

Crawlers

Crawlers (2) Info

Last updated (UTC)
November 14, 2024 at 19:04:35

↺

Action ▾

Run

Create

View and manage all available crawlers.

Q Filter crawlers

< 1 >

<input type="checkbox"/>	Name ▾	State ▾	Schedule	Last run ▾	Last run time... ▾	Log
<input type="checkbox"/>	carbon-emission-crawler	✔ Ready		✔ Succeeded	November 8, 20...	View log 🔗
<input type="checkbox"/>	project-climate-watch-glue-crawler	✔ Ready		✔ Succeeded	November 6, 20...	View log 🔗

Glue Database

Databases (1)

Last updated (UTC)
November 14, 2024 at 19:06:25

↺

Edit

Delete

Add databas

A database is a set of associated table definitions, organized into a logical group.

Q Filter databases

< 1 >

<input type="checkbox"/>	Name ▲	Description ▾	Location URI ▾	Created on (UTC)
<input type="checkbox"/>	project-climate-watch-db	database for project climate watch	-	November 5, 2024 at 19:32:12

Glue Tables

Tables (3)

Last updated (UTC)
November 14, 2024 at 19:05:03

↺

Delete

Add tables using crawler

Add table

View and manage all available tables.

Q Filter tables

< 1 > ⚙

<input type="checkbox"/>	Name ▲	Database ▾	Location ▾	Classificat... ▾	Depreca... ▾	View data
<input type="checkbox"/>	raw_carbon_emission	project-climate-watch-db	s3://project-climate-watch/raw-carbo	CSV	-	Table data
<input type="checkbox"/>	state_energy_related_carbon_dioxide_u	project-climate-watch-db	s3://project-climate-watch/raw-data/	CSV	-	Table data
<input type="checkbox"/>	storm_event_details	project-climate-watch-db	s3://project-climate-watch/raw-data/	CSV	-	Table data

Schema

Schema	Partitions	Indexes	Column statistics - <i>new</i>
--------	------------	---------	--------------------------------

Schema (51)

View and manage the table schema.



#	▼	Column name	▼	Data type
1		begin_yearmonth		bigint
2		begin_day		bigint
3		begin_time		bigint
4		end_yearmonth		bigint
5		end_day		bigint
6		end_time		bigint
7		episode_id		bigint
8		event_id		bigint
9		state		string
10		state_fips		bigint
11		year		bigint
12		month_name		string
13		event_type		string
14		cz_type		string
15		cz_fips		bigint
16		cz_name		string
17		ufo		string

Glue to RDB Connection

Connections (2) Info

Actions ▼

Create connection

Create job

You can manage your connections or use a connection in a job.

Q Filter connections by property

< 1 >

	Name ▼	Status ▼	Type ▼	Last modified ▼
<input type="radio"/>	Sqlserver connection	✓ Ready	JDBC	Nov 09, 2024
<input type="radio"/>	Sqlserver connection 2	✓ Ready	JDBC	Nov 09, 2024

Glue Databrew Datasets

Datasets (2) Info

View details

Create project with this dataset

▶ Run data profile

Actions ▼

Q Find datasets

<input type="checkbox"/>	Dataset name ▼	Data type ▼	Data profile ▼	Source ▼	Location ▼
<input type="checkbox"/>	carbon-dioxide-emissions	Data Catalog table	-	Data Catalog	state_energy_related_carbon_dioxide_emissions_by_year_1970_2022_csv
<input type="checkbox"/>	storm-event-details	Data Catalog table	-	Data Catalog	storm_event_details

Databrew Project – Data Preparation

climate-watch-data-prep

Dataset: storm-event-details Sample: Random sample (1,000 rows)

Last job run 4 days ago, no job runs scheduled

▶ Run job

JOB DETAILS

LINEAGE

...

UNDO REDO

FILTER SORT COLUMN

FORMAT CLEAN EXTRACT

MISSING INVALID DUPLICATES OUTLIERS

SPLIT MERGE CREATE

FUNCTIONS CONDITIONS

NEST-UNNEST PIVOT GROUP JOIN UNION

TEXT SCALE MAPPING ENCODE

...

13 RECIPE

Viewing 36 columns 1,000 rows

SAMPLE

GRID SCHEMA PROFILE

ABC STATE	# YEAR	ABC MONTH_NAME
OKLAHOMA	1968	June
MINNESOTA	1968	May
NEBRASKA	1968	July
MICHIGAN	1968	August
NEBRASKA	1968	June
MARYLAND	1968	June
MISSOURI	1968	July
ALABAMA	1968	November
ALABAMA	1968	November
ALABAMA	1968	November
ALABAMA	1968	August
ALABAMA	1968	August
ALABAMA	1968	August
ALABAMA	1968	August

Recipe (13)

climate-watch-data-prep-recipe

Working version

Applied steps (13) Clear all

10 more recipe steps

11. Change type of damage_property_2x to Integer

12. Create column Damage_Property using math function MULTIPLY

13. Delete column DAMAGE_PROPERTY_1, DAMAGE_PROPERTY_2, damage_property_2x

Zoom 100%

Last updated 4 days ago, on November 10, 2024, 1:43:33 pm

Databrew Project Recipe



Recipe (13)



5. **Create column** BEGIN_DATE_TIME_extracted Extract value between positions 10 and 12 from BEGIN_DATE_TIME

6. **Change type** of BEGIN_DATE_TIME_extracted to **Short**

7. **Create column** Begin_Day_Period using Logical functions CASE

8. **Split column at position** 1 from end in DAMAGE_PROPERTY

9. **Change type** of DAMAGE_PROPERTY_1 to **Short**

10. **Create column** damage_property_2x using Logical functions IF

11. **Change type** of damage_property_2x to **Integer**

12. **Create column** Damage_Property using math function MULTIPLY

13. **Delete column** DAMAGE_PROPERTY_1, DAMAGE_PROPERTY_2, damage_property_2x

Job run history

<div><div></div><div>Search by job run ID</div></div>		<div>Show all<div></div></div>		
	Job run ID <div></div>	Last job run status <div></div>	Run time <div></div>	Output
<div></div>	climate-watch-storm-data-prep_2024-11-10-09:49:42	<div></div> Succeeded	15 minutes, 29 seconds	2 outputs
<div></div>	climate-watch-storm-data-prep_2024-11-06-20:33:29	<div></div> Succeeded	14 minutes, 27 seconds	1 output

A1:A52,
B1:B52,
C1:C52,
D1:D52,
E1:E52,
F1:F52,
G1:G52,
H1:H52,
I1:I52,
J1:J52,
K1:K52,
L1:L52,
M1:M52,
N1:N52,
O1:O52,
P1:P52,
Q1:Q52,
R1:R52,
S1:S52,
T1:T52,
U1:U52,
V1:V52,
W1:W52,
X1:X52,
Y1:Y52,
Z1:Z52,
AA1:AA52,
AB1:AB52,
AC1:AC52,
AD1:AD52,
AE1:AE52,

AF1:AF52,
AG1:AG52,
AH1:AH52,
AI1:AI52,
AJ1:AJ52,
AK1:AK52,
AL1:AL52,
AM1:AM52,
AN1:AN52,
AO1:AO52,
AP1:AP52,
AQ1:AQ52,
AR1:AR52,
AS1:AS52,
AT1:AT52,
AU1:AU52,
AV1:AV52,
AW1:AW52,
AX1:AX52,
AY1:AY52,
AZ1:AZ52,
BA1:BA52,
BB1:BB52,
BC1:BC52,
BD1:BD52,