Scalable Vector Search with Weaviate

Laura Ham, Community Solution Engineer at SeMI Technologies

{codemotion}
Online Tech Conference
- Italian edition -

23-24-25 Marzo, 2021

Agenda

- Introduction to Weaviate: Vector Search Engine
- Demo
- Weaviate's core features
- Use cases

Weaviate

The AI-based vector search engine

Weaviate - why a vector search engine?

```
{ "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

The question is very **abstract.**

And how so extremely **fast**?

How did Google find exactly this **data node**?
How can we predict this **relation?**

How to do this on your **own data**?

SeMI

What's the problem?

- There is an estimated 40 trillion gigabytes in data. (Christo Petrov, 25+ Impressive Big Data Statistics for 2020)
- 80% to 90% of data is **unstructured**, that's about **34 trillion gigabytes**! (Christo Petrov, 25+ Impressive Big Data Statistics for 2020)
- Google Search has indexed over 100,000,000 gigabytes, that's only 0.00025%. (Google Search)
- We keep storing more and more data but businesses can't get **valuable insights** from it.
- 95% of businesses have issues **getting value** from their unstructured data. (Christo Petrov, 25+ Impressive Big Data Statistics for 2020)

How to achieve **search** and automatically make **relations** within your **own data**?

In a **easy**, **fast**, **secure** and **scalable** way?

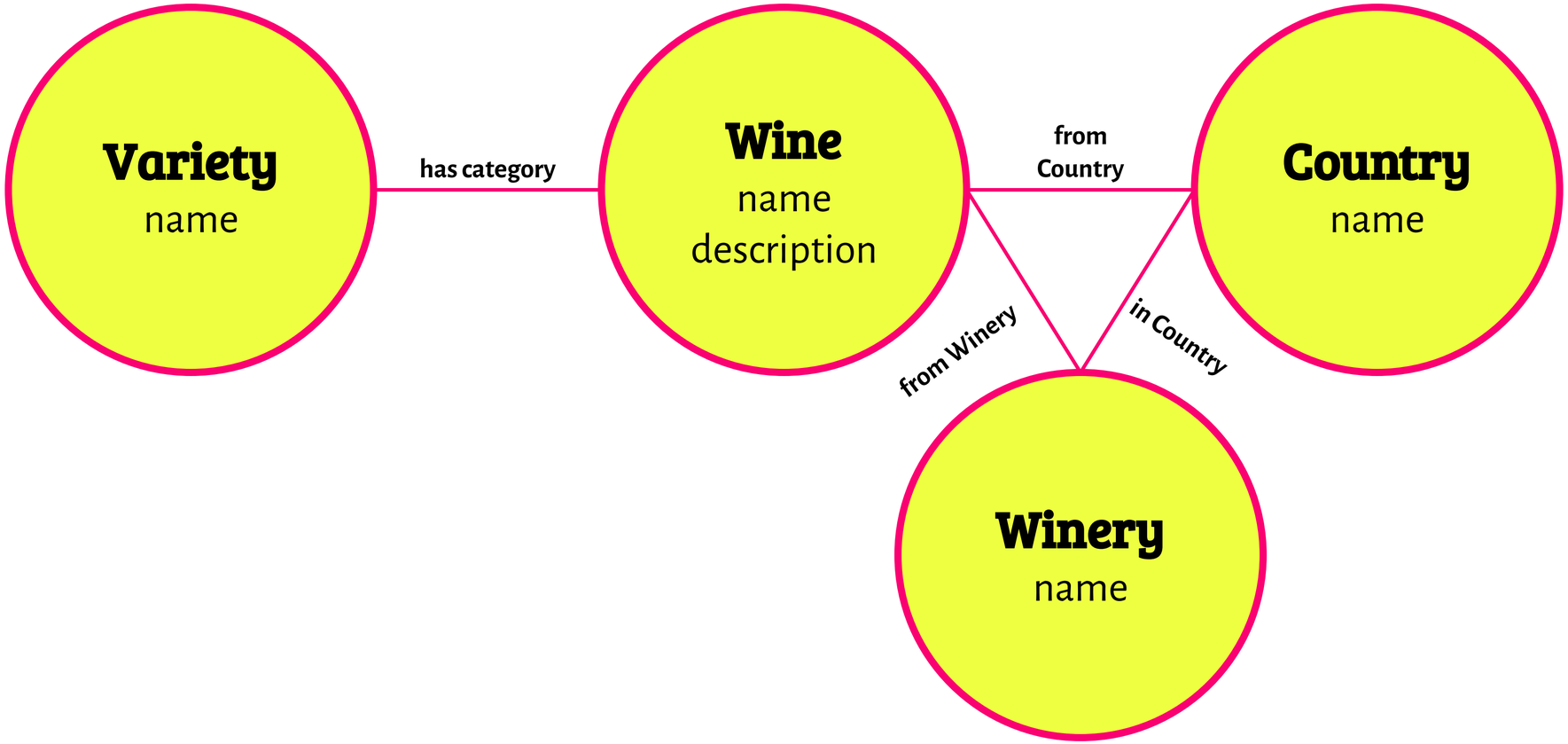
Weaviate is a cloud-native, modular,
real-time vector search engine
built to scale your machine learning models

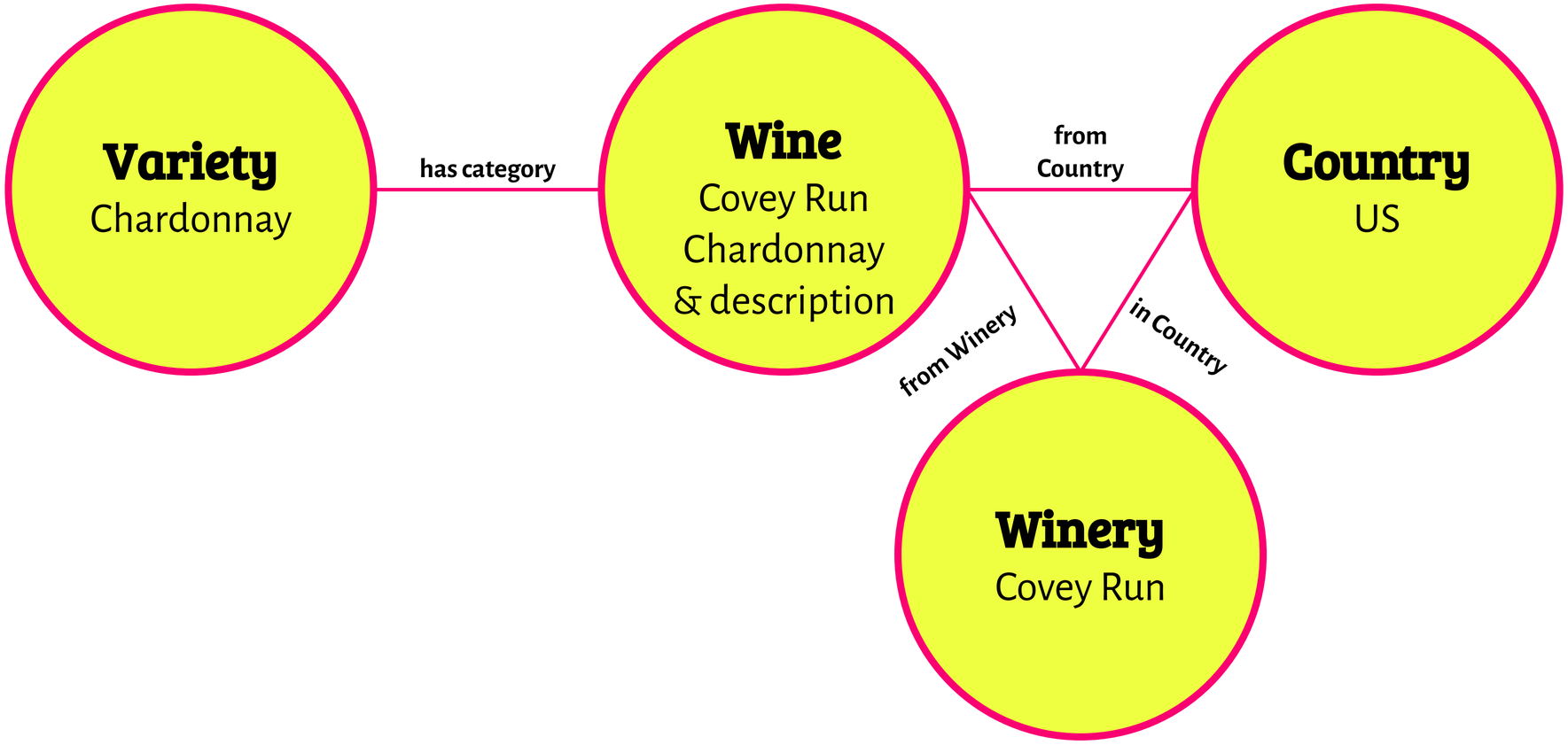
Weaviate is a cloud-native, modular,
real-time vector search engine
built to scale your machine learning models

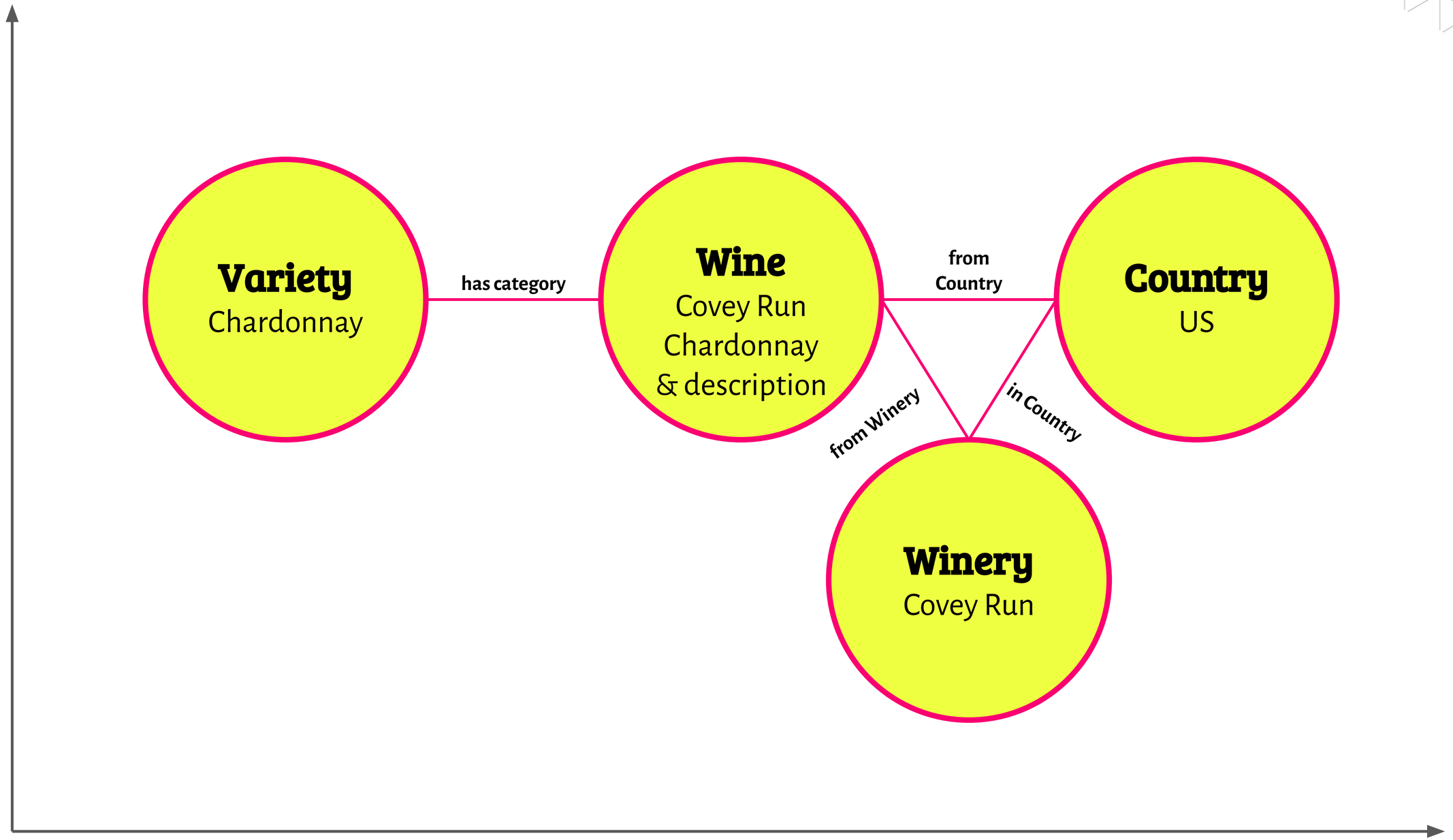
Weaviate is a cloud-native, modular,
real-time vector search engine
built to scale your machine learning models

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

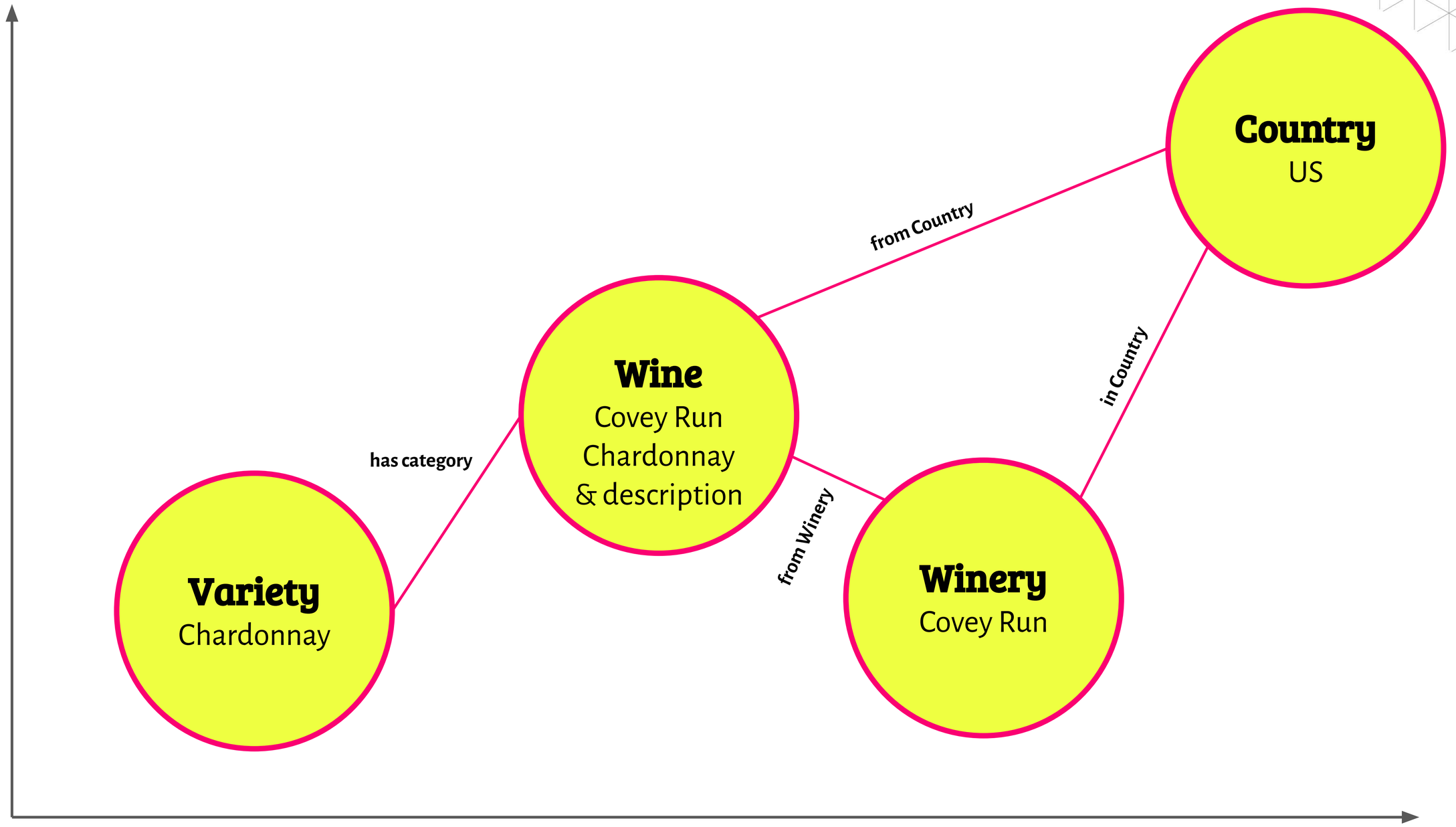
Weaviate is a cloud-native, modular,
real-time vector search engine
built to scale your machine learning models


Variety
name
has category
Wine
name
description
from
Country
Country
name
from Winery
in Country
Winery
name

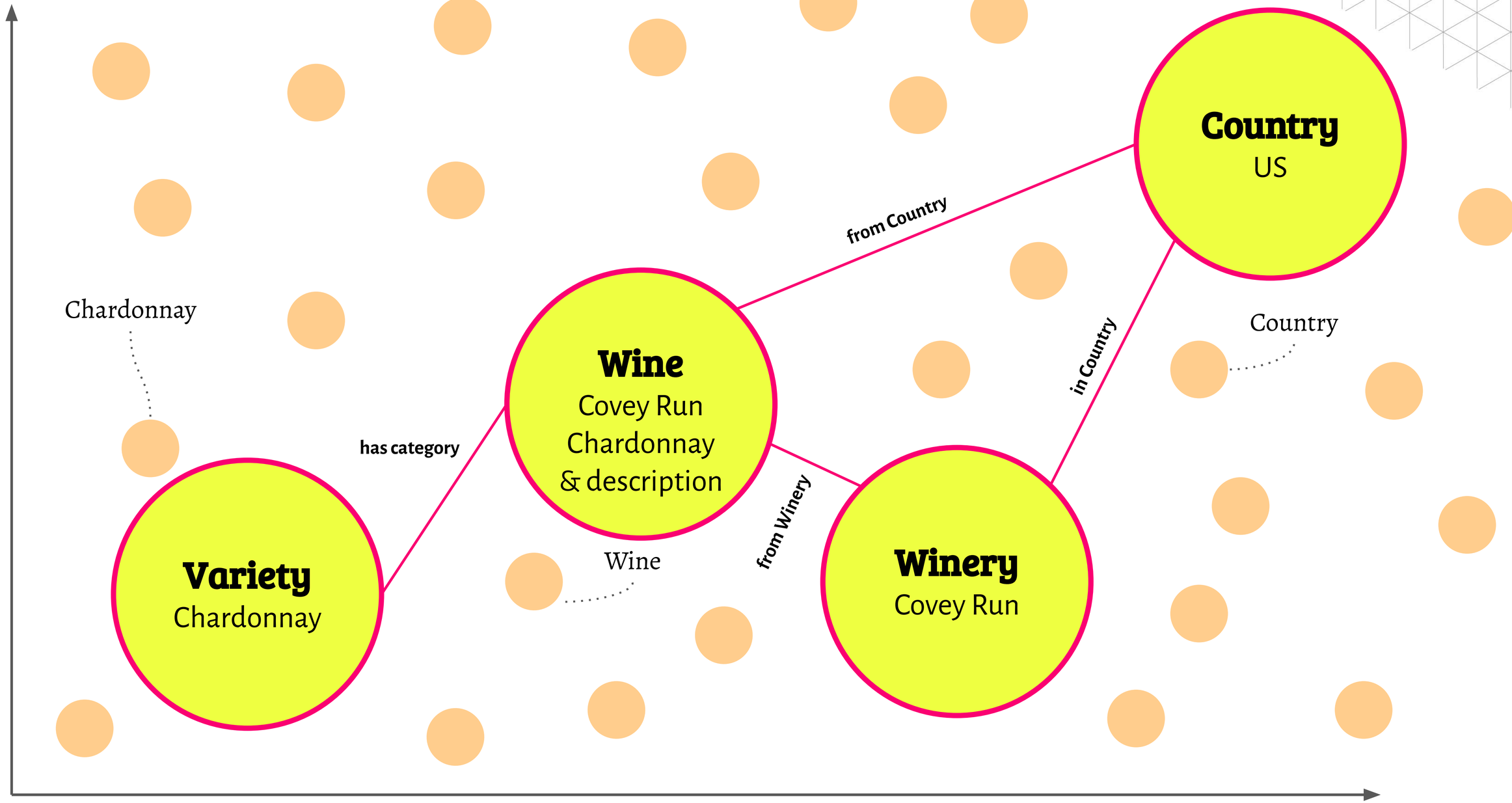

Variety
Chardonnay
has category
Wine
Covey Run
Chardonnay
& description
from
Country
Country
US
from Winery
in Country
Winery
Covey Run


Variety
Chardonnay
has category
Wine
Covey Run
Chardonnay
& description
from
Country
Country
US
from Winery
in Country
Winery
Covey Run

SeMI


Country
US
from Country
Wine
Covey Run
Chardonnay
& description
has category
in Country
from Winery
Variety
Chardonnay
Winery
Covey Run

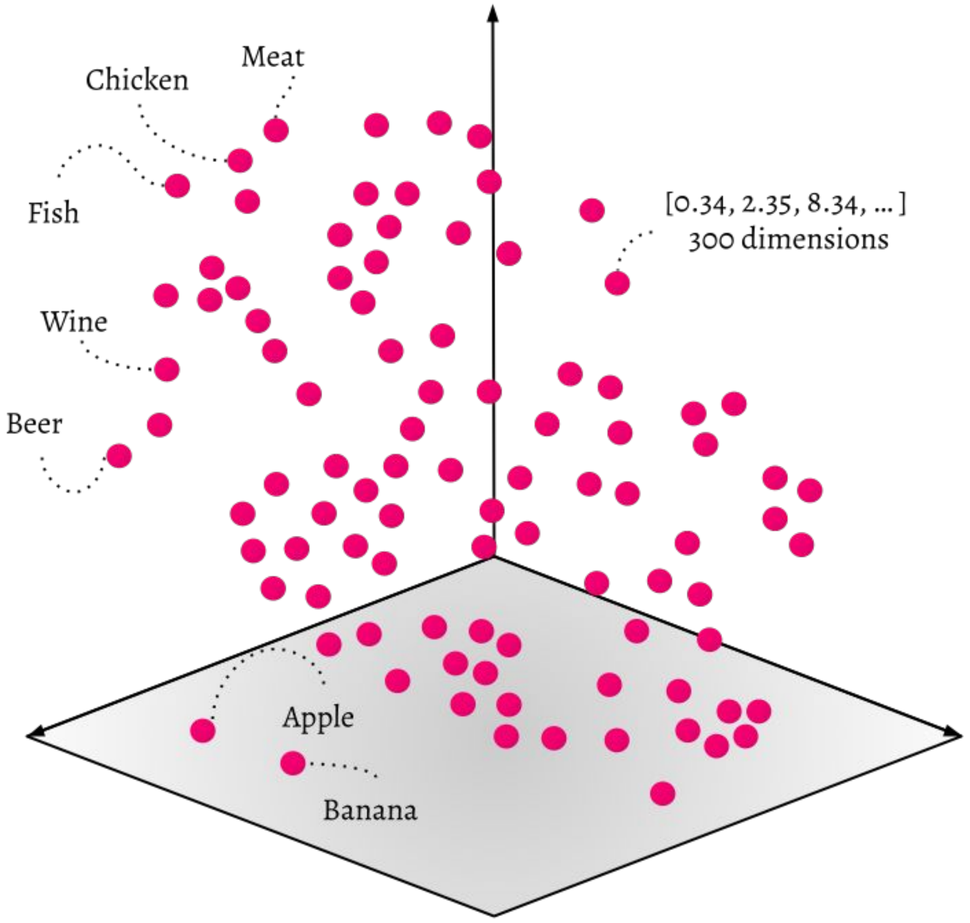
SeMI


Country
US
from Country
Chardonnay
Wine
Covey Run
Chardonnay
& description
in Country
Country
has category
from Winery
Variety
Chardonnay
Wine
Winery
Covey Run

Weaviate - How does it work?

- Data is stored as high dimensional vectors
- Vector positions capture data meaning and context
- Pre-trained NLP module for automatic
  - Vectorization
  - Classification
  - Nearest neighbor search

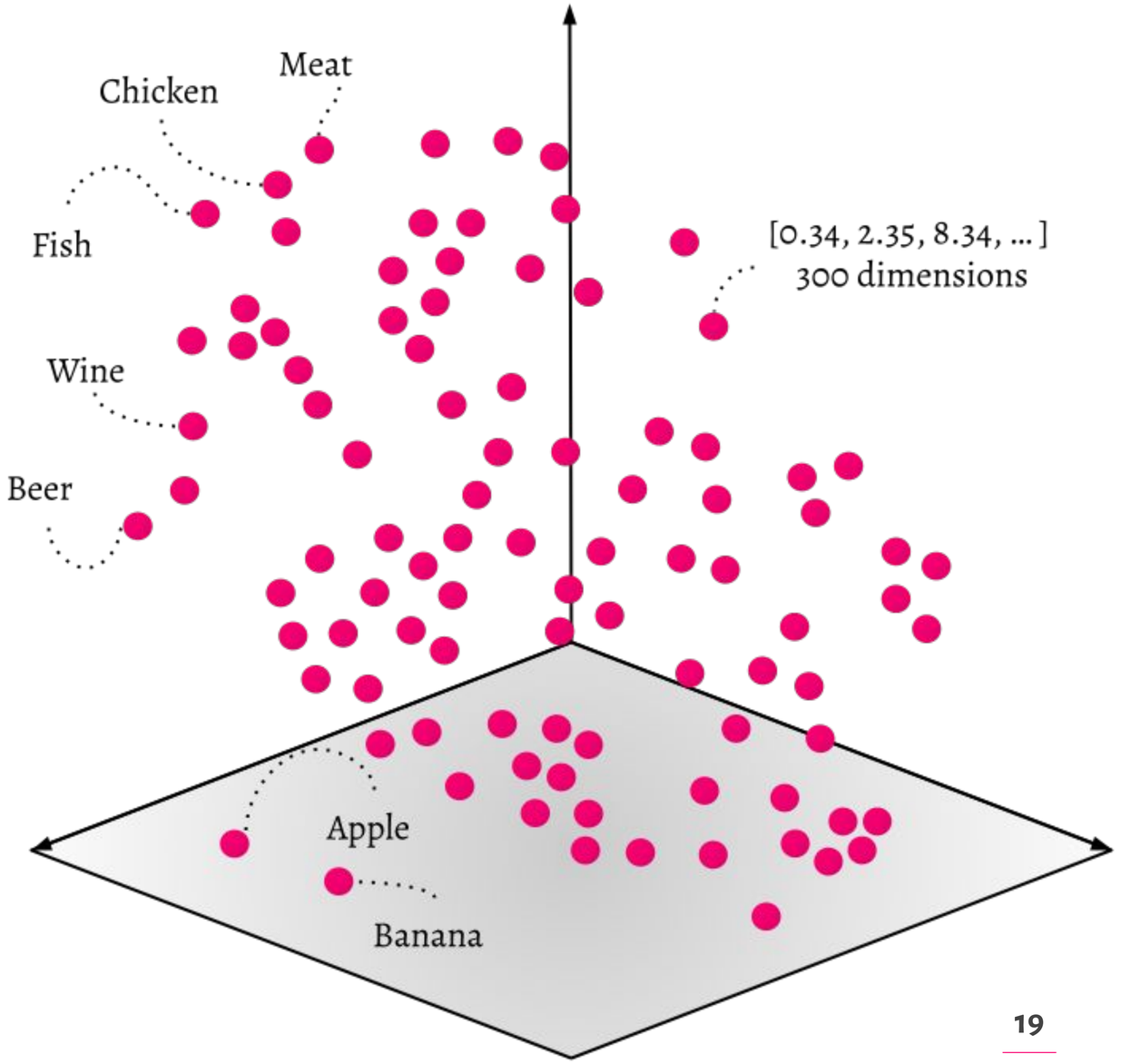
⇒ Weaviate *understands* the data



Weaviate - How does it work?

1. Weaviate Machine Learning Modules

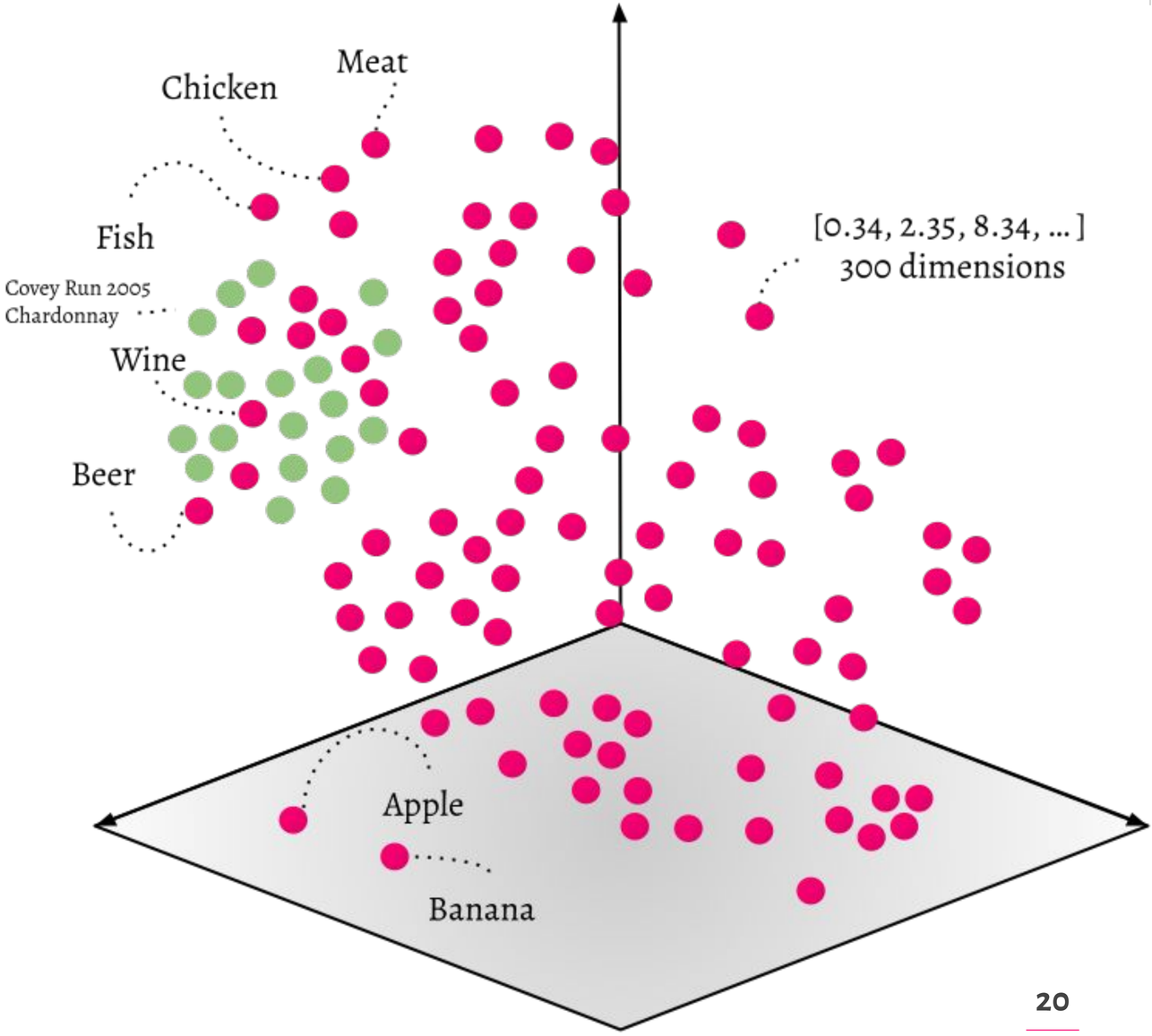
- E.g. NLP module trained with *fasttext*
- This language model represents all words and concepts in the hyperspace.



Weaviate - How does it work?

2. Automatically vectorize and index your data

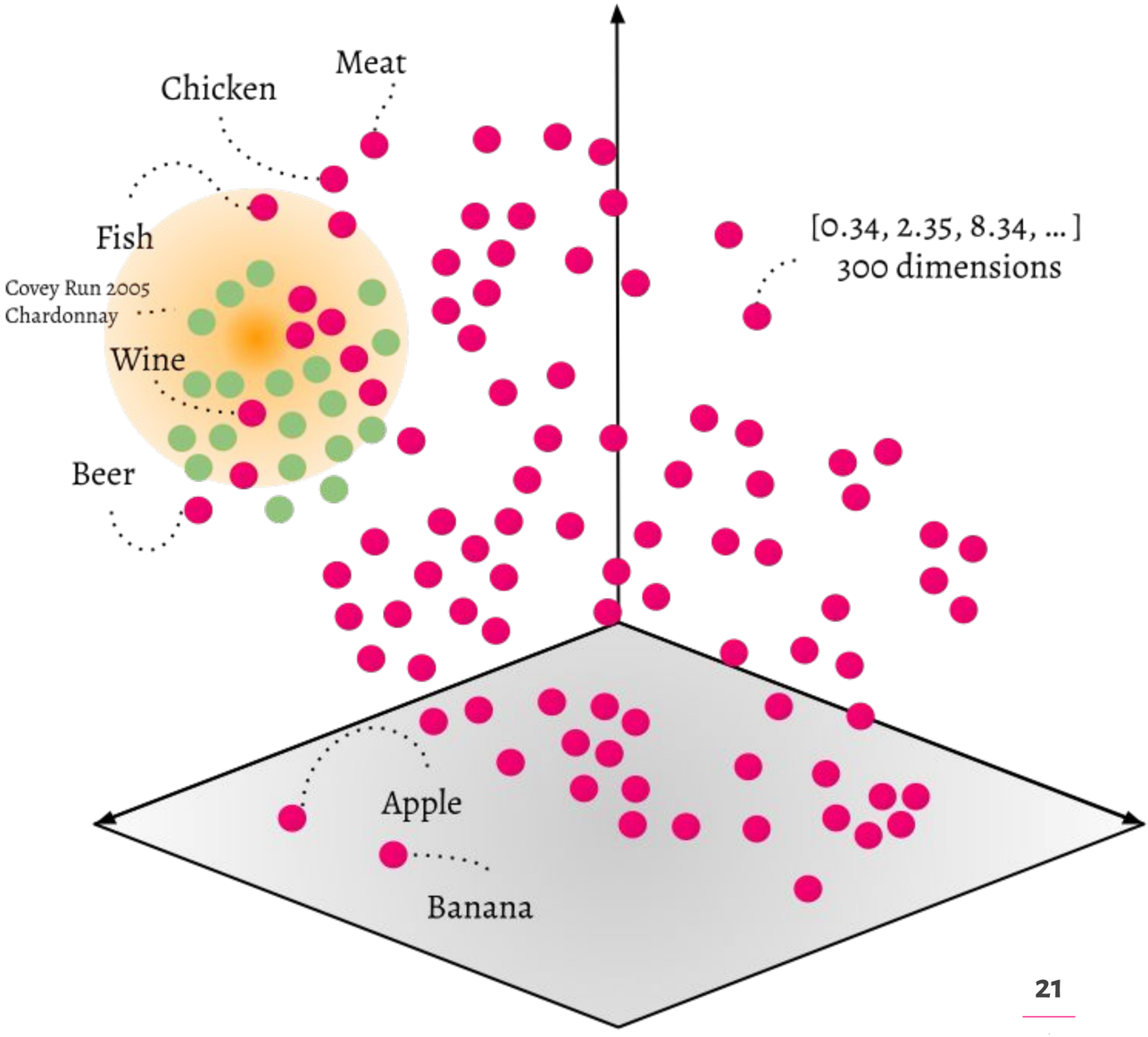
- Weaviate *understands* your data
- When you import data: Weaviate looks at the language in your data object
- E.g. a *Chardonnay* is closely related to *Wine*, *White* and the *food* it fits with



Weaviate - How does it work?

3. Search query

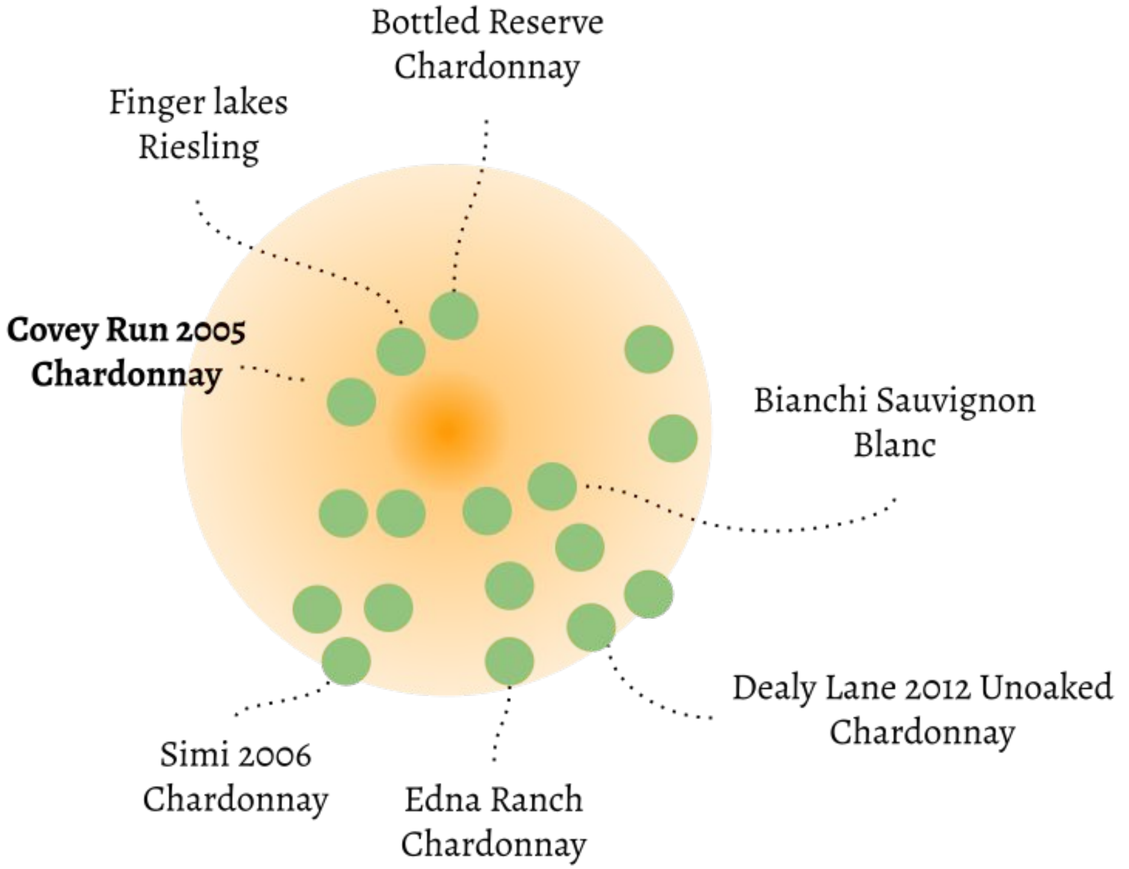
- Your search queries in natural language will also be vectorized and understood by the machine learning module of Weaviate
- It is places close to the words and data object that are semantically related to the query
- E.g. "*Wine that fits with a seafood dish*"



Weaviate - How does it work?

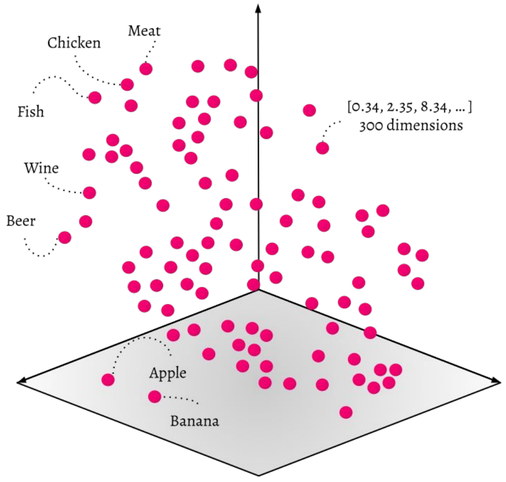
4. Results

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

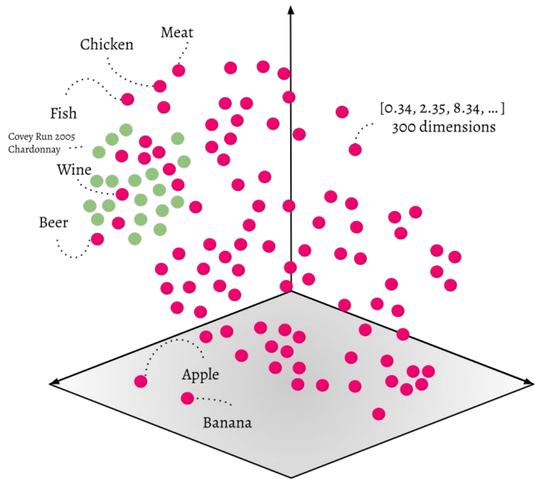


Weaviate - How does it work?

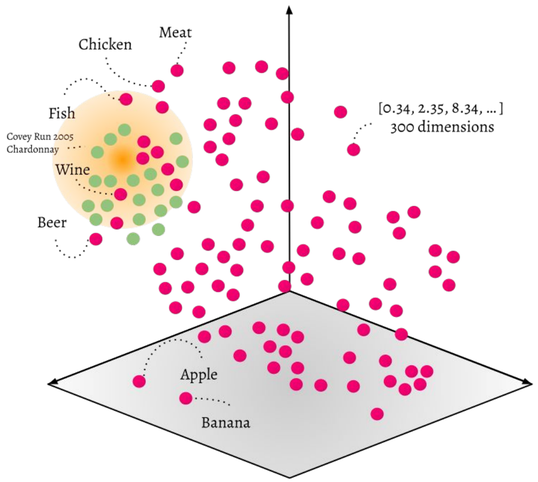
Pretrained ML model + Your own data + Your search query = Results



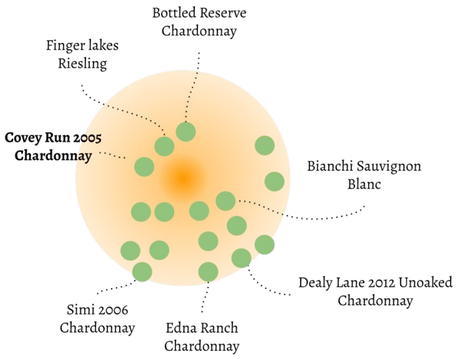
(e.g. NLP)



(e.g. Wines)



(e.g. Wine for seafood)



(e.g. Chardonnay)

SeMI

Demo

Weaviate - Scalable Vector Search Engine

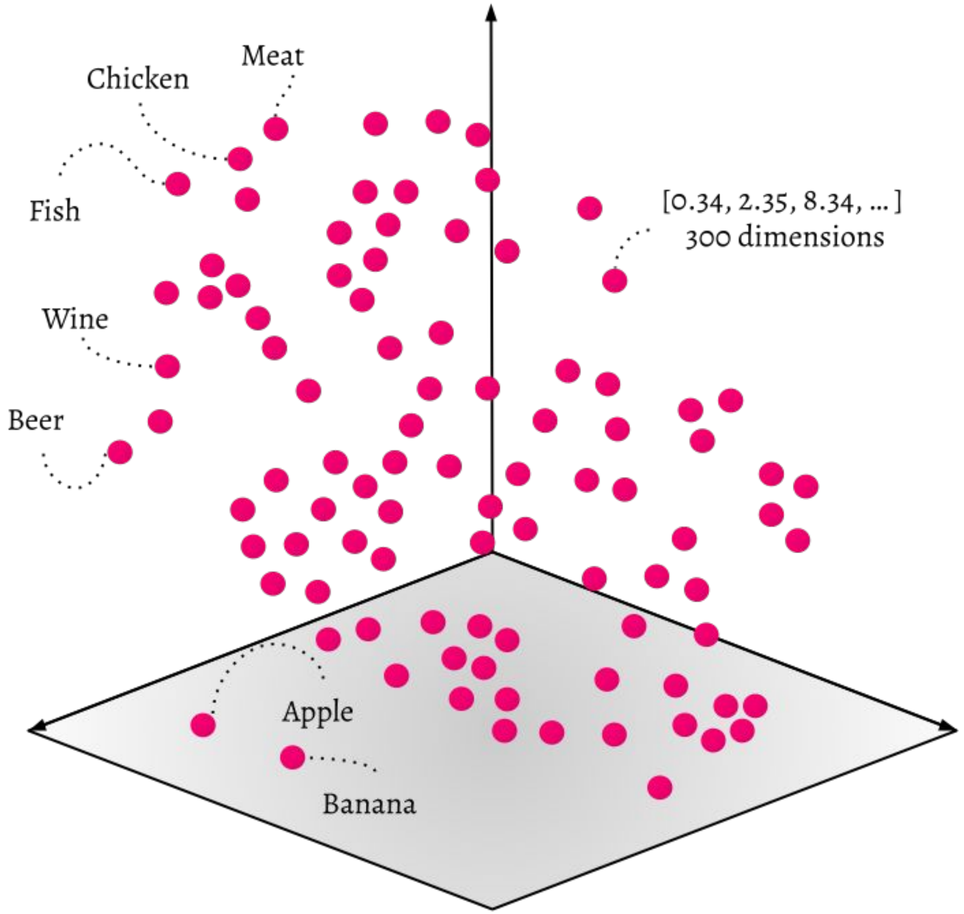
Weaviate Modules

out-of-the-box modules you can use
e.g. Weaviate NLP modules,
Transformer modules
Custom modules

Weaviate Vectorization Modules

Machine learning models

- **Weaviate's NLP module: Contextionary**
  - Trained with *fasttext*
  - Multiple languages: English, Italian, German, Dutch, etc
  - Extendable with custom concepts
  - Semantic Search
  - Contextual Classification



Weaviate Vectorization Modules

Machine learning models

- Weaviate's NLP module: Contextionary
- **Transformer inference module**
  - Support any PyTorch or Tensorflow transformers model, from **Huggingface Model Hub** or from **your own disk**.
    - e.g. BERT, RoBERTa,, DistilBert, etc
  - An easy way to deploy your Weaviate-optimized transformers NLP inference model to production using Docker or Kubernetes
  - Use **pre-built images of public** models selected to be good for semantic search tasks
  - Use custom build with any **Huggingface model** (after fine tuning for example)
  - Use custom build with a **private or local model**

Weaviate Vectorization Modules

Machine learning models

- Weaviate's NLP module: Contextionary
- Transformer inference module
- **Custom modules**
  - For custom and unique use cases, other data types, adding (ML) logic, etc

Weaviate - Scalable Vector Search Engine

Weaviate Modules

out-of-the-box modules you can use
e.g. Weaviate NLP modules,
Transformer modules
Custom modules

End-to-end

Complete solution for industry
API driven
Cloud-native
Secure

Weaviate is an End-to-End solution

- **Complete solution for industry use cases**
  - Easy to integrate
  - Fast, reliable & scalable
- **API driven**
  - RESTful and GraphQL
  - Connect to your existing infrastructure
  - Build custom applications on top
- **Cloud-native**
- **Secure**
  - On-premise

Weaviate - Scalable Vector Search Engine

Weaviate Modules

out-of-the-box modules you can use
e.g. Weaviate NLP modules,
Transformer modules
Custom modules

End-to-end

Complete solution for industry
API driven
Cloud-native
Secure

Search

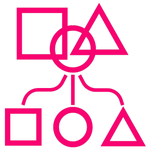
how to search the vector space

Open Source

all code and models are open

Scalable and fast

Vectorizing and querying big data

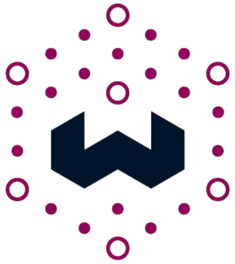


Classification

how to classify in the vector space

Example use cases

Weaviate
Enterprise Search



Weaviate
ERP classifications

Weaviate
Cybersecurity

Weaviate
Biology

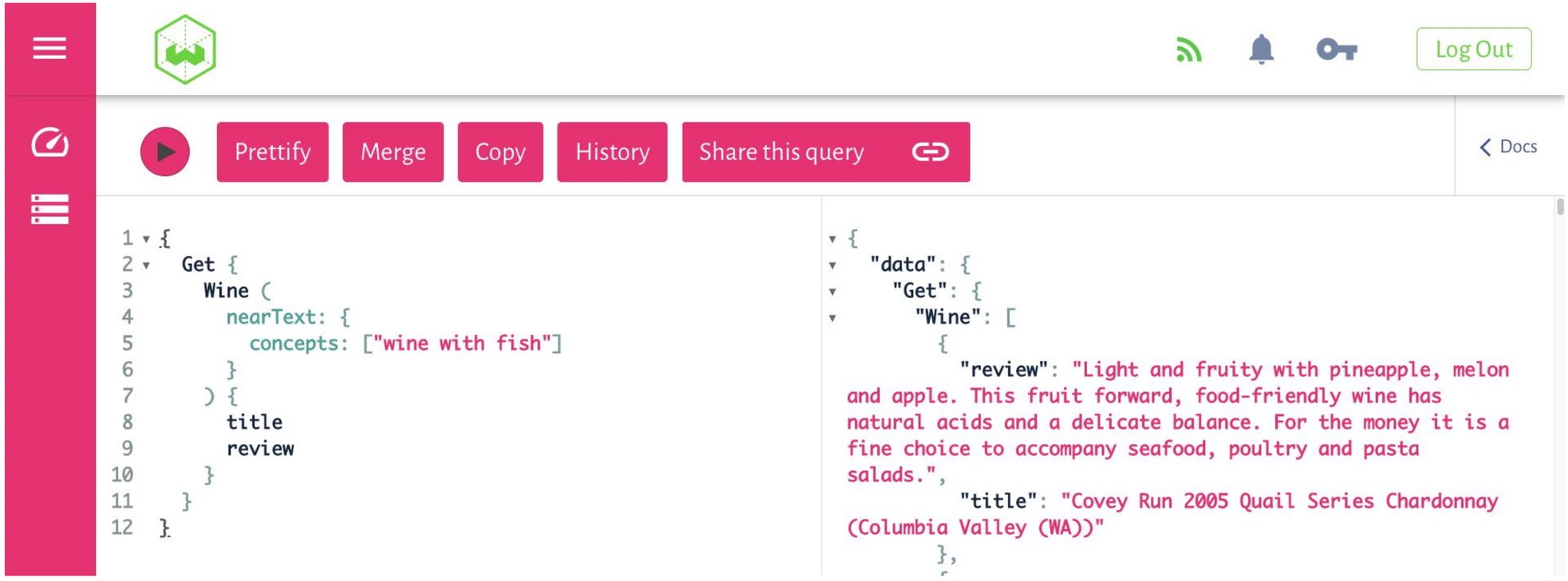
Weaviate
Data Harmonization

Weaviate
Social Media

Weaviate
Service Management

Try out Weaviate yourself!

- Go to www.semi.technology/developers



SeMI Technologies

Team

Etienne

Donna

Laura

Micha

Bob

Stefan

Marcin

Henrique

Amar

Alicja

Thank you!

www.semi.technology

Laura Ham

laura@semi.technology

SeMI