

Para evaluar el cambio en la opinión pública del electorado después del último debate presidencial entre Massa y Milei, utilizando datos recopilados de diversas plataformas sociales, propongo los siguientes pasos al momento de determinar la infraestructura mas conveniente, considerando un escaso presupuesto:

Paso 1: Definir los objetivos y la información necesaria para alcanzarlos

1. **Objetivos:**

- Análisis de sentimiento.
- Identificación de temas clave.
- Seguimiento de la popularidad de candidatos.

2. **Requerimientos de Datos:**

- Tweets relevantes al debate presidencial.
- Información sobre interacciones (likes, retweets, respuestas).
- Datos demográficos de usuarios.

Paso 2: Identificar Fuentes de Datos

1. **Twitter API:**

- Utilizar la API de Twitter para recopilar datos en tiempo real.
- Almacenar tweets en un sistema de almacenamiento de eventos como Apache Kafka.

Paso 3: Almacenamiento de Datos

1. **Kafka:**

- Usar Apache Kafka como canal de entrada para los tweets.
- Permite la ingestión en tiempo real y escalabilidad.

2. **Data Lake:**

- Almacenar datos brutos y procesados en un Data Lake (por ejemplo, AWS S3 o Azure Data Lake Storage).

Paso 4: Procesamiento de Datos

1. **Apache Flink/Spark:**

- Procesar datos en tiempo real para análisis de sentimiento.
- Identificar hashtags y menciones relevantes.

2. **Apache Hadoop (Hive):**

- Utilizar Hive para consultas ad-hoc y análisis batch.

- Estructurar datos para facilitar el análisis.

Paso 5: Almacenamiento de Metadatos

1. Metadata Store (por ejemplo, Apache Atlas):

- Gestionar metadatos para rastrear la procedencia de los datos.
- Facilitar la descubrimiento y comprensión de datos.

Paso 6: API y Consultas

1. GraphQL/REST API:

- Desarrollar una API para consultas y acceso a datos.
- Permitir a los analistas acceder a datos específicos según sus necesidades.

Paso 7: Análisis de Sentimiento y Temas Clave

1. Natural Language Processing (NLP) Library (NLTK, SpaCy):

- Integrar bibliotecas NLP para el análisis de sentimiento.
- Identificar temas clave usando técnicas de procesamiento de lenguaje natural.

Paso 8: Dashboard y Visualización

1. Power BI/Tableau:

- Crear paneles de visualización para análisis en tiempo real.
- Monitorizar la popularidad de candidatos, evolución de temas, etc.

A TENER EN CUENTA:

➤ Escalabilidad a Otras Redes Sociales

Instagram y Facebook API:

- Ampliar la arquitectura para integrar datos de Instagram y Facebook.
- Asegurar el cumplimiento de políticas de privacidad y términos de servicio.

➤ Monitoreo y Mantenimiento

ELK Stack (Elasticsearch, Logstash, Kibana):

- Implementar para monitorear el rendimiento del sistema.
- Alertas para problemas y capacidad de respuesta rápida.

➤ **Seguridad:**

- Encriptar datos en reposo y en tránsito.
- Implementar control de acceso y auditorías.

➤ **Evolución Continua:**

- Diseñar la arquitectura para permitir fácil escalabilidad y adición de nuevas fuentes de datos.