

12 Pitfalls of Delphix Data Virtualization

> Delphix Test Data Management is fine for some, but those with complex data environments should beware of the 12 pitfalls of data virtualization.

01 Support for a limited set of databases

When we speak with large enterprises about test data management, we ask, "Which database technologies do you use?" The usual response is, "At least one of each!"

Most large organizations use legacy systems, SQL and NoSQL databases, flat files, unstructured and semi-structured documents, and more, while striving to regularly upgrade their tech stack and integrate the latest technologies.

Data virtualization tools, such as Delphix TDM, support a limited set of database vendors and versions (less than 10). According to analyst firm, Gartner: "Modern applications rely on an increasing number of interconnected data stores, applications, and APIs to function, requiring tools to coordinate and synchronize changes while ensuring relational consistency and addressing security and speed mandates".

Data virtualization technology needs to account for the specific nuances of the underlying database file system (per database technology) and carefully replicate this functionality in their software. It's therefore little wonder that introducing support for a new database is a major undertaking for data virtualization vendors, typically requiring many calendar months, at least.

Organizations should carefully assess which technologies are supported by test data management vendors, and check review sites where customers report a lack of transparency and misinformation about supported integrations.



Recommendation

Implementing a virtualized test data management solution can limit a company's access to many data stores technologies, both present, and future. Evaluate Delphix competitors that integrate with any kind of data source.

02

Provisioning a subset of test data

Typically, QA personnel need a specific, well-defined dataset to ensure that they fully test a given scenario. Identifying and preparing subsets is a challenge when data is fragmented across multiple data sources. For example, a desired subset of data could be customers who live in a specific area (data coming from one data source), subscribe to a specific plan (data from a second data source), have spent a certain amount of money (data from a third data source), etc.

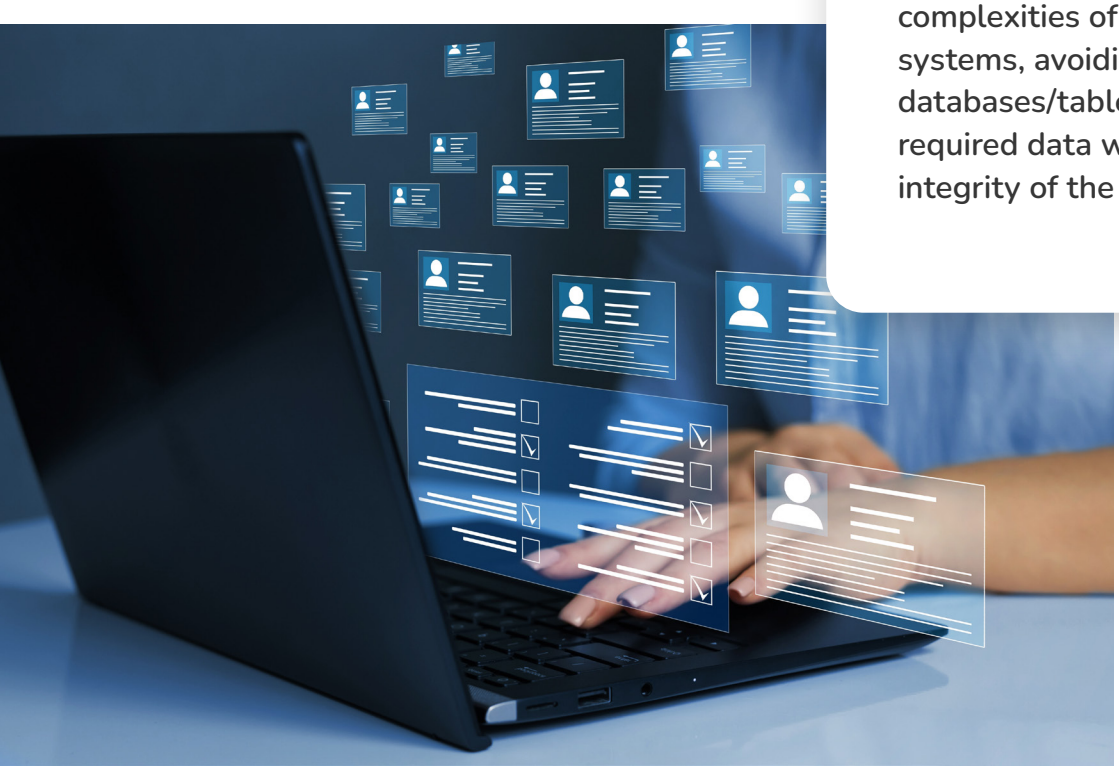
Finding these specific customers can be very time-consuming.

According to research, testers spend 46% of their time analyzing, searching, and manipulating test data.

Unfortunately, virtualized test data management tools are unable to provision a subset of test data (without resorting to complex scripting or 3rd party tools) because they can only provide a 100% replica of the data, by design.

Recommendation

Consider test data management solutions that can subset data based on parameters and business rules, without requiring any code, enabling testing teams to spend their time testing software rather than provisioning test data subsets. Test data subsetting should hide the technical complexities of the underlying source systems, avoiding the need to know which databases/tables/columns contain the required data while ensuring referential integrity of the test data subset.



03 Moving test data from one test environment to another

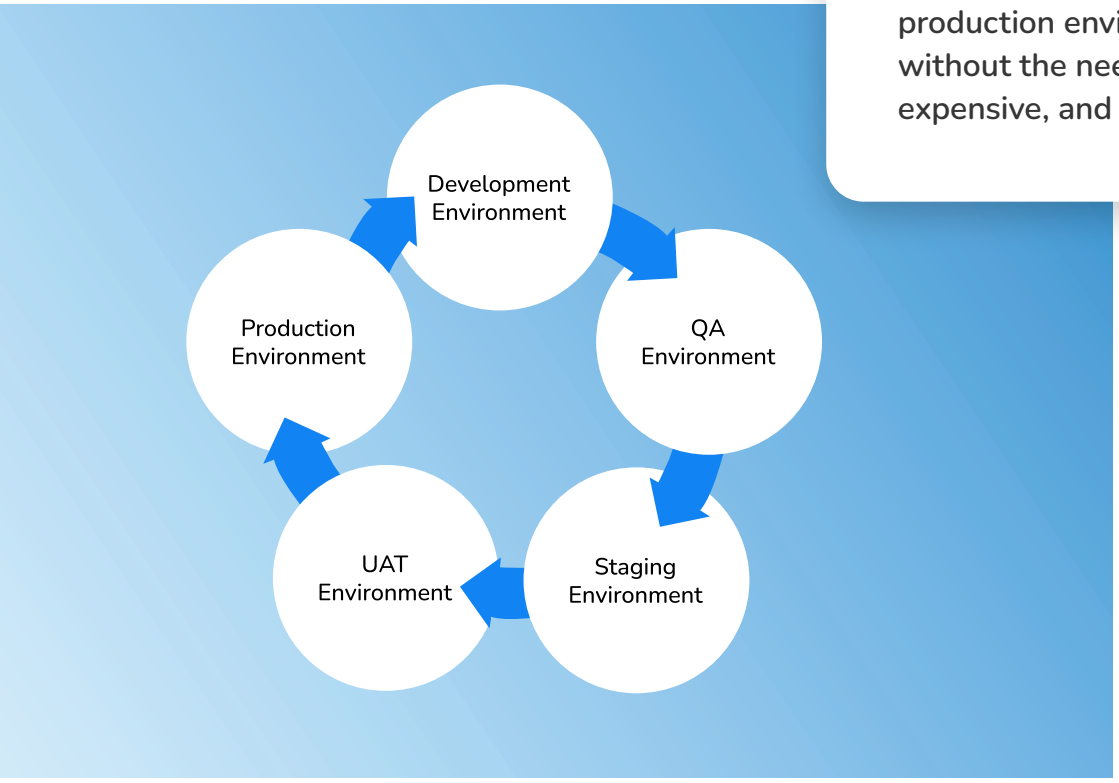
Having spent much time and effort preparing the “perfect” test data set for a given sprint (or test cycle, such as Integration Test, UAT, or Performance Test), testing teams are required to prepare a new testing environment for the next sprint (or test cycle).

With data virtualization solutions, which are dependent on a master replica of the production sources, test data cannot be copied or moved from one non-prod environment to another. This means that the new test environment needs to be rebuilt, requiring a dditional effort and time overhead, and sometimes additional hardware.

Additionally, setting up a new virtual environment in a staging environment, masking the PII, then virtualizing the data is a process that typically takes several days.

Recommendation

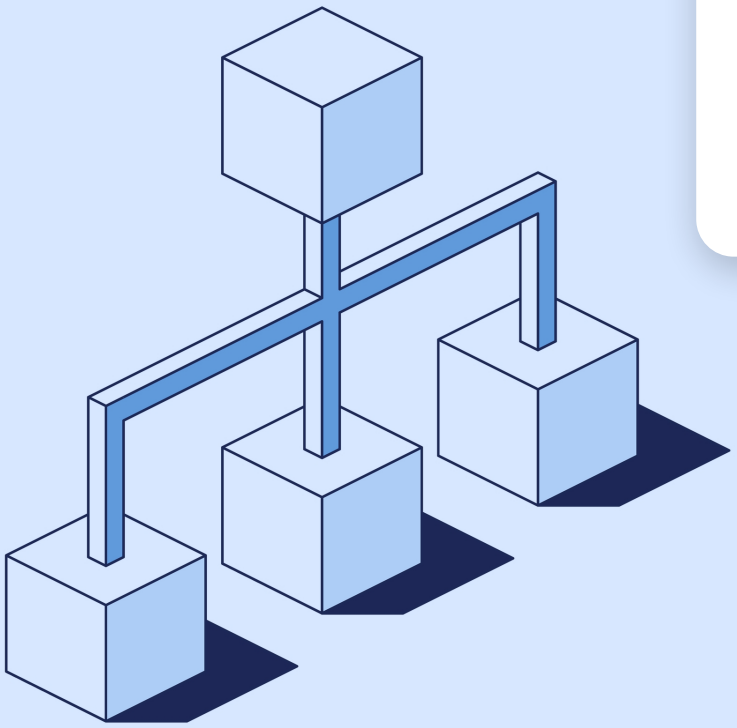
Seek a test data management solution that enables test data to be provisioned directly from any source (including non-production environment) to any target without the need for a complicated, expensive, and long set-up process.



04 Transforming production data into the test environment

Software development is a dynamic, iterative process, especially when an agile approach is used.
New software in a test environment may introduce new database tables or columns that do not exist in the prior production version.

One of the more serious data virtualization pitfalls is that test data cannot be easily transformed to match the target test version. Companies with a data virtualization test data management solution (e.g., Delphix Test Data Management) must write scripts to format and manipulate data, or use additional 3rd party tools for the job.



Recommendation

Ensure the chosen test data management solution is capable of making simple and complex transformations on-the-fly, schema changes, and moving test data from any source to any target.

05 Near-real-time test data

Initial setup and refresh of the staging area for a virtual environment is a time-consuming process: from ingesting or refreshing the data from the source systems to the staging environment, running masking jobs, and then presenting the data to lower environments.

Often times, QA teams need to test software functionality with up-to-date data from production sources. For example, when recreating bugs reported in production, having a near real-time refresh of source data becomes a necessity.

Recommendation

Prefer a test data management solution that can provision near-real-time data from production.

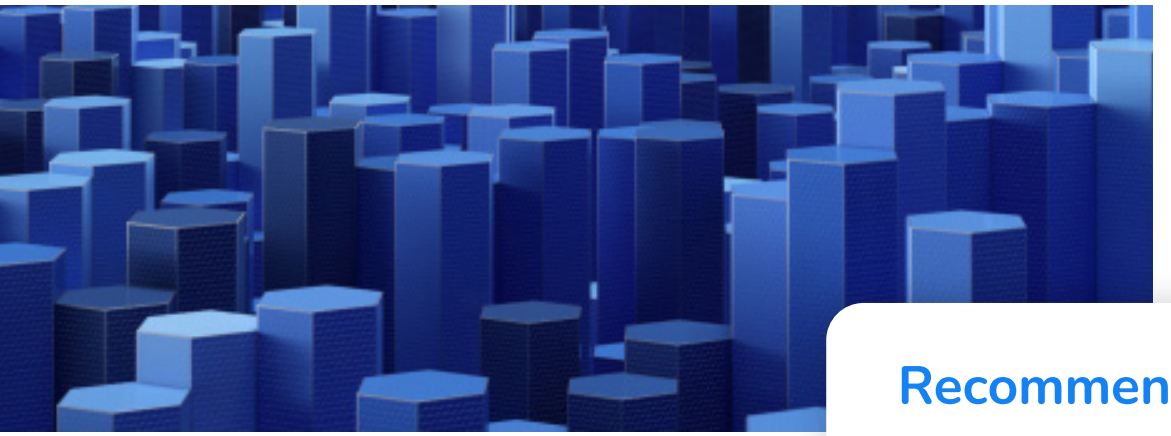


06 Reserving test data for each tester

It's not uncommon for testers to inadvertently override each other's test data, resulting in corrupted test data, lost time, and wasted efforts. Test data must be provisioned again, and tests need to be re-run.

Test data management tools attempt to solve this issue by segregating test data between individuals to mitigate the risk of corruption. Another data virtualization pitfall is that test data can be reserved only per testing environment. Setting up a dedicated test environment per individual tester, with all the time and cost overheads needed to support this, is neither practical nor efficient.

As a result, multiple testers share an environment, and if the test data gets corrupted, the entire environment must be rolled back to a previous version, wiping out all recent changes.



Recommendation

Consider a test data management solution that seamlessly supports data reservation/versioning per individual tester, without requiring additional time and cost overheads for setup.

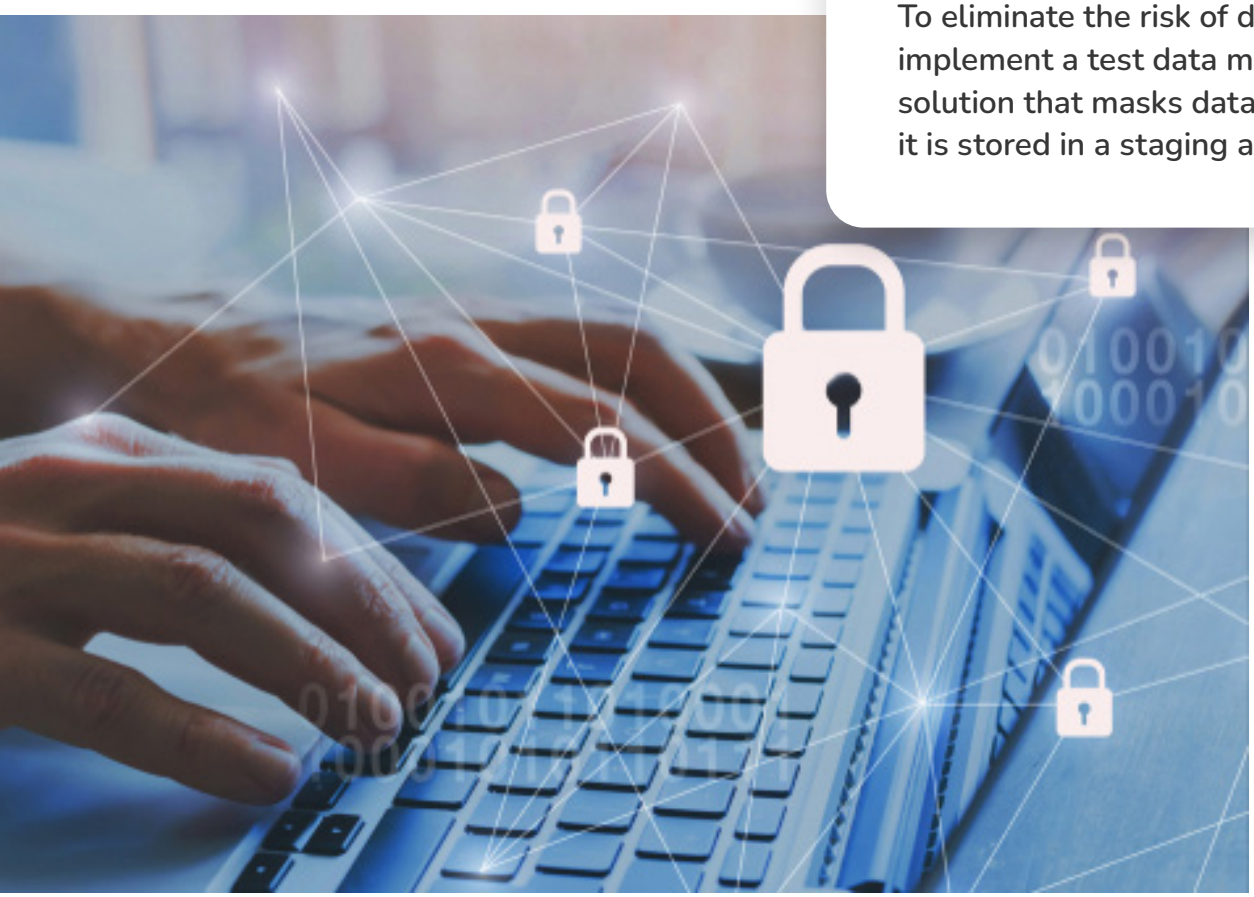
07 Potential data breach for unprotected PII

Virtualized test data management tools require a complete replica of production data, which is compressed into a staging environment.

Once there, PII data must be identified, then masked, and then loaded into development and testing environments. The process can take several days, especially when various data sources are involved, of different technologies. This lengthy process creates a period of vulnerability while the data waits to be masked in the staging environment.

Recommendation

To eliminate the risk of data breaches, implement a test data management solution that masks data in flight, before it is stored in a staging area.



08 Masking unstructured data

It's easy to forget that PII is abundant not only in structured formats within relational databases but also in unstructured formats within files and NoSQL data stores. Privacy regulations do not differentiate between customer information formats.

It's common to find PII in images (such as checks, invoices, or prescriptions), as well as in files such as XML, PDFs and Microsoft Word documents.

Recommendation

Prefer test data management solutions that can implement data masking for both structured and unstructured data, in any database or file system.



09 Consumption-based pricing

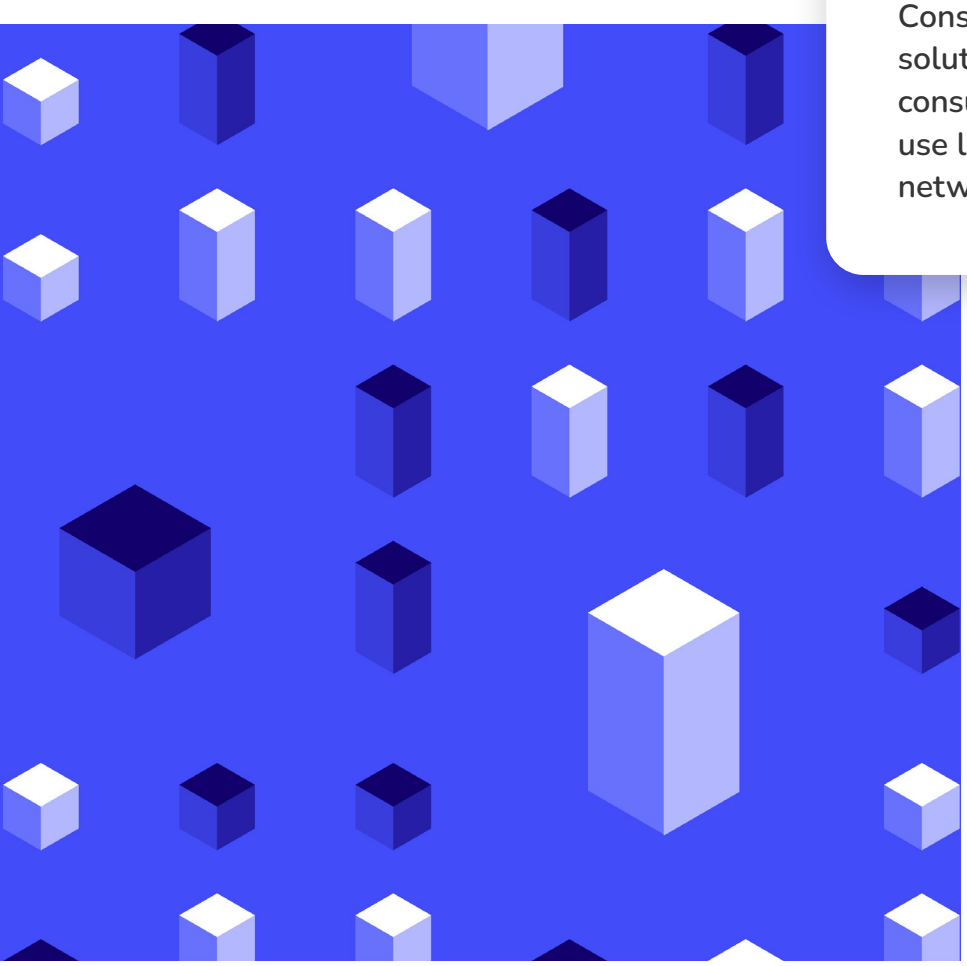
The ability to compress production data to a much smaller footprint within a staging environment is one of the benefits of data virtualization solutions.

When the price of a data virtualization solution is quoted by terabyte, that may seem like an advantage. In larger enterprises, however, the volume of data that must be virtualized can quickly reach gargantuan proportions.

The total cost of ownership can rapidly increase if an organization migrates to the cloud.

Recommendation

Consider test data management solutions where pricing is not based on consumption if your testing environments use large amounts of data and/or if your network traffic is expected to be high.



10

Test scenarios that change test data, at scale

With virtualized test data management tools, the more changes are made to the data during testing, the greater the volume of “deltas” that must be managed in each of the target environments.

This leads to a drain on performance and hardware resources. At some point, the test environment may become unusable, and require a full data refresh.



Recommendation

Before selecting a test data management solution, assess the volume of data changes that will be made to the test data. The more changes are made, the greater the volume of test data that is physically managed by data virtualization solutions per test environment.
The additional costs of storing and transferring test data changes, especially when cloud-based systems are tested, should be factored into the total cost of ownership.

11

Expensive hardware

Don’t forget to include hardware costs when evaluating Delphix Test Data Management and other test data management solutions and calculating TCO. As the scale, the number of data sources, and the complexity of tasks increase, so does the need for performance from the system. Shared virtual resources often require substantially more time to complete a task.

It’s often necessary to use high-end hardware when implementing a data virtualization solution, especially if there are numerous databases and environments involved. The cost of servers and storage is typically higher than advanced alternatives.

Unintended server sprawl is a major cause of concern since physical servers require time and resources to set up and manage.

Finally, because virtual servers can be provisioned quickly, DBAs create new ones every time the QA team requests a new dataset since it allows them to provide the service quickly and easily. The server administrator who should be handling five or six servers must handle over 20 virtual servers.

This can lead to a major disruption to operations and forced termination of certain servers, leading to a test data loss.



12 Synthetic test data generation is a separate product

Synthetic test data is growing in popularity as it is useful for testing new functionality and new applications.

Data virtualization vendors such as Delphix must partner with 3rd party synthetic data generation solutions to address test data generation.



Recommendation

Ensure that your test data management solution embeds synthetic test data generation capabilities, without requiring separate product implementation and license costs.

Alternatives to Delphix Test Data Management

While data virtualization platforms can initially appear “magical”, with virtual instances spun up quickly, there are many pitfalls to consider when implementing a test data management solution based on data virtualization.

A new business entity approach to test data management challenges provisions test data by business entities, such as customers, orders, loans, devices, or anything else that’s central to the applications being tested.

Business entity data is automatically collected from any and all underlying sources, masked in-flight, and stored in a compressed, secure test data store. DevOps and QA teams can then instantly subset, reserve, generate, roll back, transform, or age the test data they need - either via a self-service portal or via APIs to integrate test data into their CI/CD pipelines.

With an entity-based test data management solution, DevOps and QA teams can create compliant datasets to handle any test scenario with full coverage, on demand.

Each tester can reserve test data for their exclusive use, thereby avoiding collisions with other testers. If something goes wrong, a tester can roll back quickly to a previous test data version.

Most importantly, users can subset any data according to business logic, without needing to know the location and underlying complexity of the data sources.

Avoid the data virtualization pitfalls associated with tools such as Delphix Test Data Management, with an entity-based test data provisioning solution that enables shift-left testing, and that significantly reduces testing costs, accelerates software delivery, and improves software quality.

K2view TDM architecture





About K2view

At K2view, we believe that every enterprise should be able to use its data to be as disruptive and agile as the best companies in its industry.

We make this possible by enabling data teams to transform fragmented data into complete and compliant data products – in real time.

Data products could be customers, products, suppliers, orders – or anything else that's important to your business. We manage every individual data product in its own secure Micro-Database™, continuously syncing it with all source systems, and making it instantly accessible to authorized data consumers. This is all made possible by our Data Product Platform, which delivers a trusted, real-time view of anybusiness entity. The platform deploys in weeks, scales linearly, and adapts to change on the fly. It supports modern data architectures, such as data mesh, data hub, and multi-domain MDM - in on - premise, cloud, or hybrid environments.

This one platform drives many use cases, including application modernization, cloud migration, customer 360, data privacy, data testing, and more – to deliver business outcomes in less than half the time, and at half the cost, of any other alternative.

K2view is honored to have been selected by Gartner as one of the world's top data integration vendors.

© 2023 K2view. All rights reserved. K2view and Micro-Database are trademarks of K2view. Content subject to change