

TWITTER SENTIMENT ANALYSIS SERVICE

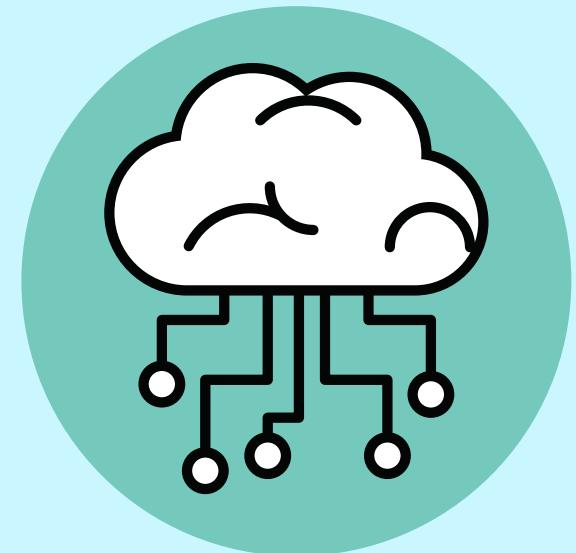
BIG DATA FOR OFFICIAL STATISTICS

**MATTEO BROGLIO
1200P39**

01

INTRODUCTION

- SENTIMENT ANALYSIS
- MAIN TECHNICAL CONCEPTS



SENTIMENT ANALYSIS

SENTIMENT ANALYSIS IS THE USE OF NATURAL LANGUAGE PROCESSING, TEXT ANALYSIS, COMPUTATIONAL LINGUISTICS, AND BIOMETRICS TO SYSTEMATICALLY IDENTIFY, EXTRACT, QUANTIFY, AND STUDY AFFECTIVE STATES AND SUBJECTIVE INFORMATION



Negative



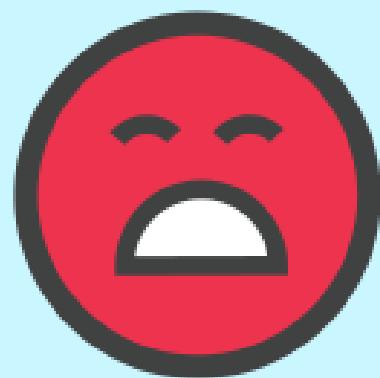
Neutral



Positive

S.A. ON TWITTER

ANALYZE GLOBAL
OPINION ABOUT A
SPECIFIC TOPIC
THROUGH TWEETS



Negative



Neutral

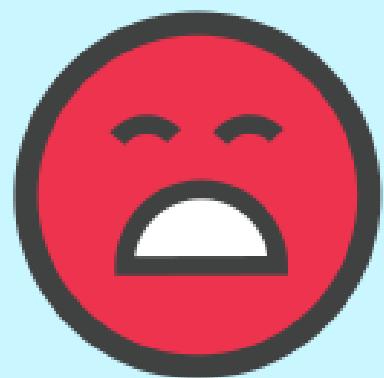


Positive

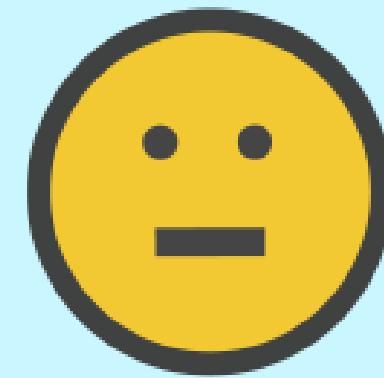


WHO NEEDS THIS?

- COMPANIES: REPUTATION MONITORING
- PUBLIC ORGANIZATIONS: (GOVERNMENTS, INSTITUTIONS, ETC) INSTITUTIONAL TOPICS MONITORING



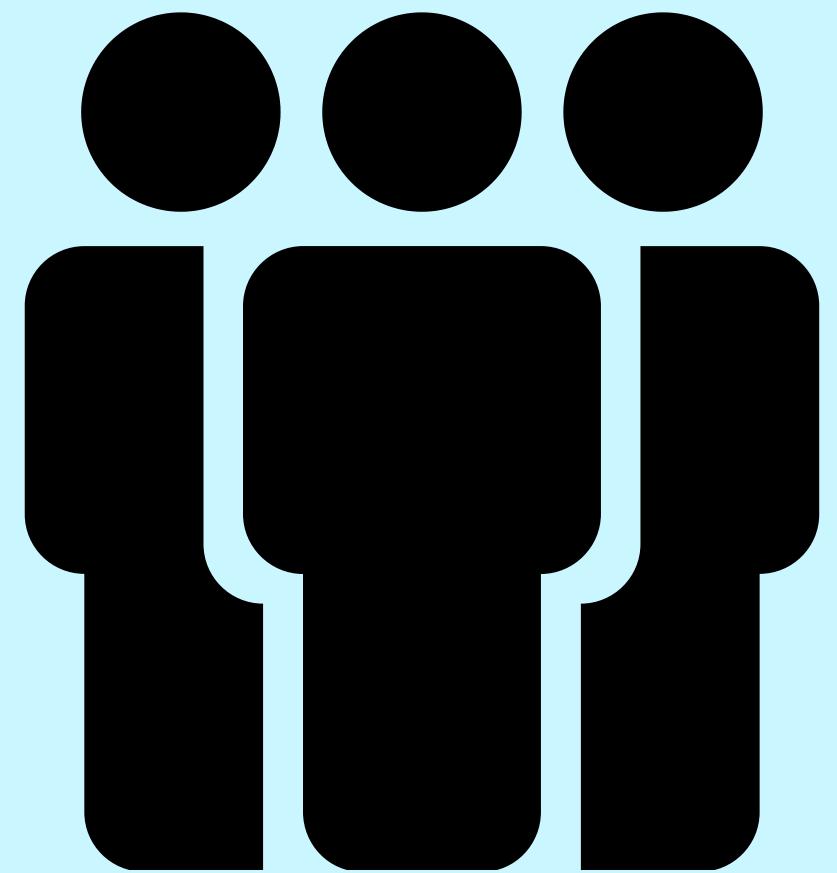
Negative



Neutral



Positive



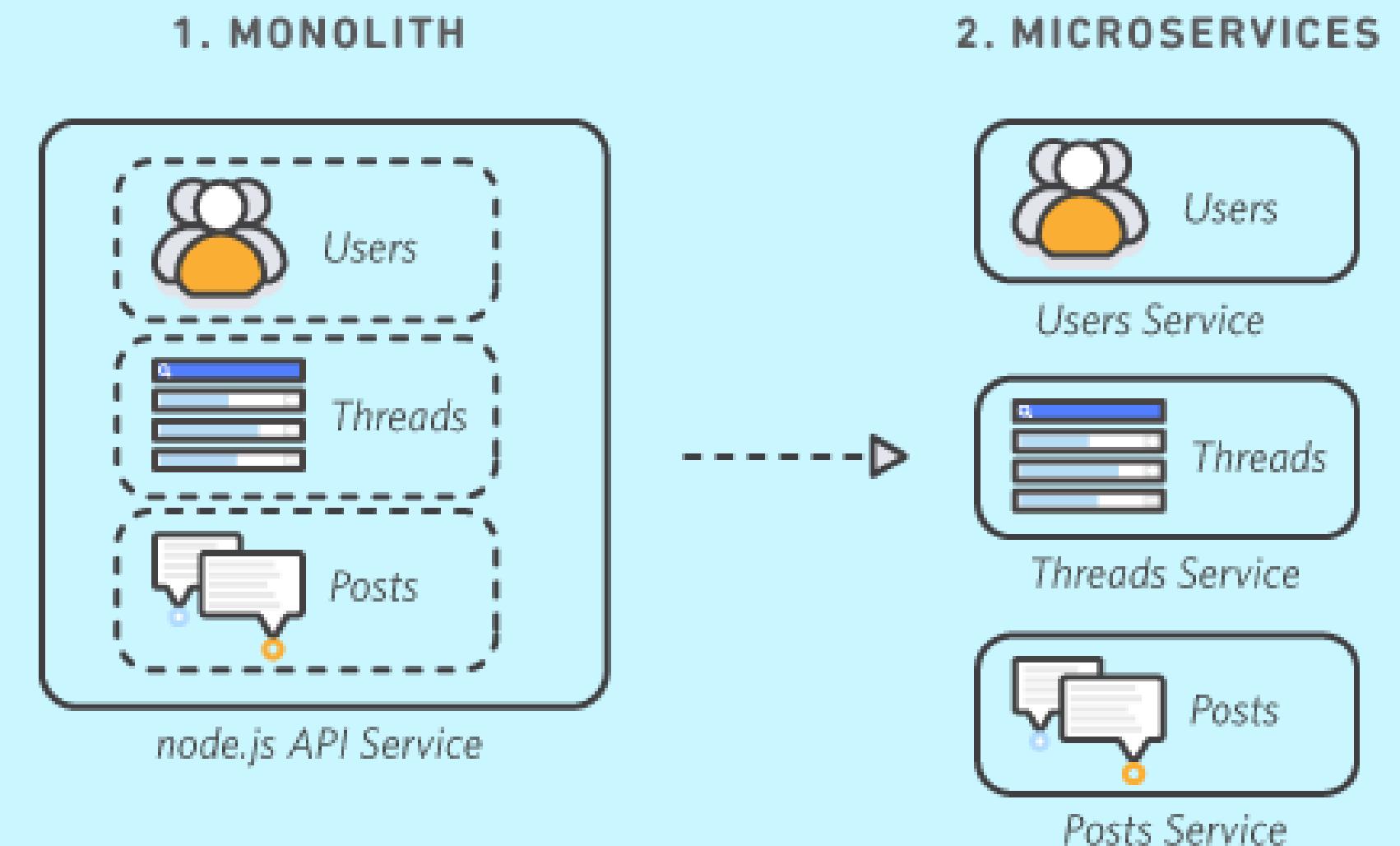
MAIN TECHNICAL CONCEPTS

- MICRO SERVICE
- REST API
- WEB SERVICE
- CLOUD SERVICE
- PYTHON LIBRARY
- DBMS

MICRO SERVICE

IS AN ARCHITECTURAL STYLE
THAT STRUCTURES AN
APPLICATION AS A COLLECTION
OF SERVICES THAT ARE:

- HIGHLY MAINTAINABLE AND TESTABLE
- LOOSELY COUPLED
- INDEPENDENTLY DEPLOYABLE
- ORGANIZED AROUND BUSINESS CAPABILITIES
- OWNED BY A SMALL TEAM

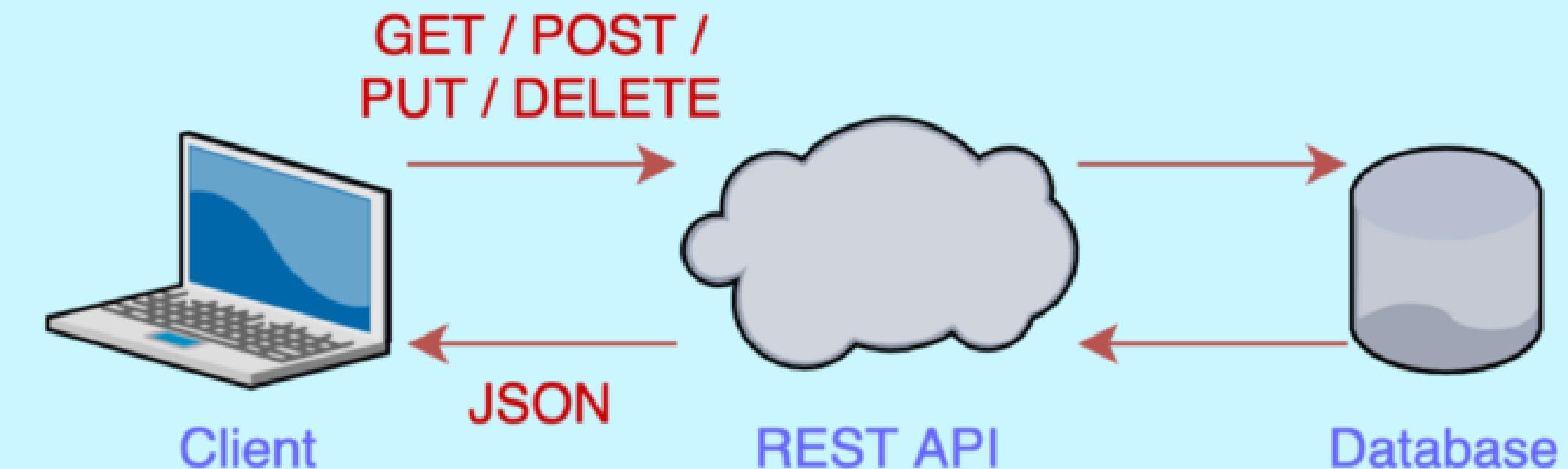


REST API

IS A SOFTWARE ARCHITECTURAL STYLE THAT WAS CREATED TO GUIDE THE DESIGN AND DEVELOPMENT OF THE ARCHITECTURE FOR THE WORLD WIDE WEB

PROPERTIES:

- CLIENT-SERVER ARCHITECTURE
- STATELESSNESS
- CACHEABILITY
- LAYERED SYSTEM



WEB FRAMEWORK

SOFTWARE FRAMEWORK
DESIGNED TO SUPPORT
THE DEVELOPMENT OF
WEB APPLICATIONS
INCLUDING WEB SERVICES,
WEB RESOURCES, AND
WEB APIs.



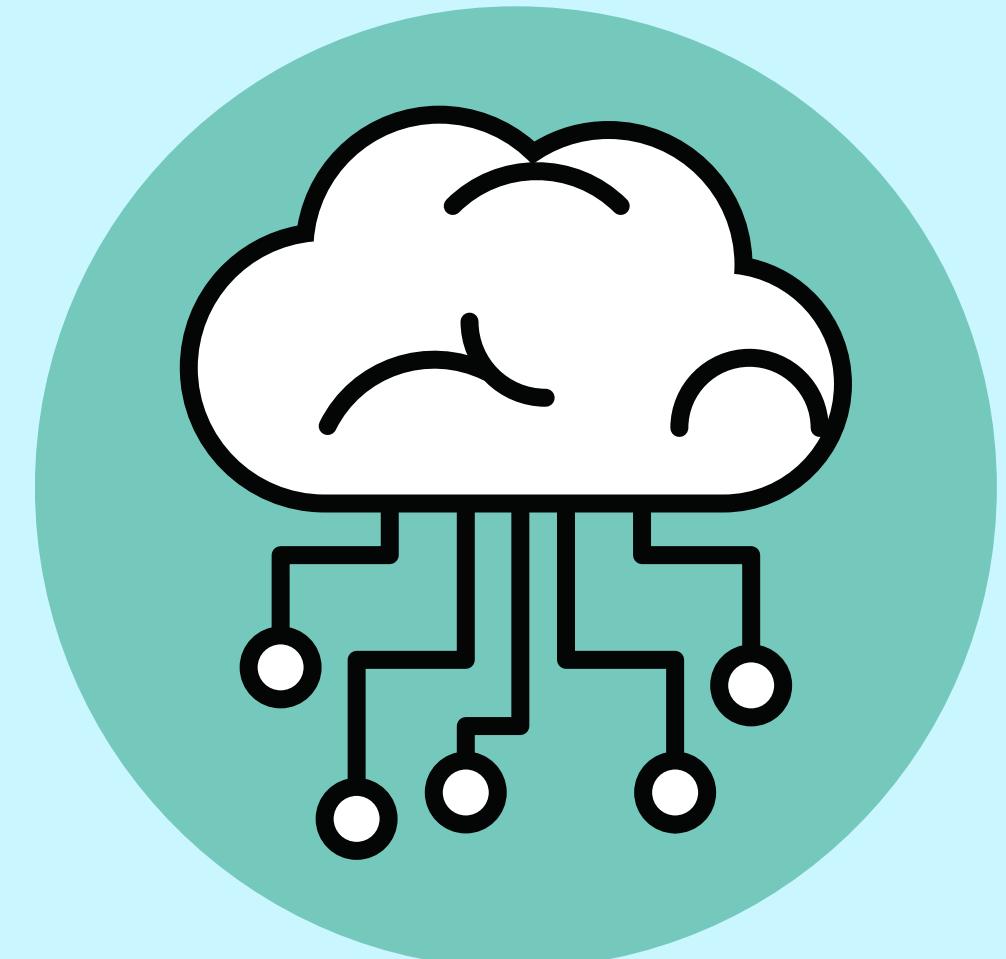


FastAPI

- PYTHON WEB FRAMEWORK
FOR BUILDING REST API
- ASGI (ASYNCHRONOUS
SERVER GATEWAY
INTERFACE)

CLOUD SERVICE

CLOUD SERVICES REFER TO ANY I.T. SERVICES THAT ARE PROVISIONED AND ACCESSED FROM A CLOUD COMPUTING PROVIDER



CLOUD COMPUTING

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" **NIST**



CLOUD COMPUTING

ESSENTIAL CHARACTERISTICS:

- ON-DEMAND SELF-SERVICE
- BROAD NETWORK ACCESS
- RESOURCE POOLING
- RAPID ELASTICITY
- MEASURED SERVICE



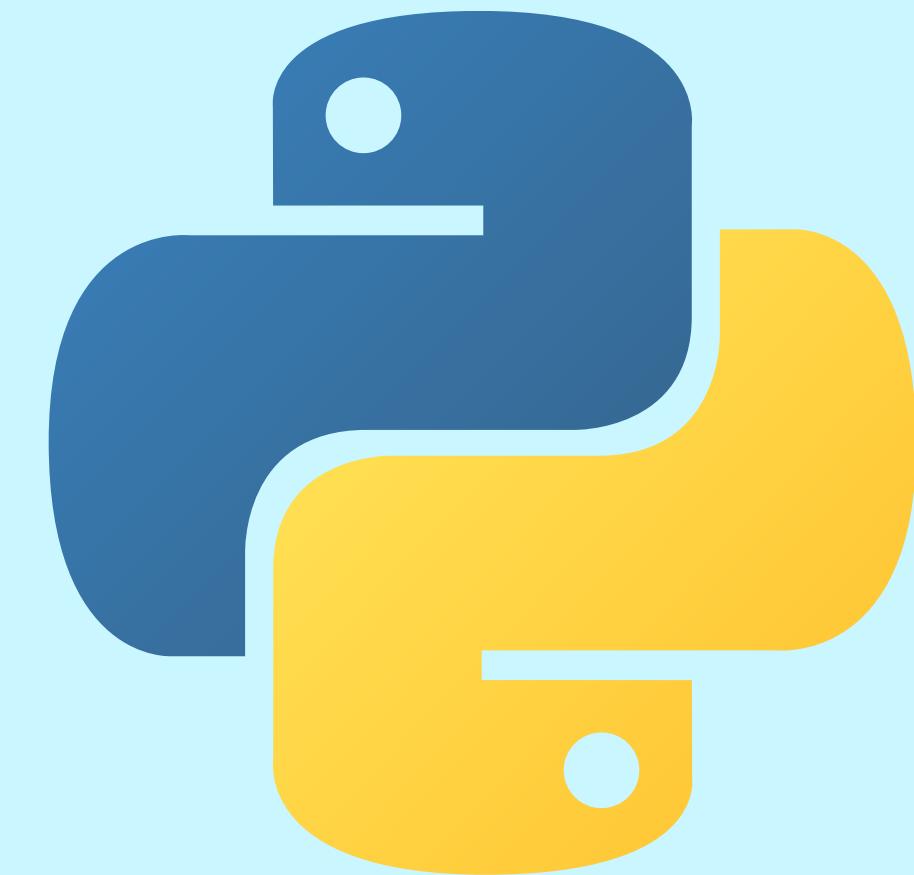
AMAZON WEB SERVICES

- SUBSIDIARY OF AMAZON
- ONE OF THE MAIN CLOUD COMPUTING PROVIDERS
- SINCE 2006



PYTHON LIBRARY

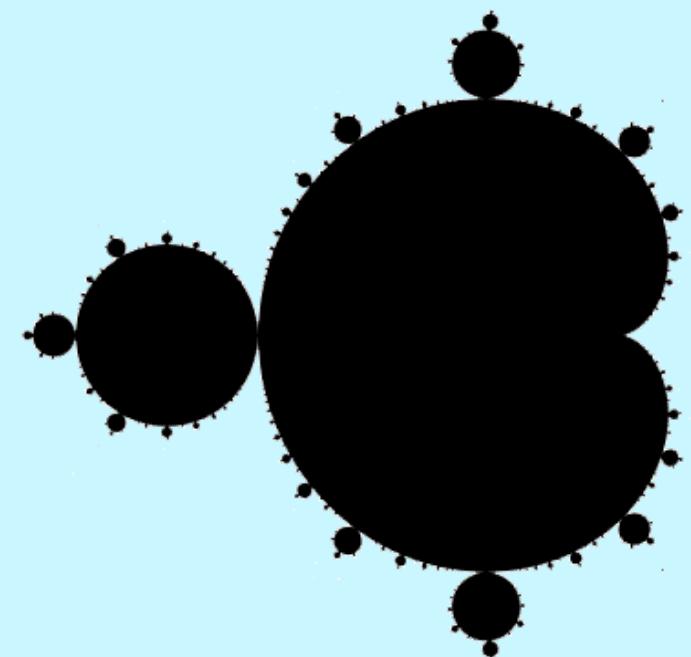
A PYTHON LIBRARY IS A COLLECTION OF MODULES, I.E. REUSABLE CHUNK OF CODE, GENERALLY DEVELOPED TO SOLVE SPECIFIC TASKS.



TEXT BLOB

PYTHON LIBRARY FOR:

- PART-OF-SPEECH TAGGING
- NOUN PHRASE EXTRACTION
- SENTIMENT ANALYSIS
- CLASSIFICATION
- TRANSLATION
- MORE

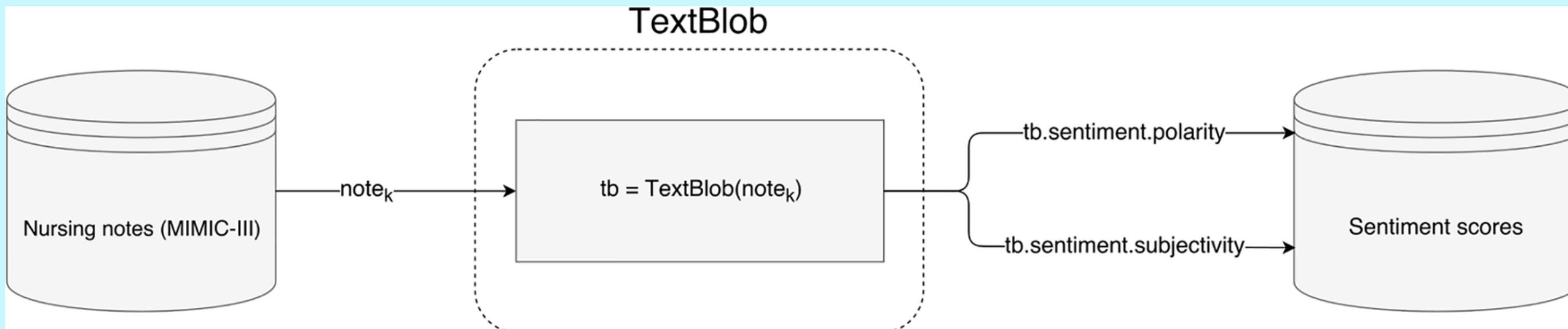


TextBlob

MODEL: TEXTBLOB

PROS: EASY DEPLOYMENT; FAST AND SIMPLE

CONS: LOW ACCURACY



DASH

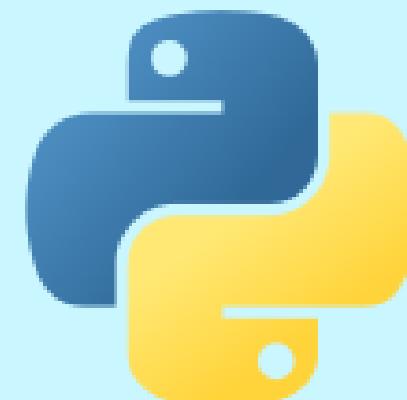
PYTHON LIBRARY
FOR INTERACTIVE
DASHBOARDS



plotly | Dash

TWEepy

AN EASY-TO-USE
PYTHON LIBRARY
FOR ACCESSING
THE TWITTER API



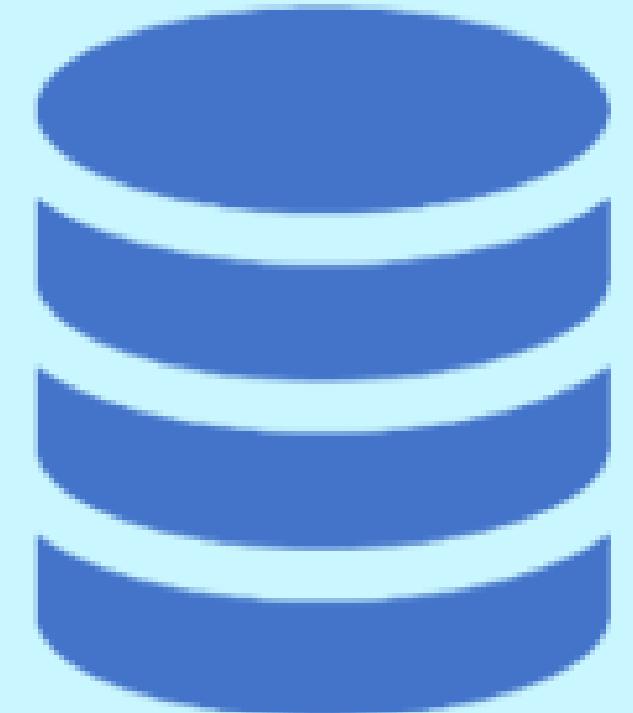
tweepyTM



RELATIONAL DATABASE

A RELATIONAL DATABASE IS A DIGITAL DATABASE BASED ON THE RELATIONAL MODEL OF DATA

THIS MODEL ORGANIZES DATA INTO ONE OR MORE TABLES (OR "RELATIONS") OF COLUMNS AND ROWS, WITH A UNIQUE KEY IDENTIFYING EACH ROW



MYSQL

DATABASE MANAGEMENT SYSTEM
(DBMS) I.E. A SOFTWARE DESIGNED TO
ENABLE EFFICIENT CREATION,
MANIPULATION AND QUERYING OF
DATABASES

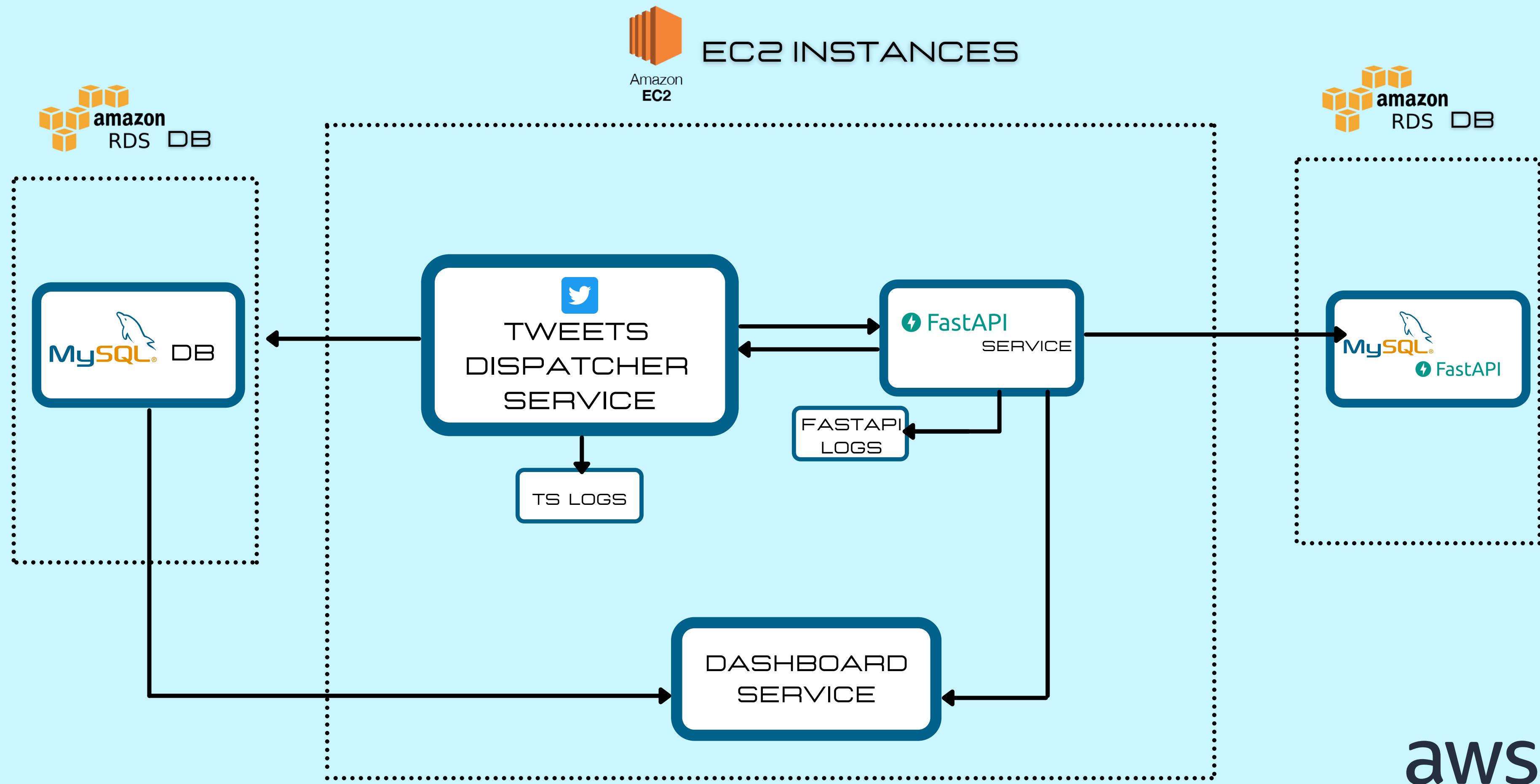


02

ARCHITECTURE

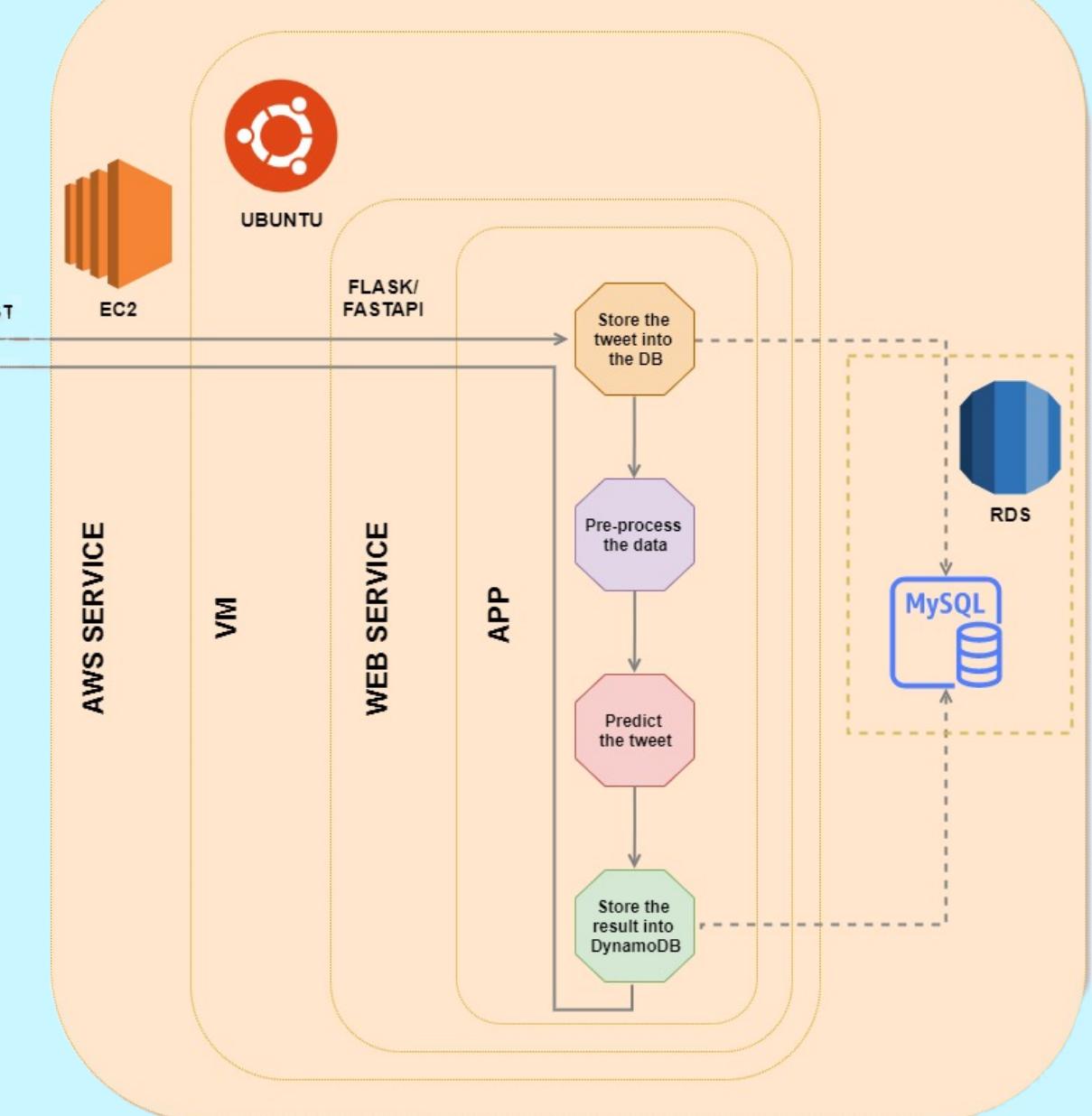
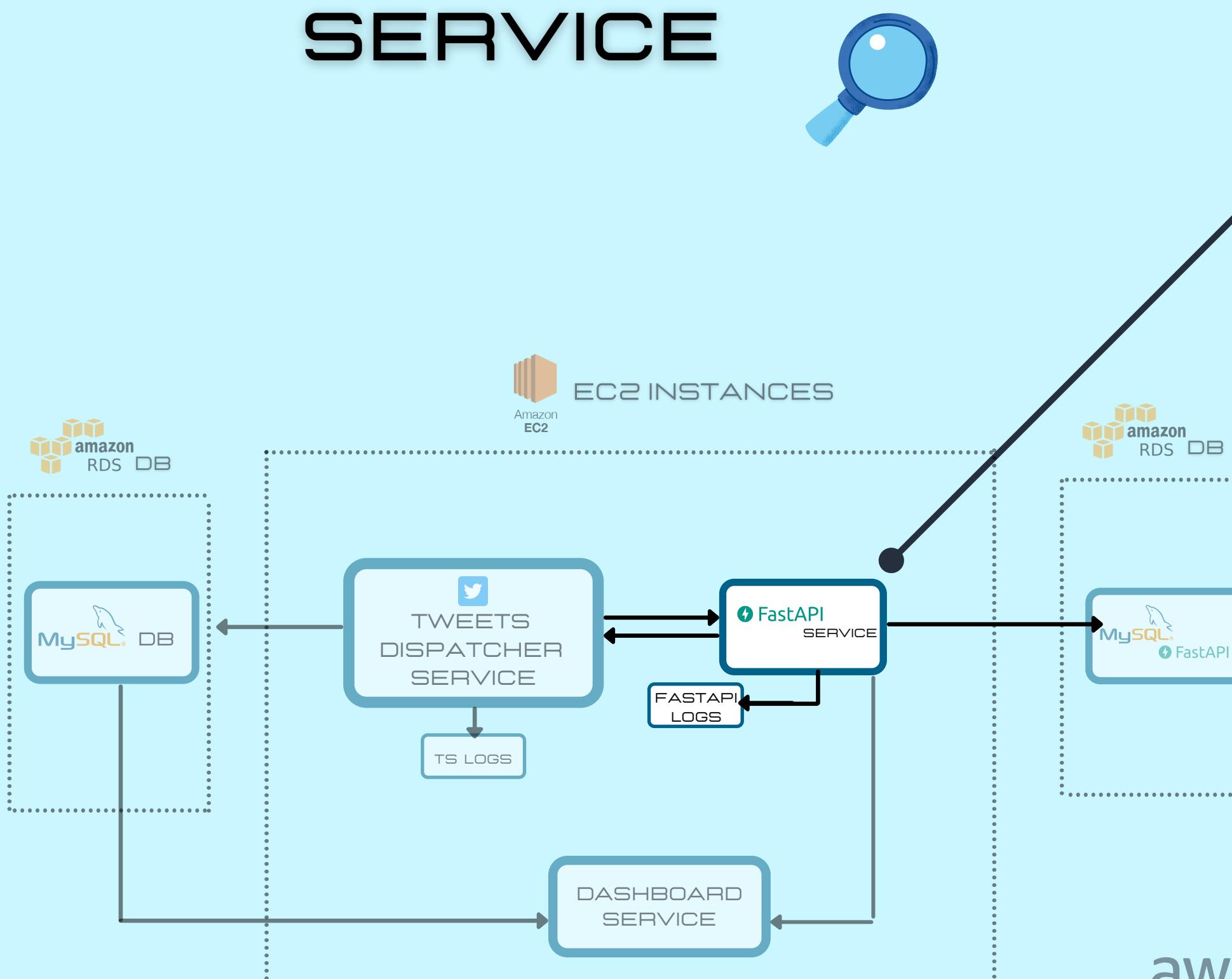
MICRO-SERVICE COMPOSITION

- 1.TWEETS DISPATCHER SERVICE
- 2.TWEETS PREDICTION SERVICE
- 3.MONITORING DASHBOARD
SERVICE



aws

TWEETS PREDICITON SERVICE



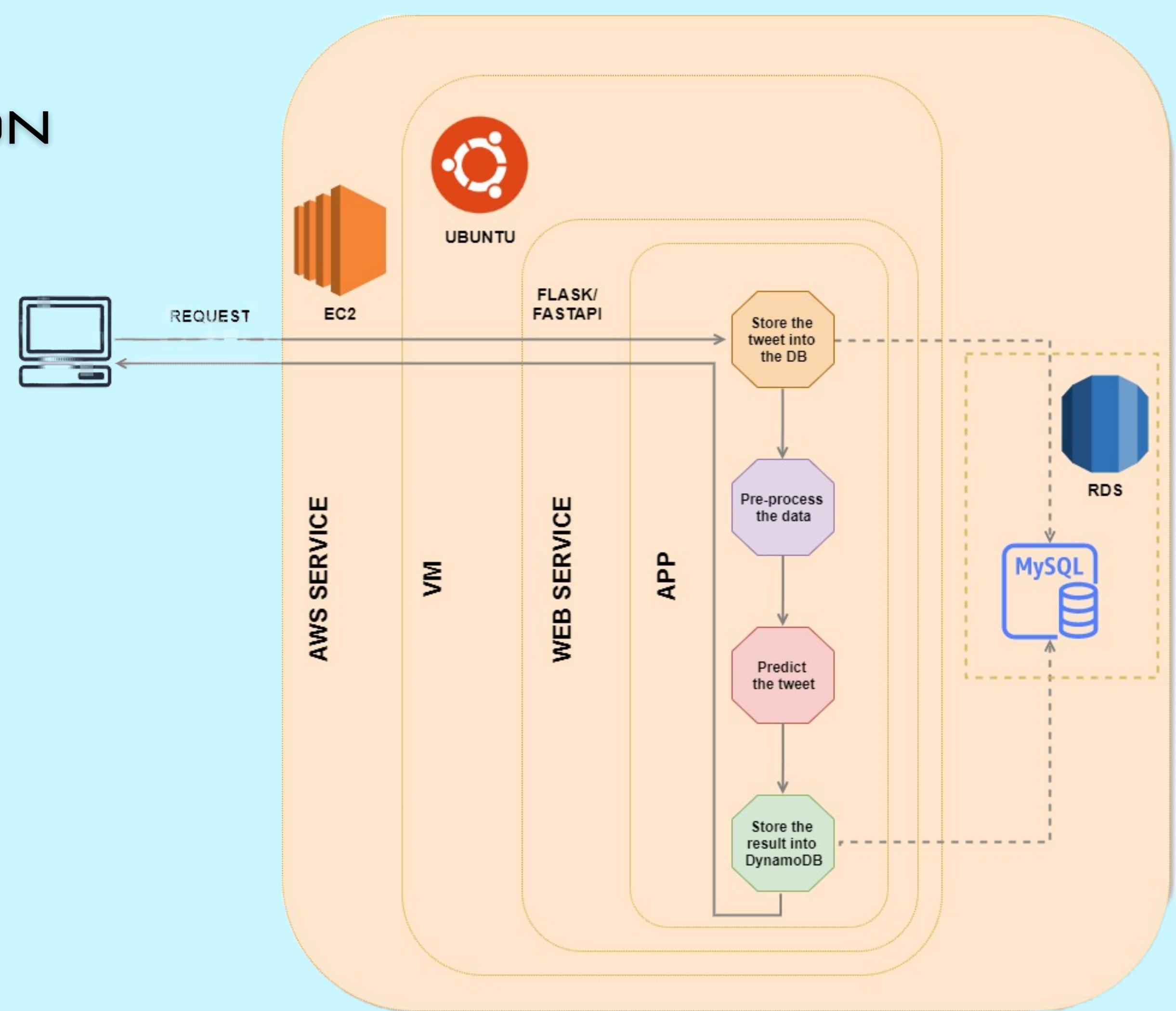
API: TWEETS PREDICTION SERVICE

SENTIMENT
PREDICTION APP

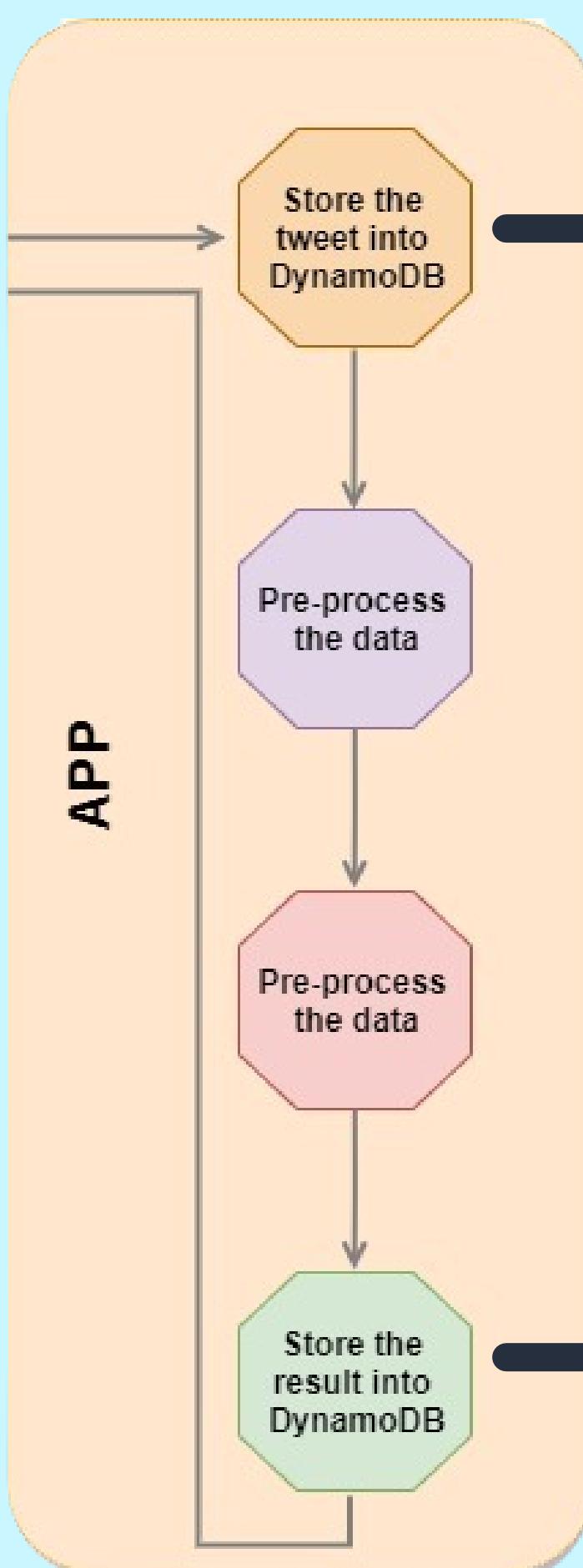
REST API
WEB-SERVICE

UBUNTU VM

AWS EC2 CLOUD
SERVICE



DB SCHEMA



TWEETS



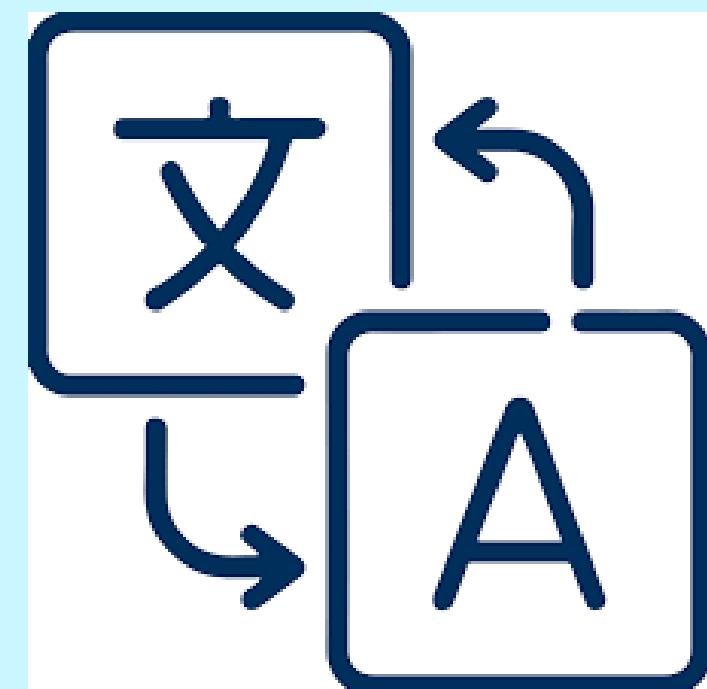
ID_TWEET
TWEET
TWEET_DATE

TWEET_PREDICT

ID_TWEET
ML_VERSION
RESPONSE
POLARITY
SUBJECTIVITY

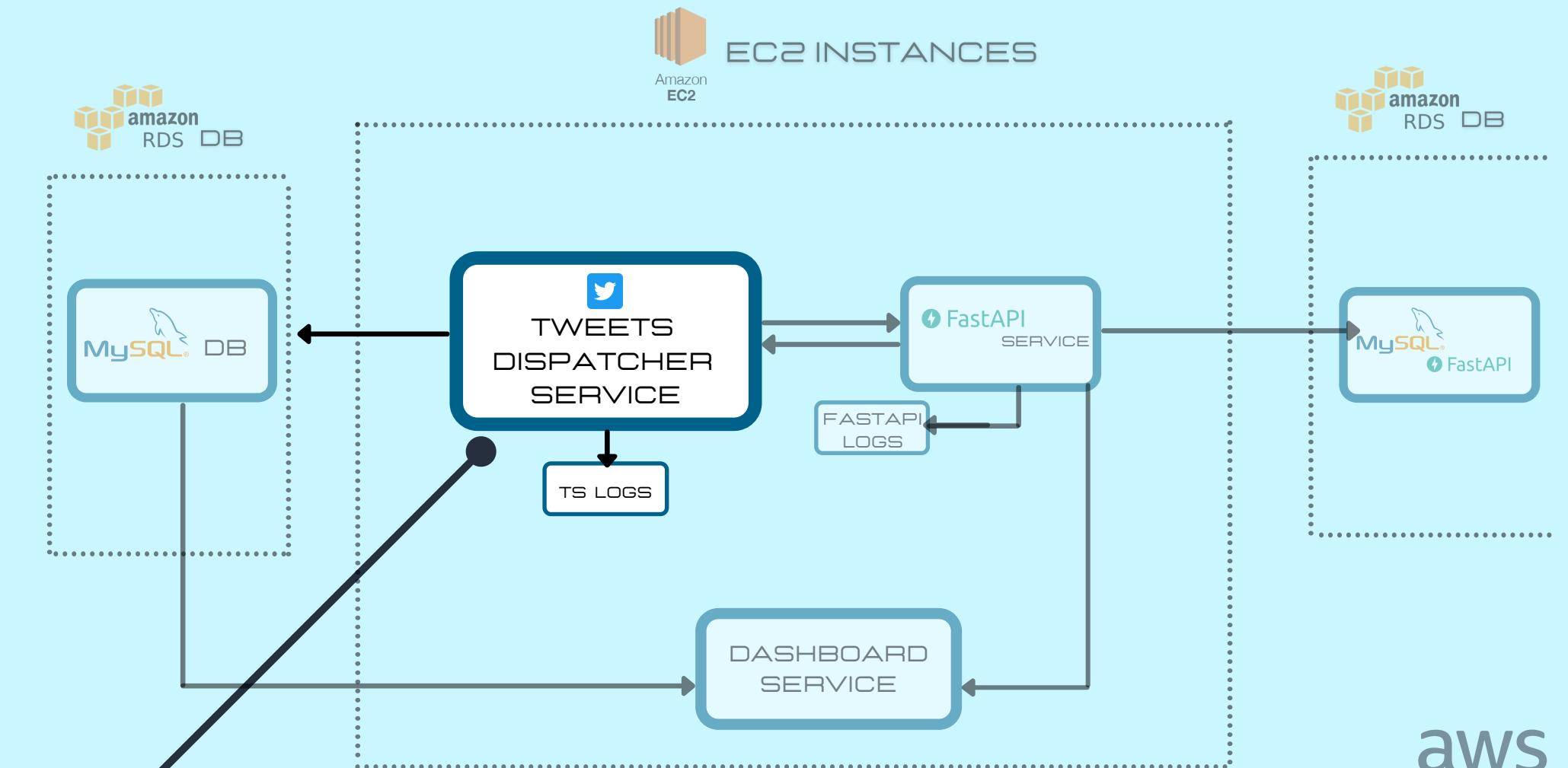
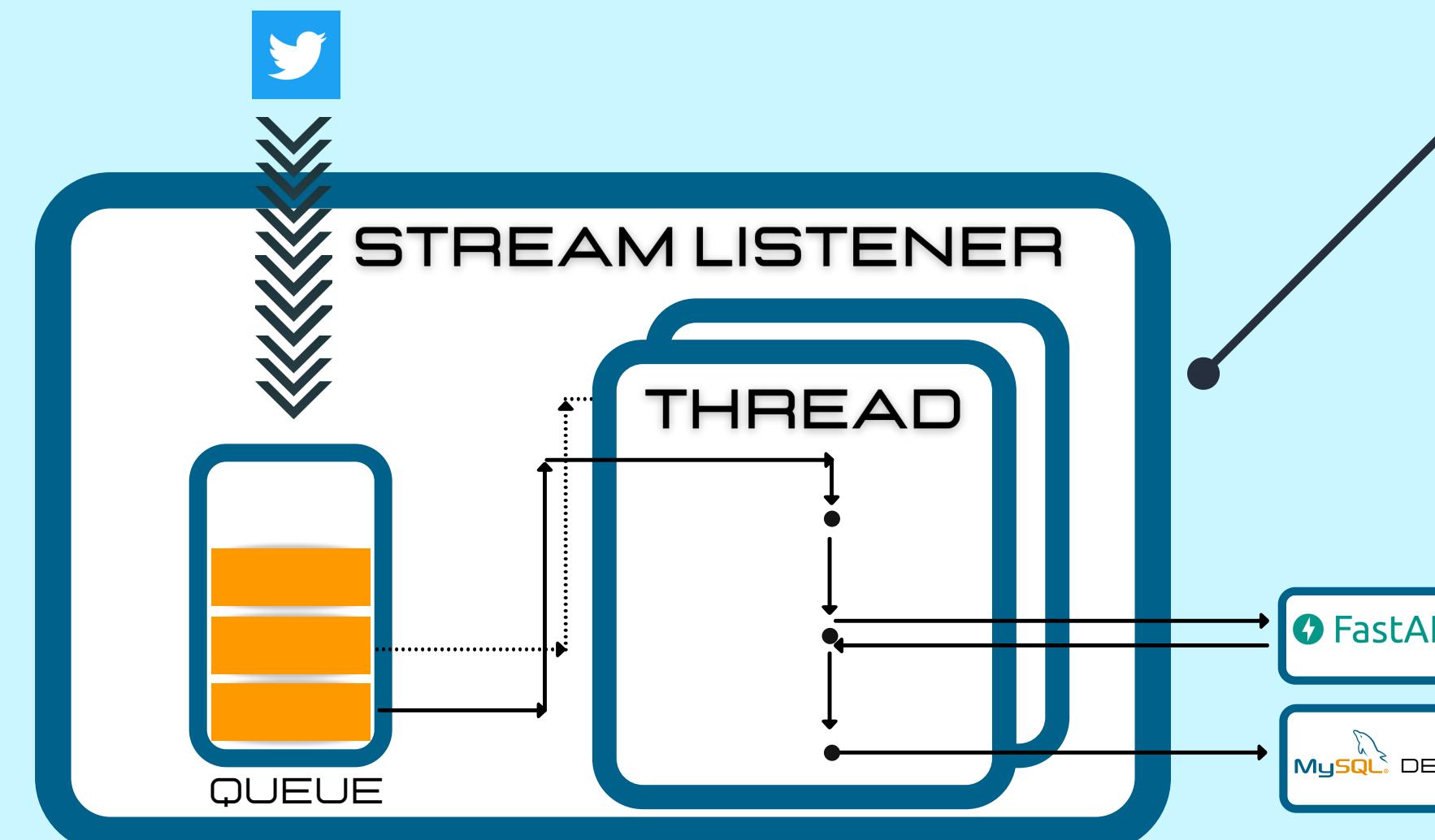
API: OTHER SERVICES

- TEXT CORRECTION
- TEXT TRANSLATION



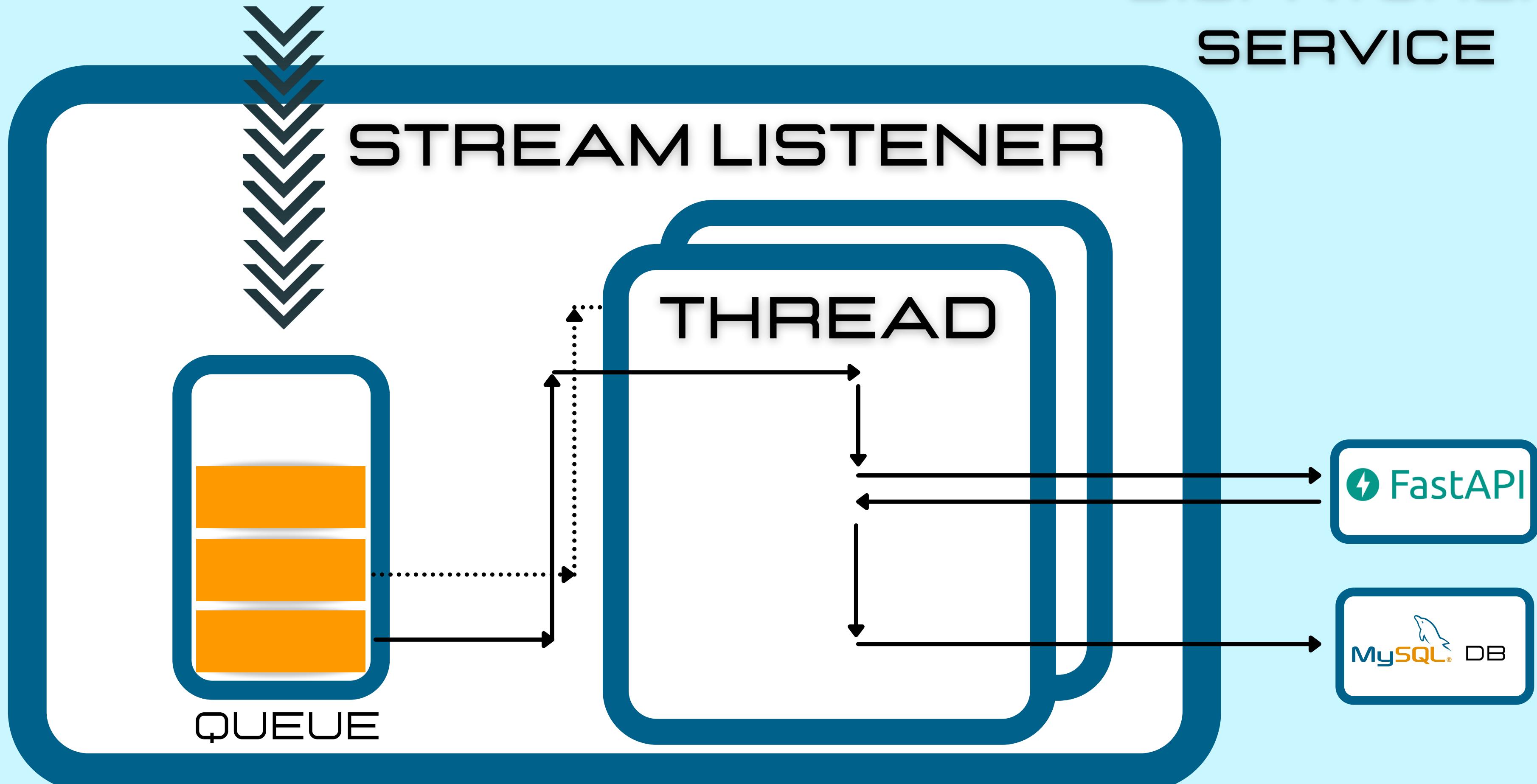


TWEETS DISPATCHER SERVICE





TWEETS
DISPATCHER
SERVICE



DB SCHEMA



SERVICE_MONITORING

TIMESTAMP

STATUS_CODE

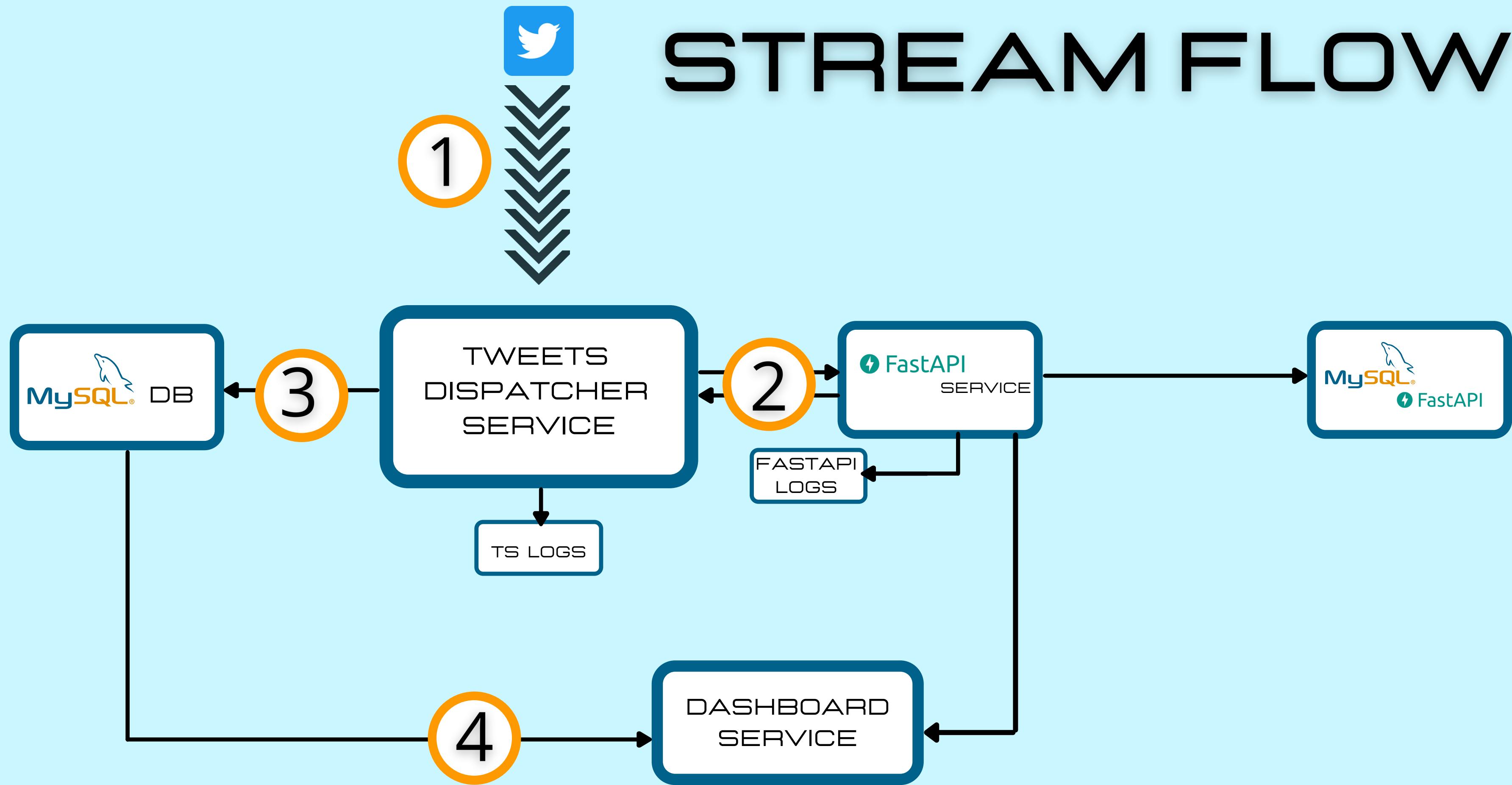
ELAPSED_TIME

RESPONSE

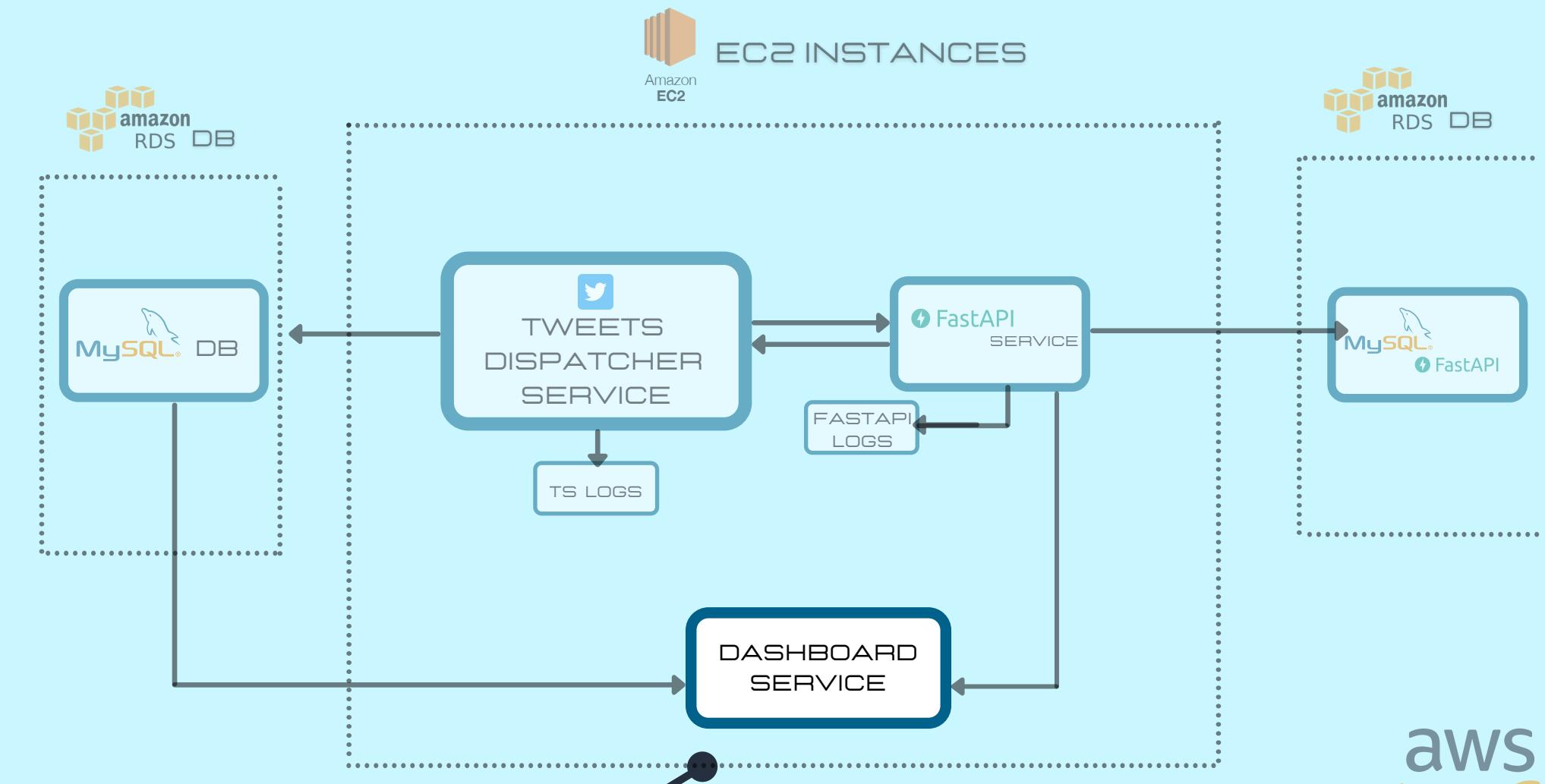
POLARITY

SUBJECTIVITY

STREAM FLOW



MONITORING DASHBOARD SERVICE



aws

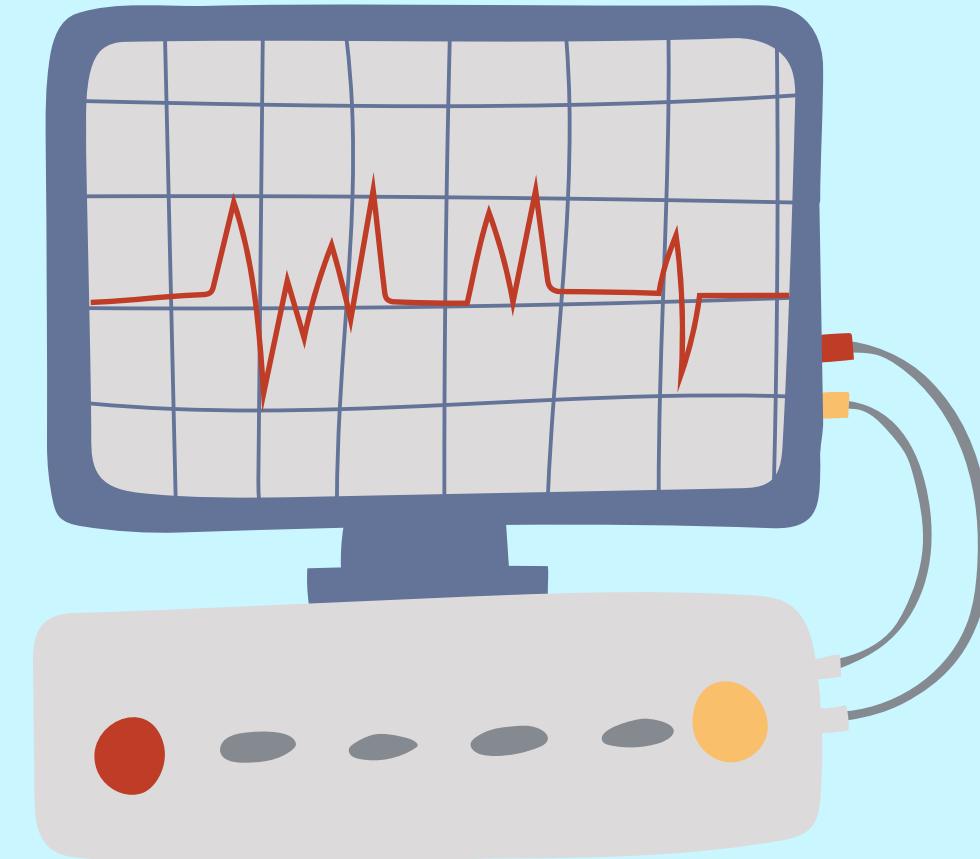
DASHBOARD SERVICE



DASHBOARD

1- REAL TIME MONITORING OF THE STREAM:

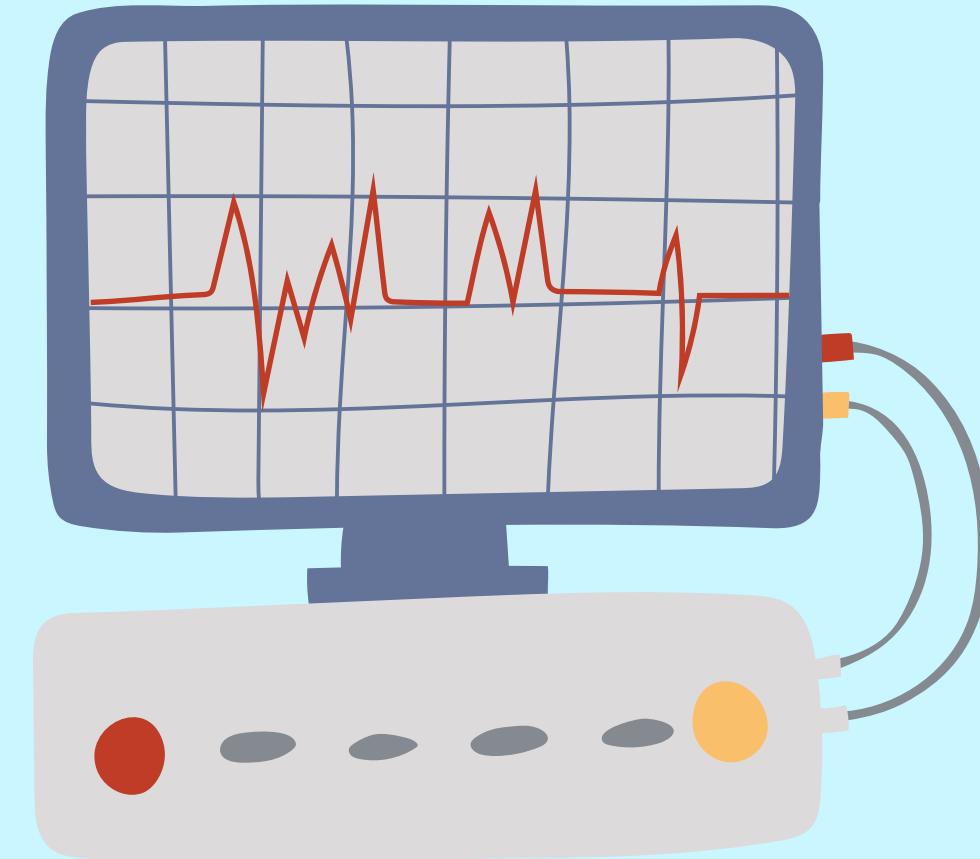
- HTTP STATUS
- ELAPSED TIME
- RESPONSE



DASHBOARD

2 - BATCH ANALYSIS
OF THE TWEETS:

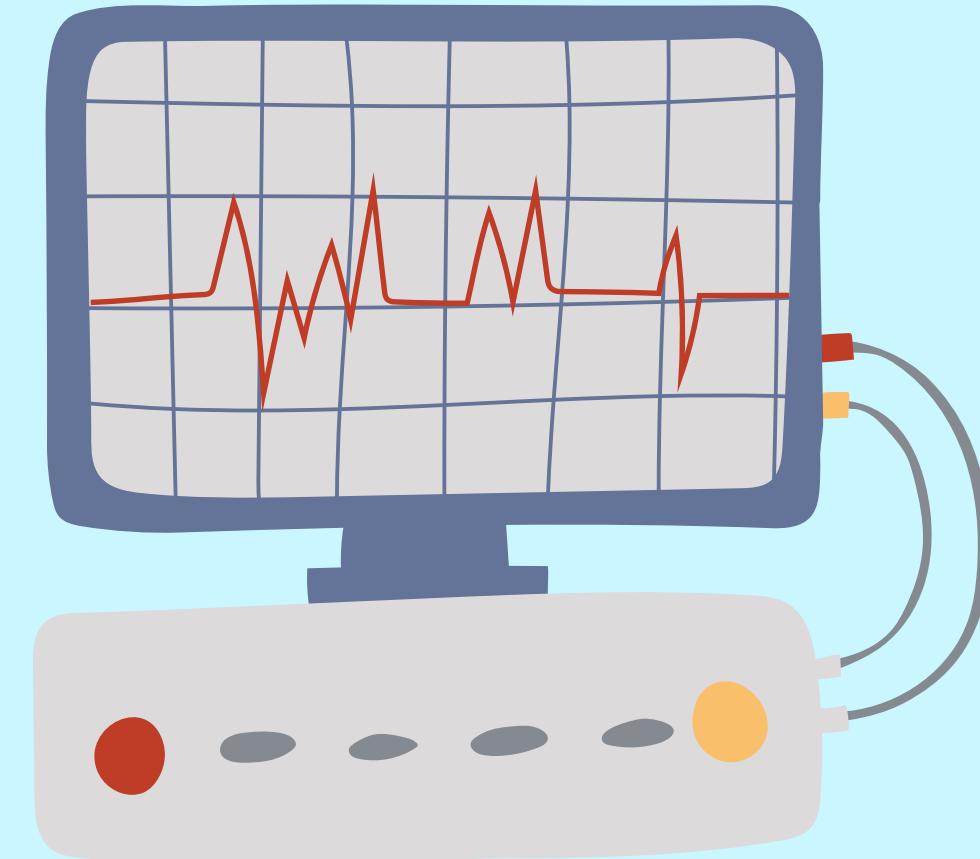
- SENTIMENT
- POLARITY
- SUBJECTIVITY



DASHBOARD

3 - SINGLE TWEET ANALYSIS:

- CORRECTION
- TRANSLATION
- POLARITY
- SUBJECTIVITY



03

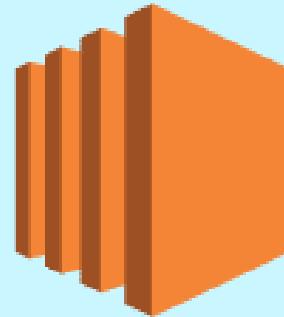
TECHNICAL ASPECT

- CONFIGURATION



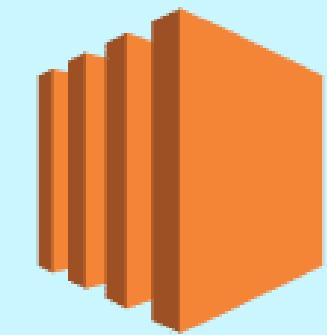
- TOOLS





Amazon
EC2

EC2 CONFIG.

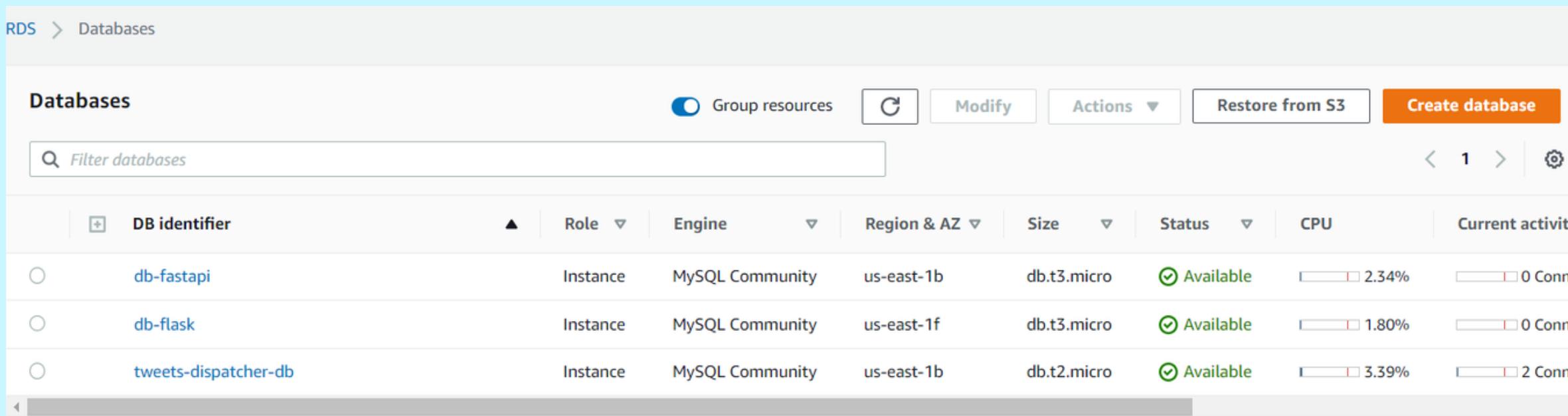


Amazon
EC2

- OS: UBUNTU
- V-CPUS : 1GB
- MEMORY(GIB): 1
- INSTANCE STORAGE(GB): EBS ONLY
- NETWORK PERFORMANCE: LOW TO MODERATE
- IPV6 SUPPORT: YES

Details	Security	Networking	Storage	Status checks	Monitoring	Tags
▼ Instance details Info						
Platform	AMI ID	Monitoring disabled				
Ubuntu (Inferred)	ami-0747bdcabd34c712a					
Platform details	AMI name	Termination protection Disabled				
Linux/UNIX	ubuntu/images/hvm-ssd/ubuntu-bionic-18.04-amd64-server-20210415					
Launch time	AMI location	Lifecycle normal				
Tue Aug 31 2021 10:40:26 GMT+0200 (Ora legale dell'Europa centrale) (5 days)	099720109477/ubuntu/images/hvm-ssd/ubuntu-bionic-18.04-amd64-server-20210415					
Stop-hibernate behavior disabled	AMI Launch index 0	Key pair name new-flask				
State transition reason -	Credit specification standard	Kernel ID -				
State transition message -	Usage operation RunInstances	RAM disk ID -				
Owner 0000000000000000	ClassicLink -	Enclaves Support -				

DB CONFIG.



The screenshot shows the AWS RDS Databases console. At the top, there are buttons for 'Group resources' (unchecked), 'Modify', 'Actions', 'Restore from S3', and a prominent orange 'Create database' button. Below this is a search bar with the placeholder 'Filter databases'. The main area is a table titled 'Databases' with columns: DB identifier, Role, Engine, Region & AZ, Size, Status, CPU, and Current activity. There are three rows of data:

DB identifier	Role	Engine	Region & AZ	Size	Status	CPU	Current activity
db-fastapi	Instance	MySQL Community	us-east-1b	db.t3.micro	Available	2.34%	0 Connections
db-flask	Instance	MySQL Community	us-east-1f	db.t3.micro	Available	1.80%	0 Connections
tweets-dispatcher-db	Instance	MySQL Community	us-east-1b	db.t2.micro	Available	3.39%	2 Connections

- **DB.T2.MICRO**
- **20 GB OF STORAGE**
- **AUTOMATIC BACKUP WITH A RETENTION PERIOD OF ONE DAY**



PuTTY

SSH AND TELNET CLIENT

PUTTY



PuTTY

```
ubuntu@ip-172-31-1-1: ~
* Support: https://ubuntu.com/advantage

System information as of Tue Aug 31 15:58:39 UTC 2021

System load: 0.0          Processes: 95
Usage of /: 28.3% of 7.69GB  Users logged in: 0
Memory usage: 20%          IP address for eth0: 172.31.1.1
Swap usage: 0%

* Super-optimized for small spaces - read how we shrank the memory
footprint of MicroK8s to make it the smallest full K8s around.

https://ubuntu.com/blog/microk8s-memory-optimisation

30 packages can be updated.
1 of these updates is a security update.
To see these additional updates run: apt list --upgradable

New release '20.04.3 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Tue Aug 31 09:10:06 2021 from 93.47.1.1
ubuntu@ip-172-31-1-1: ~$
```



WinSCP

WINSSCP



tweets_dispatcher - tweets-dispatcher - WinSCP

Local Mark Files Commands Session Options Remote Help

Synchronize Queue Transfer Settings Default

tweets-dispatcher X New Session

Desktop Upload Edit Properties New +

C:\Users\Public\Desktop\Desktop\Downloads\tweets_dispatcher\

Name	Size	Type	Changed
..		Parent directory	9/4/2021 5:45:15 PM
main_app.py	4 KB	PY File	9/4/2021 5:45:15 PM
params.py	1 KB	PY File	9/3/2021 6:37:28 PM
requirements.txt	1 KB	Text Document	9/3/2021 6:35:11 PM
tweets-dispatcher.pem	2 KB	PEM File	9/3/2021 12:16:22 AM
tweets-dispatcher.ppk	2 KB	PutTY Private Key ...	9/3/2021 12:17:45 AM

ubuntu

Download Edit Properties New +

/home/ubuntu/

Name	Size	Changed	Rights	Owner
..		9/3/2021 12:27:56 AM	rwxr-xr-x	root
__pycache__		9/3/2021 6:38:36 PM	rwxrwxr-x	ubuntu
app_logs.log	172,142 KB	9/4/2021 6:44:41 PM	rw-rw-r--	ubuntu
main_app.py	4 KB	9/4/2021 5:45:15 PM	rw-rw-r--	ubuntu
params.py	1 KB	9/3/2021 6:37:28 PM	rw-rw-r--	ubuntu
requirements.txt	1 KB	9/3/2021 6:35:11 PM	rw-rw-r--	ubuntu

0 B of 7.61 KB in 0 of 5 0 B of 168 MB in 0 of 5 10 hidden

10:00:16

OPEN
SOURCE
SFTP, FTP
AND SCP
VISUAL
CLIENT
FOR
WINDOWS.