


```
{
  "id": "/RuLr/[-.docx",
  "name": "[-.docx",
  "bucket": "webdav"
}
```

□□

WebDAV □□□□ Dify□

```
“ □□□□
```

□□□□ Dify□

```
“ □□□ docx □□□□□□□□□□
```

□□ pipeline □□□□

```
online_drive
  ↓
download
  ↓
text_model
  ↓
general_chunker
  ↓
1 chunk
```

????? file_id

□□ documents □□□□

file_id | text

□□□□□□□□

☐☐☐

```
select
id,
file_id,
name
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

☐☐☐☐

??1?????

☐☐☐

```
file_id | NULL
```

☐☐☐☐☐

WebDAV☐☐☐☐☐☐☐☐☐☐

Dify storage☐

??2

☐ file_id☐

☐☐☐

```
file_id
-----
xxxx-xxxx
```

☐☐☐☐☐

```
select *
from upload_files
where id='file_id';
```

??? segments

■■■■■■■

12■■■■■■■

```
select
  id,
  position,
  content
from document_segments
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

■■■

■■■

```
position | content
-----+-----
1        | 1-■■■■■■■■■■■■■■■■■■.docx
```

■■■■■

100%■■■

????

■■■■

```
Dify 1.15.0
plugin_daemon 0.6.3
shaba/webdav
```

■■■ Dify ■■■

■■■

■■■

```
max_chunk_length=800
```

■■■

```
max_chunk_length=1024
```

■■■

max_chunk_length=1024

□□□□

1 paragraph

□□

chunker□□□□□□□□□□ 12□□

????

??A???????? shaba/webdav ??

□□□□

□□

shaba/webdav

□□

online_drive_download_file

□□□□

□□□□□□□□

```
{
  "type": "file",
  "url": "...
}
```

□□ Dify datasource plugin □□□□□□

□□□□□ + □□ + mime□

□□□□

filename:

[-].docx

mime:

application/vnd.openxmlformats-officedocument.wordprocessingml.document

??B??? datasource??????

WebDAV

WebDAV

↓

rclone sync

↓

↓

Dify API

upload_file

docx parser

chunk

??C????????

plugin

```
docker exec -it docker-plugin_daemon-1 bash
```

```
ls /app/plugins
```

shaba/webdav

□□□

□□□□□□

□□□□□□□□

1.

```

select
id,
file_id,
name
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';

```

2.

```

select
position,
content
from document_segments
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';

```

□□□□□□□□□□

□□□□□□

□□ □□□□□□

□□□□□□□□□□

□□□□□□□□□□□□□□□□

□□ **WebDAV** □□□□□□

□□ `upload_files` □□□□□

```

□-□□□□□□□□□□□□□□□□.docx
extension = docx
mime_type = application/vnd.openxmlformats-officedocument.wordprocessingml.document

```

□□□

□ WebDAV □□□□□□

□ □□□□□□□

□ upload_files □□□□□□

□□□

```
documents.file_id = NULL
```

□□□

□□□□□□ **storage**□□□□ **document** □□□□□ **file_id**□

□□□□□

```
doc_type = NULL
doc_metadata = NULL
word_count = 12
1 paragraph
```

□□□□□

??????????

Dify 1.15 □□□□□□

```
documents.file_id
    ↓
upload_files.id
```

□□□

□□□□

```
documents
  id:
  4e9e360a-5ab3-4957-8de9-75b76e173bdb

  file_id:
  □
```

□□ upload_files □□

```
9dc4b86e-7f4c-4804-a5a8-f81742148ddc
```

■■■■■■■■

???? upload_file

■■■

```
select
  id,
  name,
  extension,
  created_at
from upload_files
where name='[-]■■■■■■■■■■■■■■■■■■■■.docx'
order by created_at desc;
```

■■■■■■■■

■■■■■

■■ document

```
created_at:
2026-07-09 07:31:48
```

■■ upload_files

■■■

```
2026-07-09 02:40:27
```

■■■■■■■■

PostgreSQL UTC ■■■■■ +8

■■■■■

```
9b2f7c06-e091-4332-a3e9-83c84af21bb8
```

■■■

```
b69aa5e5-f81d-447f-a13f-8b106c03f5f4
```

????

□□

```
select
d.id as document_id,
d.name,
d.file_id,
u.id as upload_id,
u.name as upload_name
from documents d
left join upload_files u
on d.file_id=u.id
where d.name='[-\s\.\*\?]{1,100}.docx';
```

□□□□□□

document_id	file_id	upload_id
4e9e360a...	NULL	NULL

????????

□□

□ WebDAV□□□□□□

□□

□ WebDAV datasource pipeline □□ document □□□□ upload_file_id □□ documents.file_id

????????????

□□

```
upload_files
  |
  |
documents.file_id
  |
docx parser
```

WebDAV

```
upload_files
  |
  X
documents.file_id=NULL
  |
text_model
```

????

??1?? shaba/webdav ???????

□□□□□

```
online_drive_download_file
```

□□□□□

□□□

```
{
  "file_id": "9dc4b86e-7f4c-4804-a5a8-f81742148ddc"
}
```

□□□□□□□

```
{
  "id": "/Ru1r□□/xxx.docx",
  "name": "xxx.docx",
  "bucket": "webdav"
```

```
}
```

??2???????????

□□□□□□

```
update documents
set file_id='□□upload_files.id'
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□□□□

□□□□□□

□□ 100%□□□□□□

△ □□□□□□□□

???????????

Dify 1.15 datasource □□□□□□□□

□□□

```
select *
from segment_attachment_bindings
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□□□□□□

□□□□□

```
select
id,
name,
created_at
from upload_files
where name='□-□□□□□□□□□□□□.docx'
```

```
order by created_at desc;
```

```
|||||
```

```
||||| document ||| upload_file_id|||||
```

```
|||||
```

```
ERROR: operator does not exist: text = uuid  
LINE 9: on d.file_id=u.id
```

```
|||
```

```
documents.file_id ||| = text  
upload_files.id ||| = uuid
```

```
|| PostgreSQL |||
```

```
|||
```

```
select  
d.id as document_id,  
d.name,  
d.file_id,  
u.id as upload_id,  
u.name as upload_name  
from documents d  
left join upload_files u  
on d.file_id = u.id::text  
where d.name='|-|||||.docx';
```

```
|||||
```

```
documents.file_id = |
```

```
|||||
```

document_id	file_id	upload_id
4e9e360a-5ab3-4957-8de9-75b76e173bdb	NULL	NULL

???????????? upload_file

WebDAV 5

2026-07-09 01:40:07
2026-07-09 02:20:31
2026-07-09 02:27:43
2026-07-09 02:40:27
2026-07-09 02:49:38

document

documents.created_at
2026-07-09 07:31:48

+8

Postgres:

2026-07-09 07:31:48

2026-07-09 15:31:48

upload_files

2026-07-09 02:49:38 UTC

2026-07-09 10:49:38

pipeline

document_pipeline_execution_logs
created_at
2026-07-09 07:31:48

upload_file

f20a94e5-19b2-4219-b991-5abbc3ea0627

□□□□□□□□

?? upload_file ??

□□

```
select
id,
name,
key,
size,
extension,
mime_type
from upload_files
where id='f20a94e5-19b2-4219-b991-5abbc3ea0627';
```

□□□□

```
key
bucket
size
```

□□□□□□

????????????

□□□□

```
create table documents_backup as
select * from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□

```
update documents
set file_id='f20a94e5-19b2-4219-b991-5abbc3ea0627'
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□□□□

- □□□□□
- □□□□

□□□

```
word_count
□ 12
□□□□

paragraphs
□ 1
□□□/□□
```

□□□□□□□□

□□□□□□□□□□

□□□□□□

```
POST plugin_daemon dispatch/tool/invoke
Tool: general_chunker

Outputs:
{
  "variable_value":["online_drive"]
}
```

□□□□□□

□□□□□

```
docx parser
  ↓
text
  ↓
general_chunker
```

□□□□

```
online_drive
  ↓
general_chunker
```

□□□□ **WebDAV datasource** □□□□□□□□ □

□□□□□□□□□□

shaba/webdav □□□□□□□□

□□□□□□

```
select
id,
name,
key,
size,
extension,
mime_type
from upload_files
where id='f20a94e5-19b2-4219-b991-5abbc3ea0627';
```

□□□

\d upload_files

□□□□ Dify 1.15 □ upload_files □□□□□□□□□□□□□□

□□□□□□□□□□

□□□□□□□□

```
upload_files
  |
  | □□
  ↓
f20a94e5-19b2-4219-b991-5abbc3ea0627
□-□□□□□□□□□□□□□□.docx
size=2847874
mime_type=docx
```

key=upload_files/...docx

documents



4e9e360a-5ab3-4957-8de9-75b76e173bdb

file_id=NULL

[][]

[][][][] **document** [][][][]

[][][][]

word_count = 12

paragraph = 1

doc_type = NULL

[] Dify [] docx parser[][][] datasource metadata [][][][]

??????????

[][][][][][][][][][]

1. ???? document

```
create table documents_backup_20260710 as
select *
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

2. ?? file_id

[][]

```
update documents
set file_id='f20a94e5-19b2-4219-b991-5abbc3ea0627'
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□

```
select
id,
name,
file_id
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□

id	file_id
4e9e360a...	f20a94e5...

????

□□□ file_id □□□□ □

□□□□ document □□□□□□ segment□

□□□

```
select
count(*)
from document_segments
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□ >0□

□□□□□□□□

```
delete from document_segments
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□□□□□□

??? WebDAV ??????????

□□□□

```
POST /dispatch/datasource/online_drive_download_file
HTTP 200
```

□□□□□□□□

□□□□

```
general_chunker
Outputs:
["online_drive"]
```

□□□□ pipeline □□□□

```
online_drive□□
```

□□□

```
docx□□
```

□□□□□□

```
WebDAV
|
download_file
|
upload_files
|
file_id
|
document.file_id
|
docx parser
|
chunk
|
```

embedding

□□□□

```
WebDAV
|
download_file
|
upload_files  []
|
document      [] file_id[]
|
general_chunker
|
1□
```

??????

□□□□

shaba/webdav

□□□

□□□□

online_drive_download_file

□□□□□

□□□□□□□□

```
{
  "file_id": "f20a94e5-19b2-4219-b991-5abbc3ea0627"
}
```

□□□□□

document □□□□□

?????

□□□□

```
update documents
set file_id='f20a94e5-19b2-4219-b991-5abbc3ea0627'
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□ Dify □□□

```
□□□
→ □□
→ □□□□
```

□□

□□□□

□□□□

```
doc_type: docx
word_count: xxxx
□□□□: xx
```

□□□□□□□□ 100% □ WebDAV □□ file_id □□ bug□

□□□□□□□□□□□□□□ Dify □□

□□□□□□□□□□□□□□

□□□□

□ documents.file_id □□□□

```
4e9e360a-5ab3-4957-8de9-75b76e173bdb
|
|
v
f20a94e5-19b2-4219-b991-5abbc3ea0627
```

□ □□□□□ segment □□□□

```
document_segments = 0
```

```
██████████
```

```
segment██████
```

```
document ██████
```

```
Dify ██████████
```

1. ????? document ??

```
███
```

```
select
  id,
  word_count,
  tokens,
  doc_type,
  indexing_status,
  parsing_completed_at,
  splitting_completed_at,
  completed_at,
  error
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

```
████
```

```
indexing_status
splitting_completed_at
completed_at
```

2. ????????

```
██████
```

```
indexing_status = completed
splitting_completed_at ██
completed_at ██
```

```
████
```

□□□

```
update documents
set
word_count=null,
tokens=null,
parsing_completed_at=null,
cleaning_completed_at=null,
splitting_completed_at=null,
completed_at=null,
indexing_status='waiting',
error=null
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

3. ???????

□□□□ UI□

```
□□□
→ □□
→ □-□□□□□□□□□□□□□□.docx
→ □□□□
```

□□ worker□

```
docker logs -f docker-worker-1
```

□□□□□□

□□□

```
parse document
loader: docx
extract text
split document
generate segments
embedding xxx texts
```

□□□□

```
general_chunker
online_drive
1 texts
```

4. ?????

□□□□□

documents

```
select
name,
word_count,
doc_type,
tokens
from documents
where id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□

□□□

name	word_count	doc_type	tokens

□-□□□□□□□□□□□□□□.docx	5000	docx	xxxx

segments

```
select
count(*)
from document_segments
where document_id='4e9e360a-5ab3-4957-8de9-75b76e173bdb';
```

□□□

□□ / □□

