Enprivacy Quick Start Guide

Enprivacy 3.0
Enprivacy Pte Ltd
www.enprivacy.com

Version 6
Updated 12 June 2026

Table of Contents

1 Executive Summary

Guided by the philosophy that every piece of data is a piece of someone's life story, Enprivacy develops solutions that bring a privacy-first design to data categorisation, redaction, and analysis, as well as right-sized governance for confidential data.

Enprivacy's Enprivacy 3.0 platform brings together Enprivacy's decades of experience in customer trust building, confidential data management, and data breach crisis resolution to offer a unique and unified platform for applying privacy-focused data analytics to your most sensitive information – your customer and strategic documents and data. It provides a control tower for managing your confidential data according to the policies and procedures you define.

The Enprivacy 3.0 platform is designed for ease of deployment in multiple environments, including on-premises platforms as well as cloud providers. The deployment model itself is purposefully simple with ready-to-deploy Open Container Initiative (OCI) images available, or replaced with cloud service provider offerings as needed.

1.1 Use Cases

The Enprivacy 3.0 platform has been proven for the below common use cases; nonetheless, the platform is highly extensible and many new use cases are supported over time.

These use cases include but are not limited to

- Retrieval Augmented Generation ("RAG") with Generative Artificial Intelligence ("AI" or "Gen AI") solutions without sharing confidential information,
- Machine Learning ("ML") using anonymised data,
- data science and statistical analysis of redacted data,
- internal training using pseudo-anonymised data,
- sharing of redacted documents to third parties, and
- other advanced use cases.

2 Enprivacy 3.0 Features

Enprivacy 3.0 unlocks immediate value of your confidential data for use in a variety of use cases, while simultaneously simplifying compliance with privacy regulations such as but not limited to the Singapore Personal Data Protection Act (“PDPA”), European Union (“EU”) General Data Protection Regulation (“GDPR”), California Consumer Privacy Act (“CCPA”), Australian Privacy Principles (“APP”) and other regional and industry-specific rules.

The platform allows non-technical end-users to upload files, run automated detection, review proposed redactions, and export compliant versions while maintaining a full audit trail.

Additionally, administrative users may add document repositories for automated monitoring, such as but not limited to Gmail inboxes, Simple Storage Service (S3)-compatible services, Microsoft Azure blob storage, network file shares, Microsoft OneDrive drives and SharePoint sites, NetDocuments repositories, and others. These document repositories may be periodically scanned for PII and CII with the option to generate redacted or anonymised documents automatically.

The Enprivacy 3.0 platform utilises a proprietary combination of optical-character recognition (“OCR”), pattern matching, Named Entity Recognition (“NER”), ML techniques, and other capabilities to generate insights into the confidential data held by structured, semi-structured, and unstructured data sets.

Underpinning the analysis is a graph or network topology which supports rapid factual and inferred analysis of the confidential data.

The features of Enprivacy 3.0 may be described under three (3) headings:

1. Manage
2. Monitor
3. Explore

2.1 Manage

Administrative users may configure the general settings of the platform, including segregation of data into workspaces, identified categories, categorisation rule sets, and redaction plans. Furthermore, administrative users may define databases and document repositories, as well as set secure credentials.

2.2 Monitor

Designated users may view reports and analytics of activities.

In future, this will include capabilities to write, publish, and link internal policies, procedures, controls, and evidence to external law and regulations, helping compliance and legal teams manage their confidential data effectively.

2.3 Explore

Designated users may access many tools designed to explore the data processed.

Using the Graph Explorer, end-users can explore the relationships between data, documents, repositories, database, etc.

Using the Documents Explorer, end-users can read redacted versions of documents as well as upload documents for analysis.

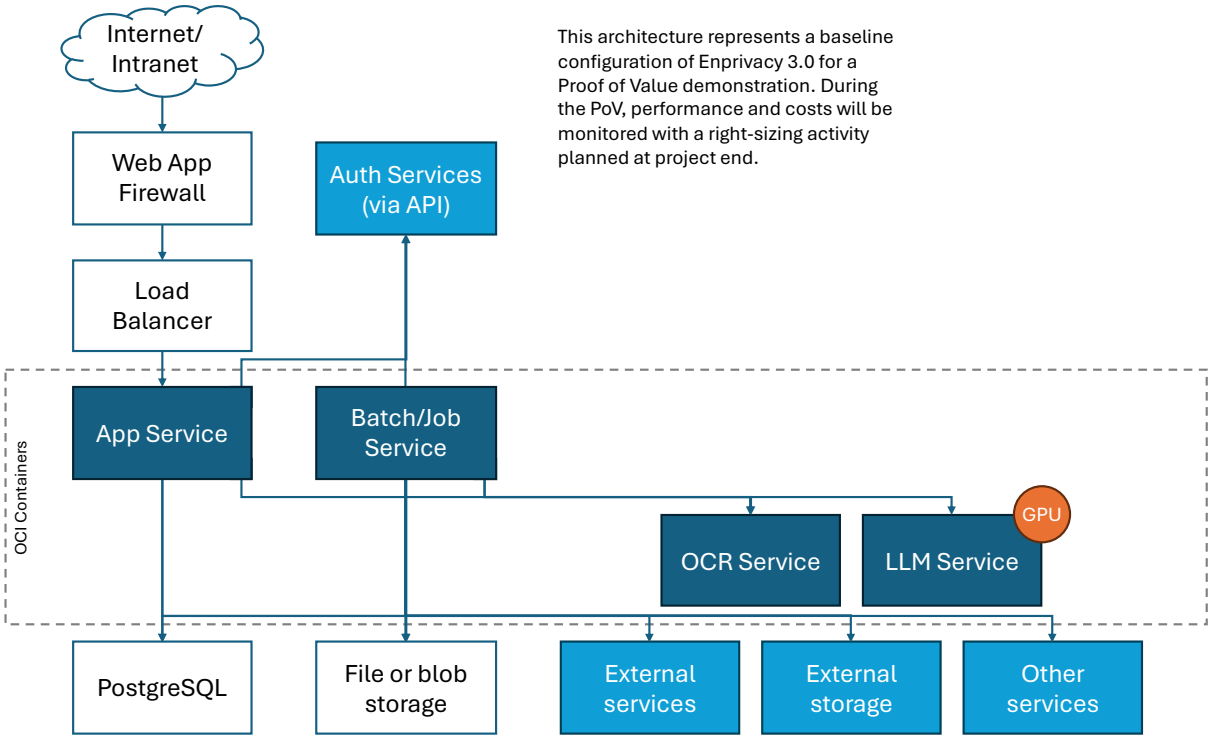
Using the Database Explorer, end-users can explore database structures.

Using the Chat feature, end-users can ‘talk’ to their data, both anonymised and original, as needed.

Using the Search feature, end-users can perform general queries across their data.

3 Overall architecture

The Enprivacy 3.0 solution is composed of four (4) services, one (1) database, and one (1) blob or file storage.



3.1 Services

All services are offered as OCI or Docker container images.

3.1.1 Web service

Serves the end-user and administrative interfaces. This service also includes public interfaces for the Job service.

The service supports horizontal scaling.

The service is stateless – all state is held within the database or the blob storage.

COMPONENT	MINIMUM	RECOMMENDED
COUNT	1	1
CPU	1 core	2 cores
RAM	2 GB	4 GB
DISK	20 GB	20 GB
GPU	Nil	Nil

3.1.1.1 Durable storage

By default, models will be downloaded to ~/.cache. It is recommended to make this path a mounted durable storage to accelerate service starts.

3.1.1.2 Allow-listing

The Enprivacy 3.0 service uses an external authentication and authorization platform for user access control.

- auth.enprivacy.com

The LLM service will download models on first use from the Hugging Face Hub. At the time of this writing, the download Content Distribution Network URLs are as follows:

- huggingface.co
- cdn-lfs.huggingface.co
- cdn-lfs-us-1.hf.co
- cdn-lfs-eu-1.hf.co
- cdn-lfs.hf.co
- cas-bridge.xethub.hf.co

Alternatively, the desired images can be loaded into the service's durable storage as a one-off action.

3.1.2 Job service

Processes background jobs and tasks. This service has no public interface.

The service supports horizontal scaling.

The service is stateless – all state is held within the database or the blob storage.

COMPONENT	MINIMUM	RECOMMENDED
COUNT	1	1
CPU	1 core	2 cores
RAM	2 GB	4 GB
DISK	20 GB	20 GB
GPU	Nil	Nil

3.1.2.1 Durable storage

By default, models will be downloaded to ~/.cache. It is recommended to make this path a mounted durable storage to accelerate service starts.

3.1.2.2 Allow-listing

The Enprivacy 3.0 service uses an external authentication and authorization platform for user access control.

- auth.enprivacy.com

The LLM service will download models on first use from the Hugging Face Hub. At the time of this writing, the download Content Distribution Network URLs are as follows:

- huggingface.co
- cdn-lfs.huggingface.co
- cdn-lfs-us-1.hf.co
- cdn-lfs-eu-1.hf.co
- cdn-lfs.hf.co
- cas-bridge.xethub.hf.co

Alternatively, the desired images can be loaded into the service's durable storage as a one-off action.

3.1.3 LLM service

Manages all inference workloads, including detection and classification processes, via vLLM and an open-source Large Language Model validated for use with Enprivacy 3.0. Generally, Enprivacy 3.0 needs one instance only.

The service is stateless; however, models are cached to local storage for performance. Enprivacy recommends attaching a durable storage for this purpose.

Given that GPUs can be expensive, the Enprivacy platform can be used with service tools such as OpenAI's ChatGPT, AWS Bedrock, Azure OpenAI, Google Vertex, and others.

The URLs to be allow-listed depend upon the service to be used. Enprivacy can work with your security teams to select and implement such hosted tools.

COMPONENT	MINIMUM	RECOMMENDED
COUNT	1	1
CPU	4 cores	8 cores
RAM	32 GB	58 GB
DISK	200 GB	200 GB
GPU	T4-equivalent	T4-equivalent

3.1.3.1 Durable storage

By default, models will be downloaded to ~/.cache. It is recommended to make this path a mounted durable storage to accelerate service starts.

3.1.3.2 Allow-listing

The LLM service will download models on first use from the Hugging Face Hub. At the time of this writing, the download Content Distribution Network URLs are as follows:

- huggingface.co
- cdn-lfs.huggingface.co
- cdn-lfs-us-1.hf.co

- cdn-lfs-eu-1.hf.co
- cdn-lfs.hf.co
- cas-bridge.xethub.hf.co

Alternatively, the desired images can be loaded into the service's durable storage as a one-off action.

3.1.4 Text Extraction (OCR) Service

Provides optical character recognition (OCR) of documents using a configured Docling open-source tool.

The service is stateless.

COMPONENT	MINIMUM	RECOMMENDED
COUNT	1	1
CPU	1 core	2 cores
RAM	2 GB	4 GB
DISK	20 GB	20 GB
GPU	Nil	Nil

3.2 Database

An always-on PostgreSQL database is critical for the Enprivacy 3.0 platform as it handles state for the Web and Job services. Additionally, it requires certain extensions to support vector and graph queries which are key to the platform.

Enprivacy recommend following any corporate standards for security related to databases.

COMPONENT	MINIMUM	RECOMMENDED
ENGINE	PostgreSQL	PostgreSQL
VERSION	16+	18+
COUNT	1	1
CPU	2 cores	4 cores
RAM	4 GB	8 GB
DISK	100 GB	200 GB
CONNECTIONS	50	100

3.2.1.1 First Set-Up

A database schema should be created.

A user with a password with all privileges over the created schema should be created.

This user will be used by the service to prepare the schema on first run, as well as any schema updates during software version upgrades. Additionally, the user will be used for all day-to-day activities of the service.

For improved security, it is recommended to perform password rotations either in a single user or dual user configuration.

3.2.1.2 Extensions

The PostgreSQL database must include the following extensions:

- hstore (https://www.postgresql.org/docs/current/hstore.html)
- PostGIS (https://postgis.net)
- pgVector (https://github.com/pgvector/pgvector)

The PostgreSQL database may include the following extensions:

- pgRouting (https://github.com/pgRouting/pgrouting; if not available, will fallback to less optimised recursive queries)

3.3 Blob storage

Blob, or file, storage is needed for holding document uploads as well as redacted documents.

Generally, this will grow over time as documents are uploaded or redacted, but there is no minimum size required.

COMPONENT	MINIMUM	RECOMMENDED
STORAGE	10 GB	10+ GB

4 Communications

4.1 Interservice communications

Services communicate with each other over the following default ports.

FROM	TO	PORT	PROTOCOL
INTERNET	Web	8080	HTTP
WEB	Database	5432	TCP
JOB	Database	5432	TCP
WEB	LLM	8000	HTTP
JOB	LLM	8000	HTTP
WEB	OCR	5001	HTTP
JOB	OCR	5001	HTTP

4.2 Public communication

As noted, only the Web service needs to be exposed for administrative and end-user access. Enprivacy recommend using HTTPS termination via load balancers or other tools as appropriate for your environment. This will require providing your own certificates for your selected domains.

However, Enprivacy 3.0 can be configured for HTTPS with certificates provisioned by Let's Encrypt. Enprivacy can work with your team to select and implement an appropriate approach.

5 Deployment Models

5.1 Recommended deployment model with containers

Enprivacy recommends deploying the Enprivacy 3.0 platform to cloud-based service providers such as AWS, Azure, Google, or Huawei, to make use of the managed service capabilities these platforms offer.

Enprivacy 3.0 is designed for ease of deployment to many environments and as such has adopted Open Container Initiative (OCI) or Docker containers. Enprivacy can provide the needed container images via GitHub Container Repository, or by pushing images to your own container repository.

Each service can then be deployed using the cloud provider's 'container application' service, or via Kubernetes.

The database can be deployed using the cloud provider's managed PostgreSQL database service, or through an assembled container with all needed extensions.

The blob storage service can likewise be deployed using the cloud provider's managed blob storage service, or through an open-source S3-compatable container (e.g. Minio or Garage).

5.2 Alternative options for LLMs

Enprivacy note that the costs of GPUs can be expensive; Enprivacy can work with your selected LLM providers in many cases to re-use existing commitments for platforms such as AWS Bedrock, Azure OpenAI, or Google Vertex.

5.3 Compliance to client requirements

The architecture is designed for flexibility; Enprivacy will work with your teams to identify and implemented any needed changes for compliance or regulatory purposes.

5.3.1 All-In-One

For demonstration purposes it is possible to operate the Web, Job, OCR, and Database services in a single environment, while durable storage may use a local disk.

The LLM service is not included in this design; an external LLM service should be used.

Please note that the Web, Job, and OCR services exhibit *ad hoc* resource utilisation and as such resource contention is possible in such configurations.

Enprivacy can provide a single Docker Compose file to enable this environment.

Enprivacy does not recommend the 'All-In-One' deployment model for production usage.

The below table provides details on the minimum and recommended computing configuration with and without the OCR service. The OCR service is required if demonstrating the Content analysis and redaction features of Enprivacy 3.0.

COMPONENT	AIO WITHOUT OCR	AIO WITH OCR
COUNT	1	1
OPERATING SYSTEM	Linux (Ubuntu 22.04+ or Debian 12), 64-bit	Linux (Ubuntu 22.04+ or Debian 12), 64-bit
DOCKER	Docker Engine 24+ with Compose v2	Docker Engine 24+ with Compose v2
CPU	4 cores	8 cores
RAM	8 GB	16 GB
DISK	150 GB	200 GB