

# Introducción a Pentaho

## Introducción

---

Pentaho Kettle es una poderosa herramienta de ETL (Extract, Transform, Load) diseñada para simplificar y optimizar el procesamiento de datos. Pentaho Kettle, también conocido como Pentaho Data Integration (PDI), es esencial para empresas que buscan gestionar grandes volúmenes de datos de manera eficiente y eficaz.

Pentaho Kettle se utiliza para extraer datos de diversas fuentes, transformarlos según las necesidades del negocio y cargarlos en un sistema de almacenamiento de datos. Esta herramienta permite manejar datos de múltiples formatos y orígenes, como bases de datos, archivos planos, servicios web, entre otros, facilitando la integración y limpieza de datos.

Entre los beneficios de utilizar Pentaho Kettle se encuentran su interfaz gráfica intuitiva, que permite diseñar flujos de trabajo de ETL sin necesidad de programación avanzada, y su capacidad para automatizar procesos repetitivos, reduciendo así el tiempo y esfuerzo requerido en el manejo de datos. Además, su flexibilidad y escalabilidad hacen posible su adaptación a las crecientes necesidades de datos de cualquier organización.

Las aplicaciones de Pentaho Kettle son diversas, abarcando desde la integración de datos para análisis y reportes hasta la migración de datos entre sistemas. Su uso es común en sectores como finanzas, salud, telecomunicaciones y retail, donde la calidad y precisión de los datos son críticas para la toma de decisiones estratégicas.

## Objetivos

---

El objetivo final de este curso es lograr que los alumnos sean capaces de desarrollar de forma autónoma procesos ETL a través de la herramienta Pentaho

## Dirigido a

---

Profesionales de las tecnologías de información, analistas de negocio o de sistemas, administradores de bases de datos, desarrolladores y cualquier profesional interesado en el mundo del BI.

## Temario

---

1. Instalación y configuración del entorno.
2. Procesos de datos y ETL
3. Extracción de datos de diferentes fuentes
  - a. Estructurados (txt, excel, zip, sql)
  - b. No estructurados (xml, json)
4. Flujos de datos y su combinación
5. Transformación de datos
  - a. Limpieza
  - b. Normalización
  - c. Agregación
  - d. Enriquecimiento
  - e. Validación
  - f. Fuzzy Match
6. Fusión de datos (join, unión)
7. Gestión de errores
8. Gestión de conexiones a fuentes internas/externas
9. Configuración de trabajos en PDI

## Duración y Desarrollo

---

10 horas teórico-prácticas

Del 23 al 24 de Septiembre de 9 a 14 horas

Modalidad virtual

## Condiciones

---

Curso enmarcado en el Digital Talent Hub. Gratuito para empresas socias de GAIA. Otra tipología de empresas pueden ponerse en contacto con [dth-academy@gaia.es](mailto:dth-academy@gaia.es).

**Cancelaciones:** Si cancelas tu inscripción con un margen mínimo de 4 días laborables previos al inicio del curso, no se aplicará ninguna penalización.

En caso de cancelar tu inscripción con un margen menor a 4 días laborables, se estudiará el % de penalización aplicable.

No informar, y/o no presentarse a la formación puede suponer un cargo de entre 150-300€.