



# Presentación de análisis y resultados de la solución

Ing. Christian Luis Lopez Bravo



# Objetivo

Trabajar con base de datos relacional para realizar el análisis de tweets enviados en archivo plano en un formato json



## Extraccion de tweet

se implemento un script python para la validacion del archivo con tweets en formato json utilizando librerias json y jsonschema para poder validar la estructura se creo un esquema y tambien se valido la existencia de campos vacios utilizando la funcion `check_empty_fields`



# Carga base de datos

se crearon 2 tablas para el cargado de los tweets

```
CREATE TABLE nao.tweet (
```

```
    `id` int(11) NOT NULL AUTO_INCREMENT, id_tweet varchar(30) NOT NULL, texto varchar(1000) NULL, usuario  
    varchar(100) NULL, fecha DATETIME NULL, retweets NUMERIC NULL, favoritos NUMERIC NULL, CONSTRAINT tweet_pk  
    PRIMARY KEY (id)) charset utf8mb4;
```

```
CREATE TABLE `hashtag` (`id` int(11) NOT NULL AUTO_INCREMENT, `hashtags` varchar(500) DEFAULT NULL, `id_tweets`  
int(11) DEFAULT NULL, PRIMARY KEY (`id`), KEY `hashtag_tweet_FK` (`id_tweets`), CONSTRAINT `hashtag_tweet_FK`  
FOREIGN KEY (`id_tweets`) REFERENCES `tweet` (`id`)
```

```
) charset utf8mb4;
```



## análisis de resultados

se desarrollaron script para identificar los siguiente

¿Cuáles son las palabras más empleadas?

¿Qué sentimiento es más intenso?

¿Qué tan objetivos son los Tweets sobre el tema?

¿Qué tan positivos se espera que sean los comentarios para el día de mañana?



## Conclusiones

- se tuvo que crear 2 tablas relacionadas para cargar la informacion y poder realizar el analisis de la informacion