



k2view

www.k2view.com

5851 Legacy Circle, 6th Floor Plano, Texas 75024, USA
Tel: +1 855 452 8439

K2view Test Data Management

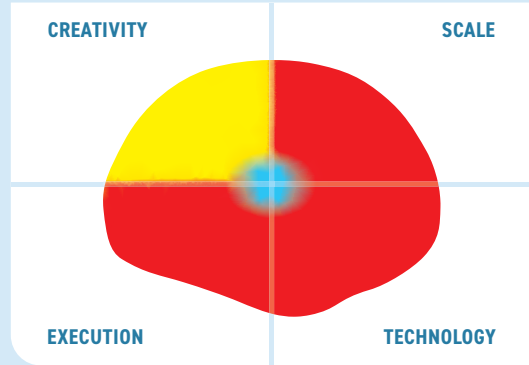
The company

K2view (named for the famously difficult-to-climb mountain) was founded in 2009, and has offices in the US, Germany, Israel, and the Netherlands. It is the developer of the K2view Data Product Platform. This platform originated as an ETL tool, but has since grown into a widely-encompassing data management offering capable of addressing a variety of enterprise use cases. In particular, K2view's Test Data Management offering (K2tdm) is designed to tackle the bottlenecks that slow down testing and development cycles. It does this by addressing test data provisioning delays, eliminating manual data masking, and simplifying complex test environment setups.

What is it?

The K2view Data Product Platform is a unified data management platform that offers numerous capabilities within a single overarching product. It includes solutions for a wide range of enterprise data management problems. Key capabilities include data integration, data governance, data fabric, cloud migration, 360° customer view, and more. Most notably, at least for the purposes of this article, is its ability to provide effective test data management (TDM). This subset of the platform's functionality is referred to as K2tdm, and it includes data subsetting, data masking, sensitive data discovery, and synthetic data generation.

“K2view test data management tools provide a self-service approach for our teams to provision test data on demand – without impacting production source systems.”
AT&T



This Mutability Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

What does it do?

K2tdm organises test data management into three key phases: extraction, organisation, and operation (as shown in **Figure 1**). Test data access follows. Notably, these phases can be repeated – either on-demand or on schedule – to selectively refresh your test data and keep it aligned with changing production environments.

In the extraction phase, K2tdm ingests data from any source, whether structured or unstructured. Business entities (such as customers) are automatically classified within the platform's data catalogue, aptly named K2catalog. This includes a sensitive data discovery process, where what qualifies as sensitive data is controlled by a series of customisable rules and parameters. These rules can match against column names and/or the values and format of the data itself. Any discovered sensitive data is automatically masked within the platform, including highly unstructured data such as documents and images. Persistent masking is available at rest

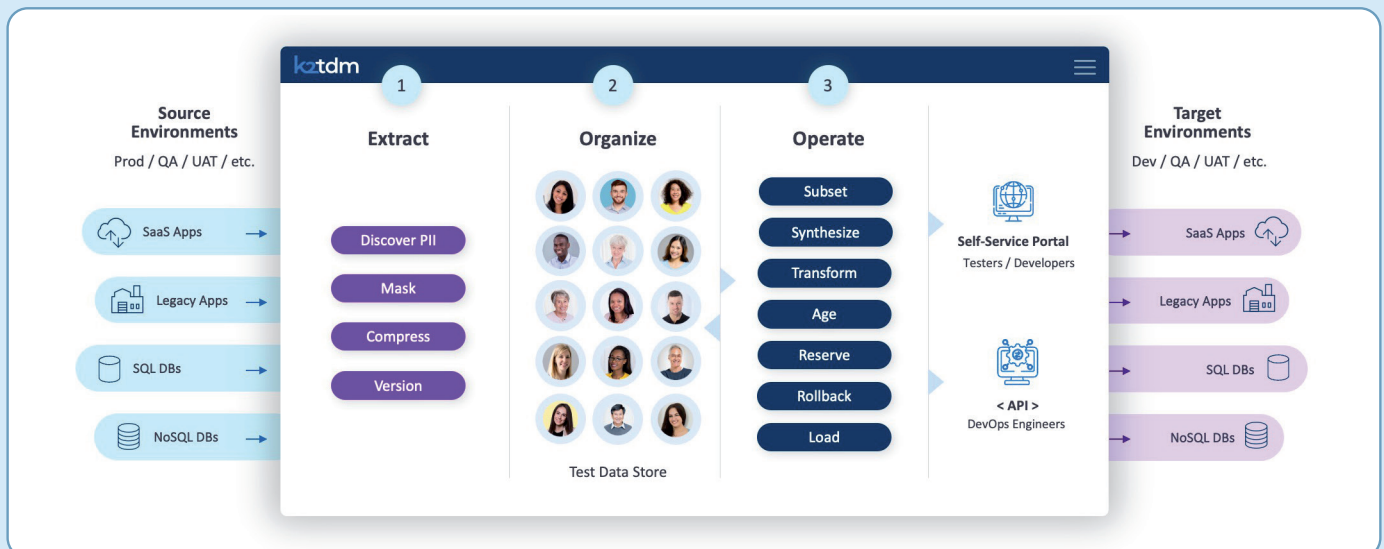


Figure 1 – K2tdm conceptual architecture

“We're collaborating closely with K2view in order to evolve our test data management tools for the purpose of driving our business agility, IT velocity, and attention to customer experience.”
Vodafone Germany

and in-flight, as is dynamic masking with role-based access control (RBAC). Masking is always consistent, maintains referential integrity, and a large number of prebuilt, customisable masking functions are provided. Facilities for data compression and versioning are also offered during this phase.

In the next phase, your imported data is organised into a structure that makes test data (sub)sets easy to create and then provision to the target environment. Specifically, it is partitioned into discrete, customisable business entities (determined by your data model, but typically representing customers), with each entity then used to create a unique “Micro-Database” that contains all the data associated with that entity, often centralising data from several different sources in the process. Each Micro-Database can be likened to a miniature data warehouse, offering a 360° view of the specific entity it represents. Notably, the centralised nature of these Micro-Databases ensures that referential integrity is always maintained further down the TDM process, because the Micro-Database itself is always referred back to as the ultimate source of truth for the entity it represents. Therefore, if you change the data in a business entity, those changes will automatically propagate elsewhere in your testing environment.

The third phase is operation. Now that your source data has been appropriately ingested, masked, compressed, and organised into entities, you can use it to create and provision your test data sets. There are two primary ways to do this in K2tdm: data subsetting and synthetic data generation. For data subsetting, you create a subset of your business entities by filtering them through various customisable business rules, which are created via dropdown menus (in other words, no SQL – or other query language – required) and as a consequence are simple to build. The corresponding Micro-Databases are fetched and combined into your test data set, then provisioned into the target environment. For example, a tester could rapidly select 1,000 customers based on location, purchase history, and loyalty program status.

For synthetic data generation, there are four methods available. The first is rules-based data generation, in which a series of specific, manually-created business rules are used to generate data sets. The second uses machine learning and generative AI to analyse your production data and create a “lookalike” data set, such that the individual entities are completely fabricated but the overall makeup of the data is very similar to the original. The third leverages data masking techniques to create “new” data, while the fourth (described as data cloning) involves duplicating business entities while changing their identifying features. These techniques vary in sophistication, and we would generally consider the latter two to be relatively ancillary:

in most cases, the choice will be between rules-based or machine learning-based data generation, depending on whether fine-grained control or automation is favoured for a particular use case.

Even so, the fact that they are all offered within a single product is a point in K2tdm’s favour and empowers you to choose the best technique for each use case. Data sets can also be created by blending different types of test data, such as masked production data and synthetic data.

Various other test data management functionality is available, including reserving data for individual testers, versioning, performing rollbacks, and so on. The platform’s ETL roots enable various capabilities, including loading and moving test data from any environment to any environment – useful for QA teams performing regression testing – and a wealth of data transformations are possible, including sophisticated masking techniques, data aging, and data enrichment.

Test data is exposed for access via a self-service web portal (see **Figure 2**) designed to isolate its users from the complexities of test data provisioning. This allows your testers, and potentially other users such as developers, to access your test data without worrying about what is going on under the hood. APIs are also provided, allowing you to integrate K2tdm into your existing CI/CD pipelines (among other things).

Why should you care?

K2tdm is a modern TDM solution that has made a big impact on the space in recent years. In particular, the fact that it is a single, unified platform that provides the lion’s share of the test data functionality you could ask for (and plenty of more general functionality, besides, such as the K2catalog), all built on its core framework, makes for a compelling offering. In addition, the organisation of data into business entities and the corresponding creation and usage of Micro-Databases is a standout feature: few others in the testing space offer anything similar. We also appreciate the emphasis on self-service test data access evident in the product’s web portal user interface design.

The bottom line

Between its unified approach and its offering of data masking, subsetting and synthetic data generation techniques driven by entity-driven data modelling, K2tdm is more than capable of addressing a diverse array of test data management use cases, up to and including those found within complex, enterprise-level data environments. In short, it is a powerful test data management solution that is more than worth your consideration.

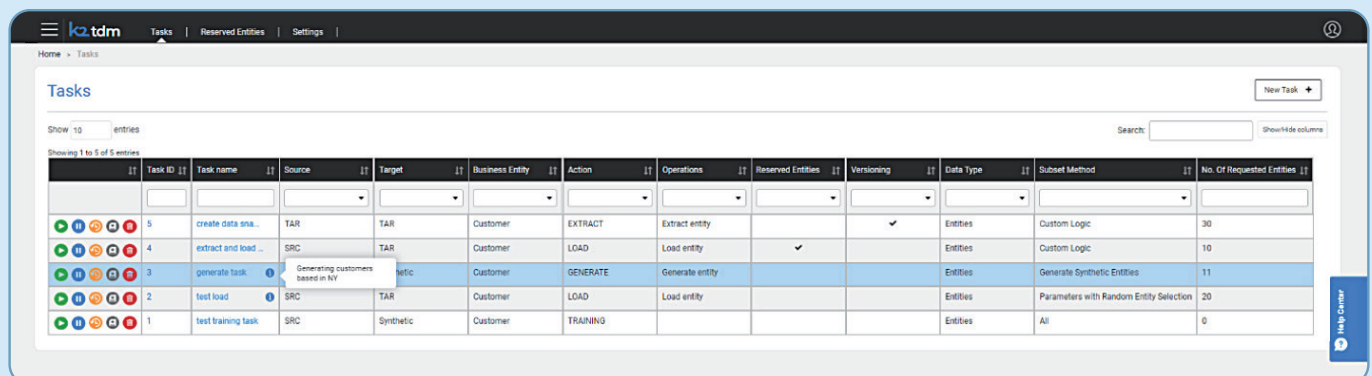


Figure 2 – K2tdm self-service web portal