

DATA FACT



SKANSKA

Introduction Vector Database



Value Quantified in Action

In the end, it's the resulting actions that prove the worth of any analytics engagement.

With our end-to-end expertise, we're with you through that last mile.



The Right Approach

We know Advanced Analytics, and we know how to make it work for you.

Our experience means our clients get successful, scalable transformations, not hype and buzzwords.



Your Secret Data Weapon

Any Analytics initiative is only as successful as the data engineering behind it.

That's why many organizations and technology partners trust us to do their heavy lifting.

Why Vector Database

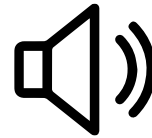
More than 80% of data around us is unstructured



Documents



Images

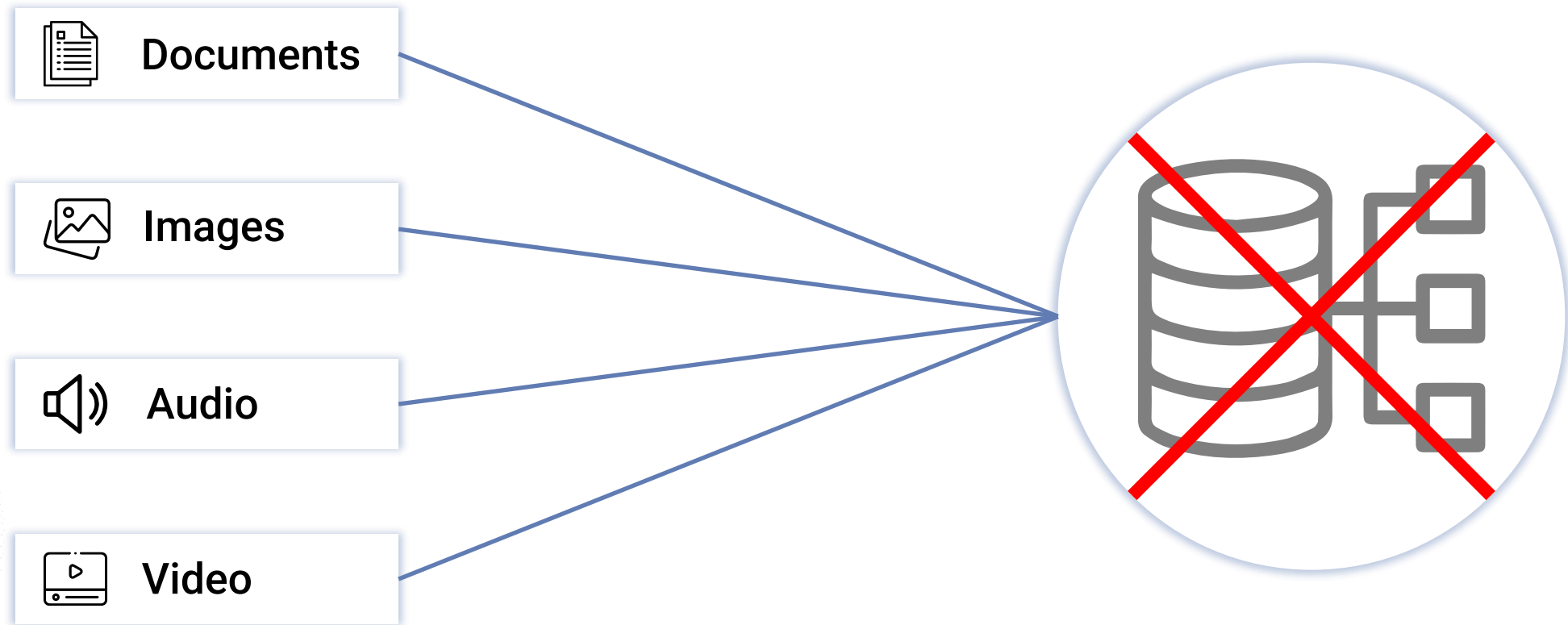


Audio



Video

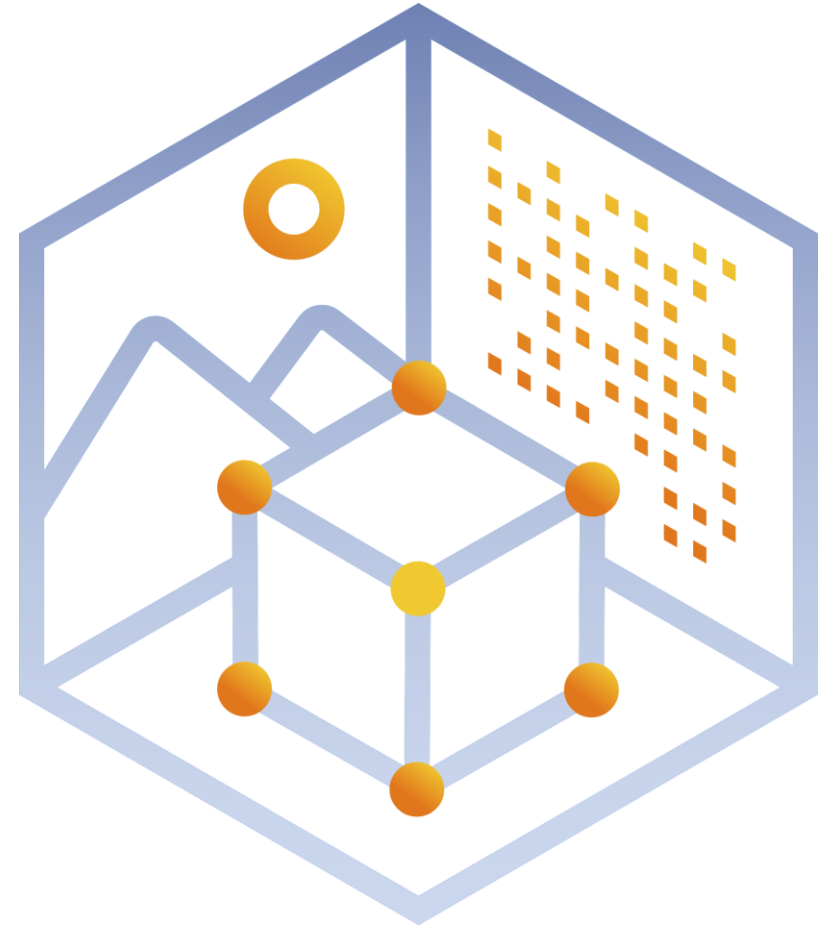
Fitting into traditional Database



Often ends up assigning attributes and tags

What is Vector Database

A vector database **indexes** and stores **vector embeddings** for fast **retrieval** and **similarity search**



Vector Embeddings

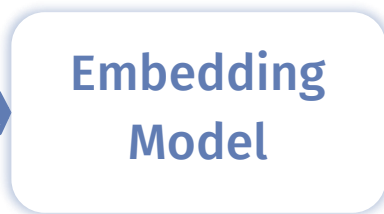
Embeddings measure the relatedness of the data

Set Objects

Object 1

Object 2

Object 3



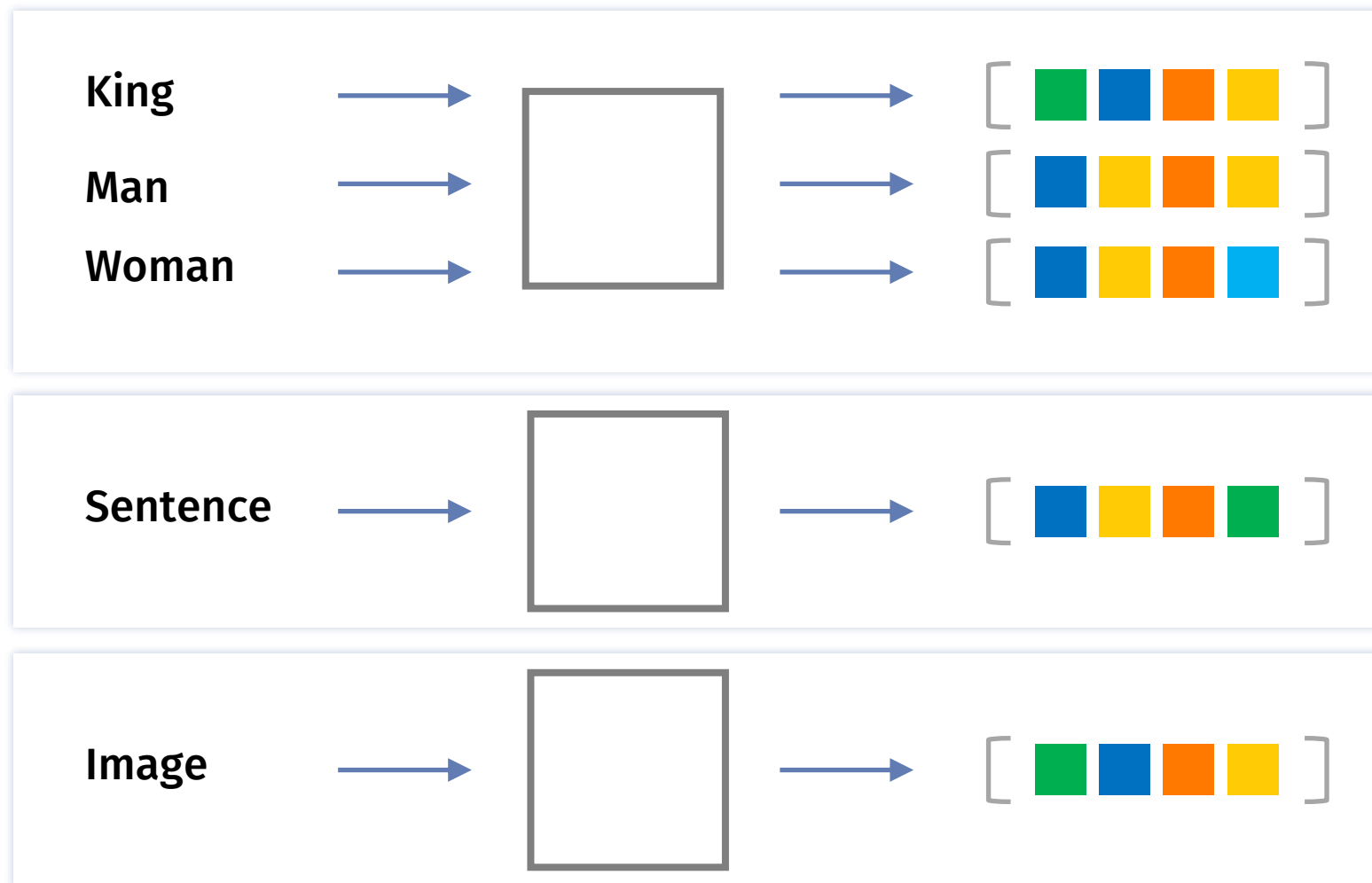
Objects as Vectors

0.6	0.3	0.1	-----
-----	-----	-----	-------

0.8	0.5	0.3	-----
-----	-----	-----	-------

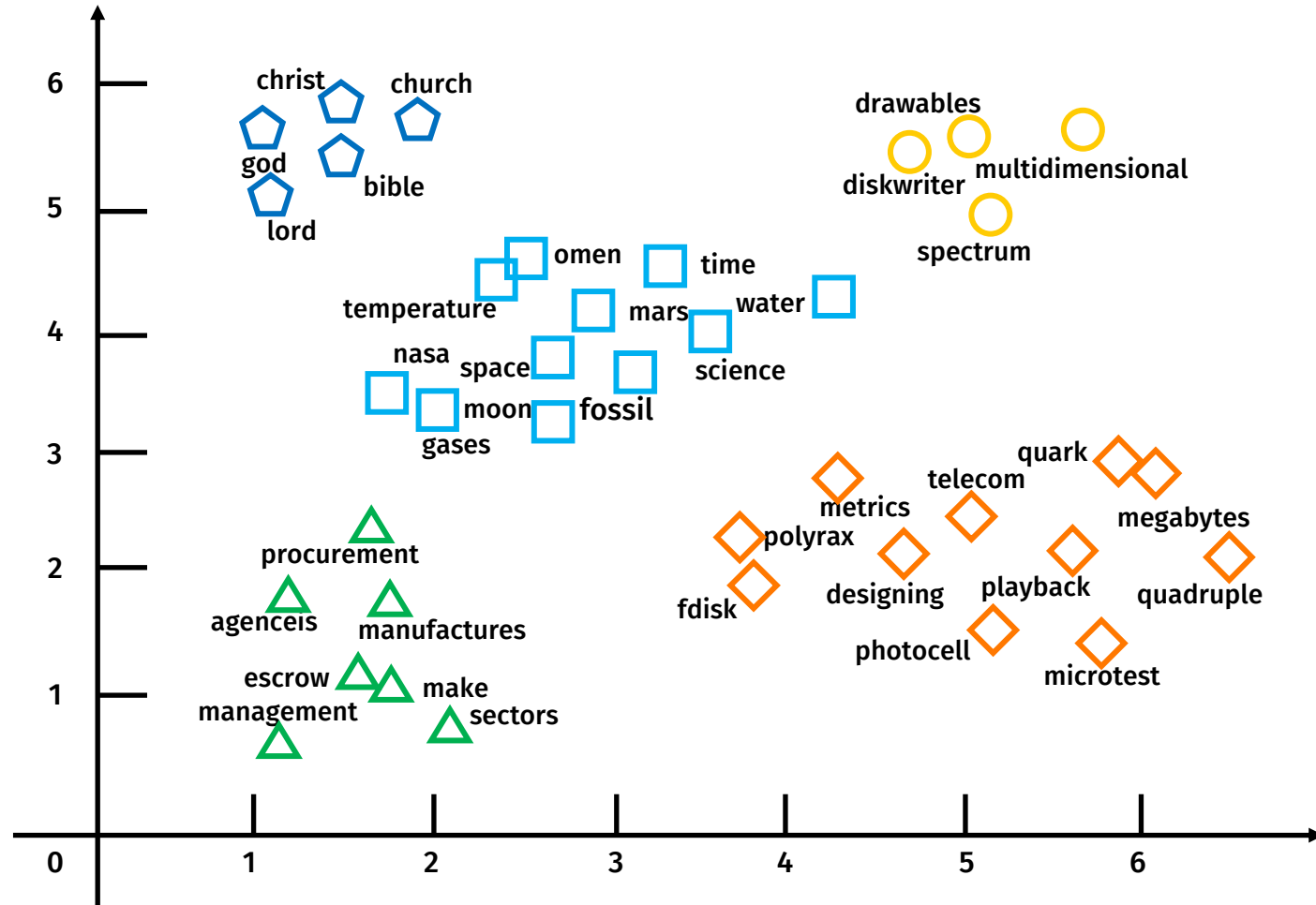
0.4	0.2	0.9	-----
-----	-----	-----	-------

We can embed words, sentences or even images



Create an embedding for “Hello world”

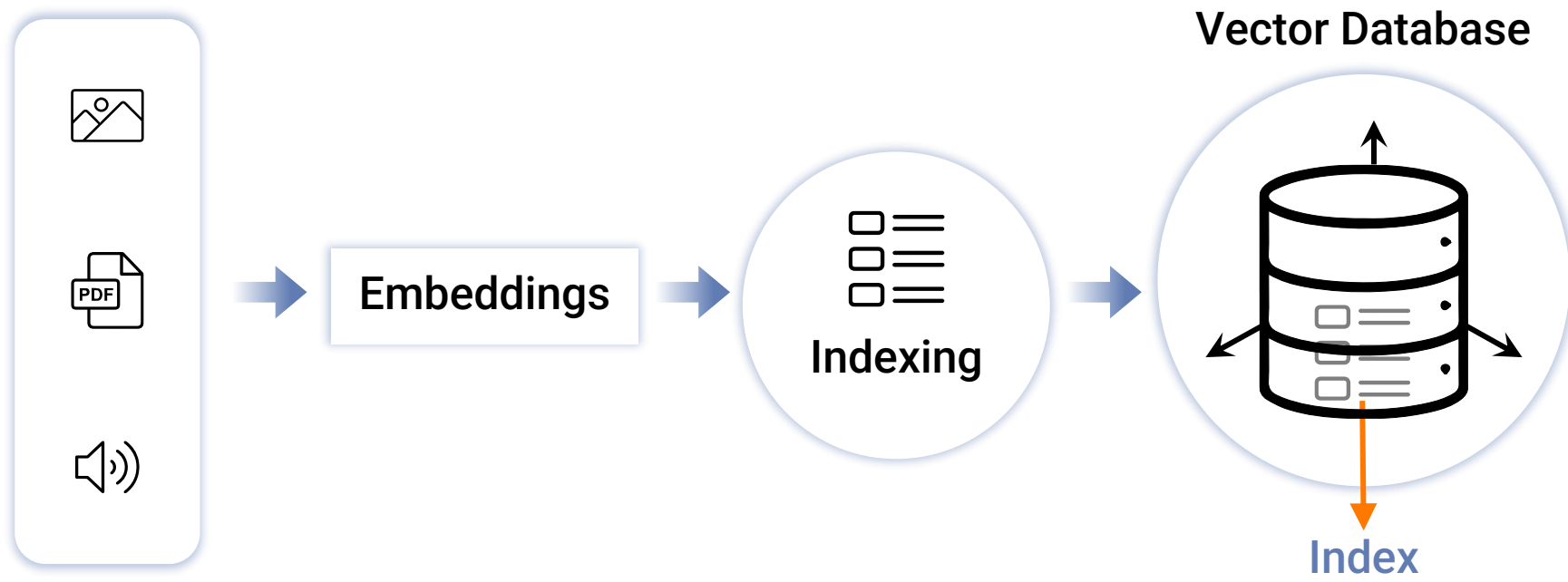
We have a representation that the machines can understand



Indexing

Performing a query across thousands of vectors based on a distance metric, on the fly, would be **extremely slow**

An index is a data structure that facilitates the search process.

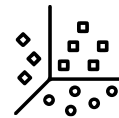


Applications



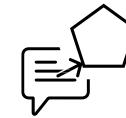
Similarity Search

Results are organized based on their relevance to a specific query.



Clustering

Groups text based on similarity.



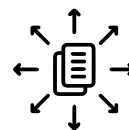
Recommendations

Suggests items that are related through text similarities.



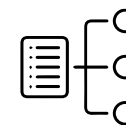
Anomaly Detection

Identifies outliers that have little in common with the main dataset.



Diversity Measurement

Analyzes the distribution of similarities.



Classification

Categorizes text into labels based on similarity.



Thank you!

DATA FACT | 

22260 Haggerty Rd, Northville, MI
1 866 739 0820
hello@datafactz.com