

Option 1: Test Data Management

Background

Organisations rely on the data they collect and store for the success of their business. Amazon has an average daily sales revenue of \$1.6 billion USD, which is powered by millions of database operations every minute. Every time a customer visits a page, performs a search, leaves a review, or makes a purchase, the database layer will be hit. As a result, performance and reliability of the database is paramount – every minute of downtime costs in excess of \$1m. An agile business such as Amazon must be able to make changes safely and quickly to continue smooth operation of the business.

SQL Developers and Database Administrators (DBAs) write SQL to query and manipulate databases, to meet the needs of the business. Since the databases are vitally important to the continuity of the business, these changes must be thoroughly tested. To test changes, they need a test environment that accurately reflects the production database estate. Traditionally this might have been a restored backup of product on a development server, however in recent years data protection legislation has restricted the use of personal data. As such SQL Developers and DBAs need to find alternative solutions to test their changes, such as data masking, sub-setting, and synthetic data generation.

Summary

We would like you to research and design a system that allows a developer or DBA to provision test environments. The tool must allow the user to customise the environment to suit their needs, for example the size of the subset, and what types of masking to perform.

The system must guide the user through the provisioning and masking process, where you need to decide how to balance the user experience and vast configurability of the process. You should consider how distinct roles might require a unique experience, for example DBAs would be much more verse in SQL than a full stack software engineer. You may also wish to consider how AI could assist with the configuration of the test data provisioning process.

You must determine clear requirements for this system and write them up together with any decisions and assumptions you are making. Some Q&A sessions will be arranged with Redgate to allow you to research the requirements. You must also design the system in pseudo code and use trace tables to test your design.

Objective

This task should encourage critical thinking, attention to detail and question asking. It is ok to make assumptions but highlighting the risk with each one. We would like the students to understand that there is more to software development than coding.

Guidelines for Test Data Management

Deliverables

A product artefact, which could be a prototype application or wireframes. The artefact must demonstrate the user interactions within the system and highlight the unique features of your idea.

There is no expectation or merit to be gained by implementing the underlying system, nor integrating with existing solutions. The prototype should use fake data or hardcoded responses. However, you should consider how integrating with third party components could be surfaced in your UI.

Your artifact must demonstrate:

- You've identified a solution to the underlying problem.
- You've considered the balance of functionality, deliverability, and quality.
- You've thought about potential issues the brief did not cover.

Requirements

- The prototype must guide the user through a workflow of generating a test data environment.
- The prototype must display relevant configuration options to the user as part of the workflow, with an appropriate level of complexity. You should research types of data masking and generation to decide which options should be surfaced.
- The prototype must allow the user to subset their data, so that they can use an appropriate amount of test data.
- The prototype must allow the user to pick which database they would like to replicate, and where this replication should exist.
- You may also wish to consider how AI could assist with the configuration of the test data provisioning process.

Suggestions

We suggest this task takes the form of a point tool that appears as a web app. Outside of that, the assignment is open to how you wish to guide the user through the process. You might wish research existing test data management solutions to help with this.

You might wish to consider how various stages of the process may take longer than others, and how notifications would enhance the user experience.

Useful Resources / Research

- Redgate SQL Clone: <https://www.red-gate.com/products/sql-clone/>
- Redgate SQL Data Generator: <https://www.red-gate.com/products/sql-data-generator/>
- Blog articles, e.g. <https://www.delphix.com/glossary/what-is-test-data-management>

Option 2: SQL Explorer

Large enterprises are complex, they have lots of teams and departments who fulfil their responsibilities to keep the business running smoothly. As a company grows, and more and more services, products, and capabilities are offered, the data estate becomes exponentially more complex. Suddenly, a single SQL Server in the office cupboard turns into hundreds of servers split across on premise and cloud deployments. Furthermore, a myriad of database engines has been employed over the years so there is a mix of SQL Server, PostgreSQL, MariaDB, and Mongo to look after.

Despite trying their best, a DBA team cannot hold all the information about where data lives and who uses it, across the whole state in their heads. Let alone a single DBA who needs to solve a problem. As a result, DBAs need a way to search and navigate through their estates, so that they can quickly identify where they need to fix something. They also need to understand the relationships between tables and databases, so that they do not accidentally delete something that another service relies upon.

Summary

Research and design a tool that allows DBAs and SQL Developers to explore the servers, databases, tables, and columns within their estate. It should also help them understand the relationships and dependencies between them.

The system must allow the user to search their estate and discover objects. You should consider how best to present the database objects to the user, especially where dependencies exist between the objects. You may also wish to consider how an AI powered chatbot might be beneficial to the workflow.

You must determine clear requirements for this system and write them up together with any decisions and assumptions you are making. Some Q&A sessions will be arranged with Redgate to allow you to research the requirements. You must also design the system in pseudo code and use trace tables to test your design.

Objective

This task should encourage critical thinking, attention to detail and question asking. It is ok to make assumptions but highlighting the risk with each one. We would like the students to understand that there is more to software development than coding.

Guidelines for SQL Explorer

Deliverables

A product artefact, which could be a prototype application or wireframes. The artefact must demonstrate the user interactions within the system and highlight the unique features of your idea.

There is no expectation or merit to be gained by implementing the underlying system, nor integrating with existing solutions. The prototype should use fake data or hardcoded responses. However, you should consider how integrating with real databases would impact your design.

Your artifact must demonstrate:

- You've identified a solution to the underlying problem.
- You've considered the balance of functionality, deliverability, and quality.
- You've thought about potential issues the brief did not cover.

Requirements

- The prototype must allow the user to connect to add a new database to the program
- The prototype must allow the user to search across the estate to discover database objects, that are relevant to their search, or the query they are writing.
- The prototype must visualise dependencies and relations between objects, using an appropriate diagram.
- The prototype must present filters to the user so they can refine their search, e.g. filter by server/database/... (you should research what filters are relevant).
- You may also wish to consider how AI could assist with the searching and discovery of database assets.

Suggestions

We suggest this task takes the form of a point tool that appears as a desktop or web app. Outside of that, the assignment is open to how you wish to present the search results and dependencies to the user. You might wish to consider the UX paradigms of search applications you use for programming.

Useful Resources / Research

- SQL Search: <https://www.red-gate.com/products/sql-search/>
- SQL Dependency Tracker: <https://www.red-gate.com/products/sql-dependency-tracker/>
- GitHub Code Search: <https://cs.github.com/>