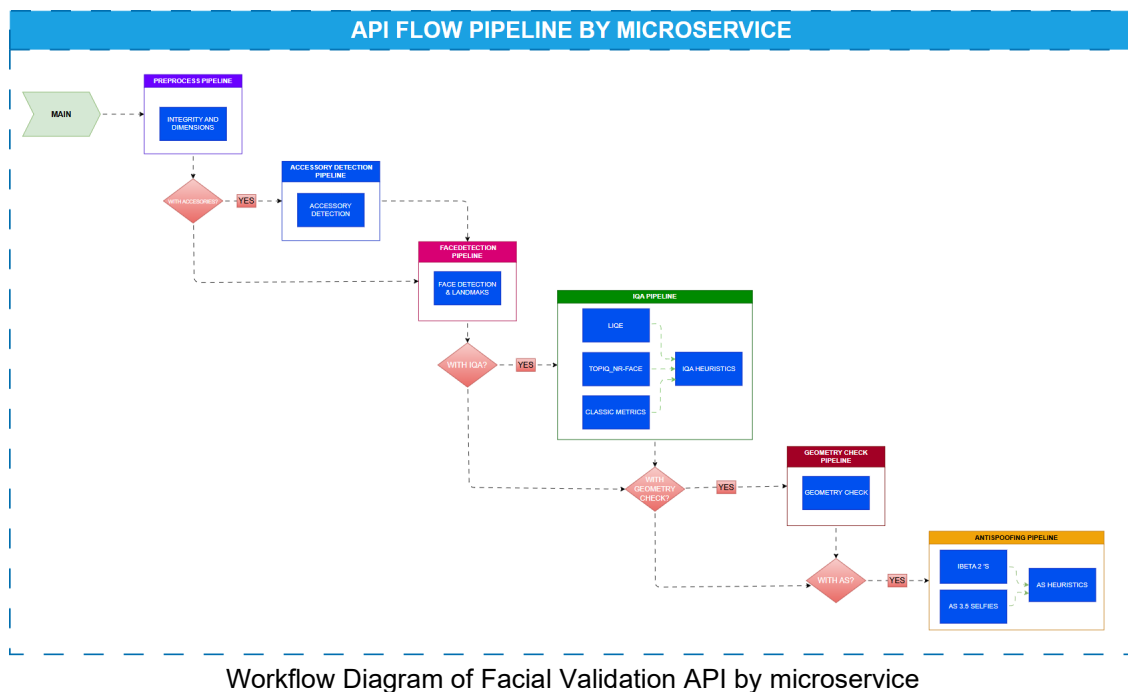


## ETL Final Workshop - Manuel Henao - [Github Repo](#)

**Problem Description:** The project aims to analyze the performance of a facial validation API using an ETL process. The API, built on a microservices architecture, includes face detection, accessory detection, image quality assessment, and Presentation Attack Detection (PAD) to prevent identity spoofing. Performance data is stored in a non-relational database. The primary objectives are to identify execution time bottlenecks of each microservice, analyze acceptance rates per client, and improve user experience by examining retry patterns.

**Context:** The facial validation API is designed to ensure secure and efficient identity verification by capturing two moments: one where the user is near the camera and another where the user is farther away. This process, known as liveness detection, aims to enhance the performance of the PAD microservice. Each microservice will have execution metrics for both images. The API comprises several microservices:

- Face Detection: Identifies and locates faces in images.
- Accessory Detection: Detects accessories like glasses or masks that might obscure the face.
- Image Quality Assessment: Evaluates the quality of the image to ensure it meets required standards using various methods.
- Presentation Attack Detection (PAD): Detects attempts to spoof the system using photos, videos, or masks.



The API's performance data is stored in a non-relational database using MongoDB, capturing metrics such as execution times, acceptance rates, and retry counts for two attempts. The pipeline workflow executes the microservices in a serial manner. For example, if the accessory detection microservice detects an accessory, it will terminate the transaction without processing the subsequent microservices.

**Dataset Description:** The dataset includes various performance metrics for each microservice. Below is a model card for the variables of interest:

feature	description	unit
api_hash	API Version	string
user_name	Username consuming the API	string
file_token	Unique identifier (hash) of the transaction	string
service	API endpoint consumed (Service)	string
summary_code	Summary response code	int
summary_status	Summary status	string
summary_desc	Summary description	string
http_code	HTTP code of the transaction	int
total_time_ms	Total transaction time in ms	ms
date	Date	date
procesoConvenioGuid	GUID of the agreement process	string
convenio	Agreement (Client ID)	string
documento	Document Identifier	string
img_size_near_img_width	Width of the near image	px
img_size_near_img_height	Height of the near image	px
img_size_far_img_width	Width of the far image	px
img_size_far_img_height	Height of the far image	px
score_liqe_near_img	LIQE microservice score of the near image	double
score_liqe_far_img	LIQE microservice score of the far image	double
score_topiq_near_img	TOPIQ microservice score of the near image	double
score_topiq_far_img	TOPIQ microservice score of the far image	double
score_sharpness_classic_metrics_near_img	Classic sharpness microservice score of the near image	double
score_sharpness_classic_metrics_far_img	Classic sharpness microservice score of the far image	double
score_colorfulness_classic_metrics_near_img	Classic colorfulness microservice score of the near image	double
score_colorfulness_classic_metrics_far_img	Classic colorfulness microservice score of the far image	double
score_contrast_classic_metrics_near_img	Classic contrast microservice score of the near image	double
score_contrast_classic_metrics_far_img	Classic contrast microservice score of the far image	double
score_brightness_classic_metrics_near_img	Classic brightness microservice score of the near image	double
score_brightness_classic_metrics_far_img	Classic brightness microservice score of the far image	double
score_blur_score_classic_metrics_near_img	Classic blur microservice score of the near image	double
score_blur_score_classic_metrics_far_img	Classic blur microservice score of the far image	double
score_svd_score_classic_metrics_near_img	Classic SVD microservice score of the near image	double
score_svd_score_classic_metrics_far_img	Classic SVD microservice score of the far image	double
prob_as_35_selfies_near_img	Probability as 3.5 selfies microservice for the near image	double
prob_as_35_selfies_far_img	Probability as 3.5 selfies microservice for the far image	double
prob_ibeta2_crops_near_img	Probability ibeta2 crops microservice for the near image	double
prob_ibeta2_crops_far_img	Probability ibeta2 crops microservice for the far image	double
prob_ibeta2_full_near_img	Probability ibeta2 full microservice for the near image	double
prob_ibeta2_full_far_img	Probability ibeta2 full microservice for the far image	double
prob_ibeta2_clip_near_img	Probability ibeta2 clip microservice for the near image	double
prob_ibeta2_clip_far_img	Probability ibeta2 clip microservice for the far image	double
prob_as_heuristics_near_img	Heuristic probability microservice for the near image	double
prob_as_heuristics_far_img	Heuristic probability microservice for the far image	double
time_ms_accessory_detector_near_img	Time in ms of accessory detector microservice for the near image	ms
time_ms_accessory_detector_far_img	Time in ms of accessory detector microservice for the far image	ms
queue_time_ms_accessory_detector_near_img	Queue time in ms of accessory detector microservice for the near image	ms
queue_time_ms_accessory_detector_far_img	Queue time in ms of accessory detector microservice for the far image	ms
time_ms_face_detector_near_img	Time in ms of face detector microservice for the near image	ms
time_ms_face_detector_far_img	Time in ms of face detector microservice for the far image	ms
queue_time_ms_face_detector_near_img	Queue time in ms of face detector microservice for the near image	ms
queue_time_ms_face_detector_far_img	Queue time in ms of face detector microservice for the far image	ms
time_ms_liqe_near_img	Time in ms of LIQE microservice for the near image	ms
time_ms_liqe_far_img	Time in ms of LIQE microservice for the far image	ms
queue_time_ms_liqe_near_img	Queue time in ms of LIQE microservice for the near image	ms
queue_time_ms_liqe_far_img	Queue time in ms of LIQE microservice for the far image	ms
time_ms_topiq_near_img	Time in ms of TOPIQ microservice for the near image	ms
time_ms_topiq_far_img	Time in ms of TOPIQ microservice for the far image	ms
queue_time_ms_topiq_near_img	Queue time in ms of TOPIQ microservice for the near image	ms
queue_time_ms_topiq_far_img	Queue time in ms of TOPIQ microservice for the far image	ms
time_ms_classic_metrics_near_img	Time in ms of classic metrics microservice for the near image	ms
time_ms_classic_metrics_far_img	Time in ms of classic metrics microservice for the far image	ms
time_ms_as_35_selfies_near_img	Time in ms as 3.5 selfies for the near image	ms
time_ms_as_35_selfies_far_img	Time in ms as 3.5 selfies for the far image	ms
queue_time_ms_as_35_selfies_near_img	Queue time in ms as 3.5 selfies for the near image	ms
queue_time_ms_as_35_selfies_far_img	Queue time in ms as 3.5 selfies for the far image	ms
time_ms_ibeta2_crops_near_img	Time in ms of ibeta2 crops microservice for the near image	ms
time_ms_ibeta2_crops_far_img	Time in ms of ibeta2 crops microservice for the far image	ms
queue_time_ms_ibeta2_crops_near_img	Queue time in ms of ibeta2 crops microservice for the near image	ms
queue_time_ms_ibeta2_crops_far_img	Queue time in ms of ibeta2 crops microservice for the far image	ms
time_ms_ibeta2_full_near_img	Time in ms of ibeta2 full microservice for the near image	ms
time_ms_ibeta2_full_far_img	Time in ms of ibeta2 full microservice for the far image	ms
queue_time_ms_ibeta2_full_near_img	Queue time in ms of ibeta2 full microservice for the near image	ms
queue_time_ms_ibeta2_full_far_img	Queue time in ms of ibeta2 full microservice for the far image	ms
time_ms_ibeta2_clip_near_img	Time in ms of ibeta2 clip microservice for the near image	ms
time_ms_ibeta2_clip_far_img	Time in ms of ibeta2 clip microservice for the far image	ms
queue_time_ms_ibeta2_clip_near_img	Queue time in ms of ibeta2 clip microservice for the near image	ms
queue_time_ms_ibeta2_clip_far_img	Queue time in ms of ibeta2 clip microservice for the far image	ms
time_ms_geometry_check	Time in ms of geometry check microservice	ms
queue_time_ms_geometry_check	Queue time in ms of geometry check microservice	ms

The ETL Workshop will be broken down into 3 main stages

- ## Evidences

## Loading / Reading Staging Table

```

db_table_name = "etl_staging_table_final_workshop"
with engine.connect() as conn:
    of_staging = pd.read_sql("SELECT * FROM {db_table_name}", conn)

    print(of_staging.info())
    print(of_staging.describe())
    of_staging.head()

```

File Edit View Help Staging in Data Manager

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14654 entries, 0 to 14653
Data columns (total 79 columns):
 #   Column                                Non-Null Count  Dtype
---  --
 0   wd_login                             14654 non-null  object
 1   user_name                           14654 non-null  object
 2   file_name                           14654 non-null  object
 3   service                             14654 non-null  object
 4   summary_code                        14654 non-null  int64
 5   summary_status                     14654 non-null  object
 6   summary_desc                       14654 non-null  object
 7   http_code                          14654 non-null  int64
 8   total_time_ms                      14654 non-null  int64
 9   date                               14654 non-null  object
10  processConvertedId                 14654 non-null  object
11  commenta                          14654 non-null  object
12  documents                          14654 non-null  object
13  img_size_near_img_width           14654 non-null  int64
14  img_size_near_img_height          14654 non-null  int64
15  img_size_far_img_width            14654 non-null  int64
16  img_size_far_img_height           14654 non-null  int64
17  score_list_near_img               13953 non-null  float64
18  score_list_far_img                13953 non-null  float64
19  score_topics_near_img             13953 non-null  float64
...
75%      1.000000
max      161.000000

```

[8 rows x 69 columns]

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

	#	wd_login	user_name	file_name	service	#	summary_code	summary_status	summary_desc	#	http_code	#	total_time_ms	date
0	020de4	user_name_1	850c204g	main		1400	ok	antispooofing OK		200		2204	2025-01-08T22:10:35.642847-05:00	
1	020de4	user_name_1	D05m7QhA	main		1400	ok	antispooofing OK		200		1937	2025-01-08T22:14:12.833131-05:00	
2	020de4	user_name_1	8Rqj4wQ	main		1400	ok	antispooofing OK		200		1958	2025-01-08T22:14:13.781434-05:00	
3	020de4	user_name_1	C6202qjg	main		1400	ok	antispooofing OK		200		1640	2025-01-08T22:18:14.887930-05:00	
4	020de4	user_name_1	7RqUqfag	main		302	error	blur/hex		200		919	2025-01-08T22:20:07.026485-05:00	

## pgAdmin 4

ObjectToolsEditViewWindowHelp

Welcomeetl\_db\_final\_workshop/postgres@PostgreSQL 17\* X

etl\_db\_final\_workshop/postgres@PostgreSQL 17

No limit

QueryQuery History

Scratch Pad X

1SELECT \* FROM etl\_staging.table\_final\_workshop;

2

Data OutputMessagesNotifications

Showing rows: 1 to 10000Page No: 1of 2

	ash_bush	user_name	file_token	service	summary_code	summary_status	summary_desc	http_code	total_time_ms	date	processo	Convenio	Guid	convenio	documento
	text	text	text	text	integer	text	text	integer	integer	text	text	text	text	text	text
1	02bc5e7	user_name_1	HLqbP2M...	main	1400	ok	antispoofting OK	200	2204	2025-01-08T22:10:36.542487-05:00	processoConvenioGuid_1	convenio_1	documento_1		
2	02bc5e7	user_name_1	D0k1wYCP...	main	1400	ok	antispoofting OK	200	1837	2025-01-08T22:14:12.65331-05:00	processoConvenioGuid_2	convenio_1	documento_2		
3	02bc5e7	user_name_1	5f6lpp7e4Q...	main	1400	ok	antispoofting OK	200	1936	2025-01-08T22:16:10.768124-05:00	processoConvenioGuid_2	convenio_1	documento_2		
4	02bc5e7	user_name_1	CalSZGg2...	main	1400	ok	antispoofting OK	200	1640	2025-01-08T22:16:14.887930-05:00	processoConvenioGuid_2	convenio_1	documento_2		
5	02bc5e7	user_name_1	7fRlU4FLag...	main	302	error	blur/haze	200	919	2025-01-08T22:28:01.054463-05:00	processoConvenioGuid_3	convenio_1	documento_3		
6	02bc5e7	user_name_1	wyZZWD...	main	302	error	blur/haze	200	763	2025-01-08T22:28:13.094088-05:00	processoConvenioGuid_3	convenio_1	documento_3		
7	02bc5e7	user_name_1	sWKmZqht...	main	302	error	blur/haze	200	699	2025-01-08T22:28:26.157333-05:00	processoConvenioGuid_3	convenio_1	documento_3		
8	02bc5e7	user_name_1	XlZxVUa7iQ...	main	406	error	non-fragmental	200	692	2025-01-08T22:41:38.050918-05:00	processoConvenioGuid_4	convenio_1	documento_4		
9	02bc5e7	user_name_1	zIdYSlG4g...	main	302	error	blur/haze	200	925	2025-01-08T22:45:42.796236-05:00	processoConvenioGuid_5	convenio_1	documento_3		
10	02bc5e7	user_name_1	jRaExYlrg...	main	203	error	hat	200	572	2025-01-08T22:49:46.404009-05:00	processoConvenioGuid_6	convenio_1	documento_5		
11	02bc5e7	user_name_1	RaMehL...	main	1400	ok	antispoofting OK	200	1615	2025-01-08T22:49:57.402437-05:00	processoConvenioGuid_6	convenio_1	documento_5		
12	02bc5e7	user_name_1	fR3aypQD...	main	1400	ok	antispoofting OK	200	1730	2025-01-08T22:50:00.843492-05:00	processoConvenioGuid_6	convenio_1	documento_5		
13	02bc5e7	user_name_1	HZDsox6H...	main	302	error	blur/haze	200	805	2025-01-08T22:51:03.447437-05:00	processoConvenioGuid_7	convenio_1	documento_3		
14	02bc5e7	user_name_1	ohtq3UpQ...	main	302	error	blur/haze	200	795	2025-01-08T22:51:40.932717-05:00	processoConvenioGuid_7	convenio_1	documento_3		
15	02bc5e7	user_name_1	c1eVQwp...	main	1400	ok	antispoofting OK	200	1662	2025-01-08T22:52:49.485790-05:00	processoConvenioGuid_7	convenio_1	documento_3		
16	02bc5e7	user_name_1	KAXqDRL...	main	1400	ok	antispoofting OK	200	1630	2025-01-08T22:53:09.765574-05:00	processoConvenioGuid_7	convenio_1	documento_3		
17	02bc5e7	user_name_1	X3l3MAd...	main	1400	ok	antispoofting OK	200	1667	2025-01-08T22:54:52.561405-05:00	processoConvenioGuid_8	convenio_1	documento_3		
18	02bc5e7	user_name_1	6LVmex4...	main	1400	ok	antispoofting OK	200	1654	2025-01-08T22:54:55.947408-05:00	processoConvenioGuid_8	convenio_1	documento_3		

Total rows: 14654Query complete 00:00:59.6

CRUF Ln 2, Col