CONSONO

WHITEPAPER

Data engines are the same all over the world...

EXCEPT ONE

Content

The problem with data...

We created more data today than we did yesterday. Tomorrow we will create more than we did today. And as it keeps coming, we are struggling to find new ways of managing it.

4 TO 1

Most data projects struggle because the effort that goes into processing the data vastly outweighs the reward - by some estimates as much as 4 to 1

Data hits us from all angles at high speed and it takes many forms. We have more specific applications creating more specific data types which are stored in more and more different ways. This makes it nearly impossible to use the data for any universally useful purpose. And, as data has moved beyond the traditional 'filing cabinet' approach of structured rows and columns, we have piled one new solution on top of another. It's a confusing landscape that keeps expanding:

- Data types: Structured, Unstructured and Semi-structured
- Approaches: ACID, BASE
- Databases: Relational, Document, Object, Graph, etc.
- Storage regimes: Data Warehouse, Data Lake, Data Hub, Data Federation, etc.
- Querying: SQL, NoSQL
- Data Management: Extraction, Integration, Mapping, Transformation, Virtualization, Quality, Master Data Management, Catalogue, etc.
- Progression: Internet of Things, AI, Machine Learning, etc.

Many of the components in this landscape have been driven by the realisation that most data is not structured at all. Even unstructured data will contain structure, if only we can find it. It is contained in the body of an email or in a document or a pdf - even the words in the cells of the rows and columns of structured data formats. And usually structured and unstructured data can't be queried in the same way.

To make use of data we need to do more than just recognise the separate elements. We need to know the semantics of the data; what those elements mean in the context in which they are used. Because without that information, we can't determine the structure and we can't connect one element to another. Recognising the context, we get a more complete picture of the data, we see the connections in the data and we take control of the way we use it.

Right now, a failure to do that comes at a high cost in terms of time and hard cash. Most data projects struggle because the effort that goes into processing the data vastly outweighs the reward - by some estimates as much as 4 to 1. It means many organisations and businesses will never achieve the benefit they seek from the investment they make. They are paying a high price to miss out on insight and opportunity.

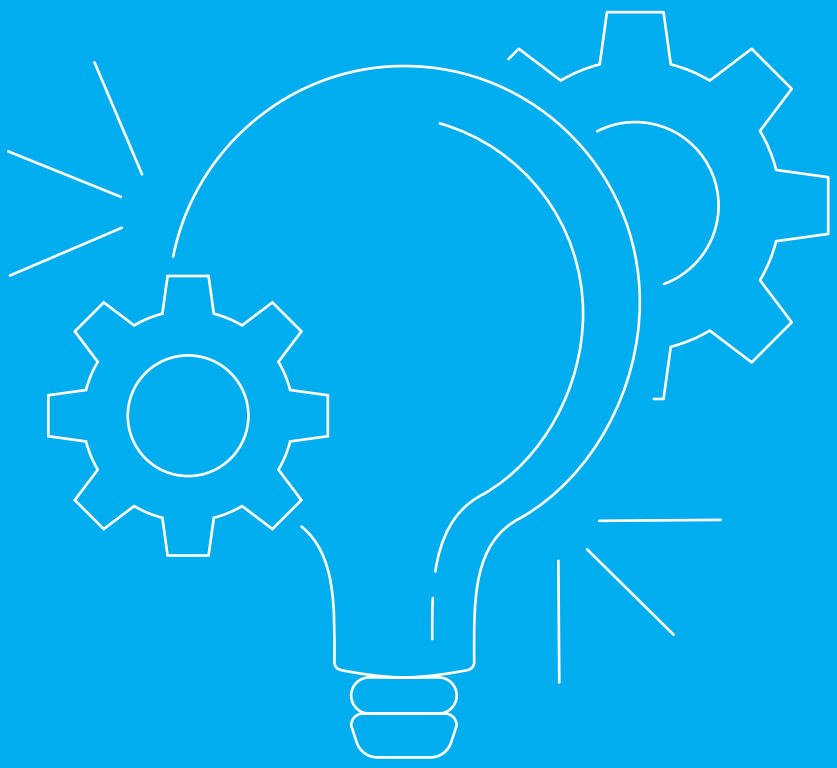
How then to take the best components of the data landscape and combine them into one insightful harmonized environment?

The Dynizer Data Harmonization Engine from Consono converts any data into one universally useful resource, automatically detects direct and indirect links between data elements without querying and stores data in uncomplicated models that won't break even if the nature of the data changes.

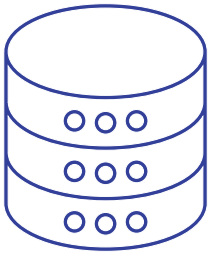
‘Harmonizing data into one universally useful resource’ is undoubtedly a bold claim.

So what differentiates the Dynizer from a Data Lake or a Data Warehouse? Where does it score over a Data Hub or a Data Federation? And is the Dynizer mutually exclusive?

To answer the last question first: In a word, ‘no’. In some cases the Dynizer displays similar characteristics to the others, in some cases it complements their functionality. We would probably say, however, that in most cases the Dynizer would improve the performance of all or any of these other technologies.



POTENTIAL SOLUTIONS

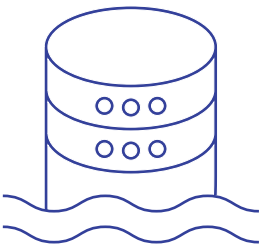


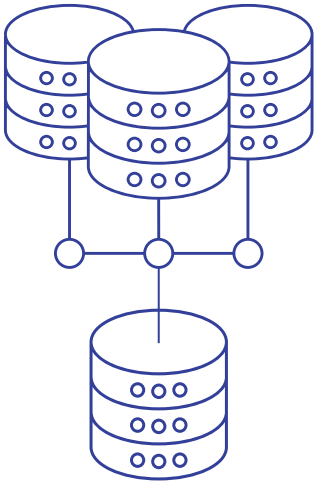
Data Warehouse

A Data Warehouse could be considered to be a specific part of the Data Integration process. It usually works well with data that is entirely structured, but estimates suggest that in reality only 20% of data within a company will be structured. While seen as a good tool to underpin reporting, a Data Warehouse almost never gives access to real time or near-real time data. A lot of data transformation is needed to make data useful. And, if reporting requirements change, if the insights needed change, then the structure of the warehouse has to change, too.

Data Lake

A Data Lake, acknowledges that not all data is structured, and that its usefulness may not be immediately apparent. It is really a big container with no concern about the structure of the data. At the moment it is stored, nothing is lost and the data remains in place. To be able to use it, however, requires the creation of some structure, which essentially just moves the problem down the process. The Dynizer stores all the data in an accessible format that's report ready. A fact-based layer can be attached to the data with no need to ever actually touch the data itself.



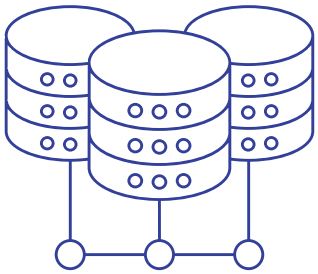


Data Hub

A Data Hub isn't fundamentally different from a Data Lake, except for the fact that many consider a Data Hub to be where data comes into the system, whereas data would go into a Data Lake after it had gone through a Hub. Also with a Hub, data may go in and come straight out again as an operational tool where there are as few transformations made to the data as possible. The Dynizer Data Harmonization Engine could be viewed as a kind of intelligent Hub. Linking the Dynizer to a Hub would make it more flexible.

Data Federation

Data Federation is a relatively old technique to query over a number of data sources. It's a common index on top of non-integrated data sources. These days a Data Federation would be seen as complementary to a Data Hub or a Data Warehouse, but not as a data integration tool in its own right. The Data Harmonization Engine supports part of Data Federation, but only in so far that it takes the data that is considered relevant for further analysis and useful in queries.



A useful starting point in describing of the Dynizer Data Harmonization Engine is to first define what we mean by 'Data Harmonization'.

It is this: Data Harmonization abstracts and contextualises any data from any source. Using simple categories and context definitions it semantically links data in a way that makes it infinitely reusable. The technical characteristics of the data don't define the way it is stored and the model doesn't define the way the data can be used.

Any data element is either a person, a thing, a location, or a moment in time: A Who, a What, a Where, or a When.

The Dynizer is a technology patented in Europe and the USA that combines many of the advantages of other data management solutions into one universally useful environment. The technology is an embodiment of the principles of Data Harmonization.

It virtualises, integrates and manages structured, unstructured and semi-structured data from any source in real time so that it can be used in any application. Data can be reused for many different purposes without needing to build specific models. The Dynizer uses separate microservices to take in, store and query data. Indexing, scaling and de-duplication are all handled automatically.

All data becomes part of the same structure based on the realisation that any data element is either a person, a thing, a location, or a moment in time: A Who, a What, a Where, or a When. Modelled this way, the links between data elements become instantly apparent and can be seen without querying.

Changes to the model have no effect on the effectiveness of the data and will not break the model, a factor that only strengthens the power of discovery within the Dynizer.

Normally, where formal queries are required, it is not possible to ask a simultaneous question of structured and unstructured data in the same operation. The Dynizer's own SQL-compliant query language, DQL, simplifies complex queries, without users even having to know the way the data is modelled.

ARCHITECTURE

Unstructured data like documents, emails, ... get processed and specific Who, What, Where, When extractions are added to the Dynizer.

Unstructured pipeline

Structured pipeline

Different structured date sources like databases or services, can be connected to the structured pipeline via the Structured Import Pipeline.

The Dynizer Core is where the actual data is stored using the Dynizer's simple Action Data model.

dyNizer

On top of this data storage is the Action Index. This is an intelligent index that is able to make all the possible links between all data points contained within.

Queries can be built using The Dynizer's own SQL-compliant query language, DQL.

Dyna Query Language (DQL)

Application Server

A specific REST-interface can be built that allows different applications to use the Dynizer in an agnostic way.

Technically speaking...

Dynizer Core

- micro-services architecture
- written in Go-lang
- distributed roaring bitmaps

Communication Protocols

- GRPC
- REST

Querying

- DQL: Fully SQL compatible

Web based Management Portal

- AngularJS

Deployment

- Kubernetes
- Docker Containers
- Argo Workflows and Pipelines

Unstructured Data Analysis Pipeline

- Python based
- Using Space and Stanford NLP
- Proprietary Algorithms

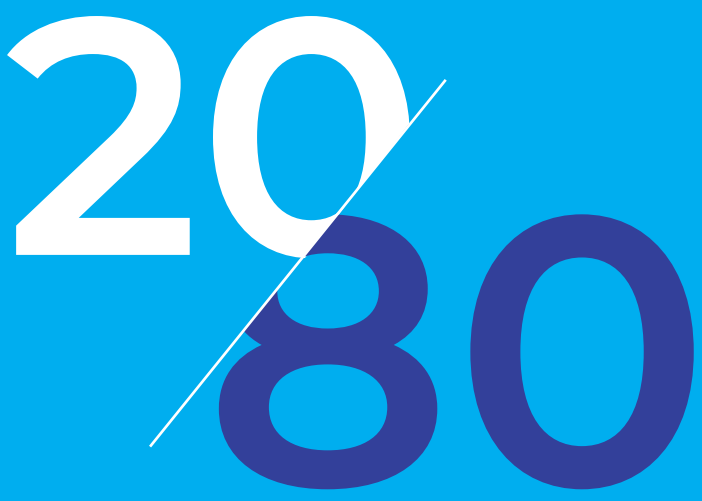
These days we need to be sure that data projects begin to return added value as soon as possible. Yet recent studies show that while we devote 80% of our time preparing data for use, we can only spend 20% of our time creating actionable insights.

The Dynizer gives your team that time back so they can focus on adding value to your business, because it lets you store data in a way that makes sure everything you add is available immediately.

But as well as meeting the challenges of today, the Dynizer is future-ready. You can continue to add any type of data at any time without having to change the way it is modelled.

And while at its core is an advanced set of specialised components, the Dynizer's public interface is a familiar one, based on SQL, which creates a very short implementation timeframe.

Ultimately the Dynizer lets you flip the 80%-20% restrictions of the past in your favour. Reduce the effort it takes to prepare data and devote more time to adding valuable insight to your business.



Want to learn more?

To find out more about how the Dynizer can harmonize your data...

Visit: www.consono.ai | Email: info@consono.ai | Call: +32 3 547 00 40

Consono HQ: Schillerstraat 8, 2050 Antwerpen, Belgium

CONSONO