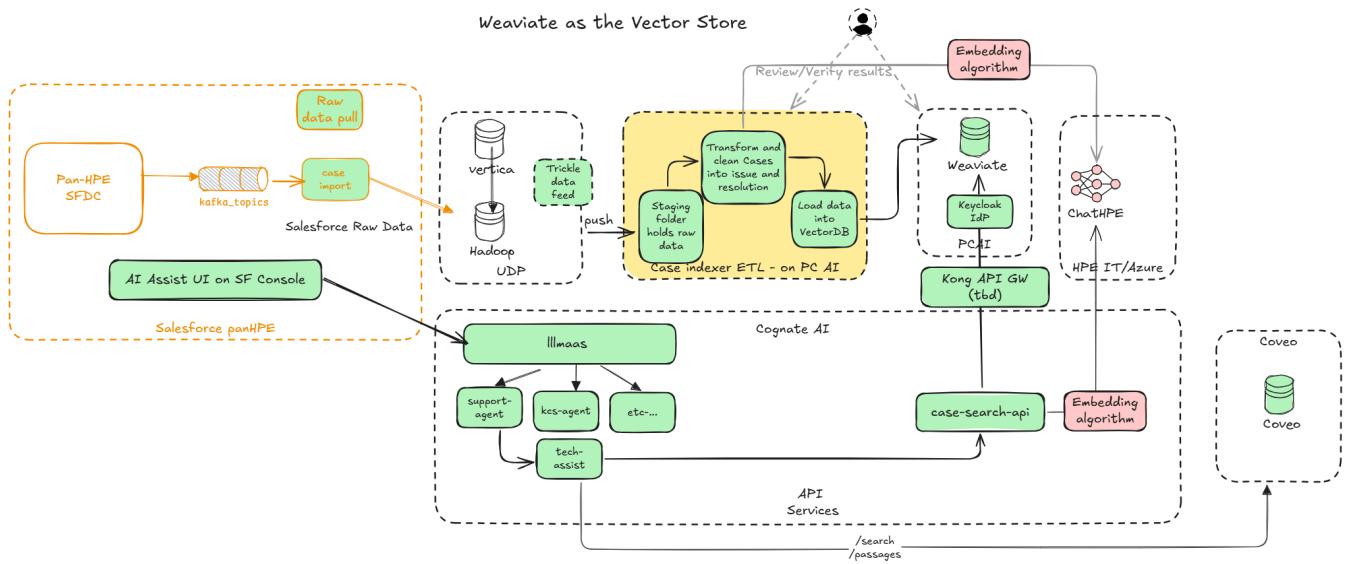


# Architecture for Semantic Historical Case Search (VectorDB)

Semantic Historical Case Search



Features are captured in Rally under **C6191**

Related epics: E1589, E1420, E1422, E1398

- Activities
- Milestone stages
- Weaviate VectorDB
- Data Transformation
- Localization
- Connections into Weaviate
- Authentication
- Authorization
  - Required roles
  - Keycloak Roles Mapping & RBAC
  - Application Clients
  - Weaviate as an OIDC client with Keycloak - application authentication example
- Embeddings
- Case-search-api
- API Gateway Kong
- Weaviate Setup
  - Requirements
  - Kubernetes Requirements
  - Sizing (a guess)
  - Enabled Modules
  - Deployment Flow
  - ITG/Staging
- Case fields
- Investigations

Activities

Activity	Owners	Notes	By When
Setup PC AI ITG developer accounts for Algoleap, DE and UDP team	DT&A	Developer accounts. Provide access to sufficient storage to host 1 years of historical case and working space for the data to be transformed	Nov 5, 2025
Install Weaviate on One AI PC AI	DT&A	Weaviate. See below for more detail on install. Decide on segregation of VectorDB - multi tenant approach needed. App connects as a tenant and may access one or more schemas. Needs to be cluster version of Weaviate that is highly available HA - multiple containers. Shared storage, DR, Sharding	Nov 10, 2025
Select embedding algorithm (see embedding section below)	DE Architects	Investigate which embedding algorithm should be used and confirm the chunking strategy. Loading VectorDB with data and searching VectorDB (case-search-api) must use same embedding algorithm	Nov 10, 2025
Metadata fields - searchable fields, filterable fields	DE Architects	Define exact list of metadata fields to be use for case search	Nov 14, 2025
Setup schema for DE CognateAI /Algoleap on Weaviate	DT&A	Assign some people to be admins on Weaviate. Provide training to them. Define process to create schemas and tenants and to link them for authorization purposes	Nov 19, 2025
PC AI PRO is ready for pipelines to run	DT&A	PC AI Pro is operational. Applications setup.	Nov 19, 2025
Enable external connections to reach Weaviate from Colo source	DT&A	Start with PC AI ITG. Configure Weaviate as an OIDC client with Keycloak. Make any needed changes to Ingres, Firewall, DNS setup, Istio to enable a connection from outside the PC AI cluster to reach Weaviate and run a simple query. Configure <a href="#">Istio Virtual Service</a> to expose the endpoint	Nov 14, 2025
Sample python client connection via SDK to Weaviate to validate setup is ready to go	DT&A	Use Python sdk on weaviate - demo a sdk client connection from outside of PC AI cluster connecting to weaviate installed inside the cluster. Application authentication must be used	Nov 24, 2025
Enable incoming OIDC connections to Weaviate on PC AI	DT&A	Configure authentication/ authorization. Keycloak is PC AI IdP. Azure AD used for users not app accounts.. Will Keycloak JWT tokens be federated with IAM Okta or standalone? Application authentication must be used. Document onboarding process to setup addition of new application accounts. Determine to what extent user/app pre-provisioning is needed for new OIDC connections. Include details in the documented process	Dec 2, 2025
Weaviate Backups/ restore process, DR	DT&A	Define and test this process	Jan 21,2026
Schedule job to push the agreed fields from UDP to PC AI file system in a Staging folder	UDP	Data would be in JSON format in a Staging folder with filenames having a timestamp appended to the name  Case Email attachments are out of scope at this stage but email body text and sender, receiver metadata would be available. See excel with list of fields  UDP need access to PC AI ITG (accounts setup, env ready to go) and 1 week of time to do this.	Nov 10, 2025
Get same push job running with PC AI Pro	UDP	Transition job to run in Pro. Note: Application is not live yet	Nov 21, 2025
Enhance UDP data pull from SF for missing fields	UDP	Gap analysis has happened on missing fields - Case emails are missing but being worked. Can preserve existing data structure 1:n between case and items like case email. Add in missing fields to the push to PC AI file system	Nov 24, 2025
Define schema for Weaviate	DE Architects	Define schema for use by Algoleap	Nov 10, 2025
Define case-search-api OpenAPI3 swagger specification	DE Architects	Define initial OpenAPI3 swagger specification	Nov 24, 2025
Validate connection to Weaviate from Colo SMART docker/swarm by building simple python client to connect with PC AI ITG and return vectors from the WeaviateDB	DE	Simple python client that calls Weaviate sdk to test connectivity. Make any firewall changes needed to allow connection from Colo swarm docker to Weaviate	Dec 10, 2025
Stand up mock version of case-search-api to enable CognateAI integration to start	DE	Define mock api stubs and generate tests - SMART docker/swarm - dev/itg	Feb 2, 2026
Setup accounts on PC AI ITG for Algoleap, DE	DT&A	Algoleap can log into PC AI ITG and run sample Spark pipelines. Can access staging folder contain case data. Look at using Projects on PC AI so the team can collaborate <a href="https://support.hpe.com/hpsc/public/docDisplay?docId=a00iae19hen_us&amp;page=Projects/Projects.html">https://support.hpe.com/hpsc/public/docDisplay?docId=a00iae19hen_us&amp;page=Projects/Projects.html</a>	Nov 3, 2025
Airflow/ spark pipeline transformation jobs	Algoleap using One AI PC AI infra	Transform data into issue and resolution concatenated text	Nov 17, 2025
Remove PII data	Algoleap using One AI PC AI infra	Remove PII data, serial numbers, ip address, customer names and addresses, personal emails in text,	Nov 24, 2025
Setup application access for Algoleap to call ChatHPE	IT ChatHPE	Algoleap need to be able to use an application account to call ChatHPE embedding api. See if they can leverage DE's existing account by reading Hashicorp or do they need own area - perhaps IT Hashicorp	Nov 18, 2025

Access to Hashicorp or similar for retrieve client secret needed to access ChatHPE embedding algorithm	IT	Get credentials into IT Hashicorp for use for calling ChatHPE embedding. Get Algoleap team setup to access that Hashicorp or similar	Nov 27, 2025
Clean issue and resolution text by using ChatHPE	Algoleap using One AI PC AI infra	Send concatenated text to ChatHPE LLM to create clean issues and resolutions for better semantic matching  Use embedding algorithm	Dec 1, 2025
NCS Changes ITG	DE	Firewall changes NCS to enable CognateAI case-search-api to connect to Weaviate ITG	Dec 3, 2025
Load cleaned up data into VectorDB - small sample data set	Algoleap using One AI PC AI infra	Data pipeline job running on PC AI. Index issue text by agreed metadata – product number, product family etc	Nov 24, 2025
Get data load pipeline running in Pro (not mtp - testing in pro to validate that env)	Algoleap using One AI PC AI infra	Initial use of Pro env for data pipeline runs. Solve any access issues. Have script to delete data in Weaviate to start over as needed	Nov 28, 2025
Replay/ recovery data load	Algoleap using One AI PC AI infra	In the event of interruption, delay of loss of data Pipeline job to write to Weaviate vectorDB must be idempotent and capable of multiple reruns at any time to cover failure scenarios  Monitoring and alerting if data pipeline has issues	Dec 3, 2025
Load cleaned up data into VectorDB - full 1 year historical dataset	Algoleap using One AI PC AI infra	Data pipeline job running on PC AI. Index issue text by agreed metadata – product number, product family etc	Dec 5, 2025
Business validation on data Algoleap has loaded	GRS	Access agreed sample of transformed data in VectorDB. Provide feedback on data quality	Dec 9, 2025
Trickle feed of new Case data and archiving of Case data older than 1 year	Algoleap using One AI PC AI infra	Data pipeline job for continuous trickle of new data and archiving old	Dec 10, 2025
Rework data clean and transform pipeline	Algoleap using One AI PC AI infra	Rework data cleanup and transform following feedback from Business GRS	Dec 12, 2025
Demo/ sign off	GRS	Demo to GRS of data quality of transformed issues and resolutions and Sign off by GRS	Dec 12, 2025
Email attachments for case emails converted into a structure format	UDP	Email attachments for case emails converted into a structure format and now included in the data that gets moved to PC AI	Jan 12, 2025
Start Pushing email attachments	UDP	Staging folder on PC AI now starts getting files for attachments	Jan 28, 2026
Configurations for Kong API GW (tbd)	DE/ Mobsters	Tbd if API GW will be used	Jan 16, 2026
NCS Changes Pro	DE	Firewall changes NCS to enable CognateAI case-search-api to connect to Weaviate.	Jan 23, 2026
Complete 11 features on NFR setup of PC AI	DT&A / Professional services	C5374 - Professional Services - OneAI Team: OneAI Infra readiness & Supportability NFRs	Jan 30, 2025
Create Case search rest api to access VectorDB	DE	Run on DE SMART infra so as to leverage existing build/deploy mechanisms. Hide embedding algorithm chosen from consumer.  Team developing the Case search rest api includes the embedding algorithm lookup – Nomic (local) or OpenAI embedding (cloud via ChatHPE)  <a href="https://huggingface.co/nomic-ai/nomic-embed-text-v1.5">https://huggingface.co/nomic-ai/nomic-embed-text-v1.5</a>	Feb 27, 2026
Integrate VectorDB rest apis into Cognate AI Agent	DE	DE Team	March 13, 2026
Complete remaining 8 features on NFR setup of PC AI	DT&A / Professional services	C5374 - Professional Services - OneAI Team: OneAI Infra readiness & Supportability NFRs	Mar 20, 2026
Complete out Epics E1589 and E1420 for Q2 deliverables	DE	Epics E1589 and E1420	Q2 2026
Integration Testing with IT AI Assist UI	DE, IT	TBD - details finalized at start of Q2	Q2 2026
Business acceptance Testing with IT AI Assist UI	GSR, DE, IT	TBD- details finalized at start of Q2	Q2 2026

## Milestone stages

Milestone	Owner	Notes	By When
PC AI ITG is ready for users to get added	DT&A/ IT		Nov 5, 2025
PC AI PRO is ready for pipelines to run	DT&A/ IT		Nov 19, 2025
Algoleap onboarded into PC AI ITG and have access to read data from UDP	DT&A/ IT and UDP		Nov 10, 2025
First data loaded by Algoleap into Weaviate VectorDB	Algoleap		Nov 24, 2025
Algoleap demo's calling ChatHPE from Spark pipeline jobs	Algoleap		Dec 1, 2025
C5374 - Professional Services demo NFR readiness on PC AI	Professional services		Jan 30, 2026
GRS signoff on Data quality of transformed data	GRS		Dec 14, 2025
Create Case search rest api to access VectorDB	DE		Feb 27, 2026
Integrate VectorDB rest apis into Cognate AI Agent	DE		March 13, 2026
Integration Testing with IT AI Assist UI	DE, IT		Q2 2026
Business acceptance Testing with IT AI Assist UI	GRS, DE, IT		Q2 2026

## Weaviate VectorDB

How is data segregated in Weaviate ? API key could be used to separate this. Multiple API keys per DB user but one DB user per app

- **Schema/Class:** Segregates data by type or domain.
- **Tenant (Multi-tenancy):** Segregates data by organization or logical group (acts like a namespace).
- **User:** Not a built-in segregation unit, but can be implemented via tenants or schema design.

Weaviate has its own rest apis for crud and semantic search. Would use weaviate database on prem not the cloud instance.

Define schema in Weaviate - table - class mapping - define vectorization, and fields on the class - metadata we want to use.

How will schema changes be handled in Weaviate? change control, update scripts - pipeline impacts - update schema in repeatable, predictable way

Should indexes be shared? Maybe not. Assume DE is the **data owner** for this data in Weaviate and exposes it through the case-search-api. Any changes needed DE will make them so to insulate callers from these changes

Tech assist could call Weaviate directly via langchain connection to the vectorDB - better to use the Weaviate PYTHON SDK as it will have more functionality. Also want to avoid direct call from CognateAI agent for abstraction purposes and go through the case-search-api

## Data Transformation

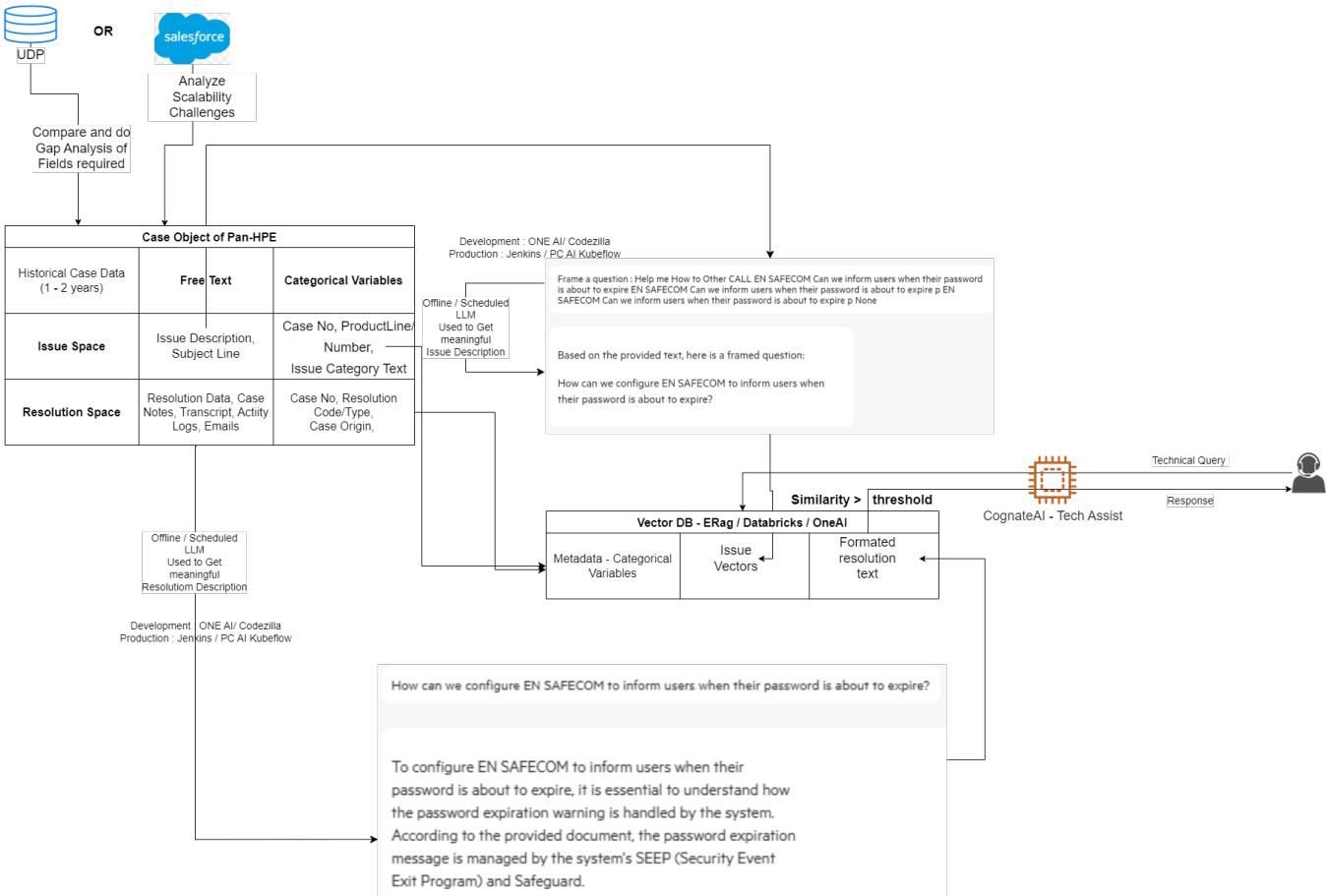
See best practices on data transformations: [Data transformation guidelines](#)

Note: need way to review transformed data by Business. Transformed data will be on a folder on PC AI. Cannot use Weaviate as the vectors are represented as numbers there and are not human readable. GSR can look at a sample of cleaned up issue and resolution and using CaseID compare it with the original.

Algoleap could produce spreadsheet showing raw and transformed for Acceptance test purposes

Define reqs to Algoleap for yellow box. ETL - data sources, fields, Transform steps- define these. Clean up steps what we expect. Initial seeding should be defined - starting chunk of data, then trickle. Reconciliation step - data missing - how to recover from data corruption - identify scenarios.

Cleaning step: strip PII data including addresses, serial numbers, IP address to give better semantic match before insertion into VectorDB - makes data more usable



## Localization

Decide on localization: English only at this time. Many cases are in other languages but out of scope for phase1. Cleanup and transformation, embeddings in multiple languages - normalize the data. Matching on issue description.

## Connections into Weaviate

Connections to Weaviate: Methods an app can connect to Weaviate

- Rest api -
- gRAPHql API
- **Client SDK - like Python**
  - SDK communicates with the Weaviate server over HTTPS
  - SDK uses Python's `requests` library or similar HTTP clients to send requests to the Weaviate endpoint.
  - The SDK manages persistent connections and retries internally.
  - You can configure timeouts, retries, and headers if needed.
  - App authentication via Bearer token or OIDC
  - AuthZ not enforced by sdk but by Weaviate server and its integration with Idp (Keycloak)
  - Scopes and Permissions: Defined in your identity provider (e.g., Auth0) and enforced by Weaviate.
  - Custom Gateways: You can place Weaviate behind an API gateway (e.g., Kong, NGINX) to enforce additional auth rules.

Using the ClientSDK may be preferable to using the rest api directly as you avoid having to know all the rest api error codes, return types and lower level details - examine further. Check if there is some reason that would favour the use of the rest api over the sdk

## Authentication

Authentication (IT may decide this - check) Check Weaviate integration with Keycloak Idp

1. API keys/Bearer token - **Did** Weaviate generate any Keycloak involvement? - Need GW or proxy out front? - put into Hashicorp secrets
2. OIDC - support for Azure AD.

The authentication requirements are as follows:

- Single domain alias
- TLS 1.3, HPE Cybersecurity-approved ciphers only
- Ingress certificate signed by HPE DigitalBadge. Enforce TLS from ingress.
- OIDC authentication with PCAI Keycloak enabled
- OIDC JWT validation is enabled
- OIDC-enabled users
  - End users
    - Interactive flow (Authorization Code), provisioned through Azure AD
  - Service client:
    - Enabled for client credentials grant type (machine-to-machine).
- Keycloak groups/roles map to Weaviate RBAC (via group claim).
- **RBAC enabled**
- Disable anonymous access.
- Enforce minimum token signature alg (RS256/RS512).
- Required users
  - cognate-admin
    - create collection, read definition, update definition, delete collections
    - create, read, update, delete objects (matching cognate\_\*)
  - cognate-ro
    - read objects (matching cognate\_\*)
- Multi-Tenancy enabled per class
- Special tenant for Cognate (data isolation) - tenant\_cognate

## Authorization

RBAC via OIDC claims - roles can be assigned via claims in the token. Custom authorization: Could assign role to DB user.

### Required roles

The following are the suggested roles we need to be present in Weaviate; names are indicative.

Role Name	Keycloak realm_access.role	Type	Purpose	Allowed Actions in Weaviate
weaviate_manager	weaviate_manager	Human or machine	Maintain data and collections	- Create schemas. Create/update/delete objects - Create and manage collections/classes (if needed) - Read data - No access to system-level features (backup, metrics, etc.)
weaviate_reader	weaviate_reader	Human	View-only access	- Read objects and collections - Query metadata and schema - No write, delete, or schema mutation
weaviate_admin_client	weaviate_manager	Machine	Used by CI/CD and automation pipelines	- Same permissions as weaviate_admin - Typically uses client credentials grant
weaviate_app_client	weaviate_client	Machine	Application backend access (data read /write)	- CRUD objects - No schema or system-level operations - Typically uses client credentials grant

## Keycloak Roles Mapping & RBAC

Map the Keycloak JWT claim **realm\_access.roles** to the Weaviate roles (role names are indicative)

- Claim: realm\_access.roles
  - Roles to claim value
    - weaviate\_manager: weaviate\_manager
    - weaviate\_reader: weaviate\_reader
    - weaviate\_admin\_client: weaviate\_admin
    - weaviate\_app\_client: weaviate\_manager

Multitenancy is based on the **tenant\_id** claim, optionally passed by Keycloak. For all Cognate clients, the claim is set to tenant\_cognate.

Keycloak must pass the necessary claims for human users and API clients.

## Application Clients

Create the following application clients

Client ID	Grant Type	Roles Assigned

weaviate-admin-client	Client Credentials	weaviate_manager
weaviate-app-client	Client Credentials	weaviate_app_client

The **weaviate\_manager** role must have the following permissions to enable automated deployments.

Operation	Alias name filter (regex)	Why is it needed
Collections: create, read definitions, update definitions, delete	^cognate_.*	Deploy and evolve collections.
Collection aliases: create, read, update, delete alias	^cognate_.*	Deploy and evolve collections.
Data objects:create, read, update, delete	^cognate_.*	Seed or migrate data
Replications:create, read, update, delete	^cognate_.*	Manage replications
backups:create / backups:restore	^cognate_.*	Run pre-deployment backups
system:get (optional)	^cognate_.*	Verify cluster state or version before deployment

## Weaviate as an OIDC client with Keycloak - application authentication example

[Weaviate as an Oidc Client with Keycloak on PC AI](#)

## Embeddings

ChatHPE Embeddings provide access to OpenAI embeddings <https://platform.openai.com/docs/guides/embeddings>

API call to use to create embeddings with ChatHPE.

[https://developer.chatpe.it.hpe.com/api-details#api=chatpe-api-wrapper-v2-9-4&operation=call\\_create\\_embedding\\_call\\_create\\_embedding\\_post](https://developer.chatpe.it.hpe.com/api-details#api=chatpe-api-wrapper-v2-9-4&operation=call_create_embedding_call_create_embedding_post)

Check if ChatHPE support open ai embeddings type in their signatures? If not will have to use external embeddings

Model	Dimensions	Multilingual Performance	Benchmarks (MIRACL / MTEB)	Cost per 1M tokens - Oct 2025	Strengths	Best Use Cases
OpenAI text-embedding-ada-002	1,536	Basic multilingual support	MIRACL: 31.4% MTEB: 61.0%	\$0.05	Cost-effective, widely adopted, good for general semantic search	Semantic search, clustering, RAG pipelines with moderate accuracy needs
OpenAI text-embedding-3-small	1,536	Improved multilingual	MIRACL: 44.0% MTEB: 62.3%	\$0.01	5x cheaper than ada-002, faster, better multilingual alignment	5x cheaper than ada-002, faster, better multilingual alignment
OpenAI text-embedding-3-large	3,072	Best multilingual alignment	MIRACL: 54.9% MTEB: 64.6%	\$0.065	Highest accuracy, richer semantic representation, supports dimension trimming	High-precision RAG, enterprise knowledge bases, multilingual semantic similarity
Nomic v1.5	~768–1,024	Strong multilingual (open-source)	Varies (competitive with SBERT)	Free (open-source)	Local deployment, privacy-friendly, good for multilingual tasks	Offline semantic search, privacy-sensitive environments, cost-free experimentation

Further discussions needed on chunking strategy and on which embedding model to use

## Case-search-api

Weaviate does a combination hybrid search of lexical + vector semantic search

```
filtered_search_results = db.similarity_search(query, filters=search_filter, k=3)
```

relevance "score" could be returned in the response

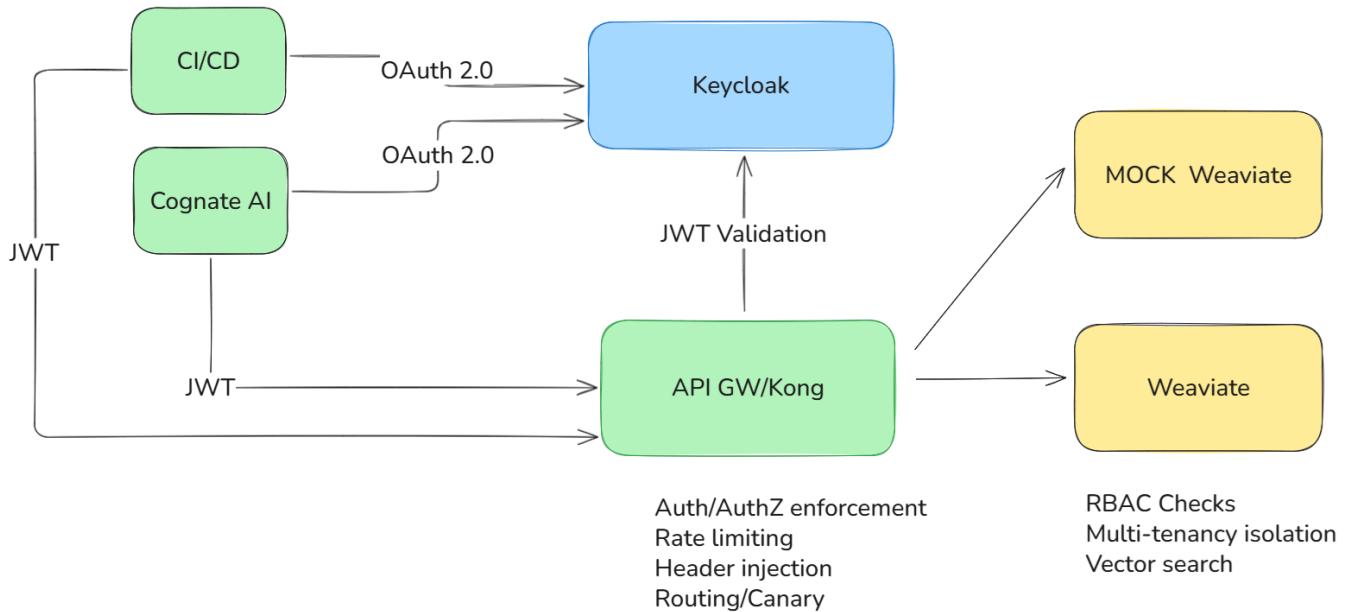
Passed arguments - (k #nearest neighbour of vectors to search for = 5, search text as text field mandatory - eg issue text, resolution text, map of following name/value pairs with at least 1 populated of these option items product line Product\_Line\_\_c, asset name, product series, remote resolution field type - onsite/csr etc, Issue\_Type\_\_c

Filter list

Responses - list Case IDs, list of cleaned issues, list of cleaned resolutions,

FilterDB

## API Gateway Kong



All DE access to Weaviate may go through the API Gateway - Kong.

- Applies rate limiting
- Keycloak integration
- Early JWT verification
- Routing changes/Canary support
- Allow for mocked backend

## Weaviate Setup

### Requirements

- Highly available Weaviate PRO instance on PCAI Kubernetes, deployment using Helm chart.
- OIDC authentication via PCAI Keycloak.
- Application (machine) access via client credentials.
- Client credentials shared via secure means.
- Support for multi-tenant collections enabled (Weaviate class-level multi-tenancy).
- Durable storage
- Backup
- Disaster recovery setup (cross-region copy).
- Secure
- Observable
- Hardening
  - Disable unused modules.
  - Set read/write timeouts (HELM: extraEnv).
  - Limit max query results (QUERY\_MAXIMUM\_LIMIT).
  - Enable request logging with redaction of secrets.
  - Rate limit at ingress (per client IP / token).

### Kubernetes Requirements

- 3+ Weaviate pods (odd number) for quorum-based cluster gossip.
- One dedicated Kubernetes namespace (e.g., cognate-weaviate-prod).
- Optional vectorizer modules (decide: built-in text2vec-transformers vs external inference). Keep modules stateless.
- PersistentVolume per pod (local data); plus shared storage for backups.

- Ingress (TLS) is terminated at the cluster gateway; mTLS (optional) is used between internal services.
- NetworkPolicy restricts ingress to Ingress Controller, Bastion, and approved app namespaces.
- Enable liveness/readiness probes (defaults).
- Horizontal scaling: manual first;
- more ... TBD

## Sizing (a guess)

- 200GB/node
- 1TB/shared

## Enabled Modules

Module	Description
text2vec-openai	Uses OpenAI embeddings API
text2vec-huggingface	Uses Hugging Face models locally or from the Hub
text2vec-cohere	Uses Cohere embeddings
text2vec-transformers	Runs transformer model locally (self-hosted)
img2vec-neural	Generates image embeddings
multi2vec-clip	Multimodal (image + text)

This is the usual set of modules that allows the database to compute the vectors on its own and use the OpenAI embeddings API

## Deployment Flow

- Allow remote deployments via Jenkins or GitHub actions - Weaviate must be accessible from COLO 1 and COLO 2.
- Allow access to health (/v1/.well-known/ready), simple query, auth token test

## ITG/Staging

- Functional parity with prod security (OIDC, RBAC) to validate auth & tenancy.
- No DR
- Integrated with ITG Keycloak
- 30% of production storage
- Monitoring, and available the same as PRO

## Case fields

### Support AI SFDC Case data

SO UR CE Tab le	SOUR CE Field	L e xi c al S e ar ch	V e ct or S e ar ch	R e s p o n se	UD P Table	UD P Field	Comment/Example
Case	Case_Resolution_Summary__c	Y	R e s ol ut ion	vw_de_sfdc_case_snapshot?	cas e_r esol uti on_s um mary		
Case	Case_Number	Y		vw_de_sfdc_case_snapshot?	cas e_n umb er		Panhpe case number in the format 5xxxxxxxxx
Case	Cause_Plain_Text__c	Y	I s s ue	vw_de_sfdc_case_snapshot?	ca use_ plai n_t ext		E.g. "DIMM failure", "HDD failure", "Storage Degradation", "4H - HPE StoreEver MSL6480 Tape Library - Tape drive broken"
Case	Close_Comments__c	Y		vw_de_sfdc_case_snapshot?	clos e_c om me nts		E.g. "Order created", "Order Converted to S4", "Record Merged Successfully", "Order approved to P20", "The sales credit for the eligible transactions has been applied in Sales Compensation Systems as an exception in accordance with the Global Sales Compensation Delegation of Authority Policy. Section 6.2.1. Please allow time for updates to be visible in MyComp Reporting. ... "

Case	Description	Y	Issue	vw_de_sfdc_case_snapshot?	description	
Case	Error_Codes_c	Y		vw_de_sfdc_case_snapshot?	error_codes	E.g. "218004", "FTO-4D", "cpStatusChanged-cpStatus-failed", "iLO_200_PowerSupplyFailure", "iLO_400_MemoryErrors", "PEAK0189".
Case	GSD_Environment_Plain_Text_c	Y	Issue	vw_de_sfdc_case_snapshot?	gsd_environm_ent_plain_text	E.g. "28.10.0.15", "Model: DL360 Gen9 OS: Windows", "2C: Drive Auto-Filling and Space Not Releasing", "HPE 3PAR StoreServ A630 2-Node Base", "[PART ONLY]//CLFP//CZ34278RY2//Dispatch Code ASAP//456096-001//NOIDA AG3//FE Not Required",
Case	Id	Y	Response	vw_de_sfdc_case_snapshot?	id	Salesforce identifier for the Case record.
Case	Issue_Plain_Text_c	Y	Issue	vw_de_sfdc_case_snapshot?	issue_plain_text	E.g. "Adapter 2 - HPE SN1200E 16Gb 2p FC HBA Status Degraded", "Blade is unreachable iLO is accessible", "CC: NSS0072024 Power Supply \$ZZCIP.S100233.2 Hardware Error", "DIMM and BIOS health degraded".
Case	Issue_Type_c	Y		vw_de_sfdc_case_snapshot?	issue_type	E.g. "Help me/How to", "HPESC Portal Issue", "Non-DBD/Non-Technical agent bypass", "Other", "Product Non-functional/Not working as Expected", "Part Order Logistics/Delivery/Administration".
Case	Parent_Id	Y		vw_de_sfdc_case_snapshot?	parentid	Salesforce identifier for the Parent Case for this Case.
Case	Product_Line_c	Y		vw_de_sfdc_case_snapshot?	product_line	E.g. "34", "35", "3S", "TI", "WB", "LA", "VL", "UY", "4A".
Case	Product_Number_c	Y			product_number	E.g. "BB958A", "HPEGLPLTSVCSPRE", "OTHSTORZ", "JW187A", "867055-B21".
Case	Product_Series_c	Y			product_series	E.g. "1008995294", "5335825", "1013297748", "None".
Case	Product_Hierarchy_c	Y		vw_de_sfdc_case_snapshot?	product_hierarchy	Salesforce identifier of the related Product Hierarchy.
Case	Resolution_c	Y		vw_de_sfdc_case_snapshot?	resolution	E.g. "<p>DIMM replaced</p>", "<p> </p><p>Event triggered due to FOS upgrade activity.</p>", "<p>13/10/2025</p><ul><li>Processed RMA.</li></ul>", "<p>Provided steps to update the server to latest version </p>".
Case	Resolution_Code_c	Y		vw_de_sfdc_case_snapshot?	resolution_code	E.g. "Electronic Case AutoClose", "Customer self-solved without HPE assistance", "IRS and 3Par STaTS False Alarm", "Onsite Repair", "Part Shipped".
Case	Resolution_Plain_Text_c	Y	Resolution	vw_de_sfdc_case_snapshot?	resolution_on_plaintext	E.g. "Engineer visited the site and performed the given POA, as per RJIL team now server health is good and working fine.", "Our records show that your HPE Primera bearing serial number CZ203607WZ has not phoned home to HPE within the last 4 hours. We will no longer be able to proactively monitor the system.", "Customer implemented certain planned changes as part of activity CHG0169095, which triggered the related alerts. After the completion of the planned activity, the RIO path state returned to normal. Customer confirmed that the issue is resolved and requested to close the case.".
Case	Root_Cause_c	Y	Resolution	vw_de_sfdc_case_snapshot?	root_cause	E.g. "CARE - Booking error", "CARE - Care pack query", "CARE - Info not available in OSV/VSA", "OM Fallout - RB6", "Missing products/ SP flag".
Case	Subject	Y	Issue	vw_de_sfdc_case_snapshot?	subject	E.g. "[AW] Order Queries/Order Status Request/3531391765", "[IMAU JOB:8102494] - OPE-0020490021", "[No account] SimpliVity Omnicube physical capacity is dangerously low. Usage: 100%.", "[Top Priority] Partner User Registration: (Scenario: New Company, Program) - DXP OPS UI", "Account[Insert Account Country]- New EMDM Account needs linkage to MDCP OPSI".

Task	Type (restrict records before Vector DB)	-	vw_de_sfdc_task_snap shot?	task	Type = Plan of Action, Trouble Shooting
Task	Description	Resolut ion	vw_de_sfdc_task_snap shot?	descript ion	E.g. "Informed the customer that we will check with the extended diagnostics team for further investigation and provide an update.", "Troubleshooting steps: MSA 2050-Falha em Disco Enc 1 Slot 5 Fault - Excessive Media Errors Model Number HUHT21212AL5204 Serial Number 5PH1SHXC", ""Issue description: Server Crashed Action Plan: Need to Replace the System Board Part needed: Yes - Part Number - P19926-001 - Part Description - Mother Board System I/O, with Subpan Customer wishes CSR for: No".
Task	Subject	Resolut ion	vw_de_sfdc_task_snap shot?	subject	E.g. "Plan of Action", "Plan of Action - Onsite", "Troubleshooting", "Next follow up", "L2 elevation - ASM involved hence elevating this now !", "History Information".
Wor dOr der	Closin g_Summar y_c	Resolut ion	vw_de_sfdc_work_order_snap shot?	clos ing_sum mary	E.g. "Happycall/HW;YES;CSC;YES;YES;disk replacement", "Visited site. Replaced faulty RAM in Proc2 Slot 12. Tested ok. Issue resolved."
Wor dOr der	Onsite _Action_c	Resolut ion	vw_de_sfdc_work_order_snap shot?	onsite_acti on	E.g. "Check the case log and contact the customer collect the log. The log shows a 8TB hard disk has error count. Waiting they provide the AHS to identify the location of the failed disk. Pass the case to DL team follow up during the office hour.",  "REVISIT PLAN OF ACTION =====Steps performed on site : Traveling to Maryland for the install. Revised plan of action : Unload, uncrate, and install racks. Document KM URL if applicable : Parts ordered without consulting CSC : false Revisit Reason : Installation Reason for ordering parts :  ORIGINAL PLAN OF ACTION =====  Cray Installation"
Wor dOr der	Proble m_De scripti on_c	Resolut ion	vw_de_sfdc_work_order_snap shot?	prob lem_descript ion	E.g. "Fault - Faulty RAM in Processor 2 Slot 12", "HPE ProLiant BL460c Gen8 Server-VHO7FUL02BL04ESX16 (NJEQ119031) - Controller has a status of failed.", "HPE Special Delivery Request/Equipment - fan board failures".
Cas eCo mm ents	Comm entBo dy	Is suse & Res olut ion	vw_de_sfdc_case_comm ent_snapshot?	comm entb ody	E.g. "The quote NQ09405645 is now Orderable and ready to work.", "10/14/25 - Order converted in Workspace- OP", "Convert to Order in NGQC application is successfully done.", "The issue was that the switch was not in sync with Aruba Central, displaying an error indicating that ""configurations were modified outside of Central. I requested minor changes to the VLAN description via Multiedit, after which the switch began to push previously made configurations, including an IP address change, causing the switch to go offline. Upon taking console access to the switch, there was no reachability to the internet. Two VLANs had IPs configured: in VLAN 1 and VLAN 106. However, the source IP connected to Aruba Central was different. After some time, VLAN 106 came up, passing through the uplink interface 1/1/14. Configured an IP route to establish internet connectivity, which allowed the switch to reconnect with Aruba Central and come back in sync. Made configuration changes in Aruba Central, which were successfully updated. The next step is to bounce the PoE ports. This can be done under Tools by running the test to bounce the PoE ports."
Wor kOr der Fee d (F edlt erm)	Title	Resolut ion	?	?	E.g. Mostly null, "urldefense.com", "support.hpe.com".
Wor kOr der Fee d (F edlt erm)	Body	Resolut ion	?	?	E.g. "Partner Notes 20251009, CE: Zaid Issue: HPE ProLiant DL380p Gen8 ServerCache Module Status Degraded Action: Replace cache battery, 660093-001, SN New: Notabletolocate, SN Old: Notabletolocate1", "<p>updated d/t per POA notes:</p><p><p>-Expected arrival date and time of the work request15/Oct/2025</p><p></p><p>-pushed to Hemmersbach for assignment to resident CE</p>", "Submitted an Onsite Work Order 01677232, the part is available P08356-001 SPS-PCA IB HDR100/EN 100G 940QSFP56 Adpt 2025-10-16 10:30".
Em ailM essage	Subject	Resolut ion	In progre ss.	In progre ss.	E.g. "RE: HPE Support Case 5392877906 [ ref!:00Dd00bUIK.!500Kh0fwcXA:ref ]", "RE: AT&T Mexico-Normal-Case ID 5392812256-USPM not Running [ ref!:00Dd00bUIK.!500Kh0e51Ui:ref ]", "Automatic process failed to apply FAN on SR 27352 (NQ)".

EmailMessage	TextBody	Resolution	In progress.	In progress.	<p>E.g. "Dear Customer,</p> <p>Thank you for choosing Hewlett Packard Enterprise. Your email response is linked to a case that has been closed.</p> <p>If your case was closed in error or you are still experiencing issues, please contact HPE Response Center (<a href="https://support.hpe.com">https://support.hpe.com</a>) to open a new case and provide your previous case ID for reference.</p> <p>THIS MESSAGE WAS SENT AUTOMATICALLY; PLEASE DO NOT REPLY TO THIS EMAIL.</p> <p>Sincerely, Hewlett Packard Enterprise",</p> <p>"&lt;l-template 18 25 fail to apply FAN -&gt; This is an automated notification sent from the Solutions Delivery ProcessMate workflow tool. Please do not reply. The Service Request No 27352 dated 10-15-2025 09:59:3 has been approved. Due to system issue the FAN could NOT applied to your Quote ID NQ09411278-02. An SFDC case was generated for the Local CPQ team for support and they will contact you shortly to provide an update. The results for the Quote ID NQ09411278-02 and Opportunity OPE-0020386876 are below: RuleUnbuildable TypeApproval StatusFANAction DateComments 81322276 COM Hard Attach Approved FAN5254158 10-16-2025 12:05:37 Regards, SD ProcessMate(Workflow Tool)"</p>
--------------	----------	------------	--------------	--------------	---

Vector db fields				
Cleaned Up Issue Text		Response?		
Cleaned Up Resolution Text		Response?		
Cleaned Up Issue Embeddings	Vector Search			
Cleaned Up Resolution Embeddings	Vector Search			
Each of the lexical search fields above	Lexical Search			

## Investigations

### Links

<https://docs.weaviate.io/weaviate>

<https://docs.weaviate.io/weaviate/connections/connect-local>

[https://docs.weaviate.io/weaviate/api?\\_\\_hstc=13542376.05bcf0de1f92292c362e4bae7099cf22.1760292817684.1760292817684.1760292817684.1&\\_\\_hssc=13542376.2.1760292817684&\\_\\_hsfp=2422918963](https://docs.weaviate.io/weaviate/api?__hstc=13542376.05bcf0de1f92292c362e4bae7099cf22.1760292817684.1760292817684.1760292817684.1&__hssc=13542376.2.1760292817684&__hsfp=2422918963)

[https://docs.weaviate.io/weaviate/client-libraries?\\_\\_hstc=13542376.05bcf0de1f92292c362e4bae7099cf22.1760292817684.1760292817684.1&\\_\\_hssc=13542376.2.1760292817684&\\_\\_hsfp=2422918963](https://docs.weaviate.io/weaviate/client-libraries?__hstc=13542376.05bcf0de1f92292c362e4bae7099cf22.1760292817684.1760292817684.1&__hssc=13542376.2.1760292817684&__hsfp=2422918963)

<https://weaviate.io/blog/weaviate-1-33-release#oidc-group-management>

<https://docs.weaviate.io/deploy/configuration/authentication#oidc-docker>

<https://docs.weaviate.io/weaviate/api/graphql/search-operators#hybrid>

### Weaviate + PC AI

[https://support.hpe.com/hpsc/public/docDisplay?docId=a00aie19hen\\_us&page=pcai/GenAI/knowledge-base.html&docLocale=en\\_US](https://support.hpe.com/hpsc/public/docDisplay?docId=a00aie19hen_us&page=pcai/GenAI/knowledge-base.html&docLocale=en_US)

### PC AI Identity management docs

[https://support.hpe.com/hpsc/public/docDisplay?docId=a00aie19hen\\_us&page=Security/identity-management.html&docLocale=en\\_US](https://support.hpe.com/hpsc/public/docDisplay?docId=a00aie19hen_us&page=Security/identity-management.html&docLocale=en_US)

### How to Setup Keycloak & Group-Based Authorization

### Examples

<https://medium.com/data-science/getting-started-with-weaviate-a-beginners-guide-to-search-with-vector-databases-14bbb9285839#cf85>

<https://python.langchain.com/docs/integrations/vectorstores/weaviate/>