# Objetivos del Taller

1. Entender qué es ETL
2. Entender qué es Pentaho Data Integration
3. Obtener materiales para capacitación
4. Instalar Pentaho Data Integration localmente
5. Cómo ingestar datos de en un lugar a otro

# 1. ¿Qué es ETL?

Extracción, transformación y carga (ETL) es una canalización de datos que se usa para recopilar datos de varios orígenes. A continuación, transforma los datos según las reglas de negocio y los carga en un almacén de datos de destino. El trabajo de transformación en ETL tiene lugar en un **motor especializado** y, a menudo, implica el uso de tablas de almacenamiento provisional para conservar los datos temporalmente a medida que estos se transforman y, finalmente, se cargan en su destino.

A diagram of a data flow

Description automatically generated

*Fuente:* [*https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl#extract-transform-load-etl-process*](https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl#extract-transform-load-etl-process)

## 2. Pentaho Data Integration

Más que un simple ETL (Extracción, Transformación, Carga), Pentaho Data Integration es una herramienta de ***orquestación*** de datos sin código que combina **diversos conjuntos** de datos en ***una única fuente*** de información como base para el análisis y la generación de informes. Se gestiona ***sin esfuerzo en una interfaz gráfica*** de arrastrar y soltar, de modo que puede realizar un seguimiento fácil de dónde provienen, hacia dónde van y cómo se transforman.

*Fuente: https://pentaho.com/products/pentaho-data-integration/*

# 3. Pentaho Learning Center – Lumada

**Plataforma Elearning:**

<https://learning.lumada.hitachivantara.com/>

**Curso Introductorio**: Pentaho Data Integration Fundamentals (PDI1000S)

<https://learning.lumada.hitachivantara.com/learn/course/pentaho-data-integration-fundamentals-pdi1000s>

**Learning Path Certif.:** Certified Specialist: Pentaho Data Integration Implementation (HCE-5920)

<https://learning.lumada.hitachivantara.com/learn/learning-path/certified-specialist-pentaho-data-integration-implementation-hce-5920>

<https://learning.lumada.hitachivantara.com/learning-paths/certified-specialist-pentaho-data-integration-implementation-hce-5920>

**Programa para certif**.: <https://www.hitachivantara.com/en-us/pdf/training/pentaho-data-integration-implementation-hce-5920-exam.pdf>

**Documentación de buenas prácticas:**

<https://support.pentaho.com/hc/en-us/articles/360000307943-Pentaho-Data-Integration#HTG>

# 4. Instalación de Pentaho Data Integration

## Pasos para la configuración en Windows

1. Descargar Java JDK 8

<https://www.oracle.com/es/java/technologies/javase/javase8u211-later-archive-downloads.html>

1. Descargar Pentaho Data Integration Developer Edition (Base Install)

A screenshot of a calendar

Description automatically generated

*Fuente: https://support.pentaho.com/hc/en-us/articles/205789159-Pentaho-Product-Lifecycle-Overview*

Repositorios:

<https://pentaho.com/pentaho-developer-edition/>

<https://github.com/ambientelivre/legacy-pentaho-ce>

1. Iniciar PDI desde spoon.bat

*Fuente:* [*https://www.hitachivantara.com/en-us/pdf/implementation-guide/three-steps-to-install-pentaho-data-integration-ce.pdf*](https://www.hitachivantara.com/en-us/pdf/implementation-guide/three-steps-to-install-pentaho-data-integration-ce.pdf)

# Extras

Dataset Supertienda en español

<https://community.tableau.com/sfc/servlet.shepherd/document/download/0694T000001ik1XQAQ?operationContext=S1>

Docker Desktop para Windows

<https://docs.docker.com/desktop/setup/install/windows-install/>

Dbeaver

<https://dbeaver.io/download/>

# Comandos

cd "C:\Users\rmonges\OneDrive - EXCELSIS S.A\Escritorio\Pentaho Data Integration - Taller 1\instaladores\pdi-ce-9.4.0.0-343\data-integration"

Pan.bat /file:"C:\Users\rmonges\OneDrive - EXCELSIS S.A\Escritorio\Pentaho Data Integration - Taller 1\xls\_a\_postgres.ktr" /level:Detailed