



PLATFORM DEEP DIVE / Plugins / Plugins setup guides / ocr-plugin-setup

Contents

- PLATFORM DEEP DIVE / Plugins / Plugins setup guides / OCR plugin setup
 - Infrastructure Prerequisites:
 - Deployment/Configuration
 - Credentials
 - Kafka configuration
 - Authorization
 - Storage (S3 configuration)
 - Performance
 - Certificates
 - Workers behavior

PLATFORM DEEP DIVE / Plugins / Plugins setup guides / OCR plugin setup

The OCR plugin is a docker image that can be deployed using the following infrastructure prerequisites:

Infrastructure Prerequisites:

- S3 bucket or alternative (for example, minio)
- Kafka cluster

IMPORTANT

Starting with `ocr-plugin 1.X` it no longer requires RabbitMQ.

The following environment from previous releases must be removed in order to use OCR plugin: `CELERY_BROKER_URL`.

Deployment/Configuration

To deploy the OCR plugin, you will need to deploy `ocr-plugin` helm chart with custom values file.

Most important sections are these, but more can be extracted from helm chart.

```
image:
  repository: <repository>/ocr-plugin

applicationSecrets: {}

replicaCount: 2

resources: {}

env: []
```

Credentials

S3 bucket:

```
applicationSecrets:
  enable: true
  envSecretKeyRef:
    STORAGE_S3_ACCESS_KEY: access-key # default empty
```

```
STORAGE_S3_SECRET_KEY: secret-key # default empty
existingSecret: true
secretName: ocr-plugin-application-config
```

Kafka configuration

You can override the following environment variables:

Environment Variable	Definition	Default Value
ENABLE_KAFKA_SASL	Indicates whether Kafka SASL authentication is enabled	False
KAFKA_ADDRESS	The address of the Kafka bootstrap server in the format <hostname>: <port>	-

Environment Variable	Definition	Default Value
KAFKA_CONSUME_SCHEDULE	The interval (in seconds) at which Kafka messages are consumed	30
KAFKA_INPUT_TOPIC	The Kafka topic from which input messages are consumed	-
KAFKA_OCR_CONSUMER_GROUPID	The consumer group ID for the OCR Kafka consumer	ocr_group
KAFKA_CONSUMER_AUTO_COMMIT	Determines whether Kafka consumer commits offsets automatically	True

Environment Variable	Definition	Default Value
KAFKA_CONSUMER_AUTO_COMMIT_INTERVAL	The interval (in milliseconds) at which Kafka consumer commits offsets automatically	1000
KAFKA_CONSUMER_TIMEOUT	The timeout (in milliseconds) for Kafka consumer operations	28000
KAFKA_CONSUMER_MAX_POLL_INTERVAL	The maximum interval (in milliseconds) between consecutive polls for Kafka consume	25000

Environment Variable	Definition	Default Value
KAFKA_CONSUMER_AUTO_OFFSET_RESET	The strategy for resetting the offset when no initial offset is available or if the current offset is invalid	earliest
KAFKA_OUTPUT_TOPIC	The Kafka topic to which output messages are sent	-

Please note that the default values and examples provided here are for illustrative purposes. Make sure to replace them with the appropriate values based on your Kafka configuration.

CAUTION

When configuring the OCR plugin, make sure to use the correct outgoing topic names that match **the pattern expected by the Engine**, which listens for messages on topics with specific names.

Authorization

You can override the following environment variables:

Environment Variable	Definition	Default Value	Example
<code>OAuth_CLIENT_ID</code>	The client ID for OAuth authentication	-	<code>your_client_id</code>
<code>OAuth_CLIENT_SECRET</code>	The client secret for OAuth authentication	-	<code>your_client_secret</code>
<code>OAuth_TOKEN_ENDPOINT_URI</code>	The URI of the token endpoint for OAuth authentication	-	<code>https://oauth</code>

Please note that the default values and examples provided here are for illustrative purposes. Make sure to replace them with the appropriate values based on your OAuth authentication configuration.

Storage (S3 configuration)

You can override the following environment variables:

Environment Variable	Definition	Default Value	
STORAGE_S3_HOST	The host address of the S3 storage service	-	minio: west-1
STORAGE_S3_SECURE_CONNECTION	Indicates whether to use a secure connection (HTTPS) for S3 storage	False	
STORAGE_S3_LOCATION	The location of the S3 storage service	-	eu-wes

Environment Variable	Definition	Default Value	
STORAGE_S3_OCR_SCANS_BUCKET	The name of the S3 bucket for storing OCR scans	-	pdf-sc
STORAGE_S3_OCR_SIGNATURE_BUCKET	The name of the S3 bucket for storing OCR signatures	-	extrac
STORAGE_S3_OCR_SIGNATURE_FILENAME	The filename pattern for extracted OCR signatures	-	extrac

Environment Variable	Definition	Default Value	
<code>STORAGE_S3_SECRET_KEY</code>	The secret key for connecting to the S3 storage service	-	

Please note that the default values and examples provided here are for illustrative purposes. Make sure to replace them with the appropriate values based on your S3 storage configuration.

Performance

Environment Variable	Definition	Default Value
<code>ENABLE_PERFORMANCE_PAYLOAD</code>	When set to true, the response payload will contain performance metrics related to various stages of the process.	<code>true</code>

Example



```
"perf": {
  "total_time": 998,
  "split": {
    "get_file": 248,
    "extract_images": 172,
    "extract_barcodes": 37,
    "extract_signatures": 238,
    "minio_signature_save": 301
  }
}
```

Certificates

You can override the following environment variables:

- `REQUESTS_CA_BUNDLE` - the path to the certificate bundle file used for secure requests

Workers behavior

You can override the following environment variables:

Environment Variable	Definition	Default Value
<code>OCR_WORKER_COUNT</code>	Number of workers	5

Environment Variable	Definition	Default Value
OCR_WORK_QUEUE_TIMEOUT	If no activity has occurred for a certain number of seconds, an attempt will be made to refresh the workers	10

! INFO

If no worker is released after `OCR_WORK_QUEUE_TIMEOUT` seconds, the application will verify whether any workers have become unresponsive and need to be restarted.

If none of the workers have died, it means they are likely blocked in some process. In this case, the application will terminate all the workers and shut down itself, hoping that the container will be restarted.

Was this page helpful?