

文档转 PDF（华为）概述：

3 种转换方式：

单一文档转为 PDF，文档是一个下载链接，用 HTTP GET 方式，见：[文档转换 GET](#)

单一文档转为 PDF，文档 POST 到服务器，用 HTTP POST 方式，见：[文档转换 POST](#)

多个图片转为 PDF，文档是多个下载链接，用 HTTP POST 方式，见：[多张图片转换 POST](#)

基本用法：

由于转换需要时间，文件越大页数越多，转换越久，故系统采用**异步**的方式获得转换结果。

调用转换接口后会获得 token，随后有 2 种方式查询转换结果：

1. 用 HTTP GET 方式，轮询“查询 query 接口”获得结果。详细见：[查询 QUERY](#)
2. 设置 callbackurl，当转换结束后，系统会回调该 URL 直接推送转换结果。详细见：[回调 URL](#)

API 调用需要签名，详细见[华为官方签名](#)：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

文档转换 GET

将文档下载地址 url 转换为 PDF，type 是源文档的 type，比如要把 docx 转为 pdf，type 就是 docx

请求参数：

参数	类型及范围	备注	是否必须发送
url	string	文件 url，支持 http(s)，ftp 开头，需要 URL Encoding	是
type	string	要转换的文档扩展名	是
linearization	int	是否需要快速 web 显示，默认 0 否，1 是	否
compress	int	是否压缩，默认 0 不压缩，压缩分 3 档：1，2，3，1 的压缩率最低，3 最高	否
excelislandscape	int	如果是 Excel 文件，是否横屏，默认 0 否（竖屏），1 是	否
exceliscenter	int	如果是 Excel 文件，是否横竖居中，默认 0 否，1 是	否
excelmargin	int	如果是 Excel 文件，四边的边距，默认 10px，单位是像素	否
excelsheetindex	int	如果是 Excel 文件，指定转换的	否

参数	类型及范围	备注	是否必须发送
		Sheet 索引，默认 0：全部，第一个 sheet 就是 1，以此类推	
excelnotshowgridlines	int	如果是 Excel 文件，不显示网格线，默认 0 显示，1 不显示	否
exceluseprintarea	int	如果是 Excel 文件，是否使用打印区域，默认 0 不使用，1 使用	否
wordshowmarkup	int	如果是 Word 文件，是否显示审阅标记，默认 0 不显示，1 显示	否
imageocr	int	如果是图片文件，是否识别图中文字并且在 PDF 中可选可搜索文字，默认 0 否，1 是	否
imagedeskew	int	如果是图片文件，是否将斜的文字矫正，默认 0 否，1 是	否
imageclean	int	如果是图片文件，是否清除图像背景只显示文字，默认 0 否，1 是	否
userpassword	string	生成 PDF 文件的 User Password，默认无	否
ownerpassword	string	生成 PDF 文件的 Owner Password，默认无	否
pdfrestriction	string	有密码时，生成 PDF 文件的权限，默认无权限。格式 3 位，比如 010，第一位代表能否打印，第二位代表能否拷贝内容，第三位代表能否编辑	否
watermark	string	添加水印，字符个数最大 10 个	否
pdffilename	string	生成的 PDF 文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，需要 URL Encoding，详细见回调 URL	否

请求示例：

`http://all2pdf.apistore.huaweicloud.com/v1/convert?url=https%3a%2f%2fxxx%2fxxx.docx&type=docx`

将所在 url 地址的 docx 文件转为 PDF，type 就是源文件的 type，这个例子里就是 docx

必须签名才能调用成功，签名见华为签名规则：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

支持多种文件格式，具体如下（type 可传入如下格式）：

微软 Office 文档：doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx

WPS 文档：wps, wpt, dps, dpt, et, ett

苹果 iWork 文档：pages, key, numbers

开放版式文档：ofd

电子刊物：caj

电子书：epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu

Markdown 格式：md

SVG 格式：svg

CAD 文档：dwg, dxf, dwt, dws

Sketch 文档：sketch

网页文件：html, htm, mht, eml

图片文件：几乎所有格式例如 png, jpg, jpeg, gif, tif, tiff, bmp，**可以统一传 img，代表一切图片**

文本文件：txt, rtf, java, js, c, cpp, jsp, css, xml, properties, log，其他任意文本文件都可以传 txt

网址网页：url（例如：<http://www.baidu.com>，不支持文件下载链接）

PDF 文件：pdf

返回数据结构：

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "result":{"token":"xxx"}
}
```

返回示例(失败状态):

```
{
  "code":40001,
  "msg":"ParmNotRight"
}
```

文档转换 POST

直接将单个文档 POST 到服务器，大小限制 12M

请求参数:

参数	类型及范围	备注	是否必须发送
file	file	要转换的文档，Content-Type 使用 multipart/form-data，最大 12M	是
type	string	要转换的文档扩展名	是
linearization	int	是否需要快速 web 显示，默认 0 否，1 是	否
compress	int	是否压缩，默认 0 不压缩，压缩分 3 档：1, 2, 3, 1 的压缩率最低，3 最高	否
excelislandscape	int	如果是 Excel 文件，是否横屏，默认 0 否（竖屏），1 是	否
exceliscenter	int	如果是 Excel 文件，是否横竖居中，默认 0 否，1 是	否

参数	类型及范围	备注	是否必须发送
excelmargin	int	如果是 Excel 文件，四边的边距，默认 10px，单位是像素	否
excelsheetindex	int	如果是 Excel 文件，指定转换的 Sheet 索引，默认 0：全部，第一个 sheet 就是 1，以此类推	否
excelnotshowgridlines	int	如果是 Excel 文件，不显示网格线，默认 0 显示，1 不显示	否
exceluseprintarea	int	如果是 Excel 文件，是否使用打印区域，默认 0 不使用，1 使用	否
wordshowmarkup	int	如果是 Word 文件，是否显示审阅标记，默认 0 不显示，1 显示	否
imageocr	int	如果是图片文件，是否识别图中文字并且在 PDF 中可选可搜索文字，默认 0 否，1 是	否
imagedeskew	int	如果是图片文件，是否将斜的文字矫正，默认 0 否，1 是	否
imageclean	int	如果是图片文件，是否清除图像背景只显示文字，默认 0 否，1 是	否
userpassword	string	生成 PDF 文件的 User Password，默认无	否
ownerpassword	string	生成 PDF 文件的 Owner Password，默认无	否
pdfrestriction	string	有密码时，生成 PDF 文件的权限，默认无权限。格式 3 位，比如 010，第一位代表能否打印，第二位代表能否拷贝内容，第三位代表能否编辑	否
watermark	string	添加水印，字符个数最大 10 个	否
pdffilename	string	生成的 PDF 文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，详细见 回调 URL	否

请求示例:

`http://all2pdf.apistore.huaweicloud.com/v1/convert_post`
Header 中的 Content-Type 必须是 multipart/form-data

必须签名才能调用成功, 签名见华为签名规则:

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

支持多种文件格式, 具体如下 (type 可传入如下格式):

微软 Office 文档: doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx

WPS 文档: wps, wpt, dps, dpt, et, ett

苹果 iWork 文档: pages, key, numbers

开放版式文档: ofd

电子刊物: caj

电子书: epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu

Markdown 格式: md

SVG 格式: svg

CAD 文档: dwg, dxf, dwt, dws

Sketch 文档: sketch

网页文件: html, htm, mht, eml

图片文件: 几乎所有格式例如 png, jpg, jpeg, gif, tif, tiff, bmp, 可以统一传 img, 代表一切图片

文本文件: txt, rtf, java, js, c, cpp, jsp, css, xml, properties, log, 其他任意文本文件都可以传 txt

网址网页: url (例如: http://www.baidu.com, 不支持文件下载链接)

PDF 文件: pdf

返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "result":{"token":"xxx"}
}
```

返回示例(失败状态):

```
{
  "code":40001,
  "msg":"ParmNotRight"
}
```

多张图片转换 POST

将多张图片转换为 pdf, url 传入多张图片的地址, 传入非图片格式无效

Header 中的 Content-Type 传入 application/json

POST Body 传入 JSON, 格式如下:

请求参数 BODY:

参数	类型及范围	备注	是否必须发送
url	string 数组	要转换的图片 URL 数组, 支持 http(s), ftp 开头	是
type	string	固定传 img	是
linearization	int	是否需要快速 web 显示, 默认 0 否, 1 是	否

参数	类型及范围	备注	是否必须发送
compress	int	是否压缩，默认 0 不压缩，压缩分 3 档：1，2，3，1 的压缩率最低，3 最高	否
ocr	int	如果是图片文件，是否识别图中文字并且在 PDF 中可选可搜索文字，默认 0 否，1 是	否
deskew	int	如果是图片文件，是否将斜的文字矫正，默认 0 否，1 是	否
clean	int	如果是图片文件，是否清除图像背景只显示文字，默认 0 否，1 是	否
userpassword	string	生成 PDF 文件的 User Password，默认无	否
ownerpassword	string	生成 PDF 文件的 Owner Password，默认无	否
pdfrestriction	string	有密码时，生成 PDF 文件的权限，默认无权限。格式 3 位，比如 010，第一位代表能否打印，第二位代表能否拷贝内容，第三位代表能否编辑	否
watermark	string	添加水印，字符个数最大 10 个	否
pdffilename	string	生成的 PDF 文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，详细见回调 URL	否

请求示例：

`http://all2pdf.apistore.huaweicloud.com/v1/convert`

传入的 Body 为 JSON 格式，如下：

```
{ "url": [ "http://xxx/xxx1.png", "http://xxx/xxx2.png" ], "type": "png" }
```

必须签名才能调用成功，签名见华为签名规则：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

例如要把多张图片 OCR，变为文字可选的 PDF，并且将斜的文字矫正，将图片背景去除，那么 JSON 就是：

```
{ "url": [ "http://xxx/xxx1.png", "http://xxx/xxx2.png" ], "type": "img", "ocr": 1, "deskew": 1, "clean": 1 }
```


支持几乎所有图片格式

返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "result":{"token":"xxx"}
}
```

返回示例(失败状态):

```
{
  "code":40001,
  "msg":"ParmNotRight"
}
```

文件操作 POST

支持以下操作:

- ◆ 文件合并
- ◆ 文件拆分
- ◆ 提取图片
- ◆ 图片签名
- ◆ 文件解密 (文件加密转换接口中已有)
- ◆ 页面删除
- ◆ 页面插入
- ◆ 页面提取
- ◆ 页面旋转
- ◆ 页面加页码

POST Body 传入 JSON, Header 中的 Content-Type 传入 application/json

通过指定 action，对 PDF 文件进行不同的操作，详细如下：

◆ **文件合并：**

action 为 merge，url 数组里放 2 个以上的 PDF 文件地址，例如：

```
{"action": "merge", "url": ["https://xxx/xxx.pdf", "https://xxx/xxx.pdf"]}
```

会合并为一个 PDF 文件

◆ **文件拆分：**

action 为 split，url 数组里放 1 个 PDF 文件地址，splitcount 是拆分的页数，例如：

```
{"action": "split", "url": ["https://xxx/xxx.pdf"], "splitcount": 2}
```

就是把 PDF 拆分为多个 PDF，每个 PDF 是 2 页

◆ **提取图片：**

action 为 extracting，url 数组里放 1 个 PDF 文件地址，例如：

```
{"action": "extractimg", "url": ["https://xxx/xxx.pdf"]}
```

就是把 PDF 里的图片提取出来

◆ **图片签名**

action 为 stampsignature，url 数组里放 1 个 PDF 文件地址，pageindex 是添加签名的页码，第一页是 1，stampurl 是签名图片，stamprect 是 4 个数字逗号分割的字符串，代表左上角为原点的签名的 x, y, width, height。例如：

```
{"action": "stampsignature", "url": ["https://xxx/xxx.pdf"], "pageindex": 1, "stampurl": "http://xxx/xxx.png", "stamprect": "10, 10, 100, 50"}
```

就是把 PDF 的第一页加上图片签名，签名位置在 x=10, y=10, 宽度=100, 高度=50, 单位像素。

◆ **文件解密（文件加密转换接口中已有）**

action 为 decrypt，url 数组里放 1 个 PDF 文件地址，password 是密码，如果没密码也可以空。例如：

```
{"action": "decrypt", "url": ["https://xxx/xxx.pdf"], "password": ""}
```

就是把 PDF 解密

◆ **页面删除**

action 为 pageremove，url 数组里放 1 个 PDF 文件地址，pageindexes 是逗号分割的页码字符串。例如：

```
{"action": "pageremove", "url": ["https://xxx/xxx.pdf"], "pageindexes": "1, 3"}
```

就是把 PDF 里的第一页，第三页删除

◆ **页面插入**

action 为 pageinsert，url 数组里放 1 个 PDF 文件地址，pageindex 是页码，指插在哪一页后面，例如 1 就是插在最后一页后面，0 就是插在最后一页前面，pageinserturl 是插入的 PDF 文件地址。例如：

```
{"action": "pageinsert", "url": ["https://xxx/xxx.pdf"], "pageindex": 1, "pageinserturl": "https://xxx/xxx.pdf"}
```

就是在 PDF 的最后一页后面，插入 pageinserturl 这个地址的 PDF

◆ **页面提取**

action 为 pageextract，url 数组里放 1 个 PDF 文件地址，pageindexes 是逗号分割的页码字符串。例如：

```
{"action": "pageextract", "url": ["https://xxx/xxx.pdf"], "pageindexes": "1, 3"}
```

就是把 PDF 里的第一页，第三页提取出来作为一个 PDF 文件

◆ 页面旋转

action 为 pagerotate, url 数组里放 1 个 PDF 文件地址, pagerotateangle 是旋转角度, 正时针, 取值 90, 180, 270。例如:

```
{"action": "pagerotate", "url": ["https://xxx/xxx.pdf"], "pagerotateangle": 90}
```

就是把 PDF 每一页正时针转 90 度

◆ 页面加页码

action 为 pageaddnumber, url 数组里放 1 个 PDF 文件地址, 例如:

```
{"action": "pageaddnumber", "url": ["https://xxx/xxx.pdf"]}
```

就是把 PDF 每一页右下角添加页码

总的请求参数 BODY (JSON) :

参数	类型及范围	备注	是否必须发送
action	string	action 可取值 merge, split, extracting, stampsignature, decrypt, pageremove, pageinsert, pageextract, pagerotate, pageaddnumber, 详细功能见上	是
url	string 数组	要操作的 PDF URL 数组, 支持 http(s), ftp 开头	是
splitcount	int	文件拆分用, 拆分的页数	否
pageindex	int	单个页码, 是否受 action 支持见上方详细操作	否
pageindexes	string	多个页码, 逗号分割, 例如: 1,3, 是否受 action 支持见上方详细操作	否
stampurl	string	图片签名用, 图片地址, 支持 http(s), ftp 开头	否
stamprect	string	图片签名用, 4 个数字逗号分割的字符串, 代表左上角为原点的签名的 x, y, width, height	否
pageinserturl	string	添加水印, 字符个数最大 10 个	否
pagerotateangle	int	页面旋转用, 角度取值 90, 180, 270, 正时针	否

参数	类型及范围	备注	是否必须发送
password	string	文件解密用，解密密码，可以空	否
pdffilename	string	生成的 PDF 文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，详见 回调 URL	否

请求示例：

`http://all2pdf.apistore.huaweicloud.com/v1/pdfaction`
 传入的 Body 为 JSON 格式，见上方详细操作。

必须签名才能调用成功，签名见华为签名规则：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

所有 action 都支持 callbackurl，文件拆分和提取图片不支持 pdffilename 参数，结果是文件名 1 开始的 pdf 或 png 文件，即 **query 返回的 result 里会变为 pdfurls 和 imageurls**。其余 action 均支持 pdffilename

返回数据结构：

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000:请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

返回示例(成功状态)：

```
{
  "code":10000,
  "msg":"",
  "result":{"token":"xxx"}
}
```

返回示例(失败状态):

```
{
  "code":40001,
  "msg":"ParmNotRight"
}
```

查询 QUERY

请求参数:

参数	类型及范围	备注	是否必须发送
token	string		是

请求示例:

<http://all2pdf.apistore.huaweicloud.com/v1/query?token=7b799e09e0838919d3ae63d0566683a2>

无需签名，无调用次数限制

由于转换需要时间，文件越大页数越多，转换越久，故需要**轮询**查询接口来获得结果。查询频率可以是1s一次，也可以更长一些。查询后先看 status，如果是 Done 或 Failed，则转换结束，停止轮询。如果是 Doing 或 Pending，则继续轮询。

返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000:请求成功
msg		string	是	
token		string	是	请求的 token
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
status	状态	string	是	Pending: 还未开始 Doing: 正在转换 Done: 转换成功 Failed: 转换失败
progress	进度	number (0.00 - 1.00)	否 (status 为 Doing 时返回)	比如 0.88 表示 88%

名称	含义	类型及范围	是否必须返回	备注
pdfurl	pdf 文件地址	string	否 (status 为 Done 时返回)	转换出来的 PDF 地址, http 和 https 都支持
count	总页数	integer	否 (status 为 Done 时返回)	PDF 页面总数
reason	失败原因	string	否 (status 为 Failed 时可能返回)	转换失败的原因

返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "token": "xxx",
  "result":
  {
    "progress":0.02,
    "status": "Doing"
  }
}
```

```
{
  "code":10000,
  "msg": "",
  "token": "xxx",
  "result":
  {
    "status": "Done",
    "pdfurl": "https://file.duhuitech.com/o/7b799e09e0838919d3ae63d0566683a2/cc9c7f1e-03f8-4742-bc37-aab9da191c26.pdf",
    "count":10
  }
}
```

返回示例(失败状态):

```
{
  "code":40000,
  "msg": "No such token"
}
```


注意:

- 上传文件大小 **不能超过 1000M**。
- 转换完成后, 在 **1 小时** 内下载文件。

回调 URL:

用途: 客户可以自行部署服务器, 系统转换结束后会调用客户提供的回调 URL, 直接发送转换结果, 从而无需再轮询 Query。

当设置了回调 URL, 转换结束后 (无论成功失败), 系统都会尝试调用该 URL, 具体如下:

以 POST 方式调用该 URL, Header 头中 Content-Type: application/json

Body 为 JSON 格式, 内容和 Query 的结果相同, 例如:

```
{"code":10000,"msg":"","token":"xxx","result":{"status":"Done","pdfurl":"https://file.duhuitech.com/o/7b799e09e0838919d3ae63d0566683a2/cc9c7f1e-03f8-4742-bc37-aab9da191c26.pdf","count":10}}
```

服务端收到该 POST 后需在 10 秒内返回 HTTP STATUS CODE 200, 视为调用成功, 否则系统认为回调失败, 会再次尝试。规则如下:

系统共计最多会调用 3 次回调 URL, 如果第一次失败, 则等待 3 秒后尝试第二次, 如果第二次失败, 则等待 5 秒后尝试第三次, 如果第三次失败, 则不再尝试。

回调 URL 超时时间 10 秒。

错误码表:

JSON 里返回的 code	错误信息
40000	通用错误
40001	参数错误
40002	参数不符合规范