SeMI Technologies

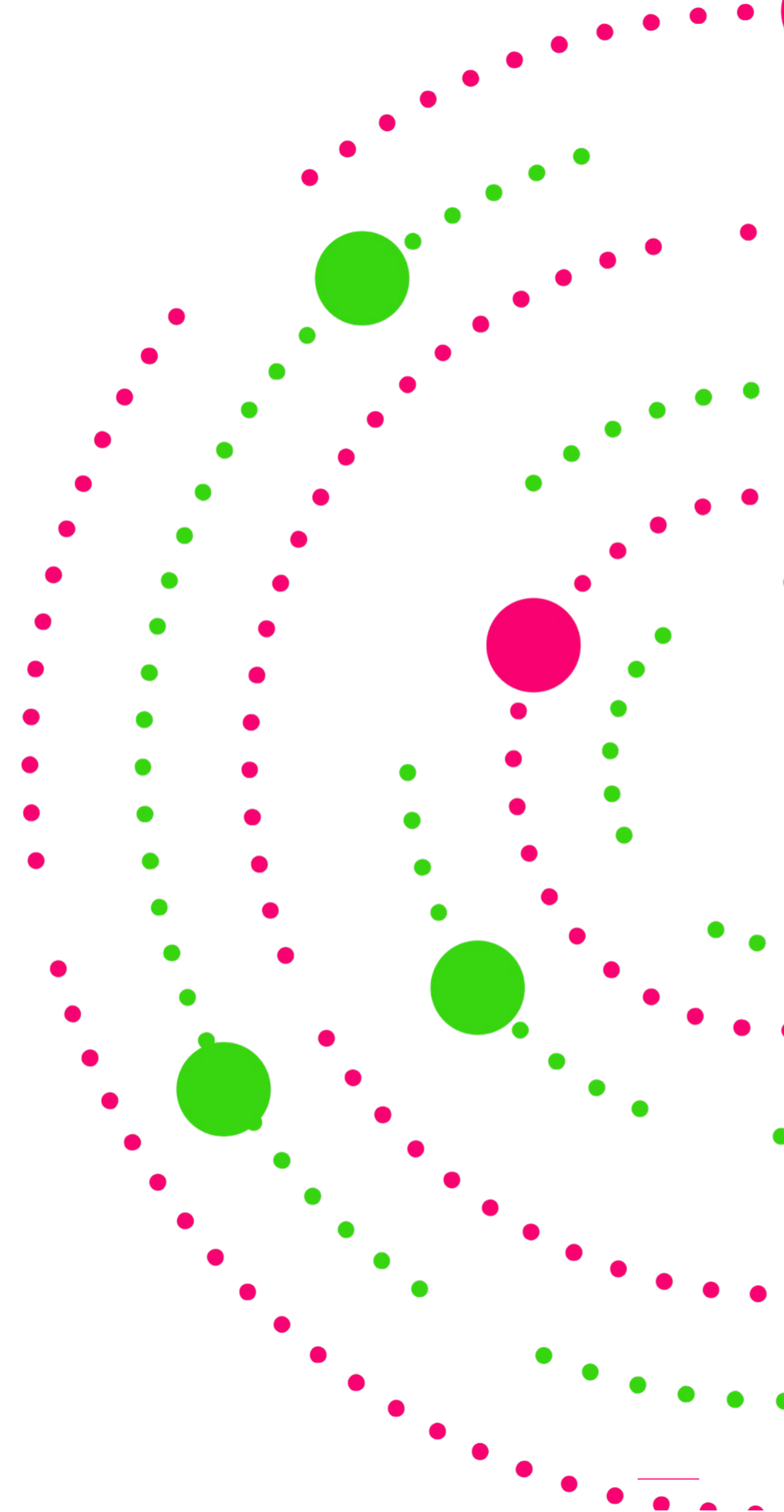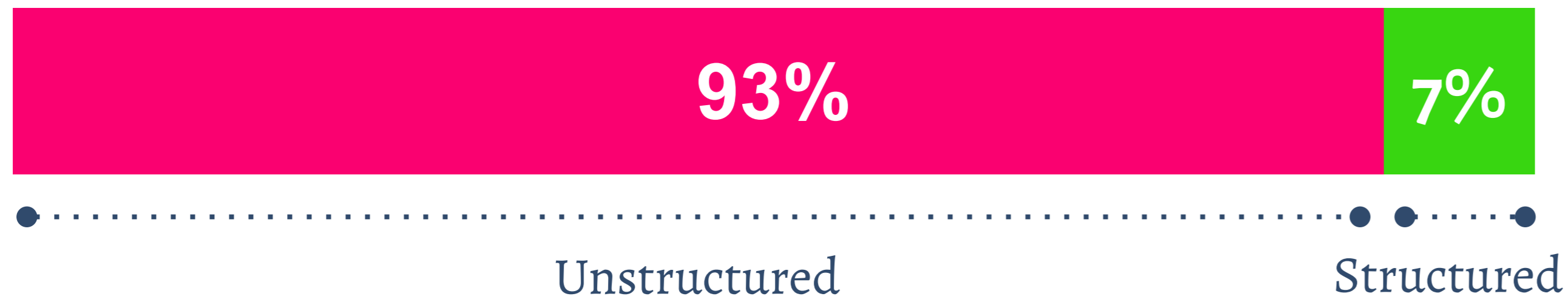# Weaviate
Introduction @Conf42

Learn how Weaviate can help with your unstructured data problems
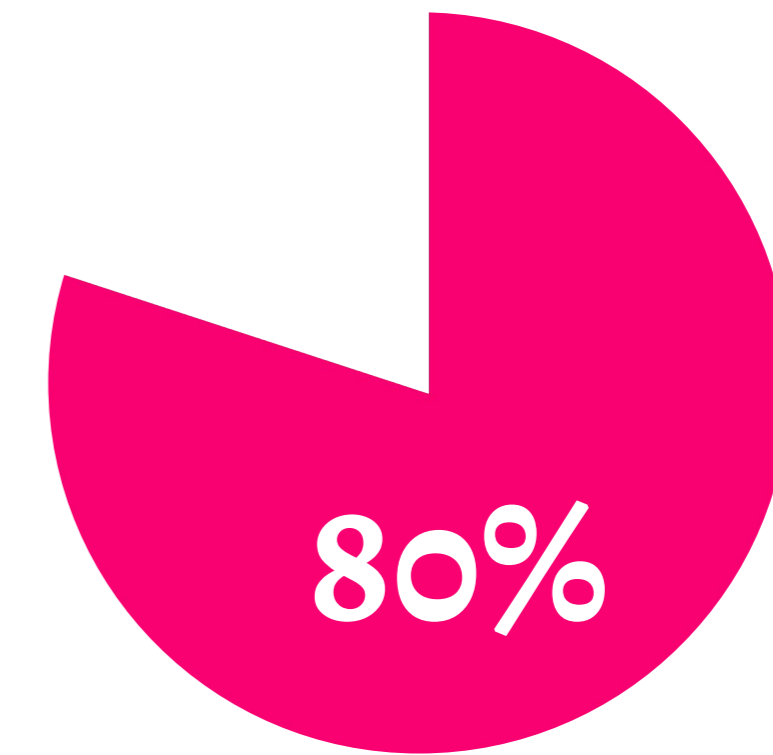
What's so difficult about **unstructured** data?

93% of your data stays unused

Unstructured data makes up about **93% of your data (Ciklum)**.
About **80% of businesses (Forbes)** don't know how to leverage it.

93%	7%

Unstructured   Structured

Business documents, ERP line items, social media posts, legal documents, emails, images, chat messages, publications, invoices help desk tickets, cyber security analysis, webpages, medical records, human input CRM data, etc.

80%

Your data analysts spend 80% of their time (HBR) preparing data to answer business questions.

SeMI

What is so difficult about unstructured data?

```
{ "data": [{
    "Wine": "Covey Run 2005 Chardonnay",
    "Description": "...  good with fish ..."
}]}
```

"Wine for seafood" 🔍

No products found ...

Traditional search engine

"Wine for seafood" 🔍

Covey Run 2005 Chardonnay

Vector search engine

SeMI

4

What is vector search?

Google is a vector search engine too!

Thanks to this, you can use Google Search to find answer in a vast amount of webpages. But did you know that it is estimated that Google only indexes about 0.004% of unstructured data? That's because companies like yours have the rest! What if you could use similar search technology for your data in a reliable and secure way?

The question is very **abstract.**

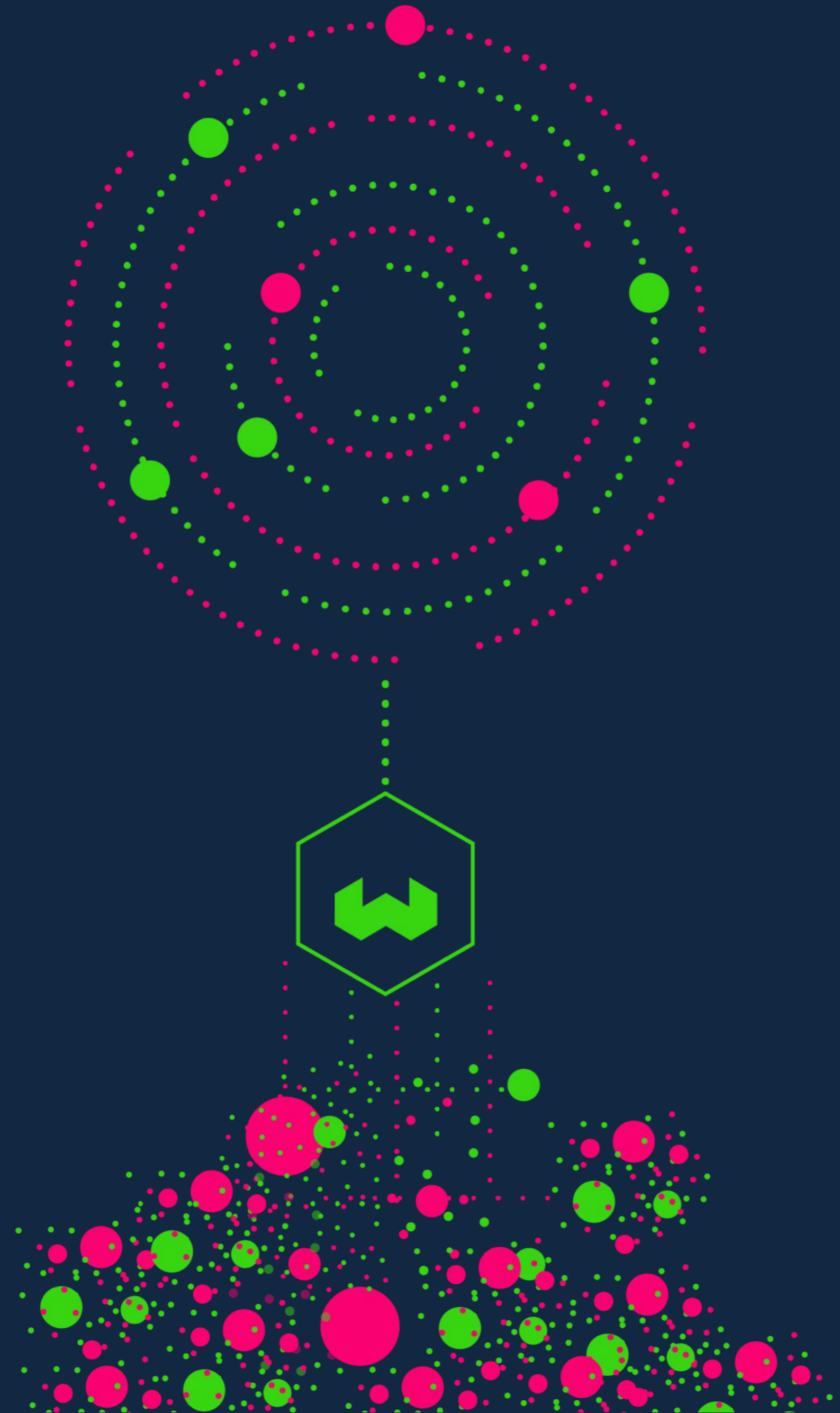And it answers **fast**!

SeMI

Google is good at working with unstructured data on the public web (which is < 0.01% of data available)

What if you could do the same with your data in a simple and secure way?

SeMI

Weaviate puts the 93%

of data to work.

Weaviate is a database based on machine learning, thanks to its easy to use API's that run in the cloud or on premise...
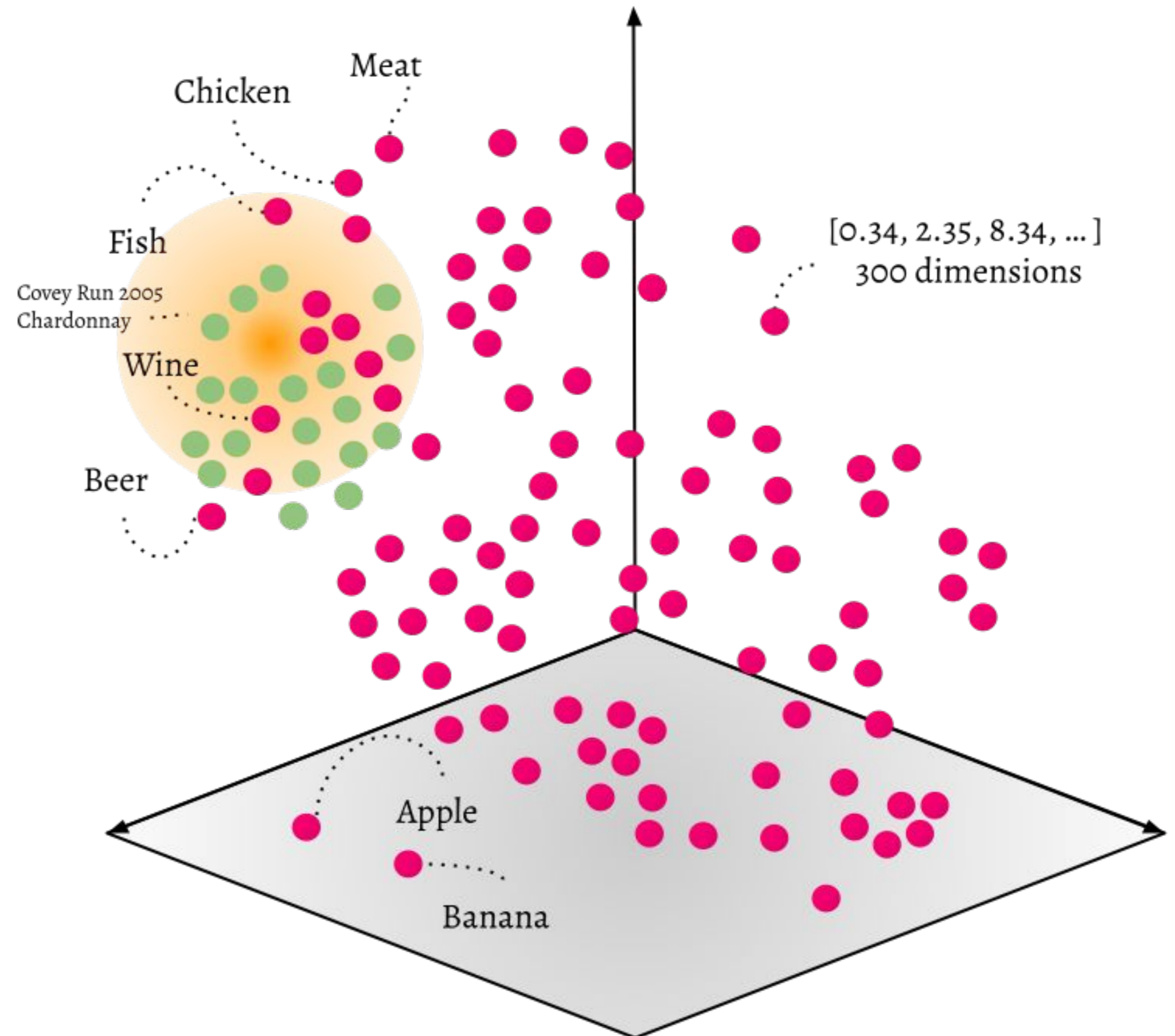
SeMI

Weaviate is a cloud-native, modular, real-time **vector search engine**, built to scale your machine learning models

SeMI

What is vector search?
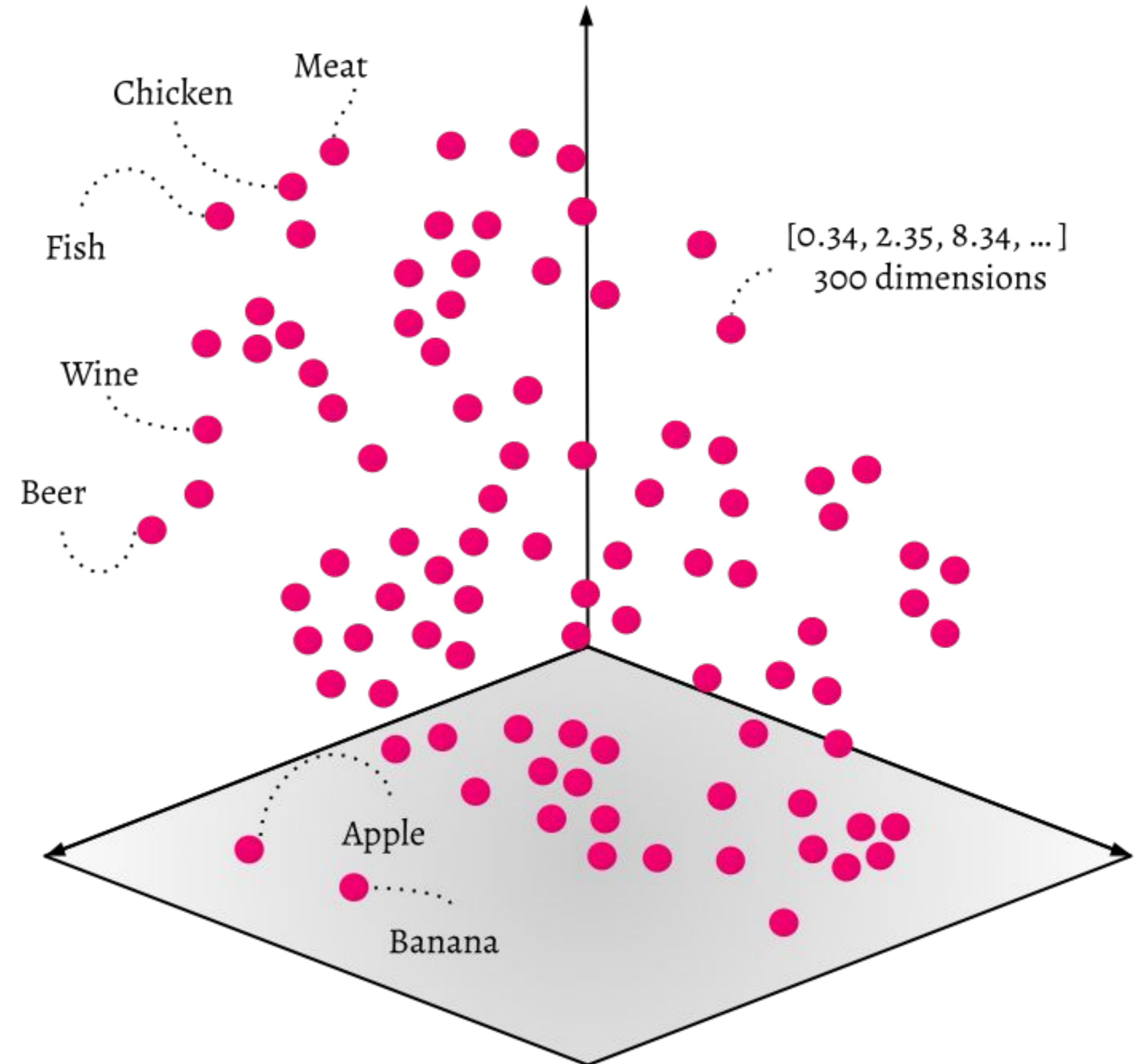
Weaviate is a *vector search engine*

Instead of just storing raw data like traditional databases do, Weaviate leverages the power of machine learning (ML) models to *vectorize* the data. What this means, is that the ML-models try to understand your data while storing it. This allows Weaviate to search, discover and classify similar results in your dataset. Simply put, Weaviate is a database that understands your data.

SeMI

What is vector search?
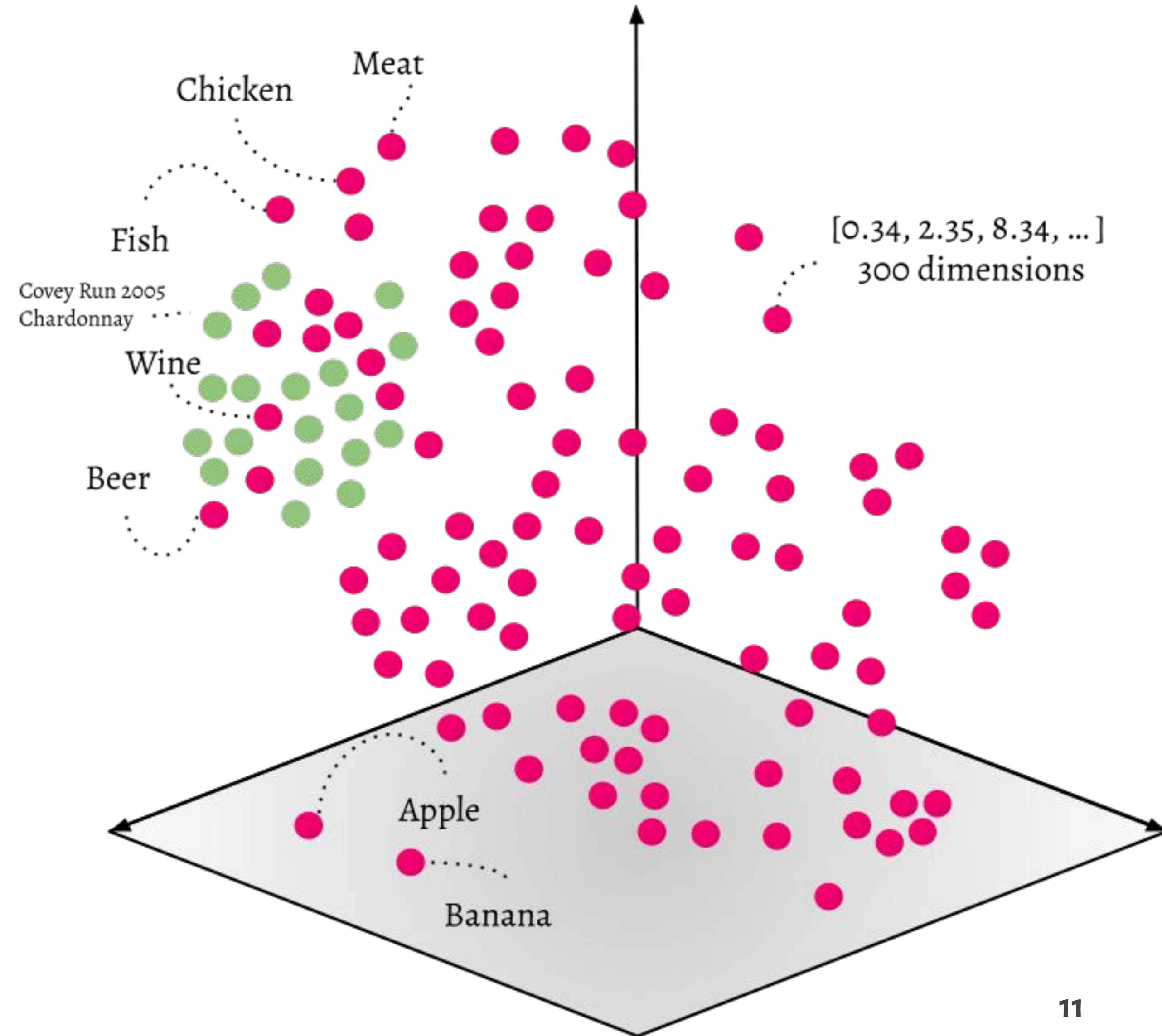
1. Weaviate Machine Learning Modules

- E.g. BERT NLP module
- This language model can represents all words and concepts in a hyperspace.

Chicken Meat

Fish

Wine

Beer

[0.34, 2.35, 8.34, ...]
300 dimensions

Apple

Banana

SeMI

What is vector search?
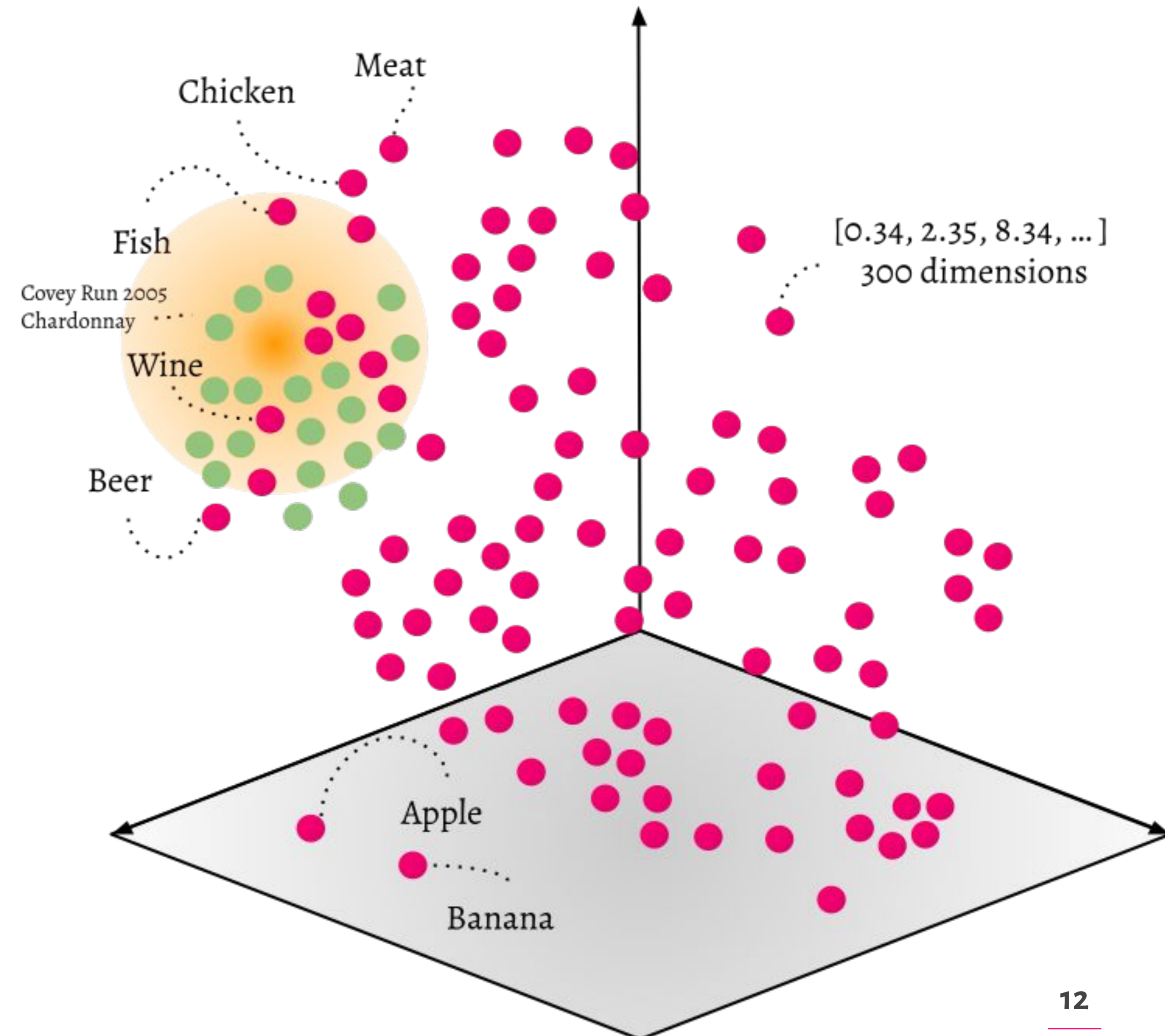
2. Automatically vectorize and index your data

- Weaviate *understands* your data
- When you import data: Weaviate looks at the language in your data object
- E.g. a *Chardonnay* is closely related to *Wine, White* and the *food* it fits with
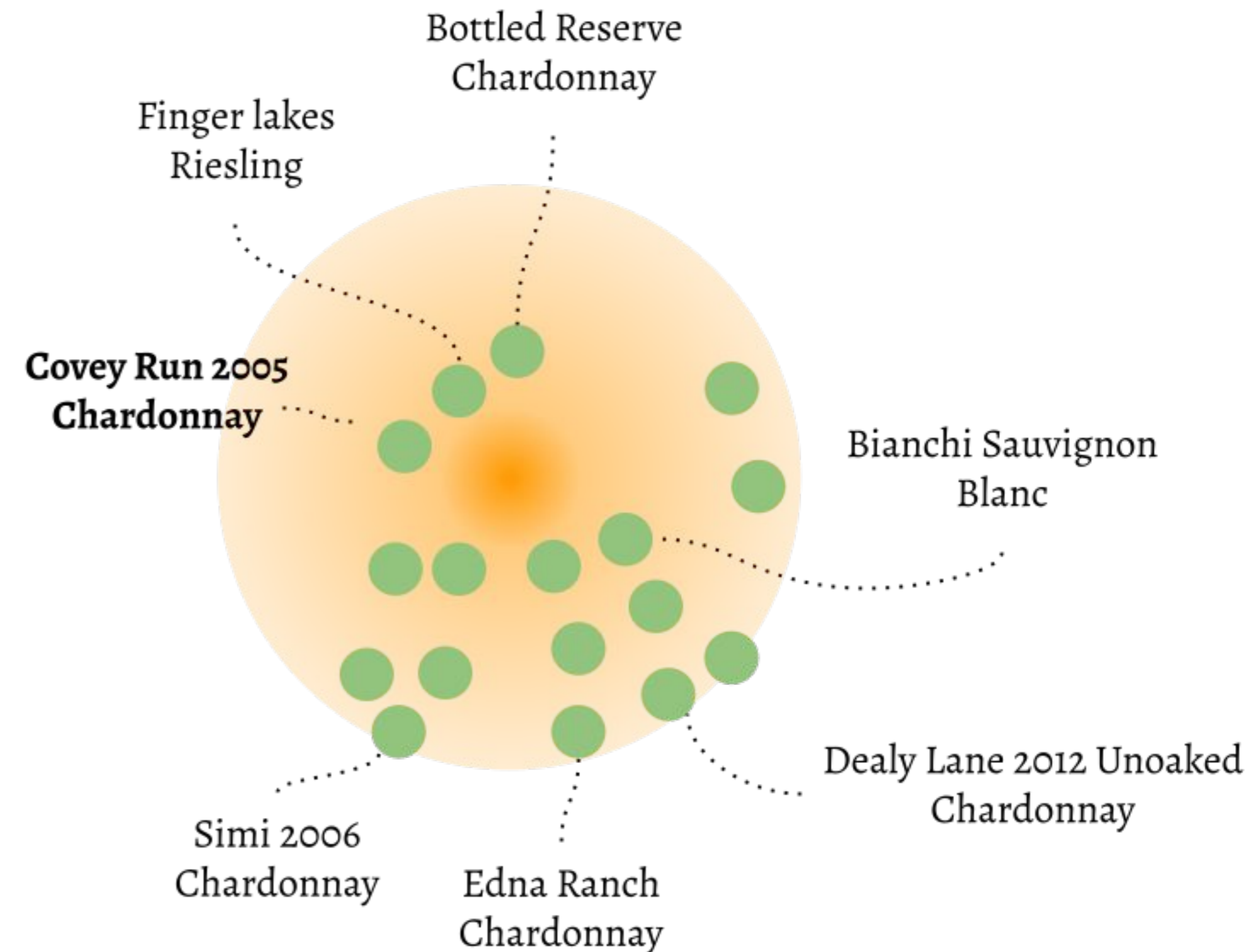
What is vector search?

3. Search query

- Your search queries in natural language will also be vectorized and understood by the machine learning module of Weaviate
- It is places close to the words and data object that are semantically related to the query
- E.g. *"Wine that fits with a seafood dish"*

Chicken
Meat
Fish
Covey Run 2005 Chardonnay
Wine
Beer
Apple
Banana

[0.34, 2.35, 8.34, ...]
300 dimensions

What is vector search?

4. Results

- The data objects that are closest to the search query are retrieved from the dataset

Bottled Reserve
Chardonnay

Finger lakes
Riesling

**Covey Run 2005
Chardonnay**

Bianchi Sauvignon
Blanc

Dealy Lane 2012 Unoaked
Chardonnay

Simi 2006
Chardonnay

Edna Ranch
Chardonnay

SeMI

What is vector search?
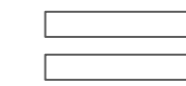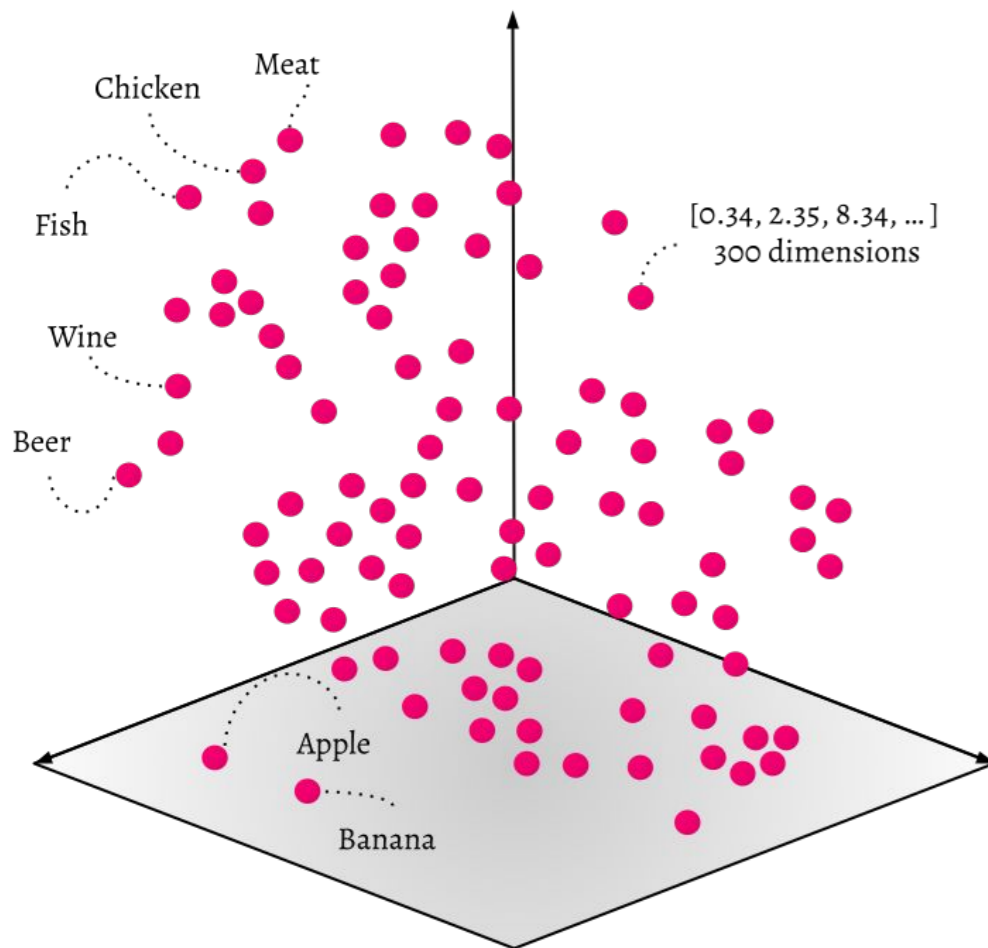
Pretrained ML model + Your own data + Your search query = Results
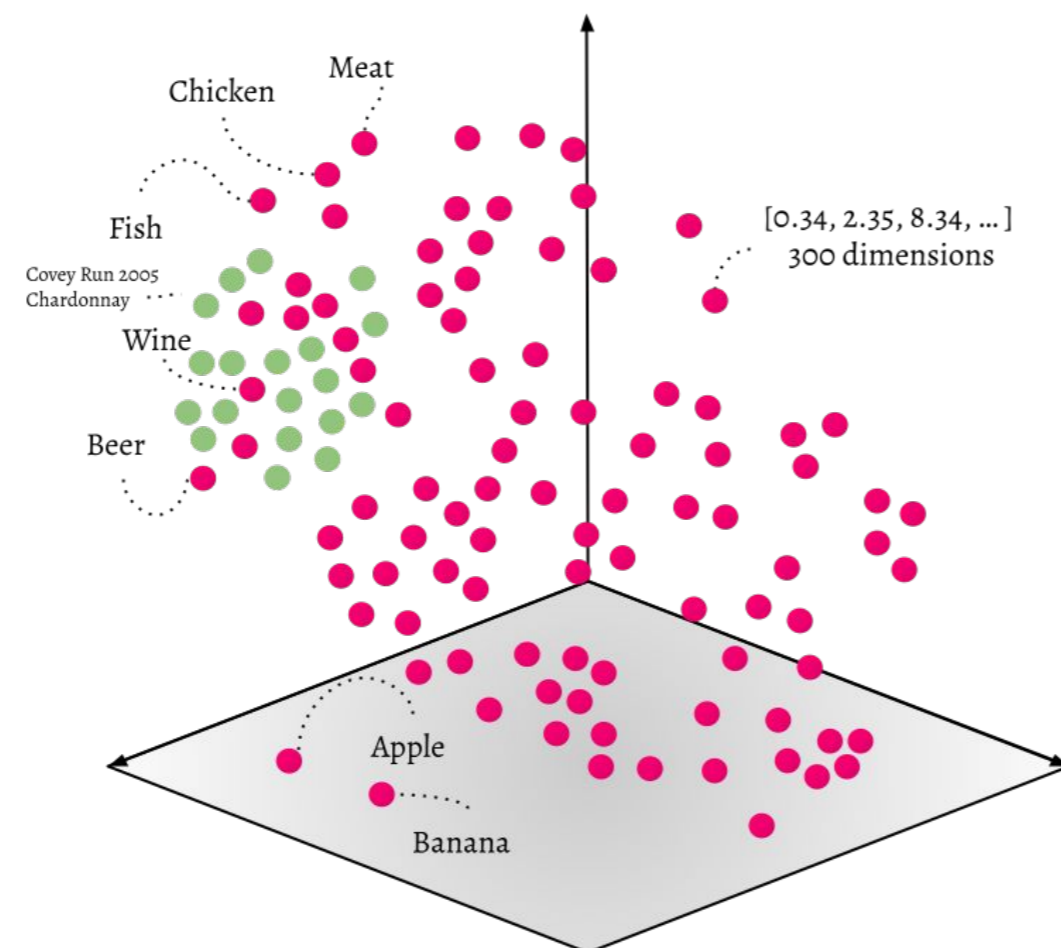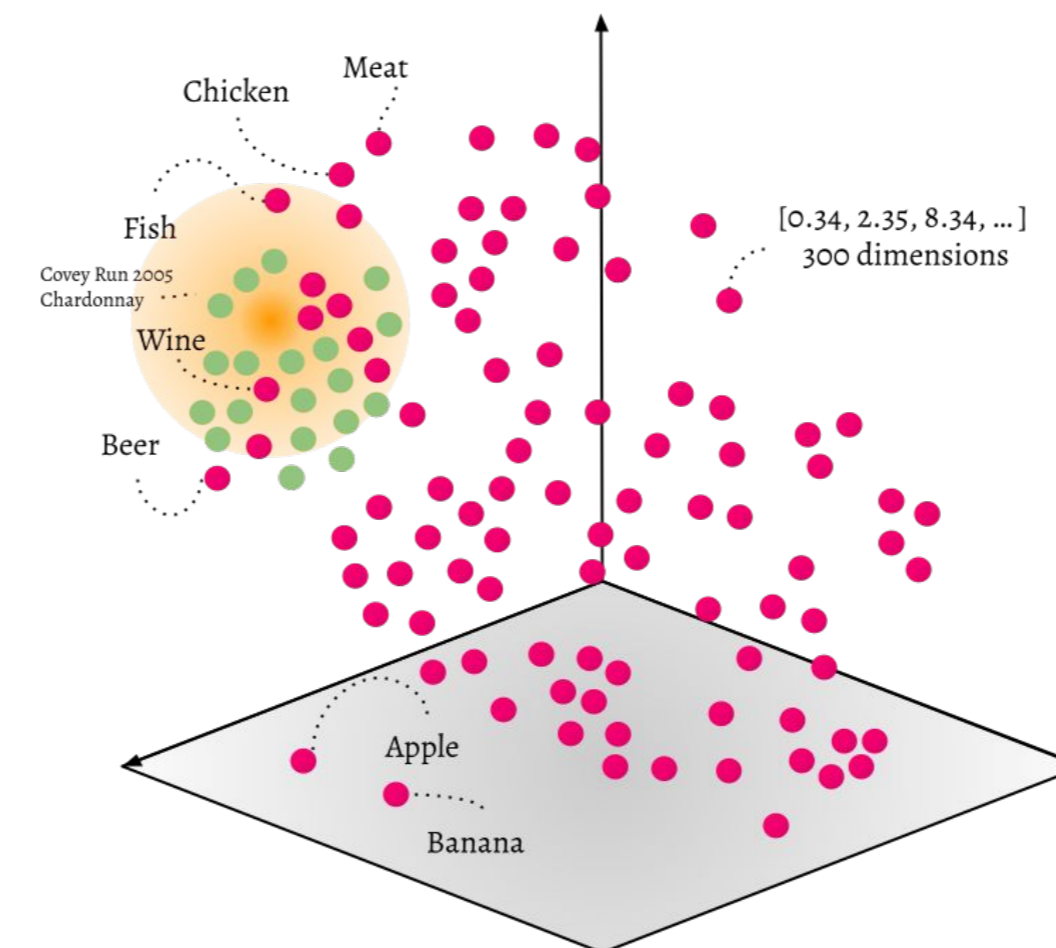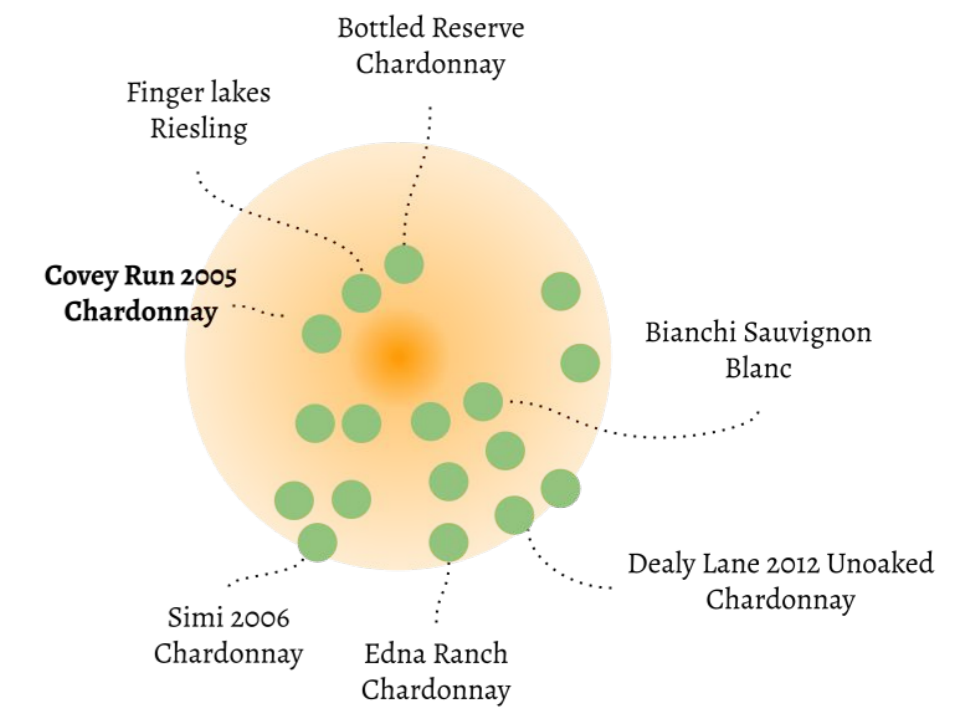

(e.g. NLP)   (e.g. Wines)   (e.g. Wine for seafood)   (e.g. Chardonnay)

It allows you to do the following with any type of unstructured data:

Search

Weaviate is able to search through your data like Google allows you to search over the web

Discover

Weaviate can answer questions based on your data

Classify

Weaviate can classify data from any data source

Predict

Weaviate can recommend relations in your data

SeMI

Weaviate Modules

SEMANTIC SEARCH

QUESTION ANSWERING

TEXT VECTORIZATION

CLASSIFICATION

FRUIT

FILTERING

OBJECT STORAGE

VECTOR SEARCH

IMAGE VECTORIZATION

IMAGE SEARCH

OTHER DATA TYPES

SeMI

Weaviate Demo

SeMI

Weaviate adds value in many industries

You can request an industry specific use-case presentation via hello@semi.technology or www.semi.technology

Retail & CPG

Healthcare & Pharma

Publishing & Media

Technology

Cyber security

Insurance

E-commerce

Helpdesk

SeMI

SeMI Technologies

Thank you

Laura Ham

Email: laura@semi.technology

www.semi.technology