Zetaris

Technology White Paper

Contents

Why Zetaris?

Data is an enterprise's most valuable asset. When data is viewed across the business and put in context, business leaders make better decisions, automated decisioning with AI is enabled, and competitive differentiation is increased. Businesses need to be able to see and work with all the data across the entire data landscape. And, as AI matures, real-time in-memory data needs to be harmonised and put in context with the historical data, or secondary memory about the business.

The ability to democratise data so that business consumers can easily access all of it, create their unique view and collaborate using their favourite tools, is crucial for generating accurate decisions and, ultimately, business value. Data-driven enterprises need to stitch data across the entire ecosystem into an analytical data fabric to stay competitive and drive differentiation.

For most organisations, this scenario is not the reality. Data is scattered all over the organisation, in silos and hidden in external clouds or partner environments. Complex data pipelines, usually multiple versions for each business process, are developed to copy data from one place to another to bring data together. This data and process duplication happens when businesses create physical data lakes or data warehouses. Massive cost-generating duplication of data, pipelining processes, people, platforms, and tools are the norm for data projects across every industry.

So, what's the way forward for business decision-makers who want a single, real-time view of all the data in an accurate and governed way? The answer is Zetaris: The Networked Data Platform – one single API for all your data.

Zetaris: The Networked Data Platform for Virtual Data Warehousing

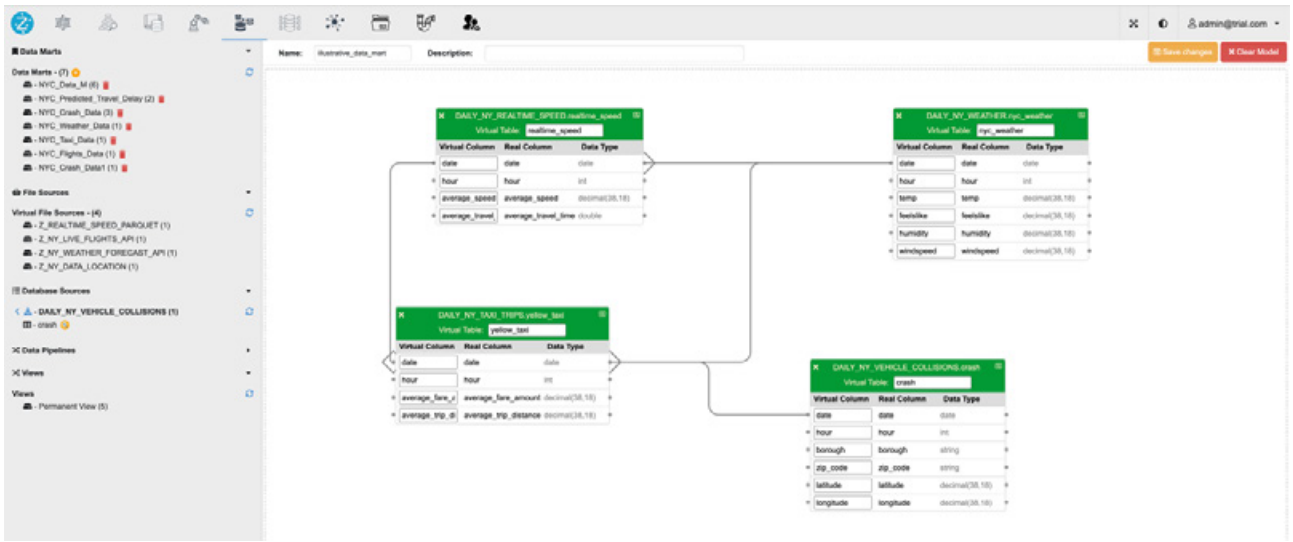
A virtual data warehouse joins data across many data stores and networks or clouds to create the views that the analytical tools need in real-time without the data, process, or systems duplication. This is a step change in the data platform and integration world, where the old approach means data has to be copied from its original source, re-structured or transformed and made consistent before any value can be created.

In the traditional data lake or data warehouse approach, data quality problems emerge from mistakes made during the costly physical data integration coding. Zetaris has removed these cost inefficiencies and technical barriers to analytics and BI.

Zetaris changes everything! We don't move or duplicate data for analysis. We implement a virtual data warehouse and deal with data quality and governance in real-time at the source.

Use of Zetaris: The Networked Data Platform with BI and analytical tools massively extends the breadth of the data that can be analysed, from a limited set within the tool or physical data warehouse to all the data across the organisation or Internet that is relevant and available. Connecting the Zetaris software layer to any data source across a company and the Internet allows new, virtual views of the data to be created without moving or copying data from place to place. We call this, "creating an analytical data mesh". An analytical data mesh put in context with a business is called a Virtual Data Warehouse.

Leveraging Zetaris' mixed-streaming analytics capability (the ability to analyse data in the stream with data in the warehouse or lake in real time) means customer-facing applications and AI can operate with all the data in real time, harmonising historical data.



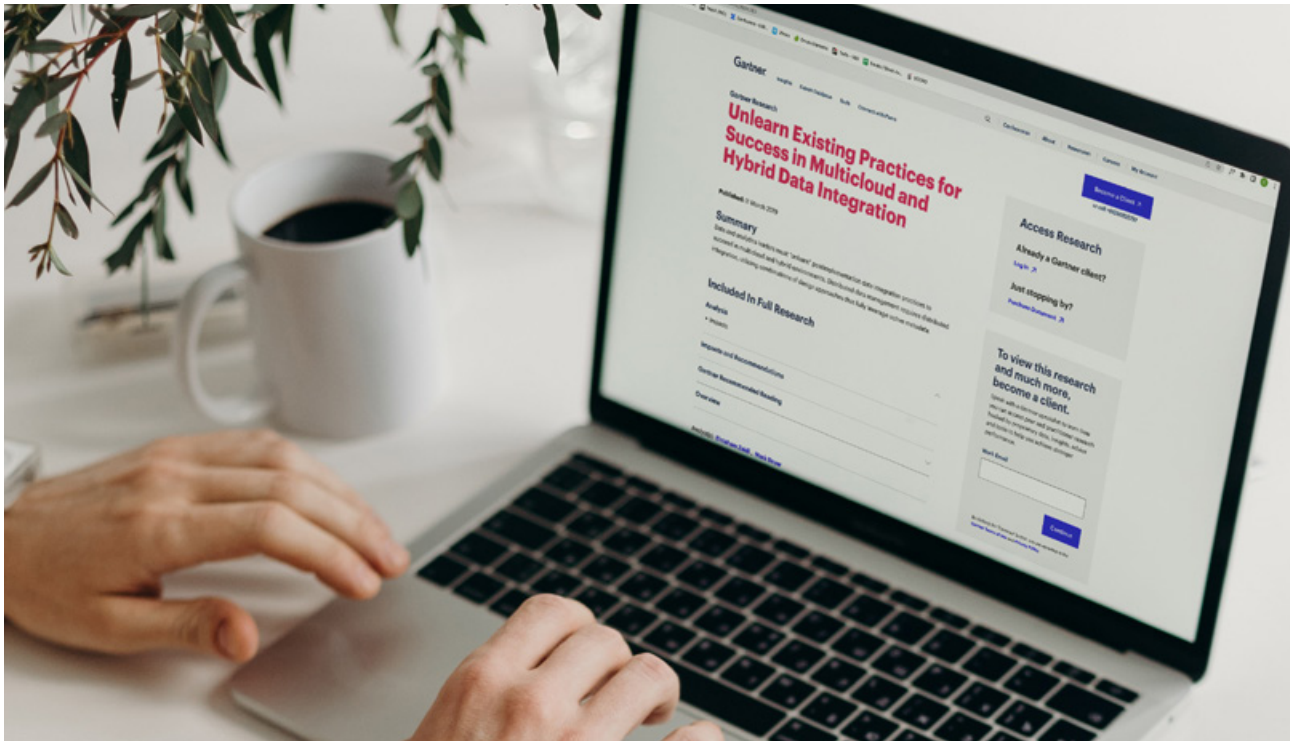
The industry is shifting – Gartner

According to Gartner, data mesh technology represents a seismic change in the data platform industry. Gartner cites Zetaris as one of three global players in the new Data Mesh space. In the Gartner research paper released in March 2019, Gartner advise that organisations must:

“unlearn existing practices for success in multi-cloud and hybrid data integration. By 2020, organisations utilizing data virtualization as a data delivery style will spend 45% less than those who do not on building and managing data integration processes for connecting distributed data assets.”

Data lakes are dead because they are slow, costly and prone to data quality issues. So, if your IT folks are planning a centralised data platform for analytics, stop!

Read the complete article from Gartner:
https://lnkd.in/fEYbVRM



Key Benefits

Zetaris reduces the need for physical data integration by virtualising data in silos while applying governance, quality, and privacy rules. Aggregations and views of data in different places (databases, API links, data marts or lakes) are simply joined and viewed virtually (no movement) as users create a data mesh for consumption by all the tools and utilities they like to use.

Zetaris enables the creation of virtual data warehouses, virtual data marts, and virtual data lakes through an industry-leading set of capabilities.

No data movement, no data duplication – Data Warehouse Virtualisation

Zetaris is a middleware layer that connects to any system, regardless of language, proprietary data structure or processing framework (streaming, disk-based, in-memory, cloud native, etc.). Zetaris takes a user query (from common apps or directly through the Zetaris Lightning client) and processes it by deploying intelligent sub-queries (re-write, optimised, streaming aggregated views, virtual schemas, etc.) to join data across systems without moving data or creating further data duplication on the source systems through the creation of local aggregations. For extreme performance, Zetaris creates temporary in-memory cached views that are deleted or persisted depending on what our Intelligent Heterogeneous Query Optimiser determines. This is a major differentiator from other data virtualisation technologies that need to aggregate data locally on source systems.

The technical benefits of the Zetaris system include:

- ✓ Creation of virtual data structures rather than costly proliferation of physical data across the organization.
- ✓ Reduction of compute and storage cost in the data centre or cloud.
- ✓ Creation of a hybrid virtual data warehouse with data in different clouds (or data centres) to reduce cloud vendor lock-in and risk.
- ✓ Movement of your business logic away from vendor infrastructure, thereby enabling a “no lock-in” architecture for your data platforms.
- ✓ Reduction of your electricity consumption for data to help save the planet.

Key Benefits

1. Streaming in-context view creation

Zetaris enables a 'No ETL' in-stream analytical capability where users can create a lens into streaming data, before it's landed, to deploy powerful data-driven applications and transformation. Define unlimited aggregation tables from your streams without extra cost or additional infrastructure. Zetaris supports aggregation tables which automatically update when new data arrives.

Benefits:

- Blending historical data with real-time data for better real-time customer interaction.
- Pre-processing data in the stream to speed up event detection and better manage data flows across the organization.
- Real-time transformation.
- Real-time data quality monitoring. No ELT/ETL.

Virtualised data extracts – avoid dangerous data breaches (stop .xls and .csv file proliferation)

With Zetaris, rogue extract files that cause data breaches are no longer in existence. We swap the need for .csv, .xls, and other extract files being distributed to users by providing them with views of the data within the Networked Data Platform.

Streaming Aggregation Tables

Data in the stream can be instantly aggregated into dynamic views and combined with historical data in many different databases, files, etc. to form a data mesh to support an analytical view for your tools. This means your BI dashboards, currently looking at historical data, can now be "always real time," instantly updating as things change.

Key Benefits

2. No need for data extracts or data marts – stop shadow IT

Customers expect you to know everything now: our boards want the reports to be up-to- date, and the regulator demands a single view of compliance and risk at the highest level of data quality. This expectation for high speed, complete and accurate decision support data on all, not a subset of the data, has resulted in most IT departments being unable to keep up with the demands of the business. More significantly, this has accelerated the growth of shadow IT, whereby the business sidesteps the corporate IT department and builds their own solution.

So what happens is that data extracts in many different formats, such as .csv or .xml etc., are all over the place with no consistent governance and oversight. This uncontrolled and poorly managed scattered data landscape, typical of most businesses, is the cause of poor data quality and bad decision-making.

With shadow IT, data is hidden (under desks), privacy is a risk and governance is non-existent. Zetaris dissolves the need for shadow IT by enabling users to build their unique view of data for their applications in a simple virtualised data access system. Users build a data mesh (virtual views) across data in multiple systems and then attach their favourite tools. IT oversees the user through a transparent data governance application within Zetaris: The Networked Data Platform.

3. No-ETL Aggregate Automation

Traditional ETL (extract, transform, and load) processes are manual and human-intensive tasks for wrangling data from its original form to the target form needed for any specific application. Zetaris eliminates the need for ETL/ELT by automating the connect, translate, and view process with its Data Mesh approach. Users build business logic views (what we call a Data Mesh) across the data landscape that Zetaris connects with to make these views available in real time to BI and analytical applications. By doing this, the job of creating data pipelines or ETL scripting for each individual application is no longer needed.

Our No-ETL Aggregate Automation means once a view (Data Mesh) is created in Zetaris, it is available for all applications that are connected to the Mesh. Importantly, the Data Mesh is optimised automatically meaning that it learns over time what the best way to process the view is. So, if a Mesh is connected to three systems, and those systems can help with the load of a user query, Zetaris will run a fully virtualised query. If one of the three systems cannot handle a sub-query, Zetaris will automatically generate an aggregated in-memory temporary table to handle the workload. This intelligent workload and aggregation capability ensures maximum query performance that improves over time.

Key Benefits

4. User behaviour-based query optimisation

Our learning algorithm identifies common usage patterns and join requirements. A granular user profile is developed over time with inputs into query optimisation, cache optimisation, and user behaviour security optimisation.

5. Heterogeneous Query Optimiser

Zetaris optimises every query at run time. Optimisation involves dynamic analysis of source data platform performance characteristics, network characteristics, availability and performance of "in-line" infrastructure and memory to disk management. Zetaris receives user queries and re-codes for best performance and employs other important data management policies, including data governance and quality exception handling. As far as the user is concerned, all optimisation processing is transparent to their query code (ANSI-SQL). Users write simple SQL using the Analytical Data Mesh within Zetaris (or using their preferred query tool) and Zetaris takes care of the rest.

6. Intelligent Semantic Engine

When you deploy Zetaris: The Networked Data Platform, the first step is to ingest all the metadata that describes your various data sources (transaction systems, databases, data marts, cloud databases, and files, etc.). Once the raw schemas or metadata are in our Zetaris Schema Store, the Zetaris software allows a unified business view to be created to support the various uses cases you have. In Zetaris, views of data that joins across multiple databases and locations is called a Data Mesh. A Data Mesh is the business logic structures that point to underlying raw data stores, and the Mesh is connected to your front-end or end-point applications just like any other database. The difference is that a Data Mesh is virtual: it's just the metadata and the business logic.

7. Infrastructure Independence

Zetaris separates the semantic layer and data structuring intelligence from the underlying physical data platforms, enabling BI logic portability. Using our real-time data integration and Data Mesh approach, moving to the cloud, changing clouds or introducing new core platform technology is simple because the business logic is stored and managed outside the core data platforms.

Common Use Cases

1. Share data without giving it away

Zetaris leaves the data where it is. There is no need for data contributors to give data away to share insight. By using the Zetaris Intelligent Push, a virtualised sharing client that is persisted within the contributor's firewall, data contributors can enable external parties to query data according to business, security and access rules. So, using Zetaris, organisations can share insight across data ecosystems with no physical data having to move.

2. BI tool data expansion

Zetaris connects to the data using common standards such as ODBC, JDBC, XML, API, supporting all the common BI and Analytics tools. Zetaris deploys automatic connection management to original data sources and creates a cross-domain view. No need to move, ingest or duplicate data.

3. Create virtual analytical views or cubes

A Data Mesh mixes data in the stream with data in legacy data marts or source databases for real-time, in-context virtual analytical views. A Data Mesh allows tool connections to enable data access across the client's data ecosystem.

4. Virtual Data Warehousing capabilities (VDW)

Zetaris: The Networked Data Platform is used to create virtual data warehouses through implementing contextual, accuracy, and consistency rules over data in AWS, Azure, Google, and other cloud vendor user accounts. Databases or files held across clouds and in the data centre can be modelled into a virtual data warehouse application for consumption by the usual tools (BI, analytical, and data management tools). Zetaris partners with major cloud vendors to provide virtual data warehouse as a service within the cloud provider's core service offering.

The Zetaris Data Fabric builder lets you develop the underlying data structures and pipeline. Users can point and click through our interface to develop a data model (data schema) and immediately query it for super fast productionisation.

Zetaris' Data Quality Engine has been designed to integrate with existing enterprise technologies and platforms. With extreme scale and performance, large-scale processing of both structured and unstructured data, with traceable data lineage and auditability, Zetaris has become the leading data quality solution in organisations requiring tight data governance to meet regulatory requirements.

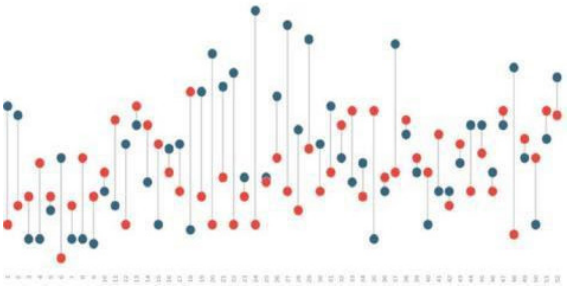
5. Traditional BI to Cloud integration

VDW enables analytical relationships between data and the tools that consume data, allowing multiple data sources in different places to be accessed and processed securely.

Common Use Cases

6. Event detection and business logic on data streams

Complex streaming analytics using Zetaris Stream Views enable the user to create aggregated views (filters/event detectives) using simple SQL commands. This a massive differentiator when handling complex and high volume streams such as JSON streams over Kafka, etc.



7. AI for development automation

- ✓ Full virtual data warehouse build and deployment automation through visual client.
- ✓ Simple SQL coding for complex data view building and data science development.
- ✓ Full analytics notebook integration using Jupyter & Zeppelin.

8. Intelligent query pipeline and caching

Zetaris deploys intelligent query pipelining and caching to enable high performance analytics and real-time data pipelining.

For further information visit:

zetaris.com