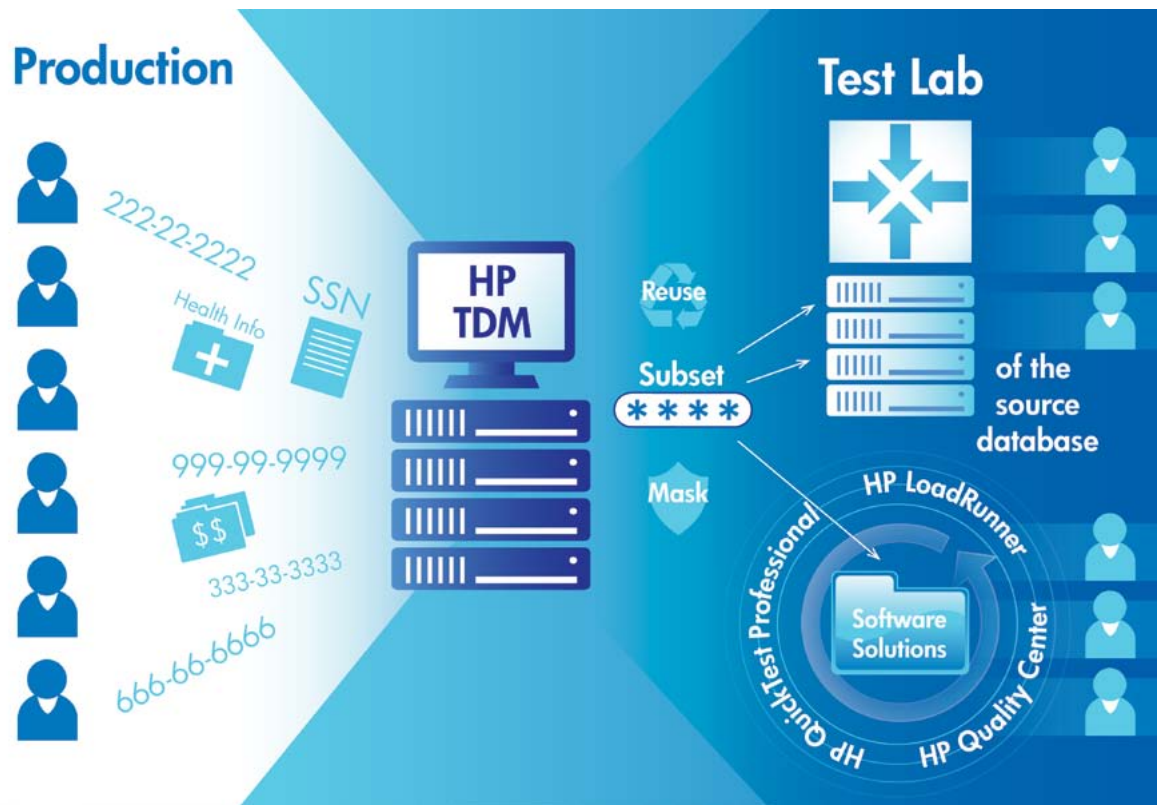
HP Test Data Management

HP Test Data Management software helps your organisation meet today's testing challenges. It provides a flexible, powerful solution for preparing test data. This standalone solution is designed to reduce the delays and costs of data-driven testing and to mitigate the risk associated with tests that use sensitive customer data. This is accomplished by accelerating test data preparation using features such as automated data extraction and masking.

HP Test Data Management complements the capabilities of the software solutions within the HP Quality Center and HP Performance Center suites, including:

- HP QuickTest Professional software for functional and regression testing automation
- HP Service Test software for simplifying and accelerating the automated functional testing of service-oriented architecture (SOA) services
- HP LoadRunner software for performance, stress and load-testing
- HP Business Process Testing software for designing tests and test cases from reusable test components

The HP Test Data Management solution is based on a methodology that involves extracting data according to pre-written rules on a demand-based schedule, detecting and masking sensitive data, uploading the data to a target, and then reusing the same data in other HP Quality Center tools.

Through this methodology and the integration of the HP solutions, HP Test Data Management makes it possible for your QA team to:

- Simplify extraction and design
- Use flexible data masking to protect sensitive information
- Reduce storage maintenance costs
- Preview data prior to download to reduce errors
- Automate data preparation with business workflows

- Manage user authentication and authorisation
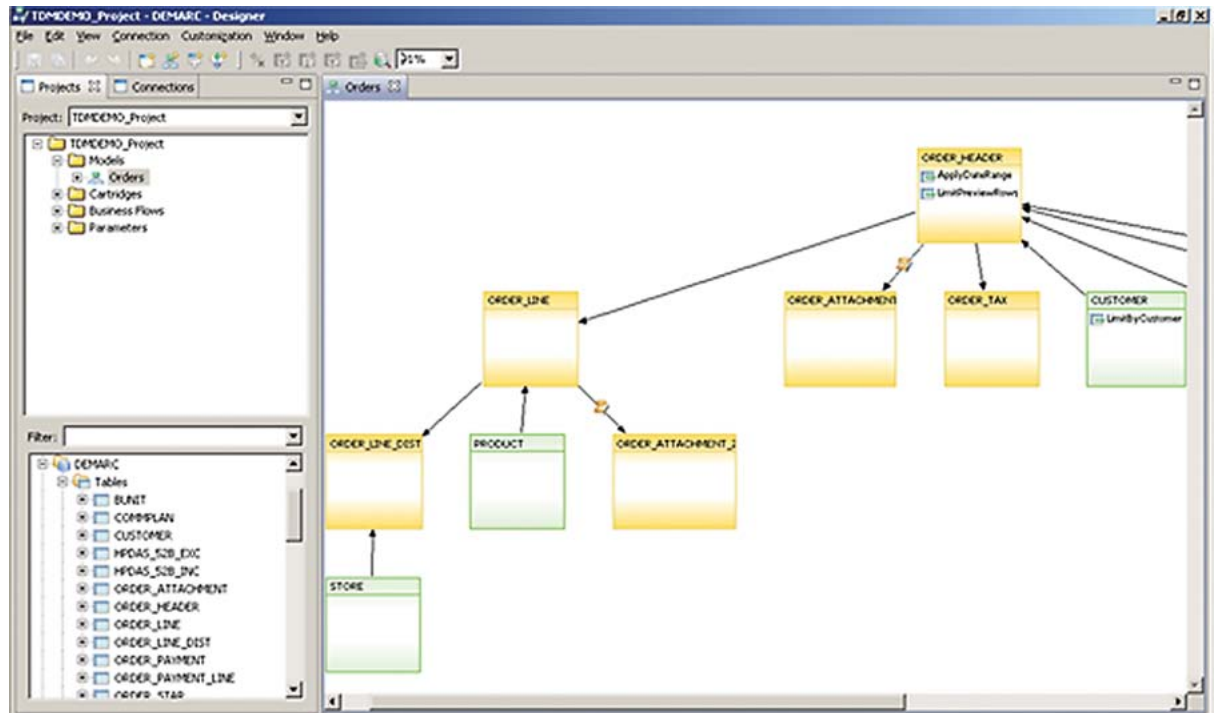- Provision test environments rapidly

With these capabilities, your QA team can efficiently acquire the test data it needs and use it effectively across HP Quality Center, manual and automated tests, and waterfall and Agile environments.

In another benefit, HP Test Data Management reduces your QA team's dependency on subject matter experts and database administrators. Typically, QA specialists have to turn to these people when they need to extract realistic and most complete data from a production database. HP Test Data Management reduces this dependency by automating many data extraction functions.

Key capabilities in HP Test Data Management

Simplify extraction and design
In a typical testing scenario, QA teams need to extract data that is spread across multiple databases. It is not cost effective to clone and store large production databases, so QA teams need to find ways to extract just relevant subsets of data. And in the process of extracting that data, they need to mask sensitive information, such as social security numbers, account numbers and balances.
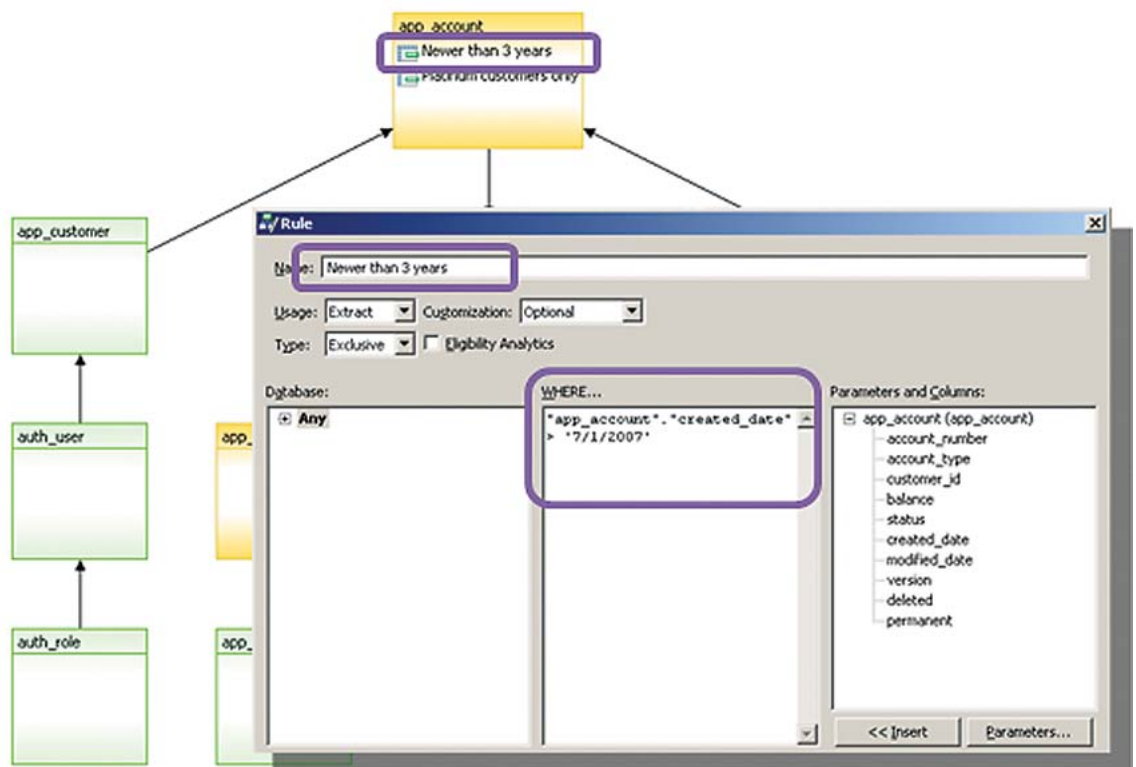
HP Test Data Management helps you address the above challenges. It gives you the ability to extract just a subset of the data that will be relevant to your tests and in the process mask sensitive data. This can save you a lot of money on storage

maintenance and it can greatly reduce your need for storage capacity.

During implementation, HP Test Data Management enables your QA team to build a data model that either fits the whole schema of the source database or models the required tables needed for testing. The HP Test Data Management Designer feature helps you accomplish this through a rich graphical user interface (GUI) that makes it possible to easily see links and dependencies between tables, maintain the model and accelerate the rest of the implementation phase. The HP Test Data Management Designer enables HP Test Data Management to see how the database is structured and how tables relate to one another, which is one of the keys to successful data extraction.

The visual interface within the HP Test Data Management Designer simplifies key tasks. With the intuitive functionality of this interface, you can easily build a data model aligned with the schema of the source database, display links and dependencies between tables, and maintain the model as changes occur. You do not need to be a SQL expert to carry out these and other extraction and design tasks.

In addition, the software's parameterized data extraction capability allows you to filter data based on sub-setting functions. This capability is particularly beneficial for data-centric testing. Filtering allows you to narrow the scope of data extraction.

A filter can be applied on any item stored in a database. You might, for example, filter data based on accounts newer than three years old, as shown in Figure 3. Or you might filter data based on accounts with balances that exceed US$100,000, all sales in a region during the previous quarter, or some other parameter.

Once the data model is built, HP Test Data Management helps your QA team manage testing data by preserving information characteristics and referential integrity.
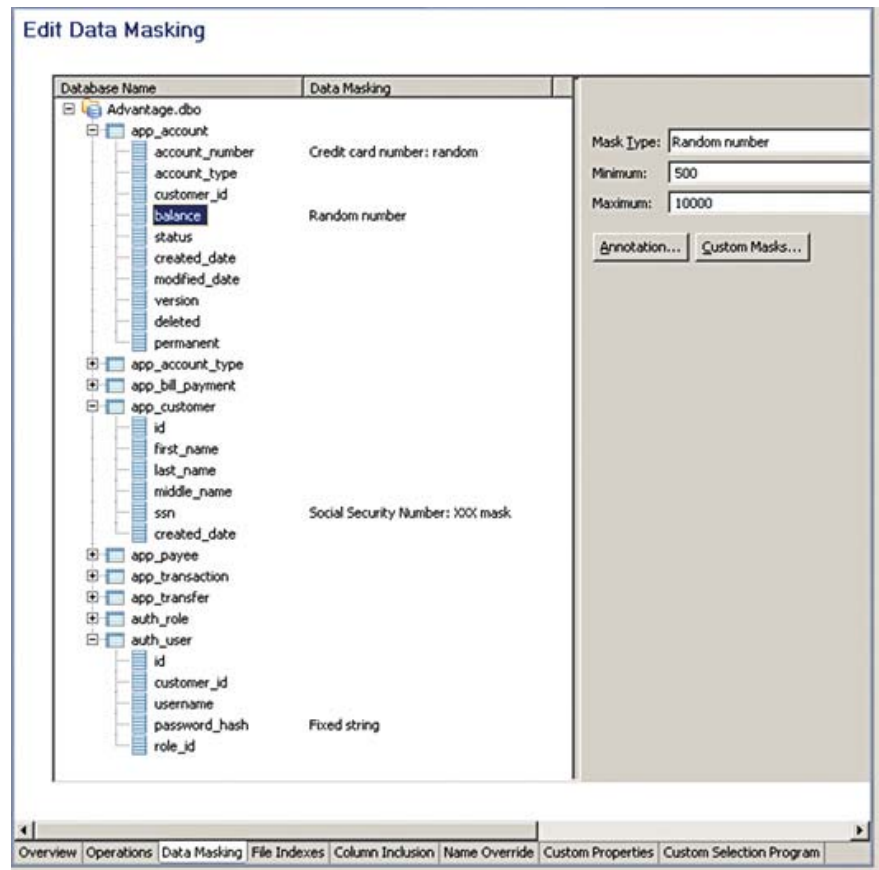
It enables you to set parameters for data extraction rules using factors such as user entitlements. These parameters simplify the process of running use cases and also make rules reusable from one test to the next.

Furthermore, HP Test Data Management allows your users to extract a relevant subset of data from a production source and output the results either as a flat file for use by testing tools or as a database to be incorporated into the test lab for the application under test.

Data extraction can be done on demand or scheduled to run automatically at intervals. As a result, the time required for extracting large amounts of distributed data for testing is greatly reduced.

Use flexible data masking to protect sensitive information

Data masking can help you mitigate the risk of compromising test data privacy in the testing process. Data masking goes by various names, including data obfuscation, de-identification, depersonalisation, scrubbing, scrambling or cleansing. By any name, it allows you to hide, remove or randomize sensitive information.

However, there is a catch. When manual methods are used, data masking is an error-prone undertaking that leaves your business at risk of data security breaches. What's more, manual data masking can be very time consuming. This is clearly an area where automated tools are essential.

HP Test Data Management solves these problems. It supports data privacy and security with built-in functions for masking and manipulating commonly used data types. You can use these functions to randomize credit card numbers and account balances and to mask social security numbers, as shown in Figure 4.

In addition to its built-in data masking and manipulation features, HP Test Data Management enables custom masking and manipulation functions for any data type. You can use scripts to create your own policies for masking or manipulating any data. All of these capabilities help you comply with your corporate data security policies and protect the company and its customers from risk.

Reduce storage maintenance costs

In the testing process, cloning, storing and maintaining an entire production database is a costly proposition. A better solution is to extract just a subset of the data that will be relevant to your tests. This capability, enabled by HP Test Data Management, can greatly reduce your overall storage requirements, decrease the footprint of your storage systems and reduce your costs associated with managing and maintaining test data.
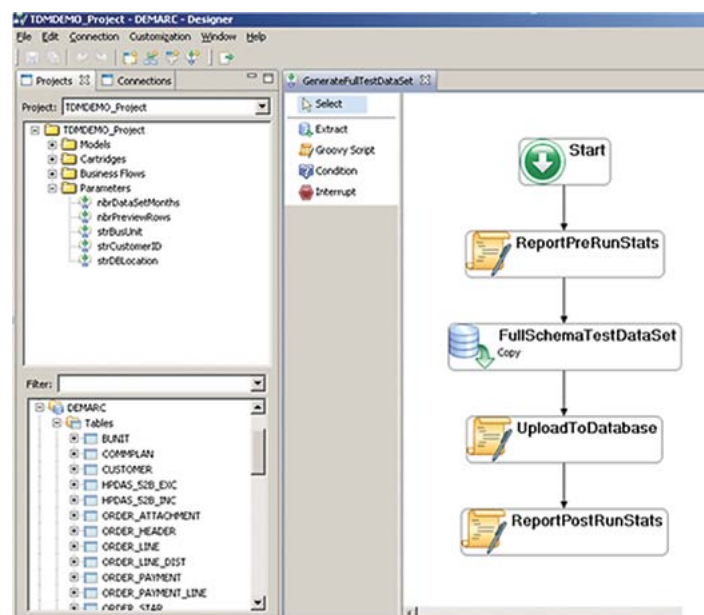
Figure 5: HP Test Data Management Extraction Preview
This feature allows you to take a quick look at your data subset prior to extraction.
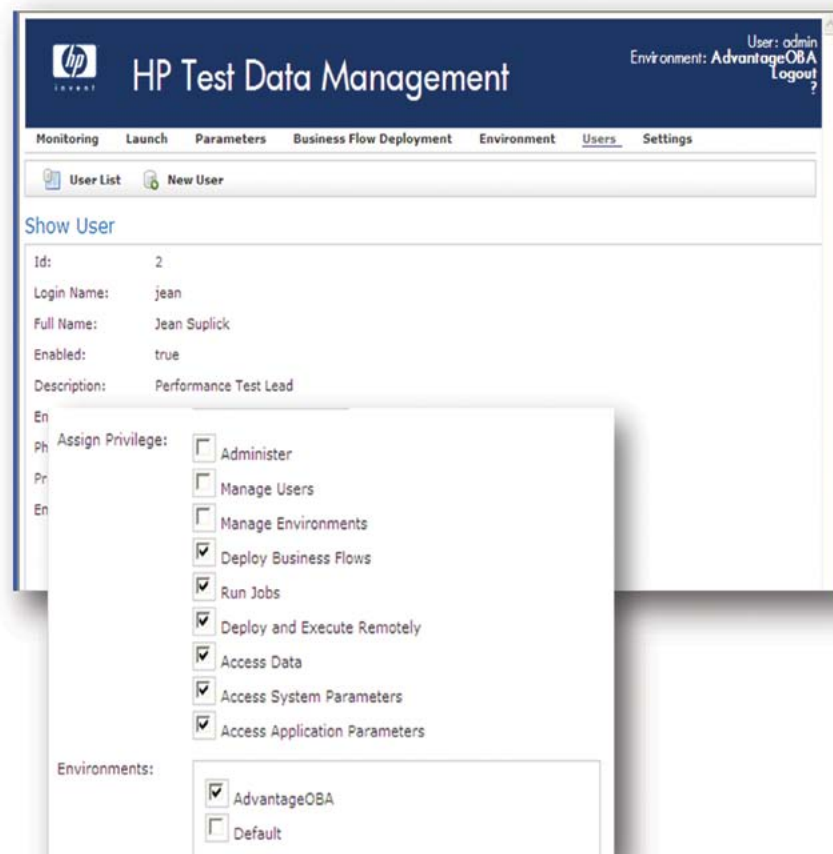

Figure 6: Automating business workflows
The HP Test Data Management Business Flow feature gives you the tools you need to automate data preparation activities.

As the software efficiently handles different data output types by generating output data as either a flat file or a database, while maintaining the structure and schema of the source database, HP Test Data Management allows you to preserve your investments in legacy databases.

Preview data prior to download
The HP Test Data Management Extraction Preview feature allows you to take a quick look at the data that you are about extract without executing the job. This capability can help you save time and resources and reduce human errors.

Automate data preparation with business workflows
The HP Test Data Management Business Flow feature is a workflow process that can automate data preparation

activities. Among other capabilities, this feature enables you to pull relevant data from multiple sources, manipulate data during extraction, apply masking policies and notify users of the end result. Intuitive drag-and-drop functionality makes it easy to automate the workflow process.
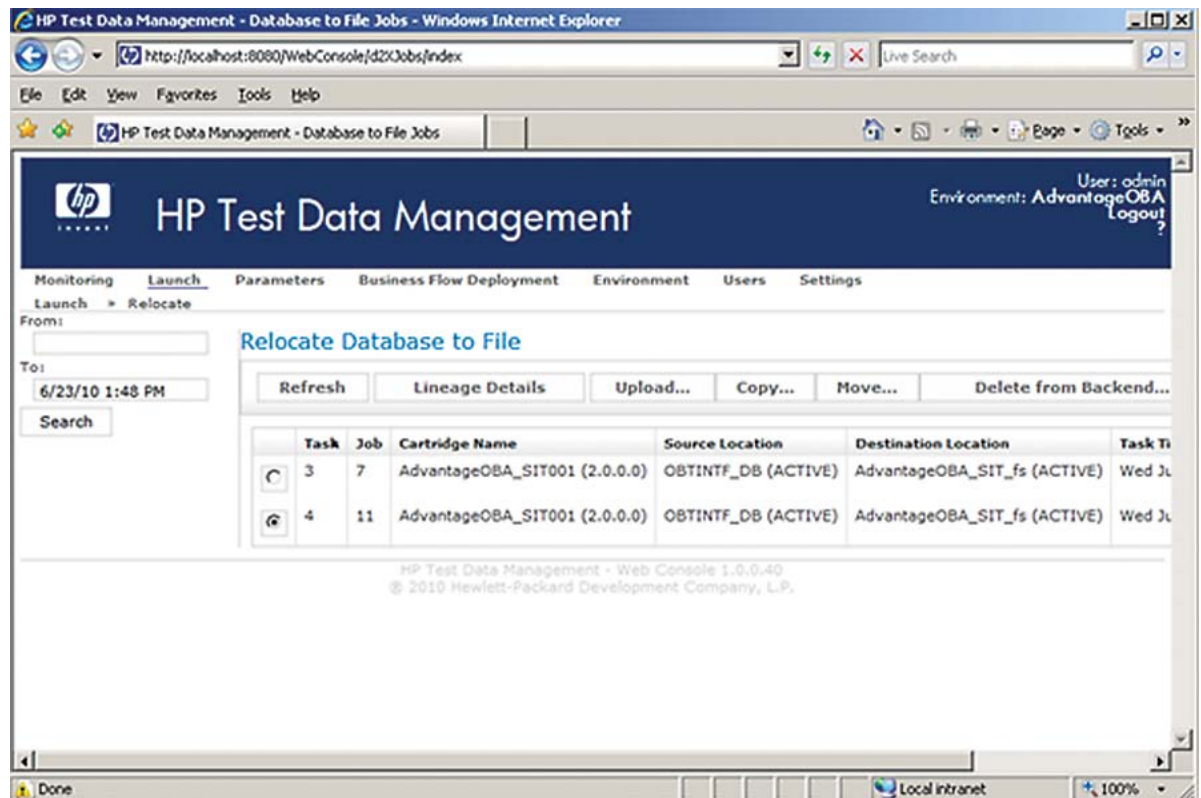
Manage user authentication and authorisation
The HP Test Data Management Administration Console provides administrative capabilities to manage user authentication and authorisation policies, as shown in Figure 7.

The console also contains business flows, configuration information and historical data that can be used to generate rich reports. In addition, users can monitor extraction jobs using the same console.

Provision test environments rapidly

HP Test Data Management makes it easy to relocate data to a file or different database so you can reuse it in another environment. After you have pulled data from a production system, the software's environment provisioning capabilities allow you to provision a test database with a few clicks of a mouse. The software maintains the referential integrity of the data.

HP Test Data Management is designed to complement the suite of HP testing tools, which includes HP Quality Center, HP QuickTest Professional, HP Service Test and HP LoadRunner. Once the data model, subset and masking rules have been defined, you can extract data on demand and use it in these other applications.

With HP Test Data Management, both developers and QA specialists gain a subset of realistic data, which increases the quality of application testing. The data is extracted to either a flat file, usable by any testing tool or a database, which can then be provisioned into the test lab for use by the application under test and corresponding testing teams.

Key takeaways

Today's manual test-preparation methods pose several challenges. Realistic data is hard to collect and sort, and it is typically not readily available for reuse. Storage maintenance costs are high and there is a serious risk that sensitive data from production systems could be compromised in the testing process. These QA challenges can slow the testing process, drive up costs and expose your business to legal and regulatory risks.

HP Test Data Management helps you overcome these challenges. Using automated data sub-setting and masking, the software enables you to accelerate test-data preparation, reduce storage costs and mitigate the risk associated with tests that use sensitive customer data. In doing so, HP Test Data Management supports the overarching goals of increasing business agility, reducing costs and mitigating risk.

Reduce costs and mitigate risks with HP Test Data Management. Visit us on the Web:
www.hp.com/go/tdm

11

Get connected