

Import.io - PDP

❖ **Index**

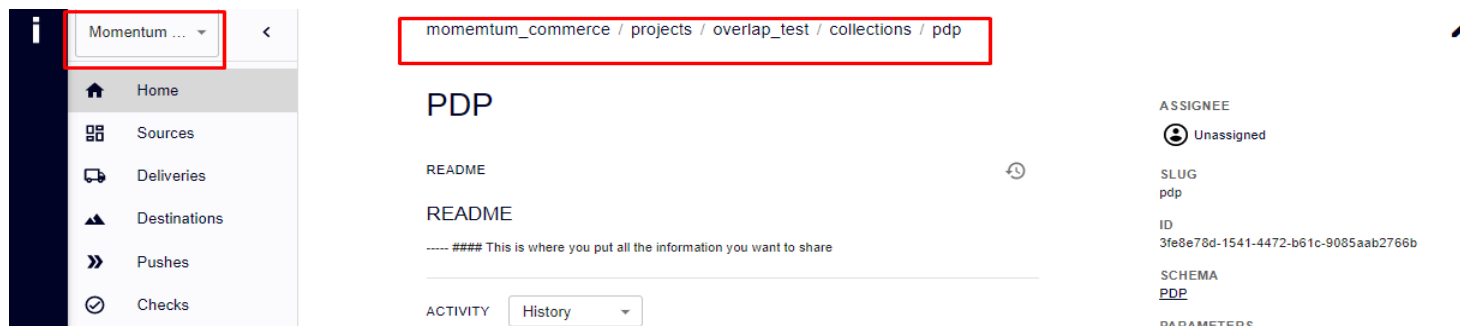
- Snapshot selection process
- Scorecard preparation
- Attributes validation in CDQA.
- The debugger verification process for issues

❖ **Snapshot selection process**

Steps to be followed :

We need to pick a source name from the work assignment sheet.

- Go to Workbench, select the project and follow the path shown in below SS : (For PDP)



➤ Here, will get the list of snapshots for which we searched.

Momentum ...

momemtum_commerce / projects / overlap_test / collections / pdp / sources / amazon_us

Last Updated on June 25th 2021, 7:03:00 pm
(a few seconds ago)

Amazon US Snapshots

Status	Assignee	Time in Status	Inputs	Breached Threshold	Rows	%
View PUSHED	Unassigned	an hour	50331	<input type="checkbox"/>	46950	20
View PUSHED	Unassigned	a day	50331	<input type="checkbox"/>	47495	
View PUSHED	Unassigned	2 days	50331	<input type="checkbox"/>	47804	

Note: Only we need to take a pushed/chunked pushed one.

- How to get the S3 path to download the t6 file from the AWS S3 bucket.

The screenshot displays the Momentum application interface. On the left is a dark sidebar with navigation icons and labels: Home, Import status, Checks, Compare, Graphs, and Drilldown. The main content area is divided into two sections. The top section shows a table with test results:

noScreenshotPct	100
failedOptionalTests	0
failedRequiredTests	0
transformFailurePct	2.861059784228408

The bottom section, titled "DESTINATION PUSHES", contains two entries. The first entry, "S3 Internal", is highlighted with a red rectangle. It shows a status of "COMPLETE" and a refresh icon. Below it is a table with the following data:

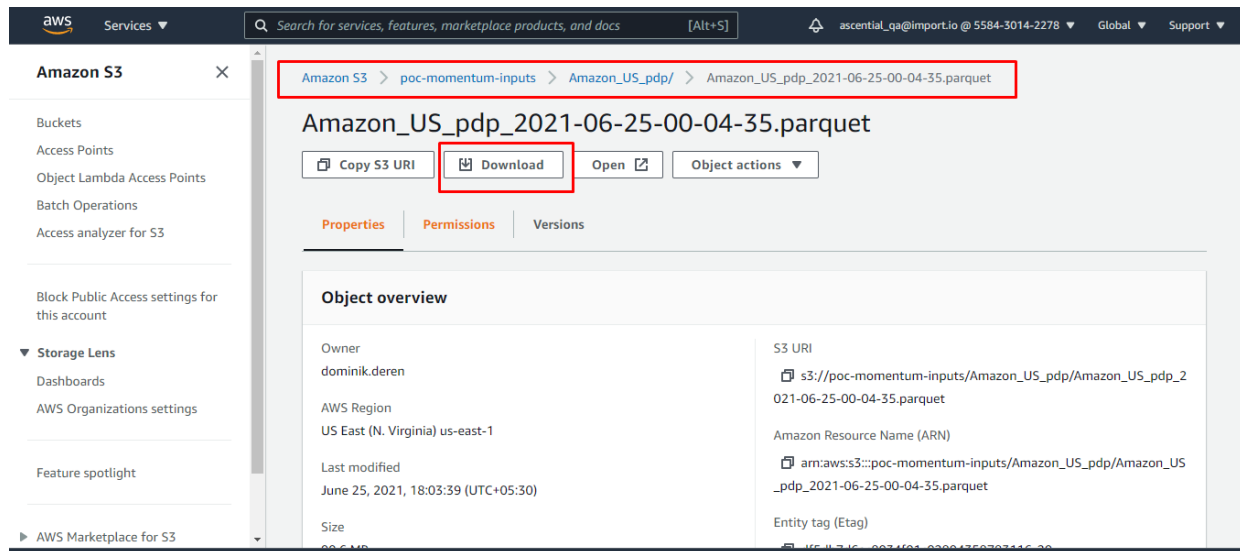
Type	URI
customParquetUri	s3://poc-momentum-inputs/Amazon_US_pdp/Amazon_US_pdp_2021-06-25-00-04-35.parquet

The second entry, "Customer External", also shows a status of "COMPLETE" and a refresh icon. Below it is a table with the following data:

Type	URI
customParquetUri	s3://imc-importio-output/Amazon_US_pdp/Amazon_US_pdp_2021-06-25-00-04-35.parquet

- S3 file path location: Only internal path we need to select.

AWS location for t6 file :



- Click on download, it will get downloaded into the system.

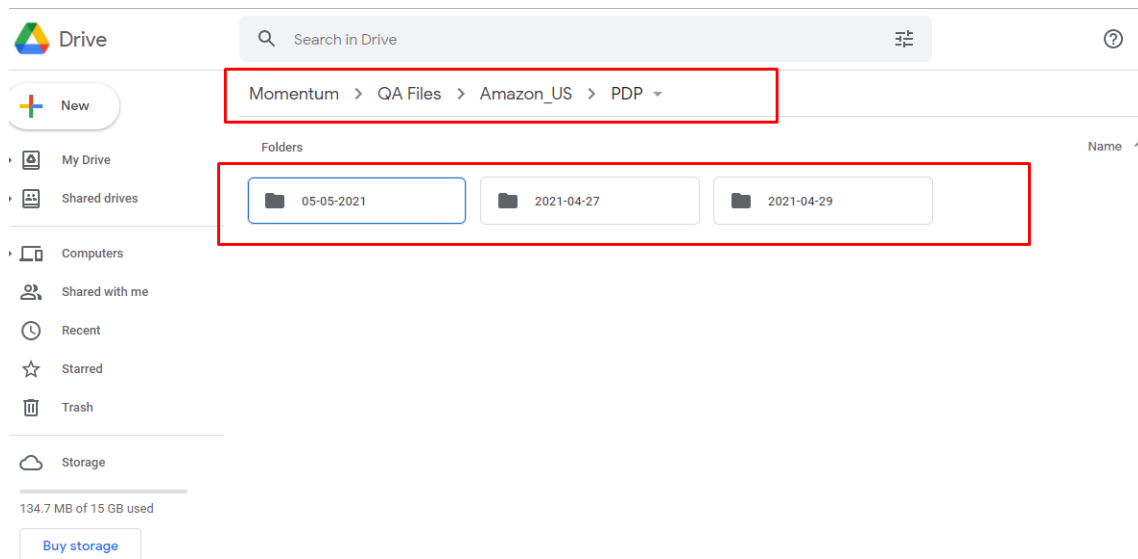
❖ Scorecard preparation

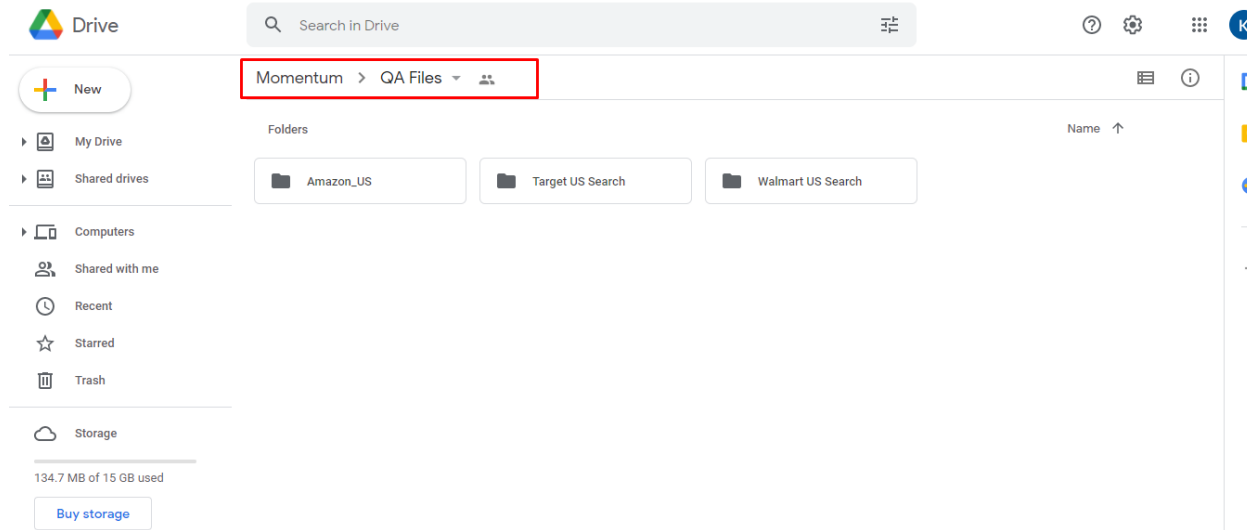
In this project, we need to perform CDQA i.e Criteria 3.

Go to momentum drive.

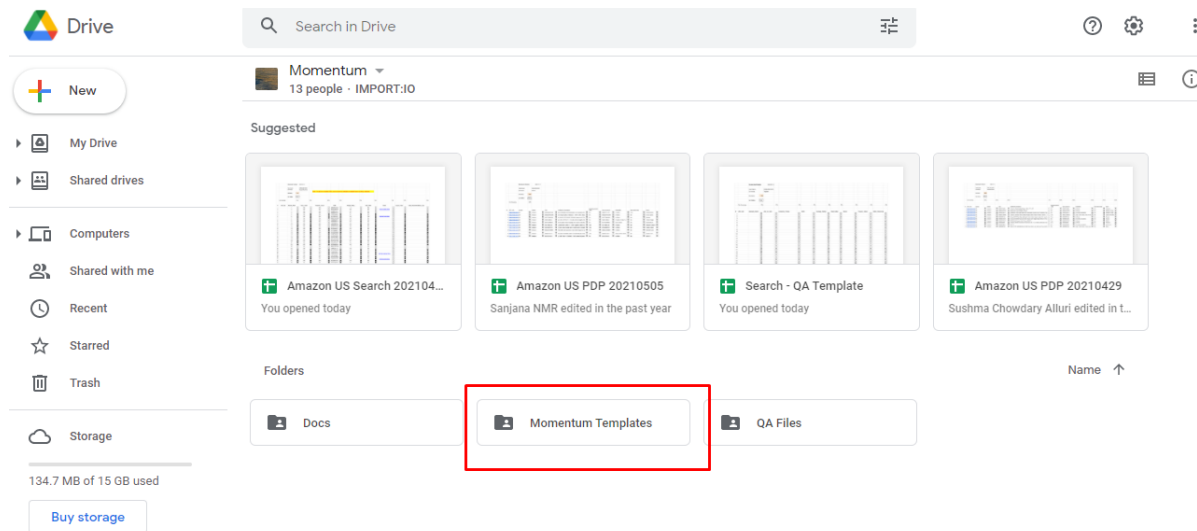
Drive link: https://drive.google.com/drive/u/0/folders/0ADlQi_yEtlcUk9PVA

- Create a source name folder in QA files.





- Create a feed type i.e PDP folder inside create a date folder
- Go to folder 'Momentum Templates'.



- Make a copy of the search template.
- Move the master scorecard to the respective domain folder.
- Template link :
 - <https://drive.google.com/drive/u/0/folders/1jEvvjbpMDSVXzhon8eH4-bseAPxePAJh>
- Rename the scorecard in the format as below:
 - Source name > Locale > Feed type > Date (YYYYMMDD)
 - EX : Amazon US PDP 20210427

❖ **Attributes validation in CDQA.**

- After downloading the t6 file into the system, upload it to the data folder.
- Paste the 10 URLs into the data sample.

Amazon US PDP 20210429

File Edit View Insert Format Data Tools Add-ons Help Last edit was on May 5

100% \$ % ,0 .00 123 Arial 9 B I T A

#	URL (ref)	amazon's choice flag	asin_canonical	availability	best_seller_flag	brand	brand_url	buy_box_owner
1	https://www.amazon.com/dp/B0002YU6E4	<input type="checkbox"/> No	<input checked="" type="checkbox"/> B0002YU6E4	<input checked="" type="checkbox"/> In Stock.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Coleman	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
2	https://www.amazon.com/dp/B000ACYW7K	<input type="checkbox"/> No	<input checked="" type="checkbox"/> B000ACYW7K	<input checked="" type="checkbox"/> In Stock.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Melissa & Doug	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
3	https://www.amazon.com/dp/B084WLBDV7	<input type="checkbox"/> No	<input checked="" type="checkbox"/> B084WLBDV7	<input checked="" type="checkbox"/> Currently unavailable.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Melissa & Doug	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
4	https://www.amazon.com/dp/B0001FSWMM	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> B0001FSWMM	<input checked="" type="checkbox"/> Only 2 left in stock (more on the page)	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Optimum Nutrition	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
5	https://www.amazon.com/dp/B000087LGA	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> B000087LGA	<input checked="" type="checkbox"/> In Stock.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Smead	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
6	https://www.amazon.com/dp/B005NICQZ0	<input type="checkbox"/> No	<input checked="" type="checkbox"/> B005NICQZ0	<input checked="" type="checkbox"/> Currently unavailable.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> Hoover	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com
7	https://www.amazon.com/dp/B0000CCXVZ	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> B0000CCXVZ	<input checked="" type="checkbox"/> In Stock.	<input checked="" type="checkbox"/> No	<input checked="" type="checkbox"/> IRWIN	<input checked="" type="checkbox"/> https://www.amazon.com/	<input checked="" type="checkbox"/> Amazon.com

Attribute list:

Attribute	Explanation	Example
country	The Locale should be retrieved.	EX: US
retailer	The Site name or the source name.	Ex: AMAZON
ASIN	The amazon-specific identifier will be present only for Amazon Websites.	Ex: B0002YU6E4
additional_description	The Additional description for the product, sometimes it will be in the form of bullets then they should be separated by Pipelines.	SS: https://prnt.sc/16qd0fn

amazons_choice_flag	If the Amazon choice flag is present then it should be retrieved Yes else No	SS: https://prnt.sc/16qew82
availability	The availability of the product should be retrieved like In stock, Out of Stock, etc...	In Stock, Currently unavailable, In stock on April 30, 2021.
best_seller_flag	Bestseller Flag will be available on the website if present it will be displayed as Yes else NO.	SS: https://prnt.sc/16qdies
brand	The actual Brand of the product.	SS: https://prnt.sc/16qf1mt
brand_url	The URL of the product original willThe timestamp be redirected to Brand.	URL: https://www.amazon.com/Coleman/b/ref=bl_dp_s_web_17713114011?ie=UTF8&node=17713114011&field-lbr_brands_browse-bin=Coleman
buy_box_owner	The Link will be redirected to the Owner like home Page.	Ex: Amazon.com
buy_box_price_raw	The original Price of the Product including the Currency Symbol	Ex: \$6.93
buy_box_price	The original price of the product without currency symbol	Ex: 6.93
sns_price	The Price which is under Subscribe and	SS: https://prnt.sc/124wfre

	Save Title with Currency symbol needs to be retrieved.	
one_time_price	The price which is under the one-time price Title can be retrieved without a Currency Symbol.	SS: https://prnt.sc/16qfree
coupon		
coupon_sns		SS: https://prnt.sc/16gg7nq
coupon_details		
coupon_value_currency		
customer_review_rating	The rating of the product originally is in the range of 1 to 5.	Ex: 4.5
deal_price		
deal_type		

download_date	The timestamp of the Download time should be populated even if the product is not found.	Ex: 04/27/2021 3:34:04 AM
enhanced_content_exists	It returns the boolean value if the enhanced content is present else No.	
enhanced_content	The extended description of the product, if it is taken from multiple places, should be separated with spaces under the heading from the manufacturer.	SS: https://prnt.sc/16qhp6z
enhanced_content_images	If in enhanced content, images are present then they should be retrieved, it will be available under the heading From the Manufacturer.	SS: https://prnt.sc/16qhibu
enhanced_content_ASIN_links	The Links which will be redirected to the enhanced content need to be retrieved in the T6 file.	
fast_track_message	The messages regarding the fastest delivery should be retrieved.	SS: https://prnt.sc/124vcq7
feature_bullets	The information which is under the bullet points should be retrieved with	Ex: ["Made of the highest quality materials' 'Camping Tents Accessories' 'Another quality Coleman product'"]

	double-quotes.	
final_price	The final price of the product after the discount.	Ex: 16.2
ingredient_image_present	If any image present near the Ingredients then it should retrieve Yes else No	Ex: Yes/No
ingredients_list	The Ingredients present for the product should be retrieved.	Ex: Maltodextrin, Protein Blend (Whey Protein Concentrate, Calcium Caseinate, Egg Albumen, Sweet Dairy Whey), Natural and Artificial Flavors, Vitamin and Mineral Blend (Dipotassium Phosphate, Magnesium Aspartate, Dicalcium Phosphate, Calcium Citrate, Ascorbic Acid, Niacinamide, Zinc Citrate, Beta Carotene, d-Calcium Pantothenate, d-alpha Tocopheryl Succinate, Selenomethionine, Copper Gluconate, Ferrous Fumarate, Manganese Amino Acid Chelate, Thiamin Mononitrate, Pyridoxine Hydrochloride, Riboflavin, Chromium Polynicotinate, Folic Acid, Biotin, Molybdenum Amino Acid Chelate, Cholecalciferol, Potassium Iodide, Cyanocobalamin), Medium Chain Triglycerides, Lecithin, FD&C Red #40, Acesulfame Potassium. SS: https://prnt.sc/16qjhce

international_shipping	If the international shipping is available for the product then that should be retrieved other than the Locale.	
is_merchant_exclusive		Ex: YES/NO
keywords	If the keywords are present then they should be retrieved in DOM we can inspect by using Keyword (Keywords).	Ex: Motherhood Maternity Women's Maternity Clip Down Nursing Tank Top Cami, White, Small, Motherhood Maternity,92305
large_image_count	Note: Only used for Amazon sites. Count the number of times text 'SL1500_.jpg' appears in the tag ImageBlockATF in DOM.	
list_price	The Struck out price is considered for the list price which is nothing but the original Price without discount.	SS: https://prnt.sc/16qk2el
manufacturer_name	The name of the Manufacturer should be retrieved sometimes. A brand can be considered as a Manufacturer.	
newer_version_asin	If a product has a newer version, the ASIN value should be retrieved.	
newer_version_title	The title of the newer version product	

	should be retrieved.	
node_id	In the Dom we can search by using the keyword node.	Ex: 7141123011 > 7147440011 > 7581675011 > 2379250011 > 2379253011 SS: https://prnt.sc/16qkuvvg
number_of_customer_reviews	The total number of customers who reviewed the product should be retrieved.	Ex: https://prnt.sc/16qlmby
online_price_currency	The Currency of the locale should be retrieved.	SS : https://prnt.sc/16ql7ff
pack_size	The number of Packs in the product should be retrieved, we can see the number of packs in the title, or else a separate heading will be present.	SS: https://prnt.sc/16qm2v5
product_size		
count		
unit_size		
unit_size_uom		

pASIN	This is the same as ASIN(Parent Amazon-specific Identifier)	Ex: B0929PBHWG
prime_flag	If the Prime flag is present on the website then it should be retrieved as Yes else No.	Ex: Yes - Ship and Sold by Amazon
secondary_images	Other than Product Main Image rest are considered as Secondary Images.	Ex SS: https://prnt.sc/16qmpax
count_of_images	Total number of Images present including Main and Secondary images.	Ex SS: https://prnt.sc/16qmtsd
product_title	The title of the Product including size, brand	Ex SS: https://prnt.sc/16qmyzl
products_per_page	The count of variants in number if available. If no variants are there should be populated as 0.	Ex SS: https://prnt.sc/16qnfcv
retailer_category	The category and Subcategories which are present on the website should be retrieved.	
best_seller_ranking_category	The category of Bestseller products should be retrieved.	Ex SS: ['Toys & Games' 'Hand Puppets']
best_seller_ranking_rank	The ranking of that Bestseller products	Ex SS: https://prnt.sc/124x60m

	should be retrieved.	
best_seller_ranking_url	The URL of the best seller product should be retrieved.	Ex SS: ['https://www.amazon.com/gp/bestsellers/toys-and-games/ref=pd_zg_ts_toys-and-games' 'https://www.amazon.com/gp/bestsellers/toys-and-games/166335011/ref=pd_zg_hrsr_toys-and-games']
shipping_weight	The total weight of the product including the package.	
buy_box_sellerid	The Merchant or Seller ID for the buy box winner.	
subscribe_and_save	Populate with 'Yes' if data attribute on the product page, else populate with 'No'.	Ex SS: https://prnt.sc/124wfre
upc	Universal Product Code consists of 12 numeric digits that are uniquely assigned to each trade item.This is the barcode used for the scanning of trade items as per GS1 guidelines. Can be gtin/ean also.	Ex: 52100005966
URI	The URL of the product should be	Ex SS:

	retrieved, even if the product is not available the URL should be retrieved.	https://www.amazon.com/dp/B000ACYW7K?th=1&psc=1&showDetailProductDesc=1
video	The video is related to the product present on the website.	Ex SS: https://prnt.sc/16qp67h
FFP	Frustration-Free Packaging. Populate with 'Yes' if data attribute on the product page, else populate with 'No'.	
BusinessOnly		
redirect_asin		
vat_relief		
gtin	Same as upc	Ex: 52100005966
item_model_number	The model number of the product should be retrieved, should be available with mpc,mpn, model.	SS: https://prnt.sc/16qpbkz
variation_type		

minimum_order_quantity		
deal_type_percent_claimed		
seller_url	The url of the product which should be redirected to Seller should be retrieved.	
product_alert_alt_text		
product_alert_image_url		
feature_bullets_count	The Total number of bullets should be retrieved.	
variant_asins	Note: Only used for Amazon sites. Variants code information separated by	Ex: B000GIPJ02 B0015R36SK B0015R6ZPQ B000GIPJ0M
climate_pledge_friendly	Populate with 'Yes' if data attribute on the product page, else populate with 'No'.	
virtual_bundle	Populate with 'Yes' if data attribute on the product page, else populate with	

	'No'.	
specificationsObj	The Name of the Specification and its value should be retrieved.	Ex: [{ 'specificationName': 'Manufacturer', 'specificationValue': 'Smead'} { 'specificationName': 'Brand', 'specificationValue': 'Smead'} { 'specificationName': 'Item Weight', 'specificationValue': '8.22 pounds'} { 'specificationName': 'Product Dimensions', 'specificationValue': '10 x 6.88 x 17.69 inches'} { 'specificationName': 'Item model number', 'specificationValue': '67070'} { 'specificationName': 'Is Discontinued By Manufacturer', 'specificationValue': 'No'} { 'specificationName': 'Color', 'specificationValue': 'Assortment'} { 'specificationName': 'Material Type', 'specificationValue': 'Paper'} { 'specificationName': 'Number of Items', 'specificationValue': '1'} { 'specificationName': 'Size', 'specificationValue': '13,000 Labels'} { 'specificationName': 'Manufacturer Part Number', 'specificationValue': '67070'}]
variantsObj	The type of the variant and that value should be retrieved.	Ex: [{ 'variantName': 'customer_package_type', 'variantValue': 'Frustration-Free

		Packaging']}]
coupon_discount_percentage	The Percentage of the Coupon should be retrieved.	
asin_canonical	It should retrieve the URL of that particular ASIN.	
fact_image_present	Only used for Amazon sites, Populate with Yes if tag FACT appears inside tag color images else no should be retrieved.	
newer_version	In the product page, information about a new version of the product.	
product_image	The image of the product.	Ex: https://prnt.sc/16qqa8p

*** Note: There is no need to verify the attributes which are kept empty.**

❖ The debugger verification process for the issues

- Go to the snapshots in the workbench.
- Click on the drill down.

Momentum ...

momentum_commerce / projects / overlap_test / collections / pdp / sources / amazon_us / snapshots / 452b530c-fd7f-4beb-aa55-91b0f2055fce

Last Updated on June 25th 2021, 7:07:27 pm (a few seconds ago)

Snapshot

PROGRESS

CRAWL RUN

Total Inputs

50331

Successful Inputs

50044

Failed Inputs

287

IMPORT

Total Stages

12

Completed Stages

12

Row Count

46950

HEALTH METRICS

Name	Value
rows	46950
okPct	96.53096501162307
pages	50331

STATE: PUSHED

ASSIGNEE: Unassigned

CRAWL RUN ID: [936909c4-3461-40ee-8ccb-702b4fb56e19](#)

RUNTIME CONFIGURATION ID: [a304757e-6262-4242-8a7e-758878203ead](#)

EXTRACTOR: [d344e286-d35f-47a8-acb1-b01c71982f60](#)

ADDED AT: 2021-06-25 00:00:17 (UTC)

COLLECT BY: 2021-06-25 10:00:00 (UTC)

DELIVER BY: 2021-06-25 12:00:00 (UTC)

STOP BY: 2021-06-25 10:00:00 (UTC)

STOPPED AT: 2021-06-25 11:24:32 (UTC)

BREACHED INPUT THRESHOLD

- Execute the query to fetch the records.
- Query : `SELECT * FROM S3Object where _input like '%ASIN%' LIMIT 200`

Orgs Help API Docs

MOMENTUM COMMERCE / TEST / SEARCH / AMAZON US / 5F5CD02B-A585-463B-8A35-9D9FEB5721CC

Summary
Import status
Checks
Timeline
Compare
Graphs
Drilldown

Query Data Top Values Validation issues

```
1 SELECT * FROM $300ject where _input like "Nail clippers" LIMIT 200
```

Run Data - internal

180 rows returned (Returning too many rows may crash your browser)

Download CSV Download JSON Download XLSX Formula to calculate

aggregating2	bagNames	country	department	departmentId	id	listPrice	listPriceCurrency	name	page	placement	price	priceCurrency	url
4.600	Amazon Prime	US	Nail Tools	17269965011	B07Q5ST1RT			Toe Nail Clipper for Ingrown or Thick Toenails, Toenails Trimmer and	1	Sponsored Product	11.990	USD	https://www.amazon.com/gp/srredirect/picassoRedir
4.500	Amazon Prime	US	Nail Tools	17269965011	B089ZCPTTJ			Nail Clippers Set - Fingernail Clippers & Toenail Clippers With FLIP OUT	1	Sponsored Product	19.970	USD	https://www.amazon.com/gp/srredirect/picassoRedir
4.800	Amazon Prime	US	Nail	17269965011	B089N92NHR			Thick Toenail	1	Sponsored Product	12.990	USD	https://www.amazon.com/en/srredirect/picassoRedir

Type here to search

30°C Light rain 3:42 PM 6/25/21

- Select the task id & copy it & then go back to the summary, click on crawl run id paste it there.
- Submit it & select the Crawl run task & the website will start loading from there we can validate our issues.

Import.io API Key (Optional, if you are logged into the application, you don't need to provide it.)

Crawlrun Guid
de38-126a-4179-833c-5d232fb0d9df790b6836-7329-41da-b63d-7de146b929bd

Environment
production

You can enter:

- crawlrun guid (eg. 560fe4bf-791e-469c-a466-9e043c9ab3b9)
- crawlrun guid together with task guid (eg. 560fe4bf-791e-469c-a466-9e043c9ab3b9:3e8bffa7-bdad-4122-9f7b-7f6eb94d8cbd)

SUBMIT LOAD