

## TAREA 1 UD2

Investigar y analizar las diferentes arquitecturas y patrones de diseño utilizados en el procesamiento y almacenamiento de Big Data, identificando las ventajas, desventajas y casos de uso más comunes de cada uno. Los estudiantes o participantes deberán presentar un análisis comparativo y proponer la arquitectura y patrones más adecuados para un escenario práctico propuesto.

### Descripción de la Actividad:

1. **Formación de Equipos:** Dividir la clase en equipos de 3 a 5 personas. Cada equipo trabajará sobre diferentes enfoques y tecnologías.
2. **Investigación:**
  - Cada equipo debe investigar las principales arquitecturas utilizadas en Big Data. Algunas arquitecturas y patrones que deben investigarse incluyen:
    - **Lambda Architecture**
    - **Kappa Architecture**
    - **Microservicios orientados a Big Data**
    - **Data Lake vs Data Warehouse**
    - **Patrones de procesamiento por lotes (batch) y en tiempo real (streaming)**
    - **Arquitecturas en la nube vs. On-premises**
    - **Patrones de diseño: ETL (Extract, Transform, Load), ELT (Extract, Load, Transform), CQRS (Command Query Responsibility Segregation)**
3. **Análisis Comparativo:**
  - Los equipos deberán realizar un análisis detallado de al menos **tres** arquitecturas o patrones. Para cada uno de los seleccionados, deben abordar:
    - **Descripción técnica:** ¿Cómo funciona la arquitectura/patrón?
    - **Casos de uso:** ¿Cuándo y por qué debería utilizarse esta arquitectura o patrón?
    - **Ventajas:** ¿Qué beneficios ofrece en términos de rendimiento, escalabilidad, resiliencia y facilidad de implementación?

- **Desventajas:** ¿Cuáles son las limitaciones o desafíos que presenta?
- **Tecnologías asociadas:** ¿Qué herramientas o plataformas suelen implementarse con este patrón? (Ej. Hadoop, Spark, Kafka, AWS Redshift, Google BigQuery, etc.)

#### 4. Escenario Práctico:

- A cada equipo se le proporcionará un **escenario práctico** en el cual deberán seleccionar la arquitectura y patrón más adecuados. Ejemplos de escenarios prácticos podrían ser:
  - **Análisis de flujos de datos en tiempo real para una plataforma de IoT.**
  - **Almacenamiento y procesamiento de grandes volúmenes de datos transaccionales para una tienda en línea.**
  - **Integración de diferentes fuentes de datos no estructurados para una empresa de medios.**
- Los equipos deberán justificar su elección basada en la investigación previa.

#### 5. Presentación de Resultados:

- Cada equipo debe preparar una **presentación oral y un informe escrito** con los siguientes puntos:
  - Resumen de las arquitecturas y patrones investigados.
  - Comparación de los tres patrones seleccionados, con su análisis de ventajas y desventajas.
  - Solución propuesta para el escenario práctico.
  - Diagramas de arquitectura sugeridos.

#### Herramientas de Apoyo:

- **Documentación Técnica:** Proporcionar enlaces a documentación oficial de herramientas de Big Data (como Apache Hadoop, Apache Spark, Kafka, etc.).
- **Lecturas Recomendadas:**
  - “Designing Data-Intensive Applications” de Martin Kleppmann.

- “Big Data: Principles and best practices of scalable real-time data systems” de Nathan Marz y James Warren.

**Criterios de Evaluación:**

- **Profundidad de la Investigación:** Calidad y detalle en la descripción de las arquitecturas y patrones. (2,5 puntos)
- **Análisis Crítico:** Capacidad para identificar pros, contras y casos de uso adecuados. (2,5 puntos)
- **Propuesta del Escenario:** Justificación clara y adecuada de la arquitectura propuesta. (2,5 puntos)
- **Presentación:** Claridad y organización de la presentación oral y el informe escrito. (2,5 puntos)