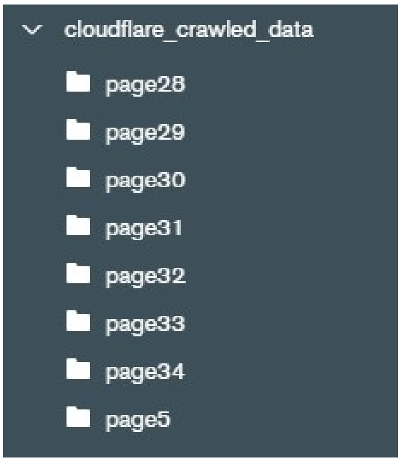
All the crawled data sets are stored by using Mongodb.
Each Cloudflare community page contains 30 posts.
For each single page, a unique collection is created to store the posts information.

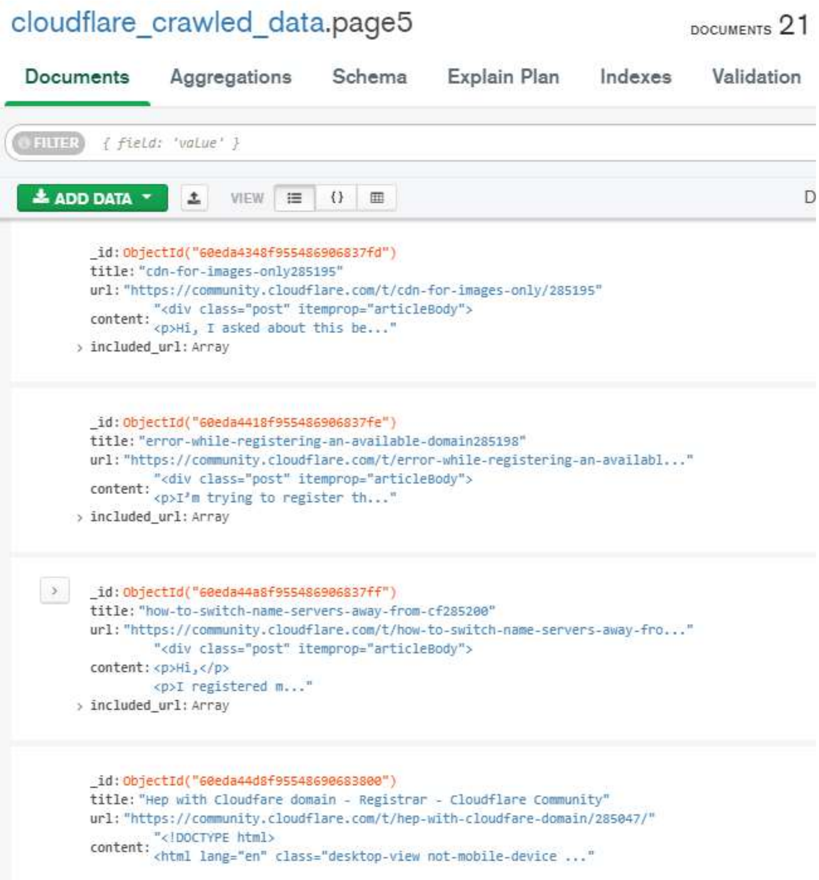


cloudflare_crawled_data

- page28
- page29
- page30
- page31
- page32
- page33
- page34
- page5

For each Cloudflare post, the database will store its title, URL, refined post contents, and included URLs.

Take page5 as a more specific example:



cloudflare_crawled_data.page5 DOCUMENTS 21

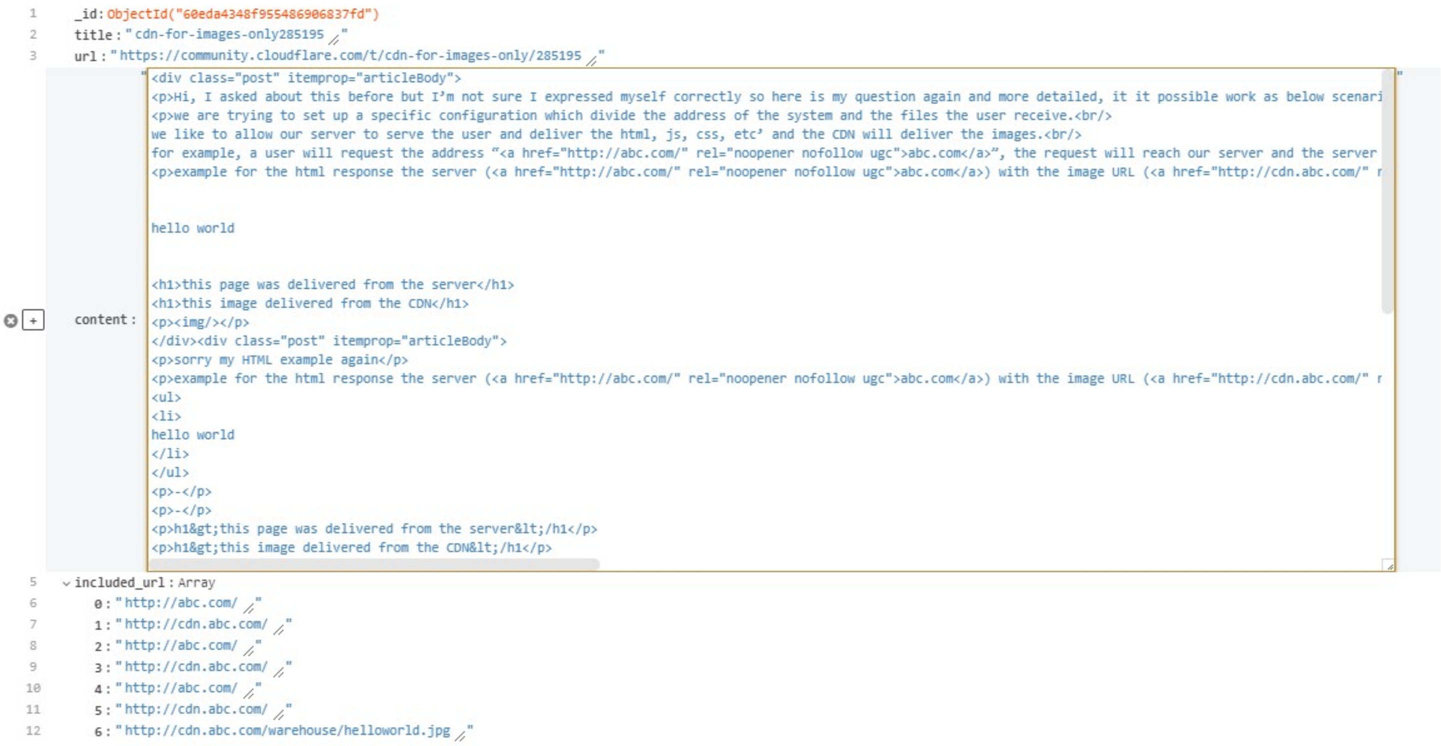
```
_id: ObjectId("60eda4348f955486906837fd")
title: "cdn-for-images-only285195"
url: "https://community.cloudflare.com/t/cdn-for-images-only/285195"
         "<div class="post" itemprop="articleBody">
content:
         <p>Hi, I asked about this be..."
> included_url: Array

_id: ObjectId("60eda4418f955486906837fe")
title: "error-while-registering-an-available-domain285198"
url: "https://community.cloudflare.com/t/error-while-registering-an-availabl..."
         "<div class="post" itemprop="articleBody">
content:
         <p>I’m trying to register th..."
> included_url: Array

_id: ObjectId("60eda44a8f955486906837ff")
title: "how-to-switch-name-servers-away-from-cf285200"
url: "https://community.cloudflare.com/t/how-to-switch-name-servers-away-fro..."
         "<div class="post" itemprop="articleBody">
content: <p>Hi,</p>
         <p>I registered m..."
> included_url: Array

_id: ObjectId("60eda44d8f95548690683800")
title: "Hep with Cloudfare domain - Registrar - Cloudflare Community"
url: "https://community.cloudflare.com/t/hep-with-cloudfare-domain/285047/"
         "<!DOCTYPE html>
content:
         <html lang="en" class="desktop-view not-mobile-device ..."
```

The first post on page 5, the raw HTML doc is refined to get the simplified post contents:

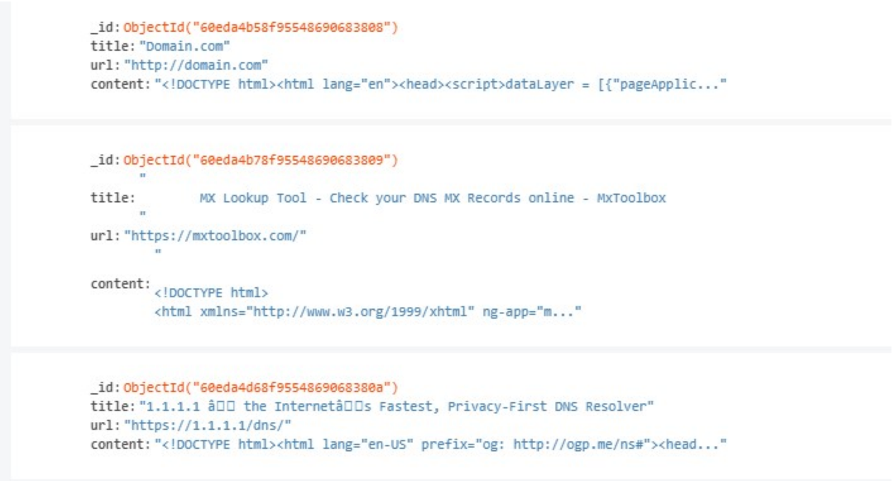


```
_id: ObjectId("60eda4348f955486906837fd")
title: "cdn-for-images-only285195"
url: "https://community.cloudflare.com/t/cdn-for-images-only/285195"
content: "<div class="post" itemprop="articleBody">
<p>Hi, I asked about this before but I’m not sure I expressed myself correctly so here is my question again and more detailed, it it possible work as below scenari
<p>we are trying to set up a specific configuration which divide the address of the system and the files the user receive.<br/>
we like to allow our server to serve the user and deliver the html, js, css, etc’ and the CDN will deliver the images.<br/>
for example, a user will request the address “<a href="http://abc.com/" rel="noopener nofollow ugc">abc.com</a>”, the request will reach our server and the server
<p>example for the html response the server (<a href="http://abc.com/" rel="noopener nofollow ugc">abc.com</a>) with the image URL (<a href="http://cdn.abc.com/" r


hello world


<h1>this page was delivered from the server</h1>
<h1>this image delivered from the CDN</h1>
<p><img/></p>
</div><div class="post" itemprop="articleBody">
<p>sorry my HTML example again</p>
<p>example for the html response the server (<a href="http://abc.com/" rel="noopener nofollow ugc">abc.com</a>) with the image URL (<a href="http://cdn.abc.com/" r
<ul>
<li>
hello world
</li>
</ul>
<p>-</p>
<p>-</p>
<p>h1&gt;this page was delivered from the server&lt;/h1</p>
<p>h1&gt;this image delivered from the CDN&lt;/h1</p>"
included_url: Array
    0: "http://abc.com/"
    1: "http://cdn.abc.com/"
    2: "http://abc.com/"
    3: "http://cdn.abc.com/"
    4: "http://abc.com/"
    5: "http://cdn.abc.com/"
    6: "http://cdn.abc.com/warehouse/helloworld.jpg"
```

For included URLs in the post contents, the titles, URLs, and raw HTML contents are stored:

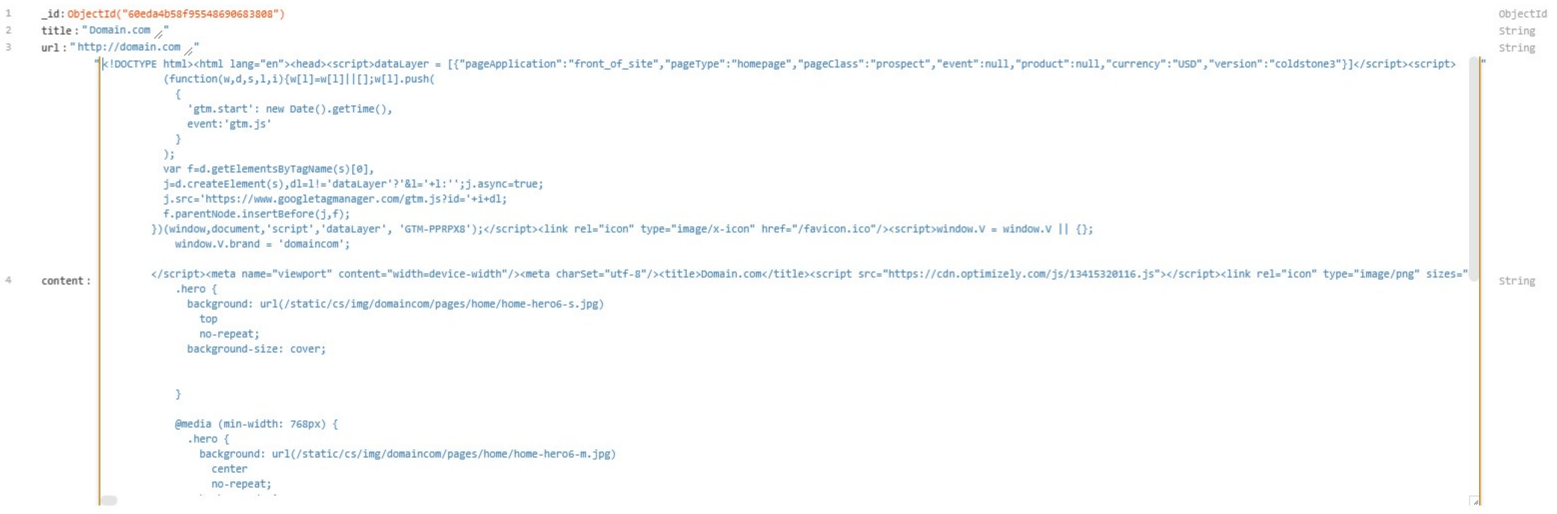


```
_id: ObjectId("60eda4b58f95548690683808")
title: "Domain.com"
url: "http://domain.com"
content: "<!DOCTYPE html><html lang="en"><head><script>dataLayer = [{"pageApplic..."

_id: ObjectId("60eda4b78f95548690683809")
        "
title:         MX Lookup Tool - Check your DNS MX Records online - MxToolbox
        "
url: "https://mxtoolbox.com/"
         "
content: <!DOCTYPE html>
         <html xmlns="http://www.w3.org/1999/xhtml" ng-app="m..."

_id: ObjectId("60eda4d68f9554869068380a")
title: "1.1.1.1 â the Internetâs Fastest, Privacy-First DNS Resolver"
url: "https://1.1.1.1/dns/"
content: "<!DOCTYPE html><html lang="en-US" prefix="og: http://ogp.me/ns#"><head..."
```

A more detailed view of the stored information of the external URL above.



```
_id: ObjectId("60eda4b58f95548690683808")                                                    ObjectId
title: "Domain.com"                                                                          String
url: "http://domain.com"                                                                     String
content: "<!DOCTYPE html><html lang="en"><head><script>dataLayer = [{"pageApplication":"front_of_site","pageType":"homepage","pageClass":"prospect","event":null,"product":null,"currency":"USD","version":"coldstone3"}]</script><script>
           (function(w,d,s,l,i){w[l]=w[l]||[];w[l].push(
             {
               'gtm.start': new Date().getTime(),
               event:'gtm.js'
             }
           );
           var f=d.getElementsByTagName(s)[0],
           j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.async=true;
           j.src='https://www.googletagmanager.com/gtm.js?id='+i+dl;
           f.parentNode.insertBefore(j,f);
         })(window,document,'script','dataLayer', 'GTM-PPRPX8');</script><link rel="icon" type="image/x-icon" href="/favicon.ico"/><script>window.V = window.V || {};
             window.V.brand = 'domaincom';

         </script><meta name="viewport" content="width=device-width"/><meta charSet="utf-8"/><title>Domain.com</title><script src="https://cdn.optimizely.com/js/13415320116.js"></script><link rel="icon" type="image/png" sizes="      String
             .hero {
               background: url(/static/cs/img/domaincom/pages/home/home-hero6-s.jpg)
                 top
                 no-repeat;
               background-size: cover;


             }

             @media (min-width: 768px) {
               .hero {
                 background: url(/static/cs/img/domaincom/pages/home/home-hero6-m.jpg)
                   center
                   no-repeat;
```