

# **Desarrollo de una aplicación para el análisis social en Twitter mediante tecnologías Big Data.**

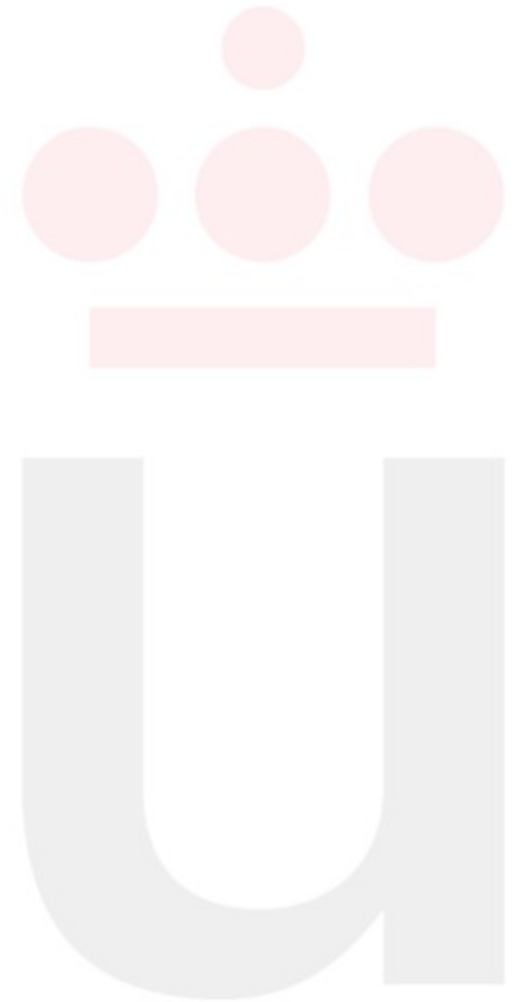
## **Caso de estudio: @urjc**

Autor: Nazareth Guerrero Yébenes

Tutor: Felipe Alonso Atienza

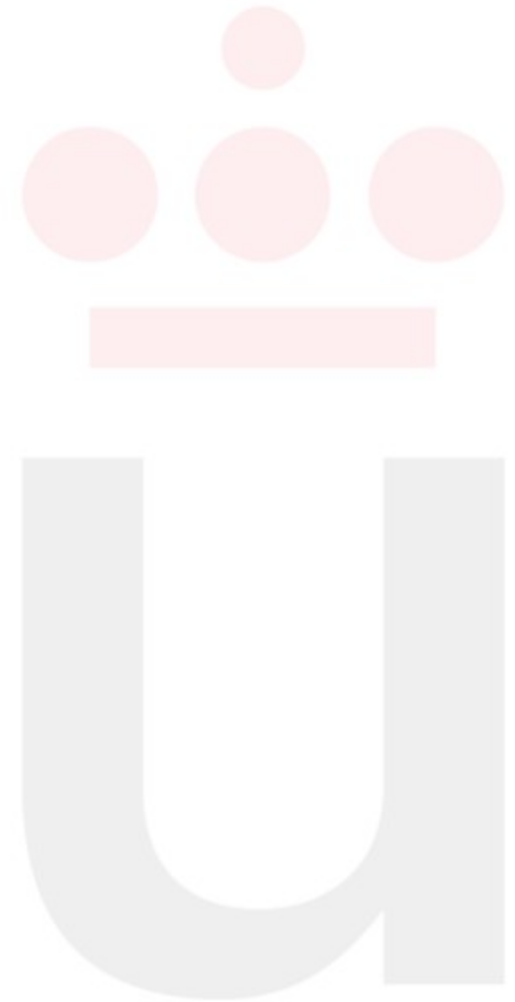
# Índice

- Introducción
- Twitter
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Índice

- **Introducción**
- Twitter
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Introducción

Twitter

Big  
Data

Bases  
de datos no  
relacionales

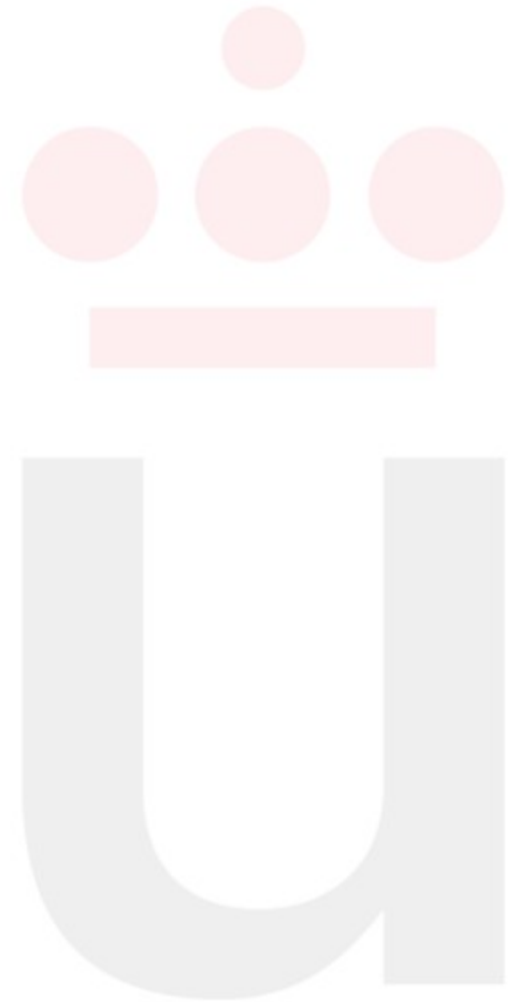
**Desarrollo de una aplicación Web que realice un análisis sobre cómo se habla de la URJC en Twitter, mostrando estadísticas en una página Web en tiempo real**

Tratamiento  
de la  
información

Visualización  
Web

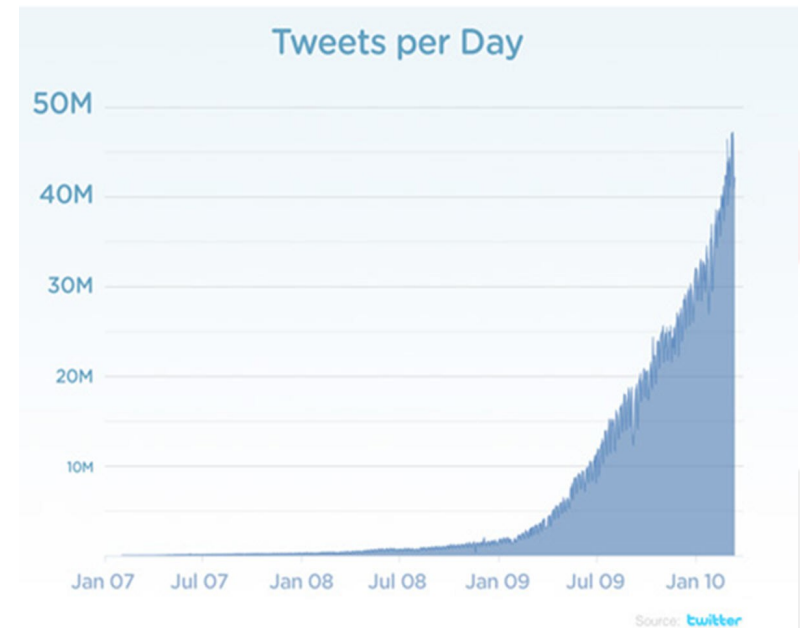
# Índice

- Introducción
- **Twitter**
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Twitter

- Jack Dorsey: mensajes cortos dentro de la compañía Odeo
- Noah Glass y Evan Williams: Primera versión de twitter: twttr
- 2009: Primera versión en español



2013: 200 millones de usuarios activos y 400 millones de tweets al día

# Twitter



**Nebrija Universidad** @Nebrija · 28m

¿Buscas algo para hacer en grupo? ¿Has estado alguna vez atrapado en el Museo Conde Duque? [madridcoolblog.com/2014/03/exit-s...](http://madridcoolblog.com/2014/03/exit-s...)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**EFEemprende** @EFEemprende · 28m

Madrid acoge I Sesión de Networking Profesional para pymes y [#emprendedores](https://www.facebook.com/emprendedores)

[ow.ly/v0gqC](http://ow.ly/v0gqC) @ComunidadMadrid

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



Retweeted by Universidad Sevilla



**Sostenibilidad\_US** @USVerde · 1h

¡3 días para la [#horadelplaneta](https://www.facebook.com/horadelplaneta) de @WWFespana! El sábado a las 20:30 apgamos el rectorado de @unisevilla. [youtube.com/watch?v=nBnKS6...](http://youtube.com/watch?v=nBnKS6...)

[View media](#)

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**EFE EMPRESAS** @EFEempresas · 35m

El crédito al consumo frenó su caída y creció un 3,25% en 2013 [@asnef](http://asnef.com) [ow.ly/v0fjw](http://ow.ly/v0fjw)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More



**Universidad Sevilla** @unisevilla · 37m

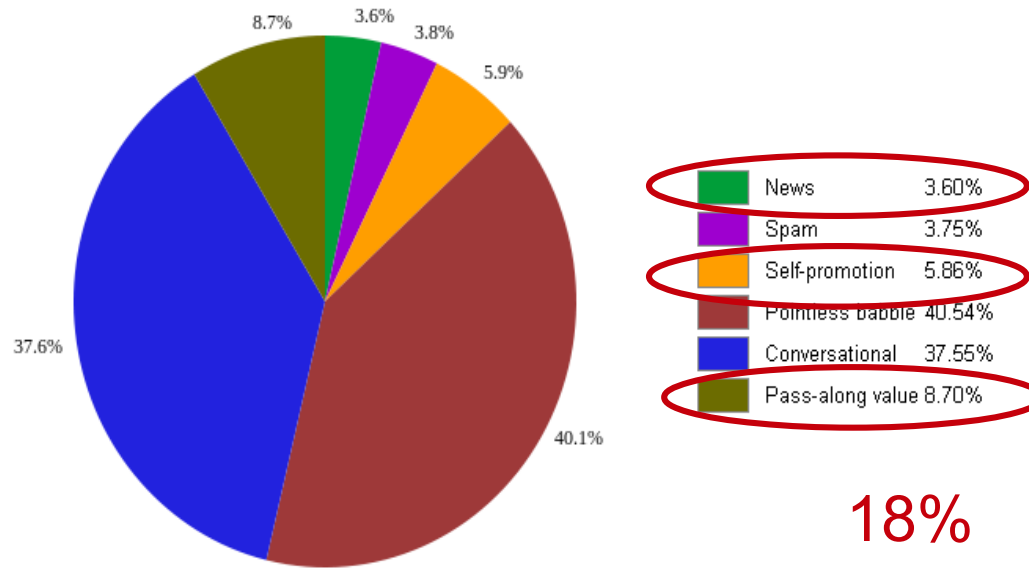
La [#ciencia](https://www.facebook.com/ciencia), muy presente en la próxima edición del [#SalónUS](https://www.facebook.com/SalonUS) y Ferisport [bit.ly/1mvbvg0](http://bit.ly/1mvbvg0) ¡Ya queda poco! Vía [@InvestigaUS](http://InvestigaUS)

Expand

↩ Reply ↻ Retweet ★ Favorite ⋮ More

- Mensajes de 140 caracteres llamados *tweets*
- Amigos y seguidores
- TimeLine
- Hashtag #
- Retweet

# Twitter

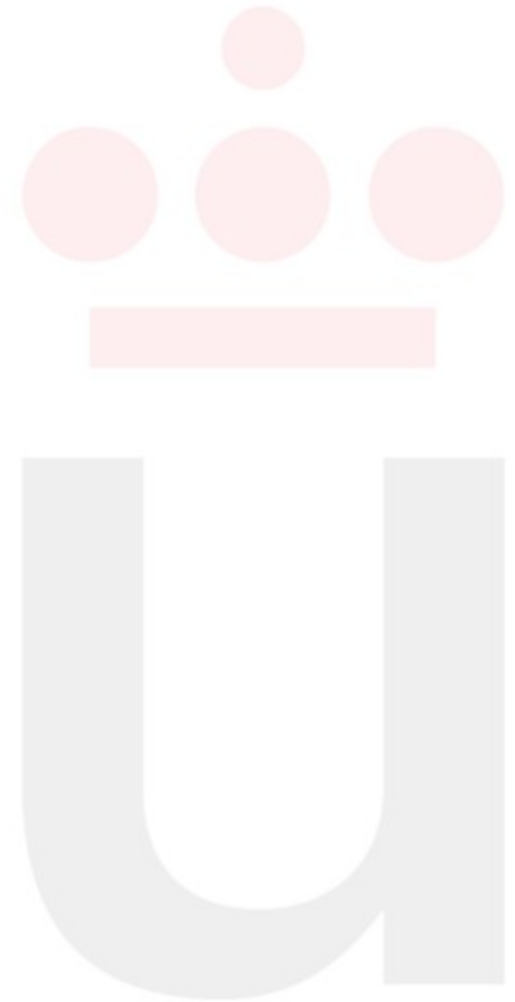


18%



# Índice

- Introducción
- Twitter
- **Big Data**
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Big Data



2001

Doug Laney

Modelo de las 3 Vs

2012

Gartner

*“Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”*

# Big Data

**amazon**



**facebook**

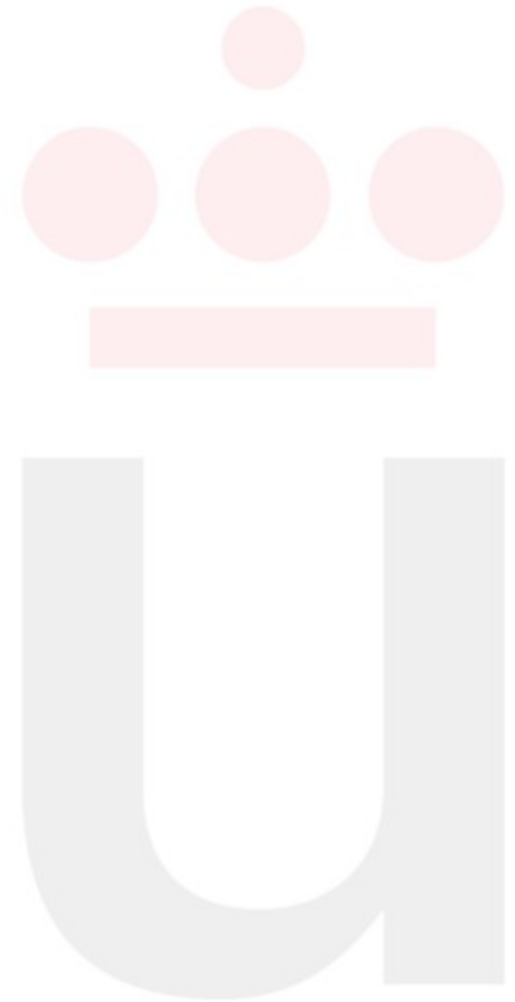


**FICO**

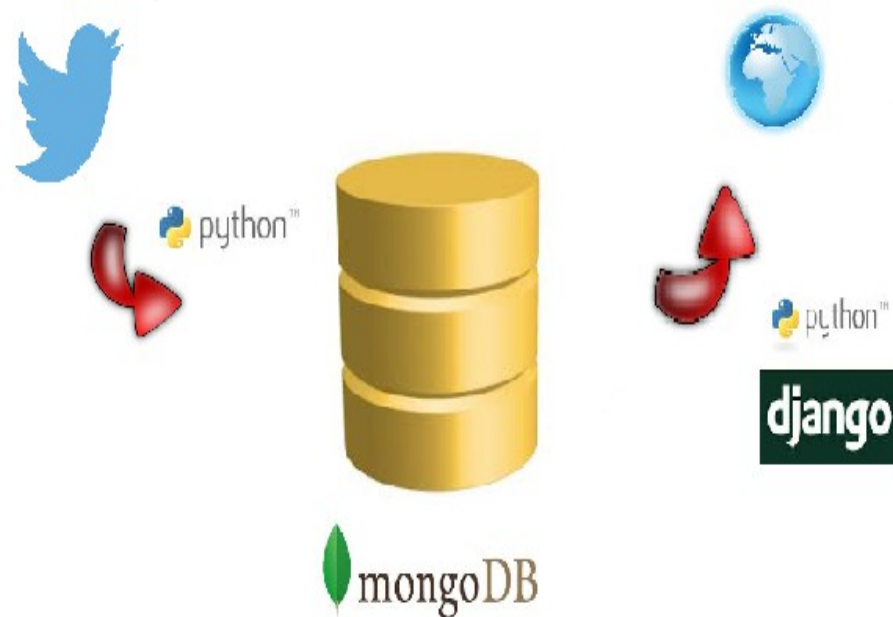
**ebay**

# Índice

- Introducción
- Twitter
- Big Data
- **Desarrollo de la aplicación**
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Desarrollo de la aplicación



Recolección de datos

Twitter - API - JSON

Almacenamiento

Python - MongoDB

Procesamiento

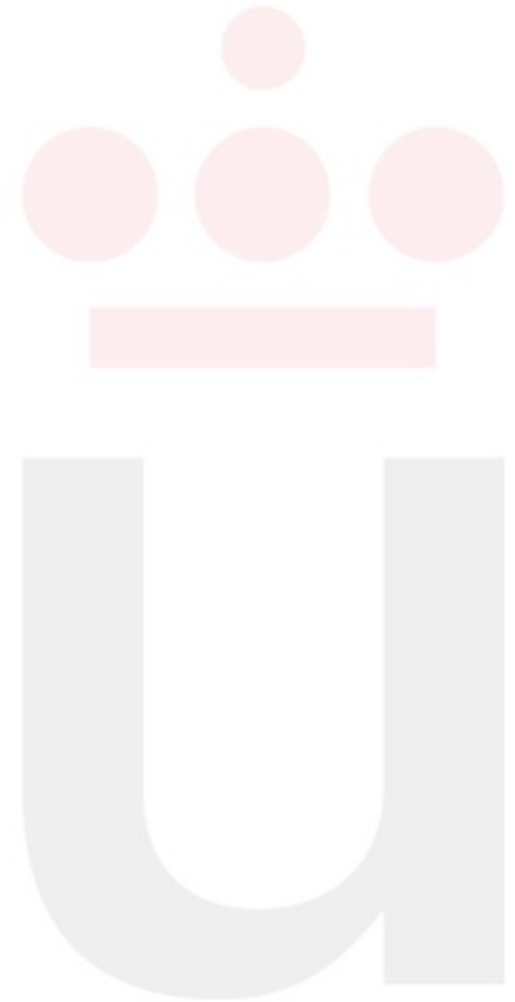
MapReduce

Visualización

Django - HTML5

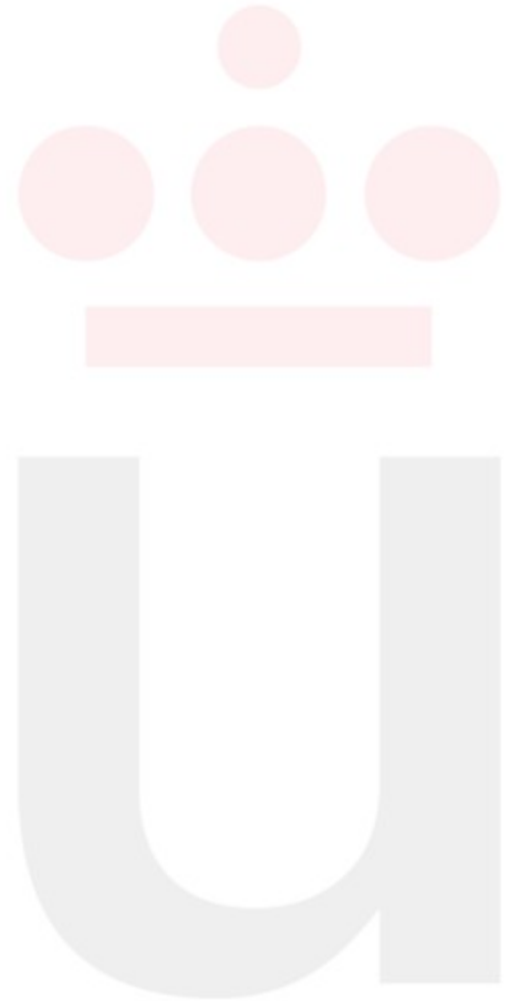
# Índice

- Introducción
- Twitter
- Big Data
- **Desarrollo de la aplicación**
  - **Recolección de datos**
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras

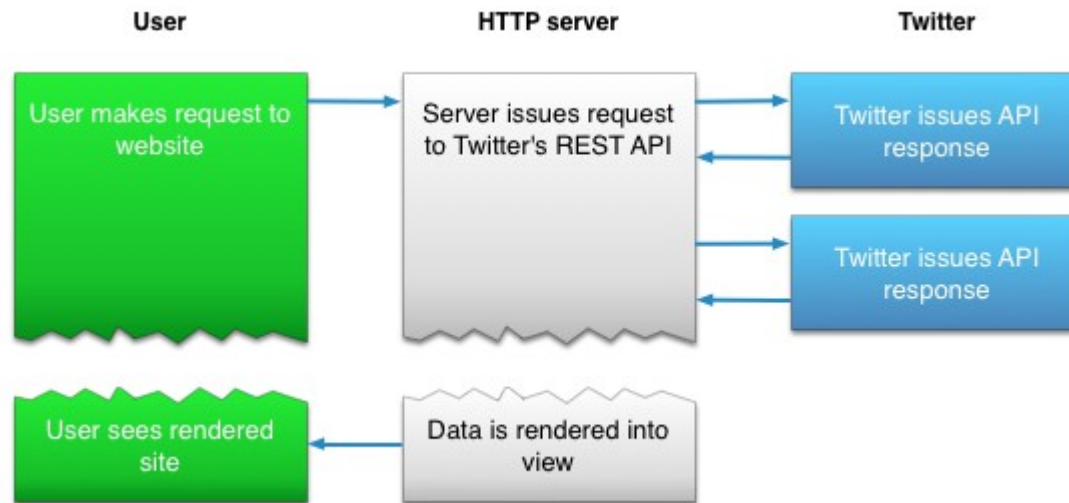


# Autenticación

oauth

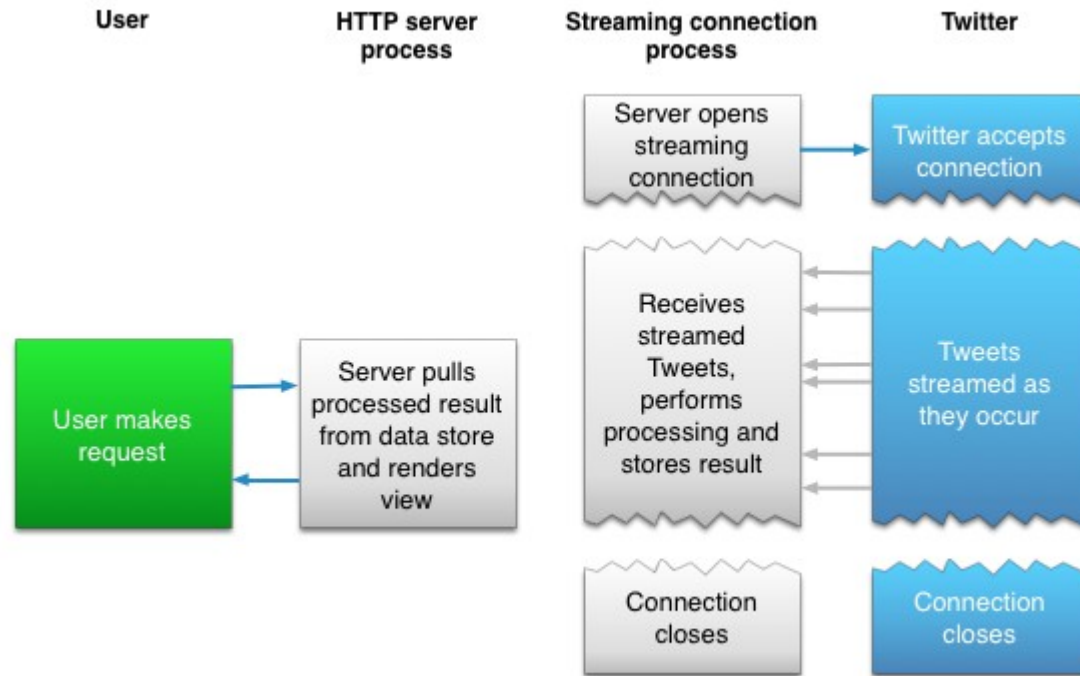


# REST API





# Streaming API

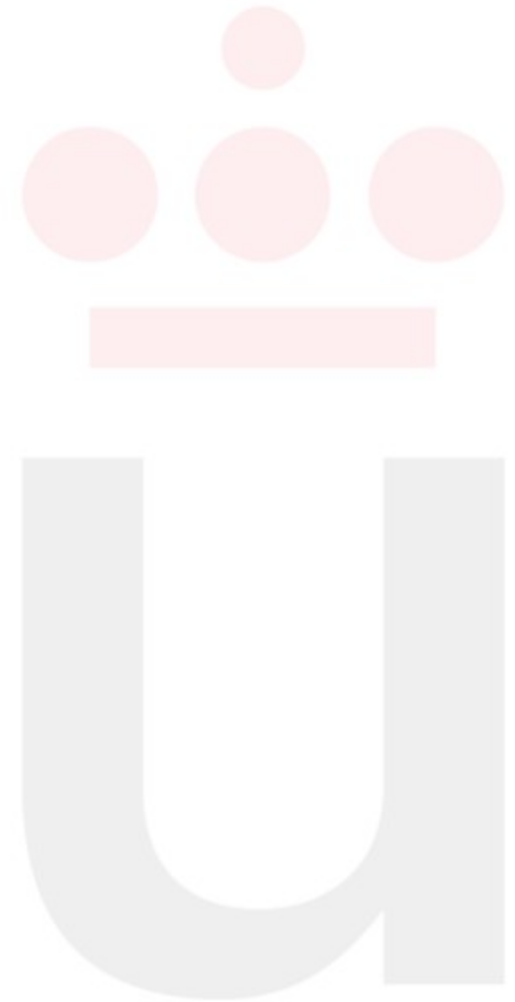


# JSON

```
"id": "123456789",
"texto": "Esto es un ejemplo de un tweet",
"Autor": "Nazareth",
"fecha": "27 de Diciembre de 2013",
"Retweet": [
  { "Numero": "2" },
  { "Usuarios": ["Amigo1", "Amigo2"] },
  { "Fechas": ["28 de Diciembre de 2013", "1 de Enero
    de 2014"] }
]
}
```

# Índice

- Introducción
- Twitter
- Big Data
- **Desarrollo de la aplicación**
  - Recolección de datos
  - **Almacenamiento**
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



# Almacenamiento

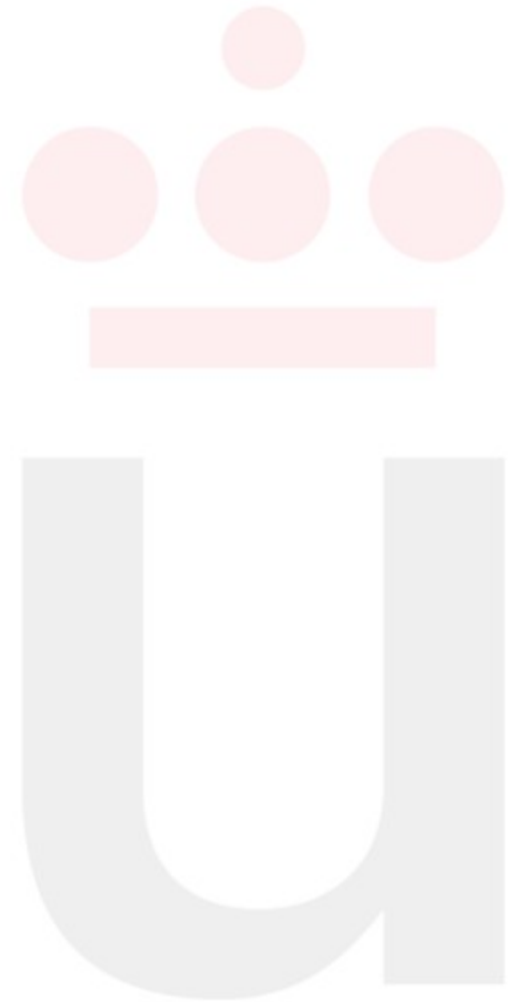
Modelo relacional	Modelo no relacional
Estructura de tablas fijas	Estructura flexible
Único nodo	Arquitectura distribuida (escalabilidad horizontal)
Pérdida de información total ante un error	Fácil recuperación frente a errores



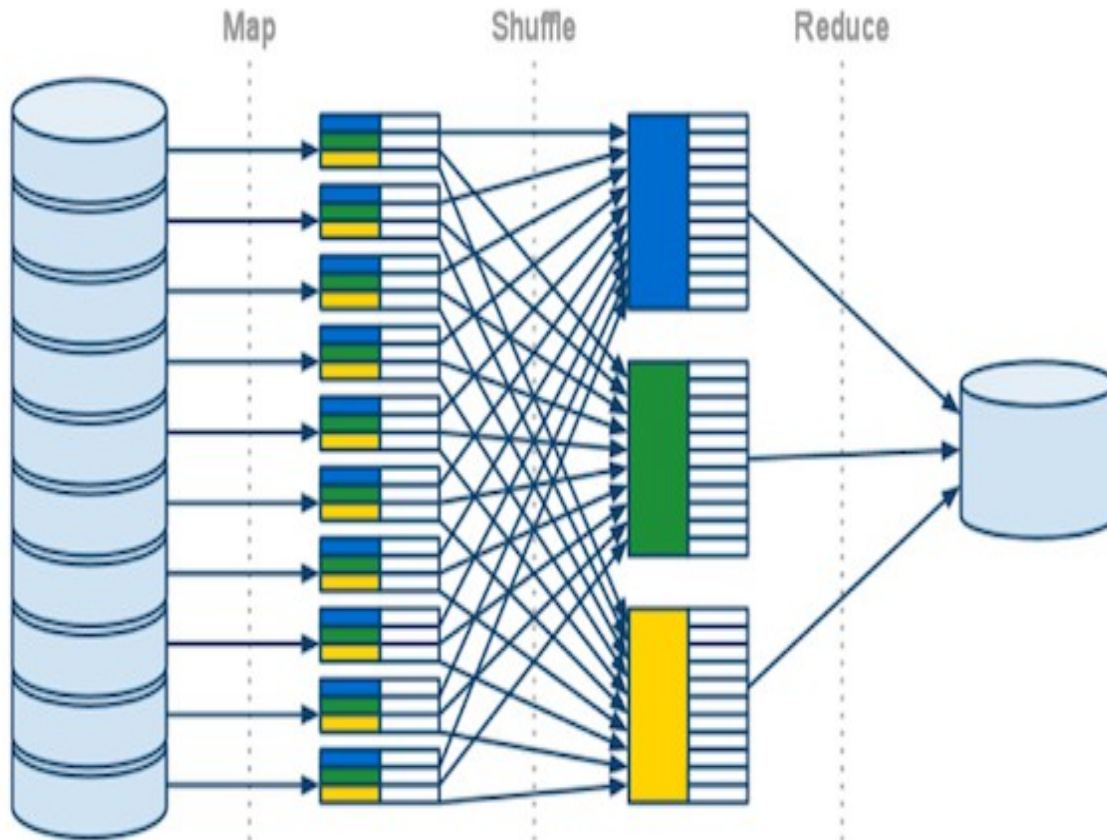
- Modelo no relacional
- Basado en documentos JSON
- PyMongo

# Índice

- Introducción
- Twitter
- Big Data
- **Desarrollo de la aplicación**
  - Recolección de datos
  - Almacenamiento
  - **Procesamiento**
  - Visualización
- Resultados
- Conclusiones
- Líneas futuras



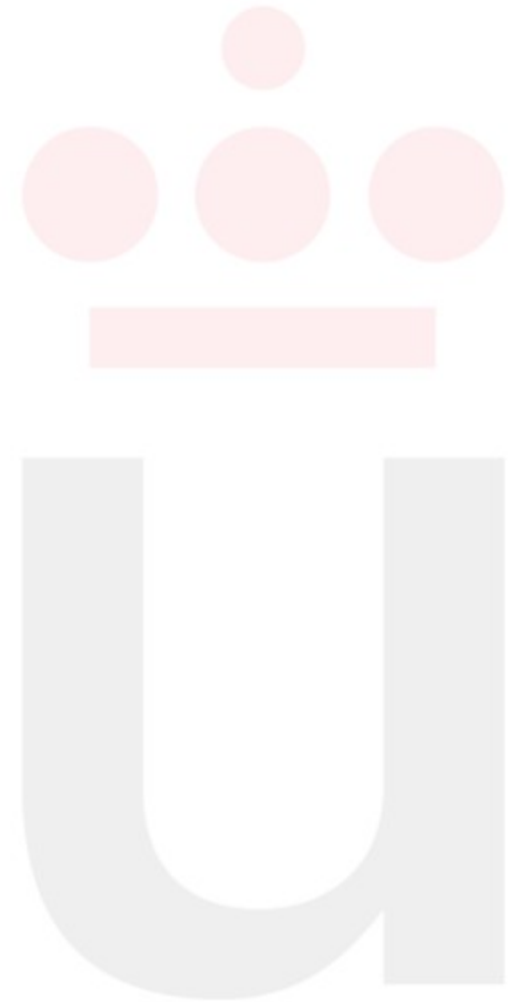
# Procesamiento - mapReduce



- Algoritmo para Big Data
- Computación paralela
- Reducción de tiempo de computación
- Métodos incorporados en PyMongo

# Índice

- Introducción
- Twitter
- Big Data
- **Desarrollo de la aplicación**
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - **Visualización**
- Resultados
- Conclusiones
- Líneas futuras



# Django

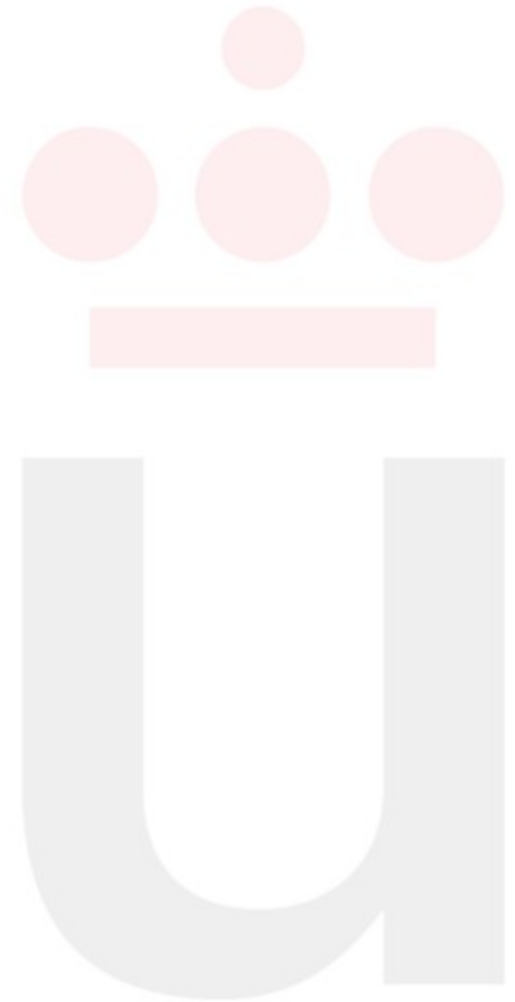


- Python
- Numerosas librerías
- Gestión de URLs
- Sistema de plantillas
- HTML5
- Bootstrap



# Índice

- Introducción
- Twitter
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- **Resultados**
- Conclusiones
- Líneas futuras



# Resultados

Twitter

Home

About us

Contact

Últimos tweets

Retweets más populares

Los usuarios que más comentan

Los usuarios más nombrados

Hashtags más populares

Últimos días

Últimos meses

[AntoanetaRad](#)

Algo más salvaje que rebajas de Armani por 5€/und? Pues sí señores, la primera fila de clase de estadística I en la #urjc !

Tue Nov 19 17:13:22 +0000 2013

[verocanero](#)

RT @Varuky13: Mucho CIU, mucha UAM, URJC, UC3M... Pero seguro que ninguno disfruta lo que tenemos por Aranjuez... <http://t.co/XFheR9IMy2>

Tue Nov 19 17:10:43 +0000 2013

[adrisanjuan\\_](#)

Volvemos a Agosto gracias a la URJC. Gracias! ;)

Tue Nov 19 17:07:46 +0000 2013

[JoseCarlosAtero](#)

Apostando por el futuro! #university #math #urjc #study #books #classmate #classroom #afternoon... <http://t.co/zsuMwObu16>

Tue Nov 19 17:07:36 +0000 2013

[futsalUAHmasc](#)

RT @WillyForeverYou: @futsalUAHmasc @urjc duelo en la cumbre, entre el segundo y tercero, ambos equipos invictos y las dos mejores defensas...

Tue Nov 19 17:04:38 +0000 2013

[WillyForeverYou](#)

@futsalUAHmasc @urjc duelo en la cumbre, entre el segundo y tercero, ambos equipos invictos y las dos mejores defensas del campeonato.

Tue Nov 19 17:00:18 +0000 2013

[jorgedems](#)

Vamonos que nos vamos (@ Departamental URJC) <http://t.co/0gDkdkEJVL>

Tue Nov 19 17:00:13 +0000 2013

[jorgedems](#)

I'm at Campus de Vicálvaro (URJC) - @urjc (Madrid) w/ 2 others <http://t.co/Lmg8KJVczo>

Tue Nov 19 16:59:50 +0000 2013

# Resultados

Twitter

Home

About us

Contact

Últimos tweets

Retweets más populares

Los usuarios que más comentan

Los usuarios más nombrados

Hashtags más populares

Últimos días

Últimos meses

## Número de tweets en los últimos días

OCT 31

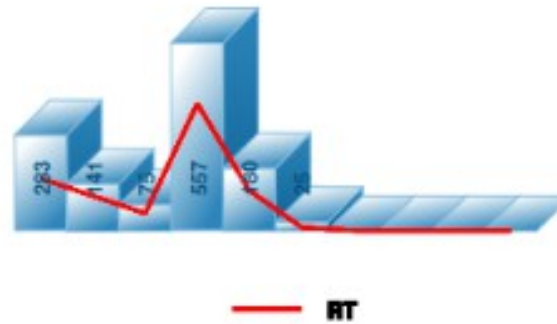
OCT 30

OCT 29

OCT 28

OCT 27

OCT 25

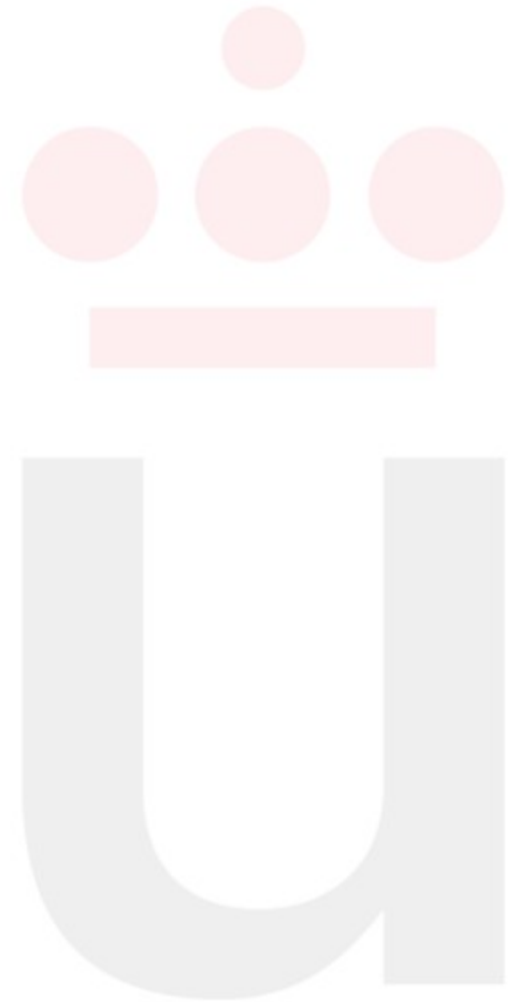


# Resultados

	500 tweets	1000 tweets	entera (2000 tweets aprox)
Complejidad 1	48.55	75.1	154.75
Complejidad 2	37.75	79.35	161.9
Complejidad 3	63.95	126.05	235.15
Media total	50.08	93.5	217.2

# Índice

- Introducción
- Twitter
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- **Conclusiones**
- Líneas futuras



# Conclusiones

Proyecto muy completo

Big Data: Tecnología en desarrollo, interesante forma de tratar los datos.

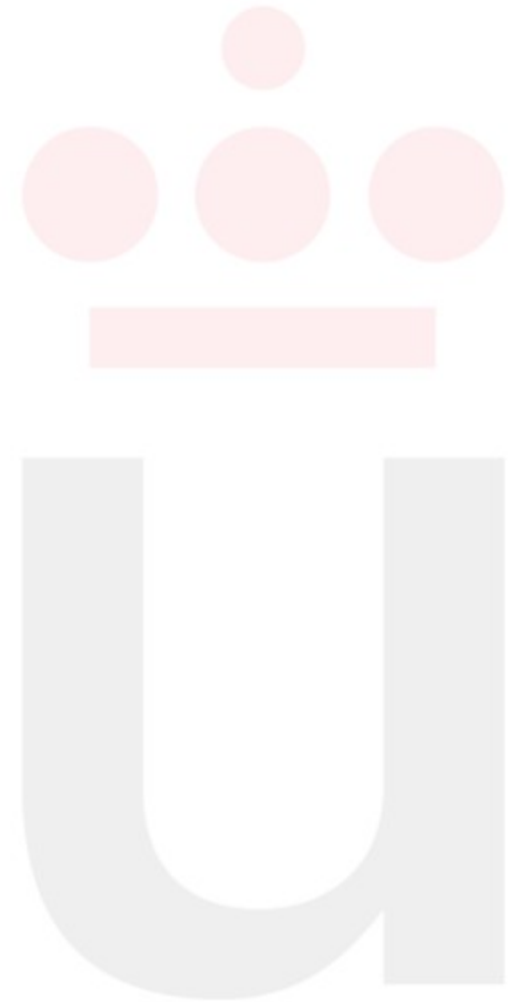
Tecnologías más complejas al tener mayor conjunto de entradas

Falta de documentación

Redes sociales: Permiten multitud de estudios sociológicos

# Índice

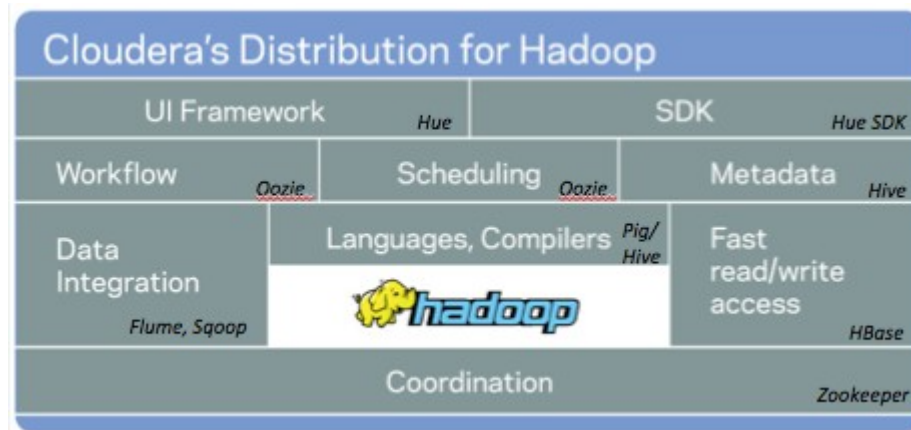
- Introducción
- Twitter
- Big Data
- Desarrollo de la aplicación
  - Recolección de datos
  - Almacenamiento
  - Procesamiento
  - Visualización
- Resultados
- Conclusiones
- **Líneas futuras**



# Líneas futuras

## Ampliar las estadísticas

- Más campos del JSON
- Estadísticas cualitativas





**Gracias**

