

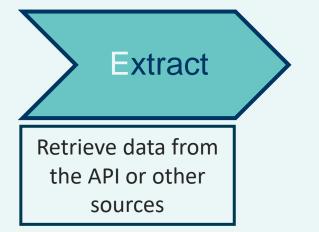
05 – Data Ingestion in FROST via Apache NiFi

Beatrice Olivari

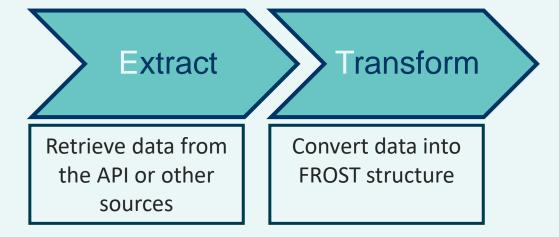


Go with the (ETL) FLOW

deda<mark>,next</mark> EDIA©I

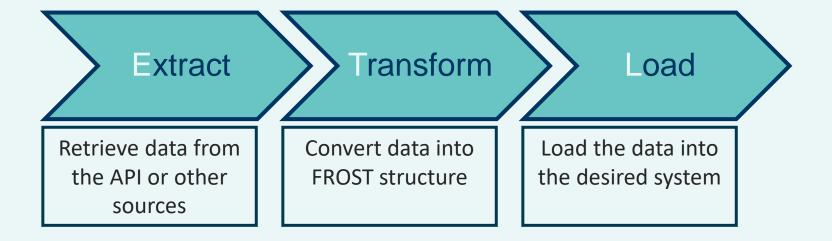


Go with the (ETL) FLOW



deda, next

Go with the (ETL) FLOW



deda<mark>, next</mark> EDIA©I

The tool: Apache NiFi

2 \bigcirc *≡*⁄/ **-**∕⊃ ⊡> ច្រៃ أصأ **111** 0 3 / 0 bytes 0 0 🥥 4 Δ 0 * 0 O O 0 0 ? 0 22:57:15 IST > 3 0 * 0 Ø Navigate GenerateFlowFile ତ୍ର୍ 🚺 ା GenerateFlowFile 1.9.2 org.apache.nifi - nifi-standard-nar 0 (0 bytes) In Read/Write 0 bytes / 0 bytes Out 0 (0 bytes) Tasks/Time 0 / 00:00:00.000 5 min Name success Queued 1 (0 bytes) GenerateFlowFile LogAttribute GenerateFlowFile 1.9.2 org.apache.nifi - nifi-standard-nar org.apache.nifi - nifi-standard-nar 👌 Operate Name success In 0 (0 bytes) 5 min 0 (0 bytes) 5 min Read/Write 0 bytes / 0 bytes 5 min Queued 1 (0 bytes) Read/Write 0 bytes / 0 bytes 5 min GenFlowFilePG Process Group Out Out 0 (0 bytes) 5 min 0 (0 bytes) 5 min Tasks/Time 0/00:00:00.000 5 min Tasks/Time 0/00:00:00.000 5 min 4d7fbd74-016f-1000-f38d-390f0b314c00 Name success * 1 * • = - -Queued 1 (0 bytes) GenerateFlowFile 🔁 🖪 🔝 🖌 🛍 DELETE \bigcirc GenerateFlowFile 1.9.2 org.apache.nifi - nifi-standard-nar In 0 (0 bytes) 5 min Read/Write 0 bytes / 0 bytes 5 min Out 0 (0 bytes) Tasks/Time 0/00:00:00.000 5 min

eda, next

Example flow: Meteorological data



- MISTRAL is an **EU financed** project **started in 2018**
- The **goal** of the MISTRAL portal is to facilitate and foster the reuse of the datasets by the weather community, as well as by its cross-area communities

deda.next

EDIA

Mistral: User Interface



		→ NC	SICN IN
17 out of 2 Regions and Autonomous		12.177 Weather Stations	9 million + Observations every day
and a state of a second block block			
T Filter	Data Statio	5 Meteograms	
Variable	Due for	and the first state	with the second se
Temperature/dry-bulb Temperature	~ ě	Stear Stear Stear	The second
Date	And the fail	218 243	(23) - Constant of
20/07/2023			
11:00 - 13:59	Bertroan - April	Arrest Total Tota	and the second sec
evel 2.000m above ground		Annual Annua	and the second s
	Carried and Carrie	Nodelle sure	and and a set of the s
Time range Analysis or observation, istantaneous value	V BUTTY BUTTY		443 an Antres An
Network Any	Traper		
			United and
Croup of Licenses	- Line	100 100 100 100 100 100 100 100 100 100	
Quality Control Filter	and a state		
	And And		
Update Map 🛃	Argen all an Unital Mirror Barrow Provinge Lange		

https://meteohub.mistralportal.it/app/maps/observations

Mistral: User Interface

deda<mark>,next</mark> EDIA©I

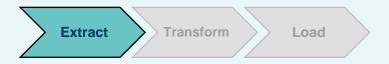
DATASETS FORECASTS • OBSERVATIONS • 17 out of 21 Regions and Autonomous Provinces	DOCUMENTATION + 12.177 Weather Stations	SICN IN SICT 9 million + Observations every day
▼ Filter	Data Stations Meteograms	
Variable	Sense Carton And Sense And Sens	arrow () Where an arrow and a set of the se
Temperature/dry-bulb Temperature ~	New Contraction of the second se	
Date		and a second and a second and a second
20/07/2023		the light way and the light wa
11:00 - 13:59	Annual Annua	The second secon
Level		A Construction of the second s
2.000m above ground		Aller and an and a second a s
Time range		
Analysis or observation, istantaneous value		men and a second s
Network		and the second s
Any ~		a call and a call and a call a c
Group of Licenses		
CCBY COMPLIANT ~		
Quality Control Filter		
•••••		
Update Map		

https://meteohub.mistralportal.it/app/maps/observations

Mistral: Data Extraction



How do we extract data in a continued, up to date and automated way?

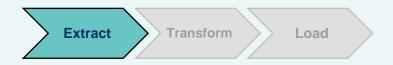


Mistral: Data Extraction



How do we extract data in a continued and automated way?





Mistral: Data Transformation



deda.next

EDIA



Mistral: Data Transformation



deda.next

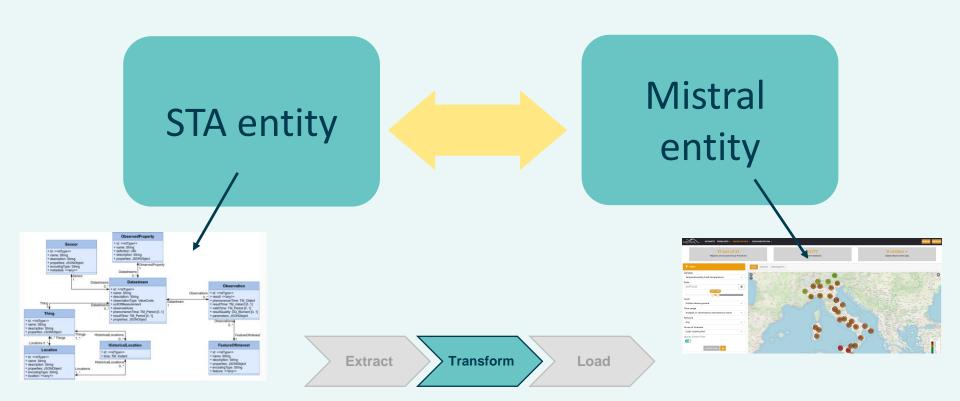
EDIA





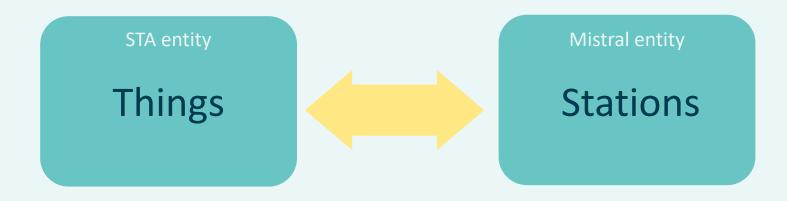
Mistral: Mapping





Mistral: Mapping

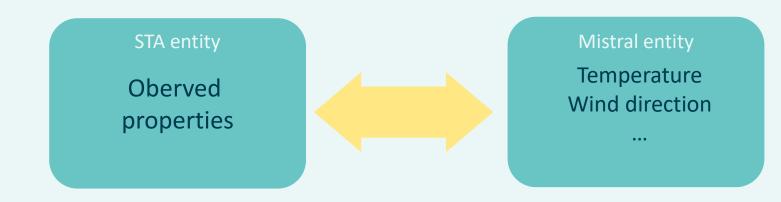


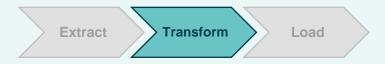




Mistral: Mapping

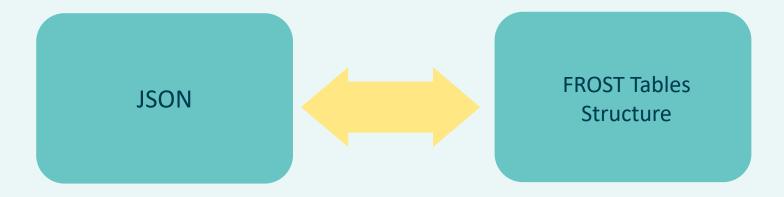






Mistral: Transformation



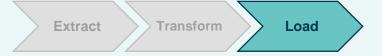




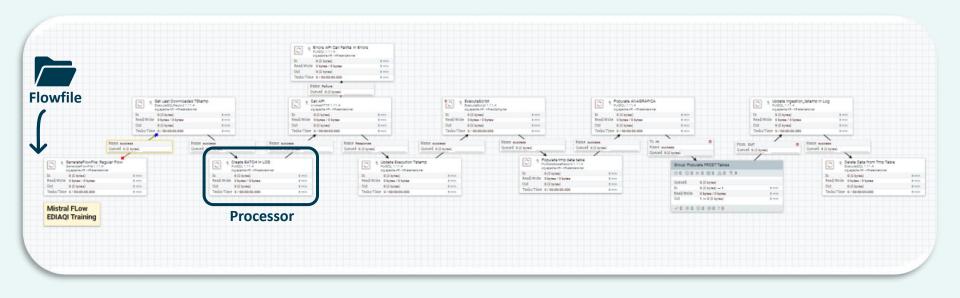
deda<mark>, next</mark> EDIA©I

Mistral: Load

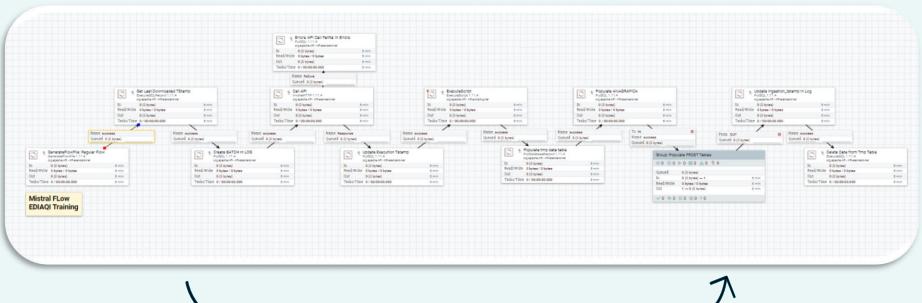
Pg Admin	File 🗸 Object 🗸 Tools 🖌 Help 🗸
Browser	ş 🎟 🖬
	✓ I Tables (18)
	> ACTUATORS
	> DATASTREAMS
	> E FEATURES
	HIST_LOCATIONS
	> ELOCATIONS
	ELOCATIONS_HIST_LOCATIONS
	MULTI_DATASTREAMS
	MULTI_DATASTREAMS_OBS_PROPERTIES
	> OBSERVATIONS
	> B OBS_PROPERTIES
	> ESENSORS
	> TASKINGCAPABILITIES
	> ETASKS
	>
	>



edia.next

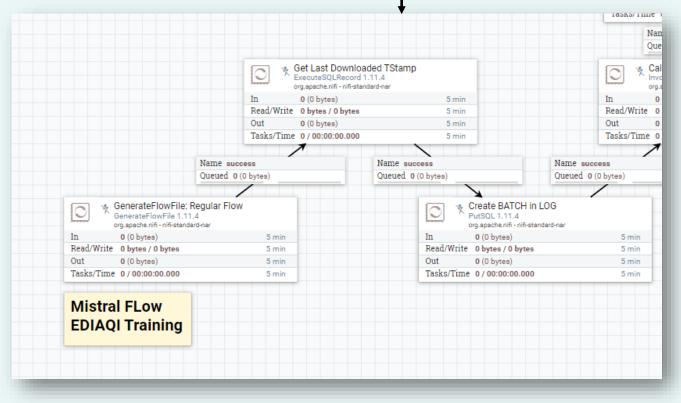


edia.next

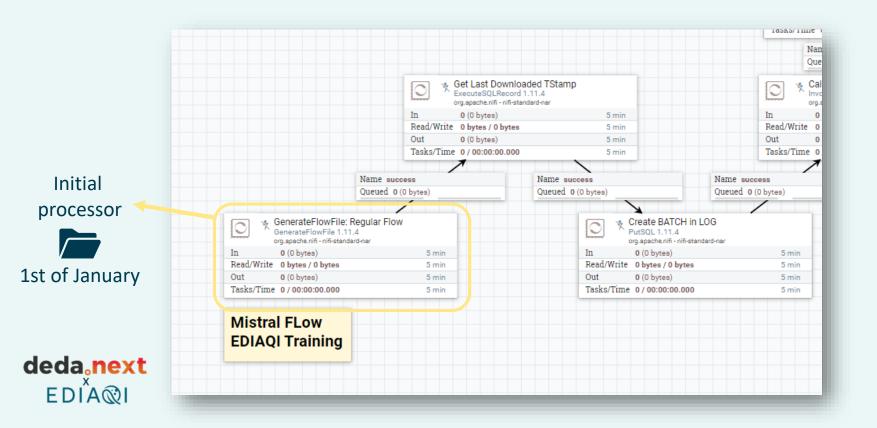


ightarrow 1 flow trip = 1 day of data ightarrow





deda, next



In the next rounds, how do we **keep track** of what day of data has already been downloaded?



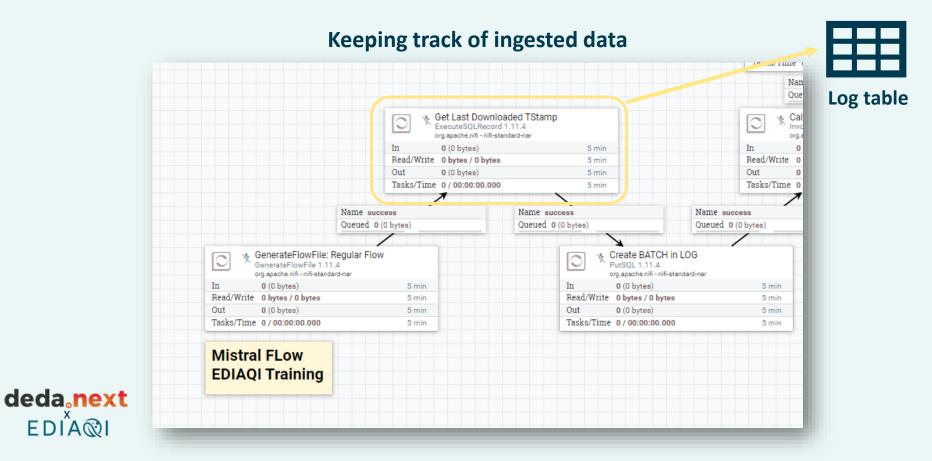


In the next rounds, how do we **keep track** of what day of data has already been downloaded?



Batch_ID	Start_Date	Stop_Date	API_Excecution_Tstamp	FROST_Ingestion_Tstamp



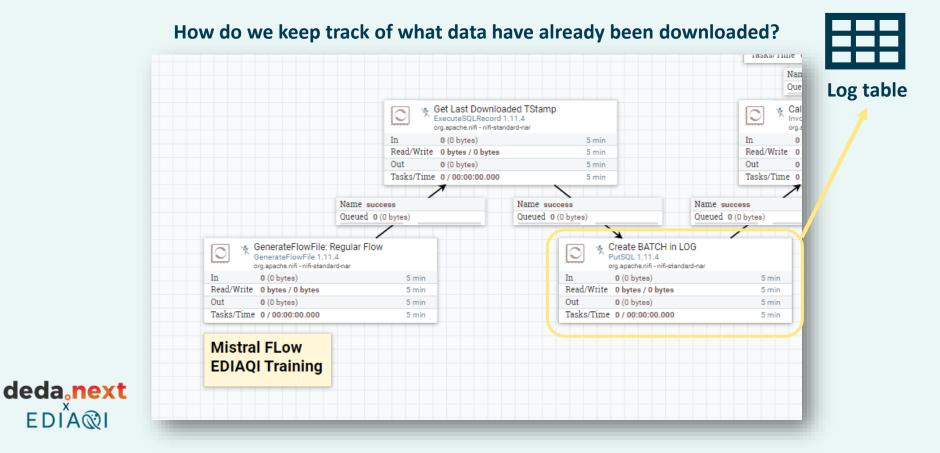


In the next rounds, how do we **keep track** of what day of data has already been downloaded?



Batch_ID	Start_Date	Stop_Date	API_Excecution_Tstamp	FROST_Ingestion_Tstamp
1	01/01/2023 00:00	01/01/2023 23:59	07/21/2023 10:00	07/21/2023 10:01





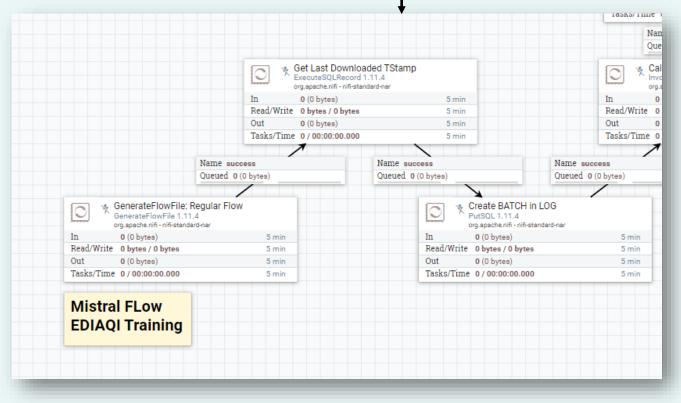
In the next rounds, how do we **keep track** of what day of data has already been downloaded?



Batch_ID	Start_Date	Stop_Date	API_Excecution_Tstamp	FROST_Ingestion_Tstamp
1	01/01/2023 00:00	01/01/2023 23:59	07/21/2023 10:00	07/21/2023 10:01
2	02/01/2023 00:00	02/01/2023 23:59		

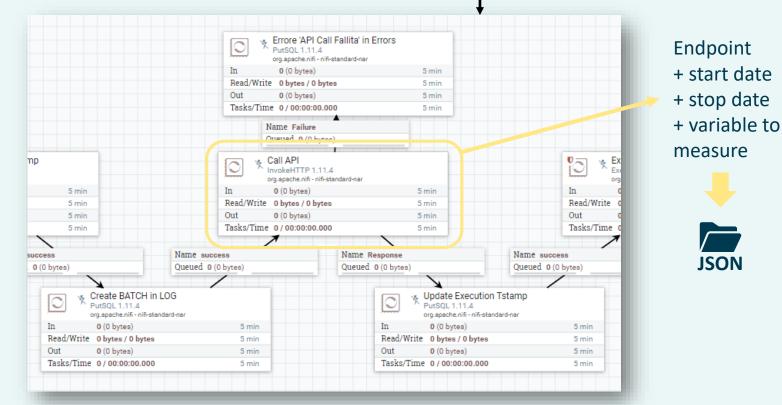






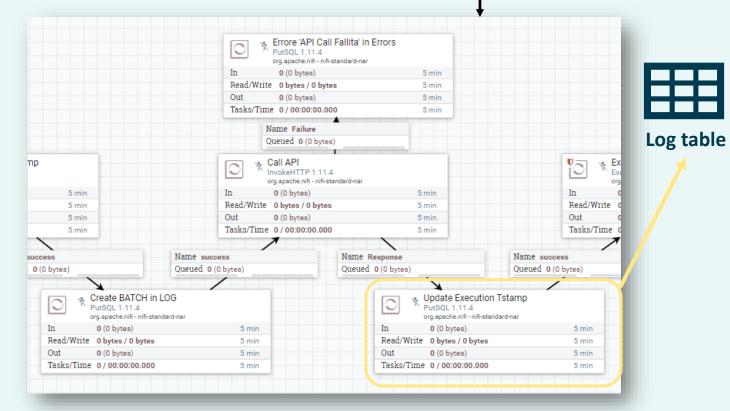
deda, next





deda, next





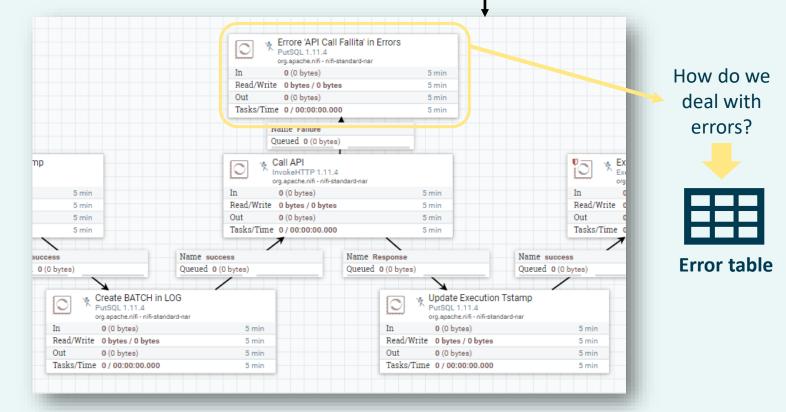
deda.next

Log table

Batch_ID	Start_Date	Stop_Date	API_Excecution_Tstamp	FROST_Ingestion_Tstamp
1	01/01/2023 00:00	01/01/2023 23:59	07/21/2023 10:00	07/21/2023 10:01
2	02/01/2023 00:00	02/01/2023 23:59	07/21/2023 10:30	







deda, next

How do you actually go from JSON data to FROST data?





	PutSQL 1	PI Call Fallita' in Errors	
	In 0 (0 byt	es) 5 min	
	Read/Write 0 bytes	/ 0 bytes 5 min	
	Out 0 (0 byt	es) 5 min	
	Tasks/Time 0 / 00:0	0:00.000 5 min	
	Name Fail Queued 0		
p	Call API InvokeHTT org.apache.n	P 1.11.4 ifi - nifi-standard-nar	
5 min	In 0 (0 bytes	s) 5 min	In
5 min	Read/Write 0 bytes /	0 bytes 5 min	Read/Write
5 min	Out 0 (0 bytes	s) 5 min	Out
5 min	Tasks/Time 0/00:00	:00.000 5 min	Tasks/Time
Iccess	Name success	Name Response	Name success
0 (0 bytes)	Queued 0 (0 bytes)	Queued 0 (0 bytes)	Queued 0 (0 bytes)
>		7	
Create BATCH in L PutSQL 1.11.4 org.apache.nifi - nifi-stan		Update Exect PutSQL 1.11.4 org.spache.nifi - n	
In 0 (0 bytes)	5 min	In 0 (0 bytes)	5 min
Read/Write 0 bytes / 0 bytes	5 min	Read/Write 0 bytes / 0 by	
Out 0 (0 bytes)	5 min	Out 0 (0 bytes)	5 min
Tasks/Time 0 / 00:00:00.000	5 min	Tasks/Time 0 / 00:00:00.	000 5 min

deda, next



Jython scritp

	ExecuteScript ExecuteScript 1.11.4 org.apache.nifi - nifi-scrip	ting-nar -		Populate ANAG PutSQL 1.11.4 org.spache.nifi - nifi-st		
	In 0 (0 bytes)	5 min		In O(Obytes)		5 min
	Read/Write 0 bytes / 0 bytes	5 min		Read/Write 0 bytes / 0 bytes	1	5 min
	Out 0 (0 bytes)	5 min		Out 0 (0 bytes)		5 min
	Tasks/Time 0 / 00:00:00.000	5 min		Tasks/Time 0 / 00:00:00.000)	5 min
	1			/		
ame		Name success	Name	success	To IN	
ueue	es)	Queued 0 (0 byt	Queue	d 0 (0 bytes)	Name suc	cess
^{IP} JSO		Put[oulate tmp data table DatabaseRecord 1.11.4 pache.nifi - nifi-standard-nar			Group: Popula
	5 min		(0 bytes)	5 min		0 🖉 0 🔍 0
	5 min	Read/Write 0	bytes / 0 bytes	5 min		0.1
	5 min	Out 0 ((0 bytes)	5 min		Queued
	5 min	Tasks/Time 0/	00:00:00.000	5 min		In Read/Write
						Out

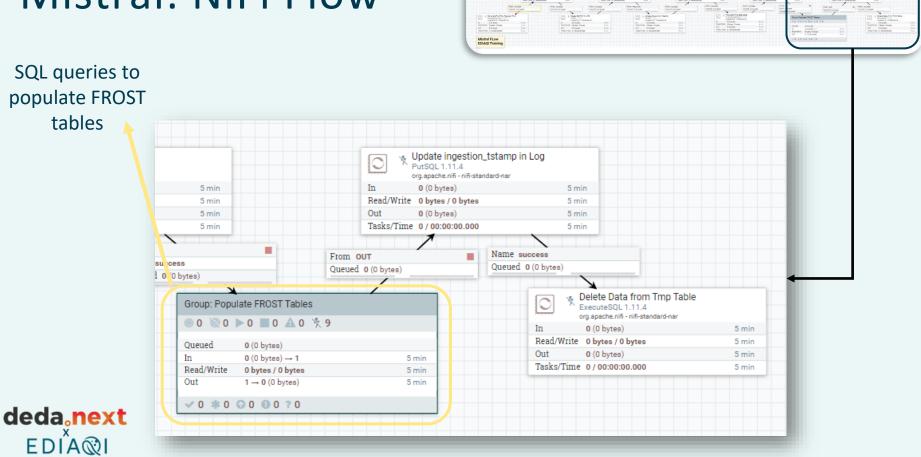
deda<mark>, next</mark> EDIA©I



Support DB table to temporarily store the data

		Populate A PutSQL 1.11 org.apache.nifi	AGRAFICA
In 0 (0 bytes)	5 min	In 0 (0 bytes)	
Tasks/Time 0/00:00:00.000	5 min	Tasks/Time 0/00:00:0	00.000 5 min
	N	News	To IN
ISS			Name success
bytes)	Queu d 0 (0 bytes)	Queuea 0 (0 bytes)	Queued 0 (0 bytes)
	PutDataba	servecord 1.11.4	Group: Popula
5 min			0 🖉 0 🔊
5 min	Read/Write 0 bytes /	0 bytes 5 min	Quand
5 min			Queued
5 min	Tasks/Time 0 / 00:00	0:00.000 5 min	In Read/Write
			Out
	Smin Smin	Smin Smin Smin Smin	ExecuteScript 1.11.4 org.spache.niff-niff-scripting-nar In 0 (0 bytes) Read/Win. 0 bytes / 0 bytes Smin 0 (0 bytes) Read/Win. 0 bytes / 0 bytes Smin 0 (0 bytes) Read/Wine 0 (0 bytes) Smin 0 (0 bytes) Smin 0 (0 bytes) Smin 0 (0 bytes) Sass Name success Dytes) Queurd 0 (0 bytes) Queurd 0 (0 bytes) Queued 0 (0 bytes) Smin Read/Write 0 bytes / 0 bytes Smin Tasks/Time 0 / 00:000

deda<mark>, next</mark> EDIA®I







deda<mark>, next</mark> EDIA©I

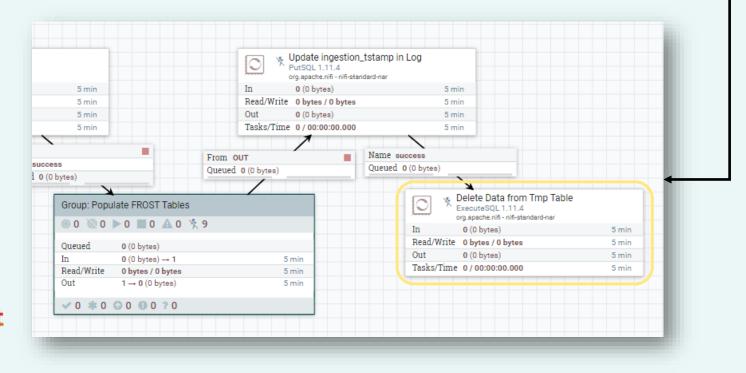
	Fulloge 1.1	gestion_tstamp in Log 1.4 ifi - nifi-standard-nar	
5 min	In 0 (0 bytes	s) 5 min	
5 min	Read/Write 0 bytes /	0 bytes 5 min	
5 min	Out 0 (0 bytes		
5 min	Tasks/Time 0 / 00:00	:00.000 5 min	
cess O bytes)	From OUT Queued 0 (0 bytes)	Queued 0 (0 bytes)	ble
Group: Populate FROST		ExecuteSQL 1.11.4 org.apache.nifi - nifi-standard-nar	
◎ 0 🔍 0 🕨 0 🔳 0 🖉	NO X 9	In 0 (0 bytes)	5 min
Queued 0 (0 bytes)		Read/Write 0 bytes / 0 bytes	5 min
In 0 (0 bytes) -	+1 5 min	Out 0 (0 bytes)	5 min
Read/Write 0 bytes / 0 b		Tasks/Time 0 / 00:00:00.000	5 min
Out 1 → 0 (0 byt	es) 5 min		
	2 0		

Log table

Batch_ID	Start_Date	Stop_Date	API_Excecution_Tstamp	FROST_Ingestion_Tstamp
1	01/01/2023 00:00	01/01/2023 23:59	07/21/2023 10:00	07/21/2023 10:01
2	02/01/2023 00:00	02/01/2023 23:59	07/21/2023 10:30	07/21/2023 10:32







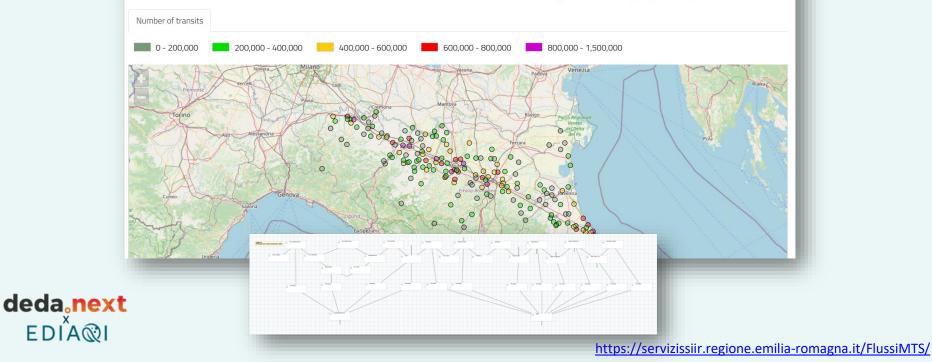
deda<mark>, next</mark> EDIA©I

Other examples: road traffic data

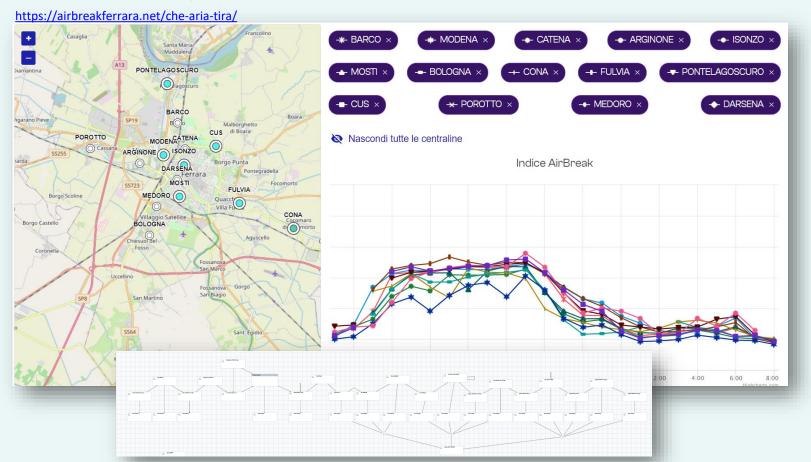
Online streams

The Regional Road Traffic Flow Monitoring System (MTS) of Emilia-Romagna, created by the Region itself, the Provinces and Anas, is made up of 283 stations, operating 24 hours a day, mainly installed on state and provincial roads.

The consultation and download of the traffic flows detected allow the use of the data recorded by the MTS System, managed by the Traffic, Logistics, Waterways and Airports Area.



Other examples: outdoor AQ data



deda.next

EDIA®I