

INTELIGENCIA SOCIAL DE NEGOCIO CON NEO4J



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Autor: Andrés Ramos Pérez
Director: Francisco M. Rangel Pardo

AUTORITAS CONSULTING S.A.

ÍNDICE

1. Autoritas Consulting.
2. Cosmos: canales de información.
3. Objetivos del proyecto.
4. Tecnologías utilizadas.
5. Modelo de relaciones.
6. Atributos de los nodos.
7. Proceso ETL.
8. Tiempos de proceso.
9. Integración de NEO4J en Cosmos.
10. Suma de tecnologías.
11. Demo: Cosmos en vivo.

AUTORITAS CONSULTING

1. Recuperación de información de fuentes abiertas (Redes sociales, Web, Blog, Foros, Wikipedia...)
2. Clasificación, filtrado y limpieza de la información.
3. Procesamiento.
4. Visualización.
5. Inteligencia.
6. Ayuda en la toma de decisiones.

DATOS -> INFORMACIÓN -> CONOCIMIENTO -> INTELIGENCIA

Herramientas y
métodos

Interpretación
humana

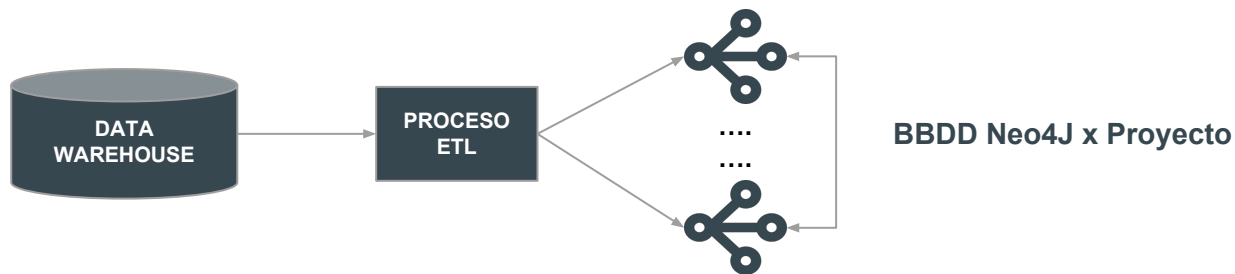
Aplicación
estratégica

COSMOS: CANALES DE INFORMACIÓN

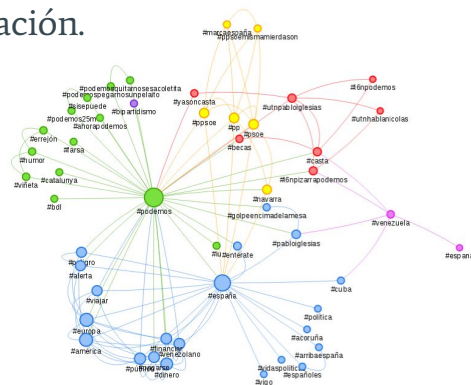
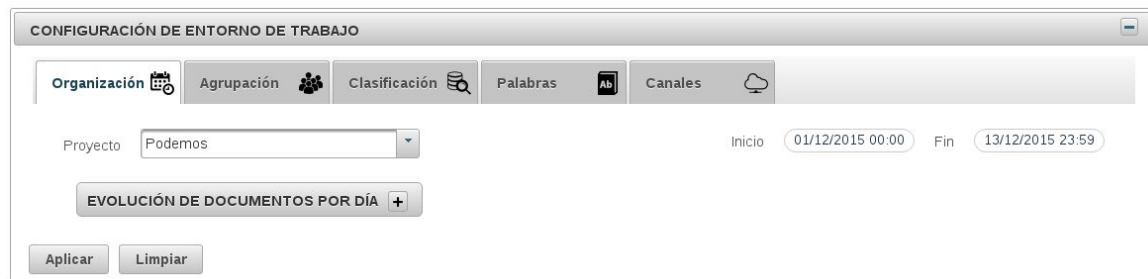


OBJETIVOS DEL PROYECTO

- Construir modelo de relaciones basado en la red social Twitter.



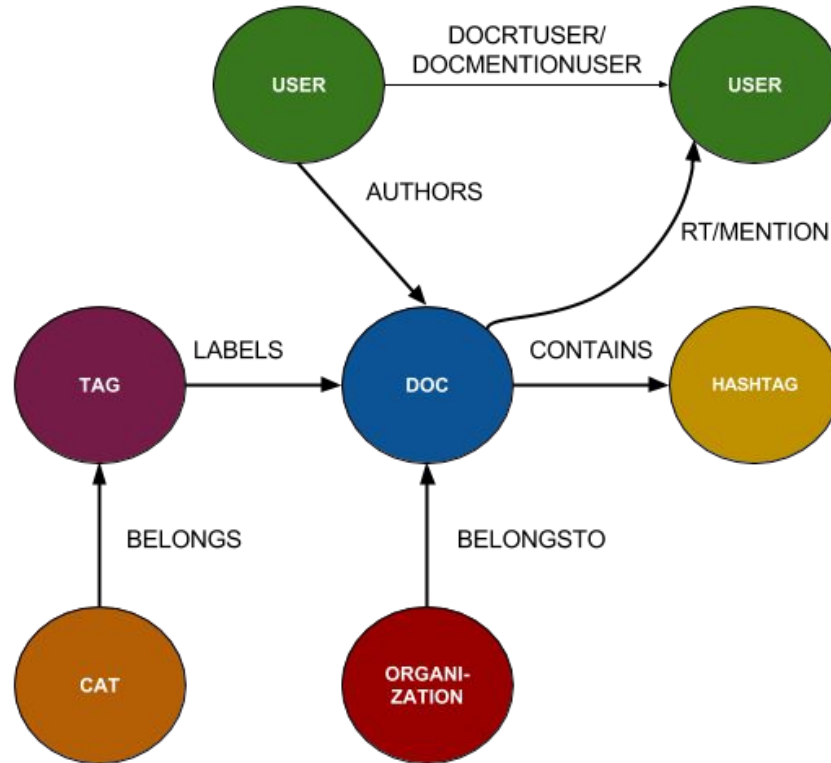
- Desarrollar una herramienta para la explotación del modelo y su visualización.



TECNOLOGÍAS UTILIZADAS

1. Proceso ETL → Java
2. BBDD de relaciones → NEO4J
3. Herramienta de explotación/visualización → Portlet Liferay (Java) + VisJS

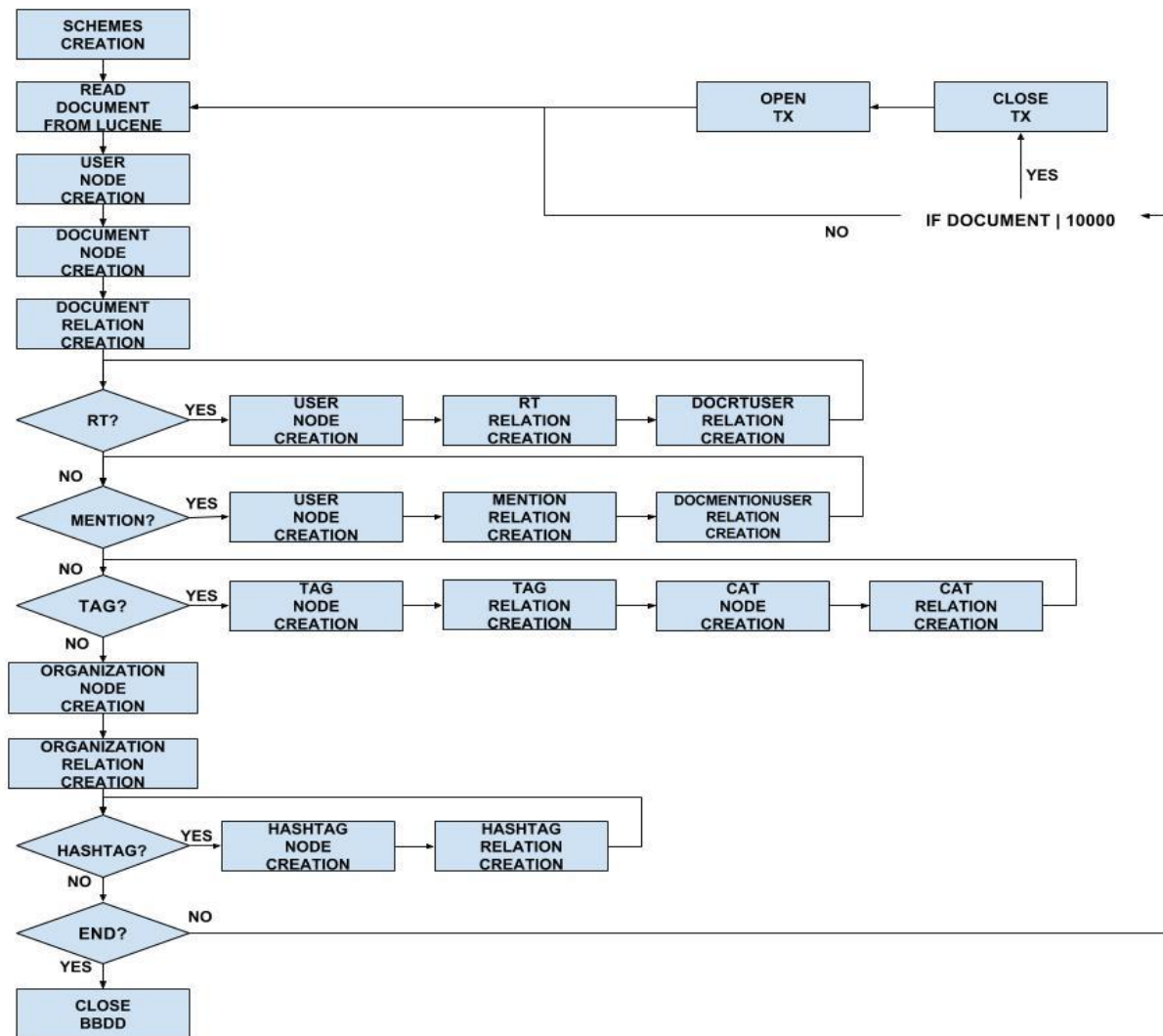
MODELO DE RELACIONES



ATRIBUTOS DE LOS NODOS

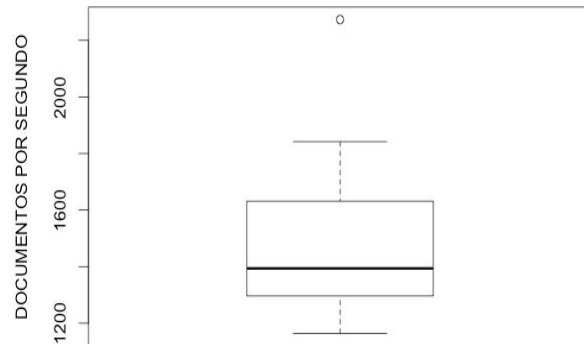
Nodo	Atributo	Significado
User	Antiquity	Antigüedad del usuario.
	Lists	Identificador de las listas del usuario.
	Friends	Usuarios a quien sigue el usuario.
	Followers	Usuarios que siguen al usuario.
	Gender	Sexo del usuario.
Doc	Rt	Indicador de si el documento es un retweet.
	Mention	Menciones del usuario a otros usuarios.
	Hashtag	Hashtag del documento.
	Question	Indicador de si el tweet es una pregunta.
	Exclamation	Indicador de si el tweet es una exclamación.
	Link	Links que contiene el documento.
	Datetime	Fecha en la que emitido el documento.
Organization	Subprjtype	Identificador del tipo de subproyecto (R,C ó I).
	Subprj	Nombre del subproyecto.
	Grouping	Agrupación de temas dentro de un subproyecto.
	Subgrouping	Subagrupación de temas dentro de un subproyecto.
Hashtag	<sin atributos>	
Tag	<sin atributos>	
Cat	<sin atributos>	

PROCESO ETL



TIEMPOS DE PROCESO

PROYECTO	Nº DOCUMENTOS	SEGUNDOS	DOCUMENTOS / SEGUNDO
Podemos	1.660.946	1.027	1.617,28
Elecciones Cataluña	6.288.008	5.000	1.257,60
Canal Panamá	602.842	518	1.163,79
Turismo Balear	174.420	106	1.645,47
Sector Energía	2.842.252	2.128	1.335,64
Sector Banca	429.168	233	1.841,92
Cumbre Mundial Comunicación Política	375.647	271	1.386,15
Cerveza Escudo	1.890.735	1.349	1.401,58
Vinos de España	606.998	267	2.273,40
TDAH	351.356	250	1.405,42
TV Social	4.406.864	3.669	1.201,11
Dilma Roussef	654.064	484	1.351,37



TIEMPO MEDIO DE PROCESO:

1.490 Documentos/segundo

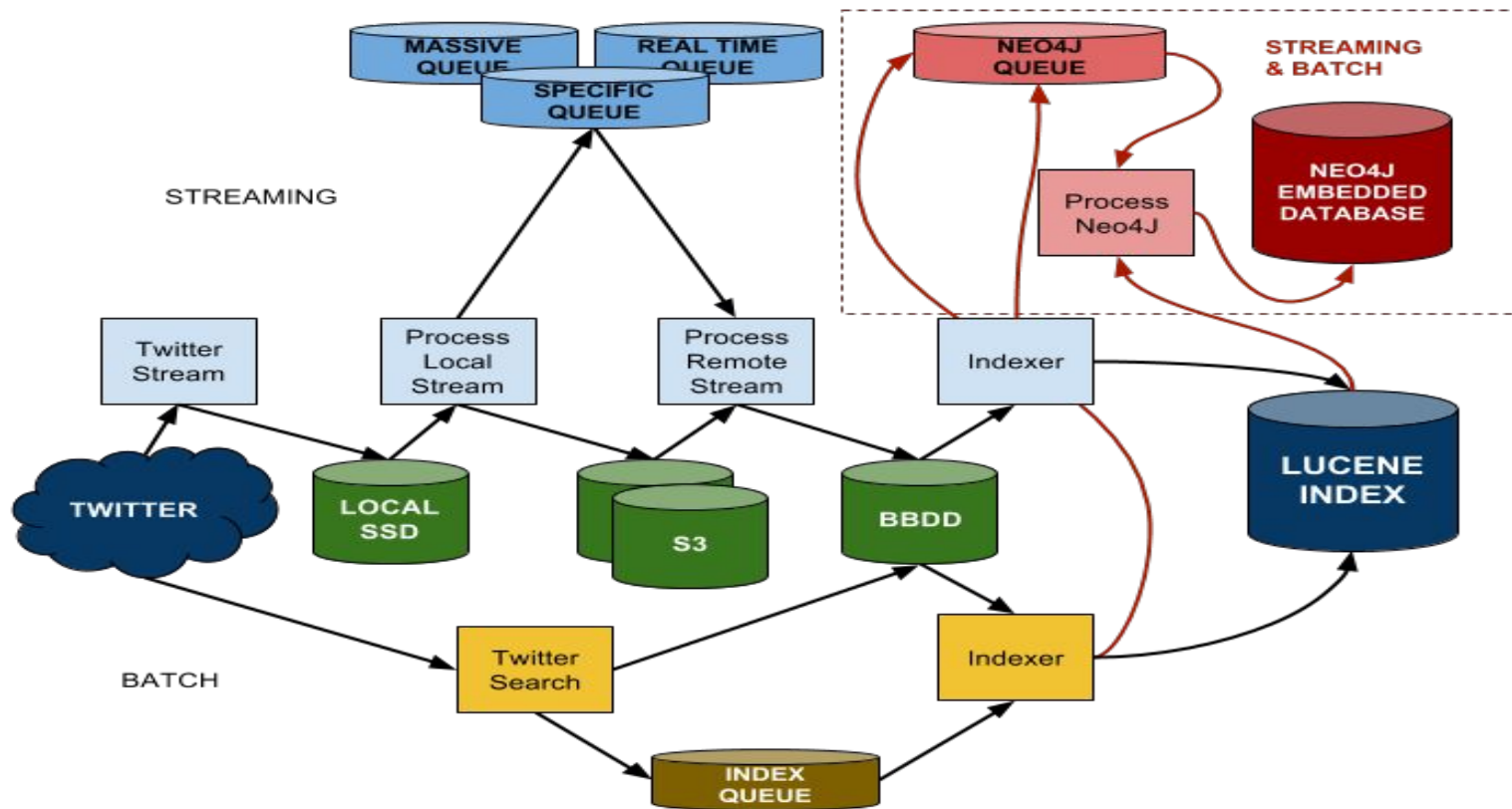
PICOS MÁXIMOS DE ENTRADA:

120 tuits/segundo (Elecciones, finales de fútbol, ...)

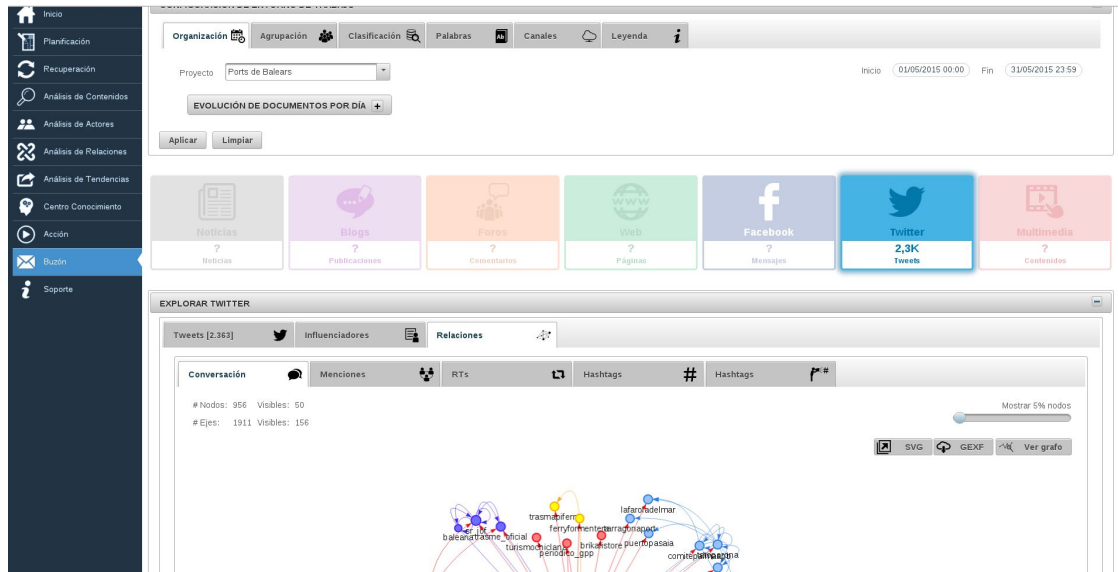


El proceso es óptimo.

INTEGRACIÓN DE NEO4J EN COSMOS



SUMA DE TECNOLOGÍAS

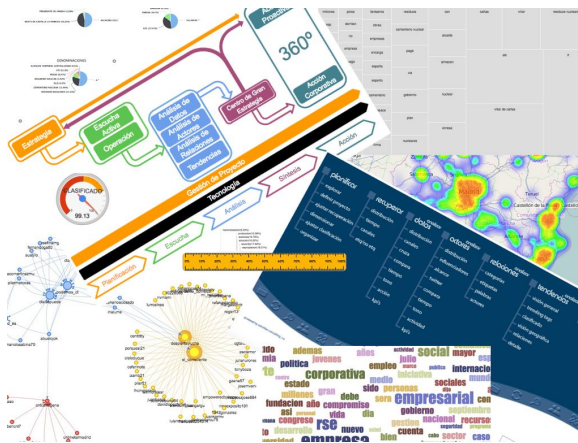


API



DEMO: COSMOS EN VIVO

La mejor muestra...



COSMOS INTELLIGENCE

DUDAS Y PREGUNTAS

