



ETL PROCESS EXPLANATION

Extraction

The `contact_collector` function performs the extraction process in the ETL (Extraction, Transformation, and Loading) pipeline by collecting contact records from the HubSpot CRM.

Here's a step-by-step explanation of the extraction process based on the function:

- 1. Initialization:** The function initializes variables such as `data_limit`, `all_records`, and `paging_count`. It also sets the access token and headers required for making API requests.
- 2. Data Query:** The function constructs a JSON payload `data` that specifies the desired properties to retrieve from the CRM. It includes properties like `raw_email`, `country`, `phone`, `technical_test__create_date`, `industry`, `address`, and `hs_object_id`. The payload also includes a filter to only collect records where the `allowed_to_collect` property is set to `True`.
- 3. API Requests:** The function sends POST requests to the specified URL (`https://api.hubapi.com/crm/v3/objects/contacts/search`) using the access token and headers. It passes the JSON payload as the request data. The function retrieves the response and checks if the status code is 200 (indicating a successful response).
- 4. Pagination:** The function checks if the response includes a `paging` key. If not, it breaks the loop as all records have been collected. If the `paging` key is present, it updates the `paging_count` and the JSON payload `data` with the new `after` value to fetch the next page of results.
- 5. Data Collection:** The function extracts the contact properties from the response JSON and appends them to the `all_records` list.
- 6. Return:** Finally, the function creates a new list `record_properties` that contains dictionaries representing the contact properties extracted from all the collected records. It returns this list as the result of the extraction process.

Subsequently, a Data Frame is generated from the data obtained.



| | address | country | createdate | hs_object_id | industry | lastmodifieddate | phone | raw_email |
|------|-------------------------|-----------|--------------------------|--------------|----------------------|--------------------------|---------------|---|
| 0 | Blackpool Rue, 6576 | Waterford | 2023-05-15T02:39:02.002Z | 416102 | Poultry and fish | 2023-07-02T03:21:09.194Z | 0-774-386-624 | Zoe <zoe_owen450104633@acrit.org> Contact Info. |
| 1 | Parkfield Avenue, 5340 | Ireland | 2023-05-15T02:39:02.003Z | 413403 | Fruit and vegetables | 2023-07-02T03:17:41.474Z | 6-777-367-783 | Zara <zara_rodwell1398442854@nickia.com> Conta... |
| 2 | Abourne Lane, 876 | Ireland | 2023-05-15T02:39:02.003Z | 417951 | Milling | 2023-07-02T03:12:00.570Z | 5-618-556-540 | Zara <zara_freeburn1593147546@gmail.com> Conta... |
| 3 | Chester Crossroad, 7070 | Dublin | 2023-05-15T02:39:02.003Z | 419852 | Dairy products | 2023-07-02T03:17:47.964Z | 1-161-604-327 | Winnie <winnie_walter538064895@sheye.org> Cont... |
| 4 | Tilloch Crossroad, 8332 | Dublin | 2023-05-15T02:39:02.003Z | 425352 | Meat | 2023-07-02T03:16:09.371Z | 5-645-416-200 | Zoe <zoe_owen1652446013@bungar.biz> Contact Info. |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 6931 | Cato Hill, 5704 | Waterford | 2023-05-15T02:39:02.968Z | 463801 | Fruit and vegetables | 2023-07-02T03:11:02.685Z | 2-301-310-740 | Rhea <rhea_grady231044837@deons.tech> Contact ... |
| 6932 | Howard Crossroad, 9998 | London | 2023-05-15T02:39:02.969Z | 454595 | Bakery products | 2023-07-02T03:22:55.259Z | 6-747-016-018 | Hailey <hailey_farrell934796609@brety.org> Con... |
| 6933 | Walnut Rue, 3859 | Dublin | 2023-05-15T02:39:02.969Z | 454631 | Animal feeds | 2023-07-02T03:21:09.195Z | 5-516-171-174 | Caleb <caleb_purvis1251615808@iatim.tech> Cont... |
| 6934 | Kimberley Pass, 3455 | Plymouth | 2023-05-15T02:39:02.969Z | 457622 | Bakery products | 2023-07-02T03:22:49.581Z | 0-605-727-343 | Gina <gina_weasley83573127@naiker.biz> Contact... |
| 6935 | Yorkshire Alley, 5843 | Cork | 2023-05-15T02:39:02.969Z | 461395 | Fruit and vegetables | 2023-07-02T03:22:56.134Z | 3-814-518-751 | Aeris <aeris_walsh769266811@gomple.com> Contac... |

6936 rows x 8 columns

Transformation

1. **Detecting country and city:** The process applies the `'country_recognition'` function. This function recognizes the country based on the city. The result of this transformation is then split into two columns, 'country' and 'city'.

2. **Applying Functions to Email Column:** The function applies the `'found_emails'` function to the 'raw_email' column. This function extracts the email address from a raw email string.

3. **Formatting phones:** This method formats the phone numbers based on the country code.

4. **Duplicates Management:** The function performs duplicates management on the transformed DataFrame using the `'duplicates_management'` function. This function manages duplicate records based on email values.

5. **Return:** Finally, the function returns the transformed DataFrame, which contains the result of the transformation process.

| | address | country | hs_object_id | industry | lastmodifieddate | phone | email | city |
|-----|-------------------------|---------|--------------|----------------------|--------------------------|--------------------|-------------------------------------|------------|
| 0 | Blackpool Rue, 6576 | Ireland | 416102 | Poultry and fish | 2023-07-02T03:21:09.194Z | (+353) 7743 86624 | zoe_owen450104633@acrit.org | Waterford |
| 1 | Parkfield Avenue, 5340 | Ireland | 413403 | Fruit and vegetables | 2023-07-02T03:17:41.474Z | (+353) 6777 367783 | zara_rodwell1398442854@nickia.com | None |
| 2 | Abourne Lane, 876 | Ireland | 417951 | Milling | 2023-07-02T03:12:00.570Z | (+353) 5618 556540 | zara_freeburn1593147546@gmail.com | None |
| 3 | Chester Crossroad, 7070 | Ireland | 419852 | Dairy products | 2023-07-02T03:17:47.964Z | (+353) 1161 604327 | winnie_walter538064895@sheye.org | Dublin |
| 4 | Tilloch Crossroad, 8332 | Ireland | 425352 | Meat | 2023-07-02T03:16:09.371Z | (+353) 5645 416200 | zoe_owen1652446013@bungar.biz | Dublin |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | Sheringham Avenue, 8029 | England | 432051 | Meat | 2023-07-02T03:17:47.962Z | (+44) 5854 686161 | wade_gates1717184332@ovock.tech | London |
| 96 | Chestnut Way, 5055 | Ireland | 439951 | Animal feeds | 2023-07-02T03:11:45.243Z | (+353) 5481 346271 | wade_alexander481189081@liret.org | Dublin |
| 97 | Adams Tunnel, 1181 | England | 440851 | Meat | 2023-07-02T03:22:54.570Z | (+44) 6174 303382 | wade_anderson2007937750@deavo.com | London |
| 98 | Gathorne Street, 9970 | England | 427453 | Bakery products | 2023-07-02T03:21:14.670Z | (+44) 7276 180387 | wade_alexander330181759@cispeto.com | Winchester |
| 99 | Adelaide Road, 5799 | England | 436653 | Animal feeds | 2023-07-02T03:22:43.185Z | (+44) 6706 678525 | vivian_russell695390295@twace.org | Winchester |

100 rows x 8 columns



Load

After the Transformation process it is necessary to upload the transformed records into the HubSpot platform using its API.

Contacts ▾
6,709 records

Data Quality

Actions ▾

Import

Create contact

My contacts

+ Add view (1/50)

All views

Contact owner ▾

Create date ▾

Last activity date ▾

Lead status ▾

Advanced filters (0)

Save view

| | NAME | EMAIL | PHONE NUMBER | ORIGINAL INDUSTRY | ORIGINAL CREATE DATE | STREET ADDRESS | COUNTRY/REGION |
|--|-------------------------|------------------------------|--------------------|----------------------|----------------------|-------------------------|----------------|
| | rufus_roth426537683... | rufus_roth426537683@acri... | (+44) 1261 765254 | Meat | Jul 2, 2023 | Dunstons Grove, 6712 | England |
| | lorraine_lindop14077... | lorraine_lindop1407779031... | (+353) 7648 4507 | Fruit and vegetables | Jul 2, 2023 | Bicknell Hill, 1202 | Ireland |
| | dani_murray67836971... | dani_murray678369716@di... | (+353) 7306 147528 | Animal feeds | Jul 2, 2023 | Paris Tunnel, 8579 | Ireland |
| | mark_lunt1841661524... | mark_lunt1841661524@sve... | (+353) 1207 121838 | Bakery products | Jul 2, 2023 | Longleigh Street, 1777 | Ireland |
| | michaela_downing15... | michaela_downing1593001... | (+44) 5304 640580 | Meat | Jul 2, 2023 | Durham Streer, 9119 | England |
| | ivette_gonzales12962... | ivette_gonzales1296254376... | (+44) 4681 760445 | Milling | Jul 2, 2023 | Arundel Tunnel, 8299 | England |
| | chester_clarkson1900... | chester_clarkson19005475... | (+353) 8067 454270 | Bakery products | Jul 2, 2023 | Thrale Lane, 1299 | Ireland |
| | isla_wilson20083659... | isla_wilson2008365983@q... | (+353) 7422 348510 | Fruit and vegetables | Jul 2, 2023 | Aspen Walk, 241 | Ireland |
| | ryan_morris15467014... | ryan_morris1546701447@... | (+44) 1333 85782 | Milling | Jul 2, 2023 | Sherlock Hill, 1032 | England |
| | maria_gilmore436638... | maria_gilmore436638082... | (+44) 5606 633403 | Dairy products | Jul 2, 2023 | Armwell Boulevard, 2333 | England |
| | enoch_kent122420385... | enoch_kent122420385@inf... | (+44) 8451 041687 | Poultry and fish | Jul 2, 2023 | Becklow Rue, 5646 | England |
| | dakota_farrow188212... | dakota_farrow1882128246... | (+44) 6238 051128 | Meat | Jul 2, 2023 | Geary Crossroad, 881 | England |