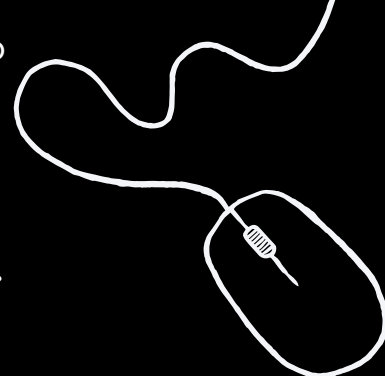
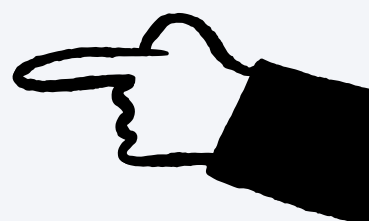


¿El dato es una semilla, sabes cómo se cosecha?



Identifica y diferencia la etapa de Adquisición de Datos (Cosecha) en el ciclo de vida de un proyecto de ciencia de datos.



El Dato: De Semilla a Conocimiento

En la ciencia de datos, todo comienza con la materia prima: el dato. Al igual que un agricultor necesita cosechar su cultivo, un científico de datos necesita adquirir o "cosechar" los datos. Esta primera etapa, crucial y a menudo la más desafiante, sienta las bases para todo el proyecto.



¿Dónde crecen nuestros datos?

La "tierra" donde encontramos los datos es variada y determina la técnica de recolección. Distinguir la fuente es el primer paso en la cosecha:

Datos

Estructurados

Organizados en un formato fijo (filas y columnas) que facilita la consulta y el análisis.

Bases de datos SQL (Transaccionales), Hojas de cálculo Excel/CSV.

Datos Semi-

estructurados

Tienen una estructura, pero no rígida como las tablas; usan etiquetas para organizar los elementos.

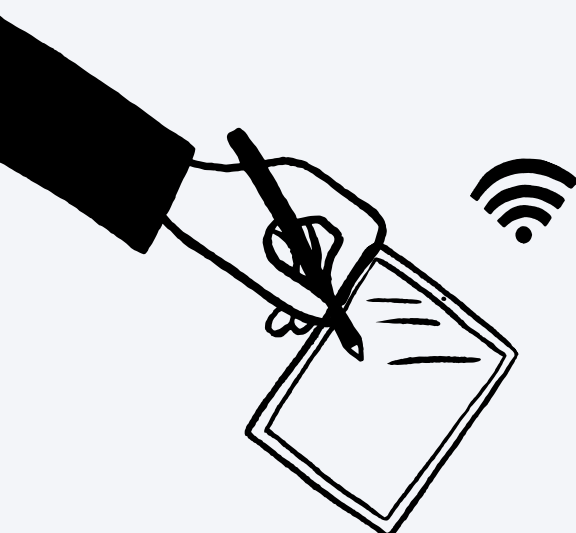
Documentos JSON o XML, Logs de aplicaciones web.

Datos No

Estructurados

Carecen de un formato predefinido y son difíciles de categorizar o analizar sin procesamiento previo.

Texto (emails, redes sociales), Imágenes, Videos o Audios.



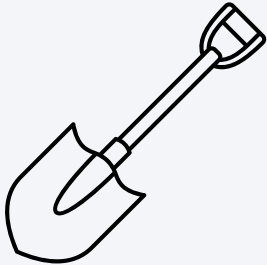
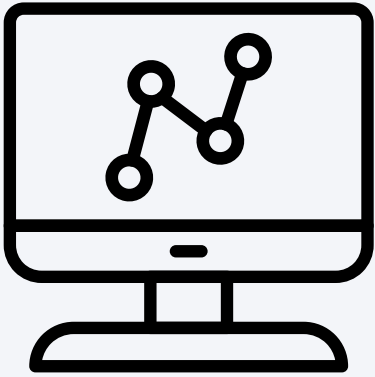
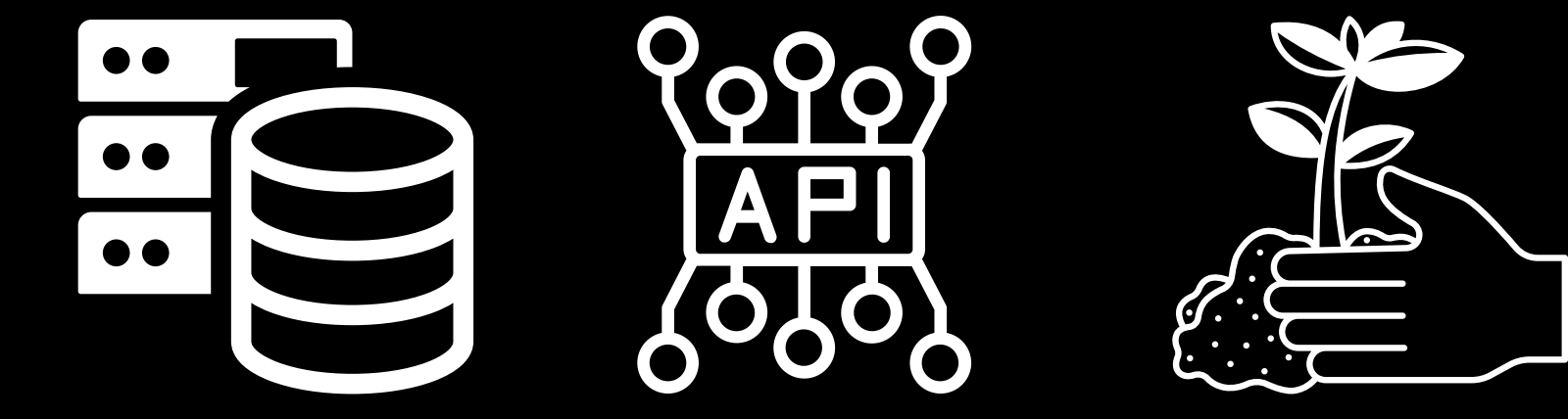
Cosecha de Datos: El Paso a Paso (Adquisición)

Esta es la etapa donde se extraen los datos de sus fuentes de origen y se preparan para ser transportados al lugar de procesamiento (almacén de datos).

Identificación y acceso

Saber qué y dónde: Se definen las fuentes de datos relevantes para el problema (BBDD, APIs, web, etc.) y se obtienen los permisos de acceso. ("El agricultor identifica la parcela y prepara sus herramientas")

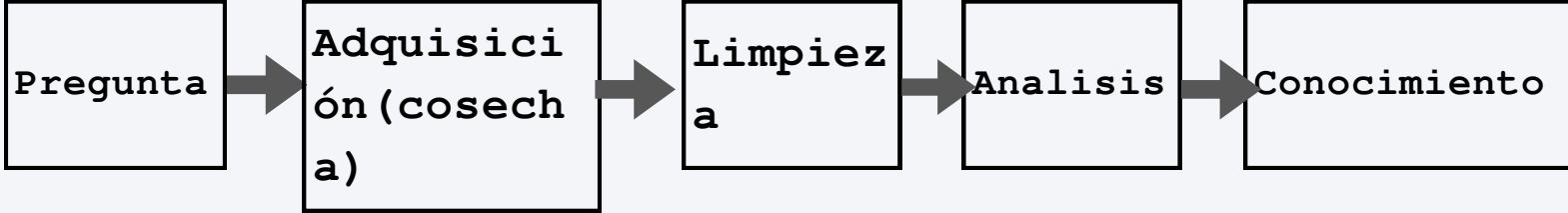
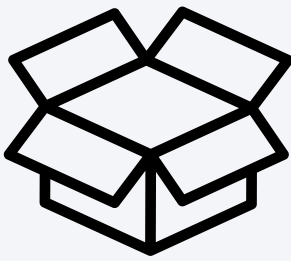




Extracción

Sacar el dato: Uso de queries (consultas SQL), scripts de web scraping o conectores API para jalar los datos del origen. Es el acto de "levantar la cosecha".

Almacenamiento Inicial
Guardar la cosecha bruta: Los datos se guardan temporalmente en un entorno seguro (Staging Area o Data Lake) en su formato original, antes de su limpieza.



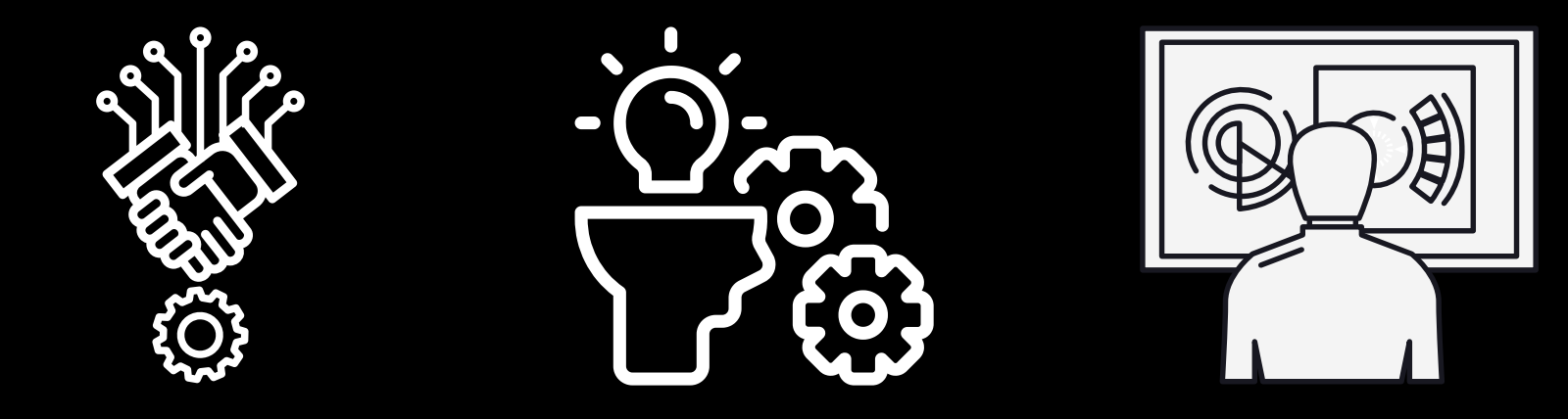
Adquisición vs. Limpieza

Aunque son consecutivas, es fundamental diferenciar la Adquisición (Cosecha) de la Limpieza (Preparación).



Características	Adquisición de Datos (Cosecha)	Limpieza de Datos (Preparación)
Propósito Principal	Obtener y reunir los datos de sus fuentes originales.	Mejorar la calidad del dato, corrigiendo errores, nulos o inconsistencias.
Analogía	La acción de arrancar el cultivo del suelo.	La acción de lavar, podar y clasificar el producto cosechado.
Resultado	Un conjunto de datos brutos (posiblement e sucios o duplicados).	Un conjunto de datos limpios, listos para el análisis.



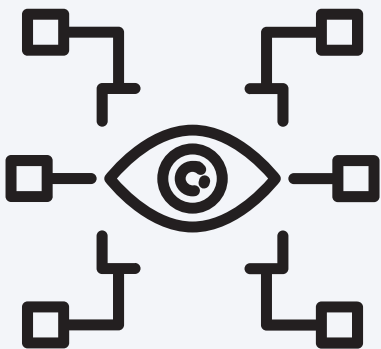


La etapa de Adquisición de Datos (Cosecha) no es solo un paso técnico; es una decisión estratégica que define el potencial y los límites de cualquier proyecto de ciencia de datos. Una cosecha de mala calidad (datos insuficientes o no relevantes) resultará en un "producto" (análisis) de mala calidad.



Reto

Imagina que eres un científico de datos que busca predecir el tráfico vehicular en tu ciudad.



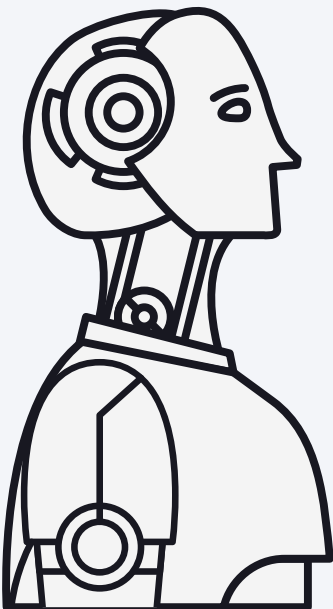
Pregunta: De las siguientes actividades, ¿cuál corresponde estrictamente a la etapa de Adquisición (Cosecha) y cuál a la de Limpieza?

- 1. Conectarse a la API de los semáforos inteligentes de la ciudad para descargar los datos de conteo vehicular de la última semana.
- 2. Rellenar los espacios en blanco (valores nulos) en la columna de velocidad donde los sensores fallaron.

Respuestas al final de la infografía.

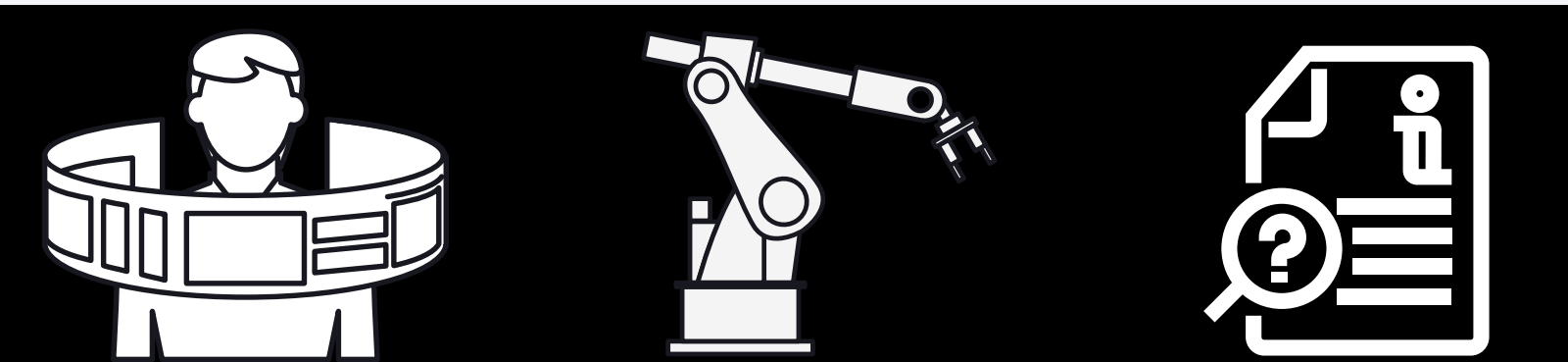
Para guiar esta aplicación, se proponen tres pasos sencillos y concretos que facilitan el inicio de un proyecto personal. Primero, Identificar un problema simple a resolver; segundo, Definir la Cosecha para determinar si el método de adquisición será manual (como una encuesta) o programático (como una API); y tercero, ¡Empezar a Extraer! utilizando herramientas accesibles.

Todo proyecto de ciencia de datos exitoso se cimienta en una cosecha intencional y bien planificada.



Respuestas

- 1. Adquisición (Cosecha): Es la acción de extraer los datos de la fuente.
- 2. Limpieza (Preparación): Es la acción de corregir o tratar la calidad de los datos.





Créditos

Este contenido se co-creó en el marco del proyecto

“Metodología para la co-creación de recursos digitales de micro aprendizaje activo, con el apoyo de IA Generativa”

liderado por las investigadoras:

Sandra Isabel Arango Vásquez

Grupo de investigación E-VIRTUAL

Bell Manrique Losada

Grupo de investigación ARKADIUS



Mariutsi Alexandra Osorio Sanabria

Grupo de investigación Automática,
Electrónica y Ciencias Computacionales



Con estudiantes de la Universidad de Medellín, la Institución Universitaria ITM
y la Institución Educativa San Juan Bautista La Salle.