

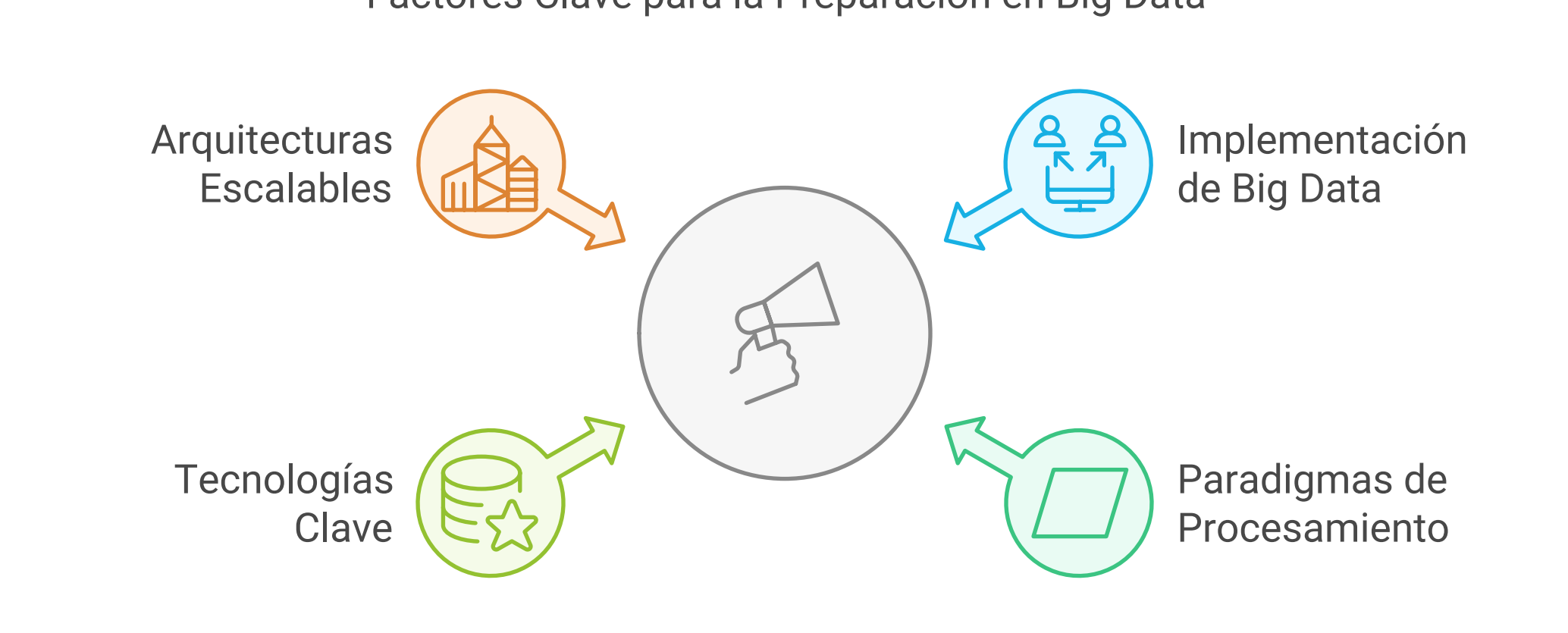
# 001 – BIG DATA

## INFORMACIÓN GENERAL

MÓDULO: **Big Data**  
PROFESORES: Rubén Juárez

## OBJETIVOS DEL MÓDULO

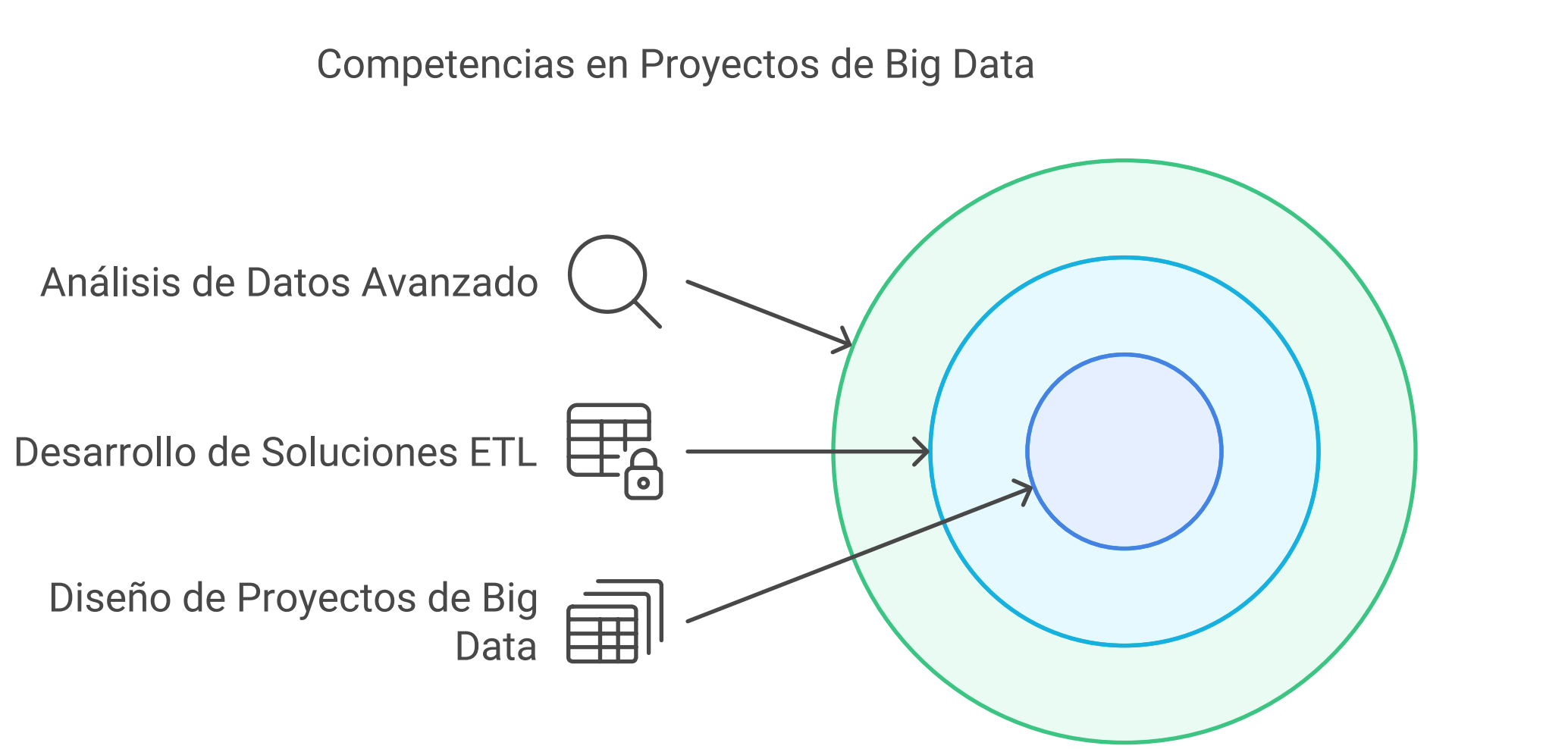
El alumno debe este módulo para afrontar con garantías las principales cuestiones asociadas a la implementación de Big Data en cualquier área de negocio, además de conocer en profundidad los diferentes paradigmas de procesamiento en sistemas Big Data y dominar las principales tecnologías y su utilización para el diseño de arquitecturas escalables adaptadas a cada proyecto.



## COMPETENCIAS DEL MÓDULO

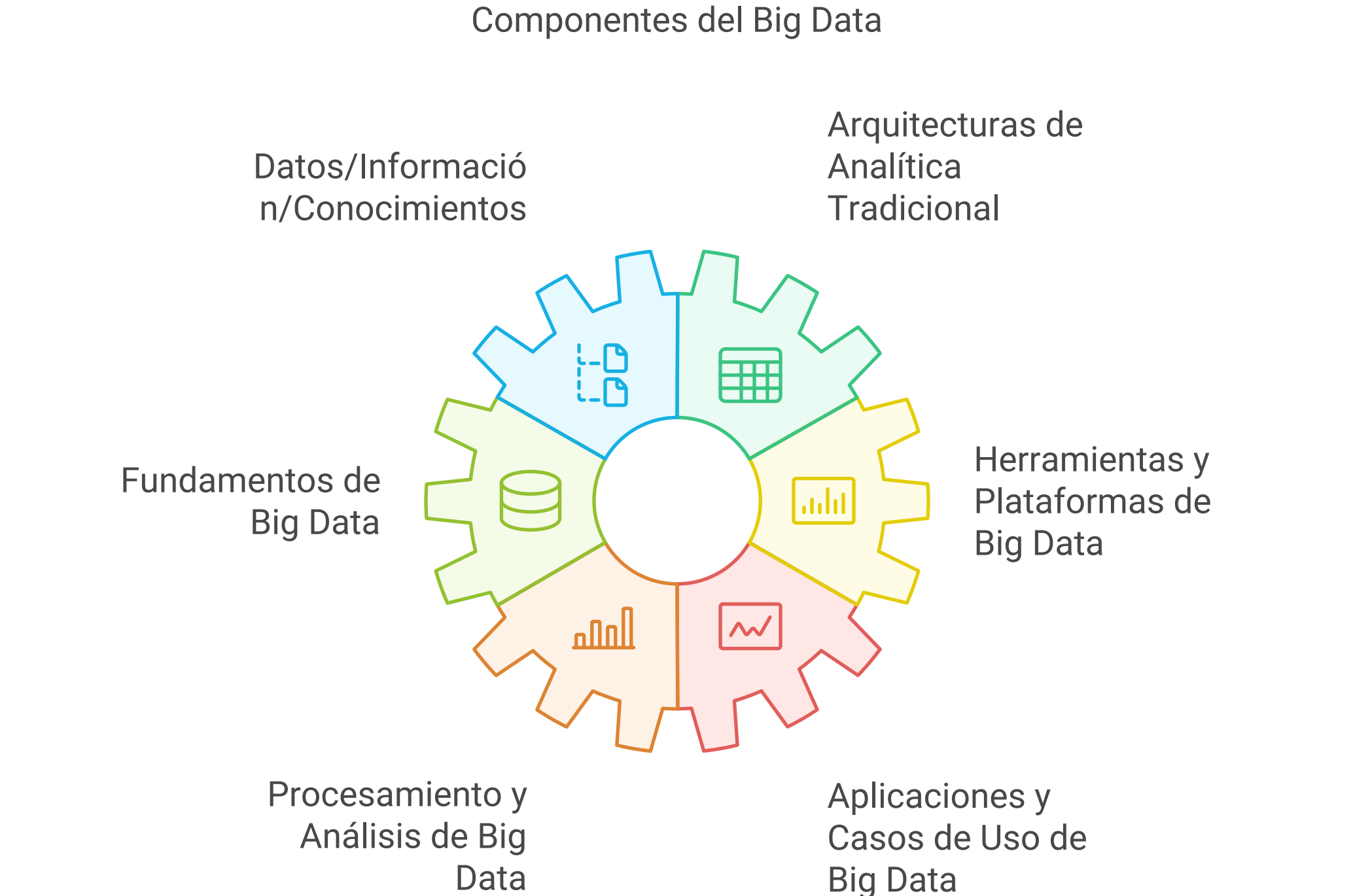
### HARD SKILLS

1. **Capacidad para analizar y diseñar proyectos de Big Data:** El alumno podrá analizar requerimientos, diseñar arquitecturas y planificar su implementación, tanto a nivel técnico como de negocio.
- Desarrollo de soluciones ETL y toma de decisiones basadas en datos:** El alumno adquirirá competencias para diseñar, desarrollar y optimizar procesos de extracción, transformación y carga de datos (ETL), así como para realizar análisis avanzados que permitan la toma de decisiones informadas en entornos reales.



## INDICE DEL CONTENIDO

- **Tema 1: Datos/Información/Conocimientos**
  - Diferencias y relaciones entre datos, información y conocimiento.
  - Tipos de datos y su importancia en Big Data.
- **Tema 2: Arquitecturas de analítica tradicional**
  - Enterprise Data Warehouse (EDW): Conceptos y estructura.
  - Proceso ETL: Fases y herramientas.
  - Cubos OLAP: Modelos y aplicaciones.
  - Informes y reports: Técnicas y mejores prácticas.
  - Cuadros de mando: Diseño y uso.
  - Master Data Management: Gestión de datos maestros y su importancia.
- **Tema 3: Fundamentos de Big Data**
  - Conceptos y características del Big Data.
  - Diferencias entre Big Data y analítica tradicional.
  - Principales desafíos del Big Data.
- **Tema 4: Herramientas y plataformas de Big Data**
  - Introducción a Hadoop: HDFS, MapReduce, YARN.
  - Introducción a Spark: RDD, DataFrame, SparkSQL.
  - Otras herramientas relevantes: NoSQL, Hive, Pig, HBase.
- **Tema 5: Procesamiento y análisis de Big Data**
  - Procesamiento por lotes vs. procesamiento en tiempo real.
  - Técnicas de análisis de Big Data: Clustering, clasificación, regresión.
  - Machine Learning y Big Data.
- **Tema 6: Aplicaciones y casos de uso de Big Data**
  - Aplicaciones en diferentes industrias: Salud, Finanzas, Retail, Telecomunicaciones.
  - Casos de éxito y lecciones aprendidas.
  - Futuro y tendencias en Big Data.



## PRÁCTICAS

- **Práctica 1:** Análisis y diseño de un proyecto de Big Data para un caso real de negocio.
  - Identificación de necesidades y requerimientos.
  - Diseño de la arquitectura del proyecto.
  - Justificación de la elección de tecnologías.
- **Práctica 2:** Implementación de un pipeline de datos utilizando herramientas de ETL.
  - Instalación y configuración de herramientas ETL.
  - Desarrollo de procesos de extracción, transformación y carga de datos.
  - Validación y optimización de los procesos ETL.
- **Práctica 3:** Desarrollo de un sistema de visualización de datos para la toma de decisiones empresariales.
  - Selección de herramientas de visualización.
  - Diseño de dashboards e informes interactivos.
  - Análisis de resultados y generación de insights.
- **Práctica 4:** Implementación de un Data Warehouse.
  - Diseño del esquema de un Data Warehouse.
  - Desarrollo de procesos ETL para la carga de datos en el Data Warehouse.
  - Creación de cubos OLAP y generación de informes.
- **Práctica 5:** Análisis de grandes volúmenes de datos utilizando Hadoop.
  - Configuración de un cluster Hadoop.
  - Procesamiento de datos con MapReduce.
  - Análisis de datos utilizando herramientas del ecosistema Hadoop (Hive, Pig).
- **Práctica 6:** Procesamiento de datos en tiempo real con Spark.
  - Configuración de Apache Spark.
  - Implementación de un proceso de streaming de datos.
  - Análisis de datos en tiempo real y visualización de resultados.
- **Práctica 7:** Desarrollo de un modelo de machine learning con datos masivos.
  - Preparación y limpieza de grandes conjuntos de datos.
  - Desarrollo de modelos de machine learning utilizando Spark MLlib.
  - Evaluación y optimización de modelos.
- **Práctica 8:** Estudio de casos de éxito de Big Data en la industria.
  - Selección y análisis de un caso de éxito de Big Data en una industria específica.
  - Presentación de los resultados y lecciones aprendidas.
  - Propuestas de mejora y adaptación del caso a otro contexto empresarial.

## TRABAJO FIN DE MÓDULO

**Título del Caso Práctico:** Implementación de un Sistema de Análisis Predictivo para la Optimización de Inventarios en una Cadena de Suministro

**El objetivo del caso práctico** es que los alumnos identifiquen, diseñen y desarrollen una solución de Big Data para optimizar la gestión de inventarios en una cadena de suministro con técnicas de análisis predictivo. Los estudiantes deberán aplicar los conceptos y herramientas aprendidos en el módulo para crear un sistema que permita prever la demanda y optimizar los niveles de inventario, reduciendo costos y mejorando la eficiencia operativa.

**Descripción del Problema:** Una empresa de retail con una amplia red de tiendas enfrenta problemas de sobrestock y desabastecimiento en sus almacenes, lo que genera altos costos de almacenamiento y pérdidas de ventas. La empresa busca implementar una solución de análisis predictivo que le permita anticipar la demanda de productos y ajustar sus niveles de inventario en tiempo real.

## HERRAMIENTAS

Indica herramientas o plataformas digitales que trabajan los alumnos en la parte práctica:

- **Apache Hadoop:** Framework para el almacenamiento y procesamiento distribuido de grandes volúmenes de datos.
- **Apache Spark:** Motor de procesamiento rápido y general para grandes volúmenes de datos.
- **Tableau:** Herramienta de visualización interactiva de datos.
- **Talend:** Plataforma de integración de datos que facilita los procesos ETL.
- **Python (Pandas, NumPy, Scikit-learn):** Lenguaje de programación y bibliotecas para análisis de datos y machine learning.

