|  |  |  |
| --- | --- | --- |
|  | Universidad de La Serena  Escuela de Ingeniería en Computación |  |

MEMÓRIA DE TÍTULO

INGENIERO EN COMPUTACIÓN

**LIITEC API: Sistema para recolección y procesamiento de datos medidos de sensores ambientales**

Alumno

ANTONY MAXIMILIANO RODRÍGUEZ CARHUACHIN

Profesor Guía

TOMÁS CARVAJAL ROJAS

Profesor Patrocinante

FRANCISCO LOPEZ

La Serena, 2023

Resumen

El presente trabajo tiene como objetivo…

Agradecimientos

Agradezco a …..

Tabla de contenido

[1. Introducción 5](#_Toc157812807)

[1.1. Motivación 6](#_Toc157812808)

[1.2. Meta 6](#_Toc157812809)

[1.3. Objetivo General 6](#_Toc157812810)

[1.4. Objetivos Específicos 7](#_Toc157812811)

[2. Marco Teórico 8](#_Toc157812812)

[2.1. Contexto Tecnológico y Fundamentos 8](#_Toc157812813)

[2.1.1. Recolección de Datos en Entornos Científicos 8](#_Toc157812814)

[2.1.2. Plataformas de Terceros y sus Limitaciones 9](#_Toc157812815)

[2.1.3. Limitaciones Comunes en Plataformas de Terceros 10](#_Toc157812816)

[2.1.4. Relevancia de las API en el Ámbito Tecnológico 11](#_Toc157812817)

[2.2. Arquitectura y Herramientas Utilizadas 13](#_Toc157812828)

[2.2.1. Arquitectura 13](#_Toc157812829)

[1. Bibliografía 15](#_Toc157812830)

# Introducción

En la era de la información y la tecnología, la capacidad de recolectar, gestionar y analizar datos de manera eficiente se ha convertido en un activo fundamental para la toma de decisiones informadas y el avance de la investigación científica. El Laboratorio de Investigación e Innovación Tecnológica para la Educación en Ciencias (LIITEC) se encuentra inmerso en este contexto, donde la obtención precisa y oportuna de datos desempeña un papel crucial en sus actividades de investigación y desarrollo.

En la búsqueda constante de mejorar sus procesos y optimizar recursos, el Laboratorio LIITEC ha dependido hasta la fecha de una plataforma de terceros para la recolección de datos. Si bien esta solución ha cumplido con su propósito básico, ha llegado el momento de considerar una alternativa que no solo permita reducir costos, sino que también ofrezca funcionalidades específicas y personalizadas para satisfacer las necesidades únicas de este laboratorio.

La presente tesis se centra en la concepción, desarrollo y aplicación de una API (Interfaz de Programación de Aplicaciones) diseñada a medida para el Laboratorio LIITEC. Esta API representa una herramienta tecnológica innovadora que tiene el potencial de revolucionar la forma en que se recopilan y manipulan los datos en el ámbito de la investigación científica. Al brindar al laboratorio su propia infraestructura de recolección de datos, se espera no solo reducir los costos asociados con la plataforma de terceros, sino también habilitar un control total sobre el proceso de adquisición de datos y la capacidad de adaptar la solución a las necesidades específicas de LIITEC……………………………………………………………………………………………………………

## Motivación

Este proyecto surge en el marco de la práctica I de la carrera de Ingeniería en Computación, durante una conversación con el supervisor de la práctica en la que se abordaba la fabricación de estaciones de medición de variables ambientales. En ese contexto, se indagó acerca del almacenamiento de los datos medidos por los sensores, revelándose el uso de ThingSpeak, un servicio de análisis de IoT que facilita la agregación, visualización y análisis de flujos de datos en tiempo real en la nube[[1]](#footnote-1).

Aunque hasta ahora esta plataforma ha demostrado eficiencia en el almacenamiento y visualización de datos, ha surgido una inquietud crucial respecto a su idoneidad a largo plazo para los objetivos del Laboratorio LIITEC. Al profundizar en la investigación, se ha encontrado respaldo en la literatura que destaca los riesgos asociados con depender exclusivamente de plataformas de terceros, en términos de flexibilidad y adaptabilidad, como se expone en el artículo "*Third-party apps: what are the risks?*” de bcs365 [1].

El artículo aborda de manera significativa los peligros vinculados a las aplicaciones de terceros, resaltando la preocupación por la seguridad cibernética, la posibilidad de fugas de datos, la complejidad de integración, y la falta de transparencia por parte de los proveedores. Específicamente, se destaca el riesgo de malware, la preocupación por fugas de datos, la complejidad en la integración, la falta de transparencia de los proveedores, y los problemas potenciales de seguridad y compatibilidad debido al abandono del soporte por parte de los desarrolladores de aplicaciones de terceros.

La motivación principal detrás de este proyecto reside en abordar los riesgos identificados y crear una solución interna para garantizar tanto el eficiente almacenamiento de los datos recolectados como un mayor control sobre su gestión y aplicación. En el entorno del Laboratorio LIITEC, donde la precisión y confiabilidad de los datos son esenciales para la investigación educativa en ciencias, la implementación de una API propia se visualiza como una oportunidad práctica y aplicativa de los conocimientos adquiridos durante la formación en Ingeniería en Computación. Asimismo, este proyecto ofrece la perspectiva de aportar de manera significativa al avance de la investigación educativa mediante la provisión de una herramienta personalizada que aborda los riesgos y desafíos previamente señalados en la literatura especializada.

## Meta

Optimizar la gestión de datos de los sensores ambientales en el Laboratorio de Investigación e Innovación Tecnológica para la Educación en Ciencias (LIITEC), facilitando su procesamiento y utilización efectiva en la toma de decisiones ambientales.

## Objetivo General

Desarrollar e implementar una API personalizada para la recolección de datos en el Laboratorio de Investigación e Innovación Tecnológica para la Educación en Ciencias (LIITEC), con el propósito de optimizar la eficiencia en el almacenamiento de datos y ofrecer mayor control sobre la gestión de la información. Esta API se diseñará con especial énfasis en la adaptabilidad a las necesidades específicas del laboratorio, buscando así proporcionar una solución técnica sólida y centrada en los requerimientos del usuario final.

## Objetivos Específicos

1. Identificar requisitos específicos del Laboratorio LIITEC para la recolección y gestión de datos.
2. Evaluar tecnologías para el desarrollo de una API adaptada a las necesidades del laboratorio.
3. Desarrollar una arquitectura de API escalable y segura.
4. Implementar un prototipo funcional con características clave para la recolección de datos.
5. Desarrollar módulo de integración para los sensores existentes.
6. Implementar medidas de seguridad para proteger la información recolectada.
7. Diseñar e implementar una interfaz de usuario intuitiva para la gestión de datos.
8. Documentar la API para facilitar la comprensión y el mantenimiento.
9. Realizar pruebas para validar la funcionalidad, seguridad y rendimiento de la API.

Marco Teórico

La presente sección constituye el cimiento conceptual y metodológico fundamental que permite una comprensión profunda del contexto en el que se enmarca este proyecto. Para facilitar la asimilación de la información, esta sección se estructura en dos componentes esenciales: el contexto tecnológico y metodológico, y las tecnologías específicas que desempeñan un papel clave en el desarrollo de la API.

## Contexto Tecnológico y Fundamentos

En esta sección del marco teórico, exploraremos el entorno tecnológico que rodea la recolección de datos en entornos científicos, examinando las plataformas de terceros y sus limitaciones, así como la relevancia crucial de las Interfaces de Programación de Aplicaciones (API) en el ámbito tecnológico.

### Recolección de Datos en Entornos Científicos

La recolección de datos en entornos científicos ha experimentado una transformación significativa gracias a los avances tecnológicos, como se evidencia en el artículo "*The technological advancements that enabled the age of big data in the environmental sciences: A history and future directions*" [2]. Este documento proporciona una visión detallada de la evolución de la recolección de datos en ciencias ambientales y destaca cómo la tecnología ha desempeñado un papel fundamental en este proceso. A continuación, se profundiza en los puntos clave presentados en el artículo:

1. Evolución de Instrumentación y Sensores:

Antes de los avances en tecnología, la recolección de datos en ciencias ambientales se basaba en evidencia visual y pruebas químicas poco específicas con límites de detección elevados. La limitada capacidad de recopilación de datos y la participación humana necesaria imponían restricciones significativas

Con el tiempo, la instrumentación mejorada, que incluye sensores ambientales, secuenciadores genéticos y otros instrumentos analíticos, ha permitido a los investigadores abordar preguntas ambientales complejas. La capacidad de tomar muestras en el campo, registrar ubicaciones mediante dispositivos móviles y analizar muestras con instrumentos analíticos avanzados ha aumentado significativamente.

1. Enfoque en Big Data en Ciencias Ambientales:

La llegada del término "Big Data" en 2005 marcó un cambio significativo en la forma en que se aborda la recolección y análisis de datos. Big Data se caracteriza por sus "cuatro V": volumen, velocidad, variedad y veracidad. Este enfoque se ha vuelto fundamental para gestionar conjuntos de datos ambientales cada vez más grandes y complejos.

1. Digitalización y Sensores Remotos:

Los avances en la digitalización de datos y la capacidad de realizar mediciones a través de sensores remotos han sido cruciales para la era de Big Data en ciencias ambientales. Desde cámaras montadas en aviones hasta sensores en satélites, UAVs y estaciones fijas, la accesibilidad a datos multiespectrales e hiperespectrales ha aumentado, permitiendo la monitorización de cambios en el uso del suelo, la identificación de fuentes históricas de contaminación y la evaluación de nutrientes en plantas

1. Mejora en Sensores Ambientales:

Los sensores ambientales han evolucionado desde dispositivos analógicos y basados en papel hasta instrumentos digitales que registran y transmiten datos de manera eficiente y precisa. Estos avances han permitido la toma de datos en intervalos específicos con mínima intervención humana, mejorando la eficiencia y reduciendo los costos de producción de sensores.

1. Biosensores y Análisis en el Campo:

La capacidad de realizar análisis ambientales que antes requerían equipos especializados en laboratorios centralizados se ha trasladado al campo mediante biosensores, reduciendo costos y aumentando la eficiencia en la toma de datos.

En resumen, la recolección de datos en entornos científicos ha evolucionado de manera significativa gracias a la instrumentación avanzada y los sensores mejorados. La capacidad de capturar grandes volúmenes de datos de manera eficiente y precisa ha allanado el camino para el enfoque de Big Data en ciencias ambientales, proporcionando nuevas oportunidades para comprender y abordar desafíos ambientales complejos.

### Plataformas de Terceros y sus Limitaciones

La recolección de datos de sensores ha experimentado una transformación significativa gracias a la proliferación de plataformas especializadas que facilitan la adquisición, almacenamiento y análisis de información proveniente de dispositivos sensoriales. Estas plataformas, que van más allá de ser simplemente herramientas, desempeñan un papel integral en la investigación científica, la monitorización ambiental y proyectos de Internet de las Cosas (IoT). A continuación, se presentan algunas plataformas destacadas junto con sus características y desafíos asociados.

1. ThingSpeak

Es un servicio de plataforma de análisis de IoT que tiene como objetivo agregar, visualizar y analizar flujos de datos en vivo en la nube. Esta solución de software está diseñada para sistemas IoT pequeños y medianos, así como para organizaciones y empresas más grandes. Según el proveedor, ThingSpeak es utilizado por profesionales de IoT, analistas de datos, ingenieros, investigadores y aquellos en el campo del monitoreo ambiental.

Limitaciones:

* Restricciones en el número de canales y campos en planes gratuitos.
* Funcionalidades avanzadas como la escalabilidad y soporte técnico requieren suscripciones pagas.[[2]](#footnote-2)

1. Ubidots

Es una plataforma de Internet de las cosas (IoT) que permite a las empresas crear aplicaciones IoT y convertir los datos de los sensores en conocimiento práctico y aplicable. Es una herramienta que facilita la transformación digital al proporcionar visualizaciones, alertas inteligentes, reportes programados, análisis de datos y más herramientas para tomar decisiones basadas en datos.

Limitaciones:

* Carece de un plan gratuito, sin embargo, ofrece una prueba de 30 días.
* Ofrece almacenamiento en la nube para los datos de los sensores, pero puede haber limitaciones en la capacidad de almacenamiento dependiendo del plan o suscripción utilizada.
* Limitaciones en la capacidad de escalar y manejar grandes volúmenes de datos.[[3]](#footnote-3)

1. Blynk

Es una plataforma de Internet de las cosas (IoT) que permite crear interfaces de control personalizadas para tus proyectos IoT. Puedes controlar y monitorear dispositivos de hardware desde tu teléfono móvil, guardar y visualizar datos, y realizar acciones basadas en eventos.

Limitaciones:

* Número limitado de widgets y proyectos, así como un límite en la cantidad de notificaciones y actualizaciones de datos por segundo en su capa gratuita.
* Es adecuado para proyectos individuales y pequeñas aplicaciones, puede enfrentar desafíos de escalabilidad en proyectos más grandes o con un gran número de dispositivos conectados.
* Utiliza servicios en la nube para la comunicación entre la aplicación y los dispositivos de hardware. Esto implica una dependencia de terceros y puede haber limitaciones o interrupciones en el servicio debido a problemas en la infraestructura de la nube.[[4]](#footnote-4)

1. Losant

Plataforma empresarial de Internet de las cosas (IoT) que proporciona una sólida base de software en el borde y en la nube. Destaca por su enfoque único y low-code, permitiendo a los desarrolladores crear y gestionar eficientemente aplicaciones IoT. Ofrece herramientas para la conexión, visualización y control de dispositivos, así como para el procesamiento y análisis de datos en tiempo real. La agilidad y velocidad de comercialización son elementos clave de esta solución integral para aplicaciones de IoT escalables y atractivas.

Limitaciones:

* Organizaciones y entornos aislados tienen límites en dispositivos y aplicaciones, con posibles costos adicionales para ajustes.
* Restricciones en la cantidad de datos que se pueden cargar, con diferenciación entre límites mensuales para organizaciones y sandboxes.
* Restricciones en velocidad para mensajes MQTT, MQTT Broker, eventos, webhooks y pulsaciones de botones virtuales.[[5]](#footnote-5)

### Limitaciones Comunes en Plataformas de Terceros

1. Restricciones en suscripciones:

Muchas plataformas de terceros ofrecen modelos de suscripción que limitan el acceso a características avanzadas y capacidades de almacenamiento, generando costos adicionales a medida que las necesidades de recolección de datos crecen.

1. Tipos de Licencia:

Las licencias pueden imponer restricciones en usuarios, dispositivos o datos, generando dificultades ante el crecimiento de actividades o una mayor participación de usuarios.

1. Dependencia de Infraestructura Externa:

La dependencia de la infraestructura y políticas de terceros puede resultar en interrupciones inesperadas en la recolección y gestión de datos.

1. Limitaciones en la Personalización:

La falta de flexibilidad en la personalización puede dificultar la adaptación a necesidades específicas del laboratorio.

1. Falta de Transparencia en Costos:

Ciertas plataformas pueden carecer de transparencia en cuanto a los costos de sus planes o los gastos vinculados al uso intensivo de recursos, lo que podría dar lugar a sorpresas financieras para los usuarios.

### Relevancia de las API en el Ámbito Tecnológico

En la presente era digital, las API se configuran como el componente conectivo esencial de nuestro entorno interconectado. La analogía pertinente sería considerar a una API como un puente que facilita la comunicación fluida entre distintos sistemas de software, permitiéndoles compartir información de manera eficiente. Este proceso posibilita la interacción sin inconvenientes entre el front-end, la interfaz de una aplicación o sitio web visible para el usuario, y el back-end, compuesto por la base de datos y el servidor, a través de diversas aplicaciones, sitios web o servicios. Esta armonía digital constituye la base para un flujo eficaz de datos y acciones en el ámbito digital.

La vertiginosa evolución tecnológica ha modificado sustancialmente la manera en que nos relacionamos con el mundo digital, abriendo nuevas perspectivas tanto para empresas como para individuos. En el epicentro de esta revolución digital, emerge una herramienta fundamental que, aunque pasa desapercibida para el usuario promedio, resulta esencial para el funcionamiento de prácticamente todos los servicios o aplicaciones en línea: las interfaces de programación de aplicaciones, conocidas como API.

#### Como funciona una API

Una forma sencilla de comprender cómo funcionan las API es observar un ejemplo común: el procesamiento de pagos de terceros. Cuando un usuario compra un producto en un sitio de comercio electrónico, es posible que se le solicite "Pagar con PayPal" u otro tipo de sistema de terceros. Esta función se basa en API para realizar la conexión.

* Cuando el comprador hace clic en el botón de pago, una API llama para recuperar información, también conocida como solicitud. Esta solicitud se procesa desde una aplicación al servidor web a través del Identificador uniforme de recursos (URI) de la API e incluye un verbo de solicitud, encabezados y, a veces, un cuerpo de solicitud.
* Después de recibir una solicitud válida de la página web del producto, la API realiza una llamada al programa externo o servidor web, en este caso, el sistema de pago de terceros.
* El servidor envía una respuesta a la API con la información solicitada.
* La API transfiere los datos a la aplicación solicitante inicial, aquí el sitio web del producto.

Si bien la transferencia de datos variará según el servicio web que se utilice, todas las solicitudes y respuestas se realizan a través de una API. No hay visibilidad en la interfaz de usuario, lo que significa que las API intercambian datos dentro de la computadora o aplicación y aparecen ante el usuario como una conexión perfecta.

#### Beneficios de las API

Las API simplifican el diseño y desarrollo de nuevas aplicaciones y servicios, y la integración y gestión de los existentes. Pero ofrecen otros beneficios importantes a los desarrolladores y a las organizaciones en general [3].

* Mejora de la colaboración

La empresa media utiliza casi 1.200 [4] aplicaciones en la nube (el enlace reside fuera de ibm.com), muchas de las cuales están desconectadas. Las API permiten la integración para que estas plataformas y aplicaciones puedan comunicarse entre sí sin problemas. Gracias a esta integración, las empresas pueden automatizar los flujos de trabajo y mejorar la colaboración en el lugar de trabajo. Sin las API, muchas empresas carecerían de conectividad, lo que provocaría silos de información que comprometerían la productividad y el rendimiento.

#### Innovación acelerada

Las API ofrecen flexibilidad, lo que permite a las empresas establecer conexiones con nuevos socios comerciales, ofrecer nuevos servicios a su mercado existente y, en última instancia, acceder a nuevos mercados que pueden generar retornos masivos e impulsar la transformación digital.

#### Monetización de datos

Muchas empresas optan por ofrecer API de forma gratuita, al menos inicialmente, para poder crear una audiencia de desarrolladores en torno a su marca y forjar relaciones con posibles socios comerciales. Si la API otorga acceso a activos digitales valiosos, la empresa   
los monetiza vendiendo el acceso. Esto se conoce como economía API.

#### Sistema de seguridad

Las API separan la aplicación solicitante de la infraestructura del servicio que responde y ofrecen capas de seguridad entre las dos mientras se comunican. Por ejemplo, las llamadas API normalmente requieren credenciales de autenticación; Los encabezados HTTP, las cookies o las cadenas de consulta pueden proporcionar seguridad adicional durante el intercambio de datos y una API Gateway [5] puede controlar el acceso para minimizar aún más las amenazas a la seguridad.

#### Seguridad y privacidad del usuario final

Al igual que las API ofrecen una protección añadida dentro de una red, también pueden proporcionar otra capa de protección a los usuarios personales. Cuando un sitio web solicita la ubicación de un usuario, que se proporciona a través de una API de ubicación, el usuario puede decidir si permite o deniega esta solicitud. Muchos navegadores web y sistemas operativos móviles, como iOS, incorporan estructuras de permisos cuando las API solicitan acceso a las aplicaciones y sus datos. Cuando la aplicación debe acceder a archivos a través de una API, los sistemas de archivos como Windows, Mac y Linux utilizan permisos para ese acceso.

La importancia fundamental de las API en la era digital va mucho más allá de la mera conectividad. Ofrecen numerosas ventajas, transformando la forma en que utilizamos la tecnología y aportando un inmenso valor tanto a los usuarios como a las empresas.

Las API ahorran tiempo, dinero y esfuerzo al reducir la necesidad de que los desarrolladores reinventen la rueda, pueden aprovechar las API existentes, evitando la necesidad de construir cada componente desde cero. Esto significa que las API permiten a las empresas centrarse en crear funciones únicas y mejorar las experiencias de los usuarios. Las API también impulsan la innovación al permitir la combinación de diferentes funcionalidades y fuentes de datos. Esta fusión tecnológica suele dar lugar a la aparición de nuevas aplicaciones y servicios que no habrían sido posibles sin estas herramientas interconectadas. Además, permiten la personalización y la adaptación a medida, lo que da a los usuarios un mayor control. Ya sea personalizando la interfaz de un smartphone o configurando sus aplicaciones favoritas, las API permiten a los usuarios moldear las experiencias digitales según sus preferencias, lo que las hace indispensables en el mundo actual, hiper personalizado y centrado en el cliente.



## Arquitectura y Herramientas Utilizadas

En este capítulo, se explorará en detalle la arquitectura subyacente y las herramientas tecnológicas que han sido fundamentales en el desarrollo de la API. La elección de una arquitectura sólida y la selección cuidadosa de herramientas desempeñan un papel crucial en el éxito y la eficiencia del sistema. A lo largo de este capítulo, se analizará la estructura general de la API, destacando los componentes esenciales y el modelo de datos subyacente. Además, se proporcionará una visión detallada de las herramientas tecnológicas utilizadas, desde el lenguaje de programación y la base de datos hasta los frameworks y estándares de seguridad implementados. Este análisis integral sentará las bases para comprender la solidez técnica y la capacidad adaptativa de la API en el contexto específico del Laboratorio LIITEC.

### Arquitectura

Al abordar la gestión de datos, surge un desafío significativo: cómo facilitar la transferencia eficiente de información entre distintos puntos. Este desafío se vuelve aún más crucial al trabajar con sensores, ya que se anticipa una generación sustancial de datos, dada la permanencia y constante actividad de estos dispositivos en su entorno laboral. Para afrontar esta complejidad y lograr una gestión óptima de la información recopilada, es esencial contar con una arquitectura sólida que ordene y facilite el tratamiento de estos datos.

En este contexto, una estrategia que cobra gran relevancia es la arquitectura ETL (Extract, Transform, Load), la cual se focaliza en extraer datos desde diversas fuentes, transformarlos según las necesidades específicas y cargarlos en un sistema de almacenamiento centralizado. Esta metodología no solo optimiza la eficiencia del proceso, sino que también posibilita una gestión más efectiva de grandes volúmenes de datos generados por sensores en su ambiente de trabajo continuo.

La implementación de una arquitectura ETL en este proyecto no solo resuelve la problemática del flujo de datos, sino que también abre la puerta a una gestión más estructurada y eficaz de la información generada por los sensores.

#### Justificación de la Elección de ETL

La elección de implementar la arquitectura ETL en el contexto de este proyecto se fundamenta en la necesidad imperante de gestionar eficientemente la ingente cantidad de datos generados por los sensores en el entorno laboral. Al adoptar la metodología ETL, se busca abordar específicamente los desafíos inherentes a la recolección, transformación y almacenamiento de esta abundante información, proporcionando una serie de beneficios clave:

* Eficiencia

Automatizar la extracción, transformación y carga de datos mejora significativamente el proceso, ahorrando tiempo y recursos esenciales al manejar grandes cantidades de datos de sensores de manera eficiente.

* Escalabilidad

Manejar volúmenes crecientes de datos de manera eficaz es esencial para adaptarse a la continua generación de datos por parte de los sensores en el laboratorio, asegurando que la solución sea flexible y sostenible a lo largo del tiempo.

* Flexibilidad

Extraer datos de diversas fuentes y cargarlos en sistemas diferentes facilita la integración efectiva de información variada. Esto es especialmente valioso al trabajar con sensores, permitiendo recopilar datos de múltiples fuentes para análisis integrales.

* Automatización

Programar y ejecutar automáticamente procesos ETL ahorra tiempo y recursos, permitiendo centrarse en análisis más significativos. La automatización es clave, especialmente en entornos donde la actualización de datos es constante.

* Mejora de la Calidad de Datos

Validar, limpiar y estandarizar datos durante la transformación contribuye significativamente a mejorar la calidad de los datos generados por sensores. Esta mejora es esencial para garantizar la precisión y confiabilidad requeridas en investigaciones educativas.

#### Fases del proceso ETL

En esta sección, exploraremos con mayor profundidad las distintas fases del proceso ETL (Ilustración 1) que conforman la columna vertebral de nuestra arquitectura para la gestión de datos provenientes de sensores.

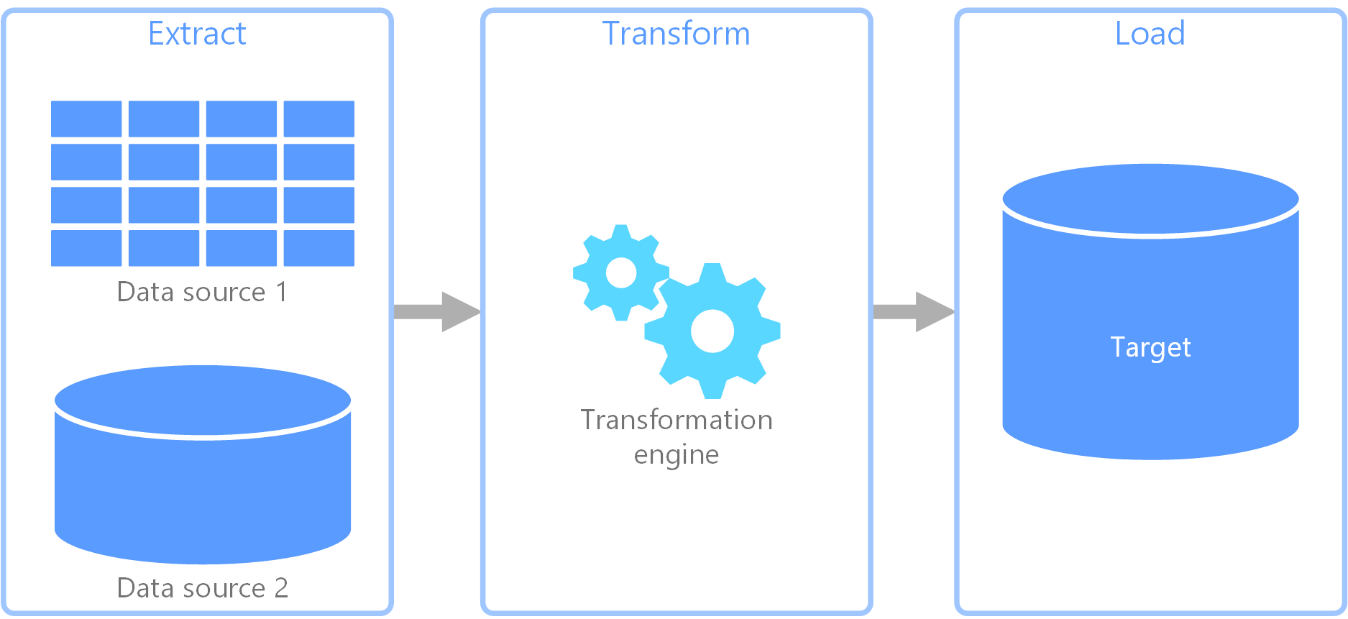


Ilustración 1- Representación gráfica del proceso ETL

1. Proceso de Extracción de datos

En el inicio, extraemos datos recopilando información en bruto de varias fuentes, como bases de datos, archivos planos o web-scraping. Estos datos pueden tener diferentes estructuras y provenir de diversas fuentes, como sistemas ERP o CRM, sitios web, tiendas de comercio electrónico, soluciones de marketing o hojas de cálculo Excel. La extracción de datos es esencial en ETL, ya que proporciona la materia prima y la información crítica para las etapas siguientes del proceso.

1. Proceso de Transformación de datos

Después de extraer los datos, el siguiente paso es la transformación. Este paso es crucial, ya que asegura que todos los datos se limpien y formateen de manera adecuada para el sistema de destino y su estructura de base de datos. La transformación adapta dinámicamente los datos extraídos al formato requerido, realizando acciones como eliminar duplicados, normalizar formatos, agregar información o aplicar cálculos específicos. Este proceso garantiza que los datos sean consistentes, precisos y libres de errores antes de cargarlos en la base de datos final del sistema empresarial.

1. Proceso de Carga de datos

El último paso es la carga de datos, donde los datos transformados se introducen en el sistema empresarial o la base de datos final. Esto puede ser un almacén de datos, un lago de datos o incluso una sencilla hoja de cálculo de Excel. Durante la carga, se estructuran los datos para facilitar su acceso, consulta y análisis. La base de datos de destino se optimiza para diferentes tipos de datos y usos finales, lo que puede implicar la creación de modelos específicos, particiones de datos o la implementación de índices.

#### Arquitectura IOT junto a proceso ETL

[Arquitectura de productos de plataforma de IoT en Google Cloud  |  Cloud Architecture Center](https://cloud.google.com/architecture/connected-devices/iot-platform-product-architecture?hl=es-419)

Para mañana.

# Bibliografía

[1 ]bcs365, "Third-party apps: what are the risks?". URL: <https://bcs365.com/insights/third-party-apps-what-are-the-risks>. Fecha de acceso: 11/12/2023.

[2] Carrie J., Samuel R., Bryn G. “The technological advancements that enabled the age of big data in the environmental sciences: A history and future directions”. URL: [The technological advancements that enabled the age of big data in the&nbsp;environmental sciences: A history and future directions (sciencedirectassets.com)](https://pdf.sciencedirectassets.com/315525/1-s2.0-S2468584420X00055/1-s2.0-S2468584420300489/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjELz%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIEN5M5%2FAXNz%2BMN6oExKxkwMjJ3r5mt7bRFbRB6Lf%2BWTzAiEAvPv%2FCIOZyL%2F6%2BjxMc2bPG6%2BjB79YZD1V%2BsnL5j%2F3GT4qswUINRAFGgwwNTkwMDM1NDY4NjUiDF5EhuJZz%2F82XmU1iyqQBdiRxMWWUSQ2gBFgUFh5DgiUU4TLhDV2tnGiMKyW09dMRqq17VunofgBpcgc8RXAloqEX6uUysCZlwxtXuDUrJLGOLuYwJJIir%2FOVMm0NxgxX7y8ocd0NIfDTsm2TLkEAn%2FwIe8bge7so91v6IqpaPTY%2BhO1hZdLWMs%2FrU%2BraZvyqQjPnDdPhykPSKi6ZKJ6KzO47I8MkTZZktU3iQiLOU8d114G26oOQ9oGbDH0iBNofsjk9mK4fXwkJohajKBCjIQ%2ByMBn1CBlwm%2FS9kyvA3%2BfV5MY56PZw%2Fl9nghos5suPD3F8XhKntshgO104iqsc3NuGpuhpy5bPH9jyxpAzFXCVRWW%2F86tsv0gm6PVaXrE%2B%2FoMF2MEvxi5qTSyPnK6ywO%2F%2B3khrYaSE4oSj5Wu54VNVEjJY4Gm2wDr9ck5zH3iUtttDI%2FgPXIisEErdANbcSgId0H4ULZbqHslMkTf1K49geFkAzcaV6uGpE9JtRDQ4k7%2Biz6gR44SimUCN7ib0t0QOzKcgpsCzOe0FFwSTBD%2Be8TsGQkquqLDAa5kJ4OD8XoXjaNc44c8KEcxGpwqnKsFfSGsgauaiNF%2BSiKW6MYMcr8oErmxHPFz2ukqM3OtPh3LD5Nt%2FqbwiCujRj5z19JRWl6IAqfwzuWSCGKlqcUAdfUzuwG8aKcNkNXQtejo%2FLROJMQHzQwiLivp3MD4OtJs5WnUGYQsQEHKKAwO7MP8orH7CQaK7aCHsRfJ6aYuLezuucozOjZCO1fXeR14VWJTUbLOC3HNWiXsufF1y2NOK6WPCdYsrc9tGRTtBH9WUlE7OMMx9Q%2BIBkaUix7GY33ZShJh%2BkrV7RNJTALTXyUraPcrFx7P8uGTKXjrDD8MMJHw4qsGOrEBZFkIXCzE%2FcAZfPXGMg0wouoMqvdzA6LxxYrOOlQZrqAuNbYubd28UmLylYiyFL2ebTifCpFm2%2B1V%2BDPEspmjENsEBof3Ls8novVRVpYzBi%2BuYReDjMoxw8oKbkWTmNkCtReKqEV8I30nwpVdqqpnusXKGeArIKNLQQiKMKEVGFlVDzIXyH0ehJtXpxOgOC31MnTEzMH2NDsXax%2Fmuu3bHHBw65PLgVzRp0KuuOuj2InT&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20231212T194949Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY3CRLZI62%2F20231212%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=347094df799c15506153b669323b511e486be5bbf77daa28f9fc24e0efaa2844&hash=b91fde2b03bfe40f82e8b710821b361e0441e440a5ec721ee1ec94e431045772&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S2468584420300489&tid=spdf-1eff6950-3308-497b-921b-66c1b5fb57de&sid=2fc2b725400f66418d786681ee22f310116cgxrqa&type=client&tsoh=d3d3LnNjaWVuY)

[3] [What is an Application Programming Interface (API)? | IBM](https://www.ibm.com/topics/api)

[4] [How APIs Can Transform Your Company (forbes.com)](https://www.forbes.com/sites/falonfatemi/2019/03/21/how-apis-can-transform-your-company/#bbbde7d668c2)

[5] [API Gateway - IBM API Connect](https://www.ibm.com/products/api-connect/api-gateway)

1. ThingSpeak, "ThingSpeak: Plataforma de Análisis IoT", MathWorks. URL: <https://la.mathworks.com/products/thingspeak.html> [↑](#footnote-ref-1)
2. ThingSpeak, “License Options”, URL: <https://thingspeak.com/prices> [↑](#footnote-ref-2)
3. Ubidots, “Plans Capacity”, URL: [https://es.ubidots.com/pricing - plans-capacity](https://es.ubidots.com/pricing#plans-capacity) [↑](#footnote-ref-3)
4. Blynk, “Pricing Plans”, URL: <https://blynk.io/pricing> [↑](#footnote-ref-4)
5. Losant, “Resource Limits”, URL: <https://docs.losant.com/organizations/resource-limits/> [↑](#footnote-ref-5)