

WHITE PAPER | APRIL 2016

# CA Test Data Manager Key Scenarios

Generate and secure all the data needed for rigorous testing,  
and provision it to highly distributed teams on demand.

Muhammad Arif

Application Delivery, CA Technologies



## Table of Contents

<b>Executive Summary</b>	<b>3</b>
<b>Section 1</b> Generate all the Data Needed for Rigorous Testing	<b>5</b>
<b>Section 2:</b> Store Data Centrally in a Test Data Warehouse	<b>6</b>
<b>Section 3:</b> The Right Data, to the Right Place, at the Right Time, in the Right Format	<b>7</b>
<b>Section 4:</b> Copy and Extract Coherent Subsets of Test Data	<b>8</b>
<b>Section 5:</b> Compliant Data in Test Environments	<b>9</b>
<b>Section 6:</b> Realistic Service and Message Virtualization	<b>10</b>
<b>Section 7:</b> Test Data Management on the Mainframe	<b>11</b>
<b>Section 8:</b> About the Author	<b>12</b>

## Executive Summary

---

### Challenge

Delivering quality software on time depends on having quality test data on demand. However, many organizations still rely on a limited number of copies of production data. These copies rarely contain the data needed for rigorous testing, are costly to maintain and often expose the business to compliance risks.

Though typically high in volume, production data is usually low in variety. The large copies incur prohibitive infrastructure costs to store and maintain and yet typically provide just 10 to 20 percent test data coverage. What's more, they rarely contain the unexpected results, negative scenarios or outliers that are needed for rigorous, real world testing. In turn defects are detected late, creating time-consuming, costly rework.

One thing production data often does contain is sensitive information. Exposing it to non-production environments therefore carries a compliance risk of fines which have previously reached into millions of dollars. Manual data masking and other in-house methods to obfuscate sensitive data are slow, adding further bottlenecks while testers still have to search through large copies of data for the exact data sets they need. Data is frequently also unavailable in parallel, so even more delays are created by upstream constraints.

---

### Opportunity

CA Test Data Manager provides a complete, end-to-end test data management solution which allows organizations to create or secure all the data they need for rigorous testing and provide it to distributed test teams on demand.

Automated data profiling can be used to detect sensitive information stored enterprise-wide while creating a multi-dimensional picture of what data exists and how it's related. Existing data can then be secured using high-performance, native masking engines.

Sophisticated coverage analysis combined with powerful synthetic data generation can be used to create the smallest set of data needed for rigorous testing. This data will then provide future scenarios, outliers and unexpected results, so that testers have all the data they need to detect bugs earlier, where they require far less time and expense to fix.

Test data is stored centrally and is re-usable from a test data warehouse. It is exposed to testers via a self-service web portal, while automated data discovery is used to find and deliver the exact data sets testers need on demand and in parallel.

## Executive Summary

---

### Benefits

- Understand exactly what data exists, and where, to mitigate the risk of a data breach and identify the right data needed for rigorous testing.
- Detect defects earlier to repair them with less expense using rich test data which includes all the combinations needed for comprehensive testing.
- Work towards regulatory compliance without creating additional bottlenecks using high-performance masking engines to secure millions of rows of data in minutes.
- Provide parallel, on-demand access to test data and deliver higher quality software earlier and at lower cost.
- Avoid prohibitively high infrastructure costs by testing using the smallest set of data needed for maximum test coverage.
- Generate rich, virtual data needed for effective message and service virtualization and provide testers with stable environments, free from cross-system dependencies and constraints.

## Section 1

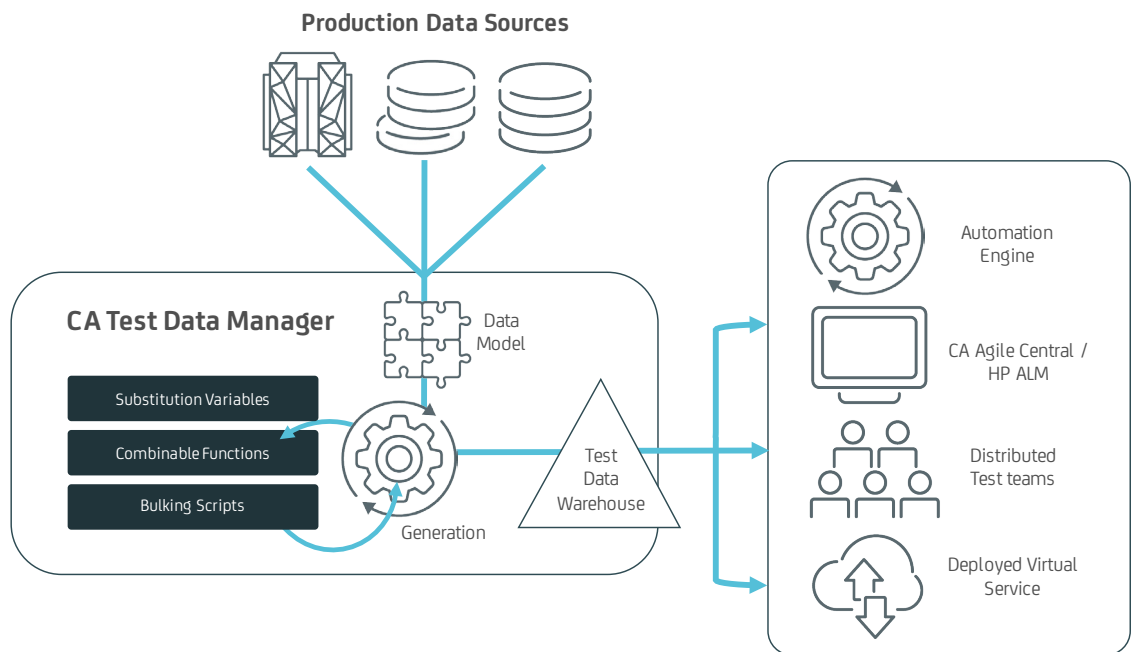
# Generate all the Data Needed for Rigorous Testing

CA Test Data Manager combines powerful synthetic data generation with sophisticated coverage analysis, enabling organizations to create the smallest set of data needed to for comprehensive testing.

Data stored across an entire enterprise can be quickly modeled. Realistic synthetic data can then be generated to fill in the gaps and provide full test coverage. The outcome is that smaller, richer sets of data can be generated for complete testing, while avoiding prohibitively high infrastructure costs. Combinable data generation functions, bulking scripts and substitution variables will automatically create millions of rows of complex, up-to-date data as fast as the database infrastructure can handle. This includes future scenarios and unexpected results, so that testers have the right data to test boundary conditions and avoid the spiraling costs and delays caused when critical bugs are detected too late.

**Figure A.**

Rich synthetic data can be created and fed into test environments, automation engines and deployed virtual services and test management tools.



With CA Test Data Manager, you can:

- Automatically generate smaller, richer sets of test data with full coverage.
- Detect and repair defects earlier, by covering all outlier data conditions and negative paths.
- Test with smaller, richer sets of test data and reduce environment infrastructure costs.

**Section 2:**

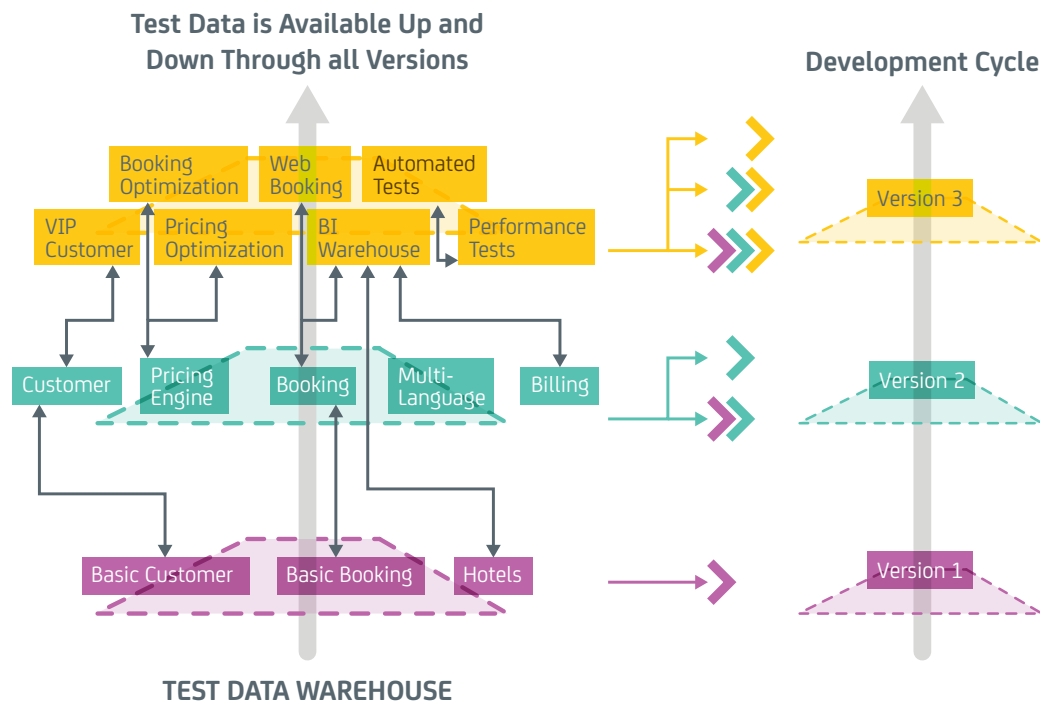
## Store Data Centrally in a Test Data Warehouse

The enhanced data is stored as re-usable assets in a central repository. Existing data and flat files stored across an organization can also be imported with the augmented data being stored and managed from a single location. Re-usable subsets of data can then be extracted, cloned and delivered to test teams on demand, eliminating data dependencies and maximizing the value of work already done.

Sophisticated version control means that relevant data sets automatically reflect and replicate changes made to system requirements across projects and versions. Test and development teams can work with the most up-to-date data, developing multiple releases in parallel, from stable, isolated environments.

**Figure B.**

Data from across an enterprise can be imported and managed centrally.



With CA Test Data Manager, you can:

- Store data pools centrally and share them across distributed teams in parallel.
- Re-use data across versions and releases for efficient regression testing.
- Eliminate delays and rework caused by data dependencies and inconsistencies.
- Assemble existing data objects into new or complex scenarios so that test teams have the data they need to fully test a system.

**Section 3:**

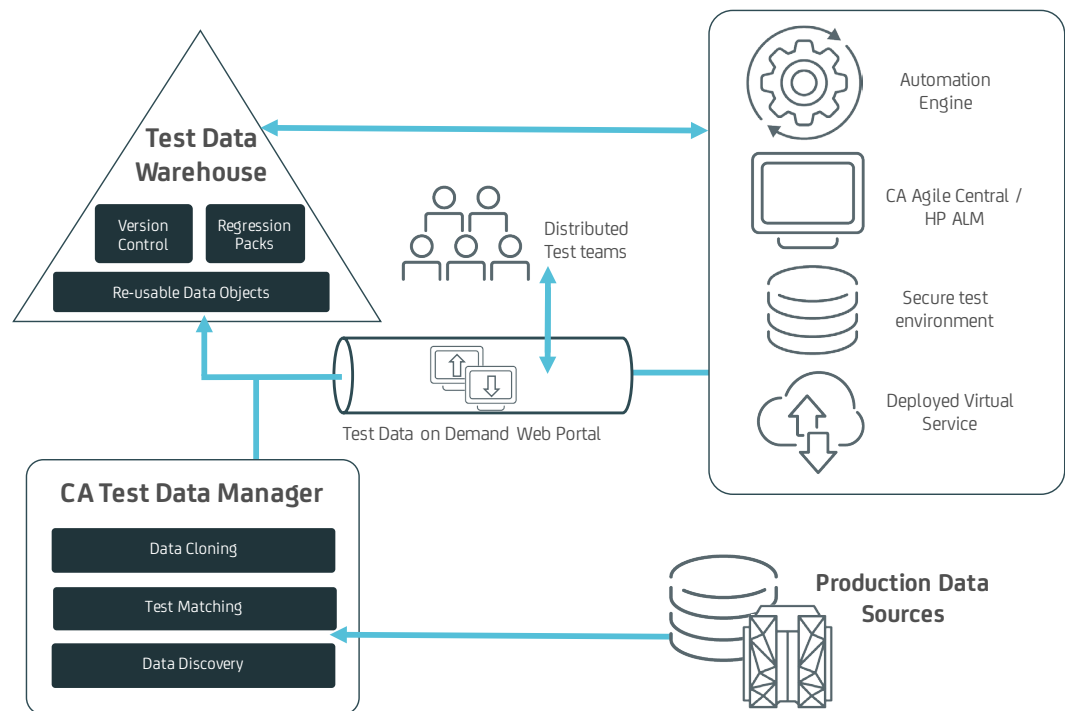
## The Right Data, to the Right Place, at the Right Time, in the Right Format

Automated data discovery means that testers receive the data they need from multiple back-end systems in minutes, eliminating the time otherwise wasted looking for or preparing data, or creating it where none exists. They can request the exact data they need, having it automatically delivered from the central repository in parallel and on demand.

Dynamic form building enables testers to mine data using pre-defined criteria, rather than fixed values, using the Test Data on Demand web portal. This data is “matched” to the exact tests it can run, allowing testers to perform more stable tests, earlier and with greater repeatability. Test teams can work in parallel, detecting defects earlier and accelerating the delivery of fully tested software at less cost to the business.

**Figure C.**

Data is provisioned in parallel and on demand from the Test Data on Demand web portal



With CA Test Data Manager, you can:

- Find and generate the correct data automatically using the self-service test data on demand Web portal.
- Automatically deliver data in minutes and avoid testing bottlenecks caused when manually searching for data or waiting for it to become available.
- Drastically reduce the time and resources required to provision data.
- Link data to specific test cases or directly into automated testing tools like HP ALM or IBM Rational.

“The customers, who had all adapted the agile software development methodology, highlighted that the [CA] Test Data Manager solution provides their teams with the right data at the right time to improve their testing velocity and increase IT resource productivity, as well as reduce the time-to-market of key business applications.”

– Forrester<sup>1</sup>

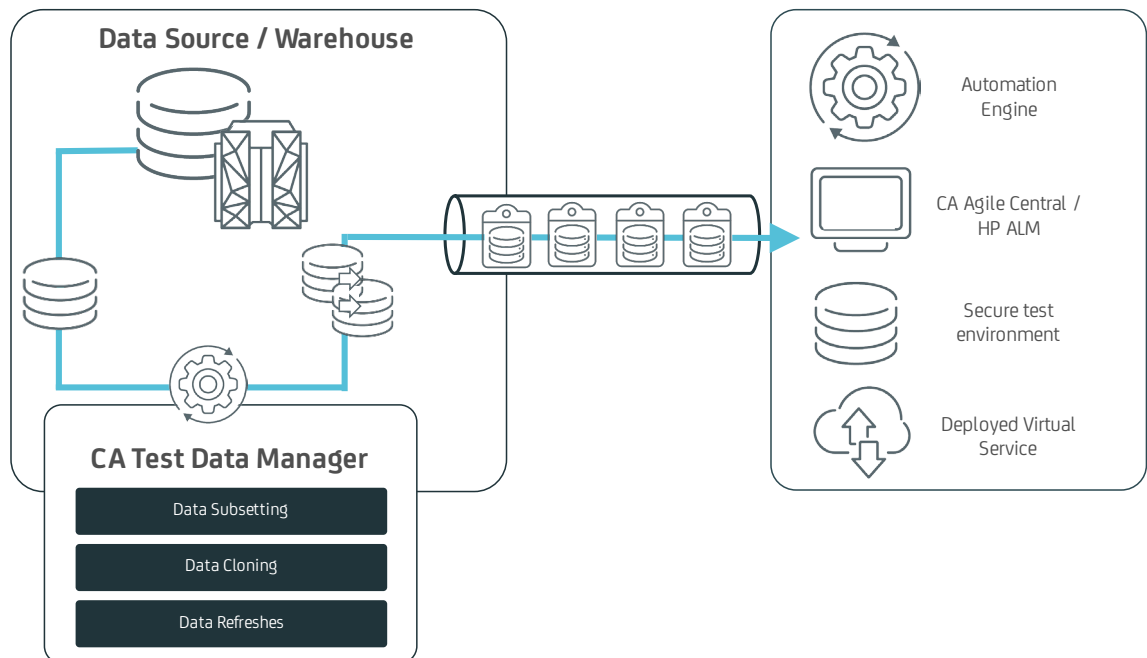
#### Section 4:

### Copy and Extract Coherent Subsets of Test Data

As data sets are provisioned, they are automatically “cloned” so that the original data remains intact. This means that teams can work in parallel, while avoiding the slow and expensive process of manually copying and moving large copies of production data. Testing using intelligent subsets of data further avoids prohibitively high infrastructure costs, while delivering “cloned” data to isolated test environments prevents bug scenarios being lost when one team makes a change. Data is likewise preserved during a database refresh, meaning that multiple test runs can be performed in parallel, delivering valuable software to market earlier and at less cost.

**Figure D.**

As they are provisioned from the test data warehouse, re-usable data sets are “cloned.”



With CA Test Data Manager, you can:

- Avoid the prohibitively high costs of storing, copying and moving large copies of production data.
- Testing using small, rich subsets of test data and reduce infrastructure costs by up to \$50,000 per database.
- Maintain and re-use data after a refresh and run multiple tests in parallel to improve the agility of testing.
- Clone and share bug scenarios, interesting and rare data.



“Our IT pipeline has new releases going into production on a weekly basis. [CA] Test Data Manager has significantly reduced our testing cycles to help meet these requirements.”

– Senior Engineer, Retail Company<sup>1</sup>

## Section 5:

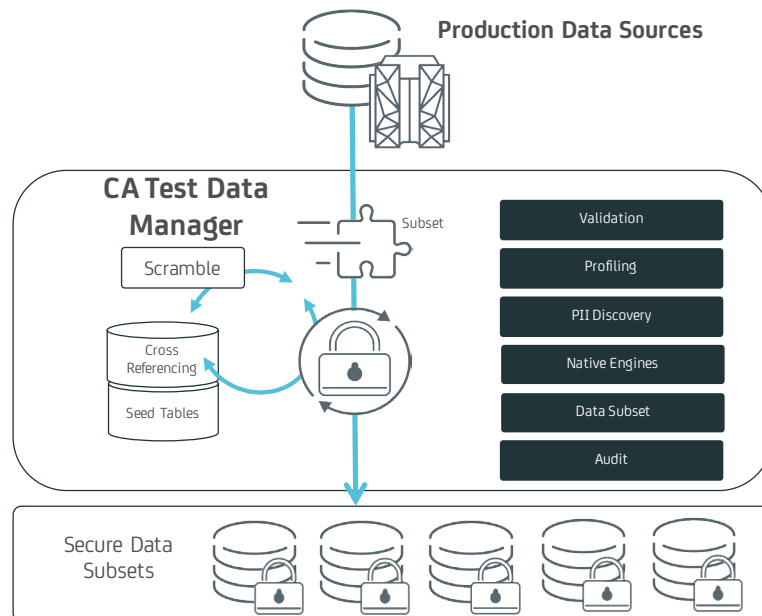
### Compliant Data in Test Environments

With CA Test Data Manager, existing data can be secured in minutes, mitigating risk while reducing the cost of compliance to the business.

CA Test Data Manager provides powerful data profiling algorithms which will automatically discover potentially sensitive information stored enterprise-wide. Over eighty combinable masking functions and four high-performance native masking engines will secure large, complex sets of data in minutes. Referential integrity and complex relationships will be maintained, producing data with all the characteristics of production but none of the sensitive content. Data can also be masked in-flight for secure service and message virtualization, allowing teams to work in parallel, free from dependencies and constraints.

**Figure E.**

High-performance, native masking engines secure large data sets in minutes.



With CA Test Data Manager, you can:

- Minimize the risk of costly data breaches by automatically finding and securing PII enterprise-wide.
- Avoid delays by masking millions of rows of data in minutes using native masking engines.
- Extract subsets of production data, mask it and move it to test environments on demand.
- Create masked services and messages in-flight for secure service and message virtualization.
- Provision compliant sets of test data to test teams and outsource providers in minutes.

“With [CA] Test Data Manager, we are now more flexible and have a strong piece of mind that our customer’s personal information is secure.”

– Manager, IT Data Architecture, Large Financial Services Organization<sup>1</sup>

## Section 6:

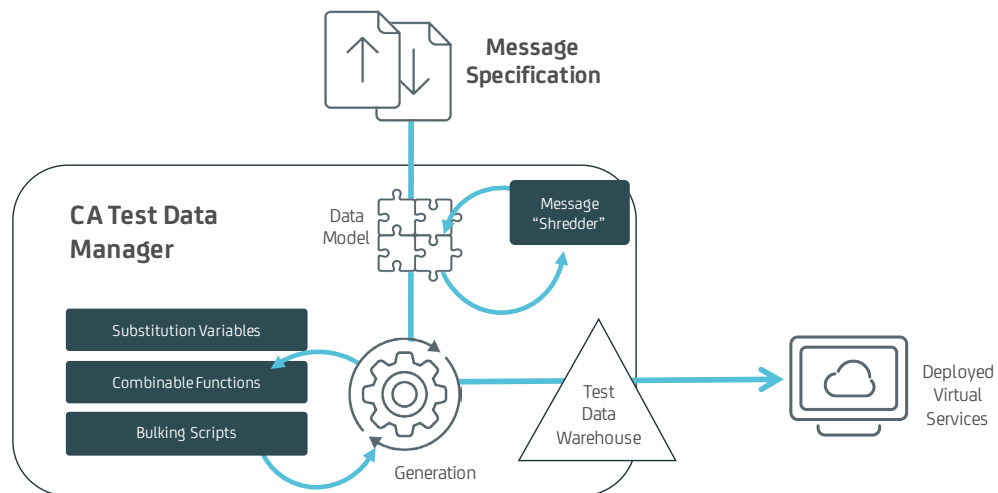
### Realistic Service and Message Virtualization

A fully integrated, end-to-end solution, CA Test Data Manager helps organizations provide testers and developers with the environments they need to deliver quality software earlier, and at lower cost. From a message specification which has been analyzed in CA Service Virtualization, realistic request-response pairs can be generated and injected into a deployed virtual service. This data is synchronized across the inter-dependent databases and services that exist in complex applications, providing testers with access to otherwise unavailable or incomplete components.

A model of the live data can also be created and imported into CA Agile Requirements Designer (formerly Grid Tools Agile Designer), CA’s complete requirements gathering and test case design tool. Automated rule-learning algorithms can then be applied to identify how the data relates and how it flows through the system. “Covered” synthetic responses can be generated so that they cover the full range of possible scenarios. This virtual data can be exposed to test teams on demand through the web-based portal, enabling them to work in parallel, free from cross-system constraints.

**Figure F.**

Realistic request-response pairs are generated and injected into a deployed virtual service.



With CA Test Data Manager, you can:

- Fully test software and APIs, using virtual data which covers every possible scenario.
- Avoid project delays by providing distributed teams with access to the secure environments they need.
- Visualize live and legacy systems, and quickly provide testers and developers with the in-depth understanding they need.
- Expose virtual data to distributed test and development teams on demand, and deliver fully tested systems earlier and at less cost.
- Drastically reduce pre-production infrastructure costs.

**Section 7:**

## Test Data Management on the Mainframe

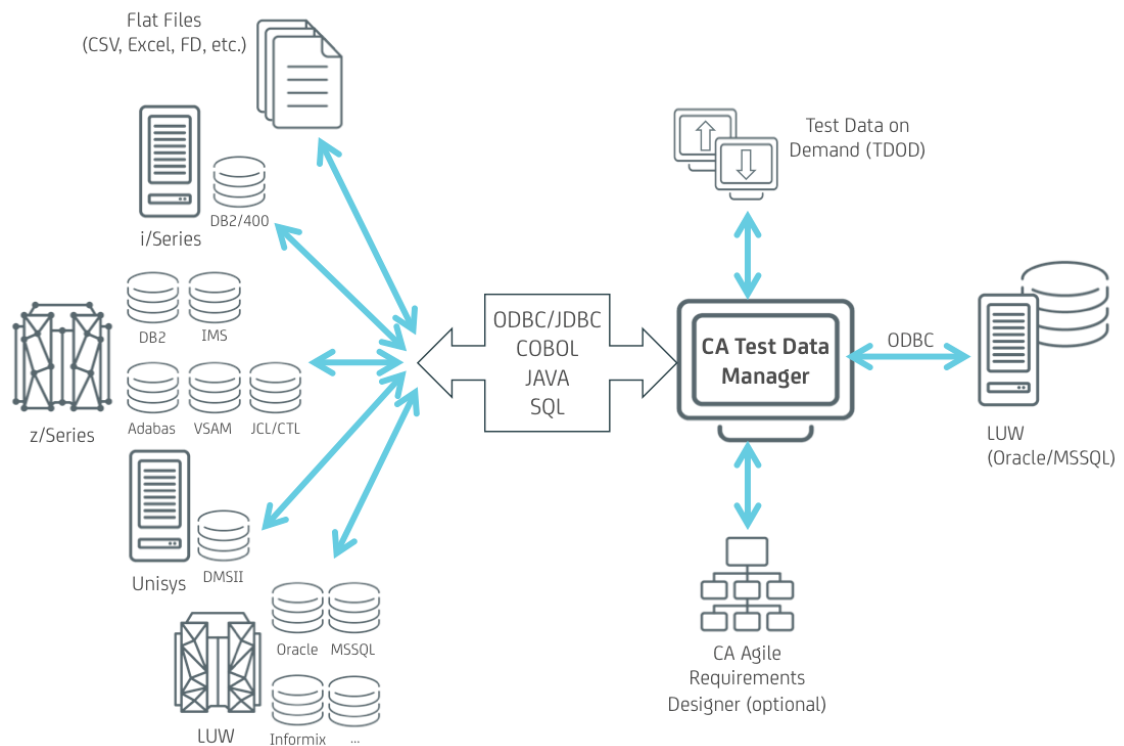
As an enterprise-wide solution, CA Test Data Manager provides a robust framework for managing test data on mainframe and distributed platforms. A single UI and repository is used to define test data engineering tasks, using mainframe batch operations and engines to execute them in mainframe runtime environments. The reliability of the mainframe can thereby be leveraged without investing in multiple tools.

All of CA Test Data Manager's capabilities are available across a broad range of platforms. Data profiling will build a multi-dimensional image of complex data stored across mainframe and legacy platforms. This includes referential information not otherwise available from mainframe sources so that the referential integrity needed for testing can be retained even when faced with massive technical debt and minimal documentation.

Sophisticated coverage analysis can also be applied, to identify and create any missing data needed for rigorous testing, while data can be masked and fed into multiple systems at once. All this data is stored and provisioned from the test data warehouse, avoiding slow manual in-house masking and extraction methods while providing testers with the access to the data they need to deliver software earlier and at less cost.

**Figure G.**

The CA Test Data Manager deployment provides a single UI and repository to define test data engineering tasks across mainframe and distributed platforms.



With CA Test Data Manager, you can:

- Leverage the reliability of the mainframe from a single tool, UI and repository.
- Fully understand mainframe data stored across platforms to pay off technical debt and retain referential integrity during test data engineering tasks.
- Secure mainframe data for testing using native engines and batch operations, or create rich synthetic data from scratch.

**Section 8:**

## About the Author

Arif Muhammad is currently part of the application delivery product management team at CA Technologies. He is responsible for defining and building solutions that can help organizations deliver innovative applications faster and more cost efficiently by leveraging current trends in DevOps, cloud and IT automation. He has more than 19 years of experience in software industry and has held many different software engineering management positions at CA Technologies' research and development organization. Arif has first-hand experience building and delivering enterprise class software applications using both waterfall and agile methodologies. Arif holds an MBA from Hofstra University and MS in Computer Science from New Jersey Institute of Technology.



Connect with CA Technologies at [ca.com](http://ca.com)



CA Technologies (NASDAQ: CA) creates software that fuels transformation for companies and enables them to seize the opportunities of the application economy. Software is at the heart of every business, in every industry. From planning to development to management and security, CA is working with companies worldwide to change the way we live, transact and communicate – across mobile, private and public cloud, distributed and mainframe environments. Learn more at [ca.com](http://ca.com).

<sup>1</sup> "The Total Economic Impact of the CA Technologies Test Data Manager Solution," a commissioned study conducted by Forrester Consulting on behalf of CA Technologies, December 2015